Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
BIOCHIMICA, BIOLOGIA MOLECOLARE E BIOTECNOLOGIE

CICLO XXI

COORDINATORE Prof. Bernardi Francesco

MOLECULAR MODELLING STUDIES OF THE NF-κB BIOLOGICAL SYSTEM AS RELEVANT DRUG TARGET

Settore Scientifico Disciplinare BIO/10

Dottorando
Dott. Piccagli Laura

Tutore
Prof. Gambari Roberto

Anni 2006/2008

Contents

CHAPTER 1. General introduction

1.1 Molecular modelling in drug discovery

An increasing number of drug discovery projects have applied bioinformatics disciplines, one of the fastest growing fields in science.[1] Major application of bioinformatics concern sequence alignment, genome assembly, gene expression prediction, protein structure prediction, protein-protein and protein-ligand docking, and molecular dynamics simulations.[2] Moreover, the number of such projects is expected to continue to rise due to the Human Genome Project and high-throughput X-ray crystallography efforts.

Molecular modelling is a general name that concerns computational techniques and theoretical methods to model or simulate the behaviour of molecules.[3] Nowadays these techniques are applied in the fields of computational biology, computational chemistry and materials science for studying molecular systems ranging from small chemical systems to large biological molecules and material assemblies. In particular, the synergy between biological experimental and computational studies has greatly benefited both fields, providing invaluable information in many different areas of the life sciences.[4]

The historical process of developing concepts leading to molecular modelling started with the quantum chemical description of molecules. This approach yields excellent results on the ab-initio level, but the size of the molecular system which can be handled is still rather limited. Therefore, the introduction of molecular modelling as a routine research tool is due to the appearance of new technologies in computer graphics together with the development of molecular mechanism about 30 years ago. Moreover, the development of more accurate and reliable algorithms, the employ of more thoughtful protocols to apply them, and greatly increased computational power nowadays permits studies to be performed with the necessary reliability and accuracy. Thus, theoretical calculations, three-dimensional visualisation and manipulation of molecular systems can be used to gain new ideas and reliable working hypotheses for molecular interactions such as drug design.

The process of drug discovery involves the identification of candidates, synthesis, characterisation, screening, and assays for therapeutic efficacy. Once a compound has shown its value in these tests, it will begin the process of drug development prior to clinical trials. Despite advances in technology and understanding of biological systems, drug discovery is still a long process with low rate of new therapeutic discovery.[5,6]

A fundamental postulate in medicinal chemistry is that most drugs produce their effect in the human body as a consequence of the molecular recognition between a small ligand (the drug itself) and a biological relevant macromolecule (the target) such as an enzyme, DNA, glycoprotein or receptor. Thus, a key activity of early drug discovery involves the identification of small molecule modulators (lead compounds) of a target function.

The virtual screening (VS) is a computational approach to screen large of small compounds virtual libraries for the discovery of interesting biological active lead candidates against biological target, especially proteins.[7,8] The expression ‘virtual screening’ , coined in the late 1990s, was adopted by the scientific community in an effort to show that this in silico method is a serious alternative (or complementary) to the well-established HTS techniques.

In some studies reported in literature, the virtual screen had 'hit rates' (number of molecules that bind at a particular concentration divided by the number of compounds experimentally tested) that were 100-fold to 1,000-fold higher than those reached by empirical screen.[9,10] When used prior to experimental screening, VS can be considered as a powerful computational filter for reducing the size of a natural or chemical library that will be further experimentally tested. Of course, experimental drawbacks, such as organic molecules availability, limited solubility, aggregate formation or any sort of physical properties of substance material that could possibly interfere with the applied assay conditions do not need to be considered in the first phase of a virtual screening application.

This mean that VS can permit a more flexible evaluation of the chemical space of molecules and the creation of focused libraries that could comprise not necessary existing different chemical structures. Moreover, VS represents a quite good alternative for growing demand of faster and relatively cheap screening of small molecule driven by genomics and combinatorial chemistry. Structured based virtual screening method is based on the application of molecular docking technique which is designed to find the correct orientation and conformation (pose) of a ligand into the binding site of a target.

The knowledge of a well defined three-dimensional (3D) structure of the biological target under study must be available as starting point for the application of this strategy for finding new leads. Docking studies have now been used with different success for decades. Once the ligand has been docked, it should be scored according to the tightness of target-ligand interactions. The capability to correctly predict ligand-protein interactions is fundamental to any accurate docking algorithm and the necessary starting point for any reliable virtual screening protocol. In fact, information about the fit of molecules into the target is then used to rank the compounds with the aim of selecting and experimentally test a small subset for biological activity. The use of multiple receptor conformations in ligand docking seems probably the best choice to date.

Molecular dynamic (MD) is a computational method for obtaining all information about the dynamic properties of a protein, among them possible conformations states. Thus, in order to take in consideration the receptor flexibility in docking application, molecular MD technique could be of help. MD is a simulation of the time-dependent behaviour of a molecular system.[11] The addition of the time as forth dimension to structural biology allows that the positions in space and time of all atom in a protein can be described in detail. Although static structures are known for many proteins, the variety of functions of complex macromolecules is governed ultimately by their dynamic character.

The NF-κB consists of a group of eukaryotic inducible proteins that have evoked widespread interest until now.[12] In spite of the well-defined solved NF-κB 3D structures, to the best of our knowledge, nowadays VS applications against this interesting target for the discovery of new NF-κB/DNA inhibitors have not been published yet. The investigations described in this thesis bear upon the application in silico methodologies focusing on structured-based virtual screening against NF-κB transcription factors family.

1.2 Aim and outlines of this thesis

The project described in this thesis is aimed to the identification of low molecular weight compounds interacting with NF-κB and able to inhibit DNA/NF-κB complex formation and biological dependent functions. The short introduction described in the previous section clearly illustrate that structured-based VS has played an important role in lead compounds

discovery. In Chapter 2 a literary survey and clearly description is given on computational methods to be applied and on the targets under study. Chapter 3 deals with the implementation of a docking protocol suitable for NF-κB p50 monomer and homodimer. As a first step towards the development of an effective procedure, a data set of 27 bioactive natural compounds and two known NF-κB p50 binders were employed. We found that the adopted sampling method and the scoring function were successful in predict the ranking of active and inactive compounds into DNA binding site of the ensemble protein. The application of docking runs for the VS of a purpose-made prepared large chemical library is presented in chapter 4. A successful identification of a micromolar active compound structurally correlated to a substituted coumarin is described together with its binding modes. In chapter 5 a focus library of 1700 furocoumarin derivatives was prepared and in silico screened against NF-κB p50. Further in depth investigations of resulting five bioactive interesting hits against NF-κB p50 homodimer and p50-p65 heterodimer are reported revealing substantial differences. Binding studies on the best poses in the homodimer showed the involvement of an essential residue for p50 in vivo activity in ligand/homodimer interaction. Finally, in chapter 6 molecular dynamic simulations on NF-κB p50-p65 in three different systems (free form, in complex with DNA and bounded to IκBα) are presented. The dynamic properties and the most relevant protein conformations in different systems have been analysed in order to enlarge and optimise the VS strategy against the most common NF-κB heterodimer p50-p65.

References

1. Setubal CJ, Verjovski-Almeidaet S: Advances in Bioinformatics and Computational Biology. Brazilian Symposium on Bioinformatics, BSB held in Sao Leopoldo, Brazil, 2005.

2. Marcus F: Bioinformatics and Systems Biology: Collaborative Research and Resources. Springer, 2008.

3. Leach. AR: Molecular Modelling: Principles and Applications. Pearson Education, 2001.

4. Ciobanu J, Rozenberg G: Modelling in Molecular Biology. Springer-Verlag, 2004.

5. Kong D, Li X, Zhang H: Where is the hope for drug discovery? Let history tell the future. Drug Discov Today in press.

6. S kraljevic, P J Stambrook, K and Pavelic: Accel. Drug discovery. EMBO **2004**, *5*, 837.

7. Chin D, Chuaqui C, Singh J: Integration of Virtual Screening into the Drug Discovery Process. Mini-Reviews in Med Chem **2004**, *4*, 1053.

8. Shoichet BK, McGovern SL, Wei B, Irwin JJ: Lead discovery using molecular docking: Curr Opinion in Chem Biol **2002**, *6*, 439.

9. Doman TN, McGovern SL, Witherbee BJ, Kasten TP, Kurumbail R, Stallings WC, Connolly DT, Shoichet BK: Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. J. Med Chem **2002**; *45*, 2213.

10. Paiva AM, Vanderwall DE, Blanchard JS, Kozarich JW, Williamson JM, Kelly TM. Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis. Biochim Biophys Acta **2001**, *1545*, 67.

11. Henzler-Wildman K, Kern D: Dynamic personalities of proteins. Nature **2007**, *450*, 964.

12. Gilmore TD: Introduction to NF-kappaB: players, pathways, perspectives. Oncog, **2006**, *25*,6680.

CHAPTER 2. Literature survey

2.1 Computational methods

2.1.1 Molecular docking

Docking simulations represent a widely employed computational tool in drug discovery field, which attempts to predict a manifold of structures of intermolecular complex between at least two objects. The first docking attempts were done manually using interactive computer modelling. The ligand is put in the binding site and minimised to avoid bad steric clashes. This approach can be effective if we have a good idea of the expected binding mode. However, X-ray structures have revealed that even very similar inhibitors can adopt different binding modes.

Usually, but not necessarily, docking programs search along the conformational degrees of freedom of a small molecule (the ligand) while the protein is treated as a rigid body.

The results of such a search are configurations, namely conformations associated to a particular spatial orientation of the ligand at the binding site, which are usually referred to as solution poses.

A standard docking protocol consists of a step-wise process. First, a proper search algorithm predicts the various configurations of the ligand within the target binding site. In the second step, each docked pose is evaluated and ranked assessing the intermolecular interaction tightness throughout an estimation of the binding free-energy. Ideally, the correlation between the most favourable free-energy values and the best predicted poses should be very straight. The ability of a standard docking protocol to achieve its ultimate goal providing a reliable binding mode prediction, strongly depends on the accuracy of the scoring function used.[1,2]

2.1.1.1 Algorithms for ligand conformational sampling

The treatment of ligand flexibility can be summarized into three basic categories: systematic, stochastic and genetic algorithms.[3]

Systematic methods attempts to cover all the conformational degrees of freedom exploring each of them in a combinatorial way. Such methods provide an exhaustive search only in the limit of very rigid or simple molecules; otherwise a combinatorial explosion of the search dimensionality occurs, yielding the approach unfeasible. In order to avoid this, termination criteria are usually implemented which focus the sampling along regions of the conformational space that are more likely to lead an effective solution ("search and grow algorithms").

Conversely, stochastic approaches operate randomly selected changes along both the conformational internal and global (orientational/translational) degrees of freedom of the ligand, attempting to reach the global minimum for the molecule inside the binding site (Monte Carlo implementations). Within the Metropolis acceptance criterion, if the *i*-th solution of the conformational search bears an energy lower than the previous one (downhill move), it is always accepted. On the contrary, if the energy increases, a Boltzmann weighted probability function is then computed:

$$\rho(\Delta E) = e^{(-\beta \Delta E)} = e^{\left(\frac{-1}{k_B T}\Delta E\right)} \tag{2.01}$$

where ΔE is the energy difference between the $i^{\text{-th}}$ and the $(i\text{-}1)^{\text{-th}}$ configurations. To accept an uphill move with the probability ρ(ΔE) a random number is uniformly generated in the range {0, 1}. If the random number is less than ρ(ΔE) then the uphill move is accepted, otherwise it is rejected. Here, an higher temperature can be usually applied to explore a wider potential energy surface. The search is then interrupted when the desired number of configurations is obtained.

Genetic algorithms also implement a different amount of randomness; hence they should be formally classified between the stochastic ones. Nevertheless, compared with the properly called stochastic methods, they differ in the sense that they are based upon the principles of biological evolution and population dynamics, rather than on the laws of physics. Model

parameters representing the degrees of freedom are encoded in data strings called "chromosomes". Such chromosomes are evaluated by a proper fitness function, and individuals whose chromosomes bear the largest fitness values have a better chance to reproduce and indeed to transmit their genetic inheritance to the next generation. Chromosomes are randomly varied by means of genetic-like operators, usually mutations and crossover, in order to increase density and prevent premature convergence. When applied to the docking problem, the genetic algorithm solution is a population of putative ligand conformations. For instance, in the software AutoDock 3.0.5, genetic algorithm represents an hybrid search technique that implements an adaptive genetic algorithm with a local search feature. The local searcher performs an energy minimization after the global sampling, hence the local changes occurred due to minimization are mapped back into the chromosomes. Since inheritance of acquired traits clearly contravenes the Mendelian genetic laws, in this sense the genetic algorithm is named "Lamarckian" after the discredited evolutionary theory of Lamarck.

Similarly, another popular docking suite such as GOLD, employs a genetic algorithm whose most remarkable feature is the migration genetic operator. At the beginning, several subpopulations of chromosomes, called islands, are created instead of a large unique population. In order to preserve diversity, individuals are allowed to move among islands through the migration operator. Finally, only a fixed number of individuals can share the same place within an island. If there are more than a specified number of individuals in the same place, then the new individuals replaces the worst scoring member in the place, and not the worst individual in the overall population.

2.1.1.2 Scoring functions for docking poses evaluation

Once a pose has been generated for a docked compound in its binding site, it needs to be scored to rank the quality of the pose with respect to other poses for the compound itself, and with respect to other molecules in the library. In order to evaluate the interaction energy between the target and each docked ligand, different scoring functions has been developed. There are several requirements a useful scoring function should satisfy. Firstly, in a virtual screening simulation, the scoring function should be able to discriminate between correct and incorrect binding modes of known ligands with reasonable accuracy and speed. Secondly, binders should be clearly distinguished from inactive compounds.

The quantitative modeling of receptor-ligand interactions can be achieved by determining the equilibrium binding constant K_{eq}, which is in turn directly related to the Gibbs free- energy:

$$\Delta G = -RT \ln K_{eq} = \Delta H - T\Delta S \tag{2.02}$$

Therefore, the ideal scoring function should be able to calculate accurate binding free energies.

The difficulties relying on the estimation of the free-energy by means of computational techniques is extensively covered in literature, hence just an overview focused on the docking field will be given here.

Docking simulations are at now usually performed in vacuo, although in principle implicit solvation models could be used as well. Nevertheless, in spite of their theoretical derivation, scoring functions are usually able to provide a proper assessment of the enthalpic contribution for the free-energy (a force field-like potential energy function), whereas the entropic contribution remains hard to estimate. The main entropic contributions to the stability of the receptor-ligand complex are provided by desolvation effects, and by the internal conformational degrees of freedom of the docked small molecule. Within the docking field, the need for a fast scoring method led to a number of different functions which bring various assumptions and approximations in the evaluation of modelled complexes. Widely employed approximations are:

scoring functions assume that the free-energy of binding can be approximated using a single structure, which is a reasonable assumption since the lower is the energy of the configuration, the larger is its contribution to the partition function; the bound state for the complex is the only explicitly considered, whereas unbound components are implicitly accounted for; the free-energy is approximated by a linear combination of several terms, while several forces involved in the complex formation are non additive.

Empirical scoring functions provide an estimation of the binding free-energy by summing up interaction terms derived from structural parameters. The development of scoring functions is based on the idea that binding energies can be approximated by means of a sum of not correlated terms, which are derived by regression analysis from experimentally determined structures whose binding mode are known. Such kind of scoring functions are simple and

intuitive, but their main drawback is that it is not clear whether they are able to predict the binding affinities for ligands whose structure is not covered among the training set.

2.1.2 Classical Molecular Dynamics simulations

The molecular structures are usually represented as static objects, but are actually dynamic entities. Dynamics simulations are useful in studies of the time evolution of a variety of systems at non-zero temperatures, including biological molecules in a variety of states, for example, crystals and aqueous solutions. The major applications of molecular dynamics include performing conformational searches, generating statistical ensembles and studying the motions of molecules. Thus many relevant macromolecular features such as phase space accessible, energetic, thermodynamic, structural, and dynamic properties can be studied. One of the most important use is to investigate the range of possible configurations that molecule may assume into phase space to numerically simulate its dynamics behaviour.[4,5] In the classical dynamic simulation this can be realized by calculating a trajectory, that is a number of time-dipendent configurations through the integration of the Newton equations of motion and using empirical force fields.

In this scheme, the fundamental equation relies on the Newton's second law:

$$\mathbf{f}_i(t) = m_i \mathbf{a}_i(t) \tag{2.03}$$

where $\mathbf{f}_i$, m_i and $\mathbf{a}_i$ stand for the force, the mass and the acceleration of atom i respectively.

The force at time t on atom i can be directly calculated from the derivative of potential energy V in respect to the coordinates $\mathbf{r}_i$.

$$\frac{\partial V}{\partial \mathbf{r}_i} = m \frac{\partial^2 \mathbf{r}_i}{\partial t_i^2} \tag{2.04}$$

where the potential V is accounted for by means of the potential energy function provided by the force field.

For conservative systems (invariant potential function in time), the force acting on the i^{th} particle is a function of the coordinates only. Since the potential energy function (which is independent of velocities and time) required for the force calculation is provided by the force field, initial velocities are solely required in order to evolve the system (as starting coordinates are obviously known).

Basically, starting velocities of all particles involved are initialized by random selecting from a Maxwell-Boltzmann distribution at the temperature of interest:

$$\mathbf{f}_{i,t} = (v_{ix}) = \left(\sqrt{\frac{m_i}{2\pi k_B T}}\right) e^{\left(-\frac{m_i v^2_{ix}}{2k_B T}\right)} \tag{2.05}$$

where k is Boltzmann's constant.

Since it represents a Gaussian distribution, it can be easily obtained from a random number generator. Therefore, dynamics runs cannot be repeated exactly, except for forcefield engines that allow setting the random number seed to the value that was used in a previous run.

As a consequence of a continuous potential, the motion of all particles is tightly correlated, giving rise to a many-body problem which can not be analytically solved. To overcome this, equations of motions are integrated using a finite difference method, where integration is performed on discrete time intervals δt. By doing so, forces are constant during each time step, and consequently collisions are elastics.

Thus for each integration step, the forces at time t are calculated by differentiating the potential energy function:

$$\mathbf{f}_{i,t} = -\frac{\partial V(\mathbf{r}_{i,t})}{\partial \mathbf{r}_{i,t}} \tag{2.06}$$

The force on an atom may include contributions from the various terms in the force field, and represents the most time consuming part for a molecular dynamics code. The equations of motions are then integrated by means of a suitable algorithm. It should be fast, ideally requiring only one energy evaluation per timestep, a little computer memory and permit the use of a relatively long timestep. There are different algorithms to use in MD, each of them

assumes that the positions and dynamic properties (velocities, accelerations, and so on) can be approximated by a Taylor series expansion. The Verlet algorithm[6] represents a third-order truncation, and uses positions and accelerations (previously computed) at time *t*, and the positions from the previous step. The variants leap-frog Verlet algorithm are perhaps the most widely used methods in molecular dynamics. Starting with $\mathbf{r}(t)$, $\mathbf{v}(t-\Delta t/2)$, and $\mathbf{a}(t)$ that represents position, velocity and acceleration at t, $t-\Delta t/2$ and t respectively, it calculates:

$$\mathbf{v}(t+\frac{1}{2}\Delta t) = \mathbf{v}(t-\frac{1}{2}\Delta t) + \Delta t\mathbf{a}(t) \tag{2.07}$$

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \Delta t\mathbf{v}(t+\frac{1}{2}\Delta t) \tag{2.08}$$

$$\mathbf{a}(t+\Delta t) = \frac{\mathbf{f}(t+\Delta t)}{m} \tag{2.09}$$

where $\mathbf{f}(t+\Delta t)$ is evaluated from $-dV/d\mathbf{r}$ a $\mathbf{r}(t+\Delta t)$

It should be point out that with this algorithm the calculated positions and the velocities are confused of an half *timestep.*

The choice of the size of the time-step is key parameter in MD primarily determined by a compromise between accuracy and speed of the calculation. Unfortunately by using a large time step to increase the speed of the calculation would lead to instability and inaccuracy of integration procedure. In most organic models, the highest vibrational frequency is that of C-H bond stretching, whose period is on the order of 10^{-14} s (10 fs). The integration timestep should therefore be about 0.5-1 fs. As a safe rule of thumb, the time-step should be approximately one tenth the time of the shortest period of motion. When using a constraint algorithm (such as SHAKE)[7,8] a longer timestep is possible.

Summarizing, the global molecular dynamic algorithm start with the input initial conditions: (i) the potential interaction V **,** as a function of atom positions, (ii) the positions $\mathbf{r}$ of all atoms in the system, and (iii) the velocities $\mathbf{v}$ of all atoms in the system. Starting by this point, the subsequent flow is repeated for the required number of steps. Thus, the forces of any atom is

computed by calculating the force between non bonded atom pairs plus the forces due to bonded interactions and plus restraining and/or external forces. The potential and kinetic energies and the pressure tensor are computed. Subsequently, the movement of the atoms is simulated by numerically solving Newton's equations of motion (see eq. 2.03). Finally, positions, velocities, energies, temperature, pressure, etc are stored, as required, in the output files.[9,10]

The important output of the MD run is the trajectory file which contains particle coordinates and, optionally, velocities at regular intervals. Since the trajectory files are lengthy, one should not save every step. To retain all information it suffice to write a frame every 15 steps, since at least 30 steps are made per period of the highest frequency in the system, and Shannon's sampling theorem states that two samples per period of the highest frequency in a band-limited signal contain all available information. But that still gives very long files. So, if the highest frequencies are not of interest, 10 or 20 samples per ps may be acceptable. The user should be aware of the distortion of high-frequency motions by the stroboscopic effect, called aliasing: higher frequencies are mirrored with respect to the sampling frequency and appear as lower frequencies.[11]

2.1.2.1 Essential dynamics

Analysis of extended molecular dynamics simulations of proteins in vacuum and in aqueous solution reveals that it is possible to separate the configurational space into two subspaces: (a) an "essential" subspace containing only a few degrees of freedom in which anharmonic motion occurs that comprises most of the positional fluctuations; and (b) the remaining space in which the motion has a narrow Gaussian distribution and which can be considered as "physically constrained". If overall translation and rotation are eliminated, the two spaces can be constructed by a simple linear transformation in Cartesian coordinate space, which remains valid over several hundred picoseconds. The transformation follows from the covariance matrix of the positional deviations. The diagonalization of the covariance matrix of the coordinate fluctuations produces a set of eigenvalues associated to eigenvectors. In mathematics, given a linear transformation, an eigenvector of that linear transformation is a nonzero vector which, when that transformation is applied to it, changes in length, but not in direction. For each eigenvector of a linear transformation, there is a corresponding scalar

value called eigenvalue for that vector, which determines the amount the eigenvector is scaled under the linear transformation. Applied to the dynamic simulation analysis context, the eigenvectors represent the direction of motions and the eigenvalues the amount of the motions.

The essential degrees of freedom seem to describe motions which are relevant for the function of the protein, while the physically constrained subspace merely describes irrelevant local fluctuations. The near-constraint behavior of the latter subspace allows the separation of equations of motion and promises the possibility of investigating independently the essential space and performing dynamic simulations only in this reduced space.[12]

2.2 Structured-based Virtual Screening strategies

Structured-based VS (SBVS) covers a variety of sequential computational phases including the application of the docking method as described above. The Figure 1 reports the classical workflow of a SBVS approach, including further experimental studies.[13] Essentially, molecules from a prepared virtual database are docked into the protein and a suitable scoring function is used to evaluate the binding affinity. Subsequently, the virtual hits are tested in biological assay for activity. Finally compounds showing biological activity (lead compounds) are further modified using computational methods, medicinal chemistry, structure–activity relationship and structural studies to get second generation compounds

The enrichment factor (EF) for a library built with the *n* top compounds of the ranked library can be defined as:[14]

$$EF(n) = \frac{Hits_n}{N_n} \Big/ \frac{Hits_{total}}{N_{total}} \tag{2.10}$$

and expresses the relative change in the probability of finding a ligand in the focused library when compared to a random pick from the complete library.

The successful application of SBVS depends on a wide range of factors based on the information accessibility and computational techniques adopted on the case-study. Thus the EF can define in manifold modalities. In common with HTS, Brian K. Shoichet[15] point out

that we can define success in VS as “finding some interesting new ligands”, and not as “correctly ranking all the molecules in a library” or “find all the possible ligands in a library”. In this chapter we describe some focusing considerations referenced in literature on the application of these strategies and of relevance for the work described in chapters 3, 4 and 5 of this thesis.[16-22]

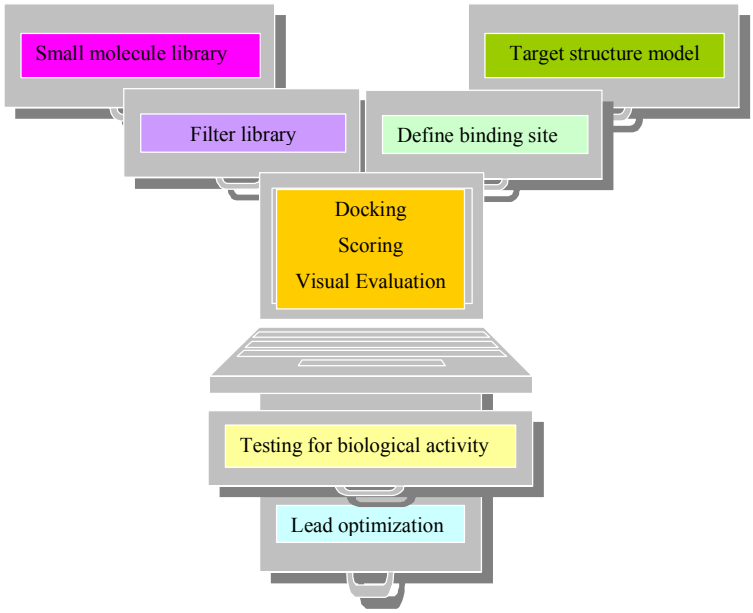


Figure 1. Typical docking-based virtual screening scheme.

2.2.1 Sources for target structure and modality of selection

High-quality structural information for a biological macromolecule can derived from X-ray crystallography and NMR techniques together with sophisticated homology modeling studies. From a practical point of view in the case of a computational scientist using molecular docking simulations to screen a database of compounds essentially means to have a file that describe the Cartesian coordinates of each atom of the protein in the 3D space. However, in the case of different configurations of the same target, the appropriate 3D geometry selection depends on the amount and the quality of experimental information available about the

system under investigation. In particular, for commonly used crystal structures, the resolution of the diffraction data is the first issue to be considered.

The RCSB protein data bank (PDB), founded in 1971, provides an online archive of experimental data and 3D coordinates of more than 40000 experimentally determined biological structures such as protein, nucleic acid and macro-molecular complex. Recently, the founding members such as RCSB PDB,[23] the Macro-molecular Structure Database at the European Bioinformatics Institute (MSD-EBI)[24] and the Protein Data Bank Japan (PDBj)[25] at Osaka University, have established a collaboration in order to ensure a single and uniform database of PDB data through the so-called worldwide Protein Data Bank (wwPDB)[26]

The BioMagRes-Bank (BMRB) [27] at the University of Wisconsin-Madison became a member in 2006 and represent a deposition site for primary experimental data and PDB data. Nowadays, as sanctioned by wwPDB committee, theoretical model depositions to PDB (such as models determined by homology or ab initio methods) are no longer accepted. Despite this, receptor structures from computer-based modelling are available through databases such as MODBASE.[28]

Docking simulation trough dedicated software, can be applied to a wide range of targets comprising well-known target sites having many examples of co-crystallized structures as well as new genomic targets. However, the availability of co-cristallized ligands from PDB is of great value for the individuation of the functional binding site and for assessing the success of protein-ligand docking pose prediction. In fact, usually the accuracy of docking poses is quantified by calculation of the root mean square deviation (RMSD) between the ligand conformations as experimentally found into the protein with the pose derived by docking runs. In some situations, as in the case of a doubted quality of representation of the complex interactions of the ligand with the target, the RMSD can not be considered an as reliable descriptor of VS performance. The binding mode shown by an experimental ligand can differ with soaking and co-crystallization conditions.[29] Thus, before selecting a particular crystal structure as reference for a virtual screening protocol, expanded analysis of parameters such as population of the bound ligand and consistency of the hydrogen bond network is advisable.

