UNIVERSITÀ DEGLI STUDI DI FIRENZE

Dipartimento di Scienze Farmaceutiche

SCUOLA DI DOTTORATO IN SCIENZE

DOTTORATO DI RICERCA IN

"Chimica e Tecnologia del Farmaco"
XXIII CICLO 2008-2010

Applicazioni di metodiche di molecular modeling allo studio ed ottimizzazione delle modalità di interazione tra macromolecole di interesse biologico e ligandi

Tesi di Dottorato di
Matteo Chioccioli

Docente supervisore:
Prof. Paola Gratteri

Coordinatore:
Prof. Elisabetta Teodori

Settore scientifico disciplinare CHIM/08

Firenze, Dicembre 2010

Alla zia Gabriella

e alla mia Famiglia

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND SYMBOLS

ADP	Adenosine Diphosphate
A-loop	Activation loop
ATP	Adenosine Triphosphate
BRACO-19	9-(4-(N,N-dimethylamino)phenylamino)-3,6-bis(3-pyrrolodinopropionamido) acridine
CPU	Central Processing Unit
3D	Three-dimensional
DNA	Deoxyribonucleic Acid
dsDNA	double-stranded DNA
EL	Extended Lagrangian
FDA	Food and Drug Administration
FF	Force Field
FLK	Fetal Liver Kinase
GA	Glucoamylase
G-quadruplex	Guanine-quadruplex
Ig	Immunoglobulin
InsR	Insulin Receptor
KDR	Kinase insert Domain-containing Receptor
KID	Kinase Insert Domain
LBDD	Ligand Based Drug Design
***m*-AMSA**	Amsacrine
MD	Molecular Dynamics
MM-PBSA	Molecular Mechanics-Poisson Boltzmann Surface Area
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
PARSE	Parameters for Solvation Energy
PBC	Periodic Boundary Conditions
PBE	Poisson Boltzmann equation
PDB	Protein Data Bank
PIGF	Placental Growth Factor

PME	Particle Mesh Ewald
PTB	Phospho-Tyrosine Binding
REM	Replica Exchange Method
r-RESPA	Reversible Reference System Propagation Algorithm
RMSD	Root Mean Square Deviation
RTK	Receptor Tyrosine Kinase
SASA	Solvent-Accessible Surface Area
SBDD	Structure Based Drug Design
SH2	Src Homology 2
SPME	Smooth Particle Mesh Ewald
Tyr(P)	Phosphorylated tyrosine
VEGF	Vascular Endothelial Growth Factor
VEGFR	Vascular Endothelial Growth Factor Receptor

1. Introduction

1.1 MOLECULAR MODELING IN MEDICINAL CHEMISTRY

One of the main goals of medicinal chemistry is the *rational design* of new compounds able to produce a biological effect. This objective is now increasingly achieved through the use of molecular modeling methodologies which allow to study *in silico* the atomic details of the interaction between ligands and biological targets and to simulate their dynamic behaviour using the equations of classical and quantum physics. This information can be successfully applied to the evaluation of new molecules in the design stage, with the aim of directing the synthesis only on the most promising compounds. Moreover, molecular modeling methods can be applied to the investigation of biophysical phenomena, such as the conformational flexibility of macromolecular targets, at atomistic level.

One of the basic principles of medicinal chemistry is that biological activity of a compound is dependent on the three-dimensional placement of specific functional groups which interacts with complementary regions in macromolecular targets (proteins or DNA molecules). The application and development of the molecular modeling methodologies in the pharmaceutical research has been mainly promoted by the increasing availability of the three-dimensional structures of biological targets solved at the atomic level through X-ray crystallography or NMR spectroscopy, as well as by the improvement in computer hardware and software.

The computational approaches employed in the drug discovery can be divided into two main classes:

- Structure Based Drug Design (SBDD), in which the starting point is the knowledge of the three-dimensional (3D) structures of biological targets, solved experimentally or built by homology modeling;

- Ligand Based Drug Design (LBDD), which uses information derived from the analysis of a series of molecules interacting with the same pharmacological profile to the same target to obtain a picture of those target features necessary for ligand interaction.

In this thesis work we deal with methodologies related to a SBDD approach, exploiting the availability of experimental three-dimensional structures for the targets

or complementing gaps in the available experimental data through the use of homology modeling. We have employed classical methods like *molecular docking* and *Molecular Dynamics simulations (MD)* and more recent methodologies which enable a more efficient exploration of the conformational space of complex macromolecules (*sampling methods*) and an accurate evaluation of the free energy ($\Delta G_{binding}$) related to the binding process between ligands and macromolecules (*free energy methods*).

Here, we will describe two applications of the aforementioned methodologies taken as an example. MD simulations and advanced sampling methods have been applied to the investigation of the tyrosine kinase domain belonging to the membrane receptor VEGFR2, while molecular docking, MD simulations and the MM-PBSA method for the evaluation of the binding free energy have been applied to the investigation of the behaviour of a series of acridine-based polyamine ligands in the binding of double helix and G-quadruplex DNA structures. In addition to these, in the appendix section of this thesis is reported a molecular docking study on the binding modes of newly synthesized iminosugars towards the enzyme glucoamylase (GA) from *Aspergillus awamori*.

1.2 RECEPTOR TYROSINE KINASES

1.2.1 Receptor Tyrosine Kinases: general features

Human Receptor Tyrosine Kinases (RTKs) represent a class of transmembrane glycoproteins involved in controlling the transduction of extracellular signals to the cytoplasm through the selective phosphorylation of tyrosine residues on proteins, often other kinases, or even themselves (autophosphorylation). [1] These receptors bind ATP-Mg^{2+} and catalyze the transfer of the γ-phosphoryl group from ATP to hydroxyls of tyrosine side chains in a protein substrate:

$$MgATP^{1-} + \text{protein-OH} \rightarrow \text{protein-OPO}_3^{2-} + MgADP + H^+$$
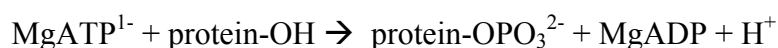
Several important targets in medicinal chemistry belong to this family of receptors including, among the others, the insulin receptor (InsR), involved in the stimulation of carbohydrate utilization, and the Vascular Endothelial Growth Factor Receptors (VEGFRs), involved in controlling physiological processes like vasculogenesis (the formation of new blood vessels) and angiogenesis (the growth of pre-existing blood vessels). The activation of these receptors is induced by the binding to their extracellular domain of some polypeptides acting as signal molecules, like the hormone insulin and the Vascular Endothelial Growth Factors (VEGFs), which include VEGF-A, VEGF-B, VEGF-C, VEGF-D and Placental Growth Factor (PIGF). [2] Over the past several years, the knowledge of the functionality of these receptors and of the structural bases for their regulation has been provided by the increasing availability of high-resolution crystal structures solved both in active and inactive states together with biochemical, genetic and modeling studies. [3] However, some fundamental aspects regarding these key events are not well understood and one of the most attractive challenges of molecular modeling is now the application of computational techniques in order to achieve a deeper knowledge of these targets and of their conformational flexibility starting from the available structural data.

Despite the great number of common features, some variability among RTKs can be observed in their overall architecture. Most of these receptors, like VEGFRs, are

monomer in their inactive forms with a single polypeptide chain passing through the membrane and undergo oligomerization after ligand binding. Other kinases consist of more than one polypeptide chain; for instance the insulin receptor is formed even in its inactive state by four subunits ($\alpha_2\beta_2$), two of them (β units) spanning the cellular membrane. [2]

1.2.2 Vascular Endothelial Growth Factor Receptors

In this thesis work the attention was focused on the investigation of the VEGFRs. These receptors have recently emerged as attractive targets in pharmaceutical and clinical research because the processes of angiogenesis and vasculogenesis in which they are involved, have been recognized as key steps in several diseases like tumor growth and spread, diabetic retinopathy and arthritis [4-7]. At the moment three different human VEGFRs are known (VEGFR1, VEGFR2 and VEGFR3), located on the surface of vascular endothelial cells and formed by a single polypeptide chain spanning the cellular membrane. [8]

Human VEGFR2 (also known as **FLK-1**, fetal liver kinase-1) is a protein constituted by 1356 amino acids and seems to be the major transducer for signals induced by VEGFs, both in physiological and pathological situations, and can be chosen as representative member of this class of proteins. [9]

Structurally it can be divided into different regions (Fig.1.2.1):
- an extracellular domain containing seven immunoglobulin (Ig)-like domains and involved in the binding of growth factors (VEGFs);
- a single short transmembrane segment;
- a juxtamembrane region;
- a cytoplasmic domain possessing tyrosine kinase catalytic activity, referred to as *KDR domain*, which is split into two portions by the so called kinase insert domain (KID). The catalytic domain is the most conserved region among RTKs;
- a carboxyl terminal domain.
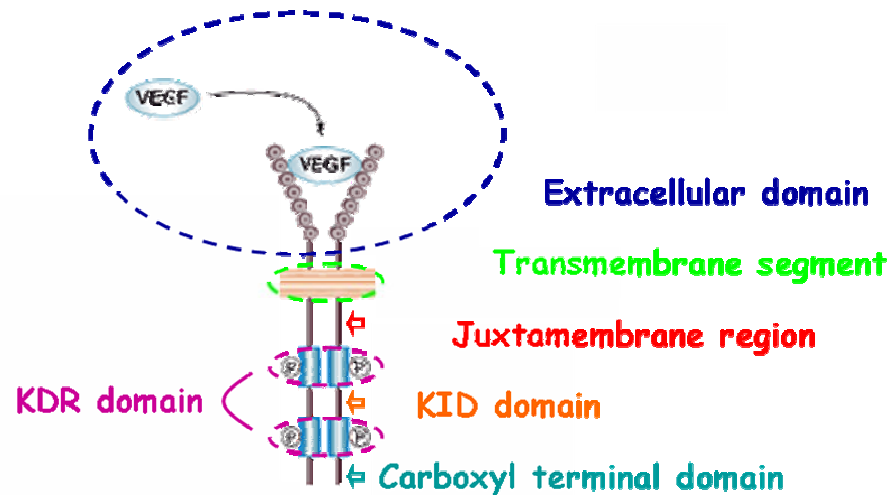
Fig.1.2.1 A simplified representation of the global architecture of the human VEGFR2 in the dimeric form.

In detail (Fig.1.2.1), growth factors bind, according to a not yet fully explained mechanism, the second and third extracellular immunoglobulin domains belonging to two monomeric receptors. This step leads to protein dimerization, activation of the intracellular kinase activity and *trans*-autophosphorylation of tyrosine residues located in the cytoplasmic subunits. [10] This *trans*-autophosphorylation requires a dimer receptor in which a monomer, with the structural features described above, assumes the function of substrate and the other serves as the enzyme. Phosphorylation of tyrosine residues located in a particular flexible segment in the center of the kinase domain, referred to as the "activation loop" (A-loop), promotes the enhancement of the kinase activity, while phosphorylated residues in other cytoplasmic regions of the target become in turn binding sites for other proteins involved in the subsequent signal propagation, such as the SH2- (Src Homology 2) and the PTB- (Phospho-Tyrosine Binding) domain containing proteins. [11]

A deeper knowledge of this mechanism of signal transduction will be possible when the three-dimensional structure of the whole human receptor VEGFR2 is available. At the moment, in fact, only some regions have been solved by X-ray diffraction for this receptor and their three-dimensional structures are collected in the Protein Data Bank (PDB) [12]:

- the three-dimensional structures of the second and third Ig-like extracellular domains in complex with VEGF-C (pdb codes 2X1W and 2X1X [13]);
- the structure of the seventh Ig-like extracellular domain (pdb code 3KVQ [14]).

Both these structures have been reported when our studies on VEGFR2 were already begun.

- the crystal structure of the cytoplasmic KDR domain, the only available at the beginning of this thesis work.

1.2.3 The structure of the KDR domain

The first crystal structure of the KDR domain was reported in 1999 by McTigue et al. [15]; at present 22 X-ray three-dimensional structures of this kinase domain are available in the PDB, both in their apo form and in complex with inhibitors (Table.1.2.1). It is worth noting that none of the crystal structures is completely solved. All the structures lack, indeed, a portion of the kinase insert domain (KID), whose removal seems to be necessary for the crystallization of the protein. [15] Moreover, only one structure (pdb code 2OH4 [16]) has the coordinates for the so-called A-loop, although with high temperature factors (B-factors) associated with the corresponding atoms. This activation segment in the other structures is either partially solved or not solved at all. For these reason the structure 2OH4 has been chosen as reference structure for the study on this macromolecular target and taken as representative for the description of KDR. Its coordinates correspond to the residues comprised between His814 and Asp1169 of the VEGFR2 primary sequence (Fig.1.2.2); the non solved region, belonging to the KID, corresponds to the residues between Tyr936 and Tyr994.

Table.1.2.1 Crystal structures of the KDR domain available, at the moment, in the PDB.

PDB CODE	Descriptor	Resolution (Å)	Source	Reference
1VR2	VEGFR2 (KDR) kinase domain	2.40	Homo sapiens	15
2OH4	VEGFR2 kinase domain in complex a benzimidazole-urea inhibitor	2.05	Homo sapiens	16
3EFL	VEGFR2 kinase domain in complex with motesanib	2.20	Homo sapiens	17
3EWH	VEGFR2 kinase domain in complex with a pyridyl-pyrimidine benzimidazole inhibitor	1.60	Homo sapiens	18
3C7Q	VEGFR2 kinase domain in complex with BIBF1120	2.10	Homo sapiens	19
3CJF	VEGFR2 in complex with a 3,4,5-trimethoxy aniline containing pyrimidine	2.15	Homo sapiens	20
3CJG	VEGFR2 in complex with a 3,4,5-trimethoxy aniline containing pyrimidine	2.25	Homo sapiens	20
3DTW	VEGFR2 kinase domain in complex with a benzisoxazole inhibitor	2.90	Homo sapiens	21
3CP9	VEGFR2 kinase domain in complex with a pyridone inhibitor	2.50	Homo sapiens	22
3CPB	VEGFR2 kinase domain in complex with a bisamide inhibitor	2.70	Homo sapiens	22
3CPC	VEGFR2 kinase domain in complex with a pyridone inhibitor	2.40	Homo sapiens	22
2RL5	VEGFR2 kinase domain in complex with a 2,3-dihydro-1,4-benzoxazine inhibitor	2.65	Homo sapiens	23
3BE2	VEGFR2 kinase domain in complex with a benzamide inhibitor	1.75	Homo sapiens	24
3B8Q	VEGFR2 kinase domain in complex with a naphthamide inhibitor	2.75	Homo sapiens	24
3B8R	VEGFR2 kinase domain in complex with a naphthamide inhibitor	2.70	Homo sapiens	25
2QU5	VEGFR2 kinase domain in complex with a benzimidazole inhibitor	2.95	Homo sapiens	26
2QU6	VEGFR2 kinase domain in complex with a benzoxazole inhibitor	2.10	Homo sapiens	26
2P2H	VEGFR2 kinase domain in complex with a pyridinyl-triazine inhibitor	1.95	Homo sapiens	27
2P2I	VEGFR2 kinase domain in complex with a nicotinamide inhibitor	2.40	Homo sapiens	27
1YWN	VEGFR2 kinase domain in complex with a novel 4-amino-furo[2,3-d]pyrimidine	1.71	Homo sapiens	28
1Y6A	VEGFR2 kinase domain in complex with a 2-anilino-5-aryl-oxazole inhibitor	2.10	Homo sapiens	29
1Y6B	VEGFR2 kinase domain in complex with a 2-anilino-5-aryl-oxazole inhibitor	2.10	Homo sapiens	29
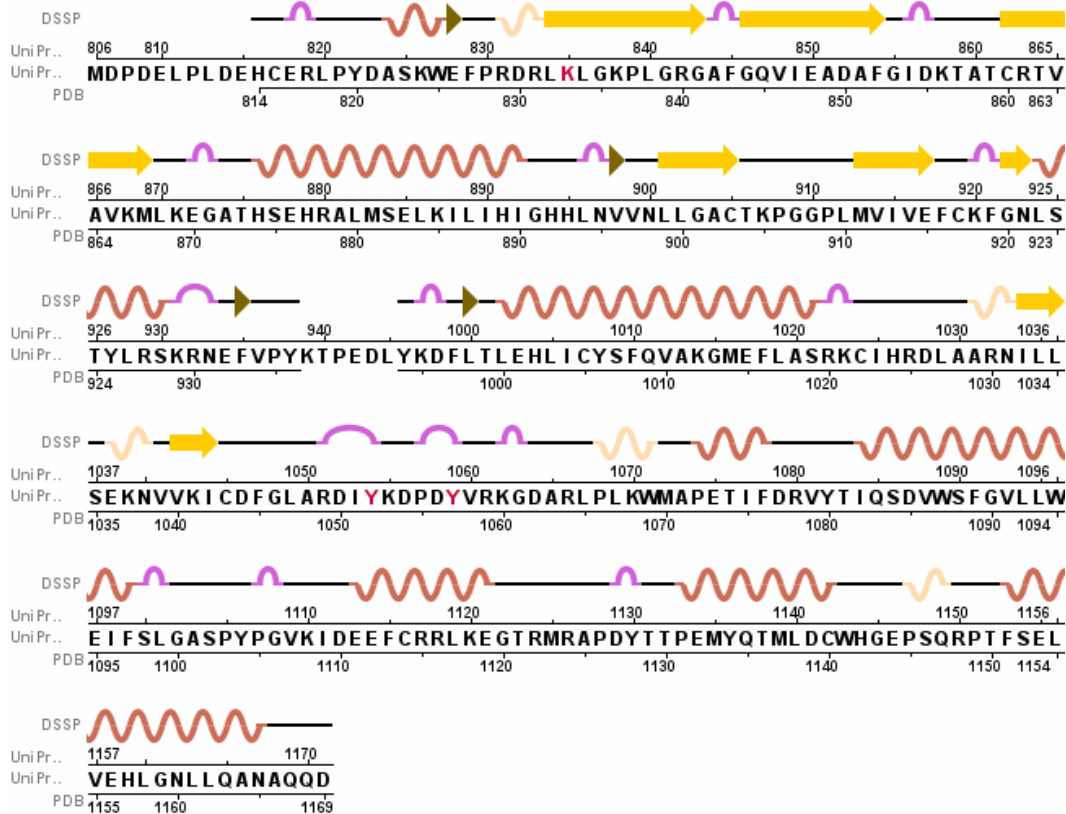
Fig.1.2.2 A schematic representation of the crystal structure 2OH4. It is worth noting that the numbering used in the structure 2OH4 is shifted by two residues compared to that used in the primary sequence of the receptor VEGFR2 (UniProtKB sequence: P35968, VEGFR2_HUMAN). The residues 814 and 1169 of 2OH4 are actually the 816 and 1171 residues of the VEGFR2 primary sequence.

