

UNIVERSITÀ DEGLI STUDI DI MILANO – BICOCCA

Facoltà di Scienze Matematiche, Fisiche e Naturali

**SCUOLA DI DOTTORATO DI SCIENZE**

*Dottorato di Ricerca in Scienze Chimiche  
XXII Ciclo*



**Development of methodologies for molecular docking  
and their applications**

*PhD Thesis*

**Alessandro Di Domizio**

Tutor: Prof. Piercarlo FANTUCCI

Coordinator: Prof. Franca MORAZZONI

Anno Accademico 2008-2009

*A mia mamma, a mio papà.*



# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Drug discovery and development process . . . . .	5
1.2	Virtual High Throughput Screening (vHTS) . . . . .	8
1.3	Molecular Docking problem: overview . . . . .	9
1.3.1	Molecular docking approximations . . . . .	10
1.3.2	Instability of the problem . . . . .	11
1.4	Motivations, purposes and outline of the thesis . . . . .	12
1.5	In-home made molecular docking software . . . . .	14
1.5.1	Why C++ . . . . .	15
<b>2</b>	<b>Receptor preparation</b>	<b>17</b>
2.1	Add hydrogens tool . . . . .	18
2.1.1	Hydrogens placement . . . . .	19
2.1.2	Technical procedure . . . . .	21
2.2	Receptorial site finder . . . . .	24
2.2.1	Technical procedure . . . . .	25
2.3	Docking box generator . . . . .	25
2.3.1	Technical procedure . . . . .	26
2.4	Potential energy grids calculator . . . . .	27
2.4.1	Technical procedure . . . . .	28
<b>3</b>	<b>Ligand preparation</b>	<b>41</b>
3.1	Ligand preparation . . . . .	43
3.2	Ligand optimizer . . . . .	45
3.3	Technical procedure . . . . .	46

3.3.1	Data input . . . . .	46
3.3.2	Determination of the ligand connectivity . . . . .	46
3.3.3	Rotatable bonds detection . . . . .	47
3.3.4	Calculation of the ligand conformational energy and start of the <i>cooling cycle</i> . . . . .	48
3.3.5	FSA - Fast simulated annealing algorithm . . . . .	49
3.3.6	FSA - Metropolis criterion . . . . .	49
3.3.7	FSA - <i>Acceptance ratio</i> method . . . . .	51
3.3.8	Execution of all <i>runs</i> and <i>cycles</i> . . . . .	52
<b>4</b>	<b>Semi flexible molecular docking</b>	<b>55</b>
4.1	Semi flexible molecular docking . . . . .	55
4.1.1	Docking energy definition . . . . .	57
4.1.2	Scoring function . . . . .	58
4.2	Input files preparation . . . . .	59
4.3	Technical procedure . . . . .	61
4.3.1	Data input . . . . .	61
4.3.2	Determination of the ligand connectivity . . . . .	61
4.3.3	Rotatable bonds detection . . . . .	62
4.3.4	Calculation of the free ligand energy reference . . . . .	63
4.3.5	Start of the <i>run</i> . . . . .	64
4.3.6	Docking energy calculation . . . . .	66
4.3.7	FSA - Fast simulated annealing algorithm . . . . .	68
4.3.8	FSA - Metropolis criterion . . . . .	71
4.3.9	FSA - <i>Acceptance ratio</i> method . . . . .	72
4.3.10	Execution of all <i>runs</i> and <i>cycles</i> . . . . .	73
4.4	Tests and discussion . . . . .	74
<b>5</b>	<b>Applications</b>	<b>77</b>
5.1	RAS: study of new Ras protein inhibitors . . . . .	77
5.1.1	Introduction . . . . .	77
5.1.2	Methods and results . . . . .	79
5.2	VPE: ranking of a protein mutants library . . . . .	87
5.2.1	Introduction . . . . .	87
5.2.2	Methods . . . . .	88
5.2.3	Conclusions . . . . .	93

<i>CONTENTS</i>	3
<b>6 Conclusions</b>	<b>95</b>
<b>Publications</b>	<b>97</b>
<b>Bibliography</b>	<b>99</b>
<b>Acknowledgements</b>	<b>103</b>



# Chapter 1

## Introduction

The treatment of disease by drugs is very important in medicine and a wide variety of drugs which are active on almost all the organs of the body are now available. Drug discovery still continues to be an area to which scientists from around the world concentrate their efforts to find remedies for diseases that haven't yet a suitable cure.

### 1.1 Drug discovery and development process

If the underlying mechanism or cause of a disease is understood and if such a disease represents a significant unmet medical need in patients, then a research program aimed at better understanding the disease and finding an effective therapy can be developed by either a pharmaceutical industry or a research institution. A typical drug discovery and development process is made up of different steps, as shown in Fig. 1.1.

#### **Target identification and validation**

Target identification and validation encompasses a wide variety of scientific activities focused on identifying new target and confirming their role in diseases. As the knowledge about human biology grows, scientists are moving towards targeting cellular pathways of proteins rather than individual proteins.



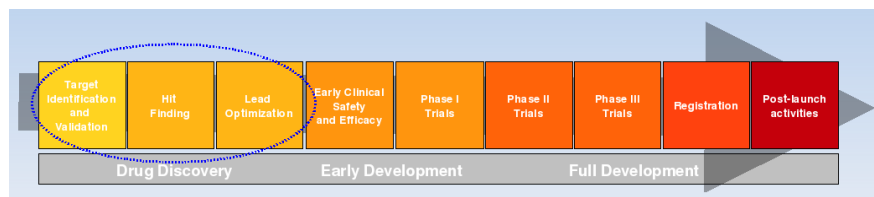


Figure 1.1: A typical drug discovery development process

### Hit finding

Hit finding entails development of robust assay to test small-molecule compounds in High Throughput Screenings (HTS). This stage entails development of ligands with affinities to chosen targets.

### Lead optimization

In lead optimization, small molecules ("hits") are chemically altered to improve their drug-like properties. At this stage ligands are modified to increase their affinity for their target.

### Early clinical safety and efficacy

To establish an initial safety profile of the drug, extensive toxicological and safety pharmacological profiles are done using *in silico*, *in vitro* and appropriate animal models.

### Phase I trials

In Phase I trials, the drug is tested in a small group of patients or healthy volunteers (e.g: 20-80) to evaluate its safety, determine a safe dosage range, and identify side effects.

### Phase II trials

In Phase II trials, the drug is given to a larger group of people (e.g: 100-300) to test its effectiveness, determine the effective dose range and to further evaluate its safety.

### Phase III trials

In Phase III trials, the drug is given to large groups of people (e.g: 1,000-3,000) to confirm its effectiveness, monitor side effects, compare it to commonly used treatments, and collect information that will allow the drug or treatment to be used safely.

### Registration

If quality, efficacy and safety of the drug are proved, a marketing authorization is granted by regulatory authorities. From then on, a new drug can be made commercially available to patients.

### Post-launch activities

Once a drug is on the market, adverse effects need to be constantly monitored and reported to the regulatory authorities.

Research and development of a drug is a long and very expensive process at high risk of failure. The pharmaceutical industry has been facing several challenges during the last years, and the optimization of their drug discovery pipeline is believed to be the only viable solution. High-throughput techniques do participate actively to this optimization, especially when complemented by computational approaches aiming at rationalizing the enormous amount of information that they can produce. *In silico* techniques, such as virtual screening or rational drug design, are now routinely used to guide drug discovery. Both heavily rely on the prediction of the molecular interaction ('*docking*') occurring between drug-like molecules and a therapeutically relevant target.

Thanks to the recent increase of performance computing systems, the first three steps of the drug development process have been supported by techniques that allow virtual experiments, with high saving of time and money. Computational methods are effectively applied to accelerate the process of lead identification and optimization: 'Computer-Aided Drug Design' (CADD) has an increasingly important role in simulating drug-receptor interactions, whose comprehension requires a deep understanding of biophysi-

cal and biochemical properties of both the ligand and the protein target at an atomic level.

## 1.2 Virtual High Throughput Screening (vHTS)

The 'chemical leads' are small potential drug like molecules which are capable of modulating the function of the target proteins that are further optimized to act as a therapeutic drug against a targeted disease.

HTS identifies lead molecules by performing individual biochemical assays with over millions of compounds. However, the huge cost and time consumed with this technology has led to the integration of cheaper and effective computational methodology, namely Virtual High Throughput Screening (vHTS).

vHTS is a computational screening method which is widely applied to screen *in silico* collection of compound libraries to check the binding affinity of the target receptor with the library compounds. This is usually achieved by using a scoring function which computes the complementarity of the target receptor with the compounds. HTS and vHTS are complementary methods and vHTS has been shown to reduce false positives in HTS [1]. Several vHTS strategies have been practiced and the technique is being continuously optimized for better performances. The successful vHTS requires the careful implementation of each phase of computational screening experiment right from target preparation to hit identification and lead optimization. Based on the availability of structural data, vHTS is carried out using receptor based screening methods, that involve usage of 3D-structure of the target receptors to search for potential candidate compounds that can modulate the target receptor function. Each of the database compounds is docked into the receptor binding site and the best fit is predicted. This is usually achieved by using a scoring function which computes the complementarity of the target receptor with the compounds.

### 1.3 Molecular Docking problem: overview

The 'docking problem' is concerned with the generation and evaluation of plausible structures of intermolecular complexes. Most docking algorithms are able to generate a large number of possible structures, and so they also require a means to score each structure to identify those of most interest.

The two main pieces of information to be obtained from a molecular docking simulation are the correct conformation of a ligand-receptor complex and a binding affinity prediction, expressed as docking energy ' $E_d$ '.

The docking energy represents an approximation of the binding free energy variation,  $\Delta G_{binding} \approx E_d$ , relevant to the complex-formation equilibrium:



starting from the free receptor (R) and the free ligand (L).

The docking problem involves many degrees of freedom: there are six degrees of translational and rotational freedom of one molecule relative to the other as well as the conformational degrees of freedom of each molecule.

In a real biological system, the system would include at least the ligand, the macromolecular receptor, and the solvent molecules. Because of the huge number of degrees of freedom associated with the solvent molecules they are normally excluded from the problem, or in special cases implicitly modeled in the scoring functions as a way to address the solvent effect. However, even the remaining part of the system, ligand and receptor, has a computational untreatable number of degrees of freedom, and therefore, the dimensionality of the problem has to be reduced through the application of different approximations, allowing the search space to be more effectively sampled.

All this is necessary to include a molecular docking algorithm in a software package for the screening (vHTS) of large ligand virtual libraries: the docking algorithm must supply a reliable solution in a short time. Supposing to screen a virtual library of  $10^6$  compounds, for each additional second needed by each molecular docking simulation an increase of about 11 days will occur on the CPU time.

### 1.3.1 Molecular docking approximations

The molecular docking problem can be faced by introducing a high number of approximations, in order to obtain a simple model able, at the same time, to represent the main properties of the real system. The combination of several expertises (in chemistry, physics, informatics, mathematics) is therefore needed to allow the development of advanced molecular docking softwares. The main approximations are:

- ◇ classical approach (force field)
- ◇ pairwise additive potential
- ◇ implicit solvent
- ◇ fixed bond lengths of both the receptor and the ligand
- ◇ fixed bond angles of both the receptor and the ligand

Furthermore, depending on whether the dihedral angles are held fixed or not, the following type of molecular docking can occur:

- ◇ **Rigid docking:** the easiest way to perform a molecular docking simulation is to hold fixed both the ligand and the macromolecule conformations. Such approximation is obviously quite strong. This is a good approximation for the ligand only if it has a small number of rotational degrees of freedom.
- ◇ **Semi-flexible docking:** current docking methods follow the assumption that protein structures are rigid entities and that it is the ligand that during the binding process changes its 3D-structure to find the best spatial and energetic fit to the proteins binding site.
- ◇ **Flexible docking:** during the course of the process, the ligand and the protein adjust their conformations to achieve an overall 'best-fit'. This approach is far from being applied because of the too high number of receptor degrees of freedom.

The main limitation in the use of 'rigid receptor' approximation is represented by the impossibility to take into consideration the 'induced fit', that is the conformational adjustment of the receptor due to the interaction with the ligand, resulting in the maximization (absolute value) of the total binding free energy variation.

### 1.3.2 Instability of the problem

As it is posed, the molecular docking suffers from the problem of instability and this is one of the reasons why the correlation between binding scores and experimentally determined binding affinities is far from trivial [12]. Because of the numerous approximations, the 'docking energy' can be affected by an error, the propagation of which can be dramatic for the calculation of the dissociation constant, if  $E_d$  is used as an approximation of  $\Delta G_{binding}$  (1.2):

$$K_d = \exp\left(\frac{E_d}{RT}\right) \quad (1.2)$$

where  $R = 1.987 * 10^{-3}$  [kcal mol<sup>-1</sup>K<sup>-1</sup>] and  $T$  /[K] is the fixed absolute temperature.

The 'percentage relative error' of the dissociation constant  $K_d$  due to a docking energy absolute error:  $\delta E_d = E_{d,i} - E_{d,TRUE}$  is given by the following equation (1.3):

$$K_d \%error \equiv \frac{K_{d,i} - K_{d,T}}{K_{d,T}} \cdot 100 = \left( \exp\left(\frac{\delta E_d}{RT}\right) - 1 \right) \cdot 100 \quad (1.3)$$

where:

- ◇  $K_{d,i}$  is the 'incorrect' dissociation constant calculated by an 'incorrect' docking energy  $E_{d,i}$
- ◇  $K_{d,T}$  is the 'True' dissociation constant calculated by the 'True' docking energy  $E_{d,T}$
- ◇  $\delta E_d = E_{d,i} - E_{d,T}$  is the absolute error on the docking energy

Assuming a temperature  $T = 298.0$  K, a plot of  $K_d \%error$  vs  $\delta E_d$  in the range  $-1.2 < \delta E_d$  / [kcal/mol]  $< +1.2$  is shown in Fig. 1.2.

The error propagation function is nonlinear (it is an exponential function) and the error on the  $E_d$  is highly amplified when  $K_d$  is calculated. Moreover, since this function is not symmetric with respect to the y-axis, nor with respect to the origin, the error propagation is different if  $\delta E_d \geq 0$ . In particular, it would be better underestimate (considering docking energy with its sign) than overestimate the docking energy, but this is not possible to be known in advance.

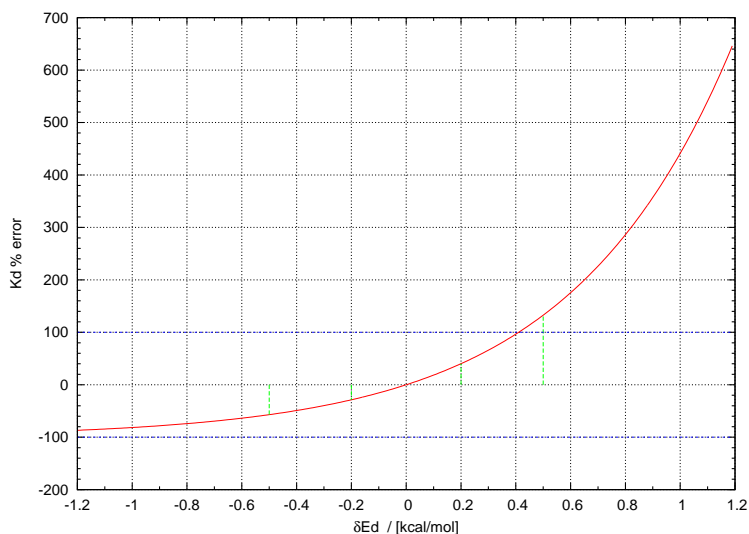


Figure 1.2: Docking energy absolute error propagation on Kd.

## 1.4 Motivations, purposes and outline of the thesis

More than 30 programs are currently available [2] and most of them are dedicated to virtual screening. The five most frequently cited ones represent 65% of the citations found in the literature: AutoDock (27%) [3], GOLD (15%) [9], FlexX (11%) [10], DOCK (6%) [27] and ICM (6%) [11].

AutoDock is by far the most cited implementation (see Fig. 1.3). It relies on a Lamarckian Genetic Algorithm (LGA) combined with a scoring function based on the AMBER force field [15], and is known for its robustness and accuracy [4]. It encompasses also a Monte Carlo simulated annealing and a traditional genetic algorithm. However, the last two are not as efficient and reliable as the LGA [2]. The program uses a five-term force field-based function that comprises a 12-6 Lennard-Jones dispersion term, a 12-10 hydrogen bonding term, a coulombic electrostatic potential, an entropic term and an intermolecular pairwise desolvation term. The scaling factor for each

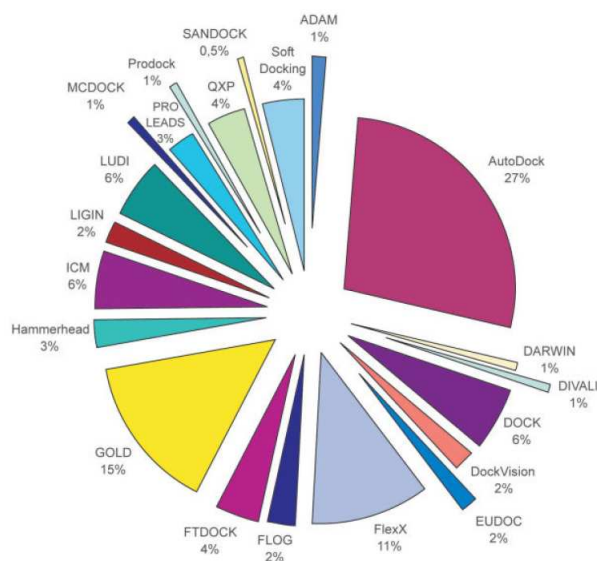


Figure 1.3: Docking software-number of citations for some of the most common docking programs, analyzed from ISI Web of Science (2005)[2]

of these five terms is empirically calibrated from a set of structurally known protein-ligand complexes. This flexible software is available for free for academic usage, and is thus often used to investigate new aspects of docking and implement new ideas [5] [6] [7] [8]; the good accuracy and high versatility shown by the program have promoted the widespread use of AutoDock, which explains the very high number of citations.

In the present work my aim was to develop a new molecular docking software ('Semi-flexible model') to be used for 'virtual High Throughput Screening' (vHTS). It would improve some of the main aspects of this type of softwares in current use. To do so my reference point was software AutoDock 3.0, where I introduced the following changes that will substantially modify the physical model of the problem:



- **Introduction of a modified scoring function** (see Section: 4.1)
- **Introduction of an univocal criterion** to calculate the ligand conformational variation energy to be used for the docking energy calculation (see Chapter: 3)
- **Implementation of a Fast Simulated Annealing (FSA) algorithm** including an *acceptance ratio* method (see Subsection: 4.3.7)

In order to obtain an efficient code to be included in a software package for the virtual screening of ligand libraries, I took care of many aspects to optimize both the algorithms and the code.

Once created, the molecular docking software can also be applied to many other fields different from the traditional computer aided drug-design.

It can, for example, be applied in the biotechnological field to design useful or valuable enzymes (Virtual Protein Engineering) in order to enable new (catalyzed) reaction pathways to occur, or to convert from some certain compounds into others (biotransformation). These products will be useful as chemicals, pharmaceuticals, fuel, food or agricultural additives (see Section: 5.2).

## 1.5 In-home made molecular docking software

The developed software is formed by the modules listed here below (see Fig.1.4):

- Add Hydrogens Tool (see Section: 2.1)
- Receptorial Site Finder (see Section: 2.2)
- Docking Box Generator (see Section: 2.3)
- Potential Energy Grids Calculator (see Section: 2.4)
- Ligand Optimizer (see Section: 3.2)
- Semi Flexible Molecular Docking (see Section: 4.1)

I also wrote some additional modules to execute an useful series of functions: an initial check of input PDB file containing the receptor structures; a module for 'combining docking and molecular dynamic simulations' in drug design [42]; a module to select the most representative binding-site conformation among a lot of conformations generated by the molecular dynamics simulations.

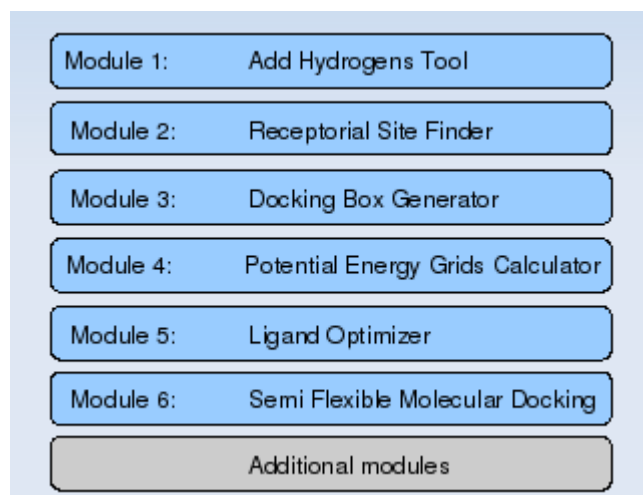


Figure 1.4: Developed molecular docking software modules

### 1.5.1 Why C++

All modules are written in C++ programming language that seems to be by far the language most commonly used for scientific programming. Because of its rich object-oriented features, C++ is rapidly becoming the programming language of choice for science and engineering applications. C++ is the object-oriented version of C that allows the use of the nice programming features of object-orientation. Features such as objects and classes introduce complex programming syntax. However, the enhanced feature set of C++ simplifies many programming tasks and naturally structures a program into logically independent units. C++ was chosen because it is a language that

gives great control over memory and it is a mature and stable language with many robust scientific libraries and tools. Most significantly, C++ provides nearly the same range of control of hardware resources through high level language constructs that is available through native machine instruction set. This range includes addressing and modifying the contents of individual memory locations and allocating and subsequently releasing the memory available to a program during execution [14].

## Chapter 2

# Receptor preparation

The structural information for the receptor are obtained by X-ray crystallography, NMR, or modelling techniques (see Figure 2.1). Since semi-flexible molecular docking algorithm is based on the ligand-receptor geometric complementarity, the 3D-structure of the receptor is of fundamental importance: the more accurate the physical description of this structure, the more relevant, accurate and useful the predicted binding mode.

The receptor structure should therefore be carefully checked regarding to two aspects. First, it should correspond to a biological conformation that is relevant to the targeted biological mechanism. For instance, the presence of crystal contacts in X-ray structures should be verified, as well as the impact of the presence/absence of other interacting partners such as cofactors. Second, the quality of the structure should be verified at an atomic level. For instance, the docking of a ligand is likely to fail if the region encompassing its native binding mode includes unresolved atoms; it also fails if the said region has a poor sequence identity with the template structure (if created by homology modeling), or encompasses flexible residues (reflected by a high B-factor if the structure has been determined by X-ray, or multiple conformations if determined by NMR). If such issues are identified, they have to be addressed during the preparation of the structures for the docking.

The structures are usually 'refined' before being used in the docking calculations with the aim to 'relax' the system towards a lower energy conformation, ideally the energy global minimum.

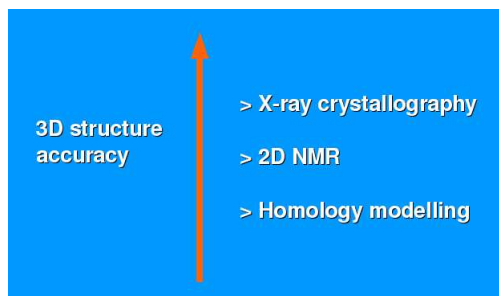


Figure 2.1: Receptor structure determination techniques

## 2.1 Add hydrogens tool

Hydrogen bonds play a major role in the stabilization of protein-ligand complexes. The ability of a functional group to form them depends on the position of its hydrogen atoms. An accurate knowledge of the positions of hydrogen atoms in proteins is therefore important to correctly identify hydrogen bonds and their properties.

The high mobility of hydrogen atoms introduces several degrees of freedom such as torsional changes, where the position of the hydrogen atom is rotated around the last heavy-atom bond in a residue, and protonation states, where the number of hydrogen atoms at a functional group may change.

X-ray crystallography cannot resolve hydrogen atoms in most protein crystals, so in most PDB files hydrogen atoms are absent. Sometimes hydrogens are added by modelling. Hydrogens are always present in PDB files resulting from NMR analysis, and usually present in theoretical models. In proteins, the average number of hydrogens per non-hydrogen atom, weighted to take into account the frequencies of amino acids, is 1.01. Thus, hydrogens are  $\approx 50\%$  of all atoms in protein. Nucleic acids have fewer,  $\approx 35\%$ .

The aim of this module is the fast automated placement of hydrogen atoms in the three-dimensional structure of a PDB file, containing either a protein or a RNA molecule, before structure-based calculations are conducted.

Molecular docking calculations highly depend on the hydrogen positions and a limit is that it is not possible to know in advance which will be the positions

$d_{C-H}$	1.02 Å
$d_{N-H}$	1.01 Å
$d_{O-H}$	0.96 Å
$d_{S_{Cys}-H}$	1.34 Å

Table 2.1: X-H bond distances.

the hydrogens will assume when the the ligand enters the binding site. The potential energy grid maps (see Section 2.4) will be built for a single structure to be decided *a priori*; consequently, a choice has to be made in order to establish a criterion to position the hydrogens.