VS approaches should take into consideration the very common phenomenon such as structural changes in the macromolecular target upon ligand binding. Ignoring this effect might have a severe impact on ligand docking and virtual screening. Moreover protein flexibility should be incorporated into structure-based drug design in order to portray a more

accurate representation of a target in solution and to accurately predict the orientation and interactions of ligands within the binding site. Furthermore, dealing with flexibility is in many cases essential as required by targets in which this property is considerable and appropriate to their function (e.g. the serine protease factor).[30] So the big challenge of scientific community is to routinely incorporate flexibility considerations into structure-based drug discovery in an affordable computing time. The implications of protein flexibility in drug discovery have been recently evaluated. Early attempts to include protein flexibility in ligand docking, such as soft docking and partial side-chain flexibility among others have been reviewed. However, most of these methods do not include backbone rearrangements, and explicit sampling of side chains is an insurmountable drawback in VS of large compounds libraries. The conclusion is that using multiple fixed protein conformations either experimental or computationally generated seems a practical alternative and probably the best choice to date.[31] Nevertheless, the carefully selection of few conformations rather than using all possible random structures, seem to be advantageous in VS application especially in order to avoid the increasing of false positive rate.[32] Molecular dynamic simulations can also be used as a source of alternative conformations even if the are some important problems relating to the applications of this computational method. May the large scale movements occur on time scales that would made MD impracticable? On contrary, what fraction of the generated alternative conformations are artefacts? However, different applications are reported in literature such as Wong et al. that docked balanol to MD trajectory snapshot of protein Kinasi A.[33]

In conclusion, the choice of the specific strategy to be used depends on the question asked and on the feasibility of the method to yield reliable results at the present state of the art.

2.2.2 Database of small compounds availability

Databases are collections of molecules that may be physically available or may exist "virtually" in the form of files. The last ones are the databases analyzed by VS applications.[34] Virtual database can be created as in house collection on the basis of a specific research project, but usually several web-accessible libraries, which in some cases contain up to 1-2 million of compounds that could represent new potential candidates for a large number of biological targets, are employed. Usually, commercial (especially larger) databases contain

molecules from different suppliers and the website, from which you get the database, can also get information on availability, cost and how to sort the compounds of interest.

The commercial databases are divided in four categories: reference database, the fine chemical database, the database of biologically active compounds and libraries for screening.

The first define the boundaries of real chemical space and generally do not report information on the availability of compounds and they are not in a format that allows the structural analysis of the content. Belong to this class the following databases: *Beilstein*[35], and the *Chemical Abstracts Services* (CAS) *Registry Database*[36] which contains all chemicals that have appeared in scientific literature from 1957 to today, with some structures dating to the early of the twentieth century.

The fine chemical database contains both biologically active compounds and molecules suitable for chemical modifications (often used as building blocks of combinatorial chemistry) which format allows the structural analysis of the content. Examples are *Maybridge*[37] and the *Available Chemicals Directory* (ACD).[38]

Databases of biologically active compounds are made up exclusively of molecules with proven biological activity and represent the current chemical space of active compounds. They are often used as a reference (eg. to determine the profile of drug-Likeness of other libraries) being in a format that allows structural analysis, even though often the substances are not "physically" available. This class includes the following databases: *Modern Drug Data Report* (MDDr),[39] *Comprehensive Medicinal Chemistry* (CMC), *World Drug Index* (WDI)[40] and *National Cancer Institute* (NCI).[41] Nowadays, NCI provides 250251 open structures reading for searching and will provide at least some of these for experimental testing.

Finally, libraries of screening are constituted of compounds, designed and selected for screening (both VS and HTS) which is not known the eventually biological activity. Some examples are the following: *Optiverse*[42]*, Compass Array*[43], *PHARMACophore*[44], and *ACD Screening Compounds* (ACD-SC)[38] which collect more than 1.8 million compounds from 40 different suppliers.

In order to employ these libraries, molecular properties such as charge, protonation states, accessible conformations and solvation must be calculated. Even details such as stereochemistry, tautomerization and protonation, which we frequently take for granted, are often ambiguous, or can change on binding to a macromolecular target.

Recently, the ZINC database[45] (http://blaster.docking.org/zinc/) containing 4,6 million commercially available 3D structures, has been prepared especially for virtual screening application. It is developed by John Irwin in the Shoichet Laboratory in the Department of Pharmaceutical Chemistry at the University of California, San Francisco. ZINC is different from other chemical databases because it aims to represent the biologically relevant, three dimensional form of the molecule. ZINC uses catalogues from over 50 compound vendors and is updated regularly and may be downloaded and used free of charge.

Although there are several commercial and freeware molecular chemical databases for virtual screening applications, there is still just few databases of natural compounds available on the web. Among them the *Traditional Chinese Medicinal Database* (TCMD)[46] consisting natural products based on reported metabolites from TCM plants or the Dictionary of *Natural Product Database* (DNP)[47], the *Marine Natural Product Database* (MNPD)[48] and the *Comprehensive Herbal Medicine Information System for Cancer* (CHMIS-C).[49]

According to the restricted free access to 3D libraries of natural products, the number of VS studies performed is limited. In fact, although there are encyclopaedias of natural compounds, the access to natural products 3D libraries is often very difficult due to often unclear availability of structures and necessity of consulting catalogues of different suppliers[50]

The *SuperNatural Database*[51] is the first public resource containing 3D structures and conformers of 45 917 natural compounds, derivatives and analogues purchasable from different suppliers. Viewing requires the free Chime-plugin from MDL (Chime) or Java2 Runtime Environment (MView), which is also necessary for using Marvin application for chemical drawing.

The profile of a database is defined primarily by two factors: diversity and drug-Likeness.

Diversity is an estimation of the percentage of the chemical area covered by Chemical Database: the greater the chemical diversity, the higher the probability of finding the database within a compound that can interact with a specific protein or nucleic acid. However, as with the criterion “similarity”, the expression “diversity” is a relative measure that depends from the specific target and the difference in binding process of two ligands to that structure. It should be pointed that in some case also a very small modification of a ligand topology, such as a methyl group, might have a dramatic effect on binding affinity.

The drug-Likeness term refers to the specific molecular characteristics that distinguish drugs from other natural and synthetic compounds. It is often used as a filter for an initial

“negative” selection of those molecules of the database that could never become drugs. The elimination of these structures also carry to a significant reduction of the calculation time involved during the VS.

There are several algorithms that evaluate the drug-like character of a molecule on the base of chemical and physical properties (logP, molecular weight, charges, etc.), the absence of reactive or toxic functional groups and the presence of functional groups considered important for biological activity. Many of these algorithms use the "rule of 5" by Lipinski[52] which assesses the bioavailability of a molecule based on logP (≤ 5), the molecular weight (≤ 500), the number of acceptor groups of hydrogen bond (≤ 10) and the number of donor groups of hydrogen bond (≤ 5).

The criteria for assessing the drug-Likeness of a molecule have been deducted from structural drugs on the market and all those entered into clinical trials: eg., the rule of Lipinski was deducted featuring a subset of 2245 molecules in the World Drug Index.

These molecules, however, define the current chemical space of compounds, which is only a small fraction of the total chemical space of active compounds. In fact, a molecule rejected by the filter as not drug-like might actually belong to a new class of active compounds not yet identified by pharmaceutical research. The drug-Likeness filters should be employed carefully and wisely, otherwise a too restrictive use should lead to the elimination of many drug-like molecules or to structures that could become drug-like with few changes or chemical substitutions (false-negative).

Important is also decide at what level in the process of drug discovery is advisable to apply filters. In forward filtering, the decrease of the initial data sample, carry to a significant reduction of calculation time because a smaller number of molecules are involved in the subsequent steps of VS and all the process is more efficient. Thus, more elaborate docking protocols can be performed, for example taking into account multiple conformational states by using this approach. The disadvantage is the production of false-negative derived from a too restrictive use of filters. Backwards filtering can instead use a greater amount of information in assessing the inhibitory and therapeutic potential of a new molecule, although these strategies will provide more false-positive hits.

2.2.3 Performance Objectives for docking-based VS

In order to be valuable in a lead-discovery context, a VS method must be sufficiently fast to screen databases on a time scale compatible with the needs of a drug-discovery project.[53] Moreover it should be able to search the active-site space of the receptor and the conformational space of the ligand and locate the correct binding mode of each legend.[54] Incorrect prediction of the binding mode can produce poor predicted binding affinities, and can provide a misleading qualitative picture of what parts of the molecule are crucial for binding. If taken seriously, such an error could result in a drawback in lead-optimization phase. A remarkable question is to predict relative binding affinities sufficiently accurately to yield acceptable levels of enrichment and to allow poor ligands to be rejected.

With these considerations in mind, which is the best software for docking and scoring tools to use on a particular target? To answer this question, a number of papers have been published, in which different docking packages and scoring functions have been compared.[55-57]

The conclusion is that no one of these methods is clearly better and that each approach works better for certain classes of target and in any case all failed to give a useful prediction of the ligand binding affinity.

Nevertheless, Emanuele Perola at al.[57] from a comparison of different docking and scoring methods on systems of pharmaceutical relevance, overall "Glide appears to be a safe general choice for docking, while the choice of the best scoring tool remains to a larger extent system-dependent and should be evaluated on a case-by-case basis".

2.2.4 GLIDE: a software for VS applications

Glide (grid-based ligand docking with energetics) is designed to address a number of problems ranging from database screening to highly accurate docking.[58-60]

Glide use a hierarchical series of filters to search for all possible locations of the ligand in the putative binding site of the protein target. The shape and the properties of the macromolecule are represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand poses. These fields are generated as preprocessing steps in the calculations and hence need to be computed only once for each protein. The Glide algorithm is based on a systematic search of positions, orientations, and conformations of the ligand in

the protein binding site using funnel type approach. The conformational flexibility is treated first dividing the ligand structure into a core region and some number of rotamer groups. Each rotamer group is attached to the core by rotatable bond, but does not contain additional rotatable bonds. During conformation generation, each core region is represented by a set of core conformations. Each core plus all possible rotamer group conformations is docked as a single object.

These conformations are selected from an exhaustive enumeration of the minima in the ligand torsion-angle space. Given these ligand conformations, initial screens are performed over the entire phase space available to the ligand to locate promising ligand poses. The search begins with the selection of "site points" on an equally space 2 Å grid that permeate the binding site area. After that the ligand centre is positioned on the site point if the matching is good enough. So, an exhaustive search of possible locations and orientations is performed over the active site of the protein. If the are too many steric clashes with the receptor, the orientation is skipped. If the score is good enough, all the interaction with the macromolecule are scored. This pre-screening drastically reduces the region of phase space over which computationally expensive energy and gradient evaluations will later be performed while at the same time avoiding the use of stochastic methods.

The scoring is carried out using the GlideScore function (Glide energy used to predict binding affinity).

Only a small number of the best refined poses are passed on the subsequent stage in the hierarchical energy minimization on the precomputed OPLS-AA van der Waals and electrostatic grids for the protein. The best poses obtained from this stage are subjected to Monte Carlo simulations and the minimized poses accepted are then rescored using the GlideScore function.

2.2.4.1 Glide Methods

Glide employs two forms of GlideScore: (i) Glide- Score SP, used by Standard-Precision Glide; (ii) GlideScore XP, used by Extra-Precision Glide. These functions use similar terms but are formulated with different objectives in mind. Specifically, GlideScore SP is a "softer", more forgiving function that is adept at identifying ligands that have a reasonable propensity to bind, even in cases in which the Glide pose has significant imperfections. This version seeks to minimize false negatives and is appropriate for many database screening

applications. In contrast, GlideScore XP is a harder function that exacts severe penalties for poses that violate established physical chemistry principles such as that charged and strongly polar groups be adequately exposed to solvent. Thus XP mode is a refinement tool designed for use only on good ligand poses. This version of Glide-Score is more adept at minimizing false positives and can be especially useful in lead optimization or other studies in which only a limited number of compounds will be considered experimentally and each computationally identified compound needs to be as high in quality as possible. Glide score is a more sophisticated version of ChemScore.

The ChemScore can be written as:

$$\Delta G_{bind} = C_0 + C_{lipo}\sum f(r_{lr}) + C_{hbond}\sum g(\Delta r)h(\Delta\alpha) + C_{metal}\sum f(r_{lm}) + C_{rotb}H_{rotb} \qquad (2.11)$$

The summation in the second term extends over all ligand-atom/receptor-atom pairs defined by ChemScore as lipophilic, while that in the third term extends over all ligand-receptor hydrogen-bonding interactions. In eq. 2.11, *f*, *g*, and *h* are functions that give a full score (1.00) for distances or angles that lie within nominal limits and a partial score (1.00-0.00) for distances or angles that lie outside those limits but inside larger threshold values.

GlideScore 2.5 SP modifies and extends the Chemscore as follows:

$$\begin{aligned}\Delta G_{bind} = {} & C_{lipo-lipo}\sum f(r_{lr}) + C_{hbond-neut-neut}\sum g(\Delta r)h(\Delta\alpha) + C_{hbond-neut-ch\arg}\sum g(\Delta r)h(\Delta\alpha) + \\ & C_{hbond-ch\arg-ch\arg}\sum g(\Delta r)h(\Delta\alpha) + C_{\max-metal-ion}\sum f(r_{lm}) + C_{rotb}H_{rotb} + \\ & C_{polar-phob}V_{polar-phob} + C_{coul}E_{coul} + C_{vdW}E_{vdW} + \ solvation\ term\end{aligned} \qquad (2.12)$$

The lipophilic-lipophilic term is defined as in ChemScore. The hydrogen-bonding term also uses the ChemScore form but is separated into differently weighted components that depend on whether the donor and acceptor are both neutral, one is neutral and the other is charged, or both are charged.

The metal-ligand interaction term (the fifth term in eq. 2.12) uses the same functional form as is employed in ChemScore but presents important differences. For example, this term considers only interactions with anionic acceptor atoms. This modification allows Glide to

recognize the evident strong preference for coordination of anionic ligand functionality to metal centers in metalloproteases.

The seventh term, from Schrodinger's active site mapping facility, rewards instances in which a polar but non-hydrogen-bonding atom (as classified by ChemScore) is found in a hydrophobic region. The second major component is the incorporation of contributions from the Coulomb and vdW interaction energies between the ligand and the receptor. The Coulomb-vdW energies used in GlideScore 2.5 (but not those used in Emodel) employ these reductions in net ionic charge except in the case of anionic ligand-metal interactions, for which Glide uses the full interaction energy.

The third major component is the introduction of a solvation model. Like other scoring functions of this type, previous versions of GlideScore did not properly take into account the severe restrictions on possible ligand poses that arise from the requirement that charged and polar groups of both the ligand and protein be adequately solvated. Charged groups, in particular, require very careful assessment of their access to solvent. In addition, water molecules may be trapped in hydrophobic pockets by the ligand, also an unfavorable situation. To include solvation effects, Glide 2.5 docks explicit waters into the binding site for each energetically competitive ligand pose and employs empirical scoring terms that measure the exposure of various groups to the explicit waters. This "water-scoring" technology has been made efficient by the use of grid-based algorithms.

Using explicit waters, as opposed to a continuum solvation model, have significant advantages. In the highly constrained environment of a protein active site containing a bound ligand, the location and environment of individual water molecules become important. Current continuum solvation models have difficulty capturing these details, but glide explicit-water approach has allowed the development of consistently reliable descriptors for rejecting a high fraction of the false positives that appear in any empirical docking calculation. These analysis also produced trial values for the coefficients of the various penalty terms. For the most part, these coefficients were used without modification in GlideScore 2.5 XP (the research on this version is not published). For Glide 2.5 SP, however, the need to make the program relatively fast limits the amount of sampling that can be done during docking and hence limits the accuracy of the docked poses. As a result, optimization of the solvation penalties led to smaller coefficients that do not too heavily penalize misdocked actives. It is

clear from enrichment factors for many database screens reported in literature that SP method in not so effective in separating “active” from “inactive” compounds.
In GlideScore 4.0 XP the sampling methods and the scoring functions have been enhanced.
In particular XP Glide scoring incorporates desolvation effects and an improved model of hydrophobic interactions designed to discriminate between different geometrical protein enviroments.[60]

2.3 Biological system: NF-κB as interesting drug target

The NF-κB family consists of a group of eukaryotic inducible transcription factors, discovered by David Baltimore’s group in1986.[61] These transcription factors have evoked widespread interest until now. In fact, the dysfunction of NF-κB is associated with many disease states such as atherosclerosis, asthma, arthritis, cancer, osteoporosis in β-thalassaemia, muscular dystrophy, and viral infections.[62-65] As a matter of fact, the NF-kB/Rel family is involved in the control of the expression of a number of mammalian genes, such as those encoding for major histocompatibility complex (MHC) proteins, interferons and growth factors.[66,67] In addition, transcription factors belonging to the NF-kB/Rel family are involved in the transactivation of viral genomes, such as HIV-1.[68,69]
With regard to viruses’ families, several of them promote their replication, prevent virus-induced apoptosis and mediate the immuno response by activating NF-kB. Thereby, the target genes under the regulation of NF-κB include a variety of cellular as well as viral genes and the appropriate regulation and control of NF-κB activity plays an essential role in several aspects of human health. Importantly, NF-κB has been shown to be the target of several anti-inflammatory and anticancer drugs.[70-74]
NF-κB comprises homo- and hetero-dimers which are related through a highly conserved DNA-binding/dimerization domain called the Rel homology. Five polypeptide subunits, p50, p52, p65 (RelA), RelB, and c-Rel constitute the mammalian NF-κB family. Many dimeric forms of NF-κB have been detected. The NF-κB dimers bind decamaric sequences known collectively as kB DNA collocated in the enhancer regions of a large number of target genes. Although the homodimers p50 are commonly observed, the classic form of NF-κB is the

heterodimer p50/p65. In particular, p50-p65 heterodimer is the predominant NF-κB complex in the T-cell in case of HIV-1.

The metabolic regulation of NF-κB biological functions involves several control levels, one of the most important being in the interaction with proteins belonging to the IκB family in the cytoplasm. Among those, IκBα plays an inhibitory role, generating a complex with the inactive and resting state form of NF-κB homo- or hetero-dimers; this molecular interaction prevents NF-κB from translocating to the nucleus and exerts its regulatory functions on transcription of NF-κB target genes. The effect of IκBα is counteracted by IκBβ, that is able to compete with IκBα in binding NF-κB, thus leading to NF-κB activation. In response to a variety of stimuli including physical and chemical stresses, cytokines, reactive oxygen intermediates and ultraviolet light, the latent cytoplasmatic NF-kB/IκB complex is activated by the I-κB Kinase (IKK) complex. IKK is formed by three distinct subunits: IKKα, IKKβ and IKKγ. In the classical or canonical pathway, activation of IKK complex leads to the phosphorylation by IKKβ of two specific serines near the N terminus of IκBα, which targets IκBα for ubiquitination and degradation by the 26S proteasome. The unmasked NF-κB can then enter the nucleus and binds to the DNA target elements present in the promoters of NF-κB regulated genes, as well as to co-activators of gene transcription to activate target gene expression.

Moreover, cellular reducing catalyst thiredoxin (Trx) is a small endogenus molecule involved in NF-κB activation. Trx include in its structure a characteristic active site amino acid sequence, Cys-Gly-Pro-Cys able to reduce disulfide bonds.[75] Some studies reported in literature shown that for the binding of NF-κB to the kB DNA, the cysteine residues must be in the reduced state.[76,77] The DNA binding region of NF-κB p50 comprises a specific sequence composed by three arginine residues and a cysteine (Cys62) that is in oxidable state in cytoplasmatic cellular compartment condition. Thus, Trx recognizes and reduces NF-κB disulfide bond involving Cys62 in the nucleus allowing the protein/DNA complex formation, even if this binding process in vivo is not yet clarify.

2.3.1 Inhibitors of NF-κB and its dependent biological functions.

Based on the different molecular targeted approaches aimed at inhibition of NF-κB dependent biological functions[78], the inhibitory molecules described in literature can be divided in the following classes:

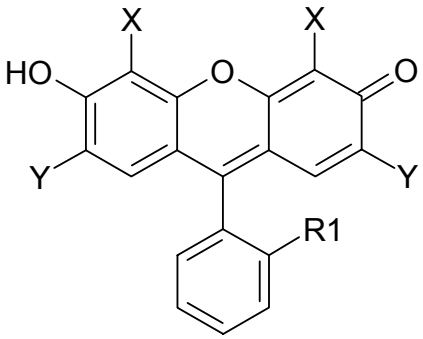
1. Anti-oxidants that have been shown to inhibit activation of NF-kB
2. Proteasome and proteases inhibitors that inhibit Rel/NF-kB
3. IkBa phosphorylation and/or degradation inhibitors
4. Miscellaneous inhibitors of NF-kB
5. DNA mimics (NF-κB PNA-DNA chimeras) that bind NF-kB[79]

Several experimental model system are available for biological validation of molecules inhibiting NF-κB function. In this respect, transcription factor (TF) decoy strategy, based on inhibition of the binding of TF-related proteins to the consensus sequences present in the promoter of specific genes, demonstrated to modulate gene expression in vitro[80,81] and in vivo.[82,83] For example in cystic fibrosis (CF), several approaches can be followed, such as the identification by electrophoretic mobility shift assay (EMSA) of transcription factors activated by P. aeruginosa infection and their putative consensus sequences in the promoters of genes activated by this pathogen in CF epithelial cells. This information will help in identifying putative target transcription factors. NF-κB is one of the transcription factors demonstrated to be involved in CF. Interestingly; NF-κB plays a crucial role in regulating the expression of genes induced during P. aeruginosa infection, including interleukin 8 (IL-8) and intercellular adhesion molecule 1 (ICAM-1). Accordingly, molecules targeting the NF-κB signal transduction pathway are expected to be active on IL-8 and ICAM-1 gene expression.[84] Besides the importance of these data for the theoretical point of view, these evidences are of great interest for the practical point of view, suggesting this treatment as a possible strategy for the therapy of inflammation associated with cystic fibrosis.[85]

Target inactivation of NF-κB by directly binding to the p50 subunit is an important point of action of structurally different known inhibitor molecules[86-92] such as Andrographolide and Selencotyanates that form covalent adducts with Cys-62 of p50. Recent studies have shown that p50 subunit of NF-κB is the one that mainly interacts with HIV-1 long terminal repeat.

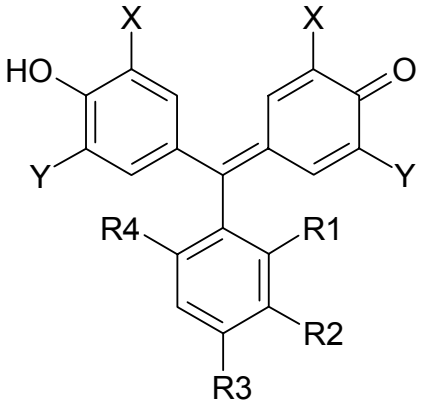
Interestingly, κB site-dependent transcriptional activation by homodimer p50 has been demonstrated in vitro.[93] In particular, homodimers p50 specifically stimulated transcription from the immunoglobulin κ-site (equivalent to the HIV-1 enhancer site). Aurintricarboxylic acid (ATA) analogues (e.g., Bromopyrogallol Red, **1i**; Pyrogallol Red, **2i**; Pyrocatechol Violet, **3i**) and polyhydroxycarboxylate derivatives (e.g., Gallic Acid, 4i; 5,7-dihydroxy-4-methylcoumarin, 5i; 7,8-dihydroxy-4-methylcoumarin, **6i**) have been reported as not-covalent inhibitors of NF-κB-p50-DNA binding (IC_{50} < 500 μM). **1i** is the most active compound of the series considered with a IC_{50} = 30 μM. Unlike, ATA (**1n**), ATA analogues (e.g., Aurin, 2n; Chrome Azurol S, 3n; Eriochrome Cyanine R, 4n; Fluorescein, **5n**) and N,N' -[4-(dimethylamino)-2,6-pyridinylidenedimethyl] bis (S,S) (histidine)-di-methylester (**6n**) have been accounted as inactive (IC_{50} > 500 μM) ligands[94,95] (see Table 1).

NF-κB p50 binders (IC_{50} < 500 μM)



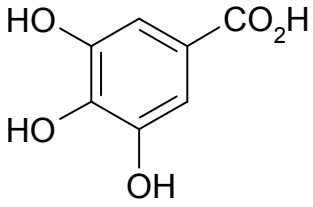
1i,2i

	X	Y	R1
1i	OH	Br	SO_3H
2i	OH	H	SO_3H

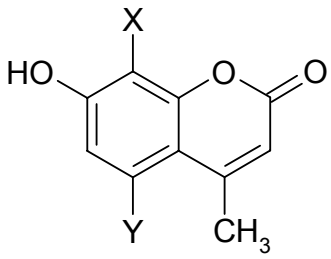


3i

	X	Y	R1	R2	R3	R4
3i	OH	H	SO_3H	H	H	H



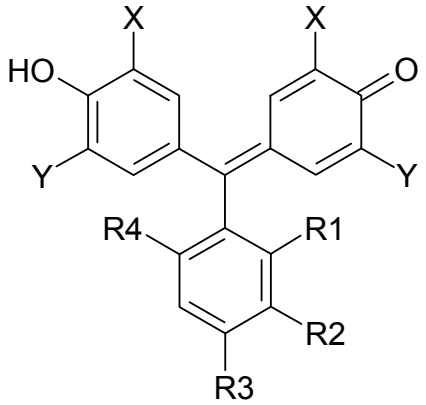
4i



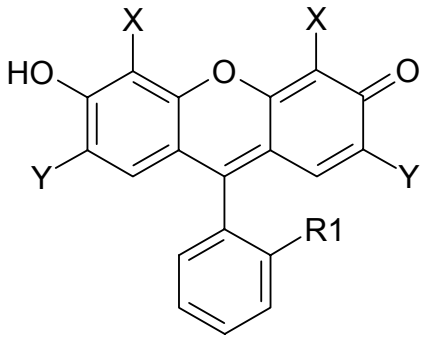
5i, 6i

	X	Y
5i	H	OH
6i	OH	H

NF-κB p50 binders (IC_{50} > 500 μM)



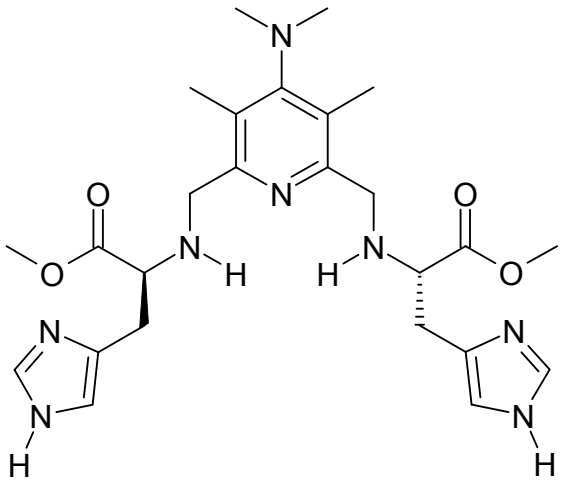
1-4n



5n

	X	Y	R1	R2	R3	R4
1n	COOH	H	H	COOH	OH	H
2n	H	H	H	H	OH	H
3n	COONa	CH3	Cl	SO_3Na	H	Cl
4n	COONa	CH3	SO_3Na	H	H	H

	X	Y	R1
5n	H	H	COOH



6n

Table 1. Chemical structures of known p50 binders

2.3.2 General structure features of NF-κB

The X-ray crystallographic structures of several Rel/NF-κB dimers on DNA (including p50-p50, p65-p65, p50-p65, c-Rel-c-Rel) have now been solved.[96-101] These proteins share high sequence identity over a 300 amino acid 'rel homolgy region' (RHR) in the amino terminus. The RHR is responsible for protein dimerization, DNA binding, and nuclear localization. The 3D structure of the highly variable C-terminal domains (not present in p50 and p52) that contain transactivation domains (Tds) have not been solved.

2.3.2.1 NF-κB p50 homodimer bound to a kB site

A crystal structure of 2.3 Å of the NF-κB p50 homodimer bound to a 10-base-pair idealized palindromic kB site (PDB id: 1NFK) has been solved by Ghosh et al.[97] The protein is derived from mouse and contain residues 39-364 including the entire RHR essential for DNA binding for each truncated p50 subunit (see Figure 2).