The structural analysis of the crystallographic data reveals, in the KDR domain, the same overall architecture which is found in most of the tyrosine kinase domains. In detail, the KDR domain (Fig.1.2.3) has the characteristic bilobed structure highly conserved among eukaryotic Ser/Thr and Tyr kinases with a smaller N-terminal lobe (residues 814-915; light blue in the Fig.1.2.3) consisting mainly of antiparallel β-sheets, apart from an important α-helix (αC-helix, corresponding to residues 874-890; blue in the Fig.1.2.3), and a larger C-terminal lobe comprising the residues 922-1169 and mainly formed by α-helices (orange in the Fig.1.2.3). The binding site for ATP-Mg^{2+} is located in the cleft between the N- and C-terminal lobes, joined together by a conserved segment, referred to as the *"hinge region"* FCKFGN (residues 916-921, using the one-letter code for amino acids; pink in the Fig.1.2.3).
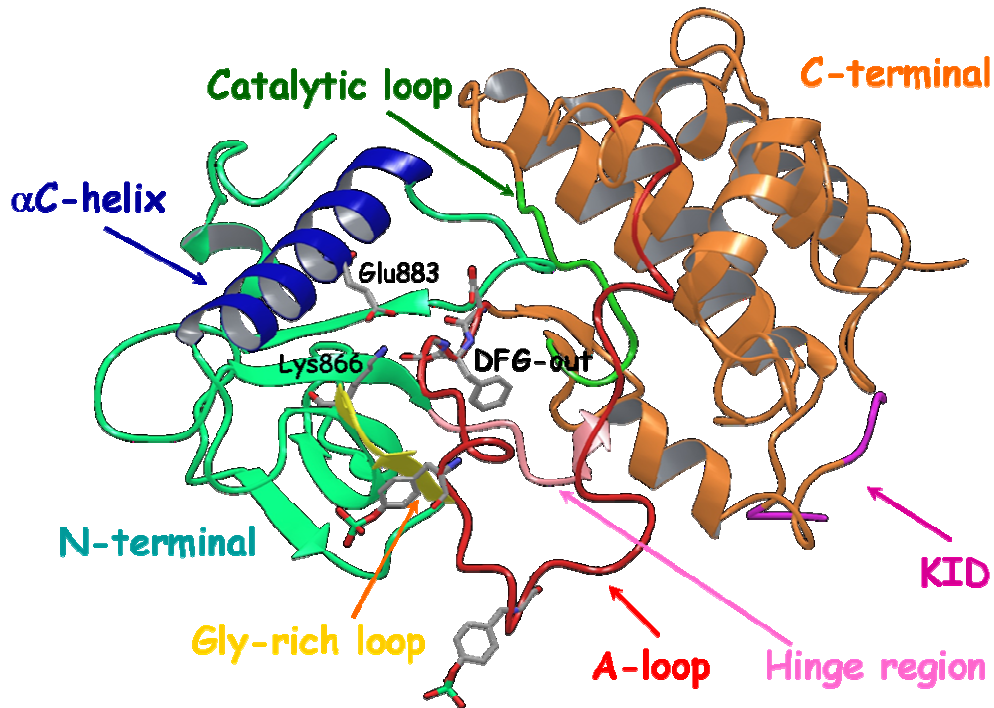
Fig.1.2.3 A schematic representation of the global architecture found in the KDR domain (crystal structure 2OH4 [16]).

Some regions of the KDR domain contain residues highly conserved in all protein kinases and involved in essential structural and dynamic roles.

In detail there are:

N-terminal lobe

- the glycine rich loop GRGAFG (residues 839-844, yellow in Fig.1.2.3) which defines the top of the ATP binding site;
- the residue Lys866, located in the so-called strand β3 and
- the residue Glu883 located in the αC-helix.

Residues Lys866 and Glu883 are involved in the complex mechanism of the A-loop transition from the inactive (closed) to the active (open or extended) form in kinases. This transition is also accompanied by the movements of the β-sheets in the N-terminal lobe as a rigid body. Moreover, the αC-helix undergoes to a rotation, upon the A-loop movement, which places the conserved Glu883 residue in proximity to

the conserved Lys866. The rotation of the N-terminal lobe towards the C-terminal lobe brings Lys866 and Glu883 into position for the proper ATP-Mg^{2+} binding.

C-terminal lobe

- the catalytic loop HRDLAARN (residues 1024-1031, green in Fig.1.2.3);
- the activation loop (residues 1044-1073, also called A-loop or T-loop, red in Fig.1.2.3), a centrally located regulatory element in protein kinases

In the catalytic loop (HRDLAARN) the first three amino acids (HRD motif, His1024, Arg1025, Asp1026) are the most conserved among protein tyrosine kinases, with the Asp1026 which acts as catalytic base in the phosphotransfer reaction (Fig.1.2.4).

The A-loop, completely solved only in the crystal structure 2OH4, begins with the characteristic DFG motif (Asp1044-Phe1045-Gly1046), highly conserved among protein kinases, and ends with the APE motif (Ala1071-Pro1072-Glu1073).

The A-loop is formed by:

(1) the magnesium binding loop, DFGLA (residues 1044-1048), in which the first residue, Asp1044, binds the Mg^{2+} ion during the phosphotranfer reaction. The Mg^{2+} in turn coordinates the phosphate group β and γ in ATP (Fig.1.2.4);

(2) the so-called strand β9;

(3) the activation segment, YpKDPDYp (residues 1052-1057), which contains two tyrosine residues undergoing phosphorylation (Tyr(P)1052-Tyr(P)1057).
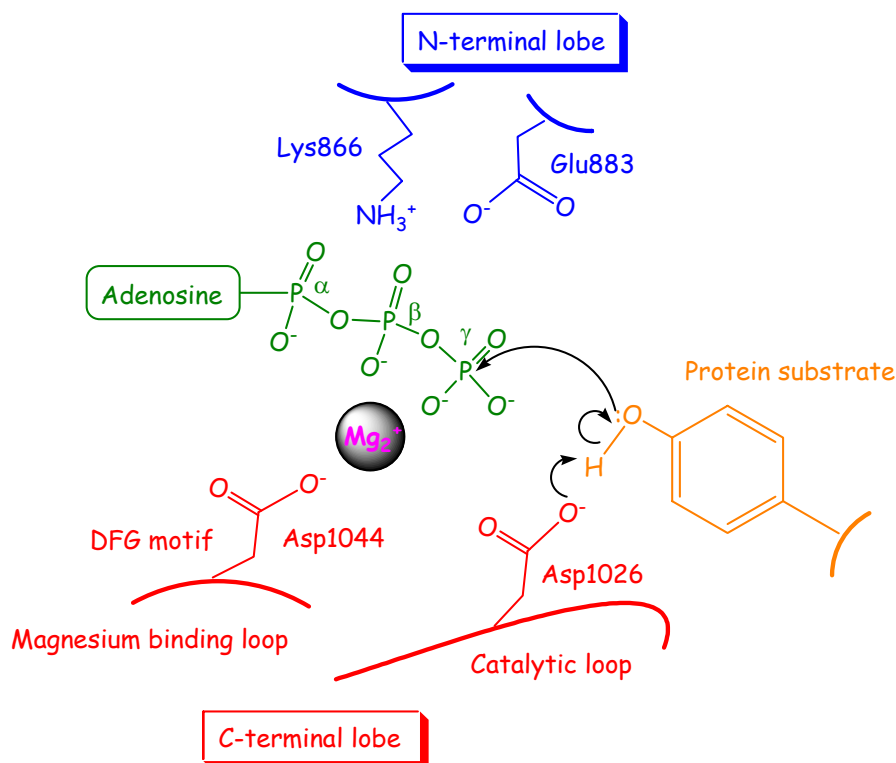
Fig.1.2.4 A schematic representation of the mechanism proposed for the reaction catalyzed by an active tyrosine kinase. (Adapted from reference [30])

The A-loop has been recognized as one of the most flexible portions of the kinase domain and can exist in different conformations, in a dynamic equilibrium controlled by the phosphorylation; this mobility is consistent with the high disorder observed in the crystal structure. Two limit conformations are supposed for this loop, corresponding to the inactivated and catalytically active state of the kinase. The double phosphorylation of the tyrosine residues should stabilize the A-loop in its active conformation. [31] In this active (open or extended) conformation, the Asp1044 residue of the DFG motif at the N-terminal end of the A-loop is positioned so as to interact properly with the Mg^{2+} which coordinates two phosphate groups of the ATP (Fig.1.2.4). The structure 2OH4 is a bit unusual in that, although double phosphorylated, its A-loop is in an inactive conformation, where the DFG motif adopts a so-called "out" conformation, with the phenyl ring of the side chain of the residue Phe1045 pointing towards the ATP binding site, in a position that actually does not allow the coordination of the magnesium ion by the residue Asp1044. [8]

In the KDR domain is also present, in the C-terminal lobe, the so-called kinase insert domain (KID), which split the kinase domain in two regions. The KID domain corresponds to the residues 931-998 and is partially deleted in all the crystallized proteins available in the PDB, including 2OH4 (residues 936-994; purple in the Fig.1.2.3). It has been demonstrated that most of the KID domain, far away from the catalytic site and from the A-loop, is not necessary for the kinase activity. KID is, however, important as binding site for other proteins, after phosphorylation of two of its tyrosine residues, and as a linker which connects two α-helices in the C-terminal lobe and maintains the overall architecture of the KDR domain. [15]

1.2.4 Conformational flexibility of the kinase domain

Structural and computational studies performed on kinases solved in active and inactive states have made possible to point out two main conformational motions associated with the transition from the inactive to the active state: a change in the orientation of both the αC-helix in the N-terminal lobe and the A-loop in the C-terminal lobe. [32-34] These movements are in turn followed by changes in other structural elements. The dynamic of the A-loop, for example, affects the conformation of the characteristic DFG motif; the conformational switch of this flexible motif is referred to as "*DFG flip*". In the active conformation, the Asp1044 (using the numbering of 2OH4) residue of the DFG motif points towards the ATP binding cleft and the phenylalanine is buried in a hydrophobic pocket adjacent to the ATP site; this orientation is referred to as *DFG-in*. In the other limit orientation, the so-called *DFG-out* conformation, the DFG aspartate and phenylalanine side chains swap positions and point in opposite directions compare to the *DFG-in* orientation. In the *DFG-out* conformation, the phenylalanine side chain points towards the ATP binding site and thus creates a hydrophobic pocket (named *allosteric hydrophobic pocket*), adjacent to the ATP-binding site. On the other side, the conformational change in the αC-helix affects the mutual orientation of two highly conserved residues, namely Glu883 and Lys866. These two residues strongly interact via a salt bridge, and Lys866 contributes to the proper orientation of the α and β phosphate ATP groups during the catalysis (Fig.1.2.4). The αC-helix movement also causes a slight shift of the *glycine-rich loop*.

Once the target is in its active state, another type of conformational change occurs in the kinase domain: the N- and C-terminal lobes move relative to each other to open and close the binding cleft.

All these conformational motions occur during the catalytic cycle, which involves in a first step the binding of ATP and protein substrate to the open conformation of the target. Once the complex is formed, the catalyzed reaction occurs in the closed conformation, and ADP and the phosphorylated substrate are released during the subsequent evolution to the open state that completes this cycle. [34]

In general, it has emerged that, while the inactive states of the kinase domains show a great structural variability among RTKs, the active state is essentially invariant from a structural perspective. [35, 36] Recently, it has also been shown that all active conformations are strongly stabilized by a "hydrophobic spine", which traverses both the N- and C-terminal lobes. This spine is instead disordered in the inactive forms of kinases and this can explain the higher variability observed for these conformations. [32, 33]

The tyrosine kinase domain of the human insulin receptor can be chosen as a suitable example to show the conformational changes which occur during the activation process. Its crystal structure, in fact, has been solved in two different conformations, corresponding to the inactive (closed) and active (open) form (pdb codes 1IRK [37] and 1IR3 [38]) (Fig.1.2.5). In the unphosphorylated structure (1IRK) the A-loop is oriented in order to prevent the binding of ATP-Mg^{2+} and of the protein substrate. In the active and phosphorylated conformation (1IR3), we observe a considerable change in the orientation of the A-loop, which enable the binding of ATP-Mg^{2+} and an efficient phosphorylation of the protein substrate. A significant change can also be observed in the orientation of the highly conserved DFG motif, which assumes respectively a *DFG-out* and a *DFG-in* conformation, and in the orientation of the αC-helix in the N-terminal lobe.
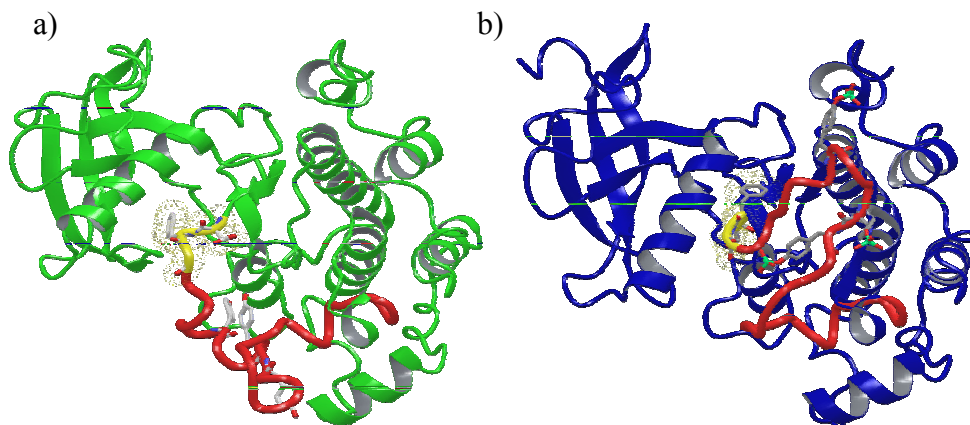
Fig.1.2.5 Ribbon representation of: a) the inactive conformation (crystal structure 1IRK) and b) the active conformation (crystal structure 1IR3) of the kinase domain of the human insulin receptor. The A-loop is coloured in red: a) closed conformation for the A-loop and b) open conformation for the A-loop. The DFG motif is highlighted in yellow: a) *DFG-out* orientation and b) *DFG-in* orientation.

To the best of our knowledge, a comprehensive (structural or computational) study which provides a detailed description of the activation mechanism at the atomistic level has not been previously reported on the KDR domain.

1.2.5 Inhibitors of Receptor Tyrosine Kinases

Receptor tyrosine kinases represent an attractive target in oncology drug discovery. In this context, the detailed knowledge of the conformational flexibility of their kinase domains is of remarkable value since the inhibition of RTKs has been mainly achieved through highly potent small organic compounds binding the ATP pocket. In particular, the conformational variability of the DFG motif has been employed in the design of protein kinase inhibitors.

These compounds can act or by mimicking the ATP binding mode to the kinase domain in its active and *DFG-in* conformation (ATP-competitive inhibitors) or by locking kinases in their inactive and *DFG-out* state (non-ATP competitive inhibitors). [39]

Compounds targeting the ATP binding site in the active and *DFG-in* conformation are generally referred to as *type-I inhibitors* (Fig.1.2.6). A common feature of the

type-I inhibitors is their ability to bind to the ATP site mimicking the adenine ring in its interactions with the "hinge" residues of the protein. The non-ATP competitive inhibitors, which bind the inactive and *DFG-out* conformations, are defined as *type-II inhibitors* (Fig.1.2.6). In the *DFG-out* form, the aromatic ring of the phenylalanine residue (Phe1045 in the crystal structure 2OH4 [16]) points towards the ATP pocket and creates, near the ATP-binding site, an allosteric hydrophobic binding pocket suitable for occupation by type-II kinase inhibitors. Type-II inhibitors bind simultaneously to the ATP binding site and to the allosteric hydrophobic pocket. Recently compounds which bind exclusively within the allosteric pocket have also been identified and are referred to as *type-III inhibitors*. [39-41]

Gefitinib

Type-I inhibitor

Imatinib
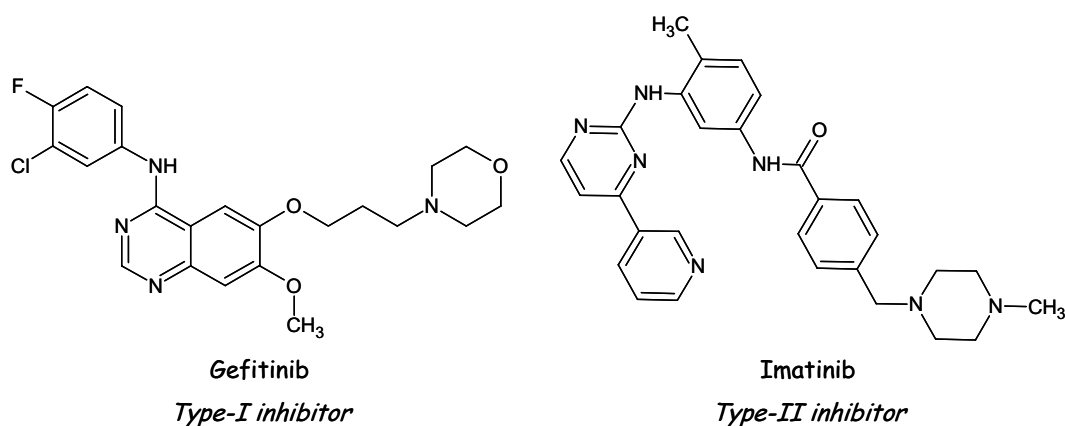
Type-II inhibitor

Fig.1.2.6 Examples of type-I and type-II kinase inhibitors approved by the FDA for cancer treatment

Because of the highly conserved architecture of the kinase domains among RTKs, the discovery of selective inhibitors, which bind only to one kinase and not to others, is a fundamental challenge in oncology drug-discovery. The deep knowledge of the conformation of the A-loop, αC-helix and other related structural elements in the active and inactive states, as well as the knowledge at atomic level of the path connecting these two limit conformations and particularly of the DFG flip, is thus crucial for providing a platform for the rational design of highly selective kinase inhibitors.

1.2.6 Motivation and Aims of the work

VEGFR2 belongs to the tyrosine kinase receptor family. It is involved both in physiological and pathological angiogenesis and vasculogenesis. It has been recognized as an attractive target in anticancer therapies. In this thesis work we have investigated the structural and dynamic features of its kinase domain, the KDR domain. In a first step, through the application of molecular modeling strategies, we have tried to characterize the active and inactive conformations of this target, completing the gaps in the available crystallographic data. In a second step, we have investigated the KDR domain in order to obtain a detailed description at atomic level of the transition from the inactive to the active conformation. In this stage, molecular dynamics simulations and advanced sampling methods have been applied to model the KDR flexibility and to sample intermediate conformations which can be used in the rational design of selective inhibitors for this target.