### 2.1.1 Hydrogens placement

Hydrogens are placed according to the geometrical and chemical characteristics of the various amino acids that are recognized by the algorithm: bond lengths are established beforehand and depend on the atom to which the hydrogen is bound (see Tab.2.1). Hydrogen positions are determined by a mixed procedure: they are established in a unique way for hydrogens bound to  $N$  and  $C_{\alpha}$  backbone atoms (see Fig.2.3) and for hydrogens bound to planar fragments, such as the aromatic rings (Phe and Tyr), guanidinium (Arg) and amide (Gln and Asn) functional groups (see Fig.2.2). When free rotation is chemically allowed, as for hydroxyl (Tyr) and sulfhydryl (Cys) functional groups, there is an initial at random orientation that is optimized afterwards by means of a random search algorithm.

The user can choose whether to add polar hydrogens only ('united atom model'), or polar and apolar hydrogens. Beside this, there is the possibility to select the pH value according to which amino acid side chains are protonated. Hydrogens atoms are added according to the pH value chosen by the user and referring to the pKa values that are normally used in a protein pKa calculation (see Tab.2.2 [19]).

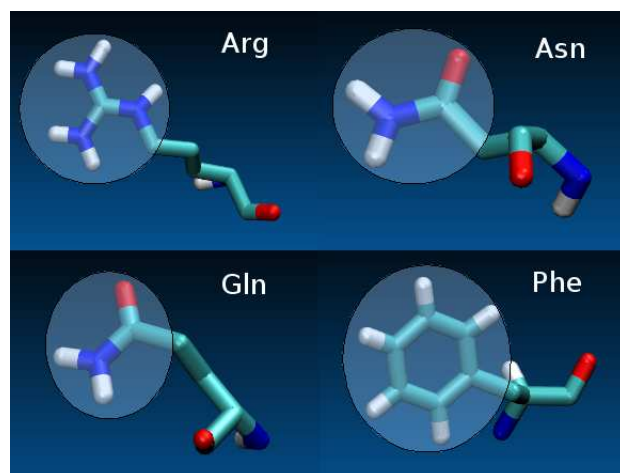


Figure 2.2: Arg, Asn, Gln and Phe side-chain planar fragments.

Amino acid	$\text{pK}_{\text{a}}_{\text{average}}$
Tyr	10.00
Cys	8.75
Asp	4.40
Glu	4.40
His	6.90
Lys	10.10
Arg	12.00

Table 2.2: Average pKa values of amino acid side chains.

### 2.1.2 Technical procedure

Here is a synthetic description of the procedure used by the algorithm to add hydrogens atoms.

#### Data input

- Acquisition of the macromolecular (protein or RNA) structure from a PDB file: storage of all macromolecular atomic coordinates
- Acquisition of the optimization cycles number ' $N_{\text{cycles}}$ '
- Acquisition of the hydrogens adding mode:
  - ◊  $flag = 0 \Leftrightarrow$  'polar and apolar' hydrogens
  - ◊  $flag = 1 \Leftrightarrow$  'only polar' hydrogens
- Acquisition of the pH value

#### Backbone hydrogens placement

- Calculation of the triangular matrix containing the macromolecular interatomic distances
- Calculation of the triangular matrix representing the macromolecular connectivity:
  - ◊  $flag = 0 \Leftrightarrow$  no covalent bond between the two atoms
  - ◊  $flag = 1 \Leftrightarrow$  presence of covalent bond between the two atoms
- Hydrogen atoms addition to  $C_{\alpha}$ :
  - ◊ the plane  $\Pi$  through the three points corresponding to:  $C_{\beta}$ ,  $N_{\text{backbone}}$  and  $C_{C=O, \text{backbone}}$  is determined
  - ◊ the normal vector for the plane  $\Pi$  is determined
  - ◊ an H atom is placed at a distance  $d = 1.02 \text{ \AA}$  from  $C_{\alpha}$
- Hydrogen atoms addition to  $N_{\text{backbone}}$ :
  - ◊ the N atom is considered  $sp^2$  hybridized



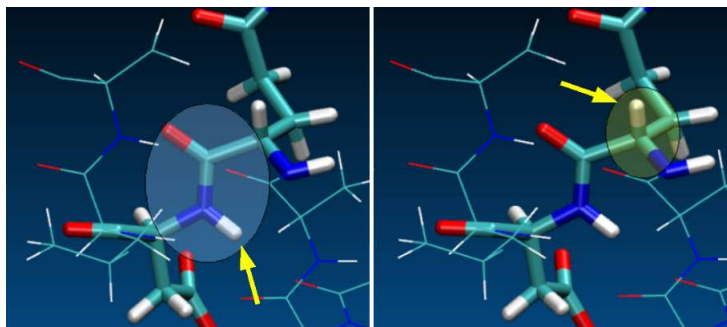


Figure 2.3: Hydrogens atoms on  $N_{backbone}$  and on  $C\alpha_{backbone}$ .

- ◇ the H atom is positioned to form a CNH angle of  $120^\circ$  (where C is  $C_{C=O,backbone}$ ), *trans* to  $O_{C=O,backbone}$ , and at a distance  $d = 1.01 \text{ \AA}$  from  $N_{backbone}$ .

### Side chain hydrogens placement

- Calculation of the new triangular matrix (including backbone hydrogens) containing the macromolecular interatomic distances
- Calculation of the new triangular matrix (including backbone hydrogens) representing the macromolecular connectivity:
  - ◇  $flag = 0 \Leftrightarrow$  no covalent bond between the two atoms
  - ◇  $flag = 1 \Leftrightarrow$  presence of covalent bond between the two atoms
- Addition of all 'polar' or 'polar and apolar' hydrogens, according to the input hydrogen adding mode
- Side chains are protonated according to the input pH value and referring to the pKa values listed in Tab.2.2 [19].
  - ◇ if  $pH < pKa \Rightarrow$  side chain is protonated
  - ◇ if  $pH \geq pKa \Rightarrow$  side chain remains in the deprotonated form

**Optimization of the orientations**

- Calculation of potential energy of the 'actual' structure (cycle 'i') of the macromolecule ' $E_i$ ':
  - ◊ 'van der Waals' and 'electrostatic' potential energy contributions are taken into account
  - ◊ a cutoff distance of 10.0 Å on interatomic distances is used
- At random generation of hydrogens new positions (where free rotation is allowed), always subject to appropriate geometric restraints:
  - ◊ bond distances are held constant
  - ◊ bond angles are held constant (e.g: COH angle = 104° for Ser, Thr and Tyr)
- Calculation of the new value (cycle 'i+1') of potential energy:  $E_{i+1}$
- Evaluation of the potential energy variation  $\Delta E = E_{i+1} - E_i$ :
  - ◊ if  $\Delta E < 0 \Rightarrow$  the new structure is accepted, and it becomes the actual structure
  - ◊ if  $\Delta E \geq 0 \Rightarrow$  the new structure is rejected
- The whole procedure is repeated up to reaching the optimization cycles number ' $N_{cycles}$ '

**Output print of the macromolecular structure with added hydrogens**

- The final structure of the macromolecule (see Fig.2.4) is printed on standard output in PDB file format.

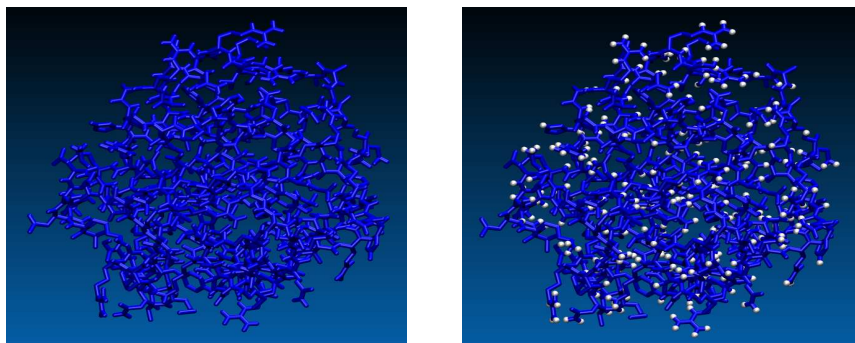


Figure 2.4: Protein structures before and after polar hydrogen atoms addition.

## 2.2 Receptorial site finder

The identification and visualization of protein cavities is the starting point for molecular docking simulations. Sites of activity in proteins usually lie in cavities, where the binding of a substrate typically serves as a mechanism for triggering some event, such as a chemical modification or conformational change. Consequently, binding sites are often targeted when attempting to interrupt molecular processes via therapeutics. Although binding site locations are often furnished by X-ray data or fold recognition, a tool that automatically predict these locations is however necessary to generate useful information to build the docking box (see Section: 2.3).

This module is a computational tool that uses geometry to characterize regions of buried volume in proteins and to identify positions likely to represent binding sites based upon the size, shape, and burial extent of these volumes.

As a modelling tool, 'Receptorial site finder' rapidly identifies favourable regions on the protein surface, simplifies visualization of residues modulating binding in these regions, and provides a means of directly visualizing buried volume, which is often inferred indirectly from curvature in a surface representation (see Figure: 2.5).

### 2.2.1 Technical procedure

To identify possible receptorial binding sites on the protein surface a purely geometrical procedure is used [18]. The algorithm is designed to fill the cavities in a protein structure with a set of spheres and to identify some such spheres as the most likely centres of the binding pockets.

An initial coating of probe spheres is calculated with the protein as substrate, then additional layers of probes are accreted onto the previously found probe spheres. Only probes with low solvent exposure are retained, and the routine finishes when an accretion layer produces no new buried probe spheres.

#### Output

This module produces the following output files:

- ◇ the structure of the macromolecule with probe spheres filling the putative binding site is printed on standard output (PDB file format)
- ◇ the estimated volume of each putative binding site is printed on standard output
- ◇ the geometric centre of each putative binding site is printed on standard output

## 2.3 Docking box generator

Once the receptorial site has been identified, this module delimits a space where docking calculations will be performed. A docking box is built to speed up subsequent calculation by drastically reducing the area on protein surface on which the ligand will be docked.

'Docking box generator' is an interactive module, which enables the user to set the 'grid spacing' (0.375 Å by default), to translate the docking box and to vary its dimensions (see Figure: 2.6).

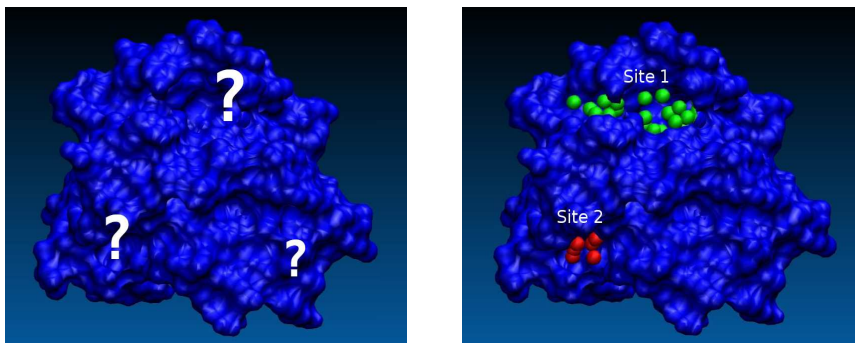


Figure 2.5: Possible binding sites on the surface of the receptor found by the module 'Receptorial site finder'

### 2.3.1 Technical procedure

The docking box is generated through a purely geometric approach: it is a parallelepiped centred in the geometric centre of the selected binding site, encompassing the whole volume of the same binding site. By means of a rototranslation, the docking box is then positioned so as to have a corner in the origin of the reference system and three edges along the x,y,z axes.

#### Output

The module 'Docking box generator' generates a file containing:

- ◇ the coordinates of the docking box centre
- ◇ the docking box dimensions
- ◇ the grid spacing
- ◇ the box versors components (they identify how the box is oriented in respect of the protein)

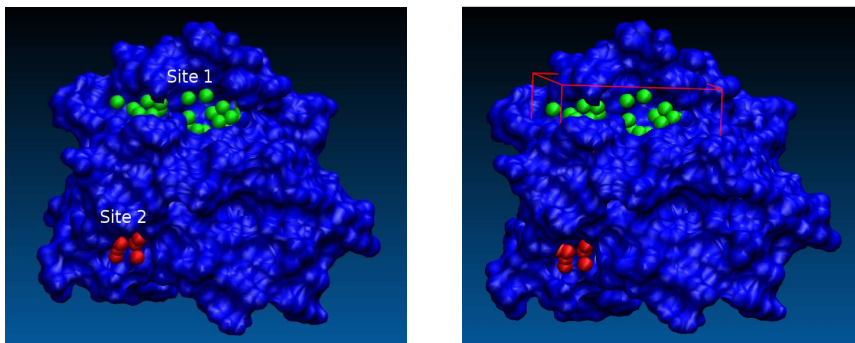


Figure 2.6: Docking box containing the first receptorial binding site.

## 2.4 Potential energy grids calculator

The aim of this module is to speed up the subsequent molecular docking simulation by simplifying the calculation of the interaction energy with the macromolecule for each generated pose of the ligand. The approach consists of generating pre-calculated potential energy grid maps, one for each possible atom-type present in a ligand being docked. This helps to make the docking calculations extremely fast, because for each generated pose the calculation of the interaction energy as a sum of the pairwise interaction energies between each ligand atom and all protein atoms is avoided.

A grid map consists of a three dimensional lattice of regularly spaced points, inside the docking box generated by the previous module (see Section 2.3). Default grid points spacing is  $0.375 \text{ \AA}$  (roughly a quarter of the length of a carbon-carbon single bond), but, if necessary, the user has the faculty to vary this value. Each point within the grid map stores the potential energy of a 'probe' atom that is due to all the atoms in the macromolecule. The interaction potential energy for the complex system (*many-body system*) is approximated by the sum of two-body contributions (*pairwise model*).

The probe's energy at each grid point is determined by the set of parameters supplied for that particular atom-type, and is the summation over all atoms of the macromolecule of all pairwise interactions.

Interaction energy is approximated by means of four main energy contributions (AMBER force field [15]):

- electrostatic
- van der Waals
- desolvation
- hydrogen-bonding

A single grid map is to be generated, for example, for the electrostatic (Coulomb) contribution: the potential energy value obtained by summing the interaction between each atom of the macromolecule and a probe atom with unit positive charge is associated to each point of the grid.

For the van der Waals contribution, instead, are generated so many maps as are the possible atomic species that can be present in a ligand. Van der Waals parameters, infact, strictly depend on the interacting couple of atoms and therefore all possible cases are to be considered beforehand.

As detailed in the next section, altogether 18 grid maps are generated:

- 1 grid map (electrostatic contribution)
- 11 grid maps (van der Waals contribution)
- 2 grid maps (desolvation contribution)
- 4 grid maps (hydrogen-bonding contribution)

### 2.4.1 Technical procedure

Here is the description of the procedure used by the algorithm to generate the potential grid maps.

#### Data input

- Acquisition of the coordinates of the docking box center ( $X_C, Y_C, Z_C$ )
- Acquisition of the box versors (they identify how the box is oriented in respect to the protein)

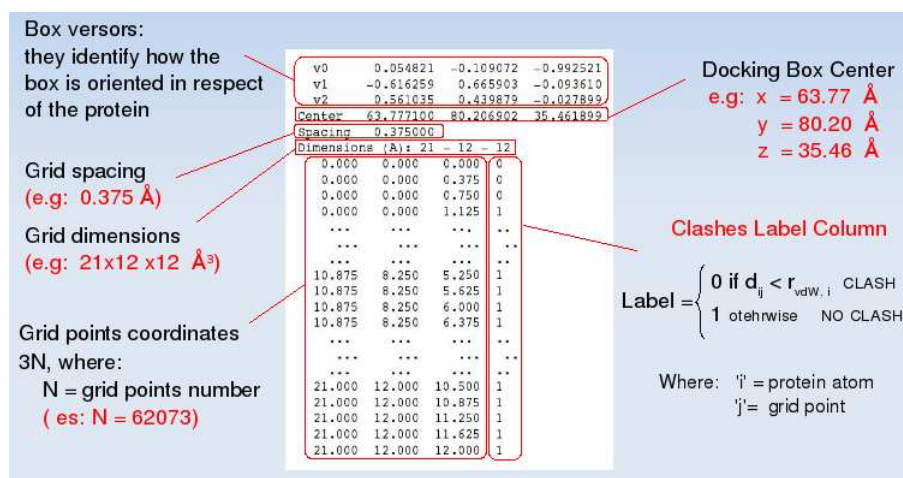


Figure 2.7: 'DockingBox.grid' output file.

- Acquisition of the grid spacing (= 0.375 Å by default)
- Acquisition of the grid dimensions ( $L_x, L_y, L_z$ )

### Generation of the grid

- Calculation of the grid points coordinates inside the box starting from the origin and moving to the three directions according to the grid spacing value. A number 'N' of points dependent on the grid spacing and on the docking box sizes is generated: each grid map will have a number of rows equal to 'N' (see Fig. 2.7).
- 'Clash' labels assignment: if a point ' $P_j$ ' of the grid is too near to any atom ' $A_i$ ' of the macromolecule, no ligand atom could be placed there (presence of a 'clash'): the label '0' is therefore assigned to the point ' $P_j$ '. On the contrary, if the distance ' $d_{ij}$ ' between ' $P_j$ ' and ' $A_i$ ' is acceptable (no 'clash'), the label '1' is assigned.

The threshold is the van der Waal radius of the atom ' $A_i$ ':



- ◇ if  $d_{ij} < r_{vdw,i} \Leftrightarrow \text{flag} = 0$  [clash]
- ◇ if  $d_{ij} \geq r_{vdw,i} \Leftrightarrow \text{flag} = 1$  [no clash]

### Generation of the electrostatic potential energy grid map

- In each point of the grid, the electrostatic contribution (Coulomb) is pre-calculated as the summation of the pairwise interactions between each atom of the macromolecule and a probe atom with unit positive charge. This is obtained through one cycle over the atoms of the macromolecule.
- Each term of the summation depends on the distance between the actual grid point and the actual atom of the macromolecule, according to the standard electrostatic Coulomb law.
- A sigmoidal distance 'd' dependent dielectric function (2.1) is used to model solvent screening ('Mehler Solmajer function' [22]).

$$\epsilon(d) = A + \frac{B}{l + ke^{\lambda\beta d}} \quad (2.1)$$

- It is so generated the electrostatic potential energy grid map (see Fig. 2.8).



**Generation of the van der Waals potential energy grid maps**

- To consider all the possible atomic species that can be present in a ligand, 11 different atomic species are taken into account:
  - ◊ C [Carbon]
  - ◊ N [Nitrogen]
  - ◊ O [Oxygen]
  - ◊ H [Hydrogen]
  - ◊ P [Phosphorus]
  - ◊ S [Sulphur]
  - ◊ Fe [Iron]
  - ◊ F [Fluorine]
  - ◊ Cl [Chlorine]
  - ◊ Br [Bromine]
  - ◊ I [Iodine]
- In each point of the grid and for each of the 11 possible atomic species, the van der Waals energetic contribution is pre-calculated as the summation of the pairwise interactions between each atom of the macromolecule and the considered atomic species. This gives rise to 11 grid maps (see Fig. 2.10) and is obtained through two nested cycles:
  - ◊ over the atoms of the macromolecule
  - ◊ over the 11 atomic species of the ligand
- Each summation term representing the Van der Waals interaction experienced between two instantaneous dipoles is calculated by using the '(6-12) Lennard-Jones' truncated function (2.2) (see Fig. 2.9):

$$E_{\text{vdW}}(d_{ij}) = \begin{cases} \frac{A_{ij}}{d_{ij}^{12}} - \frac{B_{ij}}{d_{ij}^6} & \text{if } d_{ij} \geq (r_{\text{vdW},i} + r_{\text{vdW},j}) \\ E(r_{\text{vdW}}) & \text{if } 0 < d_{ij} < (r_{\text{vdW},i} + r_{\text{vdW},j}) \end{cases} \quad (2.2)$$

where:

- ◇ 'A' and 'B' are, respectively, the repulsive and the attractive parameters
- ◇ 'i' is the protein atom index
- ◇ 'j' is the index for the hypothetical ligand atom in the actual grid point

The function is truncated for  $0 < d_{ij} < (r_{\text{vdW},i} + r_{\text{vdW},j})$  because of the presence of the clash. The algorithm already takes into account the clash in the previous phase of grid points generation, therefore the calculation of the van der Waals potential in that zone of space is unnecessary.

- A and B parameters are obtained from  $r_{ii}$ ' and ' $\epsilon_{ii}$ ', which represent, respectively, the sum of van der Waals radii of two like atoms and the van der Waals well depth of the same two atoms.

#### Generation of the desolvation potential energy grid maps

- Carbon atoms only are considered in order to generate desolvation potential energy grid maps, making a distinction between the 2 possible cases:
  - ◇  $C_{al}$  [Aliphatic carbon]
  - ◇  $C_{ar}$  [Aromatic carbon]
- In each point of the grid and for each of the 2 carbon atom-types of the ligand, the desolvation contribution is pre-calculated as the summation of the pairwise interactions between each atom of the macromolecule and the considered carbon atom-type.



This gives rise to 2 grid maps (see Fig. 2.11) and is obtained through two nested cycles:

- ◇ over the atoms of the macromolecule
- ◇ over the 2 carbon atom-types of the ligand
- Each summation term is calculated by using the relation (2.3), obtained by a heuristic approach [20][21]:

$$E_{\text{desolv}}(d_{ij}) = (S_i V_j + S_j V_i) \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right) \quad (2.3)$$

where:

- ◇ ' $S_i$ ' and ' $S_j$ ' are the 'solvation terms' for the receptor and the ligand atoms respectively
- ◇ ' $V_i$ ' and ' $V_j$ ' are the 'atomic fragmental volume' for the receptor and the ligand atoms respectively
- ◇ ' $\sigma$ ' is the 'gaussian distance constant' ( $= 3.5 \text{ \AA}$ )
- ◇ ' $d_{ij}$ ' is the distance between atom 'i' of the receptor and atom 'j' of the ligand

#### Generation of the H-bonding potential energy grid maps

- To model the hydrogen-bonding interaction between the receptor and the ligand, the 11 possible donor-acceptor couples listed in Tab.2.3 are taken into account. They correspond to 10 possible atom-types of the receptor and to 4 possible atom-types of the ligand.
- The H-bonding potential energy grid maps are generated in correspondence with the 4 atom-types of the ligand:
  - ◇ H [hydroxyl]
  - ◇ H [ammonium]
  - ◇ O [carbonyl]
  - ◇ O [hydroxyl]



- In each point of the grid and for each of the 4 possible atom-types of the ligand, the H-bonding potential is pre-calculated as the summation of the pairwise interactions between each atom of the receptor chosen among those possible 10 and the considered atom-type of the ligand.

This gives rise to 4 grid maps (see Fig. 2.12) and is obtained through two nested cycles:

- ◊ over the atoms of the macromolecule
- ◊ over the 4 atom-types of the ligand

- Each summation term representing the H-bonding interaction experienced by the donor-acceptor couple is calculated with a heuristic method, by using the '(10-12)' form of the 'Lennard-Jones' function (2.4):

$$E_{\text{H-bonding}}(d_{ij}) = \frac{C_{ij}}{d_{ij}^{12}} - \frac{D_{ij}}{d_{ij}^{10}} \quad (2.4)$$

where ' $C_{ij}$ ' and ' $D_{ij}$ ' are respectively the repulsive and attractive parameters for the considered couples of atoms.

- Hydrogen bond directionality is not taken into account in the model used inside this software.

### Complete interaction energy expression

In each point 'j' of the grid formed by N points, the interaction energy value can be directly calculated by using the following expression:

$$\begin{aligned} E_{\text{interaction}}(j) &= w_{\text{el}} \sum_{i=1}^M E_{\text{el}}(d_{ij}) + w_{\text{vdW}} \sum_{i=1}^M E_{\text{vdW}}(d_{ij}) + \\ &+ w_{\text{desolv}} \sum_{i=1}^M E_{\text{desolv}}(d_{ij}) + w_{\text{H-bonding}} \sum_{i=1}^M E_{\text{H-bonding}}(d_{ij}) \end{aligned}$$

where:

- ◊ 'M' is the total protein atoms number



<b>Receptor (10 atom-types)</b>	<b>Ligand (4 atom-types)</b>
O (carbonyl)	H (hydroxyl)
N (primary amine)	H (hydroxyl)
N (secondary amine)	H (hydroxyl)
N (secondary amine)	H (ammonium)
O (hydroxyl)	H (ammonium)
O (hydroxyl)	H (hydroxyl)
H (secondary amine)	O (carbonyl)
H (primary amine)	O (hydroxyl)
H (ammonium)	O (hydroxyl)
H (hydroxyl)	O (carbonyl)
O (carboxyl)	H (hydroxyl)

Table 2.3: Hydrogen-bonding donor-acceptor couples taken into account by the program.





## Chapter 3

# Ligand preparation

The more accurate the physical description of the ligand structure is, the more relevant, accurate and useful the predicted binding mode will be. Consequently, particular attention must be given to the file preparation of the ligand, of which molecular docking simulation will be made.

The ligand structure files usually come from ligand-receptor complex structures determined by X-ray diffraction experiments and NMR experiments, or from models generated by means of a calculator. On-line there are large ligand databases, from which it is possible to download the compound records that contain experimental or calculated three-dimensional coordinates and sometimes information about biological activity.

However, in these files are often present some mistakes that can considerably affect the simulations of molecular docking; these simulations are dramatically sensitive to the propagation of even small errors (see Section 1.3.2).

For instance, the assignment of the hydrogen atoms has to be carefully checked. It is possible that one of these hydrogen atoms clashes with the receptor, so making the native pose unfavourable. For many small molecules, there is no single 'correct' protonation state valid in all cases, and it may happen that one alternative protonation state could be more suitable.