The homodimeric p50 binding site (5'-TGGGAATTCCC-3'; 3'-CCCTTAAGGGT-5') is derived from the consensus sequences of intronic enhancer element of kB light chain (Ig-kB), human β-interferon gene enhancer κB site (β-IFN-κB) and major histocompatability class 1κB element (H2-κB). The 3D structure of p50 subunit consists of two immunoglobulin-like domains: the carboxyl-terminal dimerization domain (domain 1: 39-240) containing the basic nuclear localization sequence (NLS), and the DNA-binding amino-terminal domain (NTD, domain 2: 248-350). These two domains are connected by a 10 residues flexible linker (238-247). The “butterfly” open conformation of the protein encloses the cylindrical body of DNA. The DNA-protein recognition surface involves the flexible linker and well defined 5 loops that connect the β-strand for each p50 subunit and buries 3.55 A^2 of solvent accessible water. In particular, DNA base-specific contacts are created by loops coming from the amino-terminal domain, while non specific DNA ribo-phospate backbone interactions involved loop amino acids from both the amino-terminal and dimerization domain (see figure 2).

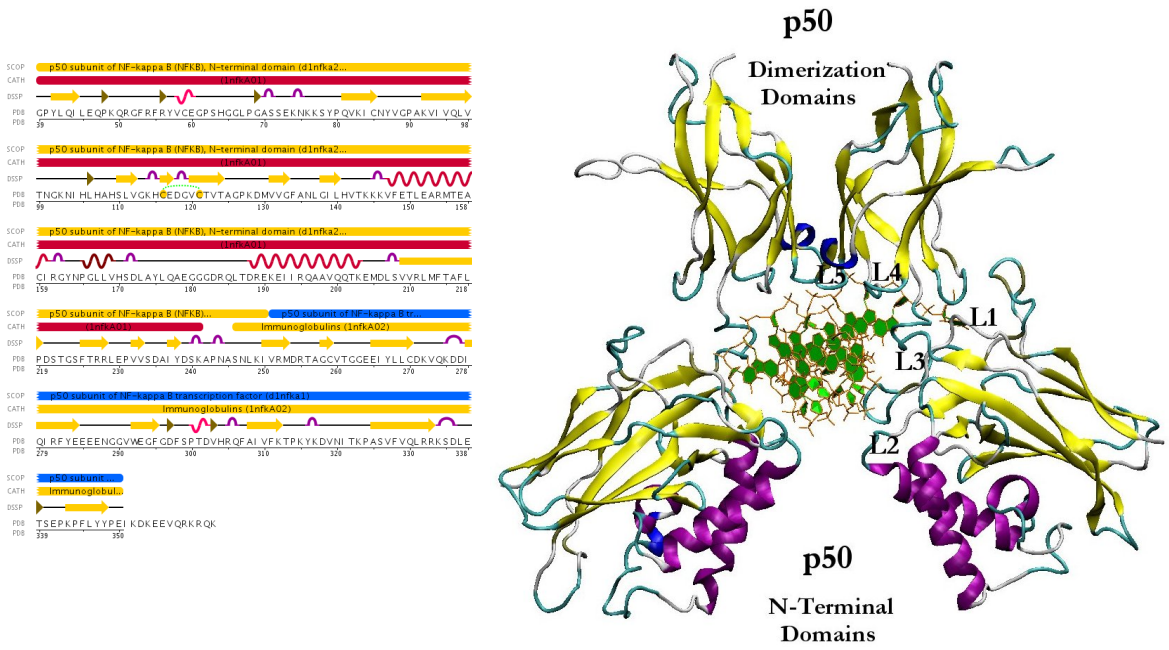


Figure 2. Structure of NF-kB p50-p50: (left) sequence of RHR of mouse p50 chain A. CATH and SCOP domain assignment, DSSP secondary structure (5 helices - 41 residues; 27 strands - 101 residues) are highlighted; (right) three dimensional structure of the protein bound to DNA. The loops (L1-L5) of p50 subunit that contact DNA are shown.

Most of the interacting residues highlighted in the figure above, have been confirmed by mutational analysis. Tyr57, Cys59 and Thr143, His64 may be involved in additional significant van der Walls contact in addition to hydrogen bonds to an extended DNA target. The contacts between the sugar-phosphate backbone and Cys59 account for the sensitivity of DNA binding to oxidation at this site.

The presence of Tyr57 in the central of the DNA/protein interface create from L1 explains its photochemical cross linking to DNA targets. Domain 1 includes 30 residues present only in p50 RHR sequence that access the DNA from the minor groove site. In particular the central phosphates form highly polarized strong hydrogen bonds with the sides chain of Lys144 and Lys145 and the NH of Lys144 backbone.

2.3.2.2 NF-κB p50-p65 heterodimer bound to a kB site

The overall topology of the NF-κB p50-p65 heterodimers is concordant with the structures of p50-p50 and other dimers crystallized on kB DNA targets. The 3 Å protein crystal structures

from mouse (PDB id: 1LE9)[96] are bound to the consensus Ig/HIV-2 kB sites 5'-TGGGACTTTCCT-3' and 5'-AAGGAAAGTCCC-3'.

As in the case of the homodimeric structure, the RHR of p50 (39-350) and p65 (19-271) subunits exist as two immunoglobulin-like domains in the active open conformation of the protein. p50 and p65 present some differences in the sequence of the DNA-binding loop and present different half-site specificity.

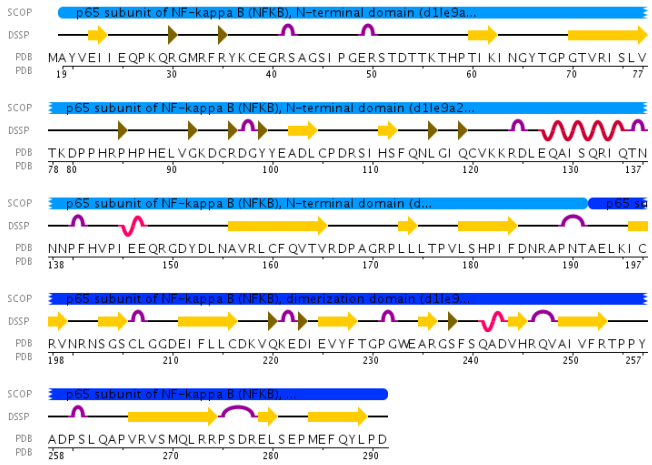


Figure 3. Sequence of RHR of mouse p65 chain A. SCOP domain assignment, DSSP secondary structure (3 helices - 15 residues; 29 strands - 90 residues) are highlighted.

In p65 subunit Arg33, Arg35, and Glu39 make base-specific DNA contact. The corresponding Arg-Arg-Glu triad in p50 plus Arg54 and Glu60 are also involved in base pairs interactions.

Considering the DNA backbone contacts, 11 p50 residues (Cys59, Lys144, Tyr57, Gln274, Gln306, and Arg305, Lys145, Lys272, Ser63, Gly65, Gly66, Asn136 and Lys77) and 12 p65 amino acids (Cys38, Lys122, Tyr36, Lys123, Gln220, Gln247, Arg246 and Lys221, Ala43, Ser42, Ser45 and Lys56) are involved.

2.3.2.3 NF-κB p50-p65 heterodimer bound to IκBα inhibitory protein

A 2.3 Å 3D crystal structure of IκBα in complex with the NF-κB heterodimer (PDB code: 1IKN) has been reported in literature.[102] The solved NF-κB structure contains the dimerization domains of p50 and p65 subunit and the amino-terminal domain of p65 subunit presenting the lack of p50 amino-terminal domain. Some experimental evidence, clearly

showed that p50 amino-terminal domain does not contribute to IκBα/NF-κB complex binding affinity. The IκBα contain a centrally located and cylindric shaped ankyrin repeat domain (ARD). Ankyrin repeats are 33 amino acid modules originally identified in the human erythrocyte protein ankyrin. The 3D structure of the complex suggests that the presence of the IkBalfa allow a strong conformational change of the DNA-bound open conformation of p50-p65 heterodimer with the adoption in the resting state of a IκBα-bound closed conformation. The p65 amino-terminal domain rotates almost 180 about its long axis and translates 38 Å towards its carboxy-terminal dimerization domain. Thus, besides interacting directly with the DNA-binding residues of NF-κB, IκBα permits allosteric regulation of the p65 subunit by locking it into a closed conformation. The relatively high temperature factor of NF-κB p65 subunit amino-terminal domain and the flexibility of IkBalfa carboxy-terminal sequence suggest that the p65 subunit conformation may be different in solution.

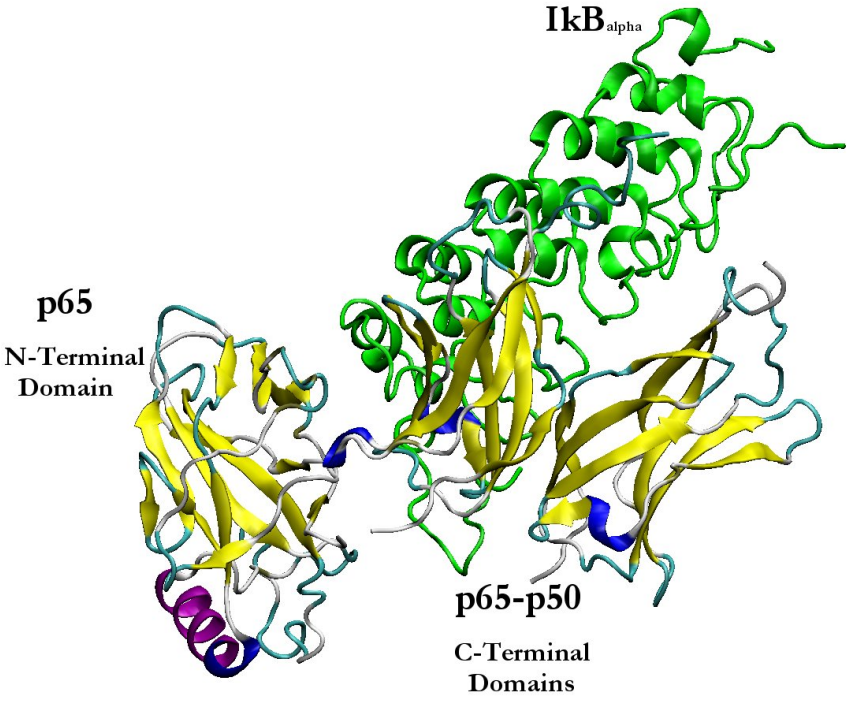


Figure 3. Three dimensional structure of NF-kB p65-p50 in complex with IκBα

References

1. Coupez B, Lewis R: Docking and Scoring-Theoretically Easy, Practically Impossible? Curr Med Chem **2006**, *13*, 2995.

2. Leach AR, Shoichet BK, Peishoff C: Docking and Scoring. J Med Chem **2008**, *49*, 5851.

3. Kontoyianni M, McClellan LM, Sokol GS: Evaluation of docking performance: comparative data on docking algorithms. J Med Chem **2004**, *47*,558.

4. Nosè S: A molecular dynamics method for simulations in the canonical ensemble. Mol Phys **1984**, *52*, 255.

5. Van Gunsteren WF and. Berendsen HJC: Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys **1977**, *34*, 1311.

6. Verlet L: Computers experiments on classical fluids. Thermodynamical properties of Lennard-Jones molecules. Phys. Rev **1967,** *159*, 98.

7. Ryckaert JP, Ciccotti G, Berendsen HJC: Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes. J Comp Phys **1977**, *23*, 327.

8. Andersen HC,). Rattle: a velocity version of the SHAKE algorithm for molecular dynamics calculations. J. Comput Phys **1983**, *52*, 24.

9. Andersen HC: Molecular dynamics simulations at constant pressure and/or temperature, J Chem Phys **1980**, *72*, 2384.

10. Beeman D: Some multistep methods for use in molecular dynamics calculations, J Comp Phys **1976**, *20* 130.

11. . Spoel Van der D, Lindahl E, Hess B, Buuren van AR, Apol E, Meulenhoff PJ,. Tieleman DP, Sijbers MALTK, Feenstra A, Drunen van R, and. Berendsen HJC. GROMACS user manual version 3.3. Department of Biophysical Chemistry, University of Groningen. Nijenborgh 4, 9747 AG Groningen, The Netherlands., 2006.

12. Amadei, Linssen AB, and Berendsen HJ: Essential dynamics of proteins. Proteins **1993**, *17*, 412.

13. Ghosh S, Nie A, An J, Huang Z: Structure-based virtual screening of chemical libraries for drug discovery. Current Opinion in Chemical Biology **2006**, *10*, 194.

14. Cavasotto CN, Kovacs JA, Abagyan RA: Representing receptor flexibility in ligand docking through relevant normal modes. J Am Chem Soc **2005**, *127*, 9632.

15. Shoichet B. Virtual screening of chemical libraries. Nature **2004**, *432*, 862.

16. Lyne PD: Structure-based virtual screening: an overview. Drug Discov Today **2002**, *7*, 1047.

17. Stahura FL, Bajorath J: Virtual screening methods that complement HTS. Comb Chem High Throughput Screen **2004**, *7*, 259.

18. Barril X, Hubbard RE, Morley SD: Virtual screening in structure-based drug discovery. Mini reviews in medicinal chemistry **2004**, *4*, 779.

19. Hou T, Xu X: Recent Development and Application of Virtual Screening in Drug Discovery: An Overview. Curr Pharm Des **2004**, *23*, 1011.

20. Kitchen D, Decornez H, Furr J, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews Drug Discovery **2004**, *3*, 935.

21. Bleicher K, Böhm H, Müller K, Alanine A: Hit and lead generation: beyond high-throughput screening. Nature Reviews Drug Discovery **2003**, *2*, 369.

22. Jain A: Virtual screening in lead discovery and optimization. Curr Opin Drug Discov Devel **2004**, *7*, 396.

23. RCSB protein data bank (PDB). (http://www.pdb.org).

24. Macro-molecular Structure Database at the European Bioinformatics Institute. http://www.ebi.ac.uk/msd

25. Protein Data Bank Japan. http://www.pdbj.org

26. Worldwide Protein Data Bank. http://www.wwpdb.org/

27. BioMagRes-Bank. http://www.bmrb.wisc.edu

28. Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, Stuart AC, Mirkovic N, Rossi A, Marti-Renom MA, Fiser A, Webb B, Greenblatt D, Huang CC, Ferrin TE, Sali A: MODBASE, a database of annotated comparative protein structure models, and associated resources. Nucleic Acids research **2004**, *32* (Database issue) D217-22.

29. Klebe G: Virtual ligand screening: strategies, perspectives and limitations. Drug Discov Today **2006**, *11*, 580.

30. Reyda S, Sohn C, Klebe G, Rall K, Ullmann D, Jakubke HD. Reconstructing the binding site of factor Xa in trypsin reveals ligand-induced structural plasticity. J of Mol Biol **2003**, 325, 963.

31. Totrov M, Abagyan R: Flexible ligand docking to multiple receptor conformations: a practical alternative. Current Opinion in Structural Biology Currenty **2008**, *18*, 178.

32. Barril X, Morley SD: Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. J Med Chem **2005**, *48*, 4432.

33. Wong CF, Kua J, Zhang Y, Straatsma TP, McCammon JA: Molecular docking of balanol to dynamics snapshots of protein kinase A. PROTEINS: Structure, Function, and Bioinformatics **2005,** *61*, 850.

34. Miller M: Chemical database techniques in drug discovery. Nature Reviews Drug Discovery **2002**, *1*, 220.

35. Beilstein. http://www.columbia.edu/cu/lweb/indiv/chemistry/index.htm

36. Chemical Abstracts Services. 2540 Olentangy River Rd Columbus, OH 43202, 614-447-3600 http://www.cas.org/expertise/cascontent/registry/index.html

37. Maybridge. http://www.maybridge.com/

38. Available Chemicals Directory. http://www.mdl.com/products/experiment/ available_chem_dir/index.jsp

39. MDL Drug Data Report. MDL Information Systems Inc., San Leandro, CA (USA). http:// www.mdl.com

40. World Drug Index. http://scientific.thomson.com/404/

41. National Cancer Institute. http://dtp.nci.nih.gov/

42. Optiverse .Tripos, Inc.1699 South Hanley Road, St. Louis, MO 63144-2319. www.tripos.com.

43. Compass Array . Arqule 19 Presidential Way Woburn, MA 01801. www.arqule.com

44. PHARMACophore . Chembridge corporation. 16981 Via Tazon, Suite G San Diego, CA 92127 www.chembridge.com

45. Irwin JJ, Shoichet BK: ZINC a free database of commercially available compounds for virtual screening. J. Chem. Inf. Model. **2005,** 45, 177. http://blaster.docking.org/zinc/

46. Lu A, Liu B, Liu H, Zhou J, Xie G: Traditional Chinese Medicinal Database. J Mol Des, **2004**, *3*, 672. http://tcm3d.com/services.htm

47. Natural Product Database. Chapman & Hall Chemical Database, Dictionary of Natural Porducts 6.2. Tripos Inc., **1998**.

48. Lei J, Zhou J: A marine natural product database. J Chem Inf Comput Sci **2002**, *42*, 742.

49. Fang X, Shao L, Zhang H, Wang S: CHMIS-C: a comprehensive herbal medicine information system for cancer. J Med Chem **2005**, *48*, 1481.

50. Rollinger J, Langer T, Stuppner H. Strategies for Efficient Lead Structure Discovery from Natural Products. Curr Med Chem **2006**, *13*, 1491.

51. Dunkel M, Fullbeck M, Neumann S, Preissner R: SuperNatural: a searchable database of available natural compounds Nucleic Acids research **2005**, *34*, 6.

52. Lipinski CA., Lombardo F, Dominy BW, and Feeney PJ: Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews **1996**, *23*, 3.

53. Schneider G, Böhm H: Virtual screening and fast automated docking methods. Drug Discov Today **2002**, 7, 1.

54. Dobson MC: Chemical space and biology. Nature **2004**, *432*, 5.

55. Cole JC, Murray CW, Willem J, Nissink M, Taylor RD, Taylor R: Comparing protein-ligand docking programs is difficult. Proteins **2005**, *60*, 325.

56. Kontoyianni M, McClellan LM, Sokol GS: Evaluation of docking performance: comparative data on docking algorithms. J Med Chem **2004**, *47*, 558.

57. Perola E, Walters WP, Charifson PS. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins **2004**, *56*, 235.

58. Friesner RA, Banks JL, Murphy RB, Halgren AT, Klicic JJ, Mainz TD, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS: Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem **2004**, *47*, 1739.

59. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL:. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem **2004**, *47*, 1750.

60. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT: Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem **2006**, *49*, 6177.

61. Sen R, Baltimore D: Inducibility of κ immunoglobulin enhancer-binding protein NF-κB by a post-translational mechanism. Cell **1986**, *47*, 921.

62. Aradhya S, Nelson DL: NF-kappaB signaling and human disease. Curr Opin Genet Dev **2001,** *11*, 300.

63. Kumar A, Takada Y, Boriek AM.; Aggarwal BB: Nuclear factor-kappaB: its role in health and disease. J Mol Med **2004,** *82*, 434.

64. Baldwin, A. S. Jr. Series introduction: the transcription factor NF-kappaB and human disease. *J. Clin. Invest*. **2001**, *107*, 3-6.

65. Voskaridouand E; Terpos E: New insights into the pathophysiology and management of osteoporosis in patients with beta thalassaemia. Brit. J. of Haematol **2004**, *127*, 127.

66. Morishita R, Sugimoto T, Aoki M, Kida I, Tomita N, Moriguchi A, Maeda K, Sawa Y, Kaneda Y, Higaki J, Ogihara T: In vivo transfection of cis element 'decoy' against nuclear factor-kappaB binding site prevents myocardial infarction. Nat. Med **1997**, *3*, 894.

67. Khaled AR, Butfiloski EJ, Sobel ES, Schiffenbauer: J. Functional consequences of the SHP-1 defect in motheaten viable mice: role of NF-kappa B. Cell. Immunol. **1998**, *185*, 49.

68. Hiscott J, Kwon H, Genin P: Hostile takeovers: viral appropriation of the NF-kappaB pathway. J Clin Invest **2001**, *107*:143.

69. Dickinson LA, Trauger JW, Baird EE, Dervan PB, Graves BJ, Gottesfeld JM: Inhibition of Ets-1 DNA binding and ternary complex formation between Ets-1, NF-kappaB, and DNA by a designed DNA-binding ligand. J Biol Chem **1999***, 274,* 12765.

70. Gilmore T, Gapuzan ME, Kalaitzidis D, Starczynowski D:. Rel/NF-kappa B/I kappa B signal transduction in the generation and treatment of human cancer. Cancer Lett **2002**, *181*,1-9.

71. Rottner M, Kunzelmann C, Mergey M, Freyssinet JM, Martínez MC: Exaggerated apoptosis and NF-kappaB activation in pancreatic and tracheal cystic fibrosis cells. FASEB J. **2007**, *21*, 2939.

72. Okamoto T: NF-kappaB and rheumatic diseases. Endocr. Metab. Immune Disord Drug Targets **2006**, *6*, 359.

73. Atreya I, Atreya R, Neurath MF: NF-kappaB in inflammatory bowel disease. J Intern Med **2008**, *263*, 591-596.

74. Lee CH, Jeon YT, Kim SH, Song YS: NF- kappaB as a potential molecular target for cancer therapy. Biofactors **2007**, *29*, 19.

75. Matthews JR, Wakasugi N, Virelizier JL, Yodoi J, Hay RT: Thioredoxin regulates the DNA binding activity of NF-kappa B by reduction of a disulphide bond involving cysteine 62. Nucleic Acids research **1992**, *20*, 3821-30.

76. Chandrasekaran V, Taylor EW: Molecular modeling of the oxidized form of nuclear factor-kappa B suggests a mechanism for redox regulation of DNA binding and transcriptional activation. J Mol Graph Model **2008**, *26*, 861.

77. Nishi T, Shimizu N, Hiramoto M, Sato I, Yamaguchi Y, Hasegawa M, Aizawa S, Tanaka H, Kataoka K, Watanabe H, Handa H: Spatial redox regulation of a critical cysteine residue of NF- kappa B in vivo. J. Biol. Chem. **2002**, *277*, 44548.

78. Gilmore TD: Introduction to NF-kappaB: players, pathways, perspectives. Oncogene **2006**, *25*, 6680.

79. Borgatti M, Finotti A, Romanelli A, Saviano M, Bianchi N, Lampronti I, Lambertini E, Penolazzi L, Nastruzzi C, Mischiati C, Piva R, Pedone C, Gambari R:. Peptide nucleic acids (PNA)-DNA chimeras targeting transcription factors as a tool to modify gene expression. Current drug targets **2004**, *5*, 735.

80. Gambari R: Biological activity and delivery of peptide nucleic acids (PNA)-DNA chimeras for transcription factor decoy (TFD) pharmacotherapy, Curr Med Chem **2004**, *11*, 1253.

81. Mischiati C, Borgatti M, Bianchi N, Rutigliano C, Tomassetti M, Feriotto G, Gambari R: Interaction of the human NF-kappaB p52 transcription factor with DNA-PNA hybrids mimicking the NF- kappaB binding sites of the human immunodeficiency virus type 1 promoter, J Biol Chem **1999**, *274*, 33114.

82. Mann MJ: Transcription factor decoys: a new model for disease intervention, Ann NY Acad Sci **2005**, *1058*, 128.

83. Mann MJ, Dzau VJ: Therapeutic applications of transcription factor decoy oligonucleotides. J Clin Invest **2000**, *106*, 1071.

84. Escotte S, Tabary O, Dusser D, Majer-Teboul C, Puchelle E, Jacquot J: Fluticasone reduces IL-6 and IL-8 production of cystic fibrosis bronchial epithelial cells via IKK-beta kinase pathway. Eur Respir J. 2003, *21,* 574.

85. Bezzerri V, Borgatti M, Nicolis E, Lampronti I, Dechecchi MC, Mancini I, Rizzotti P, Gambari R, Cabrini G. Transcription factor oligodeoxynucleotides to NF-kappaB inhibit transcription of IL-8 in bronchial cells. Am J Respir Cell Mol Biol **2008**, *39*, 86.

86. Xia YF, Ye BQ, Li YD, Wang JG, He XJ, Lin X, Yao X, Ma D, Slungaard A, Hebbel R, Key NS, Geng JG: Andrographolide Attenuates Inflammation by Inhibition of NF-κB Activation through Covalent Modification of Reduced Cysteine 62 of p50 J. Immunol **2004,** *173*, 4207.

87. Wang YJ, Wang JT, Fan QX, Geng JG: Andrographolide inhibits NF-kappaBeta activation and attenuates neointimal hyperplasia in arterial restenosis. Cell. Res **2007,** *17*, 933.

88. Chen KM, Spratt TE, Stanley BA, De Cotiis DA, Bewley MC, Flanagan JM, Desai D, Das A, Fiala ES, Amin S, El-Bayoumy K: Inhibition of nuclear factor-kappaB DNA binding by organoselenocyanates through covalent modification of the p50 subunit. Cancer Res **2007**, *67*, 10475.

89. Park HJ, Lee SH, Son DJ, Oh KW, Kim KH, Song HS, Kim GJ, Oh GT, Yoon DY, Hong JT:. Antiarthritic effect of bee venom: inhibition of inflammation mediator generation by suppression of NF-kappaB through interaction with the p50 subunit. Arthritis Rheum **2004,** *50*, 3504.

90. Kim, H. M.; Bae, S. J.; Kim, D. W.; Kim, B. K.; Lee, S. B.; Lee, U. S.; Kim, C. H.; Moon, S. K. Inhibitory role of magnolol on proliferative capacity and matrix metalloproteinase-9 expression in TNF-alpha-induced vascular smooth muscle cells. *Int. Immunopharmacol.* **2007,** *7*, 1083-1091.

91. Lambert C, Li J, Jonscher K, Yang TC, Reigan P, Quintana M, Harvey J, Freed BM. Acrolein inhibits cytokine gene expression by alkylating cysteine and arginine residues in the NF-kappaB1 DNA binding domain. J Biol Chem **2007**, *282*, 19666.

92. Park MH, Song HS, Kim KH, Son DJ, Lee SH, Yoon DY, Kim Y, Park IY, Song S, Hwang BY, Jung JK, Hong JT Cobrotoxin inhibits NF-kappa B activation and target gene expression through reaction with NF-kappa B signal molecules. Biochemistry **2005**, 44, 8326.

93. Lin R, Gewert D, Hiscott J: Differential transcriptional activation in vitro by NF-kappa B/Rel proteins. J Biol Chem 1995, 270, 3123.

94. Sharma RK, Pande V, Ramos M.J, Rajor, H.K.; Chopra, S.; Meguro, K.; Inoue, J.; Otsuka, M. Inhibitory activity of polyhydroxycarboxylate chelators against recombinant NF-kappaB p50 protein-DNA binding. *Bioorg. Chem.* **2005**, *33*, 67-81.

95. Sharma, R. K. Chopra, S.; Sharma, S. D.; Pande, V.; Ramos, M. J.; Meguro, K.; Inoue, J.; Otsuka, M. Biological evaluation, chelation, and molecular modeling studies of novel metal-chelating inhibitors of NF-kappaB-DNA binding: structure activity relationships. *J. Med. Chem.* **2006**, *49*, 3595-3601.

96. Muller, C. W.; Rey, F. A.; Sodeoka, M.; Verdine, G. L.; Harrison, S. C. Structure of the NF-kappa B p50 homodimer bound to DNA. *Nature* **1995**, *373*, 311-317.

97. Moorthy AK, Huang DB, Wang VY, Vu D, Ghosh G. X-ray structure of a NF-kappaB p50/RelB/DNA complex reveals assembly of multiple dimers on tandem kappaB sites. J Mol Biol **2007**, *26*; 373.

98. Chen FE, Huang DB, Chen YQ, Ghosh G: Crystal structure of p50/p65 heterodimer of transcription factor NF- kappaB bound to DNA. *Nature* **1998**, *391*, 410.

99. Chen YQ, Ghosh S, Ghosh G: A novel DNA recognition mode by the NF-kappa B p65 homodimer. Nat Struct Biol **1998**, 5, 67.

100. Chen YQ, Ghosh S, Ghosh GA: Novel DNA recognition mode by the NF-kappa B p65 homodimer. Nat Struct. Biol **1998**, *5*, 67.

101. Huang DB, Chen YQ, Ruetsche Mv, Phelps CB, Ghosh G: X-ray crystal structure of proto-oncogene product c-Rel bound to the CD28 response element of IL-2. Structure **2001**, *9*, 669.

102. Huxford T, Huang DB, Malek S, Ghosh G: The crystal structure of the IkappaBalpha/NF-kappaB complex reveals mechanisms of NF-kappaB inactivation. Cell **1998**, *95*, 759.