1.3 DNA: DOUBLE HELIX AND G-QUADRUPLEX STRUCTURES

1.3.1 DNA double-stranded

The activity of many antimalarial, antibacterial and anticancer agents is based on their interaction with helical double-stranded DNA (dsDNA). [42-44] As a consequence, much effort has been devoted over the past few decades to the design and synthesis of new molecules that can reversibly bind and/or react with dsDNA. The double helical structure model for the DNA was proposed for the first time by Watson and Crick in 1953. [45] In this canonical DNA structure, always referred to as the B-DNA structure, two strands of DNA are coiled around a common axis and linked together by specific hydrogen-bonds involving a purine and a pyrimidine base according to the scheme Adenine (A) – Thymine (T) and Guanine (G) – Cytosine (C), known as Watson and Crick base pairing. This double helix is right handed with antiparallel backbone chains and ten nucleotides per turn. The bases are located on the inner part of the helix and the phosphate groups are on the outside. On average, the distance between adjacent base pairs is 3.4Å. The two strands make two grooves of different widths, referred to as the major groove and the minor groove (Fig.1.3.1).
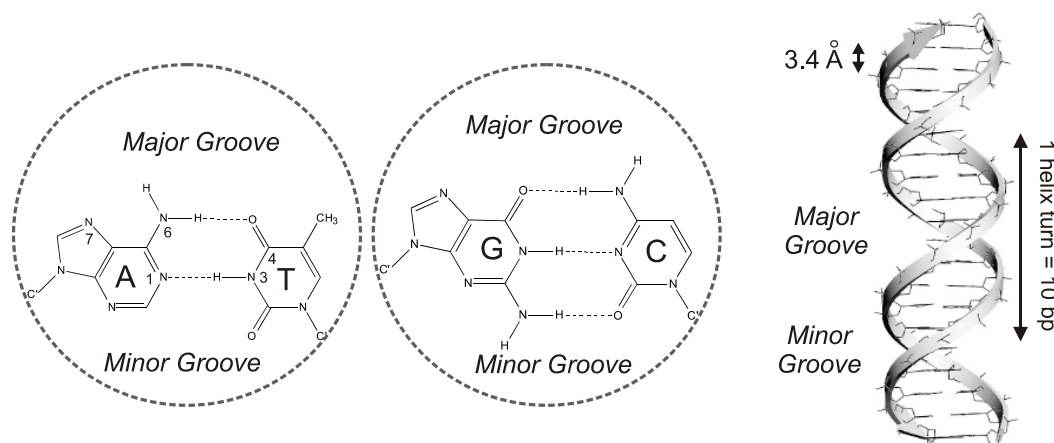
Fig.1.3.1 Features of the B-DNA Watson and Crick model. (Adapted from reference [46])

Several studies on DNA sequences, subsequent the assumption of Watson and Crick about the double-stranded structure, have revealed a high flexibility and polymorphism of DNA which is able to adopt different three-dimensional folded

structures. [47] Non-canonical DNA structures can arise from either a different base pairing A-T and C-G compared to that of Watson and Crick, such as the Hoogsteen and reverse Hoogsteen hydrogen bonding schemes, or a purine-purine base pairing, such as the G-G base pairing (Fig.1.3.2). [48-50]

A-T Hoogsteen

A-T Reverse-Hoogsteen

A-T Watson and Crick
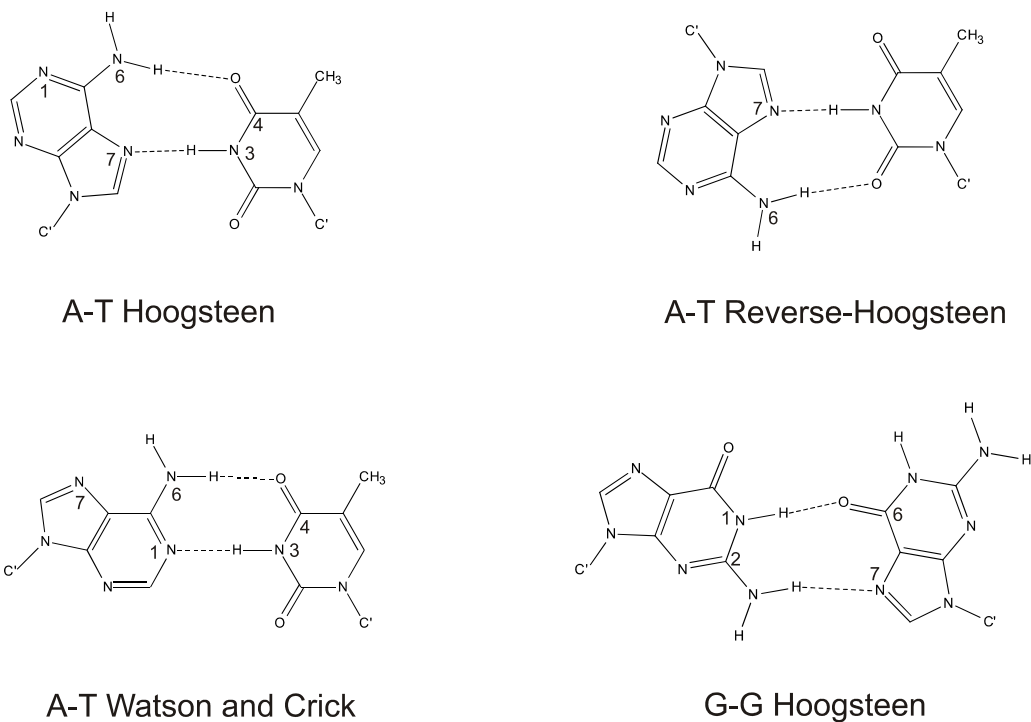
G-G Hoogsteen

Fig.1.3.2 Hoogsteen and reverse-Hoogsteen A-T base pair hydrogen bonding schemes, compared with A-T Watson and Crick hydrogen bonding scheme in B-DNA. Hoogsteen hydrogen bonding scheme for G-G (Adapted from reference [46])

Among these non-standard structures, recently a great deal of attention has been devoted to clarifying the structural features of G-quadruplexes observed in the telomeric DNA. [51-55]

1.3.2 G-quadruplex DNA structures

Guanine quadruplexes (G-quadruplexes) are unconventional DNA secondary structures that consist of at least two stacked building blocks, referred to as G-quartets or G-tetrads, formed in turn by four guanines associated together into a cyclic and planar arrangement through a network of eight Hoogsteen hydrogen bonds. [56] These coplanar cyclic arrays are connected and stabilized by monovalent

cation such as K^+ or Na^+ located in the negatively charged hole between stacked G-tetrads. The different G-quartet plans are at an average distance of 3.4Å (Fig.1.3.3).
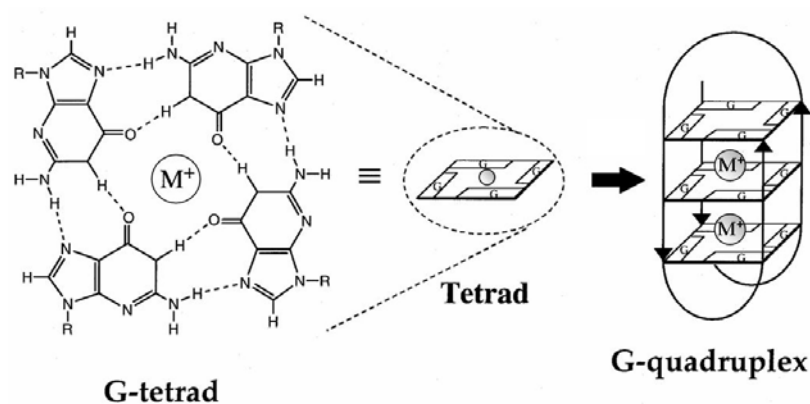
Fig.1.3.3 A schematic drawing of an intra-molecular G-quadruplex folding induced by a metal ion. A G-tetrad plane is composed of four guanines connected by eight Hoogsteen hydrogen bonds.

A higher polymorphism, compared to double stranded DNA, can be found in G-quadruplex structures. First, G-quadruplexes can be formed either from a single tract of G-rich sequence of DNA (intramolecular G-quadruplexes), or from the association of two or four separate strands (dimeric and tetrameric G-quadruplexes). Moreover, the four-stranded phosphodiester chains of all G-quadruplexes can be found in different relative arrangements: all parallel (Fig.1.3.4 a)), three-parallel and one anti-parallel (Fig.1.3.4 b)), adjacent parallel (Fig.1.3.4 c)) and alternating parallel (Fig.1.3.4 d)).

a)   b)   c)   d)
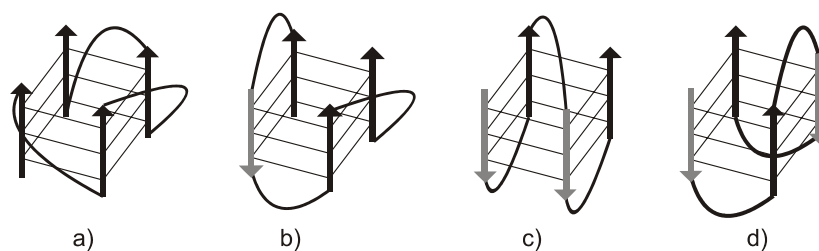
Fig.1.3.4 Strand polarity in G-quadruplexes. (Adapted from reference [46])

Another contribution to the resultant diversity in G-quadruplex topologies is given by a high variability in the sequences connecting the different G-quartets in dimeric and monomeric G-quadruplexes. These loops (TTA loops) are involved in determining the stability and the flexibility of the overall G-quadruplex structure and can be

classified into four groups: diagonal loops, connecting two opposing antiparallel strands (Fig.1.3.5 a)), edgewise or lateral loops, connecting two adjacent antiparallel strands (Fig.1.3.5 b)), double-chain-reversal or propeller loops, connecting adjacent parallel strands (Fig.1.3.5 c)), and V-shaped loops, connecting two corners of a G-tetrad core in which one supporting column is lacking (Fig.1.3.5 d)). It has also been demonstrated that these TTA loops can act as important targets for drug-binding.[57] Finally, the G-quadruplex topology can be influenced by the solution environment, such as the presence of different metal ions (Na^+ or K^+).
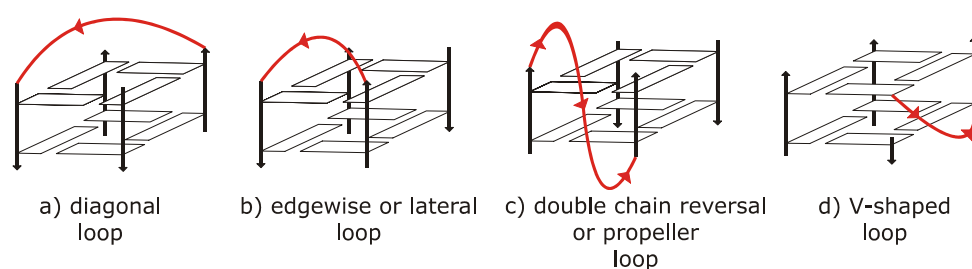
Fig.1.3.5 Loops topology in G-quadruplex structures. (Adapted from reference [46])

1.3.3 Human telomeric DNA

The human genome contains a lot of sequences able to form G-quadruplex structures. [58] Some of these G-rich sequences have been recognized in human telomeric DNA, which is involved in the immortalization process of cancer cells [59-64] and genetic stability. [65–67] Therefore, it represents a potentially suitable target for anticancer strategies.

Telomeres are non-coding DNA regions located at the end of chromosomes and whose function is to preserve chromosomal integrity. In humans and vertebrates telomeres consist of repeats of the guanine-rich hexanucleotide motif d(TTAGGG)5´-3´, toward the chromosome end. Shortening of the human telomeric DNA occurs in normal somatic cells as a function of the replication cycle and leads to cell-cycle arrest and eventually apoptosis when a critical length minimum is reached. [68] In contrast, cancer cells have evolved mechanisms to maintain telomere length, the most common of which is based on the activation of a reverse transcriptase called telomerase. The unlimited proliferative potential of cancer cells depends on this telomere maintenance mechanism. This enzyme is activated in the 85

- 90 % of tumor cells and functions as a reverse transcriptase by adding multiple copies of the 5´-d(GGTTAG)-3´ motif to the end of telomeres. [69-71] Novel antitumor strategies are aimed at interfering with the mechanisms of telomere maintenance by targeting telomerase directly according to an enzyme recognition process, or indirectly by means of telomere-interacting agents. [72]

The human telomeres are largely double-stranded DNA sequences, but the terminal 100-200 nucleotides at the 3´ end are generally single stranded. This portion is able to easily fold into an intramolecular G-quadruplex arrangement. Optimal telomerase activity requires an unfolded single-stranded DNA as substrate, while the formation of G-quadruplexes at the end of telomeres prevents the telomerase enzymatic process. [73, 74] Therefore, compounds that selectively bind to and stabilize G-quadruplex structures may interfere with telomere elongation and thus act as potential drugs in anticancer therapy. [75, 76]

In general, the development of an efficient and selective drug needs the knowledge of the biological target in terms of chemical and biological properties as well as of structural information. In this context, human telomeric DNA in its G-quadruplex folding is not an easy target, because of the high polymorphism of the G-quadruplex structures. [55]

This high polymorphism can be observed in the crystal and NMR-derived structures of human telomeric repeats, obtained both in K^+ and in Na^+ solution, available in the Protein Data Bank (PDB). [12] Since a certain difficulty exists in determining the overall fold of the long telomeric sequences present *in vivo*, the crystal and NMR-derived structures have been obtained starting from short human telomeric sequences with not more than four repeats, such as d[AGGG(TTAGGG)$_3$] referred to as Tel22. These sequences in Na^+ solution assume a unique structure, referred to as "basket-type" [77], which presents anti-parallel strands connected by diagonal and lateral loops (Fig.1.3.6 a)).

a) BASKET b) HYBRID

c) PROPELLER (ALL PARALLEL)
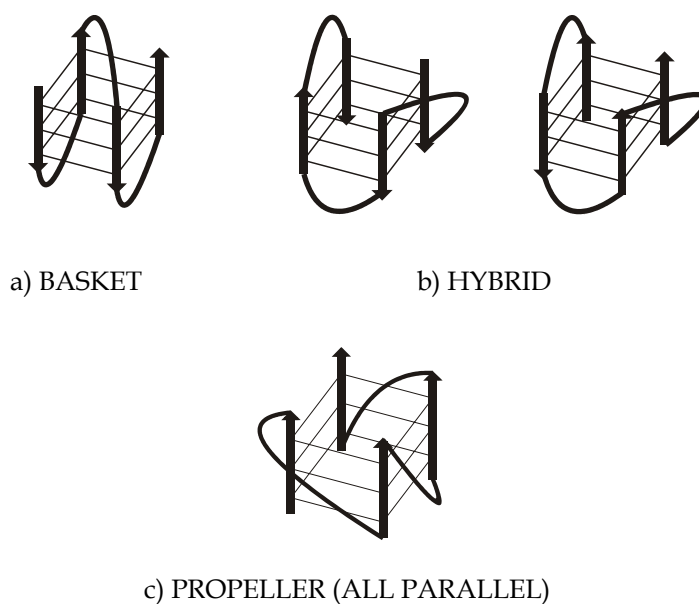
Fig.1.3.6 G-quadruplex structures formed by short telomeric sequences.

Although the cellular environment features both the monovalent cations, K^+ is largely predominant, and, as a consequence, experiments involving telomeric sequences have been usually performed in K^+ solutions. Two different G-quadruplex structures, named "hybrid-1" [78] and "hybrid-2" [79] (Fig.1.3.6 b)), have been determined for short telomeric sequences in potassium solution. It has been suggested that both hybrid-1 and hybrid-2 form and coexist in equilibrium in potassium solution. [80] A completely different conformation has been found instead for these telomeric sequences in the crystal structures solved in presence of potassium ion; this conformation is referred to as "propeller-type" or "all-parallel" (Fig.1.3.6 c)). [81]

1.3.4 G-quadruplex ligands

G-quadruplex binders, due to their ability to inhibit telomerase enzymatic activity, are now considered potential anti-cancer drugs. The number of identified G-quadruplex ligands has grown rapidly over the past few years. Extensive efforts have been made to establish reliable structure-activity relationships with the aim of identifying effective and selective telomerase inhibitors. Some of these ligands have been found to stack on the termini G-quartets of the quadruplex (external π-

stacking), [82] while others interact with quadruplex grooves or external loops without any specific π-stacking interaction with G-quartet bases. [83, 84]

As common structural features, G-quadruplex ligands generally share a large, flat, aromatic surface, and the presence of protonatable side chains. [72] These requirements are fulfilled provided that the molecule exhibits both hydrophobic and hydrophilic characteristics.

This pharmacophoric model has often been identified in some classical dsDNA intercalators, whose derivatives subsequently have been recognized as G-quadruplex binders. An interesting example is represented by acridine derivatives (Fig.1.3.7 a)). Effective anticancer drugs containing this heterocyclic moiety (such as amsacrine, *m*-AMSA; Fig.1.3.7 b)) are generally able to efficiently recognize DNA according to an intercalation binding mode.
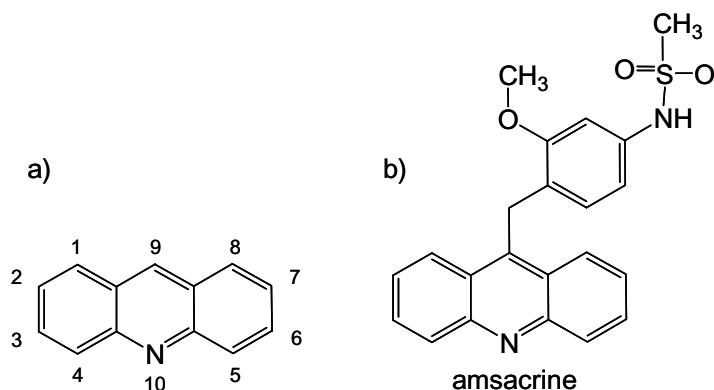
Fig.1.3.7 a) Schematic representation of the acridine skeleton, with the relative numbering system; b) structure of amsacrine (*m*-AMSA)

However, several studies have shown that by modulating the substitution pattern on this aromatic ring system, it is possible to preferentially direct such ligands toward G-quadruplex structures. In particular, a series of 3,6-disubstituited [85-87] and 3,6,9-trisubstituted acridines, [88-91] as well as some compounds characterized by a macrocyclic skeleton containing the acridine moiety [92, 93] have been reported as efficient G-quadruplex binders endowed with pharmacological anticancer properties.

In this contest, a great contribution has been made by Neidle and coworkers through the design of the 3,6,9-trisubstituted acridine ligand BRACO-19 (Fig.1.3.8), and the subsequent resolution of the crystal structure of the complex between BRACO-19 and the bimolecular human telomeric G-quadruplex of sequence d(TAGGGTTAGGGT), obtained in presence of potassium ion (pdb code 3CE5).[54]

The crystal structure of this complex features an "all-parallel" G-quadruplex arrangement and can be considered as a model of the interaction between quadruplexes and acridine ligands. In addition to strong G-quadruplex binding ability, BRACO-19 inhibits telomerase enzymatic activity [94] and cancer cells proliferation. [95]
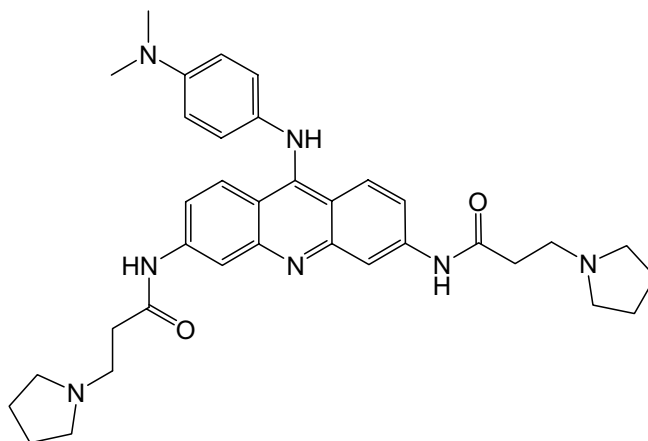
Fig.1.3.8 Structure of the ligand BRACO-19

On the other hand, different substitution patterns on the acridine skeleton, for example at 2,7 position, have, to date, been poorly investigated as G-quadruplex binders, because they apparently do not provide relevant G-quadruplex recognition.[96]

1.3.5 Computational investigations of G-quadruplex-ligand complexes

Quadruplex–ligand interactions can be investigated by both experimental and computational techniques; in this thesis work we have employed computational approaches. In this context, a series of molecular modeling methodologies applied to the investigation of G-quadruplex structures and their complexes have been recently reviewed by Neidle and Haider [97], highlighting advantages and limitations of these approaches. Some problems in simulating G-quadruplexes can arise, indeed, from their high structural polymorphism and flexibility as well as from their highly charged environment, resulting from the combination of the charges on the phosphate groups and on the central ion-containing channel. The absence of a well-defined binding site is another challenge that must be considered in the computational investigation of these molecular systems. However, this review points

out that recent improvements in force fields parameters for nucleic acids and computational algorithms make possible a more confident application of modeling methodologies to G-quadruplex complexes.

1.3.6 Aim of the work

Acridine ligands have shown the ability to bind and stabilize G-quadruplex structures and have received attention as potential anti-cancer drugs. To our knowledge, however, 2,7-disubstituted acridine derivatives have, to date, been poorly investigated as G-quadruplex binders. Recently, a series of polyaza ligands (Fig.1.3.9) have been synthesized, which feature an acridine moiety inserted in an open-chain (L1) or macrocyclic (L2 and L3) aliphatic polyamine framework via functionalization of the 2- and 7-positions of the heteroaromatic system. These ligands enable the exploitation of the effect that a cyclic organization of the aliphatic polyamine chain has on the recognition properties for various DNA structures.