Also, bond distances and bond orders of the molecule are to be carefully checked: if there is an error in the structure file, and a single bond is taken in place of a double bond, hydrogens will be added in excess. To avoid

meaningless connectivity, a further check is needed. Furthermore, before docking a ligand, atom types and charges for every atom in the ligand are needed.

Another particularly important aspect concerns the choice of the ligand initial conformation to be used for the molecular docking calculation. Although the conformational variation energy of the ligand:

$$\Delta E_L = E_{L_{\text{bound}}} - E_{L_{\text{unbound}}}$$

when passing from the unbound to the bound conformation is a crucial term in the docking energy calculation, a standard procedure that establishes how to generate the ligand initial conformation has not been defined yet.

The said variation (*in vacuo*) must always be positive, as the docked ligand conformation is necessarily less stable than that of the free ligand. In the semiflexible docking model, what leads the docking process is the compromise between the energy gained from the ligand-receptor interaction and the energy used for the ligand deformation (see Section 4.1), being the energy variation associated to the receptor equal to zero.

What induced me to develop a global optimizer (see Section 3.2) of the ligand conformation was the attempt to establish an univocal criterion to generate a single reference point for each ligand; reference point to be used in the subsequent docking energy calculation.

To my opinion, this aspect of molecular docking should deserve more consideration in as much as the absence of such a tool will lead to results that depend on the initial ligand conformation.

I'm not in agreement with AutoDock 4.0, because instead of generating a single reference point for each ligand, it gives the possibility to the user to choose "how the internal energy of the ligand should be treated when estimating the free energy of binding, and can be set to one of the following strings":

- 'unbound\_same\_as\_bound': this assumes the internal energy of the ligand is the same before and after binding
- 'extended': this assumes the internal energy of the ligand is that of an extended conformation when unbound
- 'compact': this assumes the internal energy of the ligand is that of a compact conformation when unbound

Having said that, if a user takes into consideration the necessity to find a single reference point, it will be difficult for him to find an optimizer homogeneous with the docking module, both from the point of view of the force field and on how to calculate the ligand energy.

The two developed modules:

- ◇ 'Ligand optimizer' (see Section 3.2)
- ◇ 'Semi flexible molecular docking' (see Section 4.1)

ensure such homogeneity from both points of view. Taking for granted that the *in vacuo* optimized ligand structure, as the whole model on which the molecular docking is based, is an approximation of the real system, my target is to put in evidence the above aspects and to propose a new possible approach.

### 3.1 Ligand preparation

Before passing to the real optimization phase, the file containing the structure of the ligand to be optimized, is to be prepared so that the following requests are satisfied:

- hydrogen atoms are to be added in advance according to one of the following models:
  - ◇ *all hydrogens*, in which all hydrogen atoms are explicit
  - ◇ *polar hydrogens*, in which only polar hydrogen atoms are explicit (*united-atom model*)
  - ◇ *polar and aromatic hydrogens*, in which only polar and aromatic hydrogen atoms are explicit
- the initial ligand conformation should be realistic from a physical and chemical point of view, i.e. it must be a structure obtained by a previous local optimization: in particular, bond lengths and bond angles are to be correct as they are held fixed during the optimization.
- the assignment of appropriate atomic partial charges to the ligand is essential to obtain meaningful results from any electrostatics calculation; partial atomic charges must be established in advance: they will

be held fixed, both during the optimization and during the subsequent procedure of molecular docking.

For these reasons, the use of a program that performs a local optimization and that attributes partial atomic charges to the ligand atoms is recommended. To this purpose MOPAC [25], that is a semiempirical quantum chemistry software package for the prediction of chemical properties, results to be highly suitable. Alternatively, if a reliable ligand structure is already available, partial atomic charges can be assigned by an empirical method such as 'Marsili-Gasteiger' [24], that is based on the difference in electronegativities of the atoms.

## 3.2 Ligand optimizer

This module is a global optimizer of the ligand conformation: its function is to determine the *in vacuo* minimum classic conformation of the free ligand. In order to reduce the system degrees of freedom, this is done in the space of the conformations which depend on the torsion angles relevant to the single bonds. The structure of a molecule can be defined with high precision by the dihedral angles between three successive chemical bond vectors: the dihedral angle varies only the distance between the first and fourth atom; the other interatomic distances are constrained by the chemical bond lengths and bond angles.

The value of minimum conformation energy is the reference to calculate the ligand conformation energy variation between the bound and unbound state in the subsequent module for the semiflexible molecular docking. A fast simulated annealing (FSA) algorithm has been implemented to find the global minimum energy ligand conformation. The objective function is the classical intramolecular interaction energy, obtained taking into account the electrostatic (Coulomb) and the van der Waals potential energy contributions. The global minimum energy conformation of the ligand is found by means of a FSA (*Fast Simulated Annealing*) algorithm. It includes an *acceptance ratio* method for an optimal control of the ligand structures generation: the control is active during the whole *cooling* phase.

The sampling rate of the conformational space results therefore in accordance with the actual temperature inside the *cycle* and the space sampling becomes more efficient from at least two points of view:

1<sup>st</sup> - the accuracy is improved: at high temperatures a number of states can be generated possibly very different from one another and belonging to separate potential energy basins of attraction. Together with the Metropolis criterion, this avoids the system to be trapped in local minima.

2<sup>nd</sup> - the precision is improved: at low temperatures the conformations generated are quite closed together, which allows to obtain a thicker sampling in the surroundings of the minimum energy conformation.

Inside each *cycle*, the temperature is the control parameter: it is progressively reduced (*cooling* phase) starting from the initial value ' $T_i$ ', up to the final value ' $T_f$ ' through decrements equal to the input *cooling rate* ' $\Delta T$ '.

The reduction of the temperature takes place each time a new confor-



mation is accepted. An equilibration phase in correspondence with each new temperature value is not scheduled: in this sense this algorithm can be defined 'Fast'.

To provide for the previous approximation (absence of equilibration phase) more *runs* and more cooling *cycles* inside each *run* can be made.

### 3.3 Technical procedure

Here is a description of the procedure used by the algorithm to perform the global ligand optimization.

#### 3.3.1 Data input

- Acquisition of Fast Simulated Annealing parameters specified by the user
  - ◊ Acquisition of the number of *runs*
  - ◊ Acquisition of the number of *cycles*
  - ◊ Acquisition of the initial temperature  $T_i$  /[K]
  - ◊ Acquisition of the final temperature  $T_f$  /[K]
  - ◊ Acquisition of the *cooling rate*  $\Delta T$  /[K]
- Acquisition of the ligand structure
  - ◊ Acquisition of all molecular atomic coordinates
  - ◊ Acquisition of the partial atomic charges of the ligand atoms

#### 3.3.2 Determination of the ligand connectivity

Starting from the molecular atomic coordinates, it is necessary an initial phase in order to establish the molecular connectivity. On every couple of the ligand atoms the interatomic distance is calculated and compared with a series of predefined possible bond distances well known in literature.

For the considered couple of atoms the absolute value of the difference between their distance in the PDB file ( $d_{calc}$ ) and the predefined distance ( $d_{predef}$ ) is calculated. If the result is lower than or equal to a predefined

threshold of  $0.2\text{\AA}$ ,  $|d_{\text{calc}} - d_{\text{predef}}| \leq 0.2\text{\AA}$ , the covalent bond is recognized and a label is assigned.

- Calculation of the triangular matrix containing the molecular interatomic distances
- Calculation of the triangular matrix representing the molecular connectivity:
  - ◊ flag = 0 no covalent bond between the two atoms
  - ◊ flag = 1 presence of covalent bond between the two atoms

### 3.3.3 Rotatable bonds detection

The aim of this function is the detection of the covalent single bonds around which the ligand can internally rotate. For each of them the dihedral angle  $\vartheta$  between three successive chemical bond vectors, where the second is relevant to the detected single bond, will be the variable that represents the free rotation. Covalent bonds belonging to blocked structures are detected and labelled.

- Detection of aliphatic and aromatic rings.
  - ◊ Dihedral angles relevant to chemical bonds belonging to cyclic structures are always held constant, even if the conformational transition between two conformations is allowed, as it happens, e.g., on the transition between the conformers of cyclohexane.
- Detection and optimization of amide functional groups (-CONH-).
  - ◊ amide groups are detected by searching a carbonyl (C=O) group where the C atom forms a single bond with the N atom of a NH group, and by evaluating the covalent bond lengths.
  - ◊ in case an amide group is detected, the value of dihedral angle (OCNH) is evaluated: the molecule is rotated around the internal CN axis up to reach the planar *trans* conformation.
  - ◊ in case a cyclic amide (*lactam*) is detected, the molecule is maintained in the planar *cis* conformation

- ◇ once the optimal amide conformation is obtained, during the subsequent ligand optimization phase each amide group is held fixed, as in the subsequent molecular docking calculations.
- Detection of double ( $C = C$ ,  $C = N$ ) and triple ( $C \equiv C$ ,  $C \equiv N$ ) bonds.
  - ◇ Double and triple bonds are detected according to their characteristic bond lengths: the rotation of the molecule around the  $CC$  or  $CN$  axes is not allowed.

### 3.3.4 Calculation of the ligand conformational energy and start of the *cooling cycle*

Each time the ligand conformational energy is to be calculated, the following procedure is used; it is exactly the same that is used in the subsequent molecular docking module (see Section: 4.1). Namely, the same force field and the same energy contributions and multiplying factors are used both in the present module and in the semiflexible molecular docking module.

- Calculation of the '1-n' (with  $n > 3$ ) electrostatic (Coulomb) potential energy contributions (with reference to the previously calculated triangular matrix representing the molecular connectivity).
- Calculation of the '1-n' (with  $n > 3$ ) van der Waals potential energy contributions (with reference to the previously calculated triangular matrix representing the molecular connectivity).
- Calculation of the ligand conformational energy  $E_L$ :

$$\begin{aligned}
 E_L &= w_{el}E_{el}(d_{kl}) + w_{vdW}E_{vdW}(d_{kl}) = \\
 &= w_{el} \sum_{k=1}^{D-3} \sum_{l=k+3}^D E_{el}(d_{kl}) + w_{vdW} \sum_{k=1}^{D-3} \sum_{l=k+3}^D E_{vdW}(d_{kl})
 \end{aligned}$$

where:

- ◇ 'k' is a ligand atom index
- ◇ 'l' is another ligand atom index

- ◇ 'D' is the total number of ligand atoms
- ◇ ' $w_{el} = 0.1146$ ' is the multiplying factor of the electrostatic term (see Section 4.1)
- ◇ ' $w_{vdW} = 0.1485$ ' is the multiplying factor of the van der Waals term (see Section 4.1)

### 3.3.5 FSA - Fast simulated annealing algorithm

Here and in the next two subsections are the FSA algorithm sequence instructions.

- The 'i+1' conformation of the ligand is generated from the 'i' conformation by varying the dihedral angles related to the previously detected *rotatable bonds*. All the dihedral angles are changed simultaneously; their new values are generated at random, but within a pre-established and temperature dependent range (see Subsection 3.3.7).
- The energy associated to the 'i+1' conformation is calculated.
- The sign of  $\Delta E_L = E_{L,i+1} - E_{L,i}$  is evaluated.

### 3.3.6 FSA - Metropolis criterion

- The 'i+1' conformation is accepted or rejected according to the 'Metropolis criterion':
  - ◇ if  $\Delta E_L < 0 \Rightarrow$  the new conformation is accepted. A new conformation is generated from the last one and the actual temperature is reduced of ' $\Delta T$ ' (*cooling rate*)
  - ◇ if  $\Delta E_L > 0 \Rightarrow$  the new conformation is accepted with a probability

defined by the following equation:

$$\begin{aligned}
 P_{\text{acceptance}} &= \frac{P_{i+1}}{P_i} = \\
 &= \frac{\frac{\exp(-\beta E_{L,i+1})}{Z}}{\frac{\exp(-\beta E_{L,i})}{Z}} = \\
 &= \frac{\exp(-\beta E_{L,i+1})}{\exp(-\beta E_{L,i})} = \\
 &= \exp(-\beta(E_{L,i+1} - E_{L,i})) = \\
 &= \exp(-\beta \Delta E_L)
 \end{aligned}$$

where:

- $P_{i+1}$  and  $P_i$  are the probabilities of the 'i+1' and 'i' conformations according to the Boltzmann distribution
- $\beta$  is equal to  $1/kT$ , where  $k$  is the Boltzmann constant and  $T$  is the absolute temperature
- $Z$  is the partition function of the system

A *random* number 'r' is therefore generated in the range  $[0,1]$ :

- ▷ if  $0 \leq r \leq P_{\text{acceptance}}$ 
  - the 'i+1' ligand conformation is accepted
  - the temperature is reduced by the quantity  $\Delta T$
  - a new ligand conformation is generated starting from the just accepted conformation
- ▷ if  $P_{\text{acceptance}} < r \leq 1$ 
  - the 'i+1' ligand conformation is rejected
  - the temperature remains unchanged
  - a new ligand conformation is generated starting from the previous 'i' conformation

### 3.3.7 FSA - *Acceptance ratio* method

This module includes an *acceptance ratio* method for an optimal control of the ligand structures generation: the control is active inside each *cycle* during the whole *cooling* phase, from the initial temperature  $T_i$  to the final temperature  $T_f$ .

- Each new 'i+1' conformation is generated from the previously accepted 'i' conformation, through a simultaneous variation of all the 'F' dihedral angles:

$$\vartheta_{i+1,j} = \vartheta_{i,j} + \delta\vartheta_{i,j} \quad \text{with } j = 1, 2, \dots, F \quad (3.1)$$

where:

- 'F' is the total number of dihedral angles of the ligand, coincident with its internal 'degrees of freedom' total number
- ' $\delta\vartheta_{i,j}$ ' is the dihedral angle variation, obtained by means of a random number generator, so as that:

$$\delta\vartheta_{i,j} \in [-\gamma, +\gamma] \quad (3.2)$$

where ' $\gamma$ ' can assume values between  $+30^\circ$  and  $+2^\circ$ , with a step variation of  $\pm 4^\circ$ , established as follows: inside each *cooling cycle*, blocks of 500 new conformations are considered and is calculated the *ratio*:

$$a_r = \frac{\textit{accepted}}{500} \quad (3.3)$$

where '*accepted*' stands for the number of accepted conformations according to the above mentioned Metropolis criterion. The  $\gamma$  value is so changed as follows:

- ◇ if  $a_r < 0.45 \Rightarrow$  the actual  $\gamma$  value is reduced by  $4^\circ$   
( $\gamma \rightarrow \gamma - 4^\circ$ )
- ◇ if  $a_r > 0.55 \Rightarrow$  the actual  $\gamma$  value is increased by  $4^\circ$   
( $\gamma \rightarrow \gamma + 4^\circ$ )

By doing so, it is possible to calibrate how to generate the ligand conformations while searching the energy global minimum.

### 3.3.8 Execution of all *runs* and *cycles*

The program goes on executing the *runs* and *cycles* selected by the user, by means of two nested cycles on:

- ◇ number of *runs*
- ◇ number of *cycles*

up to the completion of the last *cycle* of the last *run*.

- Each *run* starts from a different ligand conformation, generated through a random variation of the input structure. All *cycles* of the same *run* start from the same initial ligand conformation.
- Ligand integrity check: for each 'i+1' conformation generated during the research of the minimum (in every *cycle* and in every *run*) a check on the intramolecular 'd' bond distances is effected. If after the numerous dihedral rotations an excessive error (from a 'numerical' point of view) on any of the bond distances is introduced in respect of the original bond distance, the actual 'i+1' conformation is rejected and a new 'i+1' conformation is generated starting from the 'i' conformation. This prevents the possible ligand fragmentation due to the propagation of computational errors.
- During the whole calculation, a temporary file containing the best ligand conformation (PDB format) updated in real time is generated.
- Eventually, the file (PDB format) containing the ligand conformation with the lower energy (see Fig. 3.1) among all those generated by each *cycle* of each *run* is printed on standard output.

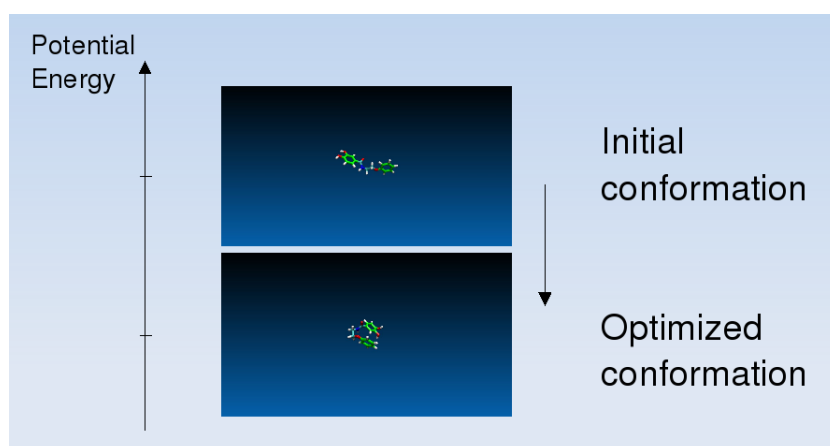


Figure 3.1: Conformational change after ligand optimization





## Chapter 4

# Semi flexible molecular docking

### 4.1 Semi flexible molecular docking

The module 'Semi flexible molecular docking' is aimed to determine the best 'conformation-pose' of a ligand inside a possible receptorial binding site located in advance on the surface of a protein (see Fig. 4.1) and to supply an estimate of the relevant binding affinity, expressed as docking energy ' $E_d$ '. The docking energy represents an approximation of the binding free energy variation  $\Delta G_{binding}$ :

$$\Delta G_{binding} \approx E_d \quad (4.1)$$

relevant to the formation of the ligand-receptor complex (RL), starting from the unbound ligand (L) and the unbound receptor (R):



where the equilibrium constant  $K_{eq}$  and the dissociation constant  $K_d$  depend on the equilibrium concentrations of [RL], [R] and [L]:

$$K_{eq} = \frac{[RL]}{[R][L]} = \frac{1}{K_d} \quad (4.3)$$

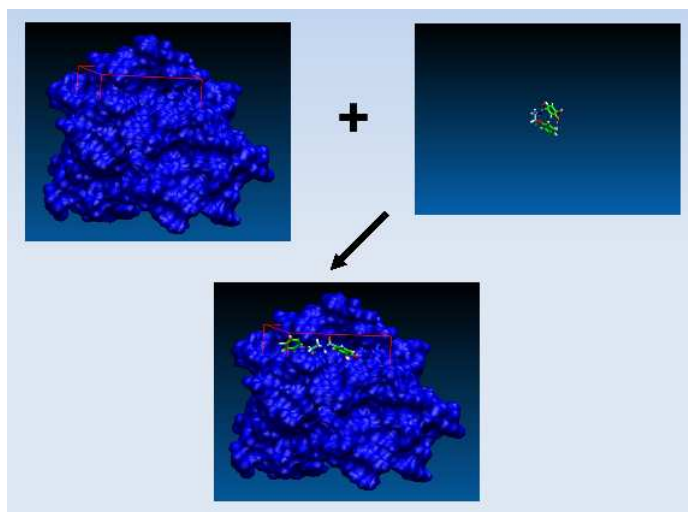


Figure 4.1: Molecular docking simulation: the best 'conformation-pose' of a ligand inside a possible receptorial binding site on the surface of a protein is searched

and where the relation between  $K_d$  and the binding free energy variation  $\Delta G_{binding}$  is:

$$\begin{aligned}
 \Delta G_{binding} &= -RT \ln K_{eq} &= \\
 &= -RT \ln \frac{1}{K_d} &= \\
 &= -RT \ln K_d^{-1} &= \\
 &= +RT \ln K_d &
 \end{aligned}
 \tag{4.4}$$

The module 'Semi flexible molecular docking' is a global optimizer of the ligand conformation-pose inside the docking box and it is based on the 'rigid macromolecule and flexible ligand' model. The variables on which the objective function ( $E_d$ ) depends are represented by the ligand roto-translational degrees of freedom into the binding site and by the ligand torsional angles relevant to the single bonds around which the internal free rotation of the molecule is allowed (*rotatable bonds*).

### 4.1.1 Docking energy definition

The docking energy is calculated through force field (classical approach) and keeps into account both the ligand-receptor interaction energy ( $E_{R-Linteraction}$ ) and the ligand conformational energy variation when passing from the 'unbound' to the 'bound' state ( $\Delta E_L$ ). It is defined as follows:

$$\begin{aligned}
 E_{\text{docking}} &= E_{\text{final}} - E_{\text{initial}} = \\
 &= (E_{R_{\text{bound}}} + E_{L_{\text{bound}}} + E_{R-Linteraction}) - (E_{R_{\text{unbound}}} + E_{L_{\text{unbound}}}) = \\
 &= (E_{R_{\text{bound}}} - E_{R_{\text{unbound}}}) + (E_{L_{\text{bound}}} - E_{L_{\text{unbound}}}) + E_{R-Linteraction} = \\
 &= \Delta E_R + \Delta E_L + E_{R-Linteraction} = \\
 &= \Delta E_L + E_{R-Linteraction} \tag{4.5}
 \end{aligned}$$

where:

- ◇  $E_{\text{final}}$  is the energy of system in the final state, after the formation of the complex
- ◇  $E_{\text{initial}}$  is the energy of the system in the initial state, before the formation of the complex
- ◇  $E_{R_{\text{bound}}}$  is the energy of the receptor in the conformation bound to the ligand
- ◇  $E_{L_{\text{bound}}}$  is the energy of the bound ligand conformation
- ◇  $E_{R_{\text{unbound}}}$  is the energy of the receptor in the unbound conformation
- ◇  $E_{L_{\text{unbound}}}$  is the energy of the unbound ligand conformation
- ◇  $\Delta E_L$  is the ligand energy variation in passing from the 'unbound' to the 'bound' conformation
- ◇  $\Delta E_R$  is the receptor energy variation of the macromolecule in passing from the 'unbound' to the 'bound' conformation
- ◇  $E_{R-Linteraction}$  is the ligand-receptor interaction energy after the formation of the complex

The model on which this module is based deals the receptor as rigid, consequently its conformational energy remains constant:

$$\bullet E_{R_{\text{bound}}} = E_{R_{\text{unbound}}} \Rightarrow \Delta E_R = 0$$

The ligand, instead, is dealt as flexible. Its conformational energy increases when passing from the unbound to the bound state:

$$\bullet E_{L_{\text{bound}}} > E_{L_{\text{unbound}}} \Rightarrow \Delta E_L > 0$$

as  $E_{L_{\text{unbound}}}$  represents the free ligand minimum energy *in vacuo* (calculated through 'Ligand optimizer' module described in Section 3.2), while  $E_{L_{\text{bound}}}$  represents the ligand energy of any other conformation. This conformation is always less favourable of when the ligand is considered isolated (see Fig. 4.2), but it could however be accepted when the energy obtained by the interaction with the receptor is considered.

### 4.1.2 Scoring function

With reference to the docking energy definition (eq: 4.5) and by splitting all contributions, the complete expression of the scoring function used in the present module is as follows:

$$\begin{aligned} E_{\text{docking}} &= (E_{L_{\text{bound}}} - E_{L_{\text{unbound}}}) + E_{R-L_{\text{interaction}}} = \\ &= (w_{\text{el}}E_{\text{el}} + w_{\text{vdW}}E_{\text{vdW}})_{L_{\text{bound}}} - \\ &- (w_{\text{el}}E_{\text{el}} + w_{\text{vdW}}E_{\text{vdW}})_{L_{\text{unbound}}} + \\ &+ w_{\text{el}}E_{\text{el}} + w_{\text{vdW}}E_{\text{vdW}} + \\ &+ w_{\text{desolv}}E_{\text{desolv}} + w_{\text{H-bonding}}E_{\text{H-bonding}} \end{aligned} \quad (4.6)$$

where four main energy contributions are taken into account:

- ◇ electrostatic
- ◇ van der Waals
- ◇ desolvation
- ◇ hydrogen-bonding

To a first approximation, the molecular mechanics-based terms are multiplied by the same coefficients of the AutoDock 3.0 [3] scoring function; the coefficient values are:

$$\begin{aligned}w_{elec} &= 0.1146 \\w_{vdW} &= 0.1485 \\w_{hbond} &= 0.0656 \\w_{desolv} &= 0.1711\end{aligned}$$

They were determined by linear regression analysis of complexes with known 3D-structures and known binding free energies.

To search the phase space of the ligand-receptor system a Fast Simulated Annealing (FSA) algorithm is used: it is applied many times ('*cycles*'), starting from several 'conformation-poses' generated at random ('*r-run*'). The aim is to find out the 'conformation-pose' of the ligand inside the binding pocket for which the energy is the global minimum. FSA algorithm works in a way that prevents the system from being trapped in local minima. Each new conformation is always accepted if its energy is lower than the previous conformation-pose. If the energy is higher, then the conformation-pose can be accepted or rejected according to a probability based on Boltzmann distribution (Temperature dependent). Consequently, each time a conformation-pose is accepted, the temperature is automatically reduced. The whole process goes on until the temperature reaches the value near to 0 K.

## 4.2 Input files preparation

Before launching the molecular docking program, it is necessary to prepare the files that will be read during the calculation.

- **Receptor preparation:** with regard to the receptor, the following files are needed (see Chapter 2):
  - ◇ file containing the points of the grid
  - ◇ file containing the electrostatic potential energy grid maps
  - ◇ file containing the van der Waals potential energy grid maps

- ◇ file containing the desolvation potential energy grid maps
- ◇ file containing the H-bonding potential energy grid maps

It is therefore requested a preliminar phase finalized to the generation of such files, obtained through the application of the module 'Potential energy grids calculator' (see Section 2.4).