CHAPTER 3. Docking of natural compounds into NF-κB p50 targets

3.1 Introduction

The recent approved natural-product-related drugs have been delineated extensively in earlier reviews[1-5] and as pointed out by Harvey[1], over 100 natural-product-derived compounds are currently undergoing clinical trials. They have inspired, at various levels, the fashioning of non natural agents of pharmaceutical import.[6-13] The studies on natural products are preponderantly being focused for use in cancer or anti-infective, but many other therapeutic research areas are comprised. A crucial factor in understanding a major probability of successful identification of interesting new lead structures from natural products in respect to synthetic compounds is that they contain a broader range of structural diversity and a large number of chiral centres. Even though combinatorial synthesis is now producing molecules that are drug-like in terms of size and property, these molecules, in contrast to the small chemical substances produced by different organisms, have not evolved to interact with biomolecules. However the applications of VS for the discovery of natural lead candidates are still much less representative in comparison to synthetic compounds. In fact, nowadays, to enable in *silico* screenings applications, large database of small synthesised molecules are available from numerous suppliers. It is not the same for database of the more chemically complex natural compounds. Other impediment, in particular for pharmaceutical companies, should concern the intellectual property in potential successful discover of interesting molecules. Despite these last considerations, the interest for natural products in drug discovery appear to be recently increasing and a comprehensive overview of VS applications has been reviewed.[14-16]

Nevertheless, no application of the widely-used molecular docking computational method has been published for the VS of chemical libraries of natural and/or synthetic compounds against

NF-κB. Indeed, the individuation of potent inhibitors of its interaction with DNA trough in "silico" virtual screening is not feasible due to the nature of the protein target under study. NF-κB as a target for docking simulations presents several potential limitations: (i) no three dimensional structure of NF-κB/small ligand complexes is available, (ii) DNA-binding site has a large contact surface area and is accessible to water, (iii) some side chains of the interaction area assumed different rotameric state depending on DNA conformation, (iv) the DNA binding site includes a flexible 10 residues' linker and (v) most ligands show rather small binding constant (micromolar range) (vi) NF- κB proteins can exist in both monomeric and dimeric states.

Recently, some molecular modelling studies have predicted possible binding mode of the inhibitors molecules to the DNA binding region of subunit p50, starting from the crystallographic structure of the NF-κB homodimer.[18-21] In particular, Sharma et al.[20] in an effort to rationalize the previously results obtained from EMSA studies on a set of aurintricarboxylic acid analogues, employed docking studies to explain the structure activity relationships observed within this class (see chapter 2). Despite the reported docking results, no studies were performed to test the predictive power of this computational method in findings new inhibitors for NF-κB.

Docking simulations to the DNA-binding site of NF-κB p50 are suitable for the identification of low molecular weight compounds interacting with this target? With this previous work, we intend to answer to this question.

In the present chapter we report docking studies on a dataset of 27 molecules from extracts of two different medicinal plants prepared in our laboratory to NF-κB -p50 with the purpose of developing a docking protocol fit for the target under study and focused on identification of natural lead compounds. After evaluation through electrophoretic mobility shift assays (EMSA) and biological effort, we obtained a fairly good agreement between experimental data and molecular modelling identification of bioactive and inactive compounds. All experimental assays have been accomplished in BioPharmaNet, Department of Biochemistry and Molecular Biology (University of Ferrara) and in the Laboratory of Molecular Pathology, Laboratory of Clinical Chemistry and Haematology (University-Hospital, Verona).[17]

3.2 Methods

3.2.1 Ligands data and preparation.

The database of 27 natural structures used in our molecular docking studies, were derived from different medicinal plant extracts (Figure 1) as prepared in our laboratory. A dataset of 12 active compounds used as references molecules were collected from four publications[18- 21] reported by one laboratory (see Table 1). Ten of these inhibitors (**1i-8i**, **11i** and **12i**) were employed in starting docking studies (protocol 1) and in the Standard Similarity Scoring for subsequently docking simulations. Two inhibitory molecules (**9i** and **10i**) were used as test set in all docking simulations. The three-dimensional models of all the molecules under investigation were built by assembling fragments from the SYBYL 7.0 software package standard library.[22] Resulting geometries were optimized and molecular charges were assigned by a semi empirical molecular orbital calculations using the AM1 Hamiltonian[23] (module MOPAC implemented in SYBYL).

3.2.2 Proteins data and preparation.

The three dimensional structure of the complex NF-kB-DNA was retrieved from the Protein Data Bank (PDB code: 1NFK). The cocrystallized DNA macromolecule was removed from the structure. p50 dimer and p50 monomers (chains A and B) were selected for the docking simulations and prepared using the graphical interface Maestro.[24] All water molecules were removed, the hydrogen atoms were added to the proteins and all atom force field (OPSL-2001) charges and atom types were assigned. Preparation and refinement were done running ProteinPrep job on the structure in a standard procedure. Minimizations were performed until the average root mean square deviation of non-hydrogen atoms reached 0.3 Å.

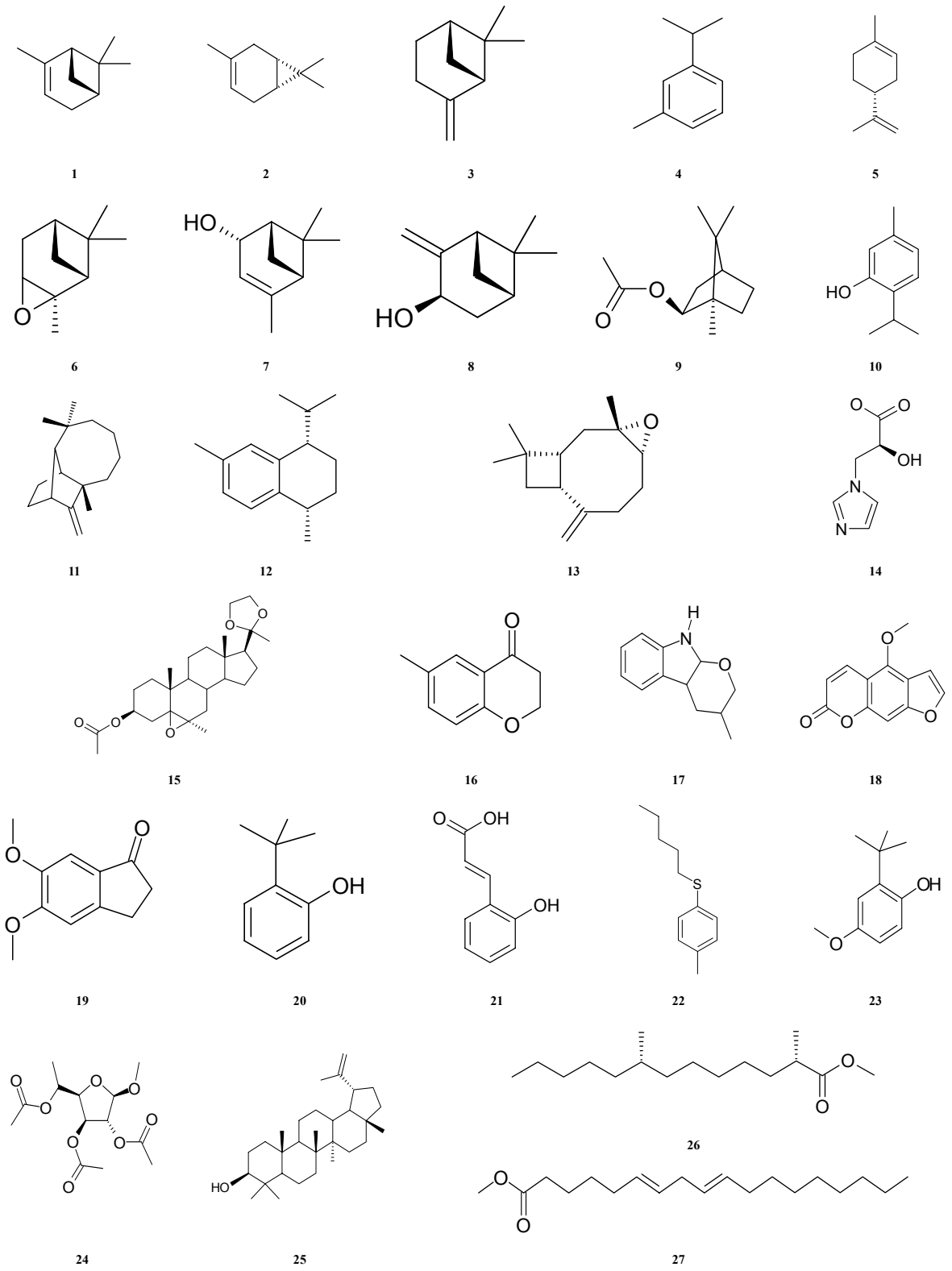
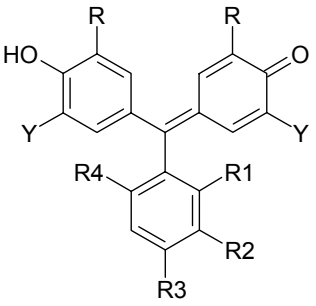
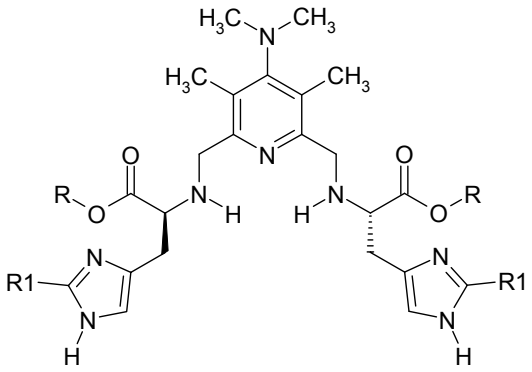


Figure 1. Structures of compounds found in Cupressus pyramidalis and Aegle marmelos extracts and used for docking simulations.



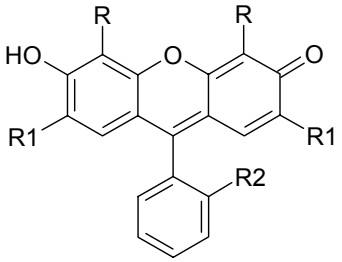
1i-3i

	R	R_1	R_2	R_3	R_4
1i	OH	H	SO_3H	H	H
2i	COONa	CH_3	Cl	SO_3Na	Cl
3i	COONa	CH_3	SO_3Na	H	H



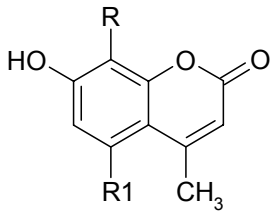
4i-6i

	R	R_1
4i	CH_3	H
5i	CH_3	$C(Ph)_3$
6i	H	$C(Ph)_3$



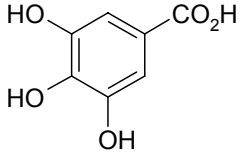
7i,**8i**

	R	R_1	R_2
7i	OH	H	SO_3H
8i	OH	Br	SO_3H

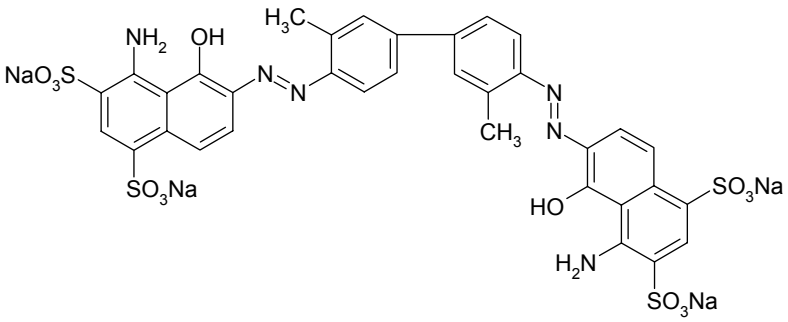


9i,**10i**

	R	R_1
9i	H	OH
10i	OH	H



11i



12i

Table 1. Chemical structure of references molecules for docking applications.

3.2.3 Docking simulations

All molecules of plant extracts (**1-27**) and the known inhibitors (**1i-12i**) under study were docked in to the binding site of the receptor (PDB ID: 1NFK) using Glide (Grid-Based Ligand Docking With Energetics) software from Schrodinger.[25,26] Grids were prepared for each proteins with the exact same center and the size of the bounding box set on 30 Å. The coordinates of the enclosing box (x=-1,1958 Å; y=9.0149 Å; z=19,7598 Å) were defined starting from the set of active site residues involved in hydrogen bonds with the NF-κB recognition site of DNA (Arg54, Arg56, Tyr57, Cys59, Lys241, Gln306 and Thr143) and optimised including the double strands DNA helices volume by visual inspection. The Glide algorithm is based on a systematic search of positions, orientations, and conformations of the ligand in the receptor binding site using funnel type approach. The search begins with a rough positioning and scoring phase that significantly limits the search space and reduces the number of poses to be selected for minimization on the precomputed OPLS-2001 van der Waals and electrostatic grids for the protein. The 5-10 lowest-energy poses obtained from this stage are subjected to Monte Carlo simulations and the minimized poses accepted are then rescored using the GlideScore function, which is a more sophisticated version of ChemScore.[27] This force field include additional terms accounting for solvation and repulsive interactions. In order to provide a better correlation between good poses and good scores, Glide Extra-Precision (XP) Mode was subsequently used on the conformations selected from Glide Standard Precision (SP) mode. The atom-pair superimposition of p50 chains A and B, prepared as described above, gave a minimum RMSD of 2,303 Å (heavy atoms). Considering the clear dependence of the docking accuracy of ligands on the protein structure, docking simulations were carried out with the same protocol on both A and B, considered as two slightly different conformations of the same structure.

Unfortunately, complexes of NF-κB cocrystallized with inhibitors has not been solved. Therefore, a common self-docking procedure to evaluate the accuracy of the docking protocol adopted was not practicable. In order to overcame this situation, two structurally similar active compounds (**9i** and **10i**) were used as test set and docked into the DNA binding site of the protein. Moreover in following docking jobs, atom pair similarity (AP) scoring (Similscore) facility as implemented in Glide, was incorporated in GlideScore (G-score),

based on the assumption that closely related chemical structure should share similar biological activity and physiochemical property. Similscore can have a value between 0 and 1 as implemented in Glide. The adjusting G-score value is illustrated here below:

- if $0.0 \leq$ SimilScore$< 0.3 \rightarrow$ G-score =G-score+6.0
- if $0.3 \leq$ SimilScore$< 0.7 \rightarrow$ G-score =G-score+(0.7-Similscore)/(0.7-0.3)*6.0
- if $0.7 \leq$ SimilScore$< 1.0 \rightarrow$ G-score =G-score+0.0

All inhibitors molecules, except for **9i** and **10i**, were used just as reference structures for AP similarity method.
Based on the best final GlideScore ranking, the similarity docking procedure for subsequently docking simulations on p50 subunits was chosen.

3.3 Results and discussion

3.3.1 Docking analysis

The docking results for all the reference inhibitory compounds and the natural compounds under study are reported in Table 2 (references compounds) and in Table 3 (natural compounds). As shown in the 2.3 Å crystal structure, the DNA/p50 complex is formed by one DNA molecule and two p50 proteins each one consisting of two distinct domains connected by a10-residue linker. Both domains and the segments that connect them, form a sequence-specific DNA-binding surface. All the five loops per subunit fill the entire major groove of the DNA. The specific interactions that stabilized the NF-κB /DNA complex, occur over 10-bp forming the kB recognition site. Unlike many dimeric protein-DNA complexes, many residues of both subunits make specific base contacts in a non-contiguous cooperative network. The plasticity of centre region of the interface carry to the lack of symmetry exhibited by the interactions of Lys 241 from the linker segment, and Lys 272 and Arg 305 from the dimerization region with the symmetrical target site. In the subsequent experimental EMSA studies, a recombinant p50 protein that probably forms a monomer-dimer mixture in binding buffer solution will be used. On the base of structural and experimental assumptions

as above mentioned, p50 dimer and monomer were employed as protein target in our molecular modelling investigation.

In order to evaluate the impact of the introduction of the similarity penalty in the docking algorithm on the results, the positions of **9i** and **10i** used as test set in the final GlideScore ranking were compared in the two different procedures (Table 3).

Docking protocol 1					
p50-p50		***p50 (A)***		***p50 (B)***	
Cpd.	**G-Score**	**Cpd.**	**G-Score**	**Cpd**	**G-Score**
8i	-5.88	**8i**	-6.06	**1i**	-6.08
7i	-5.88	**7i**	-5.83	**3i**	-5.35
2i	-5.84	**11i**	-5.20	**7i**	-5.06
3i	-5.78	**1i**	-5.07	**2i**	-4.58
11i	-3.87	**2i**	-2.79	**11i**	-3.59
1i	-3.57	**3i**	-2.77	**4i**	-1.82
4i	-0.76	**4i**	>0	**8i**	>0
1i	>0	**1i**	>0	**1i**	>0
5i	-	**5i**	>0	**5i**	-
6i	-	**6i**	>0	**6i**	-

Table 2. Ranking of the poses of references inhibitory molecules (**1i-8i** and **11i-12i**) in the target NF-kappaB p50 both as dimer (p50-p50) and as monomers (p50 A and B).

Docking protocol 1		Docking protocol 2					
p50-p50		*p50-p50*		*p50 (A)*		*p50 (B)*	
Cpd.	**G-score**	**Cpd.**	**G-score**	**Cpd.**	**G-Score**	**Cpd.**	**G-Score**
10i	*-5.61*	***21***	*-2.34*	***9i***	*-3.81*	***9i***	*-2.13*
15	-5.50	***9i***	*-2.29*	***10i***	*-2.27*	***10i***	*-1.43*
25	-5.24	***10i***	*-2.19*	***21***	*-1.40*	***21***	*-0.32*
23	-4.94	**18**	-0.50	**18**	-0.07	**18**	>0
21	*-4.91*	**23**	-0.05	**10**	>0	**15**	>0
24	-4.74	**20**	>0	**23**	>0	**23**	>0
9i	*-4.44*	**15**	>0	**7**	>0	**24**	>0
10	-4.40	**25**	>0	**8**	>0	**10**	>0
20	-4.37	**24**	>0	**20**	>0	**23**	>0
7	-4.28	**10**	>0	**15**	>0	**20**	>0
22	-4.13	**16**	>0	**25**	>0	**16**	>0
16	-3.48	**7**	>0	**24**	>0	**7**	>0
19	-3.71	**22**	>0	**13**	>0	**8**	>0
2	-3.39	**19**	>0	**12**	>0	**2**	>0
27	-3.09	**4**	>0	**16**	>0	**22**	>0
5	-3.05	**2**	>0	**6**	>0	**4**	>0
26	-2.89	**5**	>0	**9**	>0	**1**	>0
18	-2.69	**27**	>0	**11**	>0	**19**	>0
14	-2.47	**26**	>0	**2**	>0	**14**	>0
4	-2.00	**14**	>0	**3**	>0	**5**	>0

Table 3. Ranking of the poses of natural compounds and test set inhibitors (**9i** and **10i**) in the target NF-kappaB p50 both as dimer (p50-p50) and as monomers (p50 A and B). In the docking protocol 1 the similarity scoring algorithm is not used

Known active compound **10i** was ranked at the top positions in both procedures, but only the introduction of the similarity parameter in the scoring function significantly increased the efficiency in **9i** ranking (Table 3). In fact the difference in glide-score among these two inhibitors with a similar inhibitory potency (500 μM)[21] was very small (ΔGlideScore=0.10). For each selected ligand the pose with best E-Model score (combination of energy grid score, GlideScore, and the internal strain of the ligand) was used for in-deph interaction analysis. Compound **21** clearly showed highest score in respect to docked plant extracts (Table 3) outranking the known inhibitors at physiological pH in docking simulation to the dimer.

Docked compounds **1-27, 9i** and **10i**, occupied a region of the binding surface creates by the spatial relationship between the N-terminal domain of p50 subunit and the 10 residues long linker loop (Figure 2). Molecules **21**, **9i** and **10i** (Figure 3) were located in a small cleft surrounded by several polar amino acids (i.e. Tyr57, His109, His141, Tyr143 Lys144, Lys145, Ser208, Asp239, Lys241 and Ser208) and the highest score poses were superimposable with minimum RMSD of 1.36 Å for compounds **9i** and **21**.

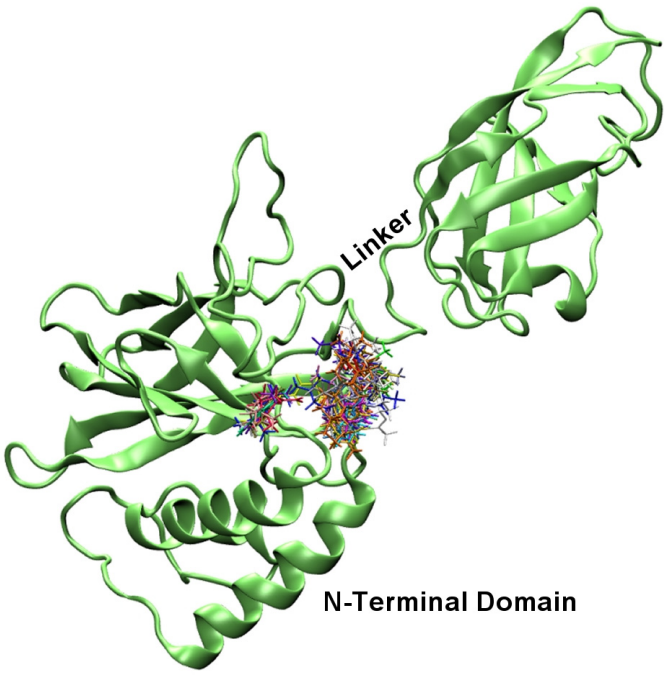


Figure 2 Stereoview of compounds **1-27**, **9i** and **10i** docked in to DNA binding region of the NF-kappaB p50 monomer chain A. The macromolecule is highlighted in green.

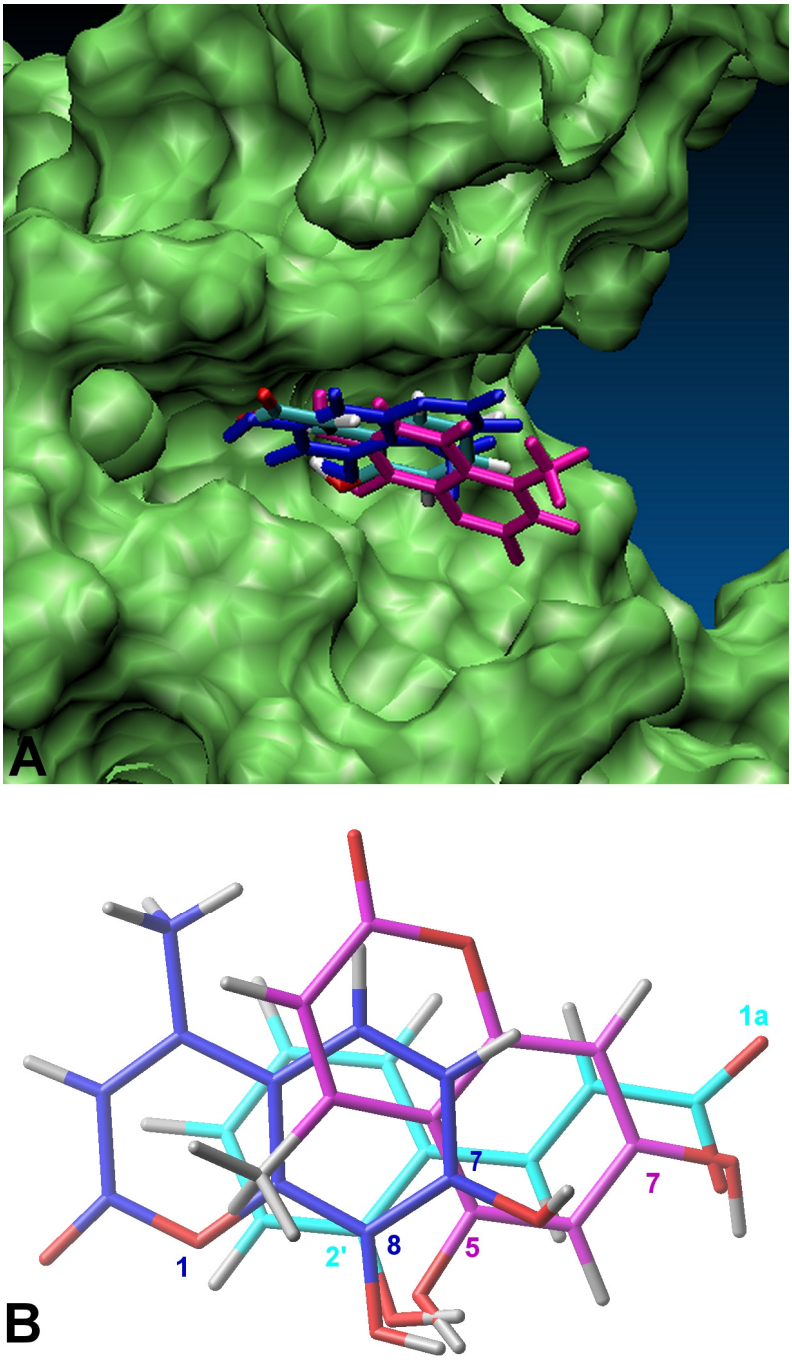


Figure 3. Superimposition of the docked poses of inhibitors **9i**, **10i** and compound **21**. A. the DNA binding site of NF-kappaB p50 (monomer chain A) is highlighted in green; B. the ligand atoms involved in hydrogen bonding are labeled. Compounds **9i**, (shown in purple), **10i** (shown in blue) and **21** are illustrated in stick representations.

The RMSD was calculated by superimposing the following atoms pairs: heteroatoms involved in hydrogen bonding with the same residues of the protein (**9i**.O8 and **21**.O2'; **9i**.O7 and **21**.O1a) (Figure 3B) and the centroid of aromatic system of coumarin structure with the centroid of benzene ring of **21.** These compounds showed slightly different binding modes in p50 (chain A), p50 (chain B) and p50-p50 targets. Here we reported the highest score poses obtained from docking protocol including the similarity function. H-bond interactions between OH groups of coumarin structures (OH of benzene ring in **21**) and both NH of His141 and the carboxylic group of Asp239 showed to be important for ligands binding. Moreover, OH groups of coumarin moiety (carboxylate group in **21**) made an additional hydrogen bond with CO of the backbone and protonated NH3 group of Lys241 (Table 4).

Residue interaction	Ligand atom	Distance (Å)
(His141) -Nε-H::O	O5 **(9i)**	2.12
	O8 **(10i)**	1.88
	O2' **(21)**	1.90
(Asp239) -COO::H	H5 **(9i)**	1.91
	H8 **(10i)**	2.09
	H2' **(21)**	2.00
(Lys241) -CO::H	H7 **(10i)**	1.96
(Lys241) -N^{+}-H::O	O7 **(9i)**	2.60
	O1a **(21)**	2.17
(Leu207) -N-H::O	O1 **(10i)**	2.18

Table 4: intramolecular hydrogen bonds of the docked poses of **9i**, **10i** and **21** with the involved residues of the DNA binding region of NF-kappaB (see Figure 3 for ligand atom labels). The interatomic distances in Angstroms are shown.

It is important to note that Lys241 could be involved in the stability of the DNA-binding conformation of the protein. In fact, as discussed above, this residue is situated in the flexible linker segment and interacts with Lys 272 and Arg 305 from the dimerization domain. Finally, carbonyl group of **10i** engage another H-bond with NH of the backbone of the Leu207. Compound **21** showed the same binding mode of active ligands in the monomer configuration of the target, with the only difference of a stronger interaction of carboxylate group with Lys241 (Table 4 and Figures 4A, 5A).