Thus, in order to obtain structural information and to elucidate the thermodynamic forces that regulate the binding process, we decided to carry out a comprehensive study on the binding properties of the three 2,7-disubstituted acridine derivatives (L1, L2, and L3, Fig.1.3.9) towards telomeric G-quadruplex structures, and to compare them with their respective dsDNA binding profiles. In this thesis work we report the detailed computational investigation performed on these molecular systems. [98]
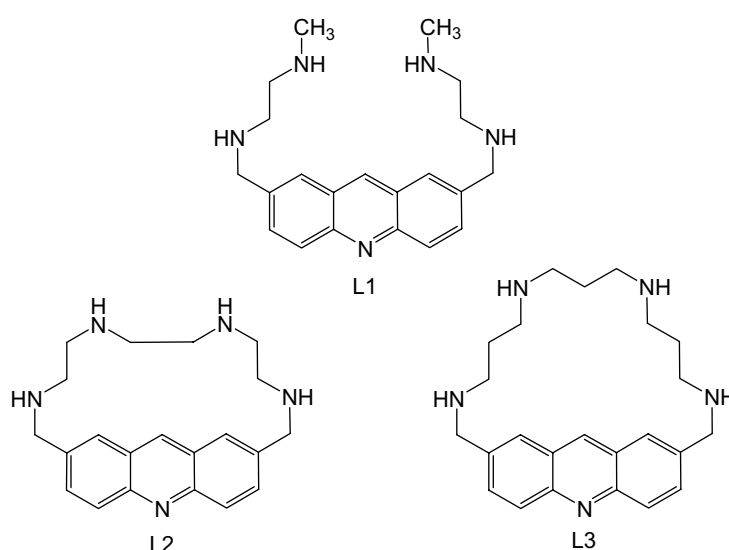
Fig.1.3.9 Acridine derivatives (L1, L2 and L3) considered in the computational study.

2. Materials and Methods

This chapter deals with theoretical bases underlying the computational methods used throughout the investigation together with the indication of the software and the hardware used. All methodologies are based on the equations of classical physics and are related to an MM (Molecular Mechanics) approach to the study of the molecules.

2.1 Hardware

- Workstation SO Linux Dual Core AMD Opteron 4 processors
  - Distribution: Linux version 2.6.9-42.0.10.ELsmp
  - CPU: i686
  - Processors: 4
  - Perl: 5.008
  - Glibc: 2.3.4

- Workstation SO Linux Intel Xeon 8 processors
  - Distribution: Linux version 2.6.18-53.1.13.e15
  - CPU: x86_64
  - Processors: 8
  - Perl: 5.008
  - Glibc: 2.5

- High Performance Computing (HPC) based on Linux cluster
  FRONT-END
  - Server: 1U E-Rack 4151
  - Processor: Intel Core 2 Duo E6400 2.13 GHz
  - RAM: 1 GB DDR2-667 ECC Reg.
  - Disk-array: 2 x SATAII 500 GB hot-swap RAID1
  - DVDRW: Dual Layer slim
  - 2 x Interface LAN 1 Gbit
  2 NODES: A002 COMPUTE NODE INTEL DUAL
  - Server: 1U E-Rack 5122
  - 2 x Processor Intel Xeon 5130 DC 2,0 Ghz 4MB
  - RAM: 4 GB DDR2-667 ECC Reg.
  - Hard-disk: SATAII 80 GB 7200 rpm hot-swap
  - DVDROM: slim
  - 2 x Interface LAN 1 Gbit

- CINECA HPC IBM P575 Power 6 SP6

  - Model: IBM pSeries 575

  - Architecture: IBM P6-575 Infiniband Cluster

  - Processor Type: IBM Power6, 4.7 GHz

  - Computing Cores: 5376

  - Computing Nodes: 168

  - RAM: 21 TB (128 GB/node)

  - Internal Network: Infiniband x4 DDR

  - Disk Space: 1.2 PB

  - Operating System: AIX 6

  - Peak Performance: 101 TFlop/s

2.2 Software

- Schrödinger® Protein Preparation Wizard (protein preparation);

- Swiss-PdbViewer (protein analysis and homology modeling);

- BioEdit (protein sequences alignment);

- ORAC (MD simulations and REM simulations);

- AMBER9 suite (MD simulations and free energy calculations);

- Schrödinger® Maestro (building and visualization of molecules);

- Schrödinger® Glide (molecular docking);

- Gaussian09 (calculation of the electrostatic potential);

- APBS (solution of the Poisson Boltzmann equation);

- Mercury (visualisation and analysis of molecular structures);

- VMD (visualization and analysis of MD trajectories).

2.3 Computational methods

2.3.1 Molecular Docking and Scoring

The computational technique referred to as ***molecular docking*** aims to predict the optimal mutual orientation resulting from the interaction between different molecules, such as two biological targets (i.e. protein-protein interaction) or a ligand and a receptor (protein or DNA fragments). [99-101] In this thesis we focused on this last approach, trying to find the optimal placement (*pose*) of small compounds (ligands) in the binding site of a biological target (receptor) and to estimate their binding affinities. Such a docking experiment requires at least three different

elements: the availability of the three-dimensional structure of the biological target, a database containing the structures of synthesized molecules or "virtual" compounds to test and a computational algorithm. In general, the available algorithms operate in two steps: at first (*pose generation or docking*) they place a ligand into the receptor binding site in the appropriate manners for optimal interactions with the macromolecules, while in a second stage (*pose evaluation or scoring*) they evaluate the ligand-target interactions in order to estimate the binding affinity by means of appropriate scoring functions. In many docking methods, during the pose generation, only the ligand is treated as flexible while the structure of the target is kept rigid. However, both molecules involved in the interaction are flexible and some docking algorithms are now able to take into account the mutual conformational changes which occur upon the binding between the ligand and the target. [102, 103]

The docking algorithms available nowadays are quite reliable in the pose generation and are generally able to reproduce the experimental pose of a compound, obtained by X-ray crystallography (*docking accuracy*). The major problem can arise on the contrary in the scoring stage, where it is very difficult to obtain an accurate prediction of binding affinities for a set of molecules (*scoring accuracy*). The scoring functions of docking algorithms are indeed a fast and simplified estimation of the binding free energy related to the ligand affinity: $\Delta G_{binding} = -RT \ln(K_{affinity})$.

Three types of scoring functions are now currently employed:

- o **Force field based**, which use electrostatic and van der Waals interaction terms derived from a molecular mechanics force field (FF) as the score, sometimes in combination with solvation or entropy terms;

- o **Empirical**, which use multivariate regression methods to determine coefficients of physically motivated structural functions, which contribute to the binding free energy, by using a training set of ligand-target complexes with experimentally determined binding affinity;

- o **Knowledge-based**, which use atom pair potentials derived from the analysis of structural databases, such as the PDB [12], as the score.

2.3.1.1 Glide

In this thesis work we have used the software Glide [104-106] (Grid-Based Ligand Docking with Energetics), developed by Schrödinger® which allows to dock "flexible" ligands within the binding site of the "rigid" structure of a known receptor (proteins or DNA fragments). The procedure implemented in Glide consists of three different steps: ***protein preparation***, ***grid generation*** and ***docking***.

The ***protein preparation*** is a crucial step for the accuracy of the subsequent docking protocol. In this stage the three-dimensional structure of the receptor is partially optimized through the neutralization of the amino acids not involved in salt bridges, the introduction of hydrogen atoms and the removal of any steric contacts (clashes) between the residues.

During the ***grid generation***, the program builds a regular lattice of points around the receptor which describe the shape and the properties of the binding site in terms of vdW and electrostatic force fields. In this stage two different grids are defined, referred to as *bounding box* or *ligand center box* or *inner box* and *enclosing box* or *outer box* which represent the boxes within which all ligand atoms must be contained. The two boxes share a common center (Fig.2.1).
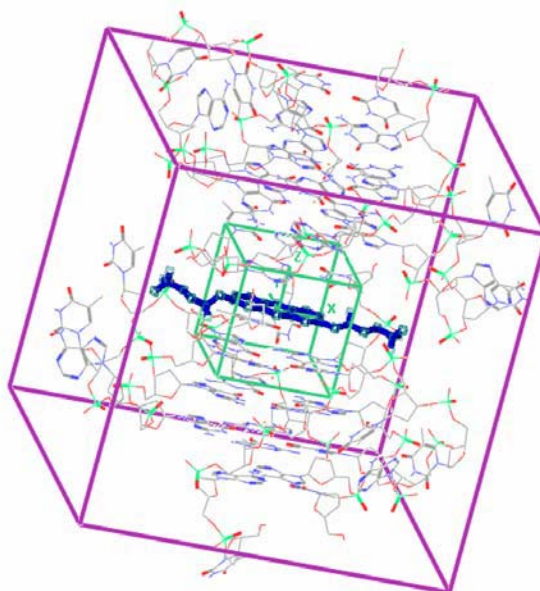
Fig.2.1 Graphical representation of the *bounding box* (coloured in green) and the *enclosing box* (coloured in purple). In the picture is represented the complex between an acridine derivative and a G-quadruplex DNA structure, considered in this thesis work.

The *bounding* (internal) *box* surrounds the area where the centroid of the ligand can move during the docking calculation, while the *enclosing* (external) *box,* which contains the *bounding box,* defines the region where the program calculates the vdW and electrostatic fields and represents the physical space where all the atoms in the ligand must be included during the docking. The residues of the receptor that are outside the enclosing boxes are not considered in the calculation.

The ***docking*** stage is characterized, in turn, by a series of steps. In a first step (*conformation generation*), Glide makes an exhaustive conformational search within the space defined within the enclosing box in order to assess the different conformations of a ligand and to eliminate those with high energy, considered unsuitable for the binding with the receptor. In this stage, the program treats each ligand as formed by a core and a rigid set of "rotamer groups" attached to this core by rotatable bonds; the various rotamers are defined as fragments which do not contain any rotatable bonds. A set of starting conformations for the "core" are thus generated, whose number generally depends on the conformational flexibility of the system, and the entire molecule is finally reconstructed by combining together all these conformations of the core with the various rotamers previously defined. At this point in order to find the best orientation of a ligand within the binding site of the receptor, a series of hierarchical filters are applied to the conformations previously obtained, according to the following scheme (Fig.2.2).
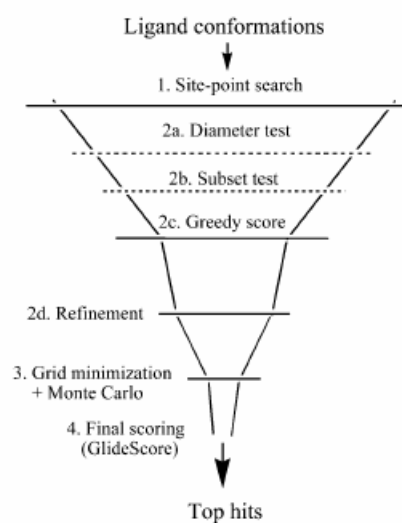
Fig.2.2 A schematic representation of the hierarchical filters applied by the software Glide.

After these steps which contribute to the pose generation, the program evaluates the affinity of each pose for the binding site (*final scoring*), using an empirical scoring function, the GlideScore or ligand-receptor interaction energy, calculated as the sum of molecular mechanics terms which account for hydrophobic, coulombic, vdW interactions, interactions with metals and hydrogen bonds. The best poses for a ligand, correctly docked into the binding site of the receptor, are finally chosen using the value of the E-model, another scoring function which combines the GlideScore with the ligand internal strain energy.

2.3.2 Molecular Dynamics Simulations

Molecular Dynamics (MD) simulation is a powerful computational technique to investigate, at the atomistic level, the dynamic behaviour of biological macromolecules in solution. This methodology allows the prediction of chemical and physical properties of these biomolecules and is currently used in the field of drug-design. [107, 108] It is essentially based on the assumption of an inter-particle potential (*V*), a function of the positions of the nuclei which represents the potential energy of the system, and on a stepwise numerical integration of the classical Newtonian equation of motion:

$$F_i = m_i a_i \quad (eq.1)$$

for each atom *i* in the system. Here m_i is the atom mass, $a_i = d^2 r_i/dt^2$ its acceleration, and F_i the force acting upon it, due to the interactions with other atoms. These forces F_i can be derived as the gradients of the potential *V* with respect to atomic displacements:

$$F_i = -\nabla_{r_i} V(r_1, r_2, ..., r_N) \quad (eq.2)$$

Molecular dynamics, based on these "classical" equations, is a deterministic technique: given an initial set of positions and velocities, the subsequent time behaviour is in principle completely determined. Nowadays, computer technology allows to perform routinely large-scale MD simulations of thousands particle biomolecular systems even on personal workstations.

MD simulations performed *in silico* are in many respects very similar to real experiments performed in a laboratory. First, the macromolecular system has to be correctly prepared. Then, through the solution of Newton's equation of motion, the system is equilibrated until its properties no longer change with time. Finally, after this equilibration stage, the actual measurement of physical properties is performed, in a situation very close to experimental conditions. The result of this computer experiment is a trajectory, which specifies the set of molecular conformations assumed by the macromolecular system during the simulation as a function of time. Typically, many interesting biophysical events occur on a time scale which goes from nanoseconds (ns) to milliseconds (ms). In order to obtain reliable results we need to simulate the biological system for a comparable time; typical "standard" MD simulations are currently carried out on biological systems for a time ranging from ns to microsecond (µs). Besides the long time of simulation, when dealing with biomolecular systems in solution, due to their polar and highly charged nature, another major challenge has to be faced, i.e. the calculation of long-range electrostatic interactions.

In this thesis work, MD simulations have been performed by using the program ORAC [109-112] and the AMBER9 suite. [113] Both the program ORAC and AMBER9 employ a similar potential function (V) to describe the interaction among solute atoms.

Following the implementation of the force field in the program ORAC, the potential V can be written as the sum of two contributions:

$$V = V_{bonded} + V_{nonbonded} \quad \text{(eq.3)}$$

The "bonded" part of the potential (V_{bonded}) is responsible for the flexibility of the system. The "non bonded" term ($V_{nonbonded}$) is dominated by electrostatic interactions. The interaction bonded potential is a function of the stretching, bending and torsion internal coordinates and has the following form:

$$V_{bonded} = V_{stretching} + V_{bending} + V_{torsion} \quad \text{(eq.4)}$$

Its explicit form is given by:

$$V_{bonded} = \sum_{bonds} K_r (r - r_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{proper-dihedrals} K_\phi [1 + \cos(n\phi - \gamma)] \quad \text{(eq.5)}$$

where, K_r and K_θ are the bonded force constants associated with bonds stretching and angles bending respectively, while r_0 and θ_0 are their respective equilibrium values. In the torsion potential, $V_{torsion}$, ϕ is the dihedral angle, while K_ϕ, n and γ are constants. In addition to this normal or *proper* torsion potential, in the bonded potential, there is another term which takes into account the so-called *improper* torsions and has the following form:

$$V_{i-tors} = \sum_{improper-dihedrals} K_\chi [1 + \cos(n\chi - \gamma)] \quad \text{(eq.6)}$$

where χ is the improper dihedral angle, while K_χ, n and γ are constants.

The "non-bonded" potential can be written as:

$$V_{non-bonded} = V_{vdw} + V_{14} + V_{el} \quad \text{(eq.7)}$$

where the V_{vdw} term is the Lennard-Jones potential, given by:

$$V_{vdw} = \sum_{ij} (A_{ij} / r_{ij}^{12}) - (B_{ij} / r_{ij}^6) \quad \text{(eq.8)}$$

The term V_{14} is a nonbonded interaction, typical of biomolecular force fields, between two atoms separated by three consecutive bonds; this interaction is composed of an electrostatic and a Lennard-Jones term. Finally, the term V_{el} in eq.7 includes all long-range electrostatic interactions and is the most difficult to handle in computer simulations of charged particles.

The standard Ewald summation method and its different implementations, such as the particle mesh Ewald (PME) procedure or the smooth particle mesh Ewald (SPME) method, [114-118] are used both in ORAC and in AMBER9 to treat long-range electrostatic interactions. These methods give the exact result for the electrostatic energy of a periodic system consisting of an infinitely replicated neutral

box of charged particles and thus are the natural choice in MD simulations of complex molecular systems with periodic boundary conditions (PBC).

Standard *equilibrium* simulations of complex biosystems are usually performed, both in ORAC and in AMBER9, by using PBC to correctly simulate a macromolecular system in an environment similar to the real one and without introducing boundary effects.

When using PBC, the macromolecular system is enclosed in a *unit cell* of suitable geometry (box) replicated to infinity by rigid translation in all the three Cartesian directions, completely filling the space. The different shapes of the box normally employed in MD simulations are cubic or rectangular, truncated octahedral and rhombic dodecahedric. The dimension of these simulation cells is very large in order to set to zero the interactions among replicated images of the system. Such a system, single solvated macromolecule in PBC, is thus representative of dilute solution of biomolecules since the solute molecules in the periodic systems can never come close to each other, thereby interacting. Moreover, PBC avoid that water molecules are in contact with the vacuum, giving rise to boundary effects.

Focusing on the program ORAC, its other features are the use of a reversible and symplectic multiple time step integration algorithm (or r-RESPA, reversible reference system propagation algorithm) [119-122] for the numerical solution of equations of motions and the possibility to perform MD simulations in different thermodynamic ensembles, by using the extended Lagrangian (EL) methods to maintain constant pressure and constant temperature.[123-125] The fundamental idea of EL methods is to represent the effect of a suitable external reservoir by adding new degrees of freedom to the system. ORAC can perform simulations in the microcanonical (NVE), canonical (NVT), isobaric (NPH), and isothermal-isobaric (NPT) ensembles. All these advanced computational techniques implemented in the program allow very efficient simulations of biomolecular systems.

2.3.3 Advanced Sampling Methods

Efficient and extensive sampling is a fundamental requirement for successful protein simulations, but large barriers between different local energy minima in rugged free energy landscapes of biological systems make it difficult to explore large areas of the

conformational space using standard MD simulations at moderate temperatures. In general, swaps between these conformers can take, in average, as long as a few microseconds.[126] Thus, in a standard MD simulation, the system may remain trapped during the whole computationally accessible simulation time in a local minimum, and the "rare event" of escaping the trap never happens. To overcome this sampling problem, a variety of advanced simulation techniques have been developed; one of these is the Replica Exchange Method (REM), also known as Parallel Tempering. [127-130]

2.3.3.1 The Hamiltonian Replica Exchange Method (Hamiltonian REM) as implemented in ORAC [112]

In its standard implementation (*Temperature REM*), Replica Exchange Method employs a set of independent and simultaneous MD simulations of the same system at different temperatures (replicas), allowing consecutive simulations (i, i±1) to exchange complete configurations of the system at regular intervals and using a probal
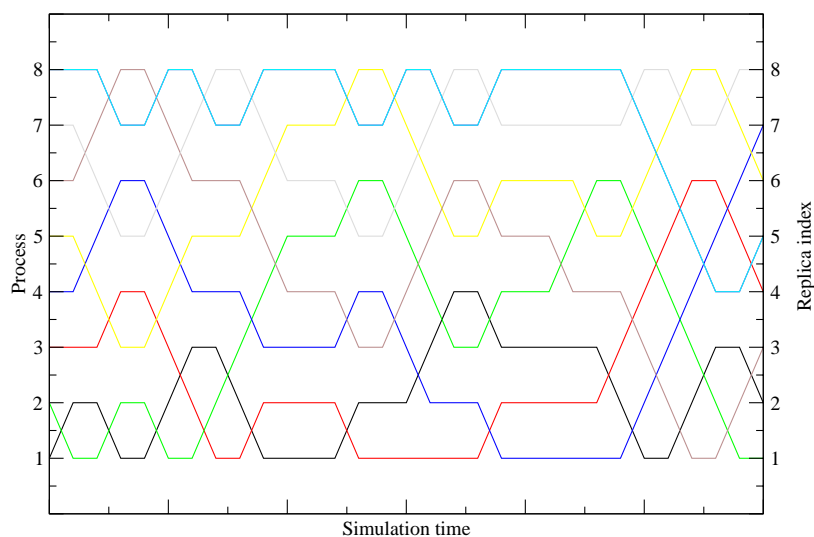

Fig.2.3 Typical REM simulation with 8 replicas. Each colour represents a process running in parallel.