- **Ligand preparation:** as far as the ligand is concerned, the following file is needed (see Chapter 3):
  - ◇ file (PDB format) containing the globally optimized conformational structure of the ligand and the partial atomic charges of the ligand atoms

The most important request is to have a globally optimized ligand structure, to be used as a reference in the calculation of the docking energy. This will be discussed in detail in the next sections. To run the molecular docking program, another preliminar phase is therefore necessary where, by using the 'Ligand optimizer' module (see Section 3.2), a file that satisfies the various requests is generated.

## 4.3 Technical procedure

Here is a description of the procedure used by the algorithm to perform the molecular docking simulation.

### 4.3.1 Data input

- Acquisition of Fast Simulated Annealing parameters specified by the user
  - ◊ Acquisition of the number of *runs*
  - ◊ Acquisition of the number of *cycles*
  - ◊ Acquisition of the initial temperature  $T_i$  /[K]
  - ◊ Acquisition of the final temperature  $T_f$  /[K]
  - ◊ Acquisition of the *cooling rate*  $\Delta T$  /[K]
- Acquisition of the precalculated potential energy grid maps of the receptor
  - ◊ Acquisition of the grid
  - ◊ Acquisition of the electrostatic potential energy grid map
  - ◊ Acquisition of the van der Waals potential energy grid maps
  - ◊ Acquisition of the desolvation potential energy grid maps
  - ◊ Acquisition of the H-bonding potential energy grid maps
- Acquisition of the globally optimized ligand structure
  - ◊ Acquisition of all molecular atomic coordinates
  - ◊ Acquisition of the partial atomic charges of the ligand atoms

### 4.3.2 Determination of the ligand connectivity

Starting from the molecular atomic coordinates, it is necessary an initial phase in order to establish the molecular connectivity. For each of all couples of the ligand atoms the interatomic distance is calculated and compared with a series of predefined possible bond distances well known in literature.



For the considered couple of atoms the absolute value of the difference between their distance in the PDB file ( $d_{\text{calc}}$ ) and the predefined distance ( $d_{\text{predef}}$ ) is calculated. If the result is lower than or equal to a predefined threshold of  $0.2\text{\AA}$ ,  $|d_{\text{calc}} - d_{\text{predef}}| \leq 0.2\text{\AA}$ , the covalent bond is recognized and a label is assigned.

- Calculation of the triangular matrix containing the molecular inter-atomic distances
- Calculation of the triangular matrix representing the molecular connectivity:
  - ◊ flag = 0 no covalent bond between the two atoms
  - ◊ flag = 1 presence of covalent bond between the two atoms

### 4.3.3 Rotatable bonds detection

The aim of this function is the detection of the covalent single bonds around which the ligand can internally rotate. For each of them the dihedral angle  $\vartheta$  between three successive chemical bond vectors, where the second is relevant to the detected single bond, will be the variable that represents the free rotation. Covalent bonds belonging to blocked structures are detected and labelled.

- Detection of aliphatic and aromatic rings.
  - ◊ Dihedral angles relevant to chemical bonds belonging to cyclic structures are always held constant, even if the conformational transition between two conformations is allowed, as it happens, e.g., on the transition between the conformers of cyclohexane.
- Detection and optimization of amide functional groups (-CONH-).
  - ◊ amide groups are detected by searching a carbonyl (C=O) group where the C atom forms a single bond with the N atom of a NH group, and by evaluating the covalent bond lengths.
  - ◊ in case an amide group is detected, the value of dihedral angle (OCNH) is evaluated: the molecule is rotated around the internal CN axis up to reach the planar *trans* conformation.

- ◇ in case a cyclic amide (*lactam*) is detected, the molecule is maintained in the planar *cis* conformation
- ◇ once the optimal amide conformation is obtained, during the subsequent ligand optimization phase each amide group is held fixed, as in the subsequent molecular docking calculations.
- Detection of double ( $C = C$ ,  $C = N$ ) and triple ( $C \equiv C$ ,  $C \equiv N$ ) bonds.
  - ◇ Double and triple bonds are detected according to their characteristic bond lengths: the rotation of the molecule around the CC or CN axes is not allowed.

#### 4.3.4 Calculation of the free ligand energy reference

Before the start of the run, the value of the input ligand conformation energy is recalculated by the present module. This value corresponds to the minimum energy conformation found by the 'Ligand optimizer' module (see Section: 3.2). It will represent the reference value ( $E_{L\text{unbound}}$ ) for the calculation of the ligand conformation energy variation when passing from the unbound to the bound state (see Fig. 4.2).

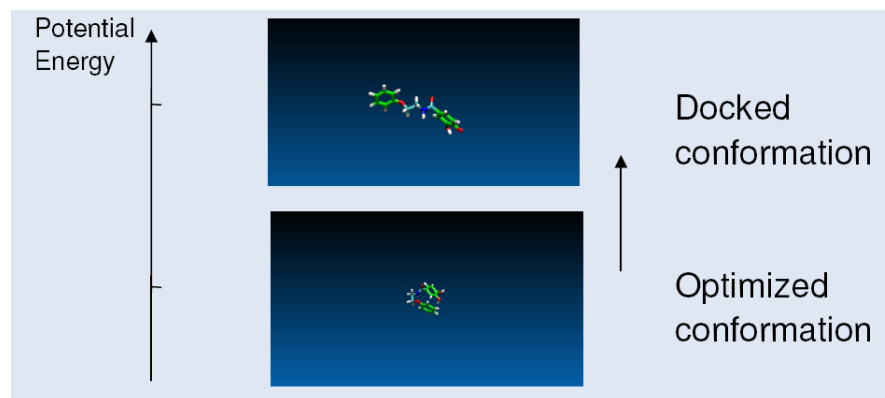


Figure 4.2: Conformational change after ligand binding

Each time the ligand conformation energy is to be calculated inside this module, the used procedure is exactly the same used in the previous 'Ligand optimizer' module (see Section: 3.2): namely, same force field [15], same energy contributions and same multiplying factors.

- Calculation of the '1-n' (with  $n > 3$ ) electrostatic (Coulomb) potential energy contributions (with reference to the previously calculated triangular matrix representing the molecular connectivity).
- Calculation of the '1-n' (with  $n > 3$ ) van der Waals potential energy contributions (with reference to the previously calculated triangular matrix representing the molecular connectivity).
- Calculation of the ligand conformational energy  $E_L$ :

$$\begin{aligned}
 E_L &= w_{el}E_{el}(d_{kl}) + w_{vdW}E_{vdW}(d_{kl}) = \\
 &= w_{el} \sum_{k=1}^{D-3} \sum_{l=k+3}^D E_{el}(d_{kl}) + w_{vdW} \sum_{k=1}^{D-3} \sum_{l=k+3}^D E_{vdW}(d_{kl})
 \end{aligned}
 \tag{4.7}$$

where:

- ◇ 'k' is a ligand atom index
- ◇ 'l' is another ligand atom index
- ◇ 'D' is the total number of ligand atoms
- ◇ ' $w_{el} = 0.1146$ ' is the multiplying factor of the electrostatic term (see Section 4.1)
- ◇ ' $w_{vdW} = 0.1485$ ' is the multiplying factor of the van der Waals term (see Section 4.1)

### 4.3.5 Start of the *run*

The first step of each *run* consists of generating a random ligand conformation-pose centred inside the docking box so that it will not overlap (*clash*) the protein atoms. The research of a conformation-pose without clashes can represent the bottle-neck of the whole molecular docking calculation. This can happen whenever sizes and/or shape of the binding site

are not suitable to host the ligand, e.g. when the site is too narrow for the ligand. The use of the 'united-atom model' (see Section 3.1) for the ligand and for the receptor can contribute to partially speed up this step because of the reduced steric hindrance.

- Generation of the initial ligand conformation. The new ligand conformation is generated from the input conformation through a simultaneous variation of all the 'F' dihedral angles:

$$\vartheta_{1,j} = \vartheta_{0,j} + \delta\vartheta_{0,j} \quad \text{with } j = 1, 2, \dots, F \quad (4.8)$$

where:

- ◇ 'F' is the total number of dihedral angles of the ligand, coincident with its internal 'degrees of freedom' total number
- ◇ ' $\delta\vartheta_{0,j}$ ' is the dihedral angle variation, obtained by means of a random number generator, so as that:

$$\delta\vartheta_{0,j} \in [-180^\circ, +180^\circ] \quad (4.9)$$

- Ligand placement into the docking box
  - ◇ calculation of the ligand geometric centre  $[c_{L,x}, c_{L,y}, c_{L,z}]$
  - ◇ calculation of the components of the  $\vec{T}$  vector that joins the ligand geometric centre  $[c_{L,x}, c_{L,y}, c_{L,z}]$  to the docking box geometric centre  $[c_{B,x}, c_{B,y}, c_{B,z}]$ .
  - ◇ ligand roto-traslation into the docking box
    - the translation is realized by the  $\vec{T}$  vector
    - the rotation is made at random
- Some useful checks are made:
  - ◇ Check on the sign of  $\Delta E_L$
  - ◇ Check on ligand position
  - ◇ Check on intermolecular clashes
  - ◇ Check on the number of attempts

The above checks are explained in detail in Subsection 4.3.7.

### 4.3.6 Docking energy calculation

Each time the docking energy ( $E_{\text{docking}}$ ) is to be calculated, the following relation is used:

$$\begin{aligned} E_{\text{docking}} &= \Delta E_L + E_{\text{R-Linteraction}} = \\ &= E_{\text{Lbound}} - E_{\text{Lunbound}} + E_{\text{R-Linteraction}} \end{aligned} \quad (4.10)$$

where:

- ◇  $E_{\text{Lbound}}$  is the energy of the bound ligand conformation inside the docking box; it is calculated as indicated in Subsection 4.3.4, equation (4.7)
- ◇  $E_{\text{Lunbound}}$  is the energy of the free ligand calculated one single time before the start of all *runs*, as indicated in Subsection 4.3.4, equation (4.7)
- ◇  $E_{\text{R-Linteraction}}$  is the ligand-receptor interaction energy for the considered conformation-pose, calculated as shown in the next Section.

### Interaction energy calculation

Each time the ligand-receptor interaction energy is to be calculated, reference is made to the precalculated energy values in correspondence with the 'j' points of the grid (see Subsection 2.4.1).

$$\begin{aligned} E_{\text{interaction}}(j) &= w_{\text{el}} \sum_{i=1}^M E_{\text{el}}(d_{ij}) + w_{\text{vdW}} \sum_{i=1}^M E_{\text{vdW}}(d_{ij}) + \\ &+ w_{\text{desolv}} \sum_{i=1}^M E_{\text{desolv}}(d_{ij}) + w_{\text{H-bonding}} \sum_{i=1}^M E_{\text{H-bonding}}(d_{ij}) \end{aligned} \quad (4.11)$$

where:

- ◇ 'M' is the total protein atoms number
- ◇ 'i' is the protein atom index
- ◇ 'j' is the index for the actual grid point

- ◇ 'w<sub>el</sub>', 'w<sub>vdW</sub>', 'w<sub>H-bonding</sub>', 'w<sub>desolv</sub>' are coefficients that come from AutoDock 3.0 scoring function and that have been determined by linear regression analysis of complexes with known 3D-structures and known binding free energies

$$\begin{aligned}
 w_{elec} &= 0.1146 \\
 w_{vdW} &= 0.1485 \\
 w_{hbond} &= 0.0656 \\
 w_{desolv} &= 0.1711
 \end{aligned}$$

As it is explained in the Section 4.1 , each potential energy term is scaled through an appropriate multiplying factor in order to assimilate the potential energy to the corresponding free energy term.

The interaction energy value ( $E_{R-L_{interaction}}$ ) to be used inside the docking energy expression (4.1) is calculated as the summation of the interaction energies between each 'g' ligand atom and the receptor atoms.

$$E_{R-L_{interaction}} = \sum_{g=1}^G E_{R-L_{interaction,g}} \quad (4.12)$$

where:

- ◇ 'g' is the ligand atom index
- ◇ 'G' is the total number of ligand atoms
- ◇ ' $E_{R-L_{interaction,g}}$ ' is the interaction energy between the 'g' ligand atom and the receptor atoms

### Three linear interpolation

Since no ligand atom will practically ever meet exactly any grid point where the interaction energy has been precalculated, the 'three linear interpolation' method [26] is used to calculate the intermediate  $E_{R-L_{interaction,g}}$  energy value within the lattice points to which the pre-calculated values  $E_{interaction}(j)$  correspond (see eq. 4.11).

The grid is formed by a three-dimensional cubic lattice with constant spacing ( $0.375\text{\AA}$  is the default value): the energy value in any  $C(x,y,z)$  point inside the cube is calculated starting from the eight corner points that surround it (see Fig. 4.3). The result of three linear interpolation is independent on the order of the interpolation steps along the three axes: any order, for instance along  $y$ , then along  $z$ , and finally along  $x$ , produces the same value.

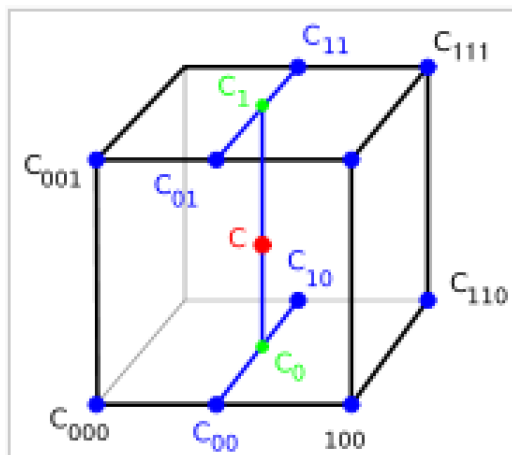


Figure 4.3: Depiction of 3D interpolation

### 4.3.7 FSA - Fast simulated annealing algorithm

Here and in the next two subsections are the FSA algorithm sequence instructions.

#### Generation of the 'i+1' conformation-pose during the *cooling phase*

The 'i+1' conformation-pose of the ligand is generated from the 'i' conformation-pose by varying the dihedral angles related to the previously detected *rotatable bonds* and by varying the parameter values that define the rototranslation inside the docking box.

All the dihedral angles and all rototranslation parameters are changed simultaneously; their new values are generated at random, but within a pre-established and temperature dependent range (see Subsection 4.3.9).

To effect the ligand rototranslation inside the docking box it is used the rototranslation operator:

$$Q = (q, p) \quad (4.13)$$

formed by the couple:

- $q = [s, x, y, z]$ , which is the quaternion representing the rotation inside the the docking box, where:
  - ◊ 's' is a scalar representing the rotation angle
  - ◊ 'x, y, z' are the components of the vector (versor) by which the rotation axis is identified
- $p = [p_x, p_y, p_z]$ , which is the vector  $\in R^3$  representing the translation inside the docking box

Quaternions [23] excel as a way of representing rotations of objects in three-dimensional space. They are economical to work with, both in terms of storage and computation; but more importantly they offer a clean conceptual framework which allows several problems involving rotations to be easily solved.

#### Some useful checks are made

- **Ligand integrity check:** for each 'i+1' conformation generated during the research of the minimum (in every *cycle* and in every *run*) a check on the intramolecular 'd' bond distances is carried out. If after the numerous dihedral rotations an excessive error (from a 'numerical' point of view) on any of the bond distances is introduced ( $|\delta d| > 10^{-3} \text{Å}$  in respect of the original bond distance), the actual 'i+1' conformation is rejected and a new 'i+1' conformation is generated starting from the 'i' conformation. This prevents the possible ligand fragmentation due to the propagation of computational errors.



- **Check on the sign of  $\Delta E_L$ :** to provide for the event that the optimizer (see Section 3.2) hasn't found the global minimum ligand conformation, the sign of:

$$\Delta E_L = E_{L,i+1_{\text{bound}}} - E_{L_{\text{unbound}}}$$

is evaluated in correspondence with each generated 'i+1' ligand conformation. In case of negative sign, the calculation is stopped and the procedure restarts from the beginning (*run 1, cycle 1*), by using as a new reference value for the unbound ligand the value  $E_{L,i+1_{\text{bound}}}$  of the just found ligand conformation.

That event, though remote, may happen since a stochastic (not a systematic) search algorithm (Fast Simulated Annealing) has been implemented in the 'Ligand optimizer' module (see Section 3.2).

The introduction of this check on the  $\Delta E_L$  makes the present module a kind of 'additional optimizer' of the ligand conformation, but not 'substitutive' of the module 'Ligand optimizer'.

In order to reduce the risk of a restart of the molecular docking calculation from the beginning (even several times), it is convenient to spend more time in the preliminary optimization of the ligand.

- **Check on ligand position:** all the ligand atoms of the new generated conformation-pose have to be within the docking box limits; in the contrary case, the actual conformation-pose is rejected and a new one is generated.
- **Check on intermolecular clashes:** the distance between each ligand atom and each receptor atom must be longer than or equal to the sum of the van der Waals radii relevant to the considered couple of atoms; in the contrary case, the actual conformation-pose is rejected and a new one is generated.
- **Check on the number of attempts:** to avoid that the program goes on and on searching a reliable conformation-pose, there is a threshold on the maximum number of attempts to be effected. In case the number of attempts overcomes the prefixed threshold of  $2 \cdot 10^5$ , the calculation is interrupted and an error message is printed.

### Calculation of $E_{\text{docking}}$

The docking energy  $E_{\text{docking}}$  associated to the 'i+1' conformation-pose is then calculated (see Section 4.1).

### 4.3.8 FSA - Metropolis criterion

The 'i+1' conformation-pose is accepted or rejected according to the 'Metropolis criterion':

- if  $E_{\text{docking}} < 0 \Rightarrow$  the new conformation-pose is accepted. A new conformation-pose is generated from the last one and the actual temperature is reduced by the quantity ' $\Delta T$ ' (*cooling rate*)
- if  $E_{\text{docking}} > 0 \Rightarrow$  the new conformation-pose is accepted with a probability defined by the following equation:

$$\begin{aligned}
 P_{\text{acceptance}} &= \frac{P_{i+1}}{P_i} = \\
 &= \frac{\frac{\exp(-\beta E_{\text{docking},i+1})}{Z}}{\frac{\exp(-\beta E_{\text{docking},i})}{Z}} = \\
 &= \frac{\exp(-\beta E_{\text{docking},i+1})}{\exp(-\beta E_{\text{docking},i})} = \\
 &= \exp(-\beta(E_{\text{docking},i+1} - E_{\text{docking},i})) = \\
 &= \exp(-\beta \Delta E_{\text{docking}})
 \end{aligned}$$

where:

- ◇  $P_{i+1}$  and  $P_i$  are the probabilities of the 'i+1' and 'i' conformation-poses according to the Boltzmann distribution
- ◇  $\beta$  is equal to  $1/kT$ , where  $k$  is the Boltzmann constant and  $T$  is the absolute temperature
- ◇  $Z$  is the partition function of the system

A *random* number 'r' is therefore generated in the range [0,1]:

- ▷ if  $0 \leq r \leq P_{\text{acceptance}}$ 
  - the 'i+1' ligand conformation-pose is accepted
  - the temperature is reduced by the quantity  $\Delta T$
  - a new ligand conformation-pose is generated starting from the conformation-pose just accepted
- ▷ if  $P_{\text{acceptance}} < r \leq 1$ 
  - the 'i+1' ligand conformation-pose is rejected
  - the temperature remains unchanged
  - a new ligand conformation-pose is generated starting from the previous 'i' conformation-pose

### 4.3.9 FSA - *Acceptance ratio* method

This module includes an *acceptance ratio* method for an optimal control of the ligand conformation-pose generation: the control is active inside each *cycle* during the whole *cooling* phase, from the initial temperature  $T_i$  to the final temperature  $T_f$ .

- Each new 'i+1' conformation is generated from the previously accepted 'i' conformation, through a simultaneous variation of all the dihedral angles 'j':

$$\vartheta_{i+1,j} = \vartheta_{i,j} + \delta\vartheta_{i,j} \quad \text{with } j = 1, 2, \dots, F \quad (4.14)$$

where:

- ◊ 'F' is the total number of dihedral angles of the ligand, coincident with its internal 'degrees of freedom' total number
- ◊ ' $\delta\vartheta_{i,j}$ ' is the dihedral angle variation, obtained by means of a random number generator, so as that:

$$\delta\vartheta_{i,j} \in [-\gamma, +\gamma] \quad (4.15)$$

where ' $\gamma$ ' can assume values between  $+30^\circ$  and  $+2^\circ$ , with a step variation of  $\pm 4^\circ$ , established as follows: inside each *cooling cycle*,

blocks of 500 new conformations are considered and is calculated the *ratio*:

$$a_r = \frac{\textit{accepted}}{500} \quad (4.16)$$

where '*accepted*' stands for the number of accepted conformation-poses according to the above mentioned Metropolis criterion. The  $\gamma$  value is so changed as follows:

- ▷ if  $a_r < 0.45 \Rightarrow$  the actual  $\gamma$  value is reduced by  $4^\circ$   
( $\gamma \rightarrow \gamma - 4^\circ$ )
- ▷ if  $a_r > 0.55 \Rightarrow$  the actual  $\gamma$  value is increased by  $4^\circ$   
( $\gamma \rightarrow \gamma + 4^\circ$ )

By doing so, it is possible to calibrate how to generate the ligand conformation-poses while searching the global minimum docking energy.

- The *acceptance ratio* method is also applied to the generation of poses: likewise the control on the range of the dihedral angle variations, the rototranslation too is modulated according to the temperature during the whole cooling phase.

The widths of the ranges within which are generated all parameters that define the rototranslation (see eq. 4.13) will vary according to the ' $a_r$ ' value (eq. 4.16).

#### 4.3.10 Execution of all *runs* and *cycles*

The program goes on executing the *runs* and *cycles* selected by the user, by means of two nested cycles on:

- ◇ number of *runs*
- ◇ number of *cycles*

up to the completion of the last *cycle* of the last *run*.

- Each *run* starts from a different ligand conformation-pose, generated through a random variation of the input structure, as described in Section 4.3.5. All *cycles* of the same *run* start from the same initial ligand conformation-pose.

- During the whole calculation, a temporary file containing the best ligand conformation-pose (PDB format) updated in real time is generated.
- To bring back the ligand from the grid to the receptor (PDB original file), an inverse rototranslation is effected in comparison to that of the module that generated the docking box (see Section 2.3): the docking box is built by the module 'Docking box generator' so that to obtain both a corner in the origin of the reference system and three edges along the x,y,z axes.
- Eventually, the file (PDB format) containing the best docked ligand structure (see Fig. 4.4) among all those generated in each *cycle* inside each *run* is printed on standard output (see Fig. 4.5).

In the first three rows of the output file are respectively printed:

- ◇ the calculated docking energy  $E_{\text{docking}}$  / [kcal/mol]
- ◇ the ligand-receptor interaction energy  $E_{\text{R-L}_{\text{interaction}}}$  / [kcal/mol]
- ◇ the ligand conformational energy variation  $\Delta E_{\text{L}}$  / [kcal/mol]

The output of the molecular docking calculation is intentionally synthetic so that it will occupy the minimum space on the disk, because this module is thought for the virtual high throughput screening (VHTS) of large ligand libraries.

## 4.4 Tests and discussion

The 'Semi flexible molecular docking' module, like the other modules for the receptor preparation and the ligand preparation were tested with a high number of ligand-receptor complexes, of which both the 3D-structures and the experimental dissociation constants are known.

Molecular docking calculations were performed at the same time with other molecular docking softwares (by using default parameters), such as AutoDock [3] and DOCK [27]. In **Appendix 1** are shown the docking energy values obtained in correspondence with 59 diverse complexes in a test set.

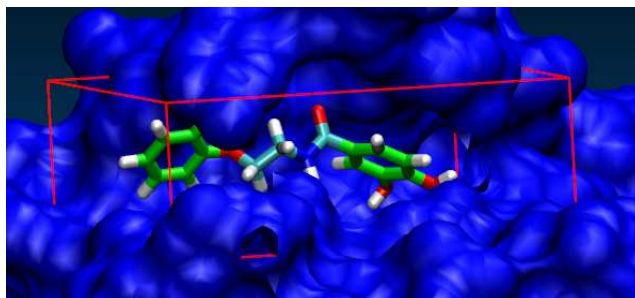


Figure 4.4: Docked ligand in the binding site

The same type of correlation between experimentally measured dissociation constant ( $K_{d,exp}$ ) and calculated dissociation constant ( $K_{d,calc}$ ) is obtained.

There is no evidence of substantial improvements in the calculation of the binding affinity. I expected this result, because, to a first approximation the AutoDock 3.0 scoring function coefficients were inserted in the scoring function of the present module. Having used a different approach to calculate the docking energy, it is now necessary to recalculate the coefficient values with a new linear regression analysis of complexes with known 3D-structures and known inhibition constants.