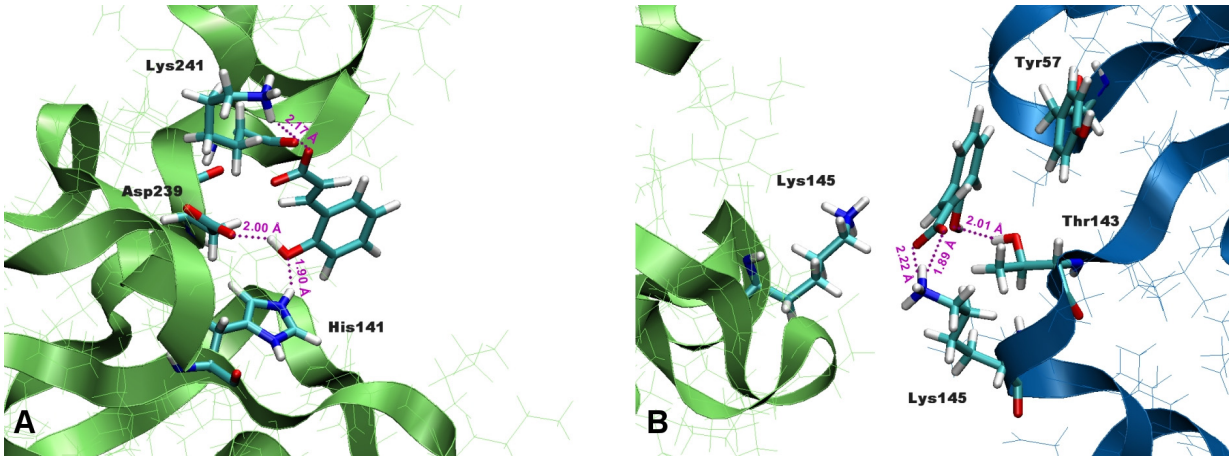


Figure 4. Binding modes of **21** docked in to the active site of NF-kappaB p50: A. monomer and B. homodimer (chain A, shown in green). The residues involved in the interaction with the ligand are shown; the hydrogen bonding and the relative distances are indicated in purple

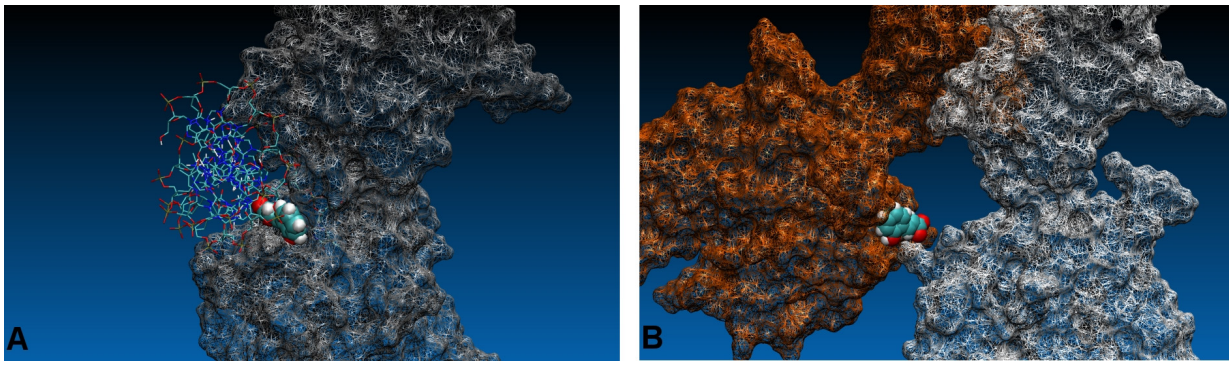


Figure 5. Poses of docked compound **21** in to the DNA binding region of NF-kappaB p50. A. monomer and B. homodimer (chain A, shown in grey). The inhibitory activity of **21** may be due to its ability to form a stable complex with the active conformation of the dimer and/or blocking the interaction of DNA with the monomer filling the protein binding site. The DNA was obtained from the crystal structure of the homodimer NF-kappaB (pdb code:1NFK). The surface of the protein is represented in wire frame, the ligand and DNA are highlighted in VDW and stick representation respectively.

Interestingly, the best pose of compound **21** occupied a region formed by residues of both p50 units (chain A and chain B) of NF-κB dimer: Lys 145 and Thr143 of chain A and Tyr57, Lys144, Lys145, Glu60, Cys59, Thr143, Lys146 of chain B. In particular, the OH group of the ligand engages a hydrogen bond with the sidechain of Thr143 (chain B), and the carboxylate group forms a salt bridge stabilized by two hydrogen bonds with the side chain of Lys 145 (chain B). Moreover the phenyl structure of compound **21** could be involved in a weak π-π stacking interaction with the aromatic moiety of Tyr57 (chain B) (centroid-centroid distance: 4.93 Å), a residue specific for kB DNA sequence 5'-GGGATTTCC-3', present in different cellular genes including HIV-LTR. Of course, further dynamics simulation on the protein-ligand complex should be necessary to validate this hypothesis. In addition, the amino group of Lys145 of the opposite p50 unit (chain A) could form an additional π-cation interaction with the aromatic group of **21** (N^+-H and centroid of benzene ring distance: 3.87 Å) (Figures 4B, 5B). These bridge structures are likely to reinforce the anchoring of this molecule to the DNA binding region of the dimer, and might account for the slight better G-score of **21** in respect to the monomer configuration of the receptor. Moreover, all the residues of the protein involved in molecular interactions with molecule **21** form hydrogen bonds also with DNA.

All compounds with higher GlideScore and E-Model score clearly showed the ability to make a maximum number of hydrogen bonding, according with the result as previously reported on a flexible docking studies of known inhibitors **9i** and **10i** (see chapter 2, table XX) even if reported residues involved in binding interaction were different. The highest ranking poses of **21**, **9i** and **10i** formed 3-4 hydrogen bonding with the target protein, whereas molecules in medium positions in docking ranking not more than 2. According, structures not involved in hydrogen bonding were ranked in the last positions (Table 3). In particular, compound **5** with a GlideScore < 0 in similarity protocol lost the ability both to occupy the same positions of active ligands and to form hydrogen bonding with the protein (not shown). In house experimental data were in good agreement with the molecular modelling findings. In accordance with docking results, **21** and **5** showed to be active and inactive respectively in further EMSA experimental studies.

3.3.2 Effects of the highest ranked compound

The effects of compound **21** on NF-κB interactions were first studied by EMSA as described elsewhere.[28-31] It is indeed well accepted that molecules binding NF-κB might retain inhibitory activity on molecular interaction between NF-κB and DNA.[31]

The results of the gel retardation analysis clearly demonstrate that compound **21** inhibit the molecular interactions between nuclear factors (Figure 6B) or isolated NF-κB p50 target double stranded oligonucleotide mimicking the NF-κB binding sites. This effect was similar to that exhibited by the reference compound **9i**. Interestingly, compound **5** and extracts from *C. pyramidalis* were found to be inactive (Figure 6A), fully in agreement with the docking data summarized in Table 3. Since NF-κB is one of the most important transcription factors regulating the expression of IL-8 gene[31] and the in silico results demonstrate that compound **21** inhibit NF-κB /DNA interactions, the activity of this compound in inhibiting the expression of IL-8 gene in IB3-1 cells infected with *Pseudomoas aeruginosa* has been tested. The biological results obtained demonstrate that compound **21** is a strong inhibitor of PAO-1 induced accumulation of IL-8 mRNA.

3.4 Conclusions

In the present work, we carried out docking studies on the dataset of 27 molecules found in different plant extracts to NF-κB -p50, with the purpose of developing a docking protocol fit for the target under study, eventually applicable for more time-consuming virtual screening of larger database of compounds. Usually, docking to protein structures that do not have a ligand present, as in the case of NF-κB , dramatically reduces the expected performance of structure-based methods. Therefore, the use of NF-κB as a target for virtual docking of natural compounds is not feasible. To overcome such a limitation, we proposed to enhance the simple docking procedure by means of a sort of combined target- and ligand-based drug design approach. Advantages of this combination strategy, based on a similarity parameter for the identification of weak binding chemical entities, are illustrated in this work with the discovery of a new lead compound for NF-κB. In this respect, this paper represents the first example of successfully individuation of a potential lead compound through molecular docking

simulations on a NF-κB target. Our results are of interest also from the practical point of view. At the same time, information derived from this structure and its different binding modes, could carry through further lead optimization to more potent NF-κB inhibitors. The transcription factor NF-κB is indeed a very interesting target macromolecule in the design on anti-tumor, anti-inflammatory, pro-apoptotic drugs.

References

1. Harvey AL: Natural products in drug discovery. Drug Discov Today **2008**, *13*, 894.

2. Newman DJ, Cragg GM: Natural products as sources of new drugs over the last 25 years. J Nat Prod **2007**, *70*, 461.

3. Butler MS: Natural products to drugs: natural product-derived compounds in clinical trials. Nat Prod Rep **2008**, *25*, 475.

4. Chin YW, Balunas M, Chai H, Kinghorn A: Drug discovery from natural sources. The AAPS J **2006**, *8*, E239–E253.

5. Lam KS: New aspects of natural products in drug discovery. Trends Microbiol **2007**, *15*, 279.

6. Balunas MJ, Kinghorn AD: Drug discovery from medicinal plants. Life Science **2005**, *78*, 431.

7. Clardy J, Walsh C: Lessons from natural molecules. Nature. **2004**, *432*, 829.

8. Cragg GM, Newman DJ, Snader KM: Natural Products in Drug Discovery and Development. J Nat Prod **1997**, *60*, 52.

9. Vuorelaa P, Leinonenb M, Saikkuc P, Tammelaa P, Rauhad JP, Wennberge T, Vuorela H: Natural products in the process of finding new drug candidates. Curr Med Chem **2004**, *11*, 1375.

10. Feher M, Schmidt JM: Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. J Chem Inf Comput Sci **2003**, *43*, 218.

11. Newman DJ, Cragg GM: Advanced preclinical and Clinical trials of natural products and related compounds from marine sources. Curr Med Chem **2004**, *11*, 1693.

12. Khan MTH, Ather A: Potentials of phenolic molecules of natural origin and their derivatives as anti-HIV agents. Biotechnology annual review **2007**, *13*, 223.

13. Wilson RM, Danishefsky SJ: Small molecule natural products in the discovery of therapeutic agents: the synthesis connection. J Org Chem **2006**, *71*, 8329.

14. Sangma C, Chuakheaw D, Jongkon N, Saenbandit K, Nunrium P, Uthayopas P and Hannongbua S. Virtual screening for anti-HIV-1 RT and anti-HIV-1 PR inhibitors from the Thai medicinal plants database: a combined docking with neural networks approach. Comb Chem High Throughput Screening **2005**, *8*, 417.

15. Franke L, Schwarz O, Müller-Kuhrt L, Hoernig C, Fischer L, George S, Tanrikulu Y, Schneider P, Werz O, Steinhilber D, Schneider G. Identification of natural-product-derived inhibitors of 5-lipoxygenase activity by ligand-based virtual screening. J Med Chem **2007**, *50*, 2640.

16. Ehrman TM, Barlow DJ, Hylands PJ: Virtual screening of Chinese herbs with random forest. J of chem info and modeling **2007**, *47*, 264.

17. Piccagli L, Fabbri E, Borgatti M, Bezzerri V, Mancini I, Nicolis E, Dechecchi MC, Lampronti I, Cabrini G, Gambari R: Docking of molecules identified in bioactive medicinal plants extracts into the p50 NF-kappaB transcription factor: correlation with inhibition of NF-kappaB/DNA interactions and inhibitory effects on IL-8 gene expression. BMC Struc Biol **2008**, *8*, 38.

18. Sharma RK, Pande V, Ramos MJ, Rajor HK, Chopra S, Meguro K, Inoue J, Otsuka M: Inhibitory activity of polyhydroxycarboxylate chelators against recombinant NF-kappaB p50 protein-DNA binding. Bioorg Chem **2005**, *33*, 67.

19. Sharma RK, Otsuka M, Pande V, Inoue J, Joao Ramos M: Evans Blue is an inhibitor of nuclear factor-kappa B (NF-kappaB)-DNA binding. Bioorg Med Chem Lett **2004**, 14, 6123.

20. Sharma RK, Chopra S, Sharma SD, Pande V, Ramos MJ, Meguro K, Inoue J, Otsuka M: Biological evaluation, chelation, and molecular modeling studies of novel metal-chelating inhibitors of NF-kappaB-DNA binding: structure activity relationships. J Med Chem **2006**, 49, 3595.

21. Pande V, Sharma RK, Inoue J, Otsuka M, Ramos MJ: A molecular modeling study of inhibitors of nuclear factor kappa-B (p50)-DNA binding. J Comput Aided Mol Des **2003**, 17, 825.

22. 1.TRIPOS, S Associates, St. Louis, MO, 1993, version7.0.

23. Dewar M, Zoebish G, Healy E: AM1: A new general purpose quantum mechanical molecular model. J Am Chem Soc **1985**, *107*, 3902.

24. Maestro. 7.0. Portland, OR:Schrodinger, LLC; 2004.

25. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL: Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem **2004**, *47*, 1750.

26. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS: Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem **2004**, *47*, 1739.

27. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP: Empirical scoring functions: The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des **1997**, 11, 425.

28. Borgatti M, Lampronti I, Romanelli A, Pedone C, Saviano M, Bianchi N, Mischiati C, Gambari R: Transcription factor decoy molecules based on a peptide nucleic acid (PNA)-DNA chimera mimicking Sp1 binding sites. J Biol Chem **2003**, 278, 7500.

29. Borgatti M, Breda L, Cortesi R, Nastruzzi C, Romanelli A, Saviano M, Bianchi N, Mischiati C, Pedone C, Gambari R. Cationic liposomes as delivery systems for double-stranded PNA-DNA chimeras exhibiting decoy activity against NF-kappaB transcription factors. Biochem Pharmacol **2002**, *64*, 609.

30. Borgatti M, Bezzerri V, Mancini I, Nicolis E, Dechecchi MC, Lampronti I, Rizzotti P, Cabrini G, Gambari R: Induction of IL-6 gene expression in a CF bronchial epithelial cell line by Pseudomonas aeruginosa is dependent on transcription factors belonging to the Sp1 superfamily. Biochem Biophys Res Commun **2007**, *357*, 977.

31. Bezzerri V, Borgatti M, Nicolis E, Lampronti I, Dechecchi MC, Mancini I, Rizzotti P, Gambari R, Cabrini G. Transcription factor oligodeoxynucleotides to NF-kappaB inhibit transcription of IL-8 in bronchial cells. Am J Respir Cell Mol Biol **2008**, *39*, 86.

CHAPTER 4. Virtual Screening of Chemical Libraries against NF-κB

4.1 Introduction

In our previously docking studies (see chapter 3) on a set of 27 compounds from bioactive plant extracts into the p50 NF- κB, a weak inhibitor of NF- κB p50/DNA interaction and correlated IL-8 gene expression has been individuated. In spite of the well defined solved NF-κB 3D structures, to the best of our knowledge, nowadays VS applications against this interesting target for discovering of new NF- κB/DNA inhibitors has not been published.
In this work, we present a successful VS of small chemical compounds library against NF-κB p50 homodimer and against NF-κB p50 monomer through ensemble docking simulations.
To this purpose, we have employed the published structures **1i-6i** and **1n-6n** (see chapter 2, Table 1) as test set in prepared compound library. As further electrophoretic mobility shift assays (EMSA) and biological evaluation of the most important hit compounds were carried out for a precise validation of the VS screening results. We obtained fair agreement between experimental and theoretical computational findings.

4.2 Methods

4.2.1 Preparation of the compounds libraries.

The web-accessible ZINC three-dimensional (3D) database (http:/blaster.docking.org/zinc/)[1] containing over 4.6 million commercially available structures was first filtered by specifying a subset of purchasable compounds sharing (2E,Z)-3-(2-hydroxyphenyl)-2-propenoate substructure (molecular weight = 161,1385) and relevant drug-like properties (molecular

weight < 500; number of hydrogen-bond acceptors ≤ 10; calculated LogP ≤ 5).[2] All 1360 remaining structures were first minimized in the OPLS_2001 force field[3] applying the conjugate gradient method by using Schrödinger, Inc script premin which applies MacroModel' s multiple minimization capabilities.[4] The compounds library was further processed with LigPrep module of Maestro in the Schrodinger suite of tools[5] to obtain a customised database of 1478 optimized structures. Relevant tautomeric states and correctly protonated forms of molecules between pH 7.2 and 7.6 were considered. A test set of 6 known inhibitors (**1i-6i**) and 6 inactive compounds (**1n-6n**) (see Chapter 2, Table 1) was prepared using the above described procedure starting from the optimization process.

4.2.2 Preparation of protein targets and binding sites.

The three dimensional structures of the NF-κB/DNA complexes was retrieved from the Protein Data Bank (PDB codes: 1NFK, 1LE9).[6,7] Starting from the cocrystallized complexes, NF-κB p50-p50 homodimer (from 1NFK), p50 monomer chain A (here referred to as p50a, from 1NFK), p50 monomer chain B (here referred to as p50b, from 1NFK) and p50 monomer (here referred to as p50c, from 1LE9) for docking were prepared using Maestro graphical interface. p50 subunits were selected by removing cocrystallized DNA and p65 subunit (Figure 1). For all the subsequent procedure referred to experimental section of chapter 2.

4.2.3 Docking Simulations and Ligands Ranking.

All small molecular weight compounds of the chemical library prepared as described above, were docked into the putative binding site of each macromolecule target using Glide .First, Standard-Precision (SP) method as implemented in Glide, was employed here to screen the starting library of 1478 compounds into NF-κB monomers (p50a, p50b and p50c) and into the dimer protein. The maximum number of poses per ligand to pass to the grid-refinement calculation was set to 10.000. Only one good pose for each molecule, established by a composite “Emodel” score, was stored. Emodel combines GlideScore (G-score), the nonbonded interaction energy and the excess internal energy of the generated ligand conformation. Subsequently, based on the G-score function, the highest-ranking 300 docked ligands into any member of the target were selected and collected in a new multifile for

further docking runs. After a minimization process in the OPLS_2001 force field by premin script, the picked out compounds were docked again into DNA recognition site, using the more CPU time-intensive and accurate Extra-Precision (XP) method. Finally, all molecules with the highest-ranking poses (G-score) against the best single target were selected for further experimental analysis.

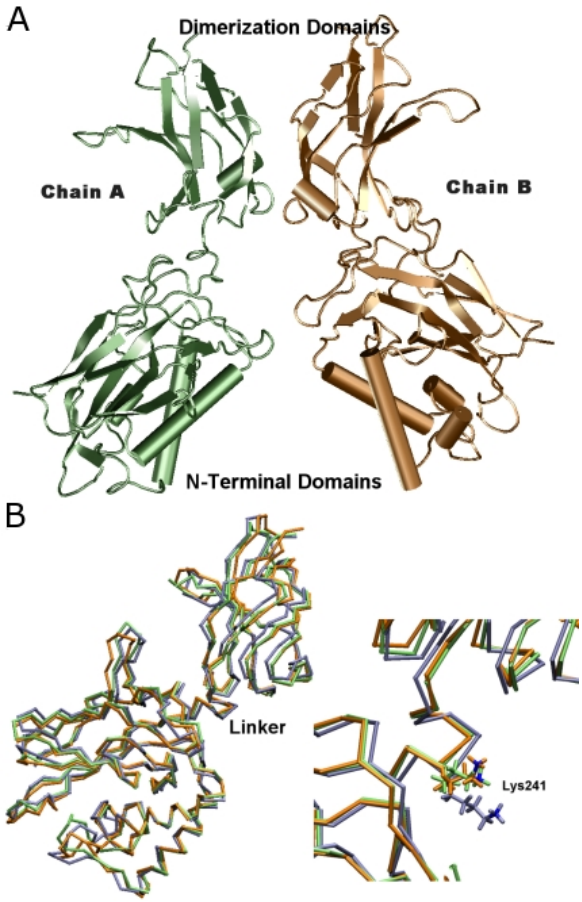


Figure 1. Stereo view of three-dimensional structures of NF-κB targets: (A) p50 homodimer (from 1NFK) and (B) superimposed different conformations of p50 monomer (from 1NFK and 1LE9): p50 (shown in green), p50b (shown in orange) and p50c (shown in blue); in the bottom, the remarkable movement of Lys241 of the linker region of p50c in respect to the other conformations is highlighted.

4.2.4 Optimization procedure on the best-scored compounds.

The promising and purchasing ZINC database compounds 1797601 [**1**], 1688817 [**2**], 1345112 [**(R)-3**], 1345109 [**(S)-3**] and 1983981 [**(R)-4**] (Figure 2) were further investigated. The three dimensional structure geometries of their resulting docked poses were optimized and charges were assigned by a semiempirical molecular orbital calculations using the AM1 Hamiltonian[8] implemented in MOPAC software.[9] Afterwards, the molecules were docked into the dimer and monomer (p50a, p50b and p50c) targets applying accurate XP method implemented in Glide as described above. In this case, the total number of docked retained poses for each compound was set to 100. The final ranking for choosing the best poses and for describe the most representative binding modes of compounds in each target, was established on E-model values.

4.3 Results and discussion

Starting with ZINC 3D database[1] specifically developed for VS application and here employed, there are more than 4,6 million commercially available structures from over 50 compound vendors. A forward filter based on a substructure searching in the ligands library preparation by using 2D chemical information derived from a biological active molecule previously discover by us (see chapter 2) was used to potentially increase its potency as well as the VS enrichment factor. Moreover relevant drug-like properties were taken in to account. After the filtering process, 1478 structures were sampled in the prepared library to be screened trough docking simulations.

Due to the lack of solved NF-κB cocrystallized inhibitors, a common RMSD comparison in the self-docking procedure was not practicable. Therefore in order to evaluate the accuracy of the docking protocol adopted and to individuate a score value threshold for selecting interesting hits, a data set of 12 compounds selected from the known inhibitors employed previously and arbitrarily divided in two clusters of 6 actives **(1i-6i,** $IC_{50} \leq 500\mu M$) and 6 inactives (**1n-6n**, $IC_{50} > 500\mu M$) in inhibition of p50/DNA binding (Chapter 2, Table 1). Indeed, most of the known molecules of the two clusters were structural analogues that differ

by single elements. The compounds here considered were selected among the ones obtained the best score in preceding docking studies on natural compounds (Chapter 3).
Finally, in order to account for p50 NF-κB different aggregation states as above mentioned, both monomer and dimer proteins have been used as targets for our VS applications.

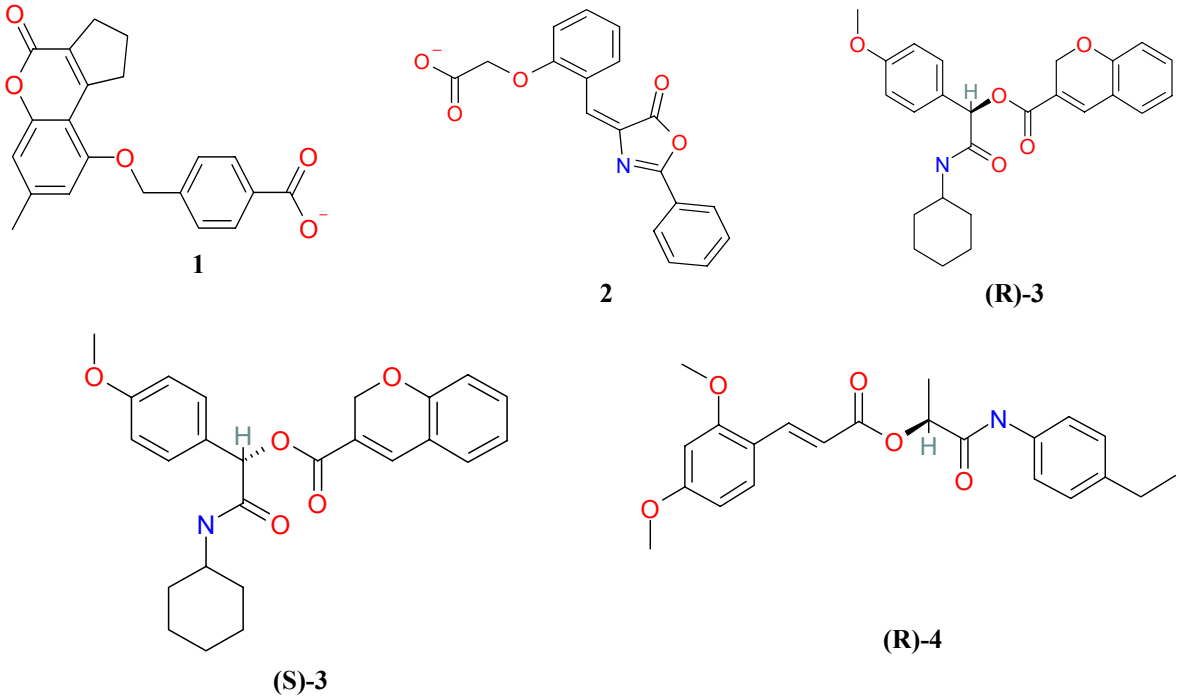


Figure 2. Selected and experimentally tested interesting hits docked compounds

4.3.1 NF-κB p50 monomer as protein target.

The alpha-carbon atom-pairs superimposition of p50a with p50b and p50c, prepared as described in experimental section, gave a overall root mean square deviation (RMSD) for non hydrogen atoms of 2,11 Å and 2.36 Å respectively. These values shift to 1.61 Å for p50b and to 2.64 Å for p50c considering only the residues selected for the definition of the binding box (see method section) (Figure 1, panel B). In particular, Lys 241 of p50c that interact with DNA in the putative binding site, clearly moves away from the position of the same residue in the corresponding subunits (Figure 1, panel B).

Thus, ensemble docking simulations with the same protocol on all p50 subunits (p50a, p50b and p50c) supposed to be as three slightly different conformations of the same structure were carried out to account for protein flexibility, a frontier problem especially for VS applications. The docking results for all the ligands under study are given in Tables 1, 2 and 3. Comparing the G-score ranking for active and inactive compounds, it is possible to partially discriminate the two class of molecules in the monomer p50a (Table 1). Nevertheless, all NF-κB inhibitors were retrieved within the first 300 selected molecules for subsequently XP docking runs in all the targets, except for 5i in p50c.

Consistently, 5i NF-κB binding involved Lys 241, as shown in our previous docking studies on p50a and p50b reported in chapter 2. As discussed above, the side chain of this residue assumed a different rotameric form in respect to p50a and p50b, forming a steric hindrance to **5i** in the NF-κB interaction area.

In our study the results of the multiple VS were compared and condensed to improve the enrichment factor (EF). EF was first calculated to quantify the docking protocols ability to assign high ranks to known active ligands.

Measurement of EF versus known inhibitors (**1i-6i**) was estimated following Pearlman and Charifson[11] as:

$$EF(s) = \frac{Hits_s}{N_s} \Big/ \frac{Hits_{total}}{N_{total}} \tag{4.1}$$

with N_{total} of 1484 (number o molecules in the library) and N_s corresponding to the rank of the last active found (Tables 1 and 2).

The EF values were calculated to recover all of the hits actives (6 inhibitors) referred to each target separately (p50a = 12.90; p50b = 6.60) and for the ensemble targets (EF = 18.10). This last measure was calculated by first find out the ensemble score for each test ligand defined as the score against the best single target of the ensemble set (Table 2). The calculated EF values versus the known inhibitors under study (**1i-6i**) showed that an ensemble of multiple conformations of NF-κB in VS studies improved enrichment of actives relative to using a single macromolecule structure. Thus, even if the ensemble molecular docking protocol cannot really discriminate between active and inactive compounds of close related structures, it could efficiently retrieve active compounds from a large molecular database.

A reasonable number of compounds for our further experimental analyses (18 molecules), over ranked the best G-score of known inhibitors under study.

Subsequently, only the compounds over ranking the best scores of known inhibitors of the screened database (compound **2i**, G-score = 6.98), were here considered as possible interesting hits. Out of 18 molecules individuated (data not shown), only ZINC database compounds **2**, **(R)-3**, **(S)-3** and **(R)-4** were purchasable and thus chosen for further analysis.