Generally, the range of employed temperatures goes from the temperature of interest T_0 to a higher temperature T_M, high enough to accelerate the crossing of local conformational barriers and to allow the sampling of metastable conformers. The

production of these independent simulations can benefit from the use of large clusters of CPUs where the different replicas are performed in parallel. The thermodynamical conditions of these simultaneous trajectories are chosen so as to span homogeneously the thermodynamic space from the ensemble of interest to a different ensemble with enhanced transition rates, where the sampling is ergodic. Due to the swapping between different replicas, a trajectory is no longer bound to an unique given equilibrium ensemble but can randomly walk in a thermodynamic space of different equilibrium conditions, visiting ensembles where an ergodic sampling is possible, and then going back to the quasi-ergodic ensemble of interest. The gain in sampling efficiency with respect to a series of uncoupled parallel trajectories comes from the exchange of information between trajectories, and the replica exchange process is the tool by which "information" (e.g. a particular configuration) is carried, for example, from a high to a low temperature.

In this thesis work we employed a more recent version of this computational method, the so-called Hamiltonian REM, [131] as implemented in ORAC. [112] In the Hamiltonian REM, the same result of an increasing of the temperature is obtained through an inverse scaling of the interaction potential energies of the system. In its simplest implementation each replica is so characterized by a different potential energy, according to the equation:

$$V_i(X) = c_i V(X) \quad \text{(eq.9)}$$

where $V_i(X)$ denotes the potential energy of the ith replica and $V(X)$ the original potential perceived by the system.

One of the main advantages of using the Hamiltonian REM, also employed in this thesis work, is the possibility to apply the scaling factor c_i only to a specific part of the potential, weakening only the interactions involving interesting regions of the simulated system. This specific approach is referred to as "solute tempering". Under these conditions the portion of interest is defined as "the solute", while the remaining regions of the simulated system are referred to as "the solvent".

According to this subdivision, the potential energy of a given macromolecular system in solution can be written as the sum of different contributions:

$$V(X) = V_{solute}(X) + V_{solute-solvent}(X) + V_{solvent}(X) \quad \text{(eq.10)}$$

where V_{solute} includes all the solute-solute interactions, while the $V_{solute-solvent}$ term includes the interactions between the solute and the solvent and $V_{solvent}$ all the solvent–solvent interactions. In this "solute tempering" implementation of the Hamiltonian REM it is possible or (1) to scale all the interactions involving the solute, according to the equation:

$$V_i(X) = c_i \cdot (V_{solute} + V_{solute-solvent}) + V_{solvent}(X) \quad \text{(eq.11)}$$

where $V_i(X)$ is the potential of the *i*th replica, or (2) to proceed to a further subdivision of the potential involving the solute and then to scale only specific terms of this potential.

In this second approach, employed in this thesis work, the solute-solute and the solute-solvent interactions are treated separately; these two terms are generally referred to as intra-solute and inter-solute potentials. In this work, we focused on the intra-solute potential, which in turn can be written as the sum of various contributions; each of these terms can be scaled in the *i*th replica according to the following equation:

$$\begin{aligned} V_{i(\text{int}ra-solute)}(X) = &c_{i1}(V_{stretching} + V_{bending} + V_{improper-torsion} + V^h_{proper-torsion}) \\ &+ c_{i2}(V_{proper-torsion} + V_{14}) + c_{i3}(V_{vdw} + V_{electrostatic}) \end{aligned} \quad \text{(eq.12)}$$

The first term of the sum, in the eq.12, collects all the potential terms generating "fast" motions: stretching ($V_{stretching}$), bending ($V_{bending}$) and improper torsional interactions ($V_{improper-torsion}$). This term includes also proper torsions involving hydrogen atoms ($V^h_{proper-torsion}$). The second term in eq.12 collects all the other proper torsional interactions ($V_{proper-torsion}$) and the so-called 1–4 interactions (V_{14}). Finally, the third term of eq.12 collects all the nonbonded interactions: the Lennard-Jones potential (V_{vdw}) and the electrostatic potential ($V_{electrostatic}$). In this context the typical choice, also employed in this thesis work, is to set in eq.12 $c_{i1} = 1$ for all replicas, as there is little advantage for conformational sampling in exchanging

configurations involving stiff degrees of freedom such as bendings, stretchings and improper torsions. On the other hand, conformational transitions in proteins are mainly driven by torsional and non bonded interactions, generating "slow" motions. The approach employed in this work was to "heat up" these degrees of freedom by scaling the corresponding potential functions in eq.12, by using $c_{i2} < 1$ and $c_{i3} < 1$.

2.3.4 Free Energy Methods

The "free energy methods" are computational tools which allow an accurate estimation of the difference in free energy ($\Delta G_{binding}$) between the two biologically relevant states involved in the binding process, corresponding to the ligand and the macromolecule in their free states (R and L) and to the resulting complex (RL). [132, 133] This binding free energy is very important in the field of drug discovery since it is related to the affinity of a compound, according to: $\Delta G_{binding} = -RT \ln(K_{affinity})$. In this thesis work $\Delta G_{binding}$ values were computed by the method referred to as Molecular Mechanics - Poisson Boltzmann Surface Area (MM-PBSA), [134] as implemented in AMBER9.

2.3.4.1 Molecular Mechanics - Poisson Boltzmann Surface Area (MM-PBSA) method

This computational approach for the evaluation of $\Delta G_{binding}$ employs a series of energy terms calculated on the basis of molecular mechanics and by implicit solvent models. In these calculations, it uses a set of conformations, for the ligand and the target alone and for their complex, taken from the output of a single MD simulation carried out on a target-ligand complex in explicit solvent. An average $\Delta G_{binding}$ is thus evaluated combining the values of the free energy of binding calculated for all the snapshots taken from the MD trajectory; in this way it is also possible to take into account the mutual adaptation ligand-target which occurs during the MD simulation and referred to as *induced-fit*.

The value of $\Delta G_{binding}$ for each snapshot is calculated according to the following equation:

$$\Delta G_{binding} = G_{complex} - (G_{target} + G_{ligand}) \quad (eq.13)$$

where the values of the free energy (G) for the ligand and the target alone and for the complex are the sum of different terms.

For each molecular species involved in the equilibrium, G is given by:

$$G = E_{MM} + G_{psolv} + G_{npsolv} - TS \quad (eq.14)$$

In the eq.14, E_{MM} is in turn the sum of several energy terms calculated on the basis of molecular mechanics ($E_{bond} + E_{angle} + E_{torsion} + E_{vdW} + E_{electrostatics}$).

G_{psolv} is the polar contribution to the solvation energy of the molecule, which is evaluated by applying a model of implicit solvent. In this thesis work, this term was calculated through the solution of the Poisson Boltzmann equation (PBE).

G_{npsolv} is the non-polar contribution to solvation energy, which depends on the solvent-accessible surface area (SASA), according to the equation:

$$G_{npsolv} = \gamma SASA + \beta \quad (eq.15)$$

where γ is the surface tension and β a coefficient.

The last term of the eq.14 accounts for the entropic contribution. Here, T is the absolute temperature and S is the entropy term calculated with the normal mode analysis (NMA). [135]

3. Experimental section

3.1 COMPUTATIONAL INVESTIGATION OF THE KDR DOMAIN

3.1.1 Molecular Modeling investigation on the closed conformation of the KDR domain

The starting coordinates for the KDR domain of the human VEGFR2 were obtained from the crystal structure of this target in a complex with a benzimidazole-urea inhibitor, refined at 2.05Å resolution (pdb code 2OH4 [16]). This X-ray three-dimensional structure is the only one, among those available in the PDB for the KDR domain, with the A-loop completely solved. This structure lacks, however, a fragment belonging to the kinase insert domain (KID) and corresponding to the residues ranging from Tyr936 to Tyr994 (according to the 2OH4 numbering system). Moreover, this structure, although double phosphorylated in two tyrosine residues (Tyr(P)1052 and Tyr(P)1057), has the A-loop in an inhibitory conformation and the DFG motif in an "out" orientation.

Computational studies require a suitable starting structure for: (1) the removal of all non-protein units, including the inhibitor and all the water molecules, from the crystal structure, (2) the selection of only one orientation for these residues (Cys860, Cys1043, Leu1067, Glu1156) with two alternative positions emerged from X-ray crystallography, (3) the reconstruction of the residues having missing atoms and (4) the assignation of the most probable protonation state to the histidine residues according to the environment at pH 7 (histidine tautomers with the hydrogen on the ε or δ nitrogen or double protonated). Moreover, the coordinates of the first residue (His814), not completely solved, were removed. All these operations were performed by the Schrödinger® Protein Preparation Wizard routine. [136]

Thus, the three-dimensional structure of the KDR domain employed in the computational study begins with the residue Cys815 and ends with the residue Asp1169, according to the 2OH4 numbering system.

In a second stage of this protein preparation, we tried to solve the problem concerning the missing region in the crystal structure 2OH4 and corresponding to 57 residues of the KID. It has been demonstrated [15] that the removal of these residues is necessary for protein crystallization but has no effect on the intrinsic kinase

activity. This region is, however, structurally important because it connects two α-helices in the C-terminal lobe and maintains the overall architecture of the KDR domain. Starting from these observations, we decided to reconstruct this not solved region (residues 936-994) by modifying the primary sequence of the KDR domain through the introduction of a short loop connecting the two solved ends, instead of modeling the whole KID. This reconstruction was performed by using the routine Build Loop of Swiss-PdbViewer. [137]

We introduced into the three-dimensional structure several loops of different length, in order to find the best "*hinge*" between the two solved α-helices in the C-terminal lobe. We chose to introduce only alanine loops because of the small and mostly inert methyl functional group of this residue. The best result was found by inserting a loop formed by five residues, according to a set of geometrical and energetic parameters provided by the routine Build Loop, such as a count of the clashes (number of bad contacts or H-bonds), an energy information (computed with a partial implementation of the GROMOS96 force field [138]) and a mean force potential value (computed from a "Sippl-like" mean force potential [139]).

After the introduction of this alanine loop into the three-dimensional structure, the number of residues of the KDR domain has changed compared to the starting crystal structure 2OH4. As a consequence, we decided to renumber the overall protein starting from 1, corresponding to the residue Cys815 in the structure 2OH4, to 303, corresponding to Asp1169. The sequence of this partially reconstructed structure is displayed in the Fig.3.1.1.
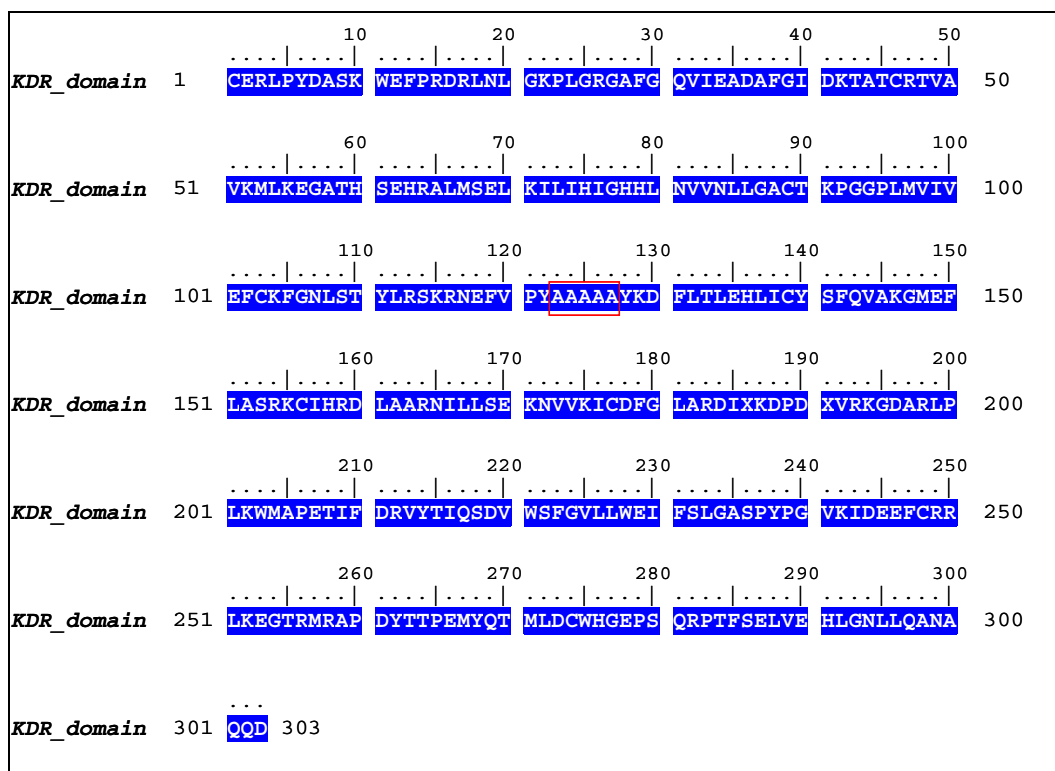
Fig.3.1.1 Primary sequence of the KDR domain after the introduction of
the alanine loop in the C-terminal lobe; this alanine loop is highlighted in
red.

According to this new numbering system employed for the KDR domain, the alanine
loop, introduced into the KID, starts with the residue Ala123 and ends with the
residue Ala127. The A-loop starts with the DFG motif, which corresponds to the
residues Asp178, Phe179, Gly180 and ends with the APE motif, which corresponds
to the residues Ala205, Pro206 and Glu207. The two phosphotyrosines, Tyr(P), into
the A-loop, identified as X in the above primary sequence, correspond to the residues
Tyr(P)186 and Tyr(P)191. Some important correspondences with the crystal
structure 2OH4 are listed in the Table.3.1.1.

Table.3.1.1 KDR important regions and correspondences between the new numbering system and the 2OH4 numbering system. One letter code is employed for amino acids; the two phosphotyrosines Tyr(P) are indicated as Yp.

Region	Primary sequence	Residue number	Structure 2OH4
Gly-rich loop	GRGAFG	25-30	839-844
β3 strand	VAVKMLK	49-55	863-869
αC-helix	HSEHRALMSELKILIHI	60-76	874-890
hinge region	FCKFGN	102-107	916-921
catalytic loop	HRDLAARN	158-165	1024-1031
A-loop	DFGLARDIYpKDPDYpVRKGDARLPLKWMAPE	178-207	1044-1073
	DFGLA (magnesium binding loop)	178-182	1044-1048
	YpKDPDYp (activation segment)	186-191	1052-1057

Since the original crystal structure 2OH4 is unusual among kinase structures having the A-loop in an inhibitory conformation although double phosphorylated, we also decided to investigate the dynamic behaviour of the corresponding structure with the two phosphotyrosines mutated into tyrosine residues (Tyr186 and Tyr191). This unphosphorylated form of the KDR domain is actually present *in vivo* when the whole receptor is inactivated. This three-dimensional structure of the target in the inactive conformation was chosen as the starting point for the investigation of the opening mechanism in the KDR domain.

3.1.2 Building of the open conformation of the KDR domain

In order to obtain structural information on the other limit conformation of the KDR domain (i.e. the active and *DFG-in* conformation whose 3D coordinates are not available), we decided to build a "chimeric" open structure of this target by using a homology modeling approach. We built a three-dimensional model for the KDR domain in its active conformation, starting from the observations that (1) the active state of the kinase domain is essentially invariant among RTKs and that (2) the main

difference between the two limits conformations is the orientation of the A-loop. [35, 36]

We used the structure of the kinase domain of the insulin receptor, crystallized in its active conformation and with the DFG motif in the "in" form (pdb code 1IR3 [38]), as template. The primary sequences of the KDR domain and of the kinase domain of the insulin receptor were aligned using ClustalW [140] (BLOSUM62 similarity matrix) multiple global alignment, implemented in BioEdit. [141] The resulting alignment is shown in the Fig.3.1.2.
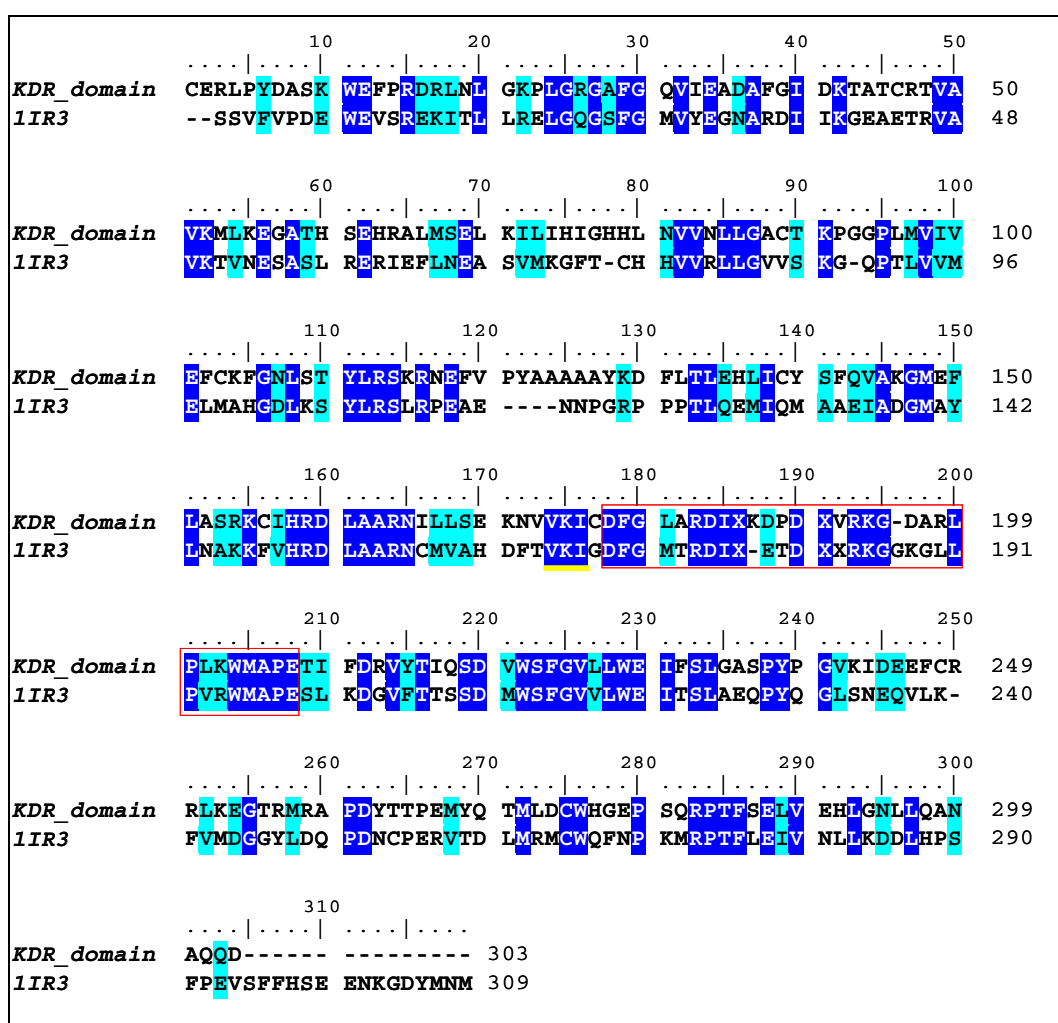
```
                10        20        30        40        50
       ....|....| ....|....| ....|....| ....|....| ....|....|
KDR_domain  CERLPYDASK WEFPRDRLNL GKPLGRGAFG QVIEADAFGI DKTATCRTVA  50
1IR3        --SSVFVPDE WEVSREKITL LRELGQGSFG MVYEGNARDI IKGEAETRVA  48


                60        70        80        90        100
       ....|....| ....|....| ....|....| ....|....| ....|....|
KDR_domain  VKMLKEGATH SEHRALMSEL KILIHIGHHL NVVNLLGACT KPGGPLMVIV  100
1IR3        VKTVNESASL RERIEFLNEA SVMKGFT-CH HVVRLLGVVS KG-QPTLVVM  96


                110       120       130       140       150
       ....|....| ....|....| ....|....| ....|....| ....|....|
KDR_domain  EFCKFGNLST YLRSKRNEFV PYAAAAAYKD FLTLEHLICY SFQVAKGMEF  150
1IR3        ELMAHGDLKS YLRSLRPEAE ----NNPGRP PPTLQEMIQM AAEIADGMAY  142


                160       170       180       190       200
       ....|....| ....|....| ....|....| ....|....| ....|....|
KDR_domain  LASRKCIHRD LAARNILLSE KNVVKICDFG LARDIXKDPD XVRKG-DARL  199
1IR3        LNAKKFVHRD LAARNCMVAH DFTVKIGDFG MTRDIX-ETD XXRKGGKGLL  191


                210       220       230       240       250
       ....|....| ....|....| ....|....| ....|....| ....|....|
KDR_domain  PLKWMAPETI FDRVYTIQSD VWSFGVLLWE IFSLGASPYP GVKIDEEFCR  249
1IR3        PVRWMAPESL KDGVFTTSSD MWSFGVVLWE ITSLAEQPYQ GLSNEQVLK-  240


                260       270       280       290       300
       ....|....| ....|....| ....|....| ....|....| ....|....|
KDR_domain  RLKEGTRMRA PDYTTPEMYQ TMLDCWHGEP SQRPTFSELV EHLGNLLQAN  299
1IR3        FVMDGGYLDQ PDNCPERVTD LMRMCWQFNP KMRPTFLEIV NLLKDDLHPS  290


                310
       ....|....| ....|....
KDR_domain  AQQD------ ---------  303
1IR3        FPEVSFFHSE ENKGDYMNM  309
```

Fig.3.1.2 Sequence alignment between the partially modified KDR domain and the kinase domain of the insulin receptor (crystal structure 1IR3). Amino acid residues conserved or similar in the two sequences are highlighted in blue and in cyan, respectively. The A-loop is highlighted in red. The phosphotyrosines of the A-loop, Tyr(P), are identified by X in the two sequences.