Docking Energy		/ [kcal/mol] = -10.7499	
Interaction Energy		/ [kcal/mol] = -11.1321	
Delta Ligand Energy		/ [kcal/mol] = +0.3822	
-----			
HET ATM	1	C1	0.280
HET ATM	2	H2	0.000
HET ATM	3	H3	0.000
HET ATM	4	H4	0.000
HET ATM	5	O5	-0.430
HET ATM	6	C6	0.659
HET ATM	7	O7	-0.570
HET ATM	8	C8	0.707
HET ATM	9	H9	0.000
HET ATM	10	N10	-0.853
HET ATM	11	H11	0.450
HET ATM	12	H12	0.450
HET ATM	13	H13	0.450
HET ATM	14	A14	-0.150
HET ATM	15	H15	0.150
HET ATM	16	A16	-0.150
HET ATM	17	H17	0.150
HET ATM	18	A18	-0.150
HET ATM	19	H19	0.150
HET ATM	20	A20	-0.150
HET ATM	21	H21	0.150
HET ATM	22	A22	-0.150
HET ATM	23	H23	0.150
HET ATM	24	A24	-0.143
END			

Figure 4.5: Example of an output file generated by the molecular docking module

## Chapter 5

# Applications

### 5.1 RAS: *in silico* study of a new class of Ras protein inhibitors

#### 5.1.1 Introduction

Ras, a key member in the super-family of small GTPases, is an essential component of signal transduction pathways that regulate cell growth, proliferation, differentiation and apoptosis [28][29]. Ras has been the subject of many pharmaceutical, genetic and biochemical studies [33] [34]. Among the many results, it was found that the proto-oncogene that codes for Ras is mutated in about 20-30% of human tumours [32]. Operating as molecular switches, Ras GTPases, assisted by guanine exchange factors (GEFs), undergo nucleotide exchange, allowing them to rapidly cycle from the inactive GDP bound state to the activated GTP bound state. The active Ras, aided by GTPase activating proteins (GAPs), hydrolyses the bound GTP to GDP and returns to the inactive GDP bound state [30]. Alteration of this deactivating reaction is a common biochemical defect associated with oncogenic Ras mutants [31]. Oncogenic Ras mutants exhibit decreased hydrolytic efficiency due to reduced affinities towards GAP proteins. Identifying small drug molecules that selectively interact with oncogenic Ras proteins and inhibit constitutive Ras activation is a challenging and a still largely unsolved task.



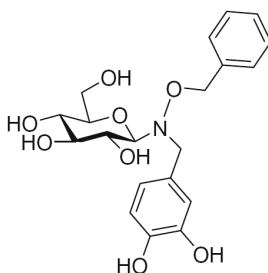


Figure 5.1: 'Compound 1' to be docked to Ras-GDP complex

To address this issue, many efforts were made to synthesize a new class of Ras protein inhibitors [38], composed of a sugar moiety or a linear spacer to which aromatic pharmacophore groups are covalently linked. These small organic molecules (whose MW ranges from 300 to 500 Da) bind to HRas-GDP with micromolar affinity and inhibit GEF-Ras interaction thereby preventing guanine nucleotide exchange [39].

The molecule 'compound 1' resulted to be experimentally the most promising because of its physical and chemical properties (see Figure 5.1): it was tailored specifically to overcome the insolubility issue associated with previously developed H-Ras inhibitors.

This work was concerned with the rationalisation of the binding modes of this and other compounds: combined techniques such as molecular dynamics and molecular docking were used to study the Ras protein inhibitors mechanism, taking into consideration also the 'induced fit' and the effect of solvent molecules [38] [39].

Despite the breadth of this study, here are reported only the descriptions on how the developed molecular docking modules (see Chapter 2, 3, 4) were here applied, besides a comparison with the software AutoDock 4.0.

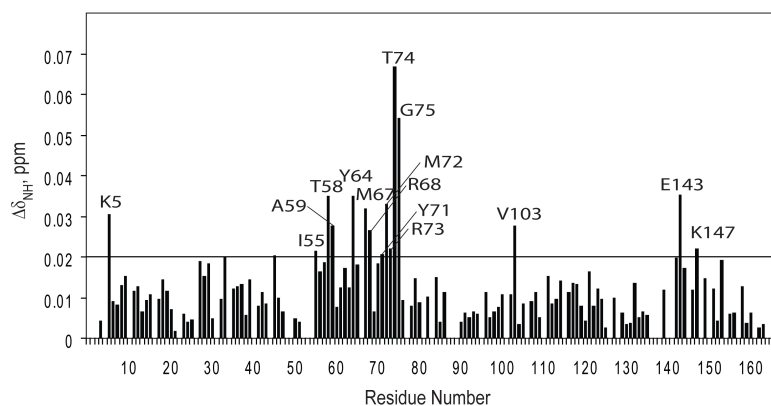


Figure 5.2:  $^{15}\text{N}$  HSQC graph of *H-Ras-GDP* complex of compound 1

### 5.1.2 Methods and results

#### Receptor preparation

- **Hydrogens addition:** the X-ray structure of the human Ras-GDP complex (PDB code: 4q21) was obtained from the Protein Data Bank and was protonated by using the module 'Add hydrogens tool' (see Section 2.1): polar hydrogens only were added ('united atom model'), having selected pH=7.
- **Binding site identification:** no structure of an inhibitor bound Ras-GDP complex is actually available. The only experimental information come from  $^{15}\text{N}$ -edited HSQC experiments [39] that allowed delineation of the Ras-ligand binding interface (see Figure 5.2).

The 'Receptorial site finder' module (see Section: 2.2) was applied to identify the putative protein binding sites. The cavity in close proximity to the 'Switch II' region (D57-E76), 'site 1' (see Figure: 5.3), was chosen to perform molecular docking calculations, because it was in accord with the results of the  $^{15}\text{N}$  HSQC experiments [39].

- **Docking box generation:** the 'Docking box generator' module (see

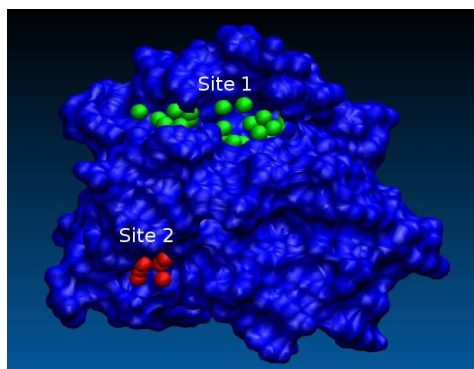


Figure 5.3: Cavities found by the 'Receptorial site finder' module on the Ras surface

Section: 2.3) was applied to build the docking box centred on 'site 1' ( $C \equiv (63.6, 80.7, 35.4)$ ). Its dimensions were set to:  $12.75 \times 12.75 \times 23.25 \text{ \AA}^3$  (see Figure: 5.4).

- **Potential energy grids maps generation:** the 'Potential energy grids calculator' module was applied to pre-calculate the 18 potential energy grid maps (see Section: 2.4). Grid spacing was set equal to  $0.375 \text{ \AA}$ .

### Ligand preparation

- **Ligand local optimization:** compound 1 was locally optimized by molecular mechanics, using the MMFF94 force field [35], since the initial ligand conformation should be realistic from a physical and chemical point of view (bond lengths and bond angles are to be correct as they are held fixed during the subsequent calculations). The model with only 'polar and aromatic' hydrogen atoms was used.
- **Ligand global optimization:** the 'Ligand optimizer' module (see Section: 3.2) was applied to compound 1 to perform a global optimization in the subspace of torsional degrees of freedom.

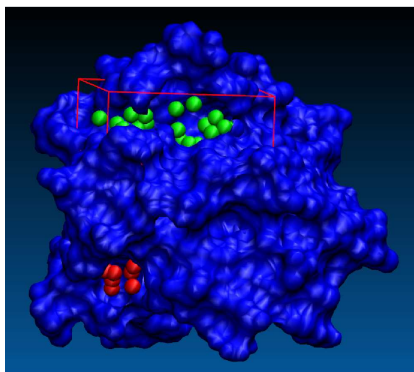


Figure 5.4: Docking box centred on 'site 1', generated by the 'Docking box generator' module

### Molecular docking simulation

The 'Semi flexible molecular docking' module (see Section: 4.1) was used to dock compound 1 to Ras-GDP complex: the torsional degrees of freedom of the ligand were explicitly considered, whereas the protein structure was kept frozen to X-ray atomic coordinates. The search for best conformation-pose was carried out using the Fast Simulated Annealing algorithm, using the following parameters:

- ◇ Number of *runs*: 50
- ◇ Number of *cooling cycles*: 5
- ◇ Initial temperature: 2000 K
- ◇ Final temperature: 5 K
- ◇ Cooling rate: 0.5 K

The best conformation-pose of the compound 1 was singled out (see Figure 5.5) having a docking energy equal to:  $E_{\text{docking}} = -9.2$  kcal/mol, with a conformational energy variation when passing from unbound to docked equal to:  $\Delta E_{\text{docking}} = +1.07$  kcal/mol.

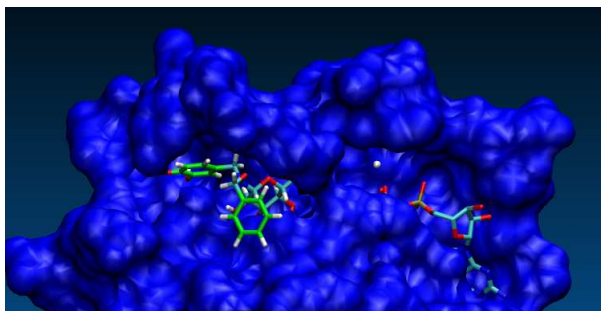


Figure 5.5: Compound 1 (on the left) docked to Ras-GDP complex, as a result of the calculation performed with the 'Semi flexible molecular docking' module (see Chapter 4)

Semi-flexible molecular docking calculations were also performed with the software AutoDock 4.0 starting from the same initial globally optimized conformation of the ligand, and by using exactly the same parameters to build the grid and the docking box. The search for best conformation-poses was carried out using a Lamarckian Genetic Algorithm (LGA), and using the following parameters:

- ◇ Number of *runs*: 100
- ◇ Population size: 200
- ◇ maximum number of evaluations: 3500000

Best poses from each run were clustered using an RMSD tolerance of 2.0 Å (see Figure 5.6)

The best conformation-pose of the first cluster has a docking energy equal to:  $E_{\text{docking}} = -6.41$  kcal/mol.

Although the docking energy values obtained by the two algorithm are different, -9.2 kcal/mol and -6.41 kcal/mol respectively, two similar results (see Figure 5.7) were obtained in respect of the prediction of the best conformation-poses (RMSD = 1.1 Å).

The best AutoDock 4.0 conformation-pose of the second cluster has a docking energy equal to:  $E_{\text{docking}} = -5.91$  kcal/mol and has a RMSD = 7.71 Å relative to the best conformation-pose of the first cluster (see Figure 5.8).

Cluster Rank	Lowest Binding Energy	Run	Mean Binding Energy	Num in Cluster	Histogram							
					5	10	15	20	25	30	35	
1	-6.41	14	-4.68	12	#####							
2	-5.91	16	-4.76	17	#####	#####						
3	-5.16	10	-4.63	2	##							
4	-4.83	21	-4.19	3	###							
5	-4.80	64	-3.73	2	##							
6	-4.64	7	-4.42	2	##							
7	-4.62	3	-3.67	12	#####							
8	-4.59	29	-4.47	2	##							
9	-4.47	62	-3.68	3	###							
10	-4.37	60	-4.02	3	###							
11	-4.35	86	-3.80	4	####							
12	-4.22	26	-4.22	1	#							
13	-4.01	66	-4.01	1	#							
14	-3.85	90	-3.85	1	#							
15	-3.83	34	-3.07	5	####							
16	-3.77	77	-3.14	2	##							
17	-3.67	44	-3.67	1	#							
18	-3.67	27	-2.77	4	####							
19	-3.52	80	-3.08	3	###							
20	-3.45	71	-3.45	1	#							
21	-3.44	20	-3.44	1	#							
22	-3.40	94	-3.40	1	#							
23	-3.23	47	-3.23	1	#							
24	-3.11	1	-3.11	1	#							
25	-2.54	42	-2.54	1	#							
26	-2.48	97	-2.10	4	####							
27	-2.48	6	-2.48	1	#							
28	-2.38	4	-2.38	1	#							
29	-2.29	41	-2.29	1	#							
30	-2.24	30	-2.24	1	#							
31	-2.03	50	-1.91	2	##							
32	-1.87	35	-1.87	1	#							
33	-1.85	85	-1.85	1	#							
34	-1.81	82	-1.81	1	#							
35	-1.45	18	-1.45	1	#							

Figure 5.6: Clustering histogram AutoDock output: it is the result of 100 runs, using an RMSD tolerance of 2.0 Å

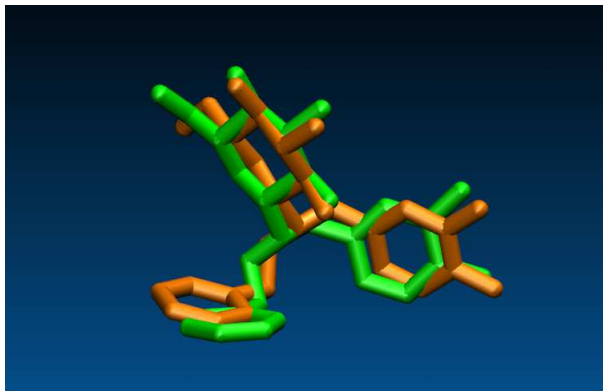


Figure 5.7: Convergence of the results: *green* for the first AutoDock conformation-pose, *orange* for the output of the 'Semi flexible molecular docking' module

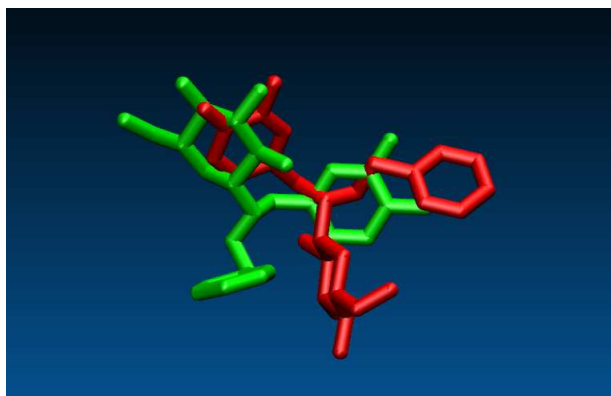


Figure 5.8: The best first (green) and second (red) AutoDock clusters conformation-poses

The same molecular docking calculations (with AutoDock 4.0) were also performed starting from an initial ligand conformation that was not globally optimized in the torsional subspace. The result was an inversion in the order of the best conformation-poses of the first two clusters.

This is a confirmation of the necessity to have a unique procedure for the ligand preparation, as I maintain in Chapter 3: independently on the procedure that I proposed, it should be avoided that the results will depend on the choice of the user.

Since the docking energy difference between the two conformation-poses obtained by AutoDock is low, about 0.5 kcal/mol, in the prosecution of the research both of them were considered as potentially reliable and were used as starting point for molecular dynamics simulations in explicit solvent [39]. The most representative protein site conformations were selected and used for structural analysis (see Figure 5.9).



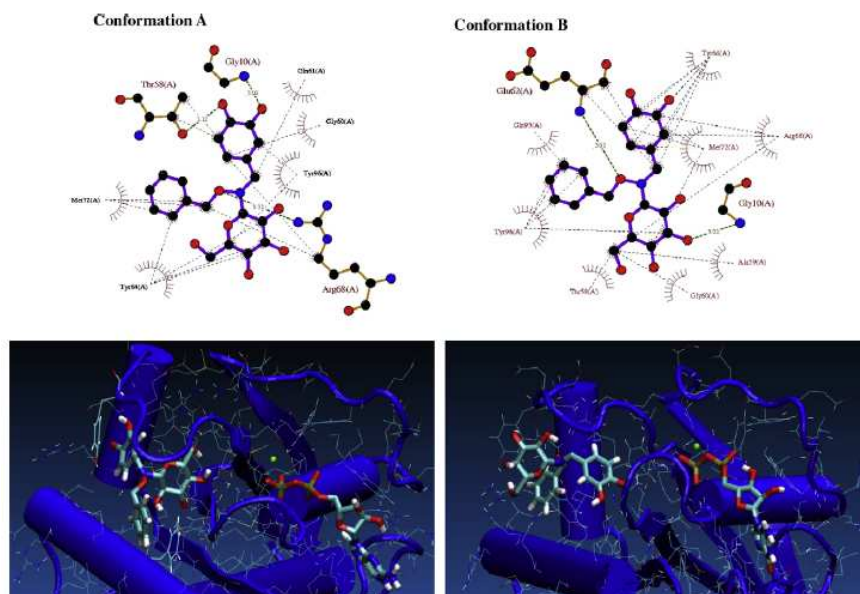


Figure 5.9: Contact maps and 3D representations of the two most representative lowest-energy protein-ligand complexes, as obtained from MD calculations starting from the best docking results of compound 1.

## 5.2 Virtual Protein Engineering: construction and ranking of a protein mutants library

### 5.2.1 Introduction

The research aims at a transposition *in silico* of the methodology used in similar researches conducted in biochemical laboratories. In particular, reference is made to the work appeared in 2004: "Production of a fully functional, permuted single-chain penicillin G acylase" [36].

PGA is an industrially important enzyme used to produce 6-aminopenicillanic acid (6-APA) from penicillin G during the manufacturing of semisynthetic antibiotics. Crystal structures of PGA are known for the dimeric and substrate-bound forms of the enzyme (Duggleby et al. 1995; Done et al. 1998; Alkema et al. 2000; Hewitt et al. 2000; McVey et al. 2001).

The object of this work was to attempt the construction of a single-chain PGA by joining the  $\alpha$  and  $\beta$ -subunits with a short peptide and generating new N and C termini: if functional, this polypeptide arrangement would no longer depend on autoproteolytic processing.

This idea can be realized since the structure of dimeric PGA (Duggleby et al. 1995) shows that the C terminus of the  $\beta$ -subunit is nearby the N terminus of the  $\alpha$ -subunit.

The present work deals with the GA protein: 'glutaryl 7-aminocephalosporanic acid acylase' (PDB code: '1jvz'), that catalyzes the hydrolysis of 'glutaryl 7-aminocephalosporanic acid' to give '7-aminocephalosporanic acid' and 'glutaric acid'.

The first step consists of analysing the three-dimensional structure of the crystal and choosing the precise points on the  $\alpha$  and  $\beta$  chains to make a junction by means of a short peptide the length of which to be defined. Once a single-chain reference structure is generated and a refinement has been made, many other mutant structures are to be produced by varying at random the linker residues, so called ' $\beta$ -linker- $\alpha$ '.

Subsequently, the effects on the binding affinity have to be evaluated: a selection will sort out the  $\beta$ -linker- $\alpha$  structures having the best affinity to the reagent and the lower affinity to the products. This is made to favour the reaction catalyzed by the permuted enzyme.

## 5.2.2 Methods

An *ad hoc* tool was written to run the various molecular docking software modules. The work was done according to the following summarized procedure.

### Construction and refinement of the single-chain structure

The starting point of the whole work has been the analysis of the crystal structure of the *wild type* GA protein (PDB code: 1jvz). It is formed by two subunits, respectively formed by 152 residues ( $\alpha$ ) and 520 residues ( $\beta$ ) and it is in complex with the ligand: 'glutaryl 7-aminocephalosporanic acid' (CEN) (see Fig: 5.10)

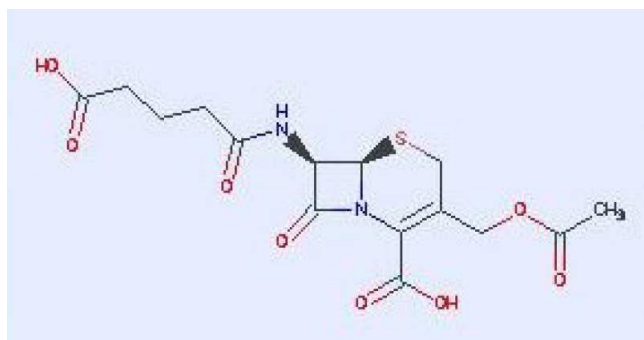


Figure 5.10: Reagent: glutaryl 7-aminocephalosporanic acid

To obtain a highly reliable reference structure, it was performed an initial refinement of the three-dimensional structure by means of a short (5.0 ns) explicit solvent molecular dynamics simulation in the NPT ensemble at 293.15 K, followed by the minimization of the most stable conformations. All optimizations and the MD simulation were performed using the GROMOS96 force field [43], as included in version 3.3.3 of the GROMACS package.

An accurate structural analysis of the reference conformation was made in the region where the N terminal of  $\alpha$ -chain and C terminal of  $\beta$ -chain are located.

On the basis of the reference work [36], I decided to attempt the construction of the single chain by removing some residues of the two subunits and by joining the new C and N termini with a tetra-peptide linker; I proceeded through the following steps by using the software MOE:

- ◇ removal of the (-688-689) C-terminal residues of the  $\beta$ -chain
- ◇ removal of the (7-8-9-10-11-12-13-14-15-16-17-) N-terminal residues of the  $\alpha$ -chain. Note: the PDB file does not include the residues from 1 to 6 of the  $\alpha$ -chain.

The distance between the ends of the two chains to be connected is now about 6.5 Å

- ◇ generation of a first tetrapeptide formed by the N,E,G and M residues.
- ◇ connection of the two  $\beta$  and  $\alpha$  chains: " $\beta$ -chain-[linker]- $\alpha$ -chain" (see Figure 5.11), i.e:  
N-terminal-...-F687-[N688-E689-G690-M691]-N18-...-C-terminal
- ◇ progressive reenumeration of the residues of the just generated single chain

### Generation of the docking box

The module 'Receptorial site finder' (see Section 2.2) has been applied to the single-chain reference structure to identify the coordinates of the centre of the binding-site. By means of the 'Docking box generator' module (see Section 2.3), the parameters to generate the docking box have been calculated (the box dimension was set to 21x21x21 Å<sup>3</sup>). The same parameters are used for all the library mutants on which molecular docking calculations will be performed.

### Preparation of the ligands: reagent and products

The files of the three ligands to be docked:

- ◇ glutaryl 7-aminocephalosporanic acid [CEN] (the reagent)

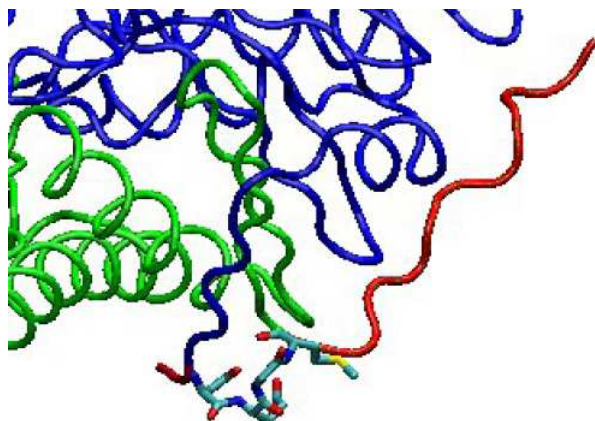


Figure 5.11:  $\alpha$ -chain (*green*) connected to  $\beta$ -chain (*blue*) with -NEGM- linker. In *red* are the removed residues.

- ◇ 7-aminocephalosporanic acid [7-ACA] (product 1)
- ◇ glutaric acid (product 2)

were prepared according to the procedure described in Chapter 3.

### Generation of the mutants library

Considering the 20 natural aminoacids, the total number of possible tetrapeptides that can be generated is  $20^4$ .

Among all these, a group of 1405 tetrapeptides were chosen at random to generate the correspondent mutant models of the single-chain. To create them, the MODELLER (version 8.1) software package [40] [41] was used with default input parameters.

Hydrogen atoms were added (see Section 2.1) to each  $\beta$ -linker- $\alpha$  model and the 'potential energy grids maps' were calculated (see Section 2.4) for each of them.

### Docking of the reagent

By means of the 'Semi flexible molecular docking' module (see Chapter 4), the reagent molecular docking simulations were effected on all the 1405 mutant structures.

The results are shown in **Appendix 2-A**: for each  $\beta$ -linker- $\alpha$  model the correspondent docking energy is indicated; rows are ordered by increasing docking energy.

### Docking of the products

The best 500 mutant models were chosen among the 1405 previously selected. By means of the 'Semi flexible molecular docking' module (see Chapter 4), the products of the reaction too were docked on these 500 mutant models. The results are shown in **Appendix 2-B**, where are listed the docking energies of the reagent and of the two products for each  $\beta$ -linker- $\alpha$  structure.

### Ranking of $\beta$ -linker- $\alpha$ structures

To select the  $\beta$ -linker- $\alpha$  structures with the best affinity towards the reagent and, at the same time, the lower affinity towards the products, the 500 models were ordered maximizing the function:

$$f = (-)E_{d,\text{reagent}} + E_{d,\text{product 1}} + E_{d,\text{product 2}}$$

In Table 5.1 are listed the best  $\beta$ -linker- $\alpha$  structures with the corresponding 'f' values, ordered according to decreasing values of the f.

The best five  $\beta$ -linker- $\alpha$  structures of Table 5.1 were selected for which both the reagent (see Figure 5.12) and the products poses in the binding site are consistent with the orientation of the reagent in wild-type enzyme:

- ◇ NPLR
- ◇ NPRR
- ◇ PALL
- ◇ NDVA
- ◇ DEAA

Finally, on the best five structures further molecular dynamics calculations were made, in order to evaluate the protein thermal stability and possible local rearrangements both at the binding site level and in the linker region.