Target								
p50a			p50b			p50c		
Comp.	Ranking[§]	Score[a]	Comp.	Ranking[§]	Score[a]	Comp.	Ranking[§]	Score[a]
5n	/	-6.88	**2i**	10	-6.98	**4i**	32	-6.31
2i	27	-6.29	**4n**	/	-5.26	**4n**	/	-6.07
1i	36	-6.14	**5n**	/	-5.21	**2n**	/	-5.33
6i	41	-6.06	**3i**	199	-4.76	**3i**	126	-5.08
3i	63	-5.85	**2n**	/	-4.65	**3n**	/	-4.61
2n	/	-5.63	**1i**	215	-4.46	**5n**	/	-4.32
5i	82	-5.63	**5i**	219	-4.28	**1i**	252	-4.02
4i	115	-5.15	**6i**	220	-4.27	**6i**	253	-4.00
4n	/	-2.79	**4i**	225	-4.11	**1n**	/	-3.83
3n	/	-1.55	**3n**	/	-3.92	**2i**	260	-3.78
1n	/	-0.3	**1n**	/	-2.72	**6n**	/	-0.47
6n	/	+0.3	**6n**	/	+2.09	**5i**	---	---

Table 1. Ranking of best poses of the scored test set ligands docked into p50 monomer.
[§]*values of inactives **1n-6n** are not reported since they are not included in the ranking evaluation.*
[a]*Glide Score (G-score) values obtained applying the XP docking method.*

Compound	*Score*[a]	*Rank*
2i	-6.98	10
4i	-6.31	32
1i	-6.14	36
6i	-6.06	41
3i	-5.85	63
5i	-5.63	82

Table 2. Final ranking of known inhibitors obtained from ensemble docking into p50 monomer for the calculations of VS enrichment factor.
[a]*Glide Score (G-score) values obtained applying the XP docking method*

Compound	*Score*[a]	*Rank*	*Target*
2	-8.01	1	p50b
	-2.64	253	p50a
	>0	>300	p50c
(R)-3	-7.61	3	p50b
	-4.36	181	p50a
	>0	>300	p50c
(S)-3	-7.63	2	p50b
	>0	>300	p50a
	>0	>300	p50c
(R)-4	-6.89	17	p50b
	-5.27	109	p50c
	>0	>300	p50a

Table 3. Ranking of the selected and experimentally tested ZINC database hits retrieved from VS studies against p50 monomer
[a]*Glide Score (G-score) values obtained applying the XP docking method*

4.3.2 NF-κB p50-p50 dimer as protein target.

The results of docking simulation of all ligands under study in VS against NF-κB dimer are shown in Tables 4 and 5.

As in the case of the NF-κB monomeric target, only the purchasable molecules were selected for further investigations. The compound **1** (Figure 2) was here retrieved as interesting hit in addition to compounds **2**, (**R)-3**, **S)-3** and **(R)-4** also individuated from preceding VS studies against the monomer protein. All docked compounds were located in the DNA-binding surface of the dimer that comprise an area formed by the 10-residues' linker and the N-terminal domain of each monomeric subunit (in common with the monomeric state of the protein) and a region formed by the spatial relationship between N-terminal domains of each subunit (Figure 1, panel A).

Compound **1** clearly showed highest score (G-score = -8.22) at physiological pH outranking all docked compounds considered in the present study (Tables 4 and 5). For the selected ligands, the pose with best E-Model score (combination of energy grid score, GlideScore, and the internal strain of the ligand) after optimisation process were used for in-depth interaction analysis.

The best pose of compound **1** (E-Model score = 47.3) occupied the region formed by residues of both p50 subunits: Lys144, Lys145, Thr143, Tyr57, Cys59 of p50a and Lys145, Lys146, Thr143 of p50b (Figure 3). In particular, the carboxylate group of **1** made two hydrogen bonds: the first one with NH_3 protonate group of Lys 146 (p50 chain B), stabilized by a salt bridge (distance CO:::HN=2.00 Å) and the second one with the OH group of the Thr143 (p50 chain B) (distance CO:::HO=2.22 Å). Moreover, the etheric oxygen atom linked to the coumarin moiety is involved in a hydrogen interaction with the protonated NH_3 function of Lys 145 (p50 chain B) (distance O:::HN=1.90 Å). In addition, the carbonyl group of coumarin function of **1** engaged a hydrogen bond with the NH group of Cys59 (p50 chain A) backbone (distance CO:::HN=1.98 Å) (Figure 3). It very interesting to note that, Cys62 (corresponding here to Cys59) is highly conserved among many members of the Rel family and it is a critical residue for the DNA binding activity of p50. In particular, as many evidences suggest, the reduced form of this residue in the nucleus is a essential prerequisite for NF-κB activation in vivo.[12]

Comp.	Ranking[§]	Score[a]
1i	145	-6.12
2i	153	-6.05
6i	154	-6.04
4n	/	-5.78
3n	/	-5.64
6n	/	-5.58
2n	/	-5.49
1n	/	-5.26
5n	/	-5.25
3i	264	-4.80
5i	269	-4.62
4i	293	-3.73

Table 4 Ranking of best poses of the scored test set ligands docked into p50 dimer in VS applications
[§]*values of inactives **1n-6n** are not reported since they are not included in the ranking evaluation.*
[a]*Glide Score (G-score) values obtained applying the XP docking method.*

Compound	*Score*[a]	*Rank*
1	-8.22	1
(S)-3	-7.62	3
2	-7.49	5
(R)-4	-7.21	15
(R)-3	>0	>300

Table 5. Ranking of the selected and experimentally tested ZINC database hits from VS studies against NF-kB dimer p50-p50

Highest ranked poses of **2**, **(R)-3, (S)-3** and **(R)-4** (Table 3) showed different binding modes. In particular, molecule **(S)-3** (G-score of -7.62) lost the ability to occupy the same position in the protein and compound **2** (G-score of -7.49) showed to form hydrogen bonds only with residues of one of the p50 subunit (chain B). (Figure 4).

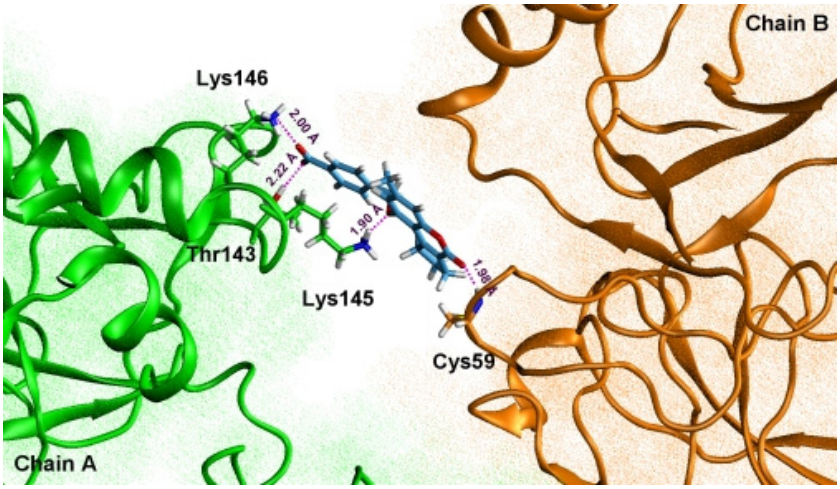


Figure 3. Binding mode of the best pose of compound **1** docked into the DNA-binding region of NF-κB p50 homodimer (chain A, shown in green). The residues involved in hydrogen bonding with the ligand are shown in purple

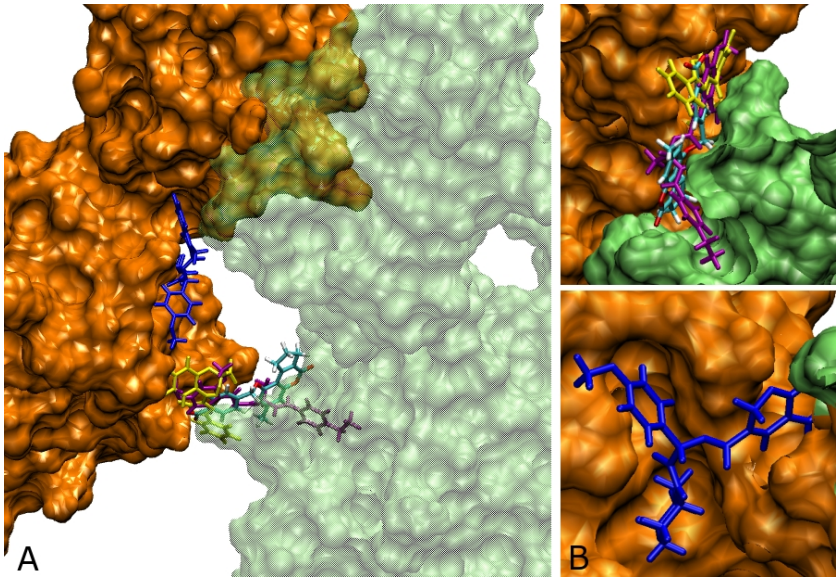


Figure 4. Stereoview of selected and experimentally tested compounds **1**, **2**, **(S)-3** and **(R)-4** docked into DNA binding region of the NF-kappaB homodimer. The macromolecule surface is highlighted (chain A in green) and compound **1**, **2** (in yellow), **(S)-3** (in blue) and **(R)-4** (in purple) are shown in licorice representation. A focus on the binding sites of **1**, **2**, **(R)-4** (panel B, above) and of **(S)-3** (panel B, below) is displayed.

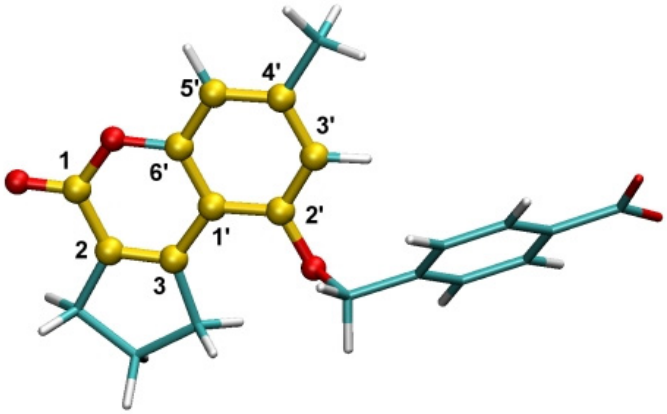


Figure 5. Active compound **1**: 4-[7-Me-4oxo-1,2,3,4-4H-cyclopenta(C)chromen-9-yloxymethyl]-benzoic acid. The reference (2Z)-3-(2-hydroxyphenyl)-2-propenoate substructure forming the coumarin moiety is highlighted in ball and stick representation (shown in yellow).

4.3.3 Effects of the best-ranked compound on NF-κB/DNA interactions

Based on computational findings, the effects of compound **1** on NF-κB interactions were first studied by EMSA in the Laboratory of Prof. R. Gambari as described for the docking studies reported in Chapter 3.

The results of the gel retardation analysis clearly demonstrate that compound **1** inhibits the molecular interactions between isolated NF-κB p50 and a target double stranded oligonucleotide mimicking the NF-κB binding sites. This inhibitory activity was higher than that exhibited by the NF-κB inhibitors **4i** and **6i**. This finding is of particular interest, since compound **1** contains the reference 3-(2-hydroxyphenyl)-2-propenoate substructure blocked in a rigid cyclic conformation (Figure 6) forming the 2H-Chromen-2-one (coumarin) moiety. Interestingly, the known coumarins inhibitors **5i** and **6i** here reported showed a lower inhibitory activity on NF-κB/DNA interactions than compound **1**.

In a second experiment, the inhibitory activity of compound **1** on NF-κB/DNA interactions was compared with the activity of compounds **2**, **(R/S)-3** and **(R/S)-4.** In agreement with the molecular modelling findings, compounds **(R/S)-3** and **(R/S)-4** exhibited low efficiency, when compared to compound **1** in inhibiting NF-κB/DNA interactions. In should be

underlined that different biological profiles might be observed in further experimental analysis to separate **(R)-3**, **(S)-3** and **(R)-4** chiral compounds as considered in molecular modeling studies.

Finally, the biological effects of compound **1** were investigated based on the inhibition of *Pseudomonas Aeruginosa* (PAO-1) mediated Increase of IL-8 mRNA. All biological assays were carried out in the laboratory directed by Prof. Cabrini (Hospital-University of Verona). The results obtained demonstrate that compound **1** is a strong inhibitor of PAO-1 induced accumulation of IL-8 mRNA. The effects were similar to those obtained using a synthetic double stranded decoy oligonucleotide against NF-κB.[13]

4.4 Conclusions

In this chapter, we have shown that VS applications against NF-κB p50 employing docking simulations are suitable for the identification of low molecular weight compounds interacting with NF-κB and inhibiting NFkB/DNA interactions and NF-κB dependent functions.

In agreement with the docking results to NF-κB dimer, the highest scored (G-score = -8.22) compound **1** showed to be active in further EMSA experimental and biological studies.

Moreover, we like to hypothesize that NF-κB p50 monomer could be an useful target for VS application through ensemble docking method here reported, since the calculated EF of known inhibitors was quite good. We acknowledge that several VS tests on different data set and targets conformations would be necessary to fully address this issue. These findings, including the resulting biological effects of inactive compound **2**, reinforce the hypothesis that the coeval involvement of both the subunits of the dimer could be improve inhibitory potency showed by molecule **1**. In particular, to account for inhibitory activity of **1**, we could propose that its binding to the dimer may masks the Cys62 and competitively blocks the access of p50 subunit to the DNA target sequence in the nucleus. Moreover, the stabilization of the dimeric form of NF-κB trough a bridge structure could reduce the amount of p50 more flexible monomeric subunit free to form different trimeric complex with DNA. Of course, both these two synergic hypothesis may contribute to the inhibition processes showed by compound **1** bound to NF-κB.

Despite the use of a reference substructure for the initial filtering process in library preparation, it should be point out that compound **1** (Figure 5) presents an important molecular diversity in respect to all known p50 NF-κB coumarin binders reported in literature. Discovered micromolar bioactive compound **1** can be considered as a "lead" for developing stronger NF-kB/DNA inhibitors. Information deduced from its chemical profile and peculiar binding mode, could carry through further rational design of more potent NF-κB inhibitors.

References

1. Irwin JJ, Shoichet BK: ZINC--a free database of commercially available compounds for virtual screening. J. Chem Inf Model **2005**, *45*, 177.

2. Lipinski CA: Drug-like properties and the causes of poor solubility and poor permeability. J. Pharmacol. Toxicol Methods **2000**, *44*, 235.

3. Jorgensen WL, Maxwell DS; Tirado-Rives J: Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc **1996**, *118*, 11225.

4. Mohamadi F, Richards MGJ, Guida WC: Liskamp, R.; Lipton M.; Caufield, G.; Chang, G.; Hendrickson T. W.; Still, C. MacroModel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. J of Comput Chem **1990**, *11*, 440.

5. Maestro (v7.0.113) – A unified interface for all Schrodinger products, developed and marketed by Schrodinger, LLC. NY, Copy-right 2005; http://www.schrodinger.com.

6. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL: Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem **2004**, *47*, 1750.

7. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M Perry JK, Shaw DE, Francis P, Shenkin PS: Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem **2004**, *47*, 1739.

8. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP: Empirical scoring functions: The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J. Comput Aided Mol Des **1997**, *11*, 425.

9. Dewar M, Zoebish G, Healy E: AM1: A new general purpose quantum mechanical molecular model. J Am Chem Soc **1985**, *107*, 3902.

10. MOPAC2007, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net, 2007.

11. Charifson PS: Improved scoring of ligand-protein interactions using OWFEG free energy grids. J Med Chem **2001**, *44*, 502.

12. Xia YF, Ye BQ, Li YD, Wang JG, He XJ, Lin X, Yao X, Ma D, Slungaard A, Hebbel R, Key NS, Geng JG: Andrographolide Attenuates Inflammation by Inhibition of NF-κB Activation through Covalent Modification of Reduced Cysteine 62 of p50 J Immunol **2004**, *173*, 4207

13. Bezzerri V, Borgatti M, Nicolis E, Lampronti I, Dechecchi MC, Mancini I, Rizzotti P, Gambari R, Cabrini G. Transcription factor oligodeoxynucleotides to NF-kappaB inhibit transcription of IL-8 in bronchial cells. Am J Respir Cell Mol Biol **2008**, *39*, 86.

CHAPTER 5. Focus library of Furocoumarin derivatives for VS against NF-κB

5.1 Introduction

Furocoumarins are a class of organic chemical compounds derived from plants. The chemical structure of furocoumarins consists of a furan ring fused in different ways with coumarin producing several isomers. Among them, the most usual are angelicin and psoralen that form respectively the two core structures of many linear and angular derivatives (Table 1).

Angelicin has been proposed in the literature for the treatment of psoriasis[1] and as antiproliferative,[2-4] antifungal,[5] and anti-inflammatory agents.[6]

It is also known that angelicin and its structural analogues, again in the fundamental state, are able to induce gamma-globin gene expression and the biosynthesis of HbK.[7] Due to this activity, these molecules can be of interest in the therapy of β-thalessemia and falciform anemia.

Recently, some of us have reported that furocoumarins photolysis products induce differentiation of human erythroid cells. These data suggest that photoproducts of psoralen warrant further evaluation as potential therapeutic drugs in β-thalassaemia and sickle cell anaemia.[8] By chemical synthesis Miolo et al. prepared new benzoquinolizin-5-one derivatives as furocoumarin analogues able to intercalate inside DNA and by subsequent irradiation with UVA light, to photoreact with DNA. Some molecular modelling studies to describe the photocycloaddition reaction mechanism have been shown.[9]

Interesting, our previously reported VS application against NF-kB, (see chapter 4) brought out the identification of the compound **1** which can be considered as a furocoumarin analogue (cyclopentane ring instead of furan function). Molecule **1** is able to inhibit the NF-kB/DNA complex formation and dependent biological functions by directly binding the protein with an IC50 of about 100 μM. Thus, we constructed a focus library of furocoumarin derivatives,

including derivative of molecule **1**, differently substituted for a in silico screening against NF-kB. The aim was finding more potent NF-κB inhibitors as well as to check whether angelicin, psoralen and derivatives were also capable of directly inactivate NF-κB target as an important point of action.

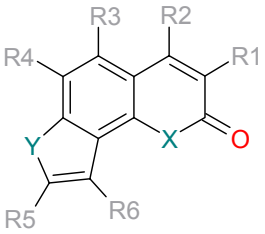
5.2 Methods

5.2.1 Preparation of a focus library of furocoumarin compounds

Natural and synthetic organic compounds available in a structure searchable database were retrieved using a substructure searching (2D) method. The distributors accessible on the network that we used are the followings: Sigma-Aldrich[10], Maybridge[11], Otava[12], TimTec[13], Chembridge[14], and DrugBank.[15]

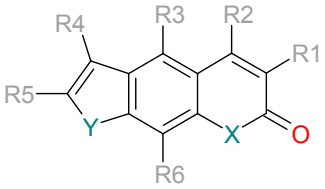
Angelcin and psolaren derivatives reported in Table 1 were build employing Maestro graphical interface. Their 3D geometry was optimized with bmin by using OPLS-AA force field.[16] The browsing hits were retrieved from the substructure match obtained drawing the chemical scaffolds reported in Table 1. The key pratical informations of each compound such as smiles, molecular weight and catalog entry were conserved as label in the structure files. All the molecules of the library were collected and prepared using LigPrep[17] from Schrodinger in a standard procedure. All the ouput files in 2D sdf format were converted in 3D maestro format by using *sdconvert*. All charged group were neutralize and subsequently ionization treatment was done generating all possible states at pH of 7.4 (+-2). Eventual counter ions in salts and water molecules were removed. Various tautomer for each structure were generate. The molecules with a molecular weight > 800 was not considered. In order to remove all the possible problematic structures premin script from Schrodinger was employed using OPLS-AA force field for minimization process. Finally, we obtained a library of 1705 optimized furocoumarin structures.

Angelicin scaffold and Derivatives



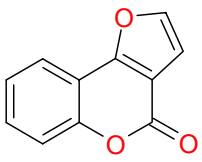
Compounds[a]	R1	R2	R3	R4	R5	R6	X	Y
Angelicin	H	H	H	H	H	H	O	O
4-Methylangelicin	H	CH_3	H	H	H	H	O	O
4,5'-Dimethylangelicin	H	CH_3	H	H	CH_3	H	O	O
6,5-Dimethylangelicin	H	H	CH_3	CH_3	H	H	O	O
4',5'-Dimethylangelicin	H	H	H	H	CH_3	CH_3	O	O
4,5',4'-Trimethylangelicin (TMA)	H	CH_3	H	H	CH_3	CH_3	O	O
4,6,2' -Trimethylfluoroquinolinone	H	CH_3	H	CH_3	CH_3	H	NCH_3	O
4,2',3'-Trimethyltioangelicin	H	CH_3	H	H	CH_3	CH_3	O	S

Psolaren Scaffold and Derivatives



Compounds[a]	R1	R2	R3	R4	R5	R6	X	Y
Psolaren	H	H	H	H	H	H	O	O
4-Methylpsolaren	H	CH_3	H	H	H	H	O	O
5-Methylpsolaren	H	H	CH_3	H	H	H	O	O
4,5'-Dimethylpsolaren	H	CH_3	H	CH_3	H	H	O	O
5,4'-Dimethylpsolaren	H	H	CH_3	CH_3	H	H	O	O
3,4-Dimethylpsolaren	CH_3	CH_3	H	H	H	H	O	O
4,5',8 -Trimethylpsolaren (TMP)	H	CH_3	H	H	CH_3	CH_3	O	O
5-Methoxypsolaren	H	H	OCH_3	H	H	H	O	O
8-Methoxypsolaren	H	H	H	H	H	OCH_3	O	O
5,8 -Methoxypsolaren	H	H	OCH_3	H	H	OCH_3	O	O
4,4',5' -Trimethylpsolaren	H	CH_3	H	CH_3	CH_3	H	O	S

Furo(3,2-c)chromen-4one Scaffold [b]



Benzoquinolizin-5-one scaffold[c]

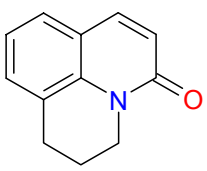


Table 1. Furocoumarin library
[a]molecules listed by Viola et al. [b]chemical structure derived by Jie Wu [c]proposed by Miolo et al.

5.2.2 Docking of the furocoumarin library into DNA binding of NF-κB

In order to screen the starting library of 17010 structures into each macromolecule, all the compound library were docked into putative binding site of the targets (NF-kb: p50ab, p50a, p50b, p50-p50, see Chapter 4) employing Glide software.[18]

The proteins, the binding box and the grids were previously prepared as described in chapter 2 and chapter 3. Moreover, NF-κB p50-p65 was prepared for docking runs in the same procedure reported for the p50 homodimer. To this purpose the two dimers were superimposed by aligning the common p50 subunit (exactly the same amminoacidic sequence). Subsequently the coordinates for define the area of the binding box in DNA-binding region were set at same values (x = -1.1958 Å; y = 9.0149 Å; z = 19.7598 Å) for grid preparation. The Standard-Precision (SP) method as implemented in Glide, was employed as first for p50 NF-kB. The maximum number of poses per ligand to pass to the grid-refinement calculation was set to 10.000. Only one good pose for each molecule was retained. Subsequently, all possible redundant structures were eliminated from the database using the table project implemented in Glide. Based on the G-score function, the highest-ranking 1000 docked ligands were selected and collected in a new multifile for further docking runs. After a minimization process in the OPLS_2001 force field by premin script, the picked out compounds were docked again into DNA recognition site, using the more CPU time-intensive and accurate Extra-Precision (XP) method. Finally, all molecules with the highest-ranking poses (G-score) against the best single target were selected for further experimental analysis and for XP docking studies on NF-κB p50-p65.

5.3 Results and discussion

The scaffolds for each one of the four classes of the furocoumarin analogues including compounds retrieved from the literature are shown in Table 1. Finally, the customized library of 1710 compounds contained 54 structures of angelicin derivatives (molecular weight range: 186-478), 1384 of psolaren derivatives (molecular weight range: 186-619), 51 of Furo(3,2-c) chromen-4one structures (molecular weight range: 268-439) and 166 of benzoquinolizin-5-one analogues (molecular weight range: 258-499).

VS approach was followed against both NF-κB p50 monomer and homodimer. The selection of interesting compounds was based on the docking XP GlideScore rank for each target. Almost one best scored ligand was chosen among each one of the four classes of chemical structures.

Moreover, the molecules with high GlideScore value in more than one configuration of the protein were preferred. Indeed, our previously VS studies on data set of known protein binders showed an increased of the enrichment factor in ensemble configurations of p50 target in respect to the docking application on just one conformation of the macromolecule.

In Table 2 in silico screening docking results are shown relatively to the first ten selected ligands (Figure 1) for each configuration of the target. Unfortunately only compounds **6f-10f** were purchasable and expect for the benzoquinolizin-5-one function, all the reference chemical scaffolds are included among them. Subsequently, interesting hits plus best-ranked compound **1f** (GlideScore= -10.30), were in deeply investigated for depicted the most representative ligand binding modes into the dimer. The results here described are focused on p50 homodimer. In fact, as shown in Table 2, compounds **1f**, **6f-10f** exhibited the highest docking ranks in the dimeric aggregation of the protein.

In agreement with further experimental EMSA studies, compounds **6f**, **7f**, **9f** and **10f** inhibited the formation of a stable NF-κB p50/DNA complex. The most active compounds **10f** and **9f** are psoralen derivatives and had an IC_{50} of about 20 μM and 25 μM respectively. The furo(3,2-c)chromen-4one derivative **7f** showed an IC_{50} of about 50 μM while the angelicin derivative **6f** showed an activity of 100 μM.. In experimental studies molecule **8f** was almost inactive (false positive).

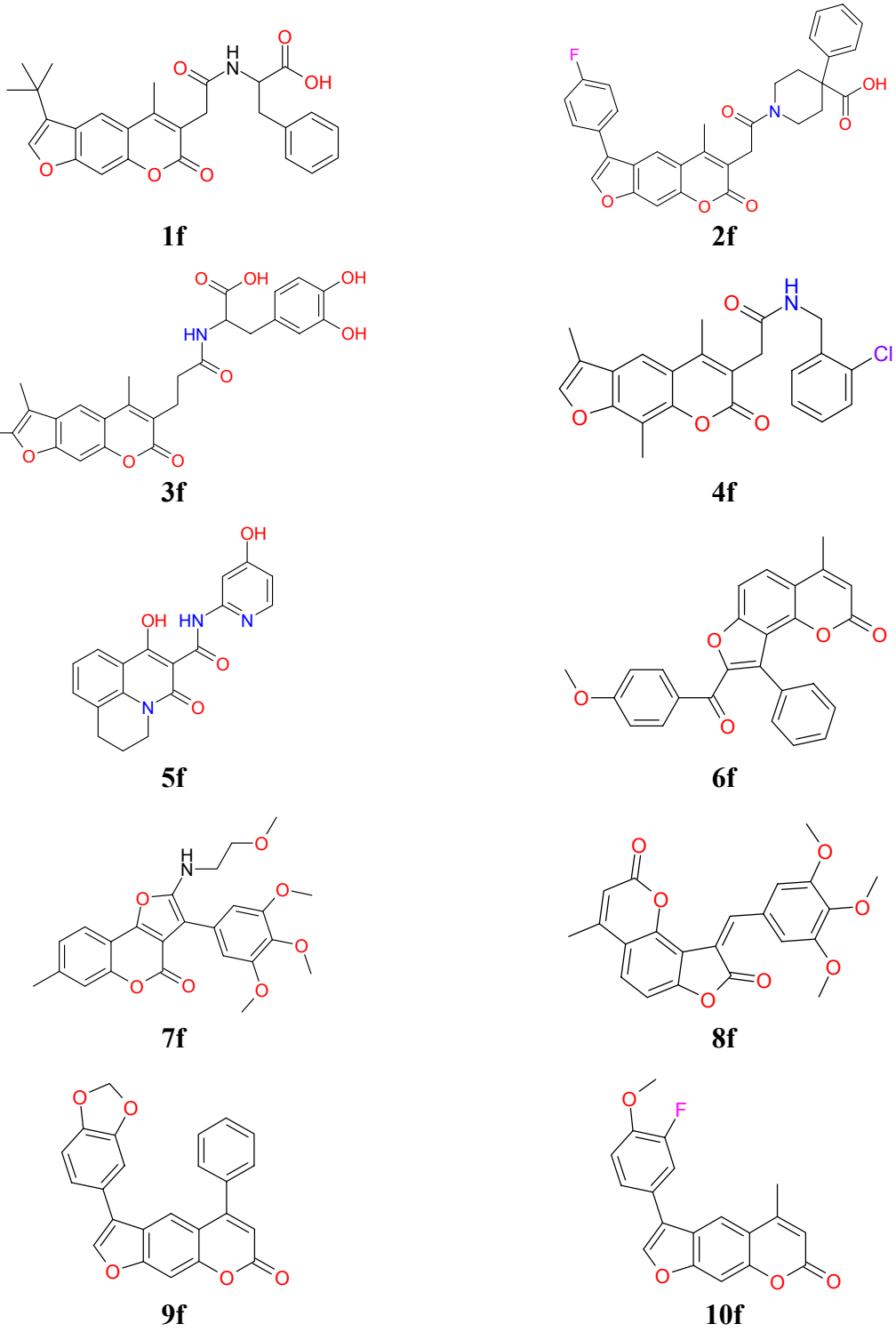


Figure 1 Chemical two dimensional structures of hits from VS against NF-kB p50 homodimer

Compound	GlideScore			
	p50-p50	*p50a*	*p50b*	*p50c*
1f	-10.30	/	-6.13	-6.424
2f	-9.65	-6.82	-6.87	/
3f	/	-8.06	-3.14	/
4f	-8.05	-7.01	-6.30	/
5f	/	-7.71	-5.49	/
6f	-7.48	/	/	-6.19
7f	-7.44	/	-6.94	/
8f	-7.30	/	-5.42	/
9f	-7.21	/	/	/
10f	-6.95	/	/	/

Table 2. Best score of the docked molecules (**1f-10f**) into both NF-kappaB p50 dimer (p50-p50) and monomers (p50a, p50b, p50c).