This sequence alignment revealed a high degree of similarity for the A-loop in the two proteins (Fig.3.1.2): (1) same number of residues between the two common end motifs, DFG and APE, (2) same position of the two phosphotyrosines (X in the primary sequences) of the KDR domain compared with two, of the three, phosphotyrosines of the insulin receptor and (3) sequence identity in the residues at the beginning (<u>VKI</u>-**DFG)** and at the end (WM**APE**) of the A-loop.

Starting from these results, we built a "chimeric" structure of the KDR domain in its active conformation, in which the coordinates for the backbone of the A-loop were obtained from the crystal structure 1IR3. The remaining residues for the "chimeric" structure were taken from the previously modified crystal structure of the KDR domain in the closed and *DFG-out* conformation (pdb code 2OH4), with the alanine loop inserted into the KID.

Although, from a structural superposition of the two three-dimensional structures (closed and *DFG-out* conformation of the KDR domain and open and *DFG-in* conformation of the InsR kinase domain) other elements show a slight different orientation, we decided to not introduce any other artificial element in this "chimeric" structure of the KDR domain, but to investigate its reliability and the possible reorganization of the different structural elements by using standard MD simulations.

In detail, the homology model was generated first by inserting the A-loop of the crystal structure 1IR3, between the motif VKI (underlined in yellow in the Fig.3.1.2) and the motif APE, into the previously modified structure of the KDR domain in its closed conformation, and then by modifying the side chains of the different amino acids according to the KDR primary sequence using the program Swiss-PdbViewer (MUTATE command). [137] Similarly to the closed conformation of the KDR domain, we modified the two phosphotyrosines of the A-loop into tyrosine residues (Tyr186 and Tyr191) to avoid the not clear effect of the phosphorylation.

3.1.3 Molecular Dynamics Simulations setup
Explicit solvent MD simulations on the previous modified structures of the KDR domain have been performed by using the program ORAC, [109, 112] the AMBER

ff03 [142] force field and the TIP3P water model. [143] The force field parameters for the two non-standard phosphotyrosines (Tyr(P)186 and Tyr(P)191) were adapted from the work of Homeyer et al. [144] The phosphorylated and unphosphorylated structures were then solvated by approximately 9100 water molecules in a rhombic dodecahedron box with periodic boundary conditions. The systems were initially equilibrated during a 100 ps NPT simulation at T = 300 K and P = 1 atm, where constant pressure was obtained using a modification of the Parrinello-Rahman Lagrangian [110] and temperature control was achieved using a Nosé thermostat. [125] The cell side was equilibrated at an average value of 77Å. The Ewald method with the smooth particle mesh algorithm [117] was used to compute electrostatic interactions. The grid spacing in each dimension of the direct lattice was $\sim$ 1.2Å, whereas the Ewald convergence parameter was set to 0.43Å^{-1}. Electro-neutrality was enforced using respectively a uniform positive and negative charge density background equalizing the -1 e and +3 e excess negative and positive charges on the phosphorylated and unphosphorylated structures. A multiple time-step r-RESPA algorithm [119, 121] with a potential subdivision specifically tuned for proteins was used for integrating the equations of motion, with time-steps 9.0, 3.0, 1.5 fs for non bonded and 0.75, 0.375 fs for bonded interactions. [109, 110, 118] The last configurations, obtained from these simulations performed on the unphosphorylated open (*DFG-in*) and closed (*DFG-out*) conformations of the target, were used as starting points for the production runs performed in the same conditions for 5 ns.

3.1.4 Hamiltonian Replica Exchange (Hamiltonian REM) simulations

The last configuration obtained from the previous equilibration of 100 ps carried out on the crystallographic derived structure of the KDR domain, in the closed and *DFG-out* conformation and with the two tyrosines phosphorylated (Tyr(P)) in the A-loop, was used as a starting point for the REM simulations performed by using the program ORAC. The Hamiltonian REM [131] implementation of the REM algorithm was adopted, in which the replicas of the system differ in their potential energy function *V*. In the Hamiltonian REM formalism, the potential of the *i*th replica is given by $V_i = c_i V_0$ where V_0 is the original potential of the system and c_i is a multiplicative factor that varies with the replica number. In this study, the application

of this methodology, in the "solute tempering" version, was limited to the residues belonging to the A-loop (the so-called "solute"), also including some residues located before the important DFG motif (residues 178-180) to better investigate its dynamic behaviour. The "solute" in our Hamiltonian REM simulations included the residues between Ile176 and Glu207 of the KDR domain (I1042-E1073 according to 2OH4 numbering). In detail, we applied the multiplicative factors only to intra-loop potential terms (torsional and non bonded interactions are weakened; paragraph 2.3.3.1). Other interactions, including those between the A-loop and the rest of system, are not affected by scaling factors. In this context, we used the knowledge of the fact that the A-loop is the most flexible region of the target, while the conformational changes in other region are more limited.

In detail, the multiplicative factors were chosen such that the effective temperatures of the 42 replicas are distributed over the temperature range from 300 K to 3000 K, with temperature progression according to the scheme given in ref [131]. Each replica was equilibrated with a different potential V_i for 5.4 ps in the NPT ensemble. Then, exchanges were attempted every 270 fs. Replica exchanges were (on average) accepted with a 60 % probability, leading to an average time of $\sim$ 450 ps between two consecutive accepted exchanges. The system was simulated for 4.5 ns, for a total simulation time of 189 ns.

3.1.5 Hot simulations

Because of the high computational demand of the REM methodology, we also performed some targeted simulations, referred to as "hot simulations", at the highest effective temperature $T_i = T_0/c_i$ employed in the previous Hamiltonian REM (3000K). With such simulations we tried to gain a qualitative understanding of the A-loop opening mechanism in the KDR domain, exploiting the accelerated sampling due to the intra-loop potential scaling. It has been shown that the A-loop is a highly flexible portion of the kinase domains, such that the closed and open loop configurations are partially stabilized by salt bridges and hydrophobic contacts with the rest of the protein. [34] We exploit this fact by altering the loop-loop interactions only, while maintaining all the other interactions unaltered. We remark that older studies using "hot" accelerated simulations of kinases were conducted heating up all

the system and applying harmonic restraints to atoms outside the kinase A-loop. [145] This approach, however, leads to weaker also solvent-solvent and hydrophobic interactions, especially between the A-loop and the rest of the protein. Our approach, on the contrary, permits to fully observe the effect of hydrophobicity on the conformational landscape of the A-loop and on the opening/closing mechanism.

In detail, we performed 16 targeted "hot simulations", starting from the closed and *DFG-out* conformation of the KDR domain in the non phosphorylated form, to avoid the unclear effect of the phosphate groups in the movement of the A-loop. All the "hot simulations" were performed for 10 ns. In these simulations, at variance with the Hamiltonian REM simulations discussed in the previous paragraph, the side chains of Asp160 and Arg164 were included in the definition of the "solute", to prevent the formation of loop-protein salt bridges, as it will be explained in detail in paragraph 4.1.4.

3.2 MODELING INVESTIGATION OF DOUBLE HELIX AND G-QUADRUPLEX DNA STRUCTURES IN COMPLEX WITH ACRIDINE-BASED POLYAMINE LIGANDS

3.2.1 Solution studies [98]

The protonation states of ligands L1, L2 and L3 (Fig.3.2.1) were investigated by potentiometric and ^{1}H NMR measurements. In all three cases, the diprotonated form $[H_2L]^{2+}$ (L = L1, L2 or L3) is the most abundant species in aqueous solution at physiological pH. 1H NMR investigations indicated that the two first protonation steps occur on the methylated nitrogen atoms in the case of L1, and on the central nitrogen atoms of the tetra-amine chain in the case of L2 and L3.
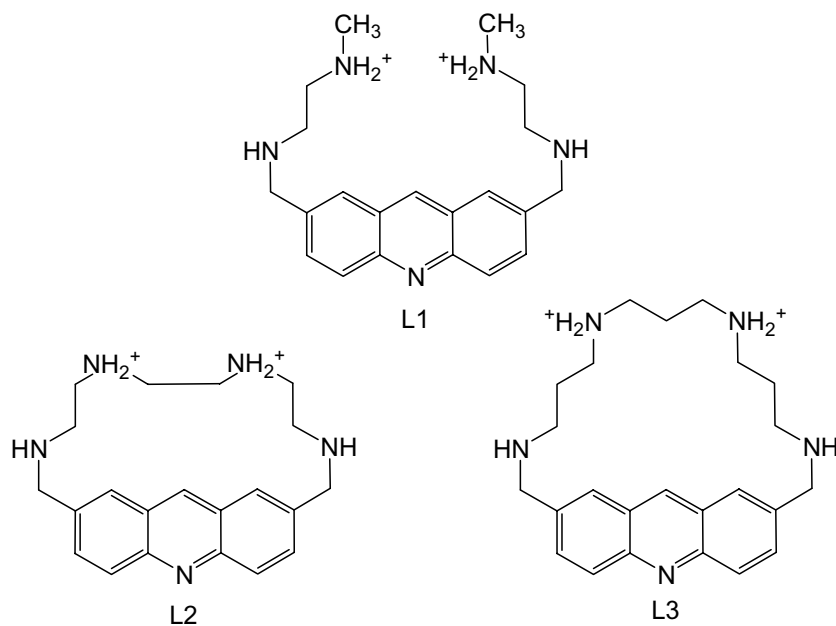

Fig.3.2.1 Acridine-based polyamine ligands considered in this thesis work (L1, L2 and L3) with the correct protonation state at physiological pH.

3.2.1.1 Acridine derivatives binding to dsDNA [98]

Because the affinity of a ligand for DNA can be affected by the DNA sequence, the binding process of the three ligands was preliminarily monitored towards oligonucleotides of different base composition, namely poly(dA-dT) and poly(dG-

dC). It was found that the binding to the tested DNA sequences is essentially equal, thus confirming a lack of sequence selectivity.

All the experimental studies performed in solution allowed us to draw a picture of the dsDNA binding pattern for these acridine derivatives (L1, L2 and L3). Although all ligands efficiently interact with DNA, the nature of the side chains plays a key role in the binding mode. In detail, experimental results seem to suggest a preferential, non-intercalative binding mode for L3. In this case, the driving force is mainly the electrostatic interaction of the charged ligand with the nucleic acid, which promotes DNA precipitation, observed under the experimental conditions employed. In contrast, L1 and L2 showed similar behaviour, thus suggesting a more favorable balance of the $\pi-\pi$ interactions between the acridine moiety and the base pairs, on the one hand, and the H-bond and salt bridge interactions, supported by the charged substituents, on the other.

It is well known that small macrocycles such as L2 and L3, which contain a rigid aromatic moiety in their structures, often adopt bent conformations. As a consequence, one could easily propose that the insertion of the acridine moiety between DNA base pairs would be more difficult in the case of macrocycles L2 and L3 than for L1. Surprisingly, whereas no such intercalation seems to be present with L3, L2, which contains an overall shorter cyclic aliphatic chain, is still able to interact with dsDNA in a manner similar to the acyclic L1.Thus, to gain a more in-depth insight, we decided to carry out a modeling investigation.

3.2.1.2 Acridine derivatives binding to G-quadruplex forming sequences [98]

The potential interaction of tested ligands (L1, L2 and L3) with G-quadruplex structures was evaluated by monitoring both the stabilization and the induction of G-quadruplex structures assumed by telomeric sequences. Experimental studies clearly illustrated how the stabilization efficiency of the three novel ligands is modulated by the nature of their side chains. In particular, L2 appeared to be the most active in stabilizing a G-quadruplex structure, followed by L3 and L1. Thus, the similarity between L1 and L2 in the dsDNA binding properties is not conserved when the target sequence assumes a G-quadruplex structure.

Other solution studies were performed in order to determine if this interaction reflects any ability of the new tested compounds to induce the formation of G-quadruplex structures, starting from an oligonucleotide containing only two human telomeric repeats (2GGG). In particular, this sequence can form G-quadruplex structures only by pairing two or four strands, thus leading to dimeric or tetrameric structures. The experimental results showed that, among the tested acridine derivatives, only L2 was able to induce the folding of 2GGG into a tetrameric G-quadruplex structure, thus confirming it as the most efficient G-quadruplex binder. The induction of the tetrameric DNA arrangement for the 2GGG telomeric sequence correlates well with recent crystallographic data showing a related trisubstituted acridine bound to a G-quadruplex [54] with two dimeric units held together by one drug molecule stacked on the terminal G-quartets. Thus this crystallographic structure (pdb code 3CE5 [54]) was considered a suitable model for our computational studies.

3.2.2 Computational studies

3.2.2.1 Building of ligands and DNA fragments

The binding capacity of H_2L1^{2+}, H_2L2^{2+} and H_2L3^{2+} (Fig.3.2.1) toward the telomeric G-quadruplex structure of sequence d(TAG-GGT-TAG-GGT) and toward double-helical DNA oligonucleotides (10-mers) of base composition poly(dG-dC) was investigated. Ligand molecules and dsDNA containing CG or GC intercalative binding sites were built by the Build module of Maestro v. 8.5. [146] Starting coordinates for the tetrameric G-quadruplex were obtained from the biological unit of the BRACO-19-d(TAG-GGT-TAG-GGT) complex X-ray crystal structure (PDB code 3CE5 [54]). The three ligands were built in the ionized state in order to simulate their behavior under physiological conditions.

3.2.2.2 Docking and Molecular Dynamics simulations

Docking calculations were performed using Glide [147] with the DNA structures kept fixed in their original conformation throughout the docking procedures. Selected poses of the ligand-target (both dsDNA and G-quadruplex) complexes were submitted to MD simulation for 10 ns in explicit solvent, and the root mean square

deviation (RMSD) values of the complexes were monitored as a function of simulation time.

The atomic electrostatic charges of the ligands were calculated by means of the RESP procedure, [148] that is, fitting them to an electrostatic potential calculated at the HF/6-31G* level of theory using Gaussian09 software. [149] General Amber force fields (GAFF) parameters were then assigned to the ligands by the *antechamber* module implemented in AMBER9 suite. [113] Each complex was immersed in a truncated octahedral box, the edges of which were located 10Å from the closest atom of the DNA fragments, and which contains ~4700 water molecules for the poly(dG-dC) dsDNA oligonucleotides, and ~5800 water molecules for the G-quadruplex complexes. To maintain neutrality in the system, 16 and 38 K^+ counterions were added to the solvent bulk of the dsDNA-water complexes and to the solvent bulk of the G-quadruplex-water complexes, respectively. In the case of the G-quadruplex structure, according to X-ray data, [54] 4 K^+ ions were placed along the axis within the central core of the complex, midway between each G-tetrad. The positively charged K^+ counterions were added to the solvent bulk in order to maintain the consistency with the crystallization conditions and to perform the simulation in a uniform K^+ ionic environment. Before starting the MD simulations, an energy minimization of the complexes was performed by setting a convergence criterion on a gradient of 0.01 kcal mol^{-1} Å^{-1}. Water shells and counterions were then equilibrated for 40 ps at 300 K, after which 10 ns of MD simulation in an isothermal-isobaric ensemble was performed without any restrain on each complex. The ff03 version of the AMBER force field was used for the DNA fragments and the counterions, [142] whereas the TIP3P model [143] was employed to explicitly represent water molecules. In the production runs, the ligand–DNA fragment systems were simulated in periodic boundary conditions. The van der Waals and short-range electrostatic interactions were estimated within a 10 Å cutoff, whereas the long-range electrostatic interactions were assessed by using the particle mesh Ewald method, [117] with 1 Å charge grid spacing interpolated by fourth-order *B*-spline, and by setting the direct sum tolerance to 10^{-5}. Bonds involving hydrogen atoms were constrained by using the SHAKE algorithm [150] with a relative geometric tolerance for coordinate resetting of 0.00001 Å. Berendsen's coupling algorithms [151] were

used to maintain constant temperature and pressure with the same scaling factor for both solvent and solutes, and with the time constant for heat-bath coupling maintained at 1.5 ps. The pressure for the isothermal–isobaric ensemble was regulated by using a pressure relaxation time of 1 ps in the Berendsen's algorithm. The simulations of the solvated complexes were performed using a constant pressure of 1 atm and a constant temperature of 300 K. A time step of 2 fs was used in the simulations, which were carried out with the AMBER9 program suite. [113]

3.2.2.3 MM-PBSA calculations of $\Delta G_{binding}$

Free energies were calculated by using the MM-PBSA method, as implemented in the AMBER9 program suite. [113] The electrostatic contribution to the solvation free energy was calculated with the nonlinear Poisson Boltzmann method as implemented in the adaptive Poisson Boltzamm solver (APBS) [152] program through the AMBER/iAPBS interface. The hydrophobic contribution to the solvation free energy was determined with terms dependent on solvent-accessible surface area. In these calculations, we used a solvent probe radius of 1.4Å to define the dielectric boundary, a physiological salt concentration of 0.154M to calculate the effect of salt on the free energies, and dielectric constants of 1.0 and 80.0, respectively, for the solute and the surrounding solvent. Atomic charges for DNA fragments and ligands are the same as those employed in the MD simulations. For atomic radii, we applied the PARSE [153] parameter set. PARSE radii (PARameters for Solvation Energy) have been properly optimized for implicit solvent models. In general, the PARSE radii are on average 0.2Å less than the corresponding van der Waals radii. In the case of the G-quadruplex structures, the cations (K^+) present within the negatively charged central channel were also explicitly included in the calculation. The K^+ ion radius was kept at 2.025Å, based on a previous study. [57] Free energies were estimated by collecting snapshots every 40 ps during the last 4 ns of the 10 ns MD simulations.