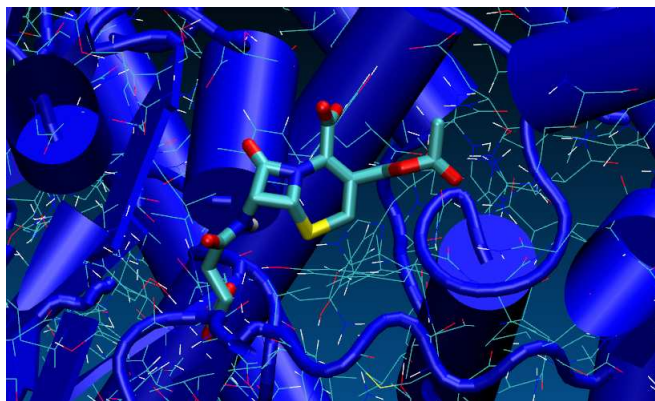


Figure 5.12: Docked reagent (CEN) on the best  $\beta$ -linker- $\alpha$  (NPLR)

### 5.2.3 Conclusions

To face the virtual protein engineering problem described in this section, all the developed molecular docking software modules (see Chapter 2,3,4) were used, besides some additional tools created *ad hoc* to run the various programs.

The virtual screening of a protein mutants library represents the inverse of the traditional molecular docking problem, in as much as, in this case, the best mutant for a fixed ligand is searched.

The models utilized to calculate the binding affinities variations due to the introduction of a  $\beta$ -linker- $\alpha$  are very simplified in respect of the real case. Nevertheless it has been possible to screen a reduced library, but however important, in respect of the total number of possible combinations ( $20^4$ ).

Any other higher level computational method would have been too much expensive from the point of view of the long computational time. A more refined treatment was given only to a small number of  $\beta$ -linker- $\alpha$  structures, namely to those resulted from the first phase of screening. Therefore, the combination of the two techniques in a protocol where docking is used for the fast screening of large libraries and MD simulations are then applied to explore conformations of the final complexes is a logical approach [42] to improve the protein engineering process.



$\beta$ -linker- $\alpha$ structure	'f' value / (kcal/mol)
QDGK	-0.92
NPLR	-0.99 *
NPRG	-1.4
NPRR	-1.49 *
PALL	-1.51 *
NDVA	-1.58 *
AGSM	-1.69
DEAA	-1.69 *
RDRM	-1.78
RAQR	-1.87
AEQV	-1.91
QGQG	-1.92
PEVL	-1.93
AARA	-1.93
GDVR	-1.93
ADAK	-1.95
QDSS	-1.97
NEGS	-1.98
AAGG	-2.00
PDSS	-2.06
GDQD	-2.06
PGSD	-2.06
PPRD	-2.07
AEER	-2.08
PGLD	-2.09
...	...
...	...

Table 5.1: Top of the list of the 500  $\beta$ -linker- $\alpha$  structures ordered according to decreasing values of the 'f' function

## Chapter 6

# Conclusions

Research and development of a drug is a long and very expensive process at high risk of failure. *In silico* techniques, such as virtual screening or rational drug design, are now routinely used to guide drug discovery. Thanks to the recent increase of performance computing systems, the first three steps of the drug development process have been supported by techniques that allow virtual experiments to accelerate the process of lead identification and optimization: 'Computer-Aided Drug Design' (CADD) has an increasingly important role in simulating drug-receptor interactions. While HTS identifies lead molecules by performing individual biochemical assays with over millions of real compounds, the more recent vHTS, that is a computational screening method, reduces false positives in HTS so lowering both costs and time in drug discovery.

Protein-ligand docking aims to predict and rank the structures arising from the association between a given ligand and a target protein of known 3D-structure. Despite the breathtaking advances in the field over the last decades and the widespread application of docking methods, several downsides still exist. Scoring functions normally used in docking programs make a number of simplifications and assumptions to allow a more computationally efficient evaluation of ligand affinity, but naturally at the cost of accuracy. Moreover, the molecular docking suffers from the problem of instability. In addition to the scoring problem, a number of other aspects bring additional complexity to the resolution of the general docking problem. The solvent

effect and the direct participation of water molecules in protein-ligand interactions, the limited resolutions of most crystallographic targets, and of course, protein flexibility, both in terms of intrinsic structural flexibility and in terms of conformational alterations upon ligand binding, are some of the most relevant ones. I contributed to develop (in C++ programming language) a new molecular docking software ('Semi-flexible model') to be used for vHTS: my reference point was software AutoDock 3.0, about which the way the ligand is treated is to my opinion criticizable. I therefore introduced some changes that have substantially modified the physical model of the problem, such as a modified scoring function, an univocal criterion to calculate the ligand conformational variation energy and a Fast Simulated Annealing (FSA) algorithm including an *acceptance ratio* method for an optimal control of the ligand conformation-pose generation. A number of checks are performed during the execution of the program.

To a first approximation, the AutoDock 3.0 scoring function coefficients were inserted in the scoring function of the semi-flexible molecular docking module. Having used a different approach to calculate the docking energy, to recalculate the coefficient values with a new linear regression analysis of complexes with known 3D-structures and known inhibition constants is now necessary.

The developed software has already been applied to the study of various real problems in the field of drug discovery, to study a new class of oncogenic Ras protein inhibitors, and of Virtual Protein Engineering, to design a modified industrially important enzyme used to better catalyze the hydrolysis of 'glutaryl 7-aminocephalosporanic acid' to give '7-aminocephalosporanic acid' and 'glutaric acid'.

In consideration of the high number of fields in which molecular docking softwares can be applied, it is worth while to invest to further ameliorate their performances. Progress can be achieved trying to enhance both the chemical-physical model of the complex problem, and the optimization of the algorithms.

# Publications

This thesis has led to the following publications:

**Chapter 5:** Palmioli A., Sacco E., Airoidi C., Di Nicolantonio F., D'Urzo A., Shirasawa S., Sasazuki T., Di Domizio A., De Gioia L., Martegani E., Bardelli A., Peri F., Vanoni M. **Selective cytotoxicity of a bicyclic Ras inhibitor in cancer cells expressing  $K-Ras^{G13D}$ .** *Biochemical and Biophysical Research Communications*, vol.386(4): 593-597; September 2009.

**Chapter 5:** Palmioli A., Sacco E., Abraham S., Thomas C.J., Di Domizio A., De Gioia L., Gaponenko V., Vanoni M., Peri F. **First experimental identification of Ras-inhibitor binding interface using a water-soluble Ras ligand.** *Bioorganic & Medicinal Chemistry Letters*, vol.19(15): 4217-4222; August 2009.



# Bibliography

- [1] Subramaniam S; Mehrotra M; Gupta D *Bioinformatics* 2008,3(1):14-17
- [2] Sousa SF; Fernandes PA; Ramos MJ *PROTEINS: Structure, Function, and Bioinformatics* 2006, 65:15-26
- [3] Morris GM; Goodsell DS; Halliday RS; Huey R; Hart WE; Belew RK; Olson AJ *J. Computational Chemistry* 1998,19:1639-1662
- [4] Park H; Lee J; Lee S *Proteins* 2006,65:549-554
- [5] Hetenyi C; vanderSpoel D *Protein Sci.* 2002,11:1729-37
- [6] Ruvinsky AM; Kozintsev AV *J. Comput. Chem.* 2005,26:1089-95
- [7] Hetenyi C; Paragi G; Maran U; Timar Z; Karelson M; Penke B *J. Am. Chem. Soc.* 2006,128:1233-1239
- [8] Moitessier N; Westhof E; Hanessian S *J. Med. Chem.* 2006,49:1023-1033
- [9] Jones G; Willett P; Glen RC; Leach AR; Taylor R *J. Mol. Biol.* 1997,267:727-48
- [10] Rarey M; Kramer B; Lengauer T; Klebe G *J. Mol. Biol.* 1996,261:470-489
- [11] Abagyan R; Totrov M; Kuznetsov D *J. Comput. Chem.* 1994,15:488-506
- [12] Wang R; Lu Y; Wang S *J. Med. Chem.* 2003,46:2287-2303
- [13] Wang R, Lai L, Wang S *Journal of Computer-Aided Molecular Design* 2002,16:11-26

- [14] Yevick D; *Computational Physics and Object Oriented Programming with C++* 2006
- [15] Duan Y; Wu C; Chowdhury S; Lee MC; Xiong G; Zhang W; Yang R; Cieplak P; Luo R; Lee T; Caldwell J; Wang J; Kollman P *J. Comput. Chem.* 2003,24:1999-2012
- [16] Harriman DJ; Deslongchamps G *Journal of Computer-Aided Molecular Design* 2004,18(5):303-308
- [17] Davis AM; Teague SJ; Kleywegt GJ *Angewandte Chemie International Edition* 2003,42(24):2718-2736
- [18] Brady GP Jr; Stouten PF *J. Comput. Aided Mol. Des.* 2000,14(4):383-401
- [19] Mathews Christopher K.; Van Holde Kensal E.; Ahern Kevin G. *Biochimica* 2004
- [20] Luty BA; Wasserman ZR; Stouten PF; Hodge NC; Zacharias M; McCammon AJ *J. Comput. Chem* 1995,16(4):454-464
- [21] Verkhivker GM; Rejto PA; Bouzida D; Arthurs S; Colson AB; Freer ST; Gehlaar DK; Larson V; Luty BA; Marrone T; Rose PW *J. Mol. Recogn* 1999,12(6):371-389
- [22] Mehler EL; Solmeyer T *Protein Eng.* 1991,4(8):903-910
- [23] Charles F.F. Karney *Journal of Molecular Graphics and Modelling* 2007,25:595-604
- [24] Gasteiger HJ; Marsili M; *Tetrahedron Lett* 1978,34:3181-3184
- [25] Dewar MJS; Thiel W *J. Am. Chem. Soc.* 1977,99:4899-4907
- [26] Rathod HT; Sridevi K *Comput. Methods Appl. Mech. Engrg.* 2001,190:3325-3368
- [27] Ewing TJA; Makino S; Skillman AG; Kuntz ID *J. Comput.-Aided Mol. Des* 2001,15:411-428
- [28] Barbacid M. *Annu. Rev. Biochem.* 1987,56:779

- [29] Lowy DR; Willumsen BM *Annu. Rev. Biochem.* 1993,62:851
- [30] Bourne HR; Sanders DA; McCormick F *Nature* 1989,341:209
- [31] Bos JL; Rehmann H; Wittinghofer A *Cell* 2007,129:865
- [32] Wittinghofer A; Waldmann H; *Angew. Chem.* 2000,112:4360-4383  
*Angew. Chem. Int. Ed.* 2000,39:4192-4214 and references cited therein
- [33] Peri F; Airoidi C; Colombo S; Martegani E; van Neuren AS; Stein M; Marinzi C; Nicotra F *Chembiochem* 2005,6(10):1839-48
- [34] Peri F; Airoidi C; Colombo S; Mari S; Barbero JJ; Martegani E; Nicotra F *Eur. J. Org. Chem.* 2006,16:3707-3720
- [35] Halgren TAJ *J. Comput. Chem.* 1996,17:490-519
- [36] Flores G; Sobern X; Osuna J *Protein Sci* 2004,13(6): 1677-1683
- [37] Nelder JA; Mead R; *Computer Journal* 1964,7: 308-313
- [38] Palmioli A; Sacco E; Airoidi C; Di Nicolantonio F; D'Urzo A; Shirasawa S; Sasazuki T; Di Domizio A; De Gioia L; Martegani E; Bardelli A; Peri F; Vanoni M *Biochemical and Biophysical Research Communications* 2009,386(4): 593-597
- [39] Palmioli A; Sacco E; Abraham S; Thomas CJ; Di Domizio A; De Gioia L; Gaponenko V; Vanoni M; Peri F *Bioorganic & Medicinal Chemistry Letters* 2009,19(15): 4217-4222
- [40] Eswar N; Marti-Renom MA; Webb B; Madhusudhan MS; Eramian D; Shen M; Pieper U; Sali A *Current Protocols in Bioinformatics* 2006,15:5.6.1-5.6.30
- [41] Marti-Renom MA; Stuart A; Fiser A; Sanchez R; Melo F; Sali A *Biophys. Biomol. Struct.* 2000,29: 291-325
- [42] Alonso H; Bliznyuk AA; Gready JE *Med. Res. Rev.* 2006,26(5):531-68
- [43] Scott WRP; Hunenberger PH; Tironi IG; Mark AE; Billeter SR; Fennen J; Torda AE; Huber T; Kruger P; van Gunsteren WF *J. Phys. Chem. A* 1999,103:3596-3607





## ACKNOWLEDGEMENTS

Dedico questo lavoro ai **miei genitori** per l'amore, l'aiuto ed il sostegno infiniti che non mi hanno mai fatto mancare. Non saró mai abbastanza riconoscente!

Al termine di questo ciclo di studi molto importante e fondamentale per la mia vita futura e che mi ha dato la possibilitá di crescere grazie ad un'esperienza significativa nel meraviglioso campo della ricerca, nei settori della chimica, fisica, matematica, biologia e sviluppo di software, desidero ringraziare tutti coloro che credono nella ricerca contro le malattie e che la sostengono in quanto valore universale per migliorare le condizioni di vita umane.

Ringrazio il mio *supervisor*, Prof. **Piercarlo Fantucci**, per la sua professionalitá e per la fiducia che mi ha sempre dimostrato in questi anni di studio.

Ringrazio la mia coordinatrice, Prof.ssa **Franca Morazzoni**, che ha sempre contribuito affinché i lavori venissero svolti nel migliore modo possibile.

Ringrazio il Prof. **Luca De Gioia** che, oltre a svolgere con estrema dedizione il suo ruolo professionale molto importante, si é sempre dimostrato anche un ottimo amico.

Ringrazio il Prof. **Francesco Peri** per l'entusiasmo e la grinta che mi ha trasmesso per approfondire temi importanti riguardanti la ricerca applicata allo studio di potenziali farmaci antitumorali.

Ringrazio la Dott.ssa **Noura Gaiji** per la Sua costante e fattiva collaborazione.

Ringrazio il Dott. **Luca Bertini** per l'amicizia e la possibilitá di confronto scientifico che mi ha sempre dato.

Ringrazio la Prof.ssa **Laura Bonati** per la passione per la chimica computazionale che mi ha trasmesso fin dai tempi del corso di laurea in Chimica. Ringrazio il Dott. **Marco Palumbo** per la grande dedizione al lavoro di ricerca e gli aiuti determinanti che mi ha fornito.

Ringrazio la Dott.ssa **Cristina Airoidi** ed il Dott. **Alessandro Palmioli** per la serietà dimostrata nel portare avanti difficili progetti di ricerca.

Ringrazio, inoltre, tutto lo staff del laboratorio di modellistica molecolare che è sempre stato caratterizzato da un'atmosfera serena e professionale: Dott. **Giuseppe Zampella**, Dott. **Claudio Greco**, Dott. **Marco Pasi**, Dott.ssa **Elena Papaleo**, Dott. **Alessandro Vitriolo**, Dott.ssa **Francesca Falcetta**, Dott.ssa **Manuela Mattara** e tutti gli altri amici e colleghi del laboratorio.

Inoltre, un ringraziamento particolare a mia **zia Isa** per la costante vicinanza ed affetto da sempre dimostrato.

Ringrazio **Elisa**, una persona per me molto speciale ed importante che mi è sempre stata vicina e sostenuto trasmettendomi entusiasmo anche nei momenti difficili.

Un ringraziamento particolare ai miei migliori amici: **Marco**, per gli incoraggiamenti e la massima disponibilità, **Paul** e **Luigi** e a tutti i colleghi di studio della biblioteca.

# Appendix 1

<b>PDB Complex</b>	<b>Dock: Ed /[kcal/mol]</b>	<b>SF D. module: Ed /[kcal/mol]</b>	<b>AutoDock: Ed /[kcal/mol]</b>
1ABE	-10.46	-8.68	-5.92
1ABF	-10.4	-8.69	-6.64
1AOE	-10.05	-10.71	-7.56
1BYG	-14.1	-16.88	-8.63
1C5C	-15.04	-8.88	-10.34
1CBX	-22.65	-14.82	-12.3
1COM	-12.56	-7.08	-7.21
1CPS	-13.74	-15.34	-9.67
1DBB	-10.62	-11.79	-9.05
1DBJ	-11.43	-13.17	-9.21
1DG5	-10.08	-13.17	-6.48
1DID	-9.49	-13.61	-5.11
1DOG	-12.6	-11.49	-7.44
1DR1	-11.65	-11.31	-6.74
1EBG	-23.91	-10.79	-20.13
1FLR	-12.76	-11.39	-9.93
1HDC	-11.3	-9.89	-9.36
1IMB	-18.83	-13.41	-5.62
1LDM	-6.23	-4.5	-5.24
1MDR	-8.94	-10.8	-11.25
1MRK	-11.53	-11.3	-7.3
1PDZ	-11.92	-8.42	-7.4
1RNT	-10.72	-10.62	-5.75
1TNG	-8.68	-8.23	-6.03
1TNH	-8.48	-7.18	-5.68
1TNI	-9.44	-11.73	-6.22
1TNL	-7.32	-8.97	-6.1
1ULB	-8.15	-6.99	-5.35

2AK3	-15.23	-9.08	-7.5
2CMD	-20.9	-7.74	-8.19
2CPP	-7.12	-9.09	-6.48
2CTC	-13.49	-9.43	-9.46
2DBL	-12.74	-11.81	-10.35
2GBP	-11.84	-7.53	-7.24
2H4N	-15.88	-8.12	-5.05
2MCP	-7.82	-6.87	-3.79
2PHH	-7.83	-8.03	-5.74
2PK4	-6.67	-11.26	-6.55
2TMN	-13.19	-9.8	-5.79
2YPI	-11.41	-5.24	-4.65
3GPB	-13.37	-10.88	-5.47
5ABP	-11.47	-9.39	-7
5CPP	-8.28	-8.53	-7.05
6RNT	-8.77	-10.53	-4.13
7TIM	-11.87	-6.78	-4.93
1CKP	-9.16	-11.67	-3.72
1OKL	-12.35	-10.03	-6.49
1PHG	-5.14	-14.11	-8.06
1FOR	-13.79	-15.75	-10.39
1FOS	-13.24	-15.18	-8.8
1FKI	-11.33	-11.57	-9.55
1LAH	-15.76	-10.12	-9.19
1QPE	-11.11	-12.72	-7.07
1YDR	-11.49	-15.32	-8.17
2CHT	-6.23	-8.39	-6.26
2PCP	-9.56	-11.82	-4.41
1ACO	-27.14	-11.99	-11.98
1HYT	-11.24	-12.16	-11.16
1SNC	-24.83	-11.45	-14.1

# Appendix 2-A

Docking of the reagent: Glutaryl 7-aminocephalosporanic acid  
to: 1405 "B-linker-A" structures

Column 1: B-linker-A structure  
Column 2: Docking energy (Ed)

Note: rows are ordered by increasing docking energy (Ed)

B-linker-A structure	Ed / (kcal/mol)
NPLR	-10.27
DPRG	-10.19
NPRR	-10.13
RGLA	-10.04
PDSS	-10.03
QDSS	-10.02
GGAV	-9.95
PALL	-9.9
EPED	-9.86
RAGG	-9.81
QDGK	-9.81
DAQK	-9.76
AARV	-9.76
DEVL	-9.73
RPAR	-9.72
EERR	-9.72
DEAA	-9.7
PEAK	-9.69
PPSD	-9.67
PEVR	-9.65
RAQR	-9.65
QGQG	-9.64
DALM	-9.64
QEGL	-9.62
GGVS	-9.62
GPED	-9.61
RDAV	-9.61
GARS	-9.6
DGQK	-9.6
NGAD	-9.59
EDRA	-9.57

NPAV	-9.54	DEVV	-9.22	RPRK	-9.09
NPRK	-9.54	QEVs	-9.22	REQD	-9.09
PPRD	-9.53	PDAK	-9.21	PEVK	-9.09
EESK	-9.53	QEVA	-9.2	QGVV	-9.09
GDVG	-9.49	PGLM	-9.2	APGV	-9.08
PGLG	-9.49	QARK	-9.2	AELD	-9.08
RDRM	-9.48	QEAK	-9.2	QGVs	-9.08
RGSS	-9.47	PGAV	-9.19	DPLK	-9.08
DPAG	-9.46	NESR	-9.19	APQL	-9.07
RPEG	-9.46	NDVD	-9.18	DELA	-9.07
GPRG	-9.45	AASV	-9.18	PGEK	-9.07
RAVR	-9.45	EPVA	-9.18	QPSG	-9.06
EPGA	-9.44	RGAK	-9.18	GDVR	-9.06
QDEV	-9.43	DESM	-9.18	AASS	-9.06
NGEG	-9.42	NGRA	-9.17	PERG	-9.06
ADLK	-9.42	RESV	-9.17	EPVK	-9.05
DEQS	-9.42	NEAK	-9.17	DGVM	-9.05
NAEV	-9.41	GDSV	-9.16	QPLD	-9.05
QPLK	-9.4	DESK	-9.15	PDVL	-9.05
PPED	-9.4	QGLK	-9.15	GPQS	-9.05
AAGG	-9.4	QPED	-9.15	GDAD	-9.05
RALD	-9.37	DPVV	-9.15	RPVS	-9.05
RERS	-9.35	EERA	-9.15	RDAA	-9.04
NPVD	-9.35	RDLG	-9.14	AEEA	-9.04
GDQM	-9.35	ADAL	-9.14	AEVV	-9.04
GGEL	-9.34	EDQA	-9.14	DPAM	-9.03
GEGV	-9.34	EPAG	-9.14	RGLM	-9.03
QGEA	-9.33	EDLR	-9.14	NPQM	-9.03
PDQS	-9.33	PPSA	-9.14	GASL	-9.03
NPQG	-9.32	NAAR	-9.13	NARM	-9.03
GDGA	-9.31	AAVL	-9.13	APLK	-9.03
EGVR	-9.31	AEQV	-9.12	QGEM	-9.03
PGLD	-9.31	RPAV	-9.12	ADAK	-9.02
PGSV	-9.28	GERA	-9.12	AEGS	-9.02
EDGK	-9.28	GAVS	-9.12	PPQV	-9.02
DAQM	-9.28	DARV	-9.12	AEVA	-9.02
PPQM	-9.26	GAGV	-9.11	RDGS	-9.01
RDER	-9.26	NPRG	-9.11	GASS	-9.01
NDGG	-9.26	EAGR	-9.11	NAGA	-9.01
QGAM	-9.25	PESS	-9.1	AAGM	-9.01
EDGR	-9.25	RPSA	-9.1	RDAG	-9
EDER	-9.24	RGRR	-9.1	GDAS	-9
QEVK	-9.23	QDSA	-9.1	RGVG	-9
PGRR	-9.23	QDSD	-9.1	GGQD	-9
AGAR	-9.22	DDSA	-9.09	DELG	-9
QELV	-9.22	NERD	-9.09	PAQS	-9

NAVV	-8.99	EAVD	-8.9	RDGV	-8.83
QPAM	-8.99	PDEG	-8.89	QAGM	-8.82
APGA	-8.99	DDLA	-8.89	REVK	-8.82
DAED	-8.99	ADRG	-8.89	AESL	-8.82
RELD	-8.99	RASR	-8.88	EELK	-8.82
PPAK	-8.98	DDLD	-8.88	PELD	-8.82
QPVG	-8.98	PAEG	-8.88	EGVM	-8.82
GGSA	-8.98	PGAM	-8.88	QESR	-8.82
EPGL	-8.98	NAAG	-8.88	GEEG	-8.82
GEQG	-8.98	QEGG	-8.88	NAVA	-8.82
RARM	-8.97	GDLD	-8.88	QASV	-8.81
NARV	-8.97	NGQR	-8.87	EDAG	-8.81
NGLR	-8.97	QGEL	-8.87	DDRM	-8.81
GESA	-8.97	DEQG	-8.87	GDQD	-8.81
EPER	-8.97	EDRV	-8.87	QPRK	-8.81
GAQA	-8.97	REQG	-8.86	DEEG	-8.81
GEVG	-8.96	NAQL	-8.86	PEVL	-8.81
PGAS	-8.96	QAEA	-8.86	PAAV	-8.81
RAGK	-8.96	AEAM	-8.86	QAGR	-8.81
EPGK	-8.95	AERA	-8.86	PAES	-8.81
QDLL	-8.95	GAQK	-8.86	NPEL	-8.8
GDQK	-8.95	RELM	-8.85	APAA	-8.8
NGRD	-8.95	ADVG	-8.85	PPSV	-8.8
QPSD	-8.95	DERA	-8.85	NEVA	-8.8
PPQS	-8.94	NERR	-8.85	RDAM	-8.8
PAQA	-8.94	EPLA	-8.85	NDRR	-8.8
DGQG	-8.94	EGRK	-8.85	PDSO	-8.8
RPAL	-8.94	NPAA	-8.85	AERM	-8.8
PDVR	-8.94	RGLV	-8.85	EARA	-8.8
EPEK	-8.94	NDLL	-8.85	NEVR	-8.8
GAES	-8.94	RERL	-8.84	GGEV	-8.79
AEVM	-8.94	GESR	-8.84	PASK	-8.79
EGVK	-8.93	EAEM	-8.84	EGRG	-8.79
RDRR	-8.93	AGSM	-8.84	ADLL	-8.79
PEGA	-8.93	AEAL	-8.84	ADAA	-8.79
DASK	-8.93	PDLM	-8.84	QEEB	-8.79
QEAG	-8.92	EALM	-8.84	GGAL	-8.79
RPVL	-8.92	DELM	-8.83	NELM	-8.78
NPVL	-8.91	NGLK	-8.83	GAER	-8.78
PPGS	-8.91	NPVS	-8.83	PPSL	-8.78
ADQS	-8.91	EASV	-8.83	EGLS	-8.78
DERV	-8.91	PAVK	-8.83	DPRR	-8.78
PEAL	-8.9	PAVM	-8.83	DAVV	-8.78
EPEG	-8.9	DELL	-8.83	QEER	-8.78
NGAV	-8.9	PEAD	-8.83	RGQK	-8.77
AEER	-8.9	DDLK	-8.83	EPRR	-8.77