5.3.1 Ligand binding modes in DNA binding region of p50 homodimer

The choice of the best-docked structure belonging to the most populated cluster of poses was made using the model energy score (Emodel) and visual inspection. As Figure 2 and Figure 3 point out, the best ranked compounds **1f** and **6f**, **7f**, **9f** and **10f** occupied a region involved in the binding of compounds **21** and **1** as previously discovered by us (see Chapters 3 and 4). This area is formed by residues of both p50 units (chain A and chain B) of NF-kB. All the residues of the protein involved in molecular interactions with the active ligands formed hydrogen bonds also with DNA. In Table 3 are shown the molecular interactions of the best-ranked compound (GlideScore = -10.30Kcal/mol) **1f** and of the active molecules **6f, 7f, 9f** and **10f** with the protein**.**

All these compounds interact preferentially with residues of p50 chain B and present a very similar binding mode of compounds **21** and **1**. In particular, highest-ranked molecule **1f** engages a hydrogen bond with the sidechains of Thr143 and its carboxylate group forms a salt bridge stabilized by one hydrogen bond with the side chain of Lys 145. Moreover the phenyl structure of compound **1f** is involved in a π-π stacking interaction with the aromatic moiety of Tyr57 a residue specific for kB DNA sequence 5'-GGGATTTCC-3', present in different cellular genes including HIV-LTR. Also molecule **1f** forms an additional hydrogen bond with

the amino group of Lys146 of the opposite p50 unit (chain A) (Figure 2). As previously reported for the binding mode of active compound **1**, and unlike the other binders here individuated, the coumarin moiety of the active molecule **7f** was involved in two hydrogen bonds with Cys59, a very important residue for NF-κB activation in vivo.

Considering the NF-κB p50-p65 as target for subsequent docking studies based on results of VS against the homodimer, compounds **1f** and **6f**, **7f**, **9f** and **10f** occupied a different region formed by the 10 residues linker and the dimerization domains of each subunit. These findings are probably due to diverse targeting residues in p65 domain and to the different accessible surface area formed by the p65 and p50 subunits.

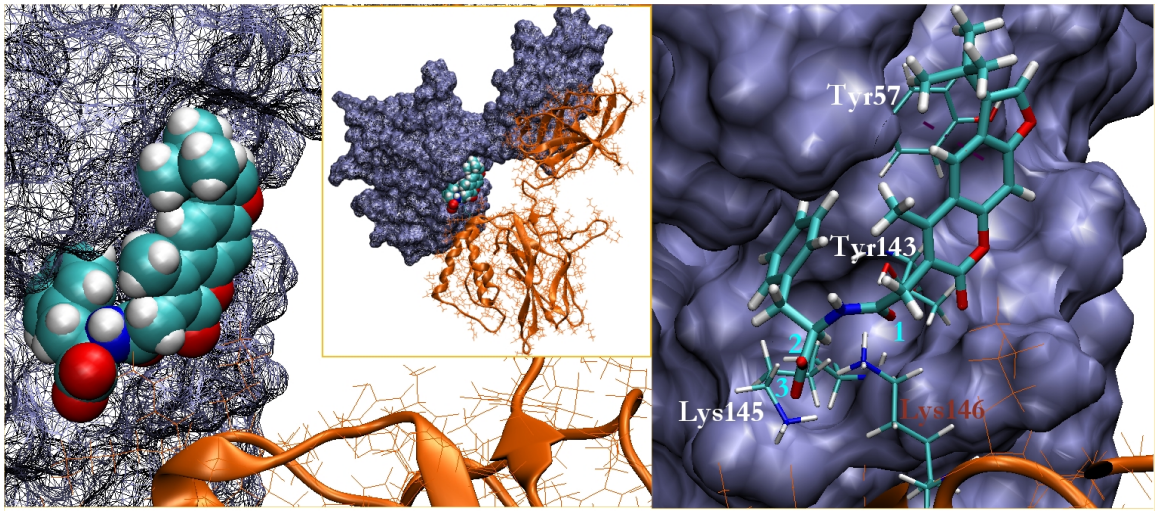


Figure 2 Stereoview of the complex formed by NF-kB p50 homodimer and the docked compound **1f**. The p50 chain B surface (purple coloured) is highlighted. Subunit p50 chain A is in NewCartoon representation with the sidechains of the residues highlighted in line drawing method. A focus of the **1f** docking pose in the DNA-binding region is shown (left and right). On the left the**1f** ligand is displayed in VDW style. On the right the key amminoacides are explicitly shown and the atoms (O.1, O.2, O.3) of compound **1f** involved in hydrogen bonds are reported. For details on protein/ligand interactions and relative distances see Table 3.

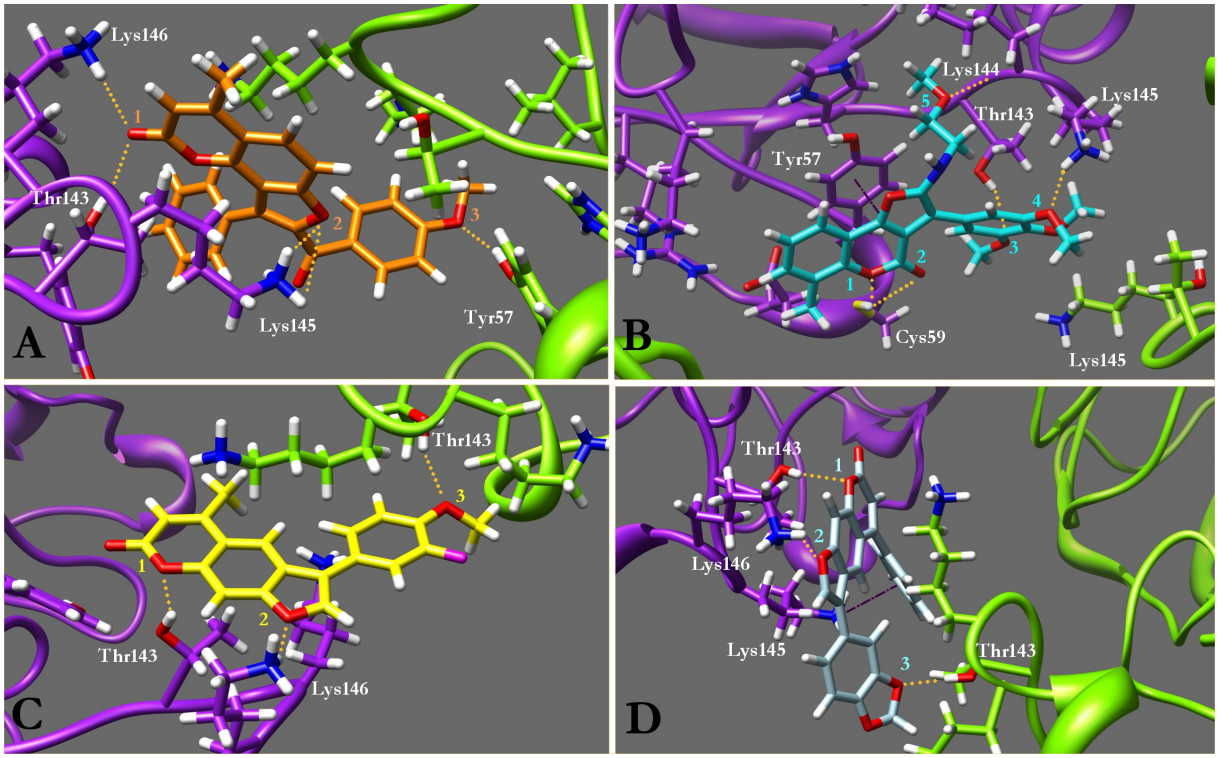


Figure 3 Close-up view of compounds **6f** (panel A, orange coloured), **7f** (panel B, cyan coloured), **9f** (panel C, yellow coloured) and **10f** (panel D, light blue coloured) binding modes into NF-kB p50 homodimer. NF-kB p50 subunits (chain A, green coloured; chain B purple coloured) are in ribbons representation. The hydrogen bonds (yellow lines), π- π interactions (violet lines) and numerations of ligand key atoms are shown. For details on protein/ligand interactions distances see Table 3.

Residue interaction		Ligand atom/s		Distance (Å)
(Cys59)(B)	-SH::O	O.1	**(7f)**	2.54
		O.2	**(7f)**	2.96
(Thr143)(B)	-OH::O	O.1	**(1f)**	2.06
		O.3	**(7f)**	2.52
		O.1	**(6f)**	2.97
		O.1	**(9f)**	2.01
		O.1	**(10f)**	2.20
(Lys145)(A)	- N^+-H::O	O.3	**(1f)**	2.73
(Lys145)(B)	-N^+-H::O	O.2	**(6f)**	2.85
		O.2	**(6f)**	2.02
		O.4	**(7f)**	2.14
(Lys145)(B)	-N^+	Aromatic moiety	**(10f)**	3.35
(Lys144)(B)	-NH::O	O.5	**(7f)**	2.09
(Tyr57)(B)	-Ph-centroid centroid	Aromatic moiety	**(1f)**	3.93
		Aromatic moiety	**(7f)**	3.57
(Tyr57)(A)	-OH::O	O.3	**(6f)**	1.80
(Lys146)(B)	-N^+-H::O	O.2	**(1f)**	2.57
		O.1	**(6f)**	2.9
		O.2	**(9f)**	1.85
		O.2	**(10f)**	2.0
(Thr143)(A)	-OH::O	O.3	**(9f)**	2.41
		O.3	**(10f)**	2.15

Table 3. Hydrogen bonds and π-π interactions of the best-ranked (**1f**) and active compounds (**6f,7f,9f,10f)** with the involved residues of the DNA binding region of NF-kB p50 homodimer (see Figure XX for ligand atom labels)

5.4 Conclusions

The VS method we used on NF-κB p50 homodimer lead to the individuation of five interesting furocumarin compounds able to inhibit the NF-kB/DNA complex formation. The forward filtering strategy based on the chemical motifs showed to be useful for the target under study. We showed that interesting motifs can be used to focusing prepared the initial database for a subset of compounds to be docked into NF-kB.

We have demonstrated that docking simulations are suitable for predictive studies on binding affinity of small compounds to a difficult target as NF-κB p50 homodimer. With these results we think to have laid the foundations for the development of further successful VS applications against NF-κB and computational lead optimizations studies.

Our starting docking experiences on NF-κB p65-p50 heterodimer shows the necessity of a more in depth computational studies in order to set up a suitable docking protocol.

References

1. Dall'Acqua F, Vedaldi D, Bordin F, Baccichetti F, Carlassare F, Tamaro M, Rodighiero P, Pastorini G, Guiotto A, Recchia G, Cristofolini: 4'-Methylangelicins: new potential agents for the photochemotherapy of psoriasis. J Med Chem **1983**, 26, 870.

2. Conconi MT, Montesi F, Parnigotto PP: Antiproliferative activity and phototoxicity of some methyl derivatives of 5-methoxypsoralen and 5-methoxyangelicin. Pharmacol Toxicol **1998**, *82*, 193.

3. Marzano C, Severin E, Pani B, Guiotto A, Bordin F: DNA damage and cytotoxicity induced in mammalian cells by a tetramethylfuroquinolinone derivative. Environ Mol Mutagen **1997**, *29*, 256.

4. Bordin F, Dall'Acqua F, Guiotto A: Angelicins, angular analogs of psoralens: chemistry, photochemical, photobiological and phototherapeutic properties. Pharmacol Ther Review **1991**, *52*, 331.

5. Sardari S, Mori Y, Horita K, Micetich RG, Nishibe S, Daneshtalab M: Synthesis and antifungal activity of coumarins and angular furanocoumarins. Bioorg Med Chem **1999**, *7*, 1933.

6. Backhouse CN, Delporte CL, Negrete RE, Erazo S, Zuniga A, Pinto A, Cassels BK: Active constituents isolated from Psoralea glandulosa L. with antiinflammatory and antipyretic activities. J. Ethnopharmacol **2001**, *78*, 27.

7. Lampronti I., Bianchi N Borgatti M., Fibach E.,Prus E., Gambari R. Eur J Haematology **2003**, *71*, 189.

8. Viola G, Vedaldi D, Dall'Acqua F, Lampronti I, Bianchi N, Zuccato C, Borgatti M, Gambari G: Furocoumarins photolysis products induce differentiation of human erythroid cells. J Photochem Photobiol B. Epub **2008**, *92*, 24.

9. Miolo G, Moro S, Vedaldi D, Caffieri S, Guiotto A, Dall'Acqua F: Derivatives as furocoumarin analogs: DNA-interactions and molecular modeling studies. Il Farmaco**1999**, *54*, 551.

10. Sigma-Aldrich, http://www.sigmaaldrich.com/chemistry/chemistry-catalog.mem.html

11. Maybridge, http://www.maybridge.com

12. Otava, http://www.otavachemicals.com/

13. TimTec, http://www.timtec.net

14. Chembridge, http://www.chembridge.com

15. DrugBank, http://www.drugbank.ca/

16. Jorgensen WL, Maxwell DS; Tirado-Rives J: Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc **1996**, *118*, 11225.

17. Maestro (v7.0.113) – A unified interface for all Schrodinger products, developed and marketed by Schrodinger, LLC. NY, Copy-right 2005; http://www.schrodinger.com.

18. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL:. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem **2004**, *47*, 1750.

CHAPTER 6. Molecular Dynamic simulation of NF-κB p50-p65 systems

6.1 Introduction

NF-κB p50-p65 is one of the most avidly NF-κB transcription factor forming dimers and is the major Rel/NF-κB complex in most cells. The X-ray crystallographic structure of RH domains of NF-κB p50-p65 on DNA or bound to IkB are known.[1,2] Nevertheless, these studies provide rather static architectures of these biological systems. Several molecular and biochemical studies indicate that Rel dimers assume distinct conformations when bound to kB DNA sites versus as free or IkB-bound. Thus, it should be of great interest to study the dynamic nature of the heterodimer in all the biological systems in which it can state. This could address for the dynamics of the pathway in response to various signals. Moreover, these informations could help virtual screening applications and strategies for drug discovery process with the individuation of representative protein conformations states and more accurate definition of putative small molecules NF-κB binding area.

There is surprisingly little information about the dynamic behaviour of the NF-κB systems. A molecular modelling study[3] has been reported on the characterization of a NF-κB model of the protein in its oxidized form in which was constructed an inter-subunit disulfide bond between the Cys62 residues of each monomer. MD simulation was used just for refinement of the model.

In this last chapter a structural analysis of NF-κB p50-p65 in free form and bound to both DNA and IκBα is presented. The analysis is based on a 14 ns MD simulations of all molecular systems under study immersed in explicit water molecules.

6.2 Methods

6.2.1 Molecular system set up.

Our starting models were derived from the solved X-ray structures previously described (chapter 1, figures XX and YY) and selected from the RCSB PDB. NF-κB p50-p65 in complex with DNA (PDB entry: 1LE9) and NF-κB p50-p65 in complex with IκBα (PDB entry: 1IKN) were first retrieved from PDB. On of the molecular systems for further dynamic simulation was the structure of protein in its active conformation in complex with DNA as get from PDB (1LE9). The NF-κB p50-p65 in its open conformation was obtained by removing the DNA from NF-kB/DNA complex (PDB entry: 1LE9).

The NF-kB/IκBα system was formed by the protein in its open conformation, as obtained above, and the IκBα inhibitory protein which was extracted from the crystallographic structure 1IKN. In order to build the NF-kB/IκBα model (Figure 1), p65 and p50 dimerization domains (C-terminal) of the protein in complex with IκBα (1IKN) were superimposed on the same domains of the protein in active conformation using Maestro graphical interface.[4] Subsequently, NF-κB in closed conformation was removed *in silico*. The missing loop residues from 66 to 100 were rebuilt by Swiss PDB viewer.[5]

Finally, VMD software[6] was used in order to prepare the molecular models for the subsequent molecular dynamics simulations. To this aim, each molecular system was located in a periodic box that allowed 5Å minimum distance between the macromolecular structures and the walls of the contiguous cells, and solvated with water molecules. Moreover Na^+ Cl^- were added to the solution until 0.5M. The resulting number of atoms for NF-κB, NF-κB /DNA and NF-κB /IκBα assemblies was of 119.980, 119.772, and 150.000 respectively.

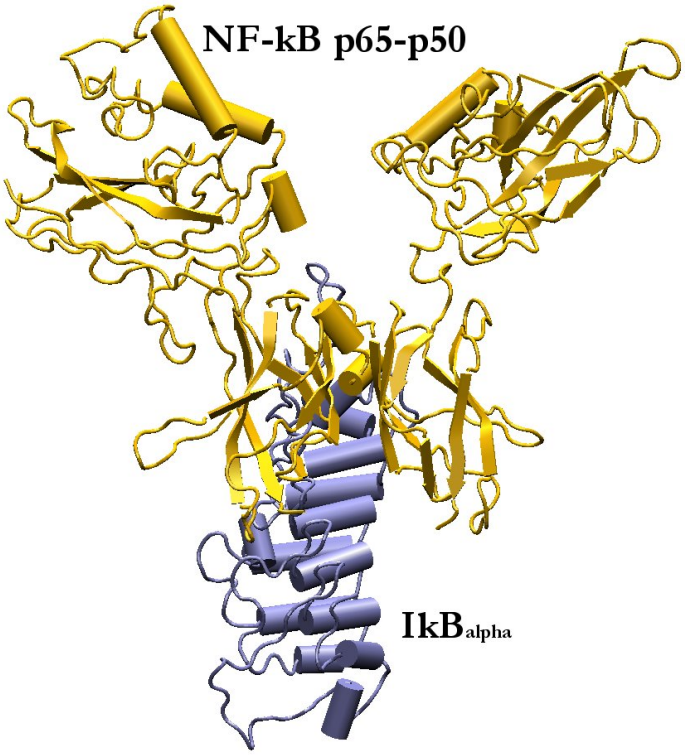


Figure 1 Three-dimensional structure of NF-kB/IκBα system in cartoon representation

6.2.2 Molecular Dynamic simulations

The NAMD[7] software package was used on the 1024 processors cluster linux IBM of at CINECA University Consortium (http://www.cineca.it/) for the main simulations.

The Charm27[8] force field was used for the description and parametrization of each molecular system. The TIP3P water model[9] was employed and periodic boundary conditions were applied. A residue-based cutoff of 10Å was used for the evaluation of the Van der Waals interactions. Electrostatic interactions were computed using the Ewald particle mesh method[10] with a dielectric constant set to 1.0. Bonds involving hydrogen atoms were constrained with the SHAKE algorithm[11] and the time step was set to 2fs. The constant temperature and volume simulations were achieved by coupling the systems with a thermostat at 300 K trough a weak coupling algorithm. The 14 ns MD productions run in the canonical (NVT) ensemble were collected for further analysis.

6.2.3 Calculated properties

The software package GROMACS 3.2.1[12-14] running on a local Linux Xeon 64bit PC was used for the properties calculation of NF-κB p50-p65 protein for each simulated molecular system under study. The GROMACS is GNU public domain and free and it presents a robust set of programs available for analysis of the trajectories.
The calculated properties are listed below.

(i) root mean square deviation (RMSD) of protein α-carbons with respect to the starting X-ray structure (1LE9) was calculated with program g_rms by least-square fitting the NF-κB to reference structure and subsequently was calculated the RMSD.

$$RMSD(t_1,t_2)=\left[\frac{1}{M}\sum_{i=1}^{N}m_i\left\|r_i(t_1)-r_i(t_2)\right\|^2\right]^{\frac{1}{2}} \tag{6.1}$$

Where $M=\sum_{i=1}^{N}m_i$ and $r_i(t)$ is the position of atom *i* at time *t*.

(ii) the root mean square fluctuation (RMSF) of the atom around its average position during the molecular dynamics were obtained using g_rmsf program. The RMSF values were converted to B-factor values, which are written to a pdb file with the coordinates of the structure file.

(iii) the secondary structure for each time frame was obtained calling the do_dssp program.

(iv) A first principal component analysis has been performed on the Cα 's trajectories using the g_covar program. The first frame (ps) to read from trajectory was at t=0. g_covar calculates and diagonalizes the (mass-weighted) covariance matrix. The diagonalization of the covariance matrix of the coordinate fluctuations led to the identification of the eigenvectors (direction of motion) with the largest eigenvalues (amount of motion). The eigenvectors were written to a trajectory file.

First the covariance matrix of the atomic coordinates (not mass wieghted) has been obtained from Cα' s trajectories through:

$$C_{ij} = \left\langle \left(x_i - \left\langle x_i \right\rangle\right)\left(x_j - \left\langle x_j \right\rangle\right) \right\rangle \tag{6.2}$$

where the averages are computed along the trajectory and the x_i and x_j are the coordinates of the Cα atoms. C is the symmetric $3N \times 3N$ matrix that has been diagonalized with an orthonormal transformation matrix.

Subsequently, the program g_anaeig was employed for analyzing eigenvectors of the covariance matrix. The analyses were performed on the first 3 eigenvectors. We calculated the full precision eigenvectors trajectory and calculated the two extreme projections along the trajectory on the average structure and interpolate 5 frames between them. We collected the first three eigenvectors in separate pdb files.

Finally, the dynamite software version v1.5 from Dynamite web site[15] at University of Oxford was also used for the principal component analysis of the NF-κB in its free dimeric form with the purpose to deepen motion studies and visualization. Starting from NF-κB dynamic simulations trajectory, a new trajectory file was prepared taking one frame every 50. The new trajectory file and the initial protein structure at t=0 were submitted to the Dynamite University of Oxford web site.

(v) A cluster analysis for NF-κB in its free form has been carried out using the g_cluster program. Distances between structures have been determined from the trajectory. RMS deviation after fitting of all selected residues atoms (except hydrogen atoms) has been used to define the distance between structures. For the selection, all the amino acids that form hydrogen bonding with DNA in the starting crystallographic structure were chosen (model numeration: p65 subunit: Arg15, Arg17, and Glu21, Cys20, Lys104, Tyr18, Lys105, Gln202, Gln229, Arg228 and Lys203, Ala25, Ser24, Ser27 and Lys38; p50 subunit: Arg286-Arg289-Glu295 Arg289 and Glu295 Cys294, Lys379, Tyr292, Gln509, Gln541, and Arg540,Lys380, Lys507, Ser298, Gly300, Gly301, Asn371 and Lys312; note Arg15 correspond to p65 Arg33 of 1LE9 and Arg286 to p50 *Arg51*)

The linkage method for clustering has been adopted using a RMSD cutoff of 0.12 nm. The most representative protein conformations were retrieved from the most populated clusters of structures.

6.3 Results and discussion

6.3.1 Structural properties of NF-kB

The overall structure of NF-κB p50-p65 heterodimer for each system appeared to be equilibrated after ≈ 7.0 ns, as deciphered by a plot of the rms deviations (Figure 2). In the trimeric forms of the protein, backbone net displacement as measured by α-carbon RMSD already quenched after 1.5 ns (NF-kB/IκBα) and after 3ns (NF-kB/DNA) to reach a plateau after 6 ns. NF-κB showed an asymptotic RMSD of 0.9, 0.55 (NF-kB/IκBα) and 0.40 (NF-kB/DNA). This underlines the higher flexibility of the NF-κB dimer with respect to the trimeric complexes.

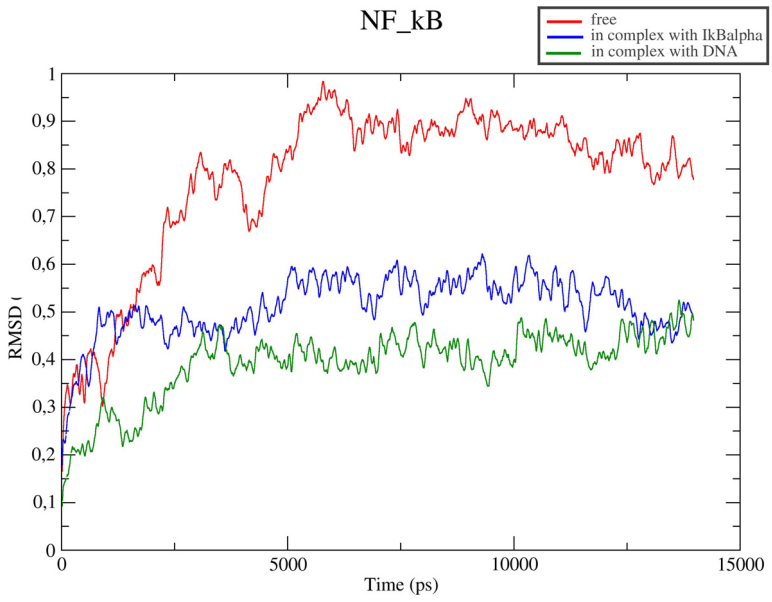


Figure 2 Cα RMSD from the initial conformation computed for structures sampled along the overall trajectory of 14ns. A plateau is reached after 6 ns for all MD simulations

Besides to the RMSD, the RMSF calculated over the Cα positions of each NF-κB p50-p65 residue was also analyzed (Figure 3). As expected, the RMSF values were highest in the free dimeric state of the protein ranging from about 0.15 to 1.0 nm followed by the trimeric states NF-κB /IκBα (values from 0.15 to 0.9 nm) and NF-kB/DNA (values from 0.15 to 0.5 nm). The differences between the protein in its free state and in complex with inhibitory protein are not so evident limiting the comparison just to the RMSF values range. The average RMSF value was significantly different as shown by visual inspection of Figure 3. Indeed, in correspondence to residues Asp35 and Thr36 of NF-κB in this trimeric form, a significant RMSF pick (0.9 nm) was observed. In any case, the NF-κB N-terminal domain, expecially in the p65 subunit, was the most mobile in all the simulations, while the short residues' linker loop was surprising quite stable. In particular, the residues from Lys104 to Arg153 featured the highest RMSF region, just before the linker in the p65 N-terminal domain of the protein in its dimeric form. Among the residues involved in hydrogen bondings, with DNA, only Lys104 and Lys105 are particularly flexible with a RMSF of 0.70 nm and 0.80 nm. The same residues were quite more stable in NF-κB complexed states.

In agreement with the dynamic simulation results, the higher stability of C-terminal domains accountc for the dimerization process (protein-protein interactions) and heterodimer formation into cellular environment.