The entropic contribution was estimated with the normal mode analysis (Nmode [135] module of AMBER9 suite) which computes vibrational, rotational, and translational entropies. The snapshots were minimized in the gas phase for a maximum number of 1×10^5 cycles to give an energy gradient of 1×10^{-4} kcal mol^{-1}

Å^{-1}. These minimized structures are then used to estimate the translational, rotational, and vibrational entropies at 300 K. Because of the extensive computational requirement, the normal mode analyses were performed by considering snapshots collected every 200 ps during the last 4 ns of the 10 ns MD simulations.

Molecular graphics were produced with Mercury [154] and Visual Molecular Dynamics (VMD). [155]

4. Results and Discussion

4.1 COMPUTATIONAL INVESTIGATION OF THE KDR DOMAIN

4.1.1 Preliminary computational studies

The human VEGFR2 is a receptor tyrosine kinase involved in many biological functions, such as angiogenesis and vasculogenesis, and recognized as an attractive target in anticancer therapies. In particular, its inhibition can be achieved through highly potent small organic compounds targeting different conformations of the cytoplasmic kinase domain. However, the structural data now available for the KDR domain of the human VEGFR2 are limited compare to other protein kinases. First of all, there are no available experimental data for the active conformation of this target and all the X-ray solved 3D structures of its inactive conformations collected in the PDB are incomplete. The only structure that has the coordinates for the entire A-loop is 2OH4 [16], although with high B-factors associated with the corresponding atoms. Here, the A-loop is in an inhibitory conformation and the DFG motif in an "out" orientation.

Thus, in order to investigate at the atomistic level a possible path involving the transition from the closed to the open conformation of the target, and to sample some possible intermediate conformations to employ in a structure-based drug discovery study, in a preliminary step we tried to solve the gaps in the experimental data and fully characterize the two limit conformations through the use of molecular modeling tools.

Affinity of inhibitors for the activate or inactivate kinase form could dictate the open or closed conformation of the A-loop, thus making the different conformations of kinases different drug targets in drug discovery projects and in this view making the conformational sampling of particular relevance.

Because of the unusual phosphorylation state of the tyrosine residues of the A-loop in the crystal structure 2OH4, computational investigations were carried out on the not phosphorylated state of the protein; similarly, the structure of the KDR domain in the open conformation obtained by the homology modeling procedure, described in the experimental section, was considered in its unphosphorylated state.

Due to incomplete and not fully reliable crystallographic data available for the KDR domain in the starting structure 2OH4 (very high B-factors values and disorder in the

electron density-maps) and in order to assess the effect of our computational modifications on the overall stability of the target, we performed on these two limit conformations 5 ns of standard MD simulations in explicit solvent. These simulations also allowed the optimization of the "chimeric" structure obtained by homology modeling for the open and *DFG-in* conformation. All MD simulations were performed on the protein structures in their unligated form and no ligands were included in the simulations. The two optimized structures can then be used in the rational design of selective type-I, type-II and type-III kinase inhibitors.

4.1.2 Characterization of the closed and open conformations of the KDR domain through MD simulations

Fig.4.1.1 shows the root mean square deviation (RMSD) of the atoms belonging respectively to the overall kinase domain and to the A-loop, calculated with respect to the corresponding atoms in the initial closed and open conformations. RMSD was monitored throughout MD simulations and evaluated by taking into account only the C_α atoms. In general, these values show a good thermal stability for the overall structure of the KDR domain in both the inactive (top panel) and active (bottom panel) conformations (black solid lines a) in the Fig.4.1.1), while a higher mobility can be observed for the A-loop (red solid lines a) in the Fig.4.1.1). It is worth noting the A-loop is more flexible in the inactive conformation than in the open structure where the coordinates for the A-loop were taken from a different kinase domain. This confirms the reliability of the model built for the open and *DFG-in* conformation and shows that this structure is preserved throughout the simulation. Moreover, the figure Fig.4.1.1 clearly shows that the A-loop is the most diverse region between the inactive and active conformations of the target (black and red dotted lines b) in both top and bottom panel in the Fig.4.1.1).
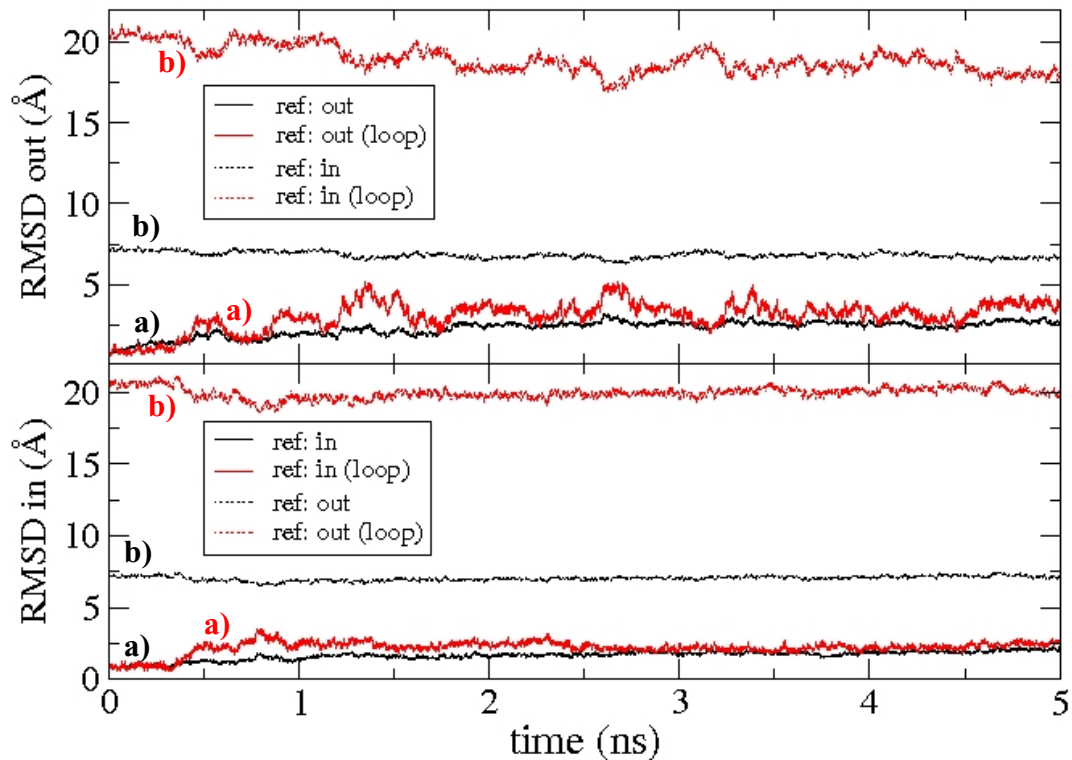
Fig.4.1.1 RMSD evaluated by taking into account only the C_α as a function of time. Top and bottom panels refer to the inactive (*DFG-out*) and active (*DFG-in*) KDR conformation, respectively.

Top: RMSD values evaluated for the closed conformation of the target with respect to a) the starting closed 2OH4 conformation (solid line) and b) to the starting open "chimeric" conformation (dotted line). Red and black lines refer to RMSD values computed for the A-loop atoms and the entire KDR domain, respectively.

Bottom: RMSD values evaluated for the open and *DFG-in* conformation of the target with respect to a) the starting open "chimeric" conformation (solid line) and b) to the starting closed 2OH4 conformation (dotted line). Red and black lines refer to RMSD values computed for the A-loop atoms and the entire KDR domain, respectively.

The visual inspection of the MD trajectories showed that the two structures remain in their starting (closed or open) conformation throughout the simulations (Fig.4.1.2). A rearrangement can, however, be observed for the αC-helix in the N-terminal lobe and for the A-loop, which is indeed the most flexible region of the target. In particular, the αC-helix, with the same coordinates in the two starting conformations, assumes after a partial and opposite rotation two different orientations, which can be referred to as αC-OUT and αC-IN (Fig.4.1.2).
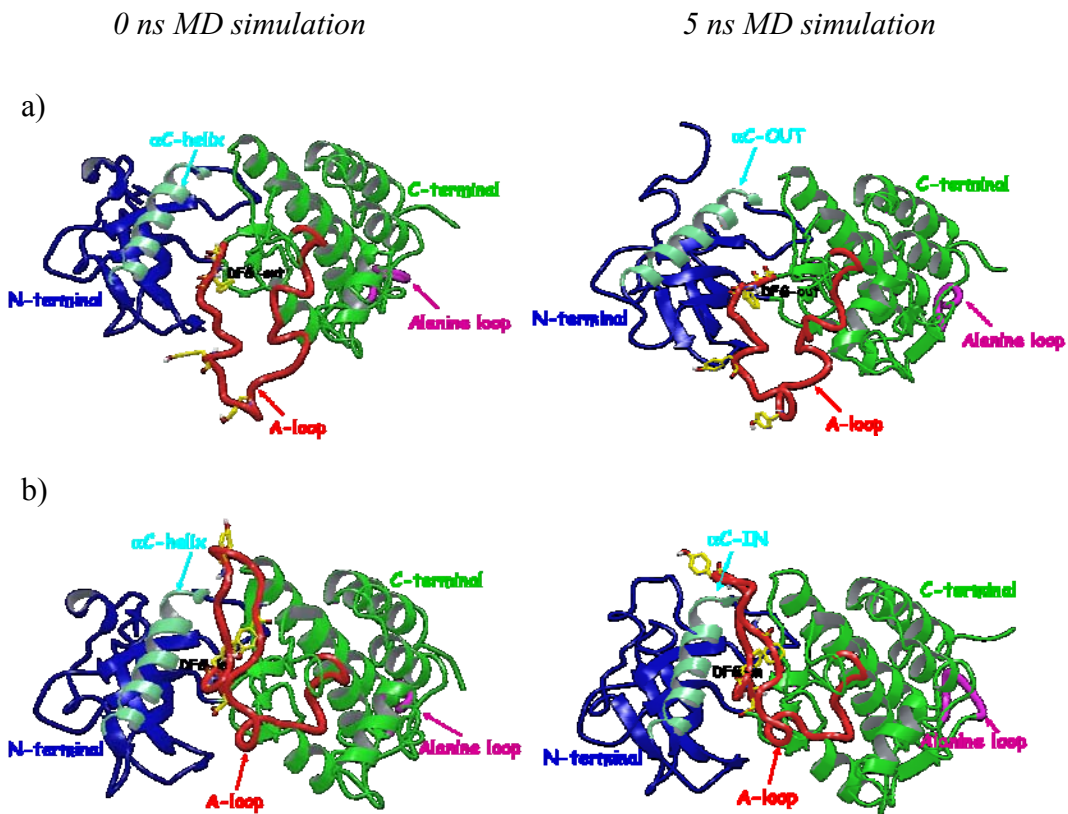
0 ns MD simulation   *5 ns MD simulation*

a)

b)

Fig.4.1.2. Ribbon representation of the closed and *DFG-out* (top) and open and *DFG-in* (bottom) conformations of the KDR domain at the beginning (left) and after (right) 5 ns MD simulation. The important regions of the protein are highlighted with different colours.

4.1.2.1 Analysis of MD trajectories

A detailed analysis of the behaviour of the A-loop region during the two MD simulations showed that the highly conserved DFG motif did not change its starting orientations (respectively *DFG-out* and *DFG-in*) and we did not observe the so-called "DFG flip". This result suggests that the time scale for this flip is much longer than the time of the standard MD simulations.

The dihedral angle (ϕ and ψ) values of the residues belonging to the DFG motif (Asp178, Phe179, Gly180; D1044 F1045 G1046 in 2OH4 numbering) were monitored throughout the simulations. These angles show characteristic values that remain constant during the simulations and can thus be used to define the two limit conformations of the A-loop. These values are reported in the Ramachandran plots in Fig.4.1.3.
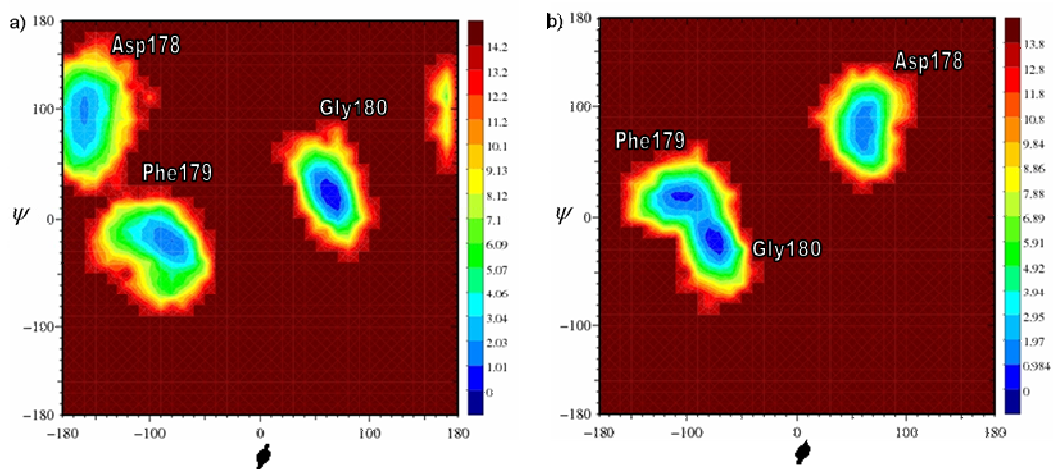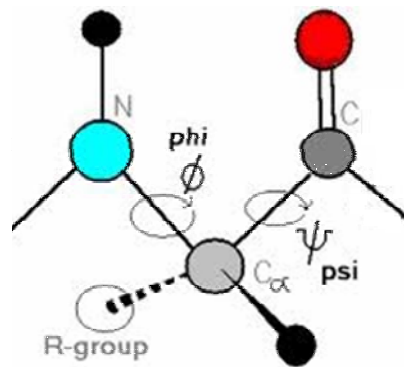
Fig.4.1.3 Ramachandran plots for the ϕ and ψ angles of the residues Asp178, Phe179 and Gly180 monitored during the two MD simulations performed on: a) the closed and *DFG-out* conformation and b) the open and *DFG-in* conformation of the KDR domain. Top: schematic representation of the ϕ and ψ angles.

In detail, in the *DFG-out* conformation (Fig.4.1.4), the aromatic ring of Phe179 (F1045) points toward the ATP binding cleft e no hydrogen bonds are formed between the three residues (Asp178, Phe179, Gly180; D1044 F1045 G1046) and other residues of the protein which contribute to the stabilization of the DFG motif in this orientation.
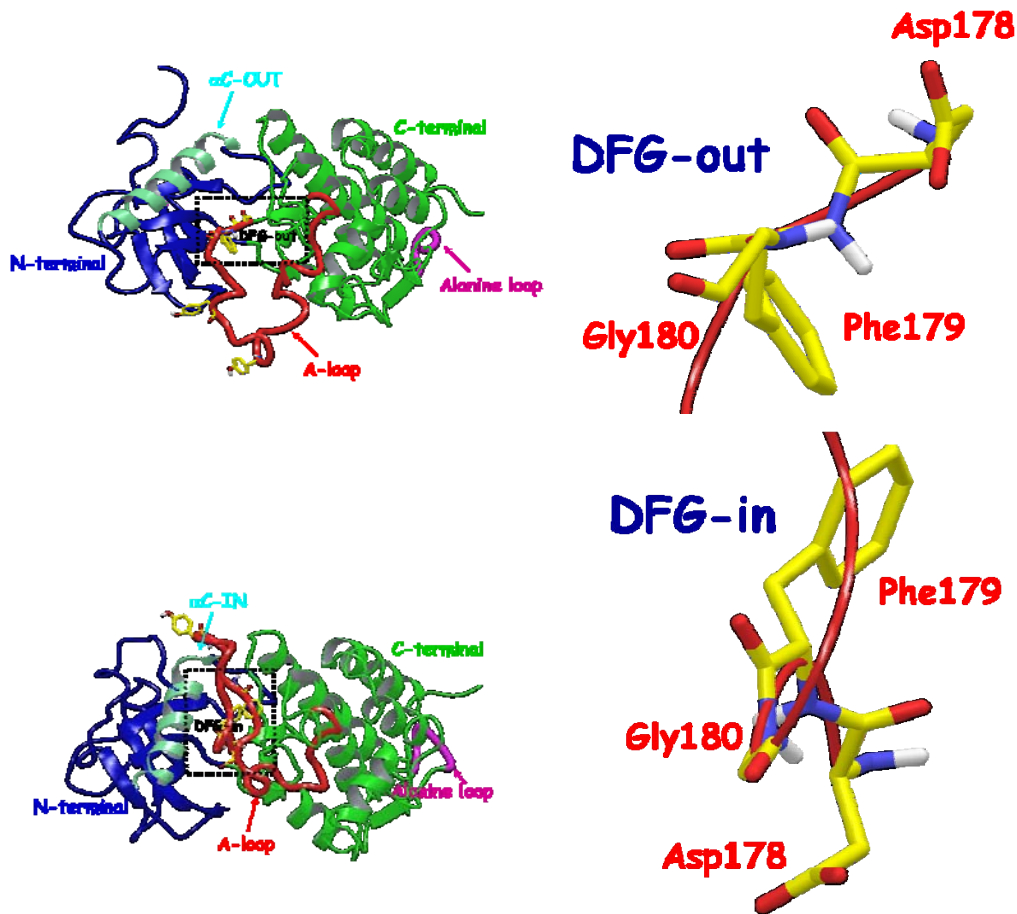
Fig.4.1.4 Representation of the *DFG-out* (top) and *DFG-in* (bottom) conformations resulting from the two standard 5ns MD simulations. The backbone of the A-loop is coloured in red.

In the *DFG-in* conformation, on the contrary, the residues of the DFG motif are involved in a network of H-bonds strengthened by salt bridges (Fig.4.1.5) with the two highly conserved residues Lys52 (K866) and Glu69 (E883) of the N-terminal lobe, which also play an important role in the kinase catalytic mechanism. The salt bridge between Lys52 (K866) and Glu69 (E883) also contributes to anchor the αC-helix in its αC-IN position.

Additional hydrogen bonds are formed between:

- the CO of the backbone of Asp178 (D1044) and the residue His158 (H1024), belonging to the highly conserved HRD motif of the catalytic domain;
- the NH of the backbone of Phe179 (F1045) and Glu69 (E883);
- the CO of the backbone of Phe179 (F1045) and the NH of Ala182 (A1048).

Moreover, in the *DFG-in* conformation, the side chain of Phe179 (F1045) is buried in the hydrophobic pocket, adjacent to the ATP binding site, which is formed by the side chains of the residues Ile72 (I886), Leu73 (L887), Val82 (V896), Leu151 (L1017), Ile176 (I1042) and Ala182 (A1048). These hydrophobic residues can be considered part of the hydrophobic network ("hydrophobic spine") which has been recognized as a stabilizing element of the active conformations of the kinase domains. [32, 34]
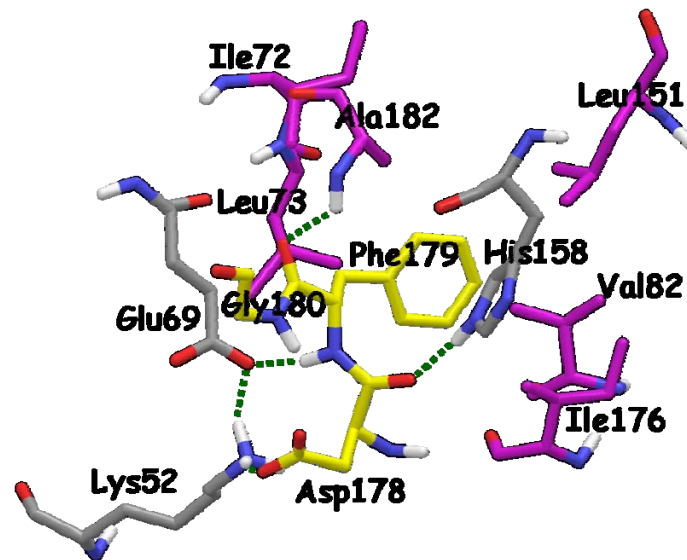

Fig.4.1.5. Network of interactions which contribute to the stabilization of the DFG motif in the "in" orientation. Residues of the DFG motif are represented in yellow. H-bonds are indicated with dotted green lines; hydrophobic residues are represented in purple.