QAVA	-8.77	NESV	-8.71	DPAK	-8.66
AGRD	-8.77	PDQV	-8.71	RDSD	-8.66
GGRM	-8.77	PGGD	-8.71	RDQV	-8.65
NGAS	-8.77	EDES	-8.7	RGRL	-8.65
AARA	-8.77	DDSK	-8.7	EEAG	-8.65
GDGV	-8.77	ADGV	-8.7	QDRD	-8.65
EGLD	-8.77	EPAL	-8.7	QGSK	-8.65
REEL	-8.77	GAQS	-8.7	GAVA	-8.65
PESA	-8.77	NDVM	-8.7	DGQA	-8.65
EDAD	-8.76	AGVA	-8.7	PAQD	-8.65
RDSR	-8.76	NARR	-8.69	GELD	-8.65
ADSA	-8.76	ADV V	-8.69	RELS	-8.65
EGLR	-8.76	QDVL	-8.69	RAQS	-8.65
EASA	-8.76	PEQM	-8.69	QPRG	-8.65
NDVA	-8.76	PAVS	-8.69	EAQG	-8.64
PEEM	-8.76	EEVG	-8.69	DEGM	-8.64
PAQM	-8.75	EGSD	-8.69	RPQR	-8.64
PPGK	-8.75	AAAD	-8.68	GEGR	-8.64
AGQK	-8.75	DAGG	-8.68	DEQK	-8.64
QELS	-8.75	QESM	-8.68	PAGS	-8.64
NPLD	-8.75	PEGS	-8.68	NGED	-8.64
NGGK	-8.74	DDVD	-8.68	RAQK	-8.64
PASG	-8.74	QPEL	-8.68	EDRG	-8.64
NGGV	-8.74	RGRS	-8.68	GDRK	-8.64
QDLK	-8.74	EAAK	-8.68	GGSG	-8.64
PPGL	-8.73	PDRA	-8.68	RAGM	-8.64
DASD	-8.73	RGVK	-8.67	GPAK	-8.64
RGRG	-8.73	PPRS	-8.67	QDLS	-8.64
PEQL	-8.73	GAGG	-8.67	AGRV	-8.64
PPEG	-8.73	QAQK	-8.67	PGSD	-8.64
PPVD	-8.73	RDRA	-8.67	PPES	-8.63
NAQK	-8.73	NGVG	-8.67	RAQM	-8.63
AEQK	-8.73	EDVK	-8.67	EPEV	-8.63
DARG	-8.72	GGRV	-8.67	PGRD	-8.63
NESD	-8.72	QAQM	-8.67	EEVL	-8.63
AGQS	-8.72	PDRK	-8.67	NPRM	-8.63
EDVD	-8.72	QDQM	-8.67	RAED	-8.63
AEVG	-8.72	NARD	-8.66	DPLS	-8.63
NGEA	-8.71	DGAK	-8.66	QAGA	-8.63
REVD	-8.71	REGV	-8.66	EAQV	-8.63
PGVV	-8.71	PPAS	-8.66	QDQK	-8.63
QGAR	-8.71	PPLD	-8.66	RPRL	-8.63
AEAS	-8.71	EPAD	-8.66	DEGS	-8.63
EPAV	-8.71	PERM	-8.66	NDRV	-8.63
QPSM	-8.71	DEEK	-8.66	REVG	-8.63
RGEK	-8.71	PESD	-8.66	EALA	-8.63

EAQD	-8.62	PDVV	-8.57	NGEM	-8.51
DEEM	-8.62	GDQR	-8.57	QGEK	-8.51
DDSR	-8.62	RGAV	-8.57	DAGK	-8.51
DDAL	-8.62	GGGR	-8.57	QDSK	-8.51
EDSA	-8.61	RDEM	-8.56	GPQG	-8.5
AARD	-8.61	PDRR	-8.56	RPVG	-8.5
PAGM	-8.61	QPQD	-8.56	RAGV	-8.5
NEQM	-8.61	DGSG	-8.55	GDED	-8.5
QAVM	-8.61	REGS	-8.55	DGER	-8.5
GELK	-8.61			GPGK	-8.5
PGEA	-8.61	GPGV	-8.55	DPAS	-8.5
PAVD	-8.61	PDRG	-8.55	NGQV	-8.5
GPVA	-8.61	QEGS	-8.55	AAED	-8.49
APAV	-8.61	EDRL	-8.55	AGGA	-8.49
DDSM	-8.61	DGQM	-8.55	NEAM	-8.49
EDSV	-8.6	RPSV	-8.55	APSR	-8.49
PPQG	-8.6	PEGV	-8.55	RDVA	-8.49
GAAD	-8.6	PGLV	-8.55	RDSK	-8.49
PERA	-8.6	REQK	-8.55	RPRS	-8.49
PELA	-8.6	RPVM	-8.54	RPSM	-8.48
GGAK	-8.6	PALM	-8.54	QALS	-8.48
EGAM	-8.6	AGQM	-8.54	RDRK	-8.48
NPEM	-8.6	RGER	-8.54	QASG	-8.48
PERR	-8.6	EDSS	-8.53	PGVK	-8.47
DDRL	-8.6	PDSV	-8.53	PEEV	-8.47
EGLK	-8.6	EGQV	-8.53	EGSA	-8.47
DGGK	-8.6	NEEM	-8.53	NDLR	-8.47
RDVM	-8.59	NAVS	-8.53	RPEK	-8.47
PAAG	-8.59	QAGL	-8.53	NGSG	-8.47
EDGA	-8.59	QAQS	-8.53	APGL	-8.47
ADSV	-8.59	EASS	-8.53	PGGG	-8.47
NEGS	-8.59	EDVL	-8.53	GEVA	-8.47
NPLM	-8.59	NASG	-8.53	QPRR	-8.47
GPVS	-8.58	ADLM	-8.53	EDLV	-8.47
RAVK	-8.58	DASG	-8.53	EGSG	-8.47
APGK	-8.58	QGRD	-8.53	GDRG	-8.47
APLR	-8.58	EPRS	-8.52	RGQL	-8.46
RAVG	-8.58	QGRR	-8.52	GGAS	-8.46
NALG	-8.58	QGGA	-8.52	NAAA	-8.46
NALL	-8.58	EPAM	-8.52	QGQR	-8.46
EDSR	-8.58	QAVV	-8.52	GPSG	-8.46
PERS	-8.58	QGRA	-8.52	RGVM	-8.46
DALG	-8.58	QPGK	-8.52	NALM	-8.46
QDVR	-8.58	QGGS	-8.52	PGQS	-8.46
NPRV	-8.58	PGLA	-8.51	PEGD	-8.46
QDGD	-8.58	DGRM	-8.51	QELD	-8.46

QDQD	-8.46	NALA	-8.4	NASS	-8.36
RDSM	-8.45	QGLA	-8.4	DDAV	-8.36
AEES	-8.45	RGAD	-8.4	QPAL	-8.36
QPLS	-8.45	PAQV	-8.4	DDGD	-8.36
PEEA	-8.45	PEVS	-8.4	DDGV	-8.36
APRK	-8.45	QEVN	-8.4	AEQM	-8.36
GAEV	-8.45	QASR	-8.4	RPLA	-8.36
EEAL	-8.44	PAGR	-8.4	PAVG	-8.35
DPVK	-8.44	PPQK	-8.39	NELL	-8.35
PPEV	-8.44	PGRL	-8.39	QGVD	-8.35
DAQD	-8.44	GERL	-8.39	QERR	-8.35
NALK	-8.44	PAAA	-8.39	QAGD	-8.35
APVK	-8.44	DGSA	-8.39	PDSR	-8.35
RERR	-8.44	PPLA	-8.39	APRV	-8.35
RAGA	-8.44	PELM	-8.39	PDVS	-8.35
DPEG	-8.43	NAVW	-8.39	EDEV	-8.35
RGED	-8.43	QAAK	-8.39	GALS	-8.35
NEGA	-8.43	RAAD	-8.39	AAGD	-8.35
GEAM	-8.43	PGVR	-8.39	AGAD	-8.35
EAAA	-8.43	RGEG	-8.39	EPRA	-8.35
AAQD	-8.43	NDSR	-8.38	APSD	-8.34
GAEM	-8.43	PGQG	-8.38	NALS	-8.34
DERM	-8.43	DDQG	-8.38	QDVG	-8.34
DDVG	-8.43	RGGR	-8.38	DEEA	-8.34
NGVV	-8.42	PPGA	-8.38	AEAA	-8.34
AAQV	-8.42	QAVR	-8.38	ADEG	-8.34
RAQD	-8.42	RGVA	-8.38	QAVD	-8.34
EELS	-8.42	AGGD	-8.38	GDAL	-8.34
EAGG	-8.42	PPGV	-8.38	RGSL	-8.34
QPVK	-8.42	ADVR	-8.38	PESR	-8.34
PDQG	-8.42	NPAD	-8.38	AALV	-8.33
EPSK	-8.42	GESK	-8.37	APGG	-8.33
PAQR	-8.41	DPEV	-8.37	DGRK	-8.33
RAQL	-8.41	PEAG	-8.37	NARG	-8.33
EAVK	-8.41	PARG	-8.37	GPGM	-8.33
DGED	-8.41	EGEL	-8.37	NDLG	-8.32
NEGD	-8.41	EPGS	-8.37	DAEG	-8.32
NDEA	-8.41	GPLG	-8.37	NDGS	-8.32
QPEA	-8.41	EAAR	-8.37	PPAV	-8.32
PDLK	-8.41	PEGM	-8.36	EPAA	-8.32
REVV	-8.41	PEVG	-8.36	RARA	-8.32
PDES	-8.41	NELG	-8.36	QPLR	-8.32
PESG	-8.41	PDSG	-8.36	PGEG	-8.32
DASR	-8.41	QEGD	-8.36	GELM	-8.31
PGRV	-8.4	EEEE	-8.36	REGM	-8.31
QPQM	-8.4	EAEK	-8.36	GDVK	-8.31

GDRL	-8.31	RPAD	-8.27	PPRM	-8.23
GPLM	-8.31	RDVL	-8.26	AGRS	-8.23
RPER	-8.31	EPQS	-8.26	DDLK	-8.23
QPQA	-8.31	PEEL	-8.26	PAGV	-8.23
APEK	-8.31	PAEA	-8.26	DDAK	-8.22
EGGG	-8.31	DPRL	-8.26	GAQL	-8.22
GGEM	-8.31	RDSS	-8.26	RAQA	-8.22
NELD	-8.31	DDED	-8.26	RGSR	-8.22
NGAL	-8.3	AGGR	-8.26	AAVG	-8.22
RDQS	-8.3	PPEA	-8.26	NEQA	-8.22
QGAG	-8.3	EGLA	-8.26	GPAV	-8.22
EELV	-8.3	RALM	-8.26	NAAK	-8.21
GERS	-8.3	GDLS	-8.26	NGQK	-8.21
NPAS	-8.3	EAVM	-8.26	DEEL	-8.21
NGGM	-8.3	GPVV	-8.26	QEEV	-8.21
AEVD	-8.3	APVA	-8.26	GAAL	-8.21
NEQR	-8.3	PEES	-8.25	EEAK	-8.21
DAVM	-8.3	NAQS	-8.25	GEGM	-8.2
RGQR	-8.3	PELK	-8.25	GGSS	-8.2
DALS	-8.3	QEQV	-8.25	QGSV	-8.2
NPAM	-8.3	DGSK	-8.25	NGSV	-8.2
RALS	-8.29	PAAL	-8.25	EAQR	-8.2
AELA	-8.29	QPVV	-8.25	EERV	-8.2
EGGL	-8.29	QGRM	-8.25	QELK	-8.2
GPLK	-8.29	PELR	-8.25	AARM	-8.2
RDED	-8.28	DDRK	-8.25	DEAK	-8.2
QDVM	-8.28	RPLD	-8.25	EPLG	-8.2
APSV	-8.28	QGLS	-8.25	QAAS	-8.2
APVR	-8.28	QEEL	-8.25	GPVG	-8.19
PGRS	-8.28	APLS	-8.24	RDLK	-8.19
NAAL	-8.28	DAGR	-8.24	GGGV	-8.19
PGQK	-8.28	EAA S	-8.24	NGQD	-8.19
EGVA	-8.28	PGLR	-8.24	NDAR	-8.19
QALD	-8.28	GGVK	-8.24	PDLL	-8.19
AGQG	-8.28	EPVL	-8.24	AAQA	-8.19
QPGA	-8.28	GDEG	-8.23	QGQA	-8.19
GPRK	-8.28	GGGS	-8.23	NEAG	-8.19
EPLS	-8.27	RDQL	-8.23	PPQL	-8.19
QGLR	-8.27	QPRD	-8.23	EARM	-8.18
PDEV	-8.27	NDAV	-8.23	RARS	-8.18
RGSG	-8.27	DDVV	-8.23	PEVV	-8.18
DDRS	-8.27	GPLR	-8.23	EASM	-8.18
PDVM	-8.27	DEV M	-8.23	AEQG	-8.18
PARM	-8.27	EEVM	-8.23	DPQG	-8.18
EASD	-8.27	AAER	-8.23	QEAD	-8.18
NPSK	-8.27	NPSA	-8.23	EALV	-8.18

PDLV	-8.18	RPAG	-8.14	QDQS	-8.08
DGAR	-8.18	PPAG	-8.14	RAER	-8.07
ADQL	-8.18	DARA	-8.14	QDRV	-8.07
EAVR	-8.18	DALL	-8.14	PAAD	-8.07
EELM	-8.18	DEAM	-8.13	NPGR	-8.07
QPRA	-8.18	AGVS	-8.13	QERK	-8.07
PDQD	-8.17	DEQA	-8.13	EEAD	-8.07
GPAL	-8.17	GDVS	-8.13	PDEL	-8.07
QERV	-8.17	EESA	-8.13	GPQM	-8.07
RGQA	-8.17	APAG	-8.13	PDGS	-8.06
RDRS	-8.17	RGAA	-8.12	NEER	-8.06
APQG	-8.17	NDGK	-8.12	PALS	-8.06
PAVA	-8.17	ADAR	-8.12	GEQD	-8.06
QERA	-8.17	PGQL	-8.12	QGGK	-8.06
QGVR	-8.16	NAAD	-8.12	QPEK	-8.06
RGLD	-8.16	QPAA	-8.12	RESS	-8.06
DELK	-8.16	EGSK	-8.12	AEAR	-8.06
PPGR	-8.16	PEAS	-8.12	NEVD	-8.06
DGEA	-8.16	QPVD	-8.12	EPVS	-8.06
NPAL	-8.16	QGGM	-8.12	EPGR	-8.05
DAER	-8.16	AESK	-8.11	PDEM	-8.05
RGGM	-8.16	EAQL	-8.11	DPAV	-8.05
NPGL	-8.16	RGAM	-8.11	EEER	-8.05
RGRM	-8.16	AALG	-8.11	AAGV	-8.05
PGAD	-8.16	PGVA	-8.11	PGRK	-8.05
NPEA	-8.15	DGSV	-8.11	AEEL	-8.04
PERD	-8.15	EPSV	-8.11	NGVR	-8.04
PGSS	-8.15	NGGR	-8.11	REQA	-8.04
AAEM	-8.15	DGEV	-8.1	NARS	-8.04
QAAG	-8.15	QAQV	-8.1	ADQA	-8.04
DDAG	-8.15	DGVK	-8.1	QELG	-8.04
DGSM	-8.15	QGES	-8.1	PDLL	-8.04
DPLG	-8.15	EAGK	-8.1	RPGV	-8.04
RPGR	-8.15	NPRL	-8.1	GERM	-8.04
DAVG	-8.15	QDEG	-8.09	NAEM	-8.04
EGAK	-8.15	PEEL	-8.09	GGRD	-8.04
EPAR	-8.15	DDVS	-8.09	PPAD	-8.04
GEGK	-8.15	NGLS	-8.09	GEGD	-8.03
DDSS	-8.15	GGAA	-8.09	GEVV	-8.03
NDAK	-8.15	PDGG	-8.09	APRL	-8.03
PDLR	-8.15	QDRM	-8.09	READ	-8.03
NDQV	-8.15	PEVA	-8.09	NGGA	-8.03
AGSA	-8.15	GGSD	-8.09	NPQL	-8.03
DGQV	-8.15	PGGS	-8.09	PAEK	-8.03
EALS	-8.14	DDQD	-8.08	AGAM	-8.03
EGQK	-8.14	PAGD	-8.08	QAVS	-8.03

GPEA	-8.03	DAEA	-7.98	DPSM	-7.92
RPQV	-8.03	PPVK	-7.98	EPLR	-7.92
DDGM	-8.03	NGLM	-7.98	QGQV	-7.92
GPRM	-8.03	NEGL	-7.98	RDGR	-7.92
PPLV	-8.03	AAAM	-7.98	ADRA	-7.92
GDAA	-8.02	AARL	-7.98	DPES	-7.91
QALV	-8.02	NDLS	-7.98	DARK	-7.91
EGEG	-8.02	QDVS	-7.98	NDRD	-7.91
RGGA	-8.02	PDLG	-7.98	AELK	-7.91
RDSV	-8.02	QDSR	-7.97	PERV	-7.91
AGED	-8.02	RPED	-7.97	EDGV	-7.91
NGSS	-8.02	EGSL	-7.97	NPED	-7.91
GAQG	-8.02	REQM	-7.96	GDER	-7.91
NARA	-8.01	PERK	-7.96	QEEK	-7.9
RPRV	-8.01	GGLM	-7.96	GGQL	-7.9
APQS	-8.01	PPAR	-7.96	NDVR	-7.9
QAQA	-8.01	DGGG	-7.96	QGLL	-7.9
QGAK	-8.01	DEES	-7.96	PPSM	-7.9
NGRL	-8.01	PARV	-7.95	NEES	-7.9
GDVA	-8.01	PDEK	-7.95	RPGD	-7.9
PDAA	-8.01	DAVR	-7.95	EGVS	-7.89
GAED	-8.01	PGRG	-7.95	PDGL	-7.89
EAAM	-8.01	GEQS	-7.94	NEAD	-7.89
GAQD	-8.01	EEQL	-7.94	PPAM	-7.89
NAQV	-8.01	NPSM	-7.94	NESG	-7.89
RDVK	-8.01	QALG	-7.94	GDSL	-7.89
PDSK	-8.01	EAEA	-7.94	RDRG	-7.88
PDGD	-8	RGEL	-7.94	APES	-7.88
RAGS	-8	RGSA	-7.94	GDLG	-7.88
DAAS	-8	AGVG	-7.94	ADSD	-7.88
RESA	-8	PEQK	-7.94	APVS	-7.87
QPVS	-8	DEQM	-7.94	GPAD	-7.87
DDQL	-8	ADRR	-7.93	DAGM	-7.87
EERD	-8	PGQV	-7.93	PGAG	-7.87
ADRM	-8	PGGV	-7.93	PDGM	-7.87
DASM	-8	GERK	-7.93	RASM	-7.87
RPAS	-8	EEVV	-7.93	GPGA	-7.86
QPGR	-8	PDAS	-7.93	EPSR	-7.86
GASK	-7.99	EDGD	-7.93	GPEN	-7.86
APLL	-7.99	EGQS	-7.93	RPRA	-7.86
GPQD	-7.99	NPGS	-7.92	PDED	-7.86
GGSK	-7.99	RDSG	-7.92	PAVL	-7.86
NAEA	-7.99	GGQG	-7.92	EESG	-7.86
RPSL	-7.99	RALK	-7.92	QESG	-7.86
NDGD	-7.99	PAED	-7.92	AGGL	-7.86
DPSG	-7.99	EPVM	-7.92	ADER	-7.85

APEV	-7.85	AGLA	-7.79	DDVL	-7.72
AAEL	-7.85	QEAS	-7.79	RPEA	-7.72
PASA	-7.85	QEAM	-7.79	RAQV	-7.71
AGVM	-7.85	GDLR	-7.79	RGLS	-7.71
DERS	-7.85	GEER	-7.79	APED	-7.71
DDGG	-7.85	NGLD	-7.79	QDAG	-7.71
AEGV	-7.85	PAAM	-7.78	NGLL	-7.7
PPAL	-7.85	PDSA	-7.78	GGSL	-7.7
PGAK	-7.85	NDQA	-7.78	GPSV	-7.7
RELR	-7.85	NASV	-7.78	EGRV	-7.69
DEVD	-7.85	GGEK	-7.78	GEGS	-7.69
EGVD	-7.85	ADLG	-7.77	QPRV	-7.68
RAEV	-7.85	PPSS	-7.77	EPSD	-7.68
EEEV	-7.85	ADRS	-7.77	DEAG	-7.68
DEGD	-7.85	QGQD	-7.77	EELL	-7.68
NPAK	-7.85	PEVM	-7.76	NAQM	-7.68
QAEM	-7.85	NAAV	-7.76	APRR	-7.67
GGED	-7.84	GGQV	-7.76	DGGL	-7.67
PGES	-7.84	NDRK	-7.76	PEAV	-7.67
AGLK	-7.84	GDEV	-7.76	PDRL	-7.66
DAVS	-7.84	GESS	-7.76	QDLG	-7.66
QASL	-7.84	PARK	-7.75	AGGM	-7.66
REAA	-7.84	PPRG	-7.75	QDGS	-7.65
AELR	-7.84	PASL	-7.75	PGLL	-7.65
AAES	-7.83	PPGM	-7.75	QPGM	-7.65
PEGG	-7.83	PGLS	-7.75	PGLK	-7.65
NGQS	-7.82	GARR	-7.75	GDAR	-7.65
PEAM	-7.82	NDGR	-7.75	DPVS	-7.65
NGGL	-7.82	EELG	-7.74	EEAR	-7.64
QPGL	-7.82	EDAM	-7.74	DESA	-7.64
PDQL	-7.82	NPLV	-7.74	PGER	-7.64
ADQG	-7.82	NGAK	-7.74	NPGM	-7.64
NGVL	-7.82	EDEM	-7.74	NASR	-7.64
EALK	-7.81	NASA	-7.74	APLV	-7.63
RAEA	-7.81	PDAG	-7.74	REAS	-7.63
QDGR	-7.81	RDGL	-7.74	ADQM	-7.63
GEVM	-7.81	PPVG	-7.74	QEVV	-7.63
ADSL	-7.81	AAAK	-7.73	QARD	-7.63
ADLR	-7.81	RPGS	-7.73	GERV	-7.63
PDVG	-7.81	PGRA	-7.73	ADVL	-7.62
AGVV	-7.8	RAVA	-7.73	PDRD	-7.62
GGRL	-7.8	AGES	-7.72	AGGK	-7.62
EGLG	-7.8	PPVS	-7.72	PEQA	-7.62
QDRR	-7.8	RPQD	-7.72	RERA	-7.62
AESV	-7.8	PPLG	-7.72	EGGD	-7.62
GPSL	-7.8	RARR	-7.72	PEGL	-7.62

PELG	-7.61	NAVL	-7.51	DPSS	-7.32
NARK	-7.61	RDLR	-7.5	QPVL	-7.31
EERK	-7.61	EEGA	-7.5	DDQS	-7.31
EEVA	-7.61	PPVA	-7.5	RELL	-7.3
GEQV	-7.61	DPLD	-7.5	NPSR	-7.3
QAVL	-7.61	GDSR	-7.5	DAEL	-7.29
EAGS	-7.61	QGGG	-7.49	EGRS	-7.28
NDED	-7.61	QGLV	-7.49	NGAG	-7.27
DDRV	-7.6	RPAM	-7.49	QASA	-7.27
ADQK	-7.6	GDAK	-7.49	APEG	-7.27
APEM	-7.6	AERG	-7.48	PDAM	-7.27
GELS	-7.6	DDSG	-7.48	PAQK	-7.25
PDVK	-7.6	GEEL	-7.47	AAEG	-7.24
DPLR	-7.59	NGRV	-7.47	EEQM	-7.24
DESS	-7.59	AEEV	-7.47	QEEM	-7.24
AGQD	-7.59	RPSK	-7.47	PPGG	-7.22
EARS	-7.59	PAER	-7.46	QEGA	-7.22
PDQK	-7.58	AGRR	-7.46	EARR	-7.22
PPRR	-7.58	QERG	-7.45	RDQA	-7.21
ADLV	-7.58	QAES	-7.45	NDES	-7.2
GGRS	-7.57	EGSM	-7.44	GDEL	-7.19
PARD	-7.57	EA SG	-7.44	PELV	-7.18
PALV	-7.57	DGRV	-7.44	RDLA	-7.18
DDEL	-7.57	RGGG	-7.43	GGVA	-7.16
PPAA	-7.56	NDQD	-7.42	GPGG	-7.14
QGGR	-7.56	DAVL	-7.41	GDLA	-7.14
NDSL	-7.56	QGRS	-7.41	DAAV	-7.14
GEQA	-7.56	EDLA	-7.41	AAAS	-7.12
PARR	-7.55	EEAV	-7.4	EAED	-7.11
PPQA	-7.55	DAAA	-7.4	GEEM	-7.11
EPVD	-7.54	PEQG	-7.39	NAVR	-7.1
PGEL	-7.54	NEQD	-7.39	PPSK	-7.1
PDAV	-7.54	QASD	-7.38	AGLD	-7.09
PEQS	-7.54	AAQL	-7.37	PAGG	-7.03
RGSD	-7.54	DPVA	-7.37	PDER	-6.97
GPQL	-7.53	GAGR	-7.37	AESD	-6.94
RESK	-7.53	PDGA	-7.36	QGRV	-6.91
NDLA	-7.53	DGLD	-7.36	PGQR	-6.89
AEAK	-7.52	EAES	-7.35	AGGG	-6.87
REQR	-7.52	GAAK	-7.34	AALM	-6.81
DGSS	-7.51	QALK	-7.34		
QDAK	-7.51	DPGK	-7.34		
NGGG	-7.51	NESK	-7.33		
RGRK	-7.51	EDQL	-7.33		
RAGR	-7.51	DGAL	-7.33		
QGSD	-7.51	EEEE	-7.33		



## Appendix 2-B

Docking to the 500 “B-linker-A” structures showing the best binding affinity towards the reagent.