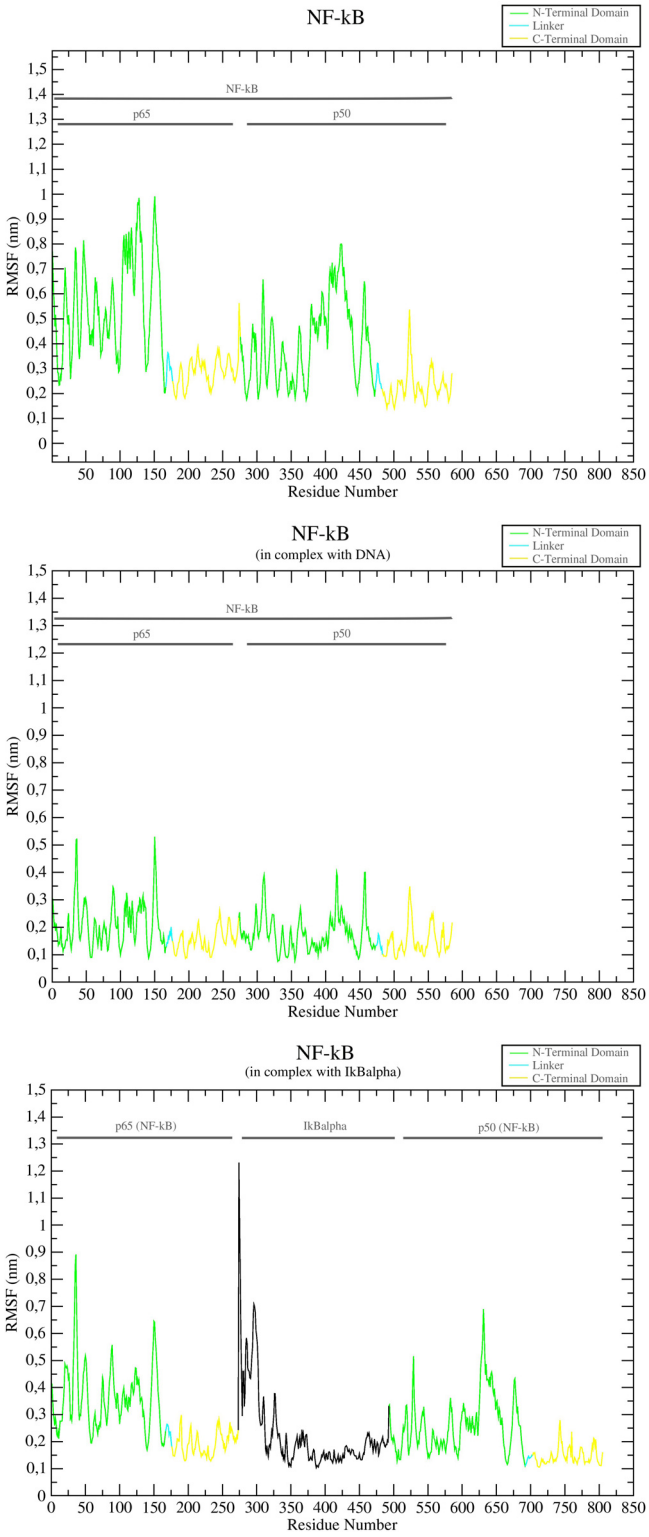
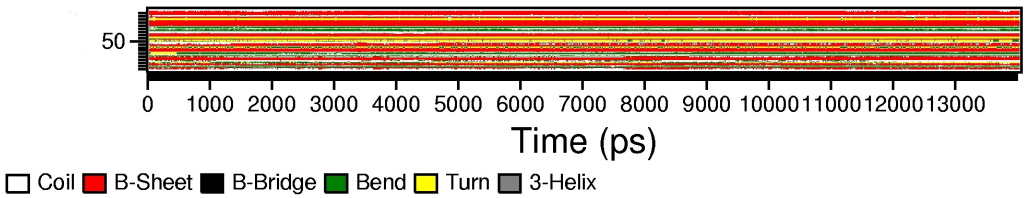


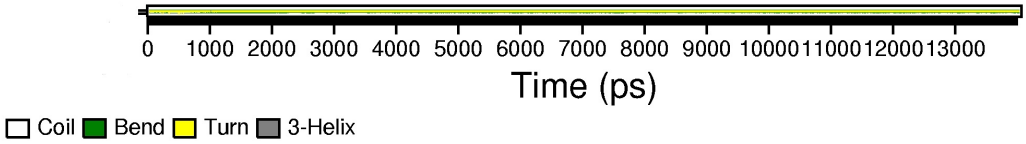
Figure 3 Structural fluctuations as a function of residue position for NF-κB heterodimer in all the systems under study. Data are averaged over all the 14 ns of MD. Both p50 and p65 subunits are shown. The N- and C-terminal domains and the ten residue's linker for each subunit are highlighted.

The secondary structure elements of the protein are well conserved during the overall dynamics for each system considered. The same behaviour is present for residues (Lys104 - Arg153) belonging to more flexible p65 region in which the α-helixe, the β-bridges and bends residues are entirely conserved (Figure 4).

(a)



(b)



(c)

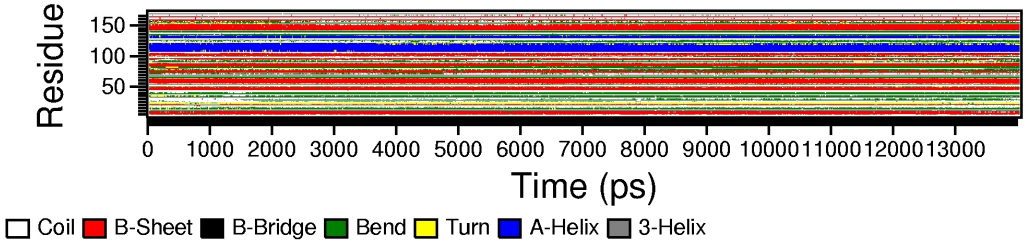


Figure 4. Plot of the time-dipendent secondary structure of NF-κB p65 subunit reported for each residues of the protein. For clarity **(a)** N-terminal domain **(b)** 10 redidues' linker **(c)** C-terminal domain are visualized separately.

6.3.2 Essential dynamics analysis

Essential dynamics (ED) analysis[16] on NF-κB was performed both on the equilibration part of (from 0 to 7 ns) and on the production part (from 7ns to 14ns) of the MD trajectory.
The long-time fluctuations can be probed by the essential modes (eigenvectors) associated to the covariance matrix (see section *Calculated Properties*). The total amount of motion was higher for NF-κB in its dimeric ensemble than its trimeric states. Interestingly, the associated eigenvalues decreased in the last part of MD trajectory while in the bounded form of the protein these values increased weakly.
As shown in Figure 5 the first eigenvector account for more than half of the fluctuations of the NF-κB both in dimeric system and in complex with DNA during the MD simulation. At the same way, the second eigenvector describe more than half of motion for NF-κB in complex with IκBα. During the production phase the results are essentially maintained. For NF-κB in its dimeric state the addition of the second eigenvector is necessary for describing more than half of motion. This allows representing a large part of the complex motions of each system just through few coordinates, i.e. the projections of the trajectory along the first or the second eigenvector (Figure 6).

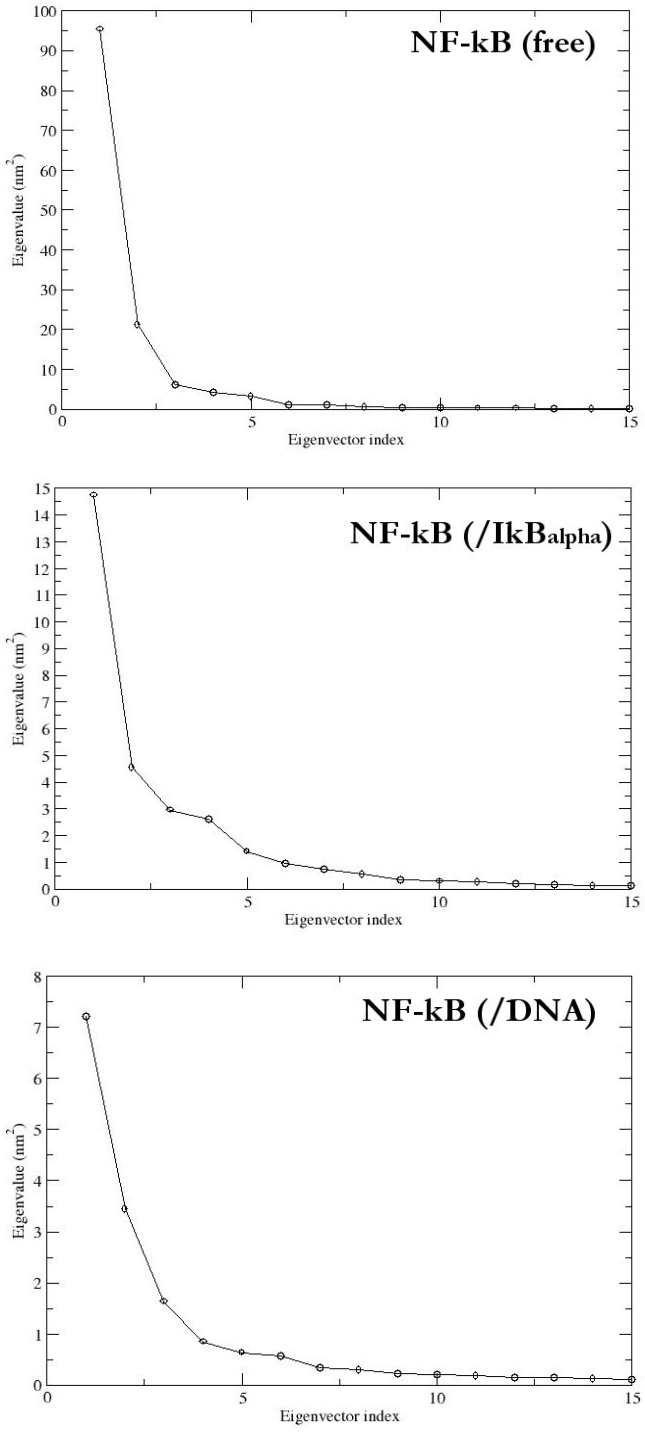


Figure 5. Plots of the eigenvalues of the covariance matrix for each NF-kB system**.**

Considering the first eigenvector for the protein in its "butterfly" dimeric state and in complex with the inhibitory protein, both the N-terminal domains showed a concerted motion along opposite directions that leads to wing span of the protein structure in the first 7ns of MD simulation. The p65 domain mainly contributes to this displacement. In the case of trimeric system, the p50 domain presents a more evident rotation motion along its long axis and out-of-plane formed by both domains towards a closer conformation. The first eigenvector for NF-κB in complex with DNA only accounts for a very small rotation of the N-terminal domains that does not lead to any modification in the distance between the p50 and p65 subunit. In all the cases, the dimerization terminal domains are quite stable in their initial positions. The second eigenvector for NF-κB in complex with the inhibitory protein describe a turning movement responsible for the observed opening of N-terminal domains. The combination of these essential motions leads protein chains to get closer to an inactive conformation moving away the residues involved in κB DNA interactions.

In the stabilization phase of the MD simulation (from 7 to 14ns) regarding the most dynamic NF-κB system, the first eigenvector represent a weak motion of the N-terminal domain in opposite direction with respect to the equilibration part of MD. This behaviour is not observed in the transcription factor when it is complexed with its inhibitor or with DNA.

We further carried out the analysis for NF-κB in its dimeric form by exploiting the program dynamite version v1.5. We displayed the eigenvectors as a vector field on the structure of the complex. The vector on each Cα atom represents the component of the motion of the eigenvector on that residue (Figure 7). This allows obtaining a visual rationalization of the meaning of the eigenvector. The eigenvectors, obtained from the diagonalization of the covariance matrix, with the largest eigenvalue (principal eigenvectors) represent collective motions of the residues. The amino acids communally move along defined directions or according to overall rotations.

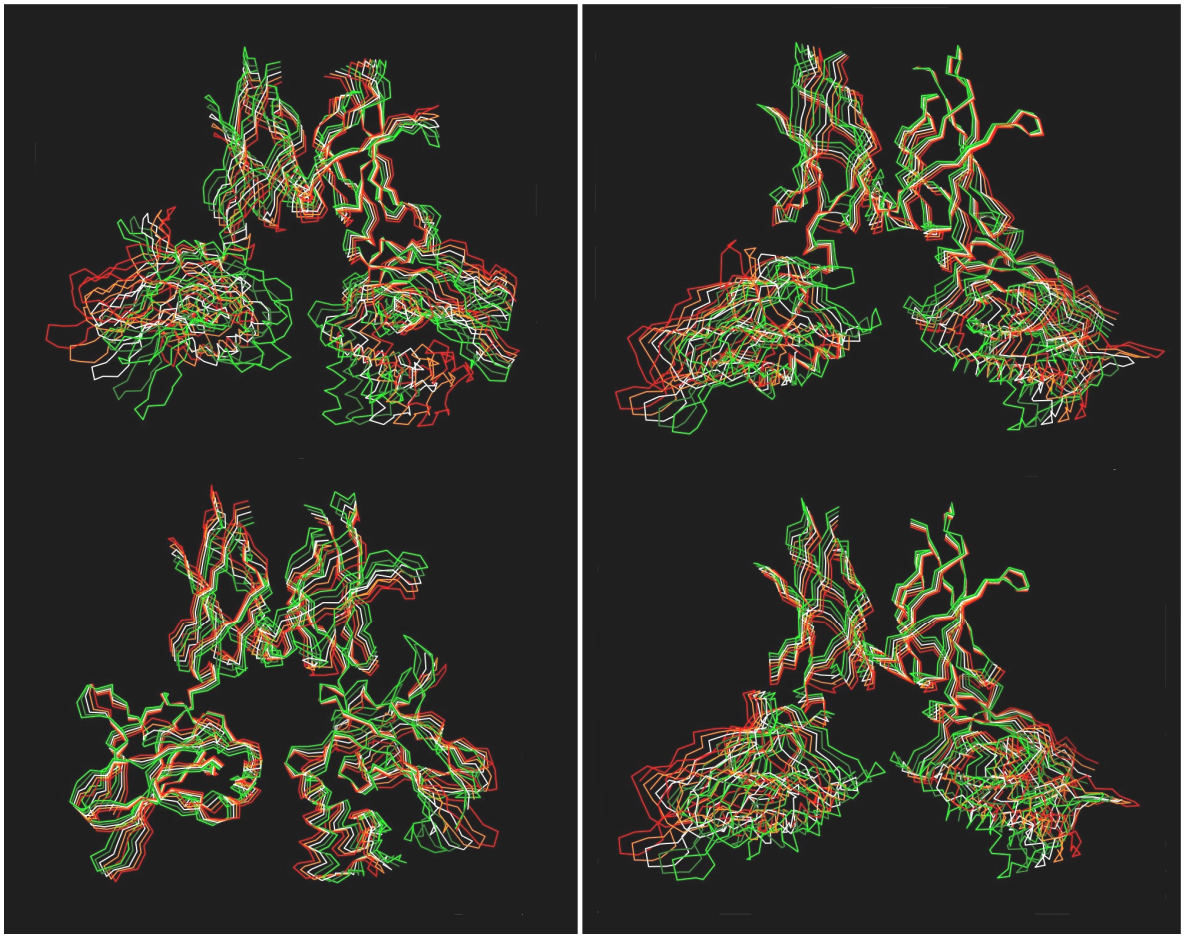


Figure 6: The first eigenvectors of the covariance matrix for NF-κB in its free state (left) and in complex with IκBα (right). On the left the five frames referred to the equilibration part of MD trajectory (from 0 to 7 ns) and on the right on the last part (7-14 ns) All the five frames are highlighted (the first one is coloured in green and the last one in red). For all NF-kB representations its p65 subunits is on the left part.

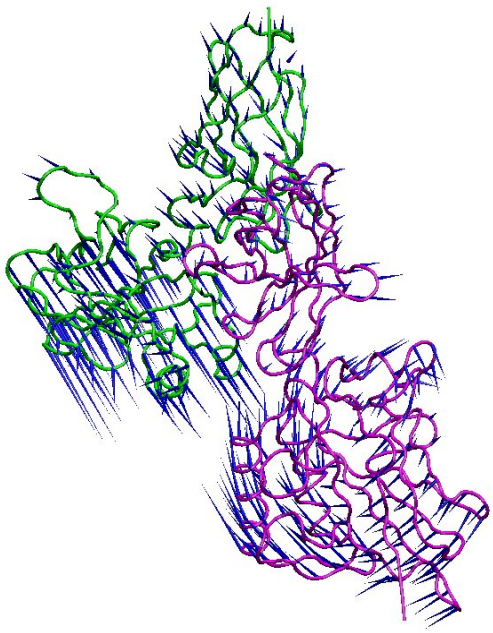


Figure 7. The first eigenvector of the covariance matrix. It is represented as vector field on the Cα atoms of the average structure of NF-κB (p65 subunit in green). The blue-ended stick departing from every Cα atom is proportional (both in length and direction) to the component of the eigenvector along that atom. This picture referred to the essential dynamics analysis on overall 14ns of MD by using software dynamite (see method section).

From the analysis of the trajectory, we could suppose that the interactions between NF-κB in heterodimeric form and the DNA is the driving force for the compaction and stabilization of the open 'butterfly' conformation of the protein. As shown in the solved crystallographic NF-kB/IκBα complex[1] the p65 N-terminal domain is rotated of almost 180° about its long axis and translates toward its C-terminal dimerization domain. In this case the NF-κB presents a 'butterfly' closed inactive conformation (see chapter 2).

The NF-κB starting reorganization towards this closed inactive conformation was more evident from dynamic studies on the NF-κB /IκBα molecular system. In fact, even if the NF-κB in its free state was more flexible and presented an higher total amount of motion, the direction of the essential motion changed towards a closed "scissors" configuration when the dynamic simulation switched from the equilibration phase to the more stable one (Figures 6 and 8). Probably this is due to the fact that the starting NF-κB model has been retrieved from a crystallographic complex of the protein with the DNA, in which the adopted conformation

may be the result of a DNA induced fit upon binding. Thus, the MD production phase should be more indicative of protein motions.

6.3.3 Most representative NF-κB p50-p65 conformations

The selection of NF-κB p50-p65 representative conformations derived from the 14ns MD simulations of the protein in its dimeric state. Starting from 1400 number of structures for each matrix, 152 clusters have been obtained. The RMSD ranges were from 0.074 to 1.370 nm with an average RMSD of 0.490 nm. In Table 1 are reported the results for the first five most populated clusters. In Figures 8 the selected NF-κB five structures are shown.

Cluster	Number of structures	Middle structure	RMSD (nm)
134	99	6110	0.165
143	65	7830	0.152
145	47	8430	0.141
148	446	12500	0.195
151	41	13710	0.144

Table 1.First five most populated clusters.

A new dimerization interface between the N-terminal domains of both p65 and p60 subunits is observed in all the most representative conformations, all collected from the production phase. This structural feature significantly distinguishes these conformations from the starting crystallographic NF-κB configuration. The NF-κB dimerization interface, involving the α-helixes of p50 and p65 subunits, exhibits a network of hydrophobic interactions. Moreover hydrogen bonds and electrostatic interactions may contribute to the overall stability of NF-κB p50-p65 conformations (Figure 9). It is obvious that in such blocked configuration the DNA binding is hindered.

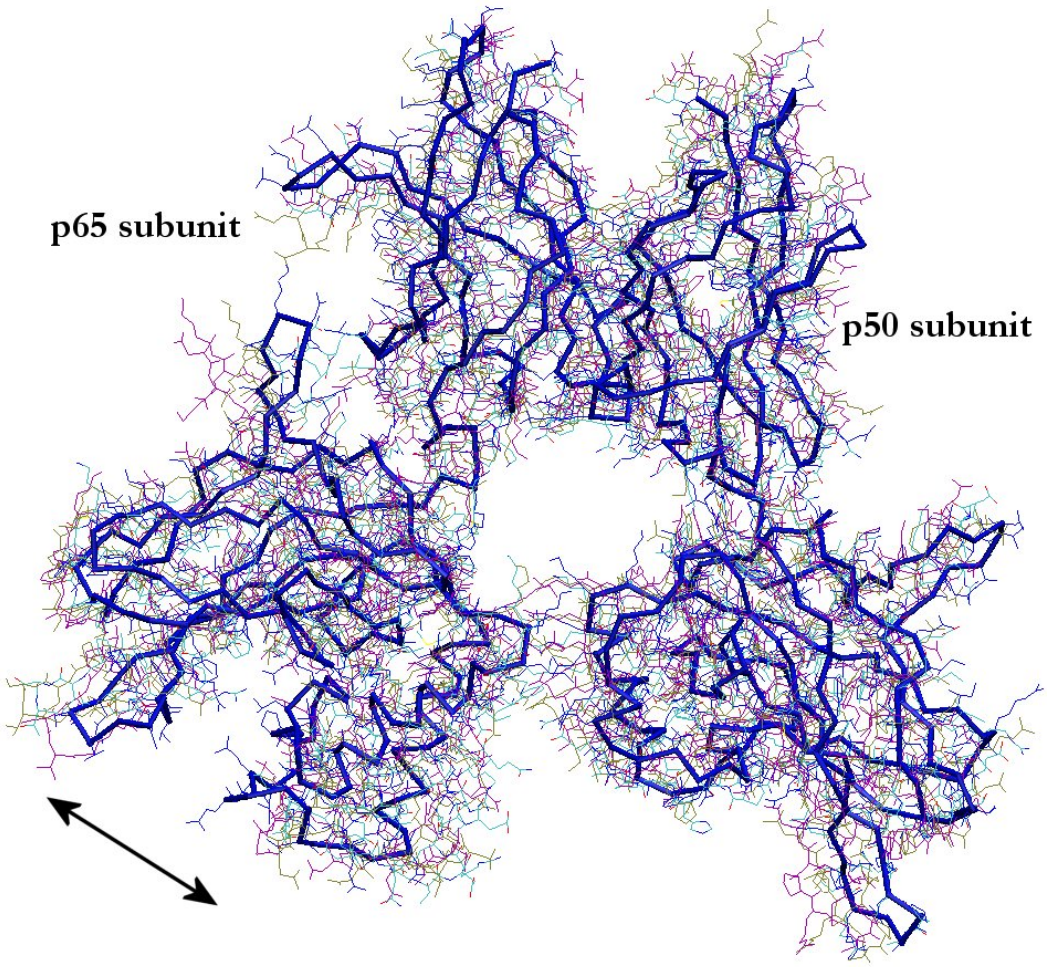


Figure 8 Superimposition between the five average structures of the most populated conformational clusters. The most relevant conformation is highlighted in blue. The arrow indicates the pocket formed between the β-sandwich core folding and the α-helixes secondary structures of the p65

This may suggests that NF-κB p50-p65 when localized into the nucleus needs to interact with an activator before binding. Another interesting finding is the volume increase of the pocket formed between the β-sandwich core folding and the α-helix secondary structure motif of the p65 N-terminal domain due to the more evident separation between these structural elements. We could intriguing suggest that small molecules able to fit into this pocket strongly interacting with targeting residues, may stabilize the closed "scissors" conformation no more able to accommodate the κB DNA.

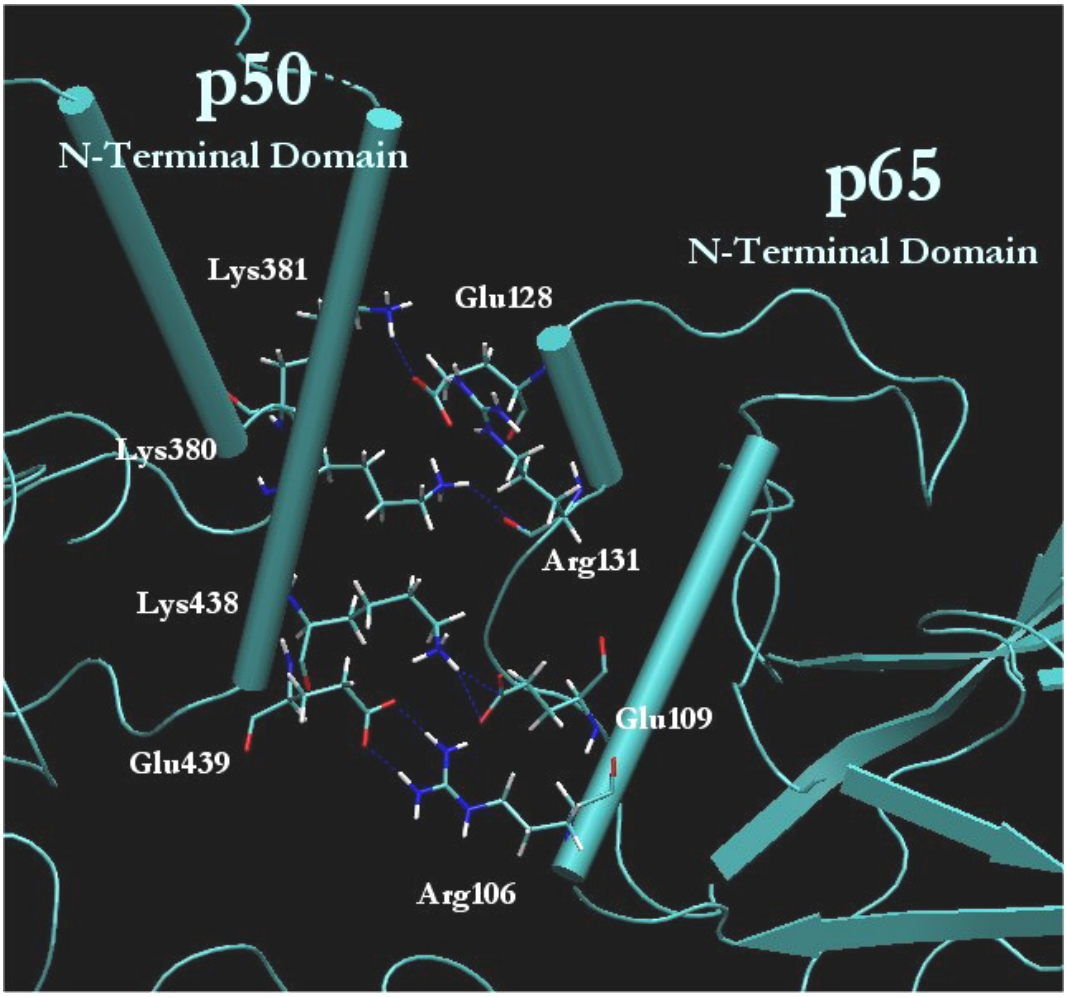


Figure 9 The NF-kB most relevant conformation. The dimer interface between the p50 and p65 α-helixes are shown

6.4 Conclusions

In chapter 6 we have carried out an analysis of the dynamical characteristics of NF-κB p50-p65 heterodimer as free *versus* DNA-bound and IκBα-bound systems, whose binding inhibition is involved in the drug discovery approach followed in this thesis. The flexibility of each domain of p50 and p65 subunits has been detailed.

The analysis of the dynamical characteristics of each system, simulated through a 14ns MD runs in explicit solvent, has been complemented by a study of the collective motions that characterized transcription factor essential dynamic behaviours. We have found relevant differences among each analyzed system, such as the directions and the amount of motions. The amino acids relevant for the DNA binding have been used as a criterion for a cluster analysis of all conformations collected in the overall simulations. The most representative conformations have been selected for further in-depth analysis and drug design studies on NF-κB p50-p65 in its heterodimeric form.

References

1. Huxford T, Huang DB, Malek S, Ghosh G: The crystal structure of the IkappaBalpha/NF-kappaB complex reveals mechanisms of NF-kappaB inactivation. Cell **1998**, *95*, 759.

2. Chen FE, Huang DB, Chen YQ, Ghosh G: Crystal structure of p50/p65 heterodimer of transcription factor NF- kappaB bound to DNA. Nature **1998**, *391*, 410.

3. Chandrasekaran V, Taylor EW: Molecular modeling of the oxidized form of nuclear factor-kappa B suggests a mechanism for redox regulation of DNA binding and transcriptional activation. J Mol Graph Model **2008**, *26*, 861.

4. Maestro (v7.0.113) – A unified interface for all Schrodinger products, developed and marketed by Schrodinger, LLC. NY, Copy-right 2005; http://www.schrodinger.com.

5. Swiss PDB Viewer http://www.expasy.org/spdbv/

6. VMD, Visual Molecular Dynamics. http://www.ks.uiuc.edu/Research/vmd/

7. NAMD, Scalable Molecular Dynamics. http://www.ks.uiuc.edu/Research/namd/

8. MacKerell AD, Banavali N, Foloppe N: Development and current status of the CHARMM force field for nucleic acids. Biopolymers **2001**, *56*, 257.

9. Jorgensen WL, Chandrasekhar J, Madura JD: Comparison of simple potential functions for simulating liquid water J Chem Phys **1983**, *79*, 926.

10. Darden T, York D, Pedersen L: Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys **1993**, *98*, 10089.

11. Ryckaert JP, Ciccotti G,. Berendsen HJC: Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes. J Comp Phys **1977**, *23*, 327.

12. Van der Spoel D, Lindahl E, Hess B, van Buuren AR, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM,. Feenstra KA, van Drunen R,. Berendsen HJC: GROMACS user manual version 3.3. Department of

Biophysical Chemistry, University of Groningen. Nijenborgh 4, 9747 AG Groningen, The Netherlands, 2006.

13. Lindahl E, Hess B, and van der Spoel D: GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model **2001**, *7*, 306.

14. Berendsen HJC, van der Spoel D, van Drunen R: GROMACS: A message-passing parallel molecular dynamics implementation. Comp Phys Comm **1995**, *91*, 43.

15. Barrett CP, Hall BA, Noble MEM: Dynamite: A Simple Way to Gain Insight into Protein Motions Acta Crystallographica **2004**, *60*, 2280.

16. Amadei A,. Linssen AB, Berendsen HJ: Essential dynamics of proteins. Proteins **1993**, *17*, 412.