Column 1: B-linker-A structure  
Column 2: Docking energy of the reagent: glutaryl 7-aminoccephalosporanic acid  
Column 3: Docking energy of product 1: 7-aminoccephalosporanic acid  
Column 4: Docking energy of product 2: glutaric acid

Note: rows are in alphabetical order

B-linker-A structure	Ed(reagent) /(kcal/mol)	Ed(product 1) /(kcal/mol)	Ed (product 2) /(kcal/mol)
AAAD	-8.68	-9.14	-2.9
AAGG	-9.4	-8.69	-2.71
AAGM	-9.01	-8.83	-3.24
AARA	-8.77	-8	-2.7
AARD	-8.61	-9.2	-2.76
AARV	-9.76	-9.53	-2.8
AASS	-9.06	-8.65	-2.82
AASV	-9.18	-8.93	-3.15
AAVL	-9.13	-9.11	-3.93
ADAA	-8.79	-9.62	-3.68
ADAK	-9.02	-8.11	-2.86
ADAL	-9.14	-9.15	-2.84
ADGV	-8.7	-9.22	-4.32
ADLK	-9.42	-8.94	-3.36
ADLL	-8.79	-9.06	-2.99
ADQS	-8.91	-9.08	-2.83
ADRG	-8.89	-8.84	-3.36
ADSA	-8.76	-8.63	-3.44
ADSV	-8.59	-8.17	-3.35
ADVG	-8.85	-8.85	-2.86
ADV V	-8.69	-9.22	-2.86
AEAL	-8.84	-8.51	-2.86
AEAM	-8.86	-9.5	-2.95
AEAS	-8.71	-9.35	-3.89
AEEA	-9.04	-9.39	-2.84
AEER	-8.9	-8.22	-2.76
AEGS	-9.02	-9.04	-3.49

AELD	-9.08	-8.69	-2.94
AEQK	-8.73	-8.97	-2.89
AEQV	-9.12	-8.2	-2.84
AERA	-8.86	-8.98	-2.94
AERM	-8.8	-8.73	-3.79
AESL	-8.82	-9.06	-2.66
AEVA	-9.02	-8.94	-2.87
AEVG	-8.72	-8.98	-3.37
AEVM	-8.94	-9.26	-2.9
AEVV	-9.04	-8.63	-2.56
AGAR	-9.22	-8.97	-3.2
AGQK	-8.75	-8.93	-3.28
AGQS	-8.72	-9.32	-3.15
AGRD	-8.77	-9.03	-3.04
AGRV	-8.64	-8.05	-3.01
AGSM	-8.84	-7.62	-2.91
AGVA	-8.7	-9.27	-2.87
APAA	-8.8	-8.72	-2.93
APAV	-8.61	-8.03	-2.85
APGA	-8.99	-8.88	-3.5
APGK	-8.58	-9.27	-2.72
APGV	-9.08	-9.34	-2.72
APLK	-9.03	-8.48	-3.09
APLR	-8.58	-9.26	-3.39
APQL	-9.07	-8.79	-3.86
DAED	-8.99	-9.01	-2.9
DAGG	-8.68	-9.5	-3.21
DALG	-8.58	-8.72	-3.72
DALM	-9.64	-9.04	-3.24
DAQK	-9.76	-9.25	-3.69
DAQM	-9.28	-8.62	-3.37
DARG	-8.72	-8.66	-2.99
DARV	-9.12	-9.63	-3.32
DASD	-8.73	-9.19	-2.88
DASK	-8.93	-9.54	-3.57
DAVV	-8.78	-9.12	-3.44
DDAL	-8.62	-9.26	-2.77
DDLA	-8.89	-8.35	-3.84
DDLD	-8.88	-8.79	-2.66
DDLRL	-8.83	-9.16	-2.94
DDRL	-8.6	-9.12	-3.04
DDRM	-8.81	-8.96	-3.13
DDSA	-9.09	-8.83	-2.86
DDSK	-8.7	-8.8	-2.78
DDSM	-8.61	-8.78	-2.98
DDSR	-8.62	-7.67	-3.07

DDVD	-8.68	-8.94	-2.78
DEAA	-9.7	-8.85	-2.54
DEEG	-8.81	-8.97	-2.77
DEEK	-8.66	-9.21	-2.73
DEEM	-8.62	-9.98	-3.23
DEGM	-8.64	-8.92	-3.21
DEGS	-8.63	-8.53	-2.65
DELA	-9.07	-8.14	-3.36
DELG	-9	-8.66	-2.98
DELL	-8.83	-8.6	-2.7
DELM	-8.83	-9.07	-3.36
DEQG	-8.87	-9.15	-3.37
DEQK	-8.64	-8.91	-3.43
DEQS	-9.42	-9.04	-2.79
DERA	-8.85	-9.27	-3.04
DERV	-8.91	-8.78	-2.76
DESK	-9.15	-8.44	-3.02
DESM	-9.18	-8.83	-3.18
DEVL	-9.73	-9.04	-3.89
DEVV	-9.22	-8.71	-3.01
DGAK	-8.66	-9.14	-3.04
DGGK	-8.6	-9.22	-3
DGQA	-8.65	-8.27	-3.28
DGQG	-8.94	-9.51	-3.03
DGQK	-9.6	-8.96	-4.46
DGSG	-8.55	-8.88	-2.72
DGVM	-9.05	-9.22	-2.93
DPAG	-9.46	-9.21	-2.68
DPAK	-8.66	-8.23	-3.14
DPAM	-9.03	-8.54	-2.76
DPLK	-9.08	-8.72	-3.22
DPLS	-8.63	-9.34	-3.34
DPRG	-10.19	-9.6	-4.58
DPRR	-8.78	-8.73	-3.35
DPVV	-9.15	-9.6	-3.66
EAAK	-8.68	-8.83	-3.28
EAEM	-8.84	-8.86	-2.81
EAGR	-9.11	-9.22	-2.95
EALA	-8.63	-9.38	-2.92
EALM	-8.84	-9.12	-2.91
EAQD	-8.62	-9.58	-4.09
EAQG	-8.64	-8.88	-2.79
EAQV	-8.63	-8.19	-3.19
EARA	-8.8	-9.11	-3.12
EASA	-8.76	-8.66	-3.17
EASV	-8.83	-8.65	-3.08

EAVD	-8.9	-9.14	-3.13
EDAD	-8.76	-9.03	-3.13
EDAG	-8.81	-8.88	-3.81
EDER	-9.24	-8.95	-2.87
EDES	-8.7	-8.68	-2.82
EDGA	-8.59	-8.39	-3
EDGK	-9.28	-9.37	-3.13
EDGR	-9.25	-9.03	-2.91
EDLR	-9.14	-9.33	-3.41
EDQA	-9.14	-9.23	-3.33
EDRA	-9.57	-9.41	-2.82
EDRG	-8.64	-9.02	-3.41
EDRV	-8.87	-9.13	-3.63
EDSA	-8.61	-9.05	-2.64
EDSR	-8.58	-9.36	-3
EDSV	-8.6	-8.15	-2.98
EDVD	-8.72	-8.27	-3.01
EDVK	-8.67	-8.66	-3.38
EEAG	-8.65	-8.17	-2.98
EELK	-8.82	-8.58	-3.07
EERA	-9.15	-8.96	-3.21
EERR	-9.72	-9.23	-3.01
EESK	-9.53	-9.69	-2.65
EEVG	-8.69	-9.34	-2.78
EEVL	-8.63	-8.81	-3.61
EGAM	-8.6	-8.65	-2.89
EGLD	-8.77	-9.1	-3.06
EGLK	-8.6	-8.56	-2.82
EGLR	-8.76	-8.48	-3.22
EGLS	-8.78	-9.66	-4.45
EGRG	-8.79	-8.92	-3.03
EGRK	-8.85	-9.14	-2.91
EGSD	-8.69	-9.08	-2.74
EGVK	-8.93	-8.86	-2.77
EGVM	-8.82	-8.8	-3.64
EGVR	-9.31	-9.59	-3.03
EPAD	-8.66	-8.19	-2.88
EPAG	-9.14	-8.98	-3.41
EPAL	-8.7	-9.33	-3.22
EPAV	-8.71	-8.15	-3.18
EPED	-9.86	-9.49	-3.94
EPEG	-8.9	-8.41	-3.07
EPEK	-8.94	-9.04	-2.85
EPER	-8.97	-9.14	-3.32
EPEV	-8.63	-8.93	-2.89
EPGA	-9.44	-8.72	-2.82

EPGK	-8.95	-8.99	-2.78
EPGL	-8.98	-8.77	-2.72
EPLA	-8.85	-9.32	-3.3
EPRR	-8.77	-8.77	-3.61
EPVA	-9.18	-8.96	-3.23
EPVK	-9.05	-8.78	-3.81
GAAD	-8.6	-8.59	-2.94
GAER	-8.78	-8.84	-2.9
GAES	-8.94	-8.89	-2.87
GAGG	-8.67	-8.82	-2.96
GAGV	-9.11	-8.78	-2.87
GAQA	-8.97	-8.51	-2.86
GAQK	-8.86	-8.81	-2.44
GAQS	-8.7	-9.35	-2.94
GARS	-9.6	-9.33	-2.9
GASL	-9.03	-8.88	-3.52
GASS	-9.01	-8.23	-3.22
GAVA	-8.65	-8.85	-3.64
GAVS	-9.12	-8.63	-3.21
GDAD	-9.05	-8.81	-4.05
GDAS	-9	-8.98	-3.03
GDGA	-9.31	-9.27	-2.62
GDGV	-8.77	-8.94	-3.22
GDL D	-8.88	-8.93	-3.06
GDQD	-8.81	-7.98	-2.9
GDQK	-8.95	-9.04	-3.18
GDQM	-9.35	-8.86	-3.29
GDQR	-8.57	-7.92	-3.38
GDRK	-8.64	-8.79	-3.21
GDSV	-9.16	-8.79	-4.08
GDVG	-9.49	-9.02	-4.06
GDVR	-9.06	-8.21	-2.79
GEEG	-8.82	-9.04	-4.07
GEGR	-8.64	-8.64	-2.89
GEGV	-9.34	-8.85	-3.28
GELD	-8.65	-8.95	-2.76
GELK	-8.61	-9.36	-3.02
GEQG	-8.98	-8.5	-3
GERA	-9.12	-9.16	-3.96
GESA	-8.97	-8.99	-4.16
GESR	-8.84	-8.04	-4
GEVG	-8.96	-9.35	-2.79
GGAK	-8.6	-9.12	-2.82
GGAL	-8.79	-9.7	-3.85
GGAV	-9.95	-9.65	-3.01
GGEL	-9.34	-8.75	-2.94

GGEV	-8.79	-8.43	-3.02
GGGR	-8.57	-9.18	-3.85
GGQD	-9	-8.49	-2.93
GGRM	-8.77	-9.04	-3.68
GGRV	-8.67	-9.55	-3.18
GGSA	-8.98	-8.71	-3.42
GGSG	-8.64	-8.83	-3.15
GGVS	-9.62	-8.72	-3.2
GPAK	-8.64	-9.15	-3.76
GPED	-9.61	-9.11	-3.98
GPQS	-9.05	-8.6	-3.25
GPRG	-9.45	-8.66	-3.57
GPVA	-8.61	-8.62	-3.62
GPVS	-8.58	-8.83	-3.13
NAAG	-8.88	-9.16	-2.85
NAAR	-9.13	-9.48	-2.67
NAEV	-9.41	-9.91	-2.99
NAGA	-9.01	-9.58	-2.79
NALG	-8.58	-9.06	-3.47
NALL	-8.58	-9.32	-3.3
NAQK	-8.73	-8.57	-2.78
NAQL	-8.86	-9.03	-3.95
NARD	-8.66	-9.37	-2.88
NARM	-9.03	-8.72	-3.05
NARR	-8.69	-9.22	-3.27
NARV	-8.97	-9.77	-3.43
NAVA	-8.82	-9.52	-3.77
NAVV	-8.99	-9.33	-2.9
NDGG	-9.26	-8.55	-3.12
NDLL	-8.85	-8.64	-2.68
NDRR	-8.8	-8.27	-3.03
NDRV	-8.63	-9.45	-2.82
NDVA	-8.76	-7.78	-2.56
NDVD	-9.18	-9.42	-3.94
NDVM	-8.7	-8.64	-3.04
NEAK	-9.17	-8.97	-3.16
NEGS	-8.59	-7.79	-2.78
NELM	-8.78	-9.19	-3.71
NEQM	-8.61	-8.84	-2.74
NERD	-9.09	-9.7	-2.72
NERR	-8.85	-8.02	-3.08
NESD	-8.72	-8.77	-2.74
NESR	-9.19	-8.66	-3.1
NESV	-8.71	-8.97	-3.06
NEVA	-8.8	-8.54	-2.72
NEVR	-8.8	-8.77	-3.35

NGAD	-9.59	-8.88	-3.26
NGAS	-8.77	-9.03	-2.86
NGAV	-8.9	-9.09	-3.26
NGEA	-8.71	-9.19	-2.8
NGED	-8.64	-8.94	-3.57
NGEG	-9.42	-9.18	-2.84
NGGK	-8.74	-9.74	-2.88
NGGV	-8.74	-9.22	-3.58
NGLK	-8.83	-9.13	-2.99
NGLR	-8.97	-9.1	-3.91
NGQR	-8.87	-9.14	-3.35
NGRA	-9.17	-9.63	-2.46
NGRD	-8.95	-9.58	-2.99
NGVG	-8.67	-8.27	-2.69
NPAA	-8.85	-8.97	-2.97
NPAV	-9.54	-9.17	-3.06
NPEL	-8.8	-9.14	-2.71
NPEM	-8.6	-8.98	-3.21
NPLD	-8.75	-8.08	-2.97
NPLM	-8.59	-9.37	-2.82
NPLR	-10.27	-8.56	-2.71
NPQG	-9.32	-9.33	-3.9
NPQM	-9.03	-8.14	-3.71
NPRG	-9.11	-7.78	-2.73
NPRK	-9.54	-9.81	-3.69
NPRM	-8.63	-9.24	-2.91
NPRR	-10.13	-8.92	-2.7
NPRV	-8.58	-9.2	-2.72
NPVD	-9.35	-9	-3.61
NPVL	-8.91	-8.9	-2.9
NPVS	-8.83	-8.79	-3.61
PAAG	-8.59	-8.95	-3.13
PAAV	-8.81	-9.31	-3.21
PAEG	-8.88	-9.8	-2.81
PAES	-8.81	-8.13	-2.99
PAGM	-8.61	-8.79	-2.95
PAGS	-8.64	-7.97	-3.05
PALL	-9.9	-8.71	-2.7
PAQA	-8.94	-8.93	-3.74
PAQD	-8.65	-9.25	-3.05
PAQM	-8.75	-8.75	-2.87
PAQS	-9	-8.86	-2.95
PASG	-8.74	-8.95	-3.36
PASK	-8.79	-9.77	-2.76
PAVD	-8.61	-8.63	-2.89
PAVK	-8.83	-8.8	-2.92

PAVM	-8.83	-8.96	-3.34
PAVS	-8.69	-9.36	-3.04
PDAK	-9.21	-8.92	-4.11
PDEG	-8.89	-9.68	-3.04
PDLM	-8.84	-9.95	-2.73
PDQS	-9.33	-8.59	-3.85
PDQV	-8.71	-9.26	-2.85
PDRA	-8.68	-8.21	-2.76
PDRK	-8.67	-9.3	-3.41
PDRR	-8.56	-7.95	-3.09
PDSO	-8.8	-8.47	-2.64
PDSS	-10.03	-9.15	-2.94
PDVL	-9.05	-9.11	-3.58
PDVR	-8.94	-8.56	-2.85
PDVV	-8.57	-9.34	-2.77
PEAD	-8.83	-9.08	-3.14
PEAK	-9.69	-9.06	-3.3
PEAL	-8.9	-9.48	-2.69
PEEM	-8.76	-8.61	-3.2
PEGA	-8.93	-9.09	-2.77
PEGS	-8.68	-9.14	-2.39
PELA	-8.6	-8.96	-3.03
PELD	-8.82	-8.71	-3.38
PEQL	-8.73	-9.16	-2.83
PEQM	-8.69	-8.97	-2.91
PERA	-8.6	-8.86	-2.92
PERG	-9.06	-8.67	-2.93
PERM	-8.66	-9.42	-3.39
PERR	-8.6	-8.48	-2.93
PERS	-8.58	-8.79	-4.2
PESA	-8.77	-8.74	-3.29
PESD	-8.66	-9.3	-3.02
PESS	-9.1	-8.47	-3.58
PEVK	-9.09	-9.17	-3.52
PEVL	-8.81	-7.91	-2.83
PEVR	-9.65	-8.97	-3.14
PGAM	-8.88	-8.78	-2.9
PGAS	-8.96	-8.98	-2.7
PGAV	-9.19	-9.09	-3.08
PGEA	-8.61	-9.58	-3.07
PGEK	-9.07	-8.93	-2.89
PGGD	-8.71	-8.84	-2.81
PGLD	-9.31	-8.49	-2.91
PGLG	-9.49	-9.7	-3.05
PGLM	-9.2	-8.97	-3.34
PGRD	-8.63	-9.21	-2.85



PGRR	-9.23	-9.05	-3.12
PGSD	-8.64	-8.07	-2.62
PGSV	-9.28	-8.59	-3.87
PGVV	-8.71	-8.49	-2.88
PPAK	-8.98	-9	-3.89
PPAS	-8.66	-9.75	-3.43
PPED	-9.4	-9.24	-3.01
PPEG	-8.73	-9.47	-3.05
PPES	-8.63	-9.21	-3
PPGK	-8.75	-8.42	-3.52
PPGL	-8.73	-8.78	-3.02
PPGS	-8.91	-9.02	-2.86
PPLD	-8.66	-8.97	-2.77
PPQG	-8.6	-9.45	-3.06
PPQM	-9.26	-9.15	-3.02
PPQS	-8.94	-8.7	-4.03
PPQV	-9.02	-9.07	-2.97
PPRD	-9.53	-8.64	-2.95
PPRS	-8.67	-9.11	-3.48
PPSA	-9.14	-9.72	-3.03
PPSD	-9.67	-9.46	-2.85
PPSL	-8.78	-8.45	-2.67
PPSV	-8.8	-7.84	-3.16
PPVD	-8.73	-9.05	-3.35
QAEA	-8.86	-8.9	-2.7
QAGA	-8.63	-8.84	-3.88
QAGM	-8.82	-8.91	-4.02
QAGR	-8.81	-9.59	-3.23
QAQK	-8.67	-9.54	-3.11
QAQM	-8.67	-8.56	-4.09
QARK	-9.2	-9.02	-3.26
QASV	-8.81	-9.08	-3.49
QAVA	-8.77	-9.19	-3.38
QAVM	-8.61	-8.93	-3.26
QDEV	-9.43	-8.86	-2.88
QDGD	-8.58	-9.23	-3.32
QDGK	-9.81	-7.74	-2.99
QDLK	-8.74	-9.03	-2.66
QDLL	-8.95	-9.48	-3.27
QDLS	-8.64	-8.86	-3.02
QDQK	-8.63	-8.75	-2.93
QDQM	-8.67	-8.76	-2.75
QDRD	-8.65	-8.93	-3.59
QDSA	-9.1	-8.92	-3.94
QDSD	-9.1	-8.86	-4.14
QDSS	-10.02	-8.56	-3.43

QDVL	-8.69	-9.5	-2.98
QDVR	-8.58	-8.34	-3.33
QEAG	-8.92	-8.58	-2.64
QEAK	-9.2	-8.63	-2.66
QEER	-8.78	-8.9	-2.79
QEES	-8.79	-8.64	-2.82
QEGG	-8.88	-9.17	-2.87
QEGL	-9.62	-9.41	-3.34
QELS	-8.75	-9.36	-4.21
QELV	-9.22	-8.73	-3.33
QESM	-8.68	-9.07	-3.23
QESR	-8.82	-9.09	-2.98
QEVA	-9.2	-9.27	-3.76
QEVK	-9.23	-9.18	-3.51
QEVS	-9.22	-9.76	-3.33
QGAM	-9.25	-8.82	-3.23
QGAR	-8.71	-8.77	-3.03
QGEA	-9.33	-8.54	-4.05
QGEL	-8.87	-8.99	-3.3
QGEM	-9.03	-9.44	-3.81
QGLK	-9.15	-8.77	-3.33
QGQG	-9.64	-8.77	-2.8
QGSK	-8.65	-9.09	-3.17
QGVV	-9.08	-8.37	-3.03
QGVV	-9.09	-8.3	-3.87
QPAM	-8.99	-9.53	-3.11
QPED	-9.15	-9.22	-2.48
QPEL	-8.68	-8.32	-3.15
QPLD	-9.05	-8.04	-3.56
QPLK	-9.4	-9.46	-4.49
QPQD	-8.56	-9.21	-2.79
QPRG	-8.65	-9.33	-2.81
QPRK	-8.81	-8.71	-3.61
QPSD	-8.95	-8.39	-3.51
QPSG	-9.06	-9.02	-3.33
QPSM	-8.71	-9.39	-4.19
QPVG	-8.98	-9.5	-3.26
RAED	-8.63	-8.69	-2.89
RAGG	-9.81	-8.82	-3.75
RAGK	-8.96	-9.29	-3.69
RAGM	-8.64	-9.14	-2.96
RALD	-9.37	-9.4	-3.33
RAQK	-8.64	-9.46	-3.01
RAQM	-8.63	-9.2	-2.95
RAQR	-9.65	-8.68	-2.84
RAQS	-8.65	-9.05	-2.85

RARM	-8.97	-8.9	-2.57
RASR	-8.88	-8.63	-2.83
RAVG	-8.58	-8.79	-3.6
RAVK	-8.58	-8.9	-3.36
RAVR	-9.45	-8.94	-3.19
RDAA	-9.04	-8.87	-3.34
RDAG	-9	-9.1	-2.68
RDAM	-8.8	-9.52	-2.93
RDVAV	-9.61	-8.92	-2.87
RDEM	-8.56	-8.72	-3.02
RDER	-9.26	-9.07	-4.15
RDGS	-9.01	-8.21	-3.28
RDGV	-8.83	-9.3	-3.13
RDLG	-9.14	-8.74	-3.01
RDQV	-8.65	-9.08	-3.97
RDRA	-8.67	-9.57	-3.16
RDRM	-9.48	-8.81	-2.44
RDRR	-8.93	-8.8	-3.01
RDSB	-8.66	-8.96	-2.62
RDSR	-8.76	-8.54	-2.99
RDVM	-8.59	-8.73	-3.37
REEL	-8.77	-7.97	-2.97
REGS	-8.55	-9.45	-3.91
REGV	-8.66	-9.01	-3.65
RELD	-8.99	-9.69	-2.92
RELM	-8.85	-9.5	-3.2
RELS	-8.65	-8.42	-3.13
REQD	-9.09	-9.29	-3.57
REQG	-8.86	-8.73	-2.75
RERL	-8.84	-9.39	-4.4
RERS	-9.35	-9.04	-2.91
RESV	-9.17	-8.46	-2.99
REVD	-8.71	-9.75	-3
REVG	-8.63	-8.29	-3.13
REVK	-8.82	-9.03	-2.83
RGAK	-9.18	-9.44	-3.02
RGAV	-8.57	-8.48	-3.47
RGEK	-8.71	-8.42	-3.02
RGLA	-10.04	-9.17	-3.05
RGLM	-9.03	-8.37	-3.89
RGLV	-8.85	-9.7	-4.35
RGQK	-8.77	-8.54	-2.77
RGRG	-8.73	-9.17	-2.93
RGRL	-8.65	-8.83	-3.33
RGRR	-9.1	-9.01	-3.11
RGRS	-8.68	-8.16	-3.33

RGSS	-9.47	-8.77	-2.9
RGVG	-9	-8.63	-3.74
RGVK	-8.67	-8.53	-3.1
RPAL	-8.94	-9.44	-2.91
RPAR	-9.72	-8.96	-3.08
RPAV	-9.12	-8.79	-3.64
RPEG	-9.46	-9.46	-3.25
RPQR	-8.64	-8.43	-2.71
RPRK	-9.09	-9.13	-3.17
RPRL	-8.63	-8.62	-3.77
RPSA	-9.1	-8.72	-4.12
RPVL	-8.92	-8.97	-3.11
RPVS	-9.05	-9.21	-3.05