The detailed analysis of the remaining portion of the A-loop in the two limit protein conformations allowed the determination of other interactions which can be responsible for the stabilization of the A-loop in its active or inactive state.

The analysis of the MD trajectory for the *DFG-out* structure shows that the conformation of the A-loop seems to be mainly stabilized by a series of hydrogen-bonds, strengthened by salt bridges, involving residues belonging to the A-loop and to other portions of the protein (inter-loop H-bonds). In addition several hydrogen-bonds inside the A-loop are formed and disrupted during the MD simulation (intra-loop H-bonds). Residues involved in inter-loop H-bond contacts are shown in Fig.4.1.6.
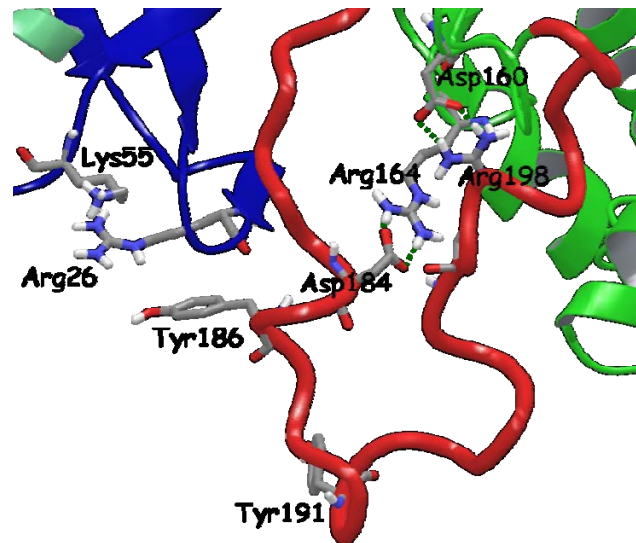
Fig.4.1.6 Conformation of the KDR domain sampled during the MD simulation, with the network of external interactions which stabilize the A-loop in its inactive state. Hydrogen-bonds are indicated with dotted green lines.

The interaction involving the charged groups of Asp184 (D1050) and Arg164 (R1030) is very stable throughout the simulation, while the salt bridge formed between Asp160 (D1026) and Arg198 (R1064) is less stable during the MD trajectory, although the charged residues remain in close proximity (Fig.4.1.7). Both these interactions can act as constraint to the free movements of the A-loop and can anchor the loop in its closed conformation, thus preventing the sampling of intermediate conformations for the target. An additional constraint to the flexibility of the A-loop can also be recognized in the inter-loop interaction involving the side chain of Tyr186 (Y1052), a residue which undergoes phosphorylation, and the positively charged region formed by the residues Arg26 (R840) and Lys55 (K869).
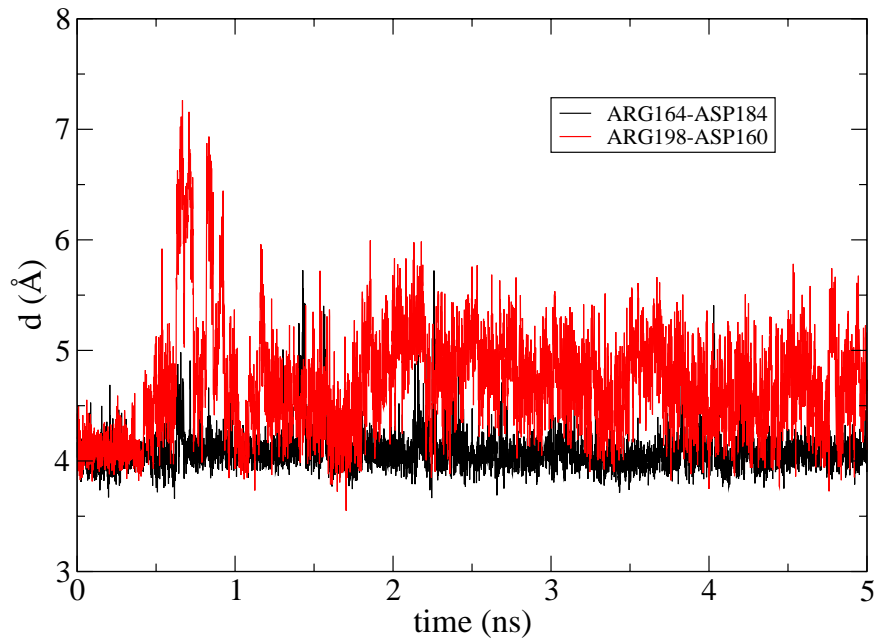
Fig.4.1.7. Monitoring of Arg164 (R1030) - Asp184 (D1050) and Asp160 (D1026) - Arg198 (R1064) distances during the MD simulation carried out on the closed and *DFG-out* conformation of the KDR domain. The distance is evaluated between the centroids of the functional groups involved in the polar interaction.

On the contrary, in the active and *DFG-in* conformation of the KDR domain, the A-loop is oriented in order to allow the access of an ATP molecule into the active site. During the MD trajectory hydrogen-bonds are in turn formed and broken, the only intra-loop stable interaction that is maintained throughout the MD simulation involves the side chains of Arg183 (R1049) and Asp188 (D1054). However, apart from these polar interactions, the open conformation of the KDR domain seems to be mainly stabilized by hydrophobic interactions as recently highlighted for other kinase domains. [32, 34]

The results obtained from these standard MD simulations are not surprising because the transition between the two limit conformations of the target involves a complete rearrangement of the region around the binding site and happens on a time scale much longer than that usually employed for standard MD. In this view the transition from the inactive to the active (or vice versa) A-loop conformation can be indeed considered a "rare event". In general, rare events are processes that occur

infrequently due to dynamical bottlenecks that separate stable states and their observation require either very long simulation times or the employment of a sampling method. Standard MD studies in explicit solvent can, in fact, access only a limited time scale (about 100 ns), while large-scale conformational rearrangements occur on a longer time scale. Thus, in order to overcome such limitation in the sampling efficiency of standard MD, we used the sampling method referred to as Hamiltonian REM. Our aim was to investigate at atomic details a possible opening mechanism involving the KDR domain, starting from the inactive and *DFG-out* conformation.

4.1.3 Hamiltonian REM simulations on the crystallographic derived structure of KDR domain

Hamiltonian REM simulations in explicit solvent were performed starting from the inactive 2OH4 derived structure of the KDR domain bearing the two tyrosines of the A-loop in their phosphorylated states: Tyr(P)186 (Yp1052) and Tyr(P)191 (Yp1057). In these simulations we gradually weakened (scaled) only the torsional and non bonded interactions involving the residues of the A-loop (residues 176-207; 1042-1073); the other interactions, including those involving simultaneously residues inside and outside the A-loop were not affected by the scaling factors. Scaling only the intra-loop interactions, in particular, permits to fully observe both the effect of hydrophobicity and the effect of the inter-loop interactions on the conformational landscape of the A-loop and on the opening/closing mechanism.

These REM simulations aimed also to test the effect of the 2OH4 unusual phosphorylation on the efficiency of sampling. Results from these simulations confirmed the importance of the dephosphorylation of the inactive conformation in order to allow a free movement of the A-loop and the sampling of intermediate conformations. The strong interaction between the negative charged phosphate group of the Tyr(P)186 (Yp1052) and the positive charged region generated by the side chains of both Arg26 (R840) and Lys55 (K869) in the N-terminal lobe is not affected by scaling factors in the different replicas, and locks the A-loop in its inactive conformation, thus acting as a real external hook (Fig.4.1.8). This polar interaction is also present in the replica performed at the highest effective temperature ($T = 3000$

K) and is not affected by the accelerated motion of the A-loop resulting from the scaling of the intra-loop potential terms.
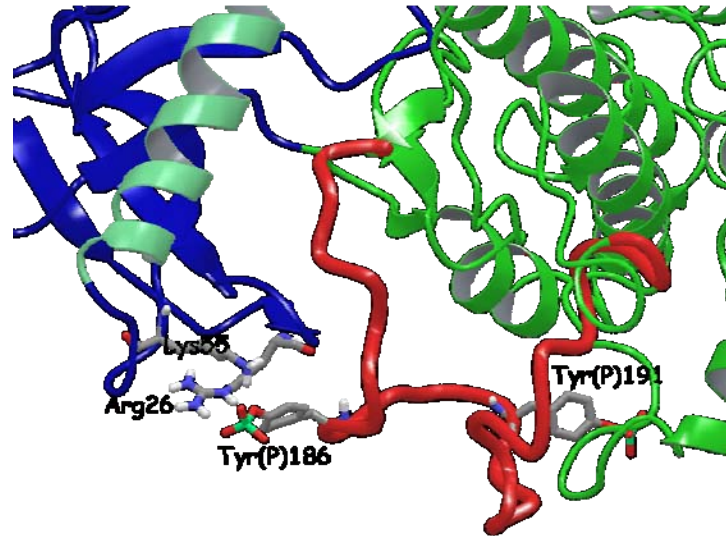

Fig.4.1.8 A conformation of the KDR domain observed in the highest replica employed in Hamiltonian REM simulations. The strong interaction involving the side chains of the residues Tyr(P)186 (Yp1052), Arg26 (R840) and Lys55 (K869) is highlighted. It is worth noting that the side chains of the two phosphotyrosines are distorted as a result of the scaling of the intra-loop potential terms.

Fig.4.1.9 shows the probability distribution of the distances between the residues involved in the aforementioned strong polar interaction (Tyr(P)186 (Yp1052), Arg26 (R840) and Lys55 (K869)), in the highest replica of Hamiltonian REM simulations. These distributions reveal, in fact, that even at T = 3000 K the two monitored distances assume values centered to 3.5Å and 4.5Å, corresponding to strong interactions, thus demonstrating that these salt bridges play a key role in anchoring the A-loop in its closed conformation.
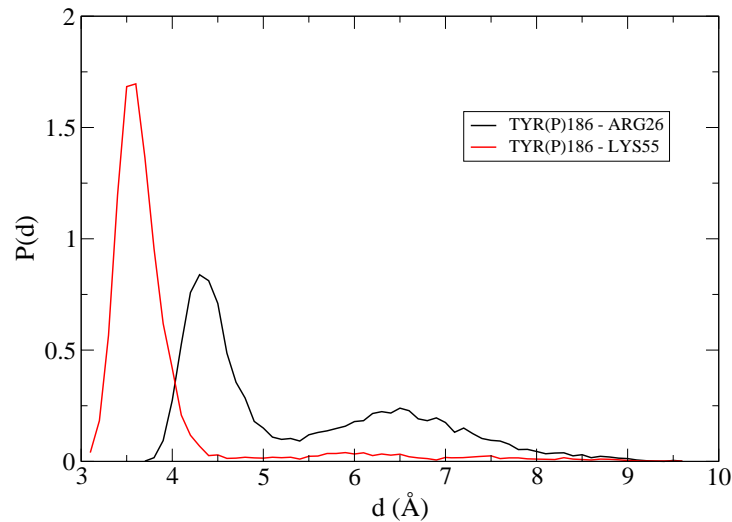
Fig.4.1.9 Structural data obtained from the highest replica of the Hamiltonian REM simulations performed on the phosphorylated inactive and *DFG-out* conformation of the KDR domain (effective temperature, T=3000 K). Black line: distribution of the distances between the phosphorus atom (P) of Tyr(P)186 (Yp1052) and the centroid of the guanidinium group of Arg26 (R840). Red line: distribution of the distances between P of Tyr(P)186 (Yp1052) and the centroid of the ε-amino group of Lys55 (K869).

Combining the results obtained from the previous simulations (standard MD and Hamiltonian REM) clearly emerged that a series of external constraints can interfere with the free movement of the A-loop and do not allow the sampling of intermediate conformations for the target. In particular, we have identified as possible hooks, the salt bridges involving the residue pairs Arg164 (R1030) - Asp184 (D1050) and Asp160 (D1026) - Arg198 (R1064), in addition to the polar interaction between the phosphotyrosine Tyr(P)186 (Yp1052) and the positive region formed by the side chains of Arg26 (R840) and Lys55 (K869). In this context, the removal of the phosphate groups from the inactive structure greatly reduces the last interaction.

4.1.4 Hot (3000 K) simulations to explore a possible opening mechanism, starting from the closed conformation of the KDR domain

In order to reduce the computational cost of a Hamiltonian REM simulation and to investigate the effect of the previously identified external salt bridges on the opening mechanism involving the A-loop, 16 independent "hot (10 ns) MD simulations" were performed starting from the closed and *DFG-out* conformation of the KDR domain in the non phosphorylated form. In these "hot simulations", the interactions involving

the salt bridges between the residue Asp160 (D1026) - Arg198 (R1064) and Arg164 (R1030) - Asp184 (D1050) were weakened, in addition to the intra-loop interactions also scaled in the previous Hamiltonian REM simulations. Among these 16 "hot simulations", 12 remain close to the initial closed and *DFG-out* conformation throughout 10 ns. Instead, the other 4 "hot simulations" have a different behaviour and the system approaches, in different ways, to the open conformation during the 10 ns.

The RMSD values, evaluated by taking into account only the C_α atoms of the A-loop from their reference positions in the open and *DFG-in* and closed and *DFG-out* conformations (same references as in Fig.4.1.1) are shown as a function of the time In detail, in the Fig.4.1.10 are shown the 4 "hot simulations" (coloured in green, yellow, blue and black) in which the A-loop reaches an almost open conformation, plus another "hot simulation" (coloured in red) taken as representative of the 12 "hot s

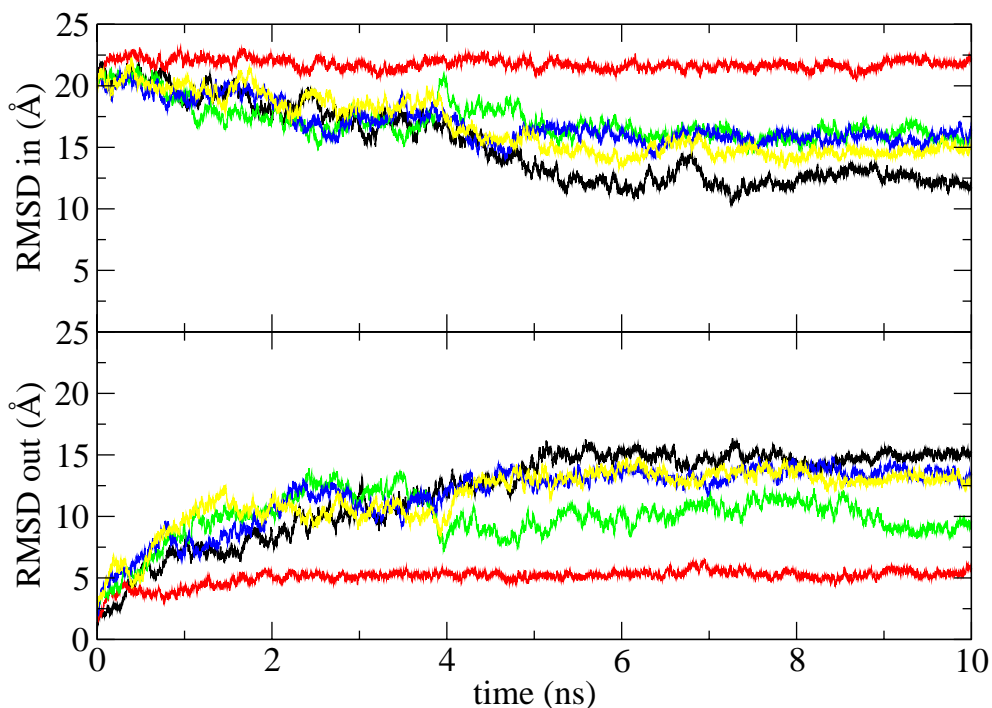**Fig.4.1.10** Conformational sampling observed during 5 independent 10 ns "hot simulations". Top: RMSD evaluated for the C_α atoms of the A-loop with respect to the corresponding atoms in the open *DFG-in* conformation as a function of the simulation time. Bottom: RMSD evaluated for the C_α atoms of the A-loop with respect to the corresponding atoms in the starting closed *DFG-out* conformation as a function of the simulation time.

The trend of RMSD for the black "hot simulation" gets away from the reference *DFG-out* conformation (bottom panel in Fig.4.1.10), while approaches to the reference *DFG-in* conformation (top panel in Fig.4.1.10), thus indicating that the conformation of the target at the end of this simulation is the one closer to the open state of the KDR domain. For this black "hot simulation", representative snapshots were taken at 500 ps intervals; the superposition of these frames (Fig.4.1.11) clearly points out that the removal of potential external constraints, obtained by scaling the salt bridges, makes it possible to sample a wide range of intermediate conformations for the A-loop.
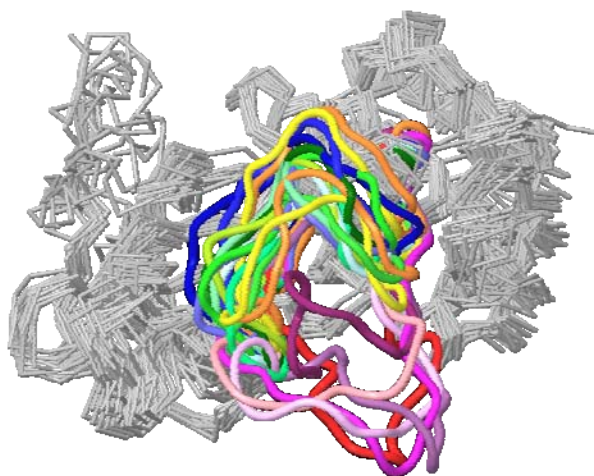

Fig.4.1.11 Sampling of the A-loop conformations during one "hot simulation"; snapshots were sampled at 500 ps intervals. In red is coloured the A-loop in the starting *DFG-out* conformation; in blue is coloured the conformation of the A-loop sampled in the last snapshot of the "hot simulation".

The visual analysis of the black "hot simulation" trajectory showed that Tyr186 (Y1052 in 2OH4) rapidly loses its interaction (already weakened by the absence of the phosphate group) with the positively charged region formed by the side chains of Lys55 (K869) and Arg26 (R840) and, as a consequence, the A-loop undergoes a rapid conformational rearrangement. However, the movement of the A-loop is also promoted by the weakening of the two external salt bridges involving the residues Asp160 (D1026) - Arg198 (R1064) and Arg164 (R1030) - Asp184 (D1050). It is worth noting that, in the first 5 ns of the black "hot simulation", the movement of the

A-loop involves also a partial rotation of the DFG motif which passes from the *DFG-out* to an *almost DFG-in* conformation (Fig.4.1.12).
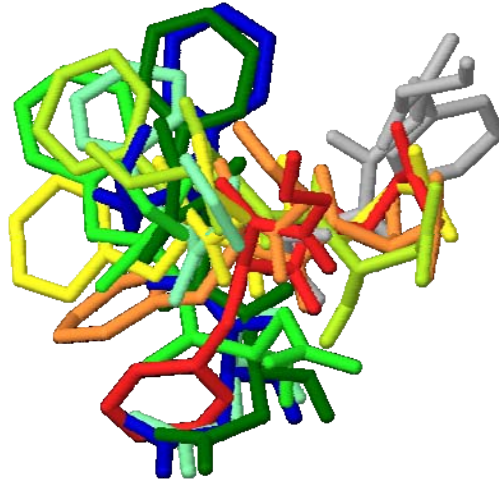

Fig.4.1.12 Conformations of the DFG motif sampled during the black "hot simulation". Snapshots are coloured in chronological order on a rainbow scale with red and blue indicating respectively the first (*DFG-out*) and the last (*almost DFG-in*) conformation sampled during the trajectory. The *DFG-in* conformation of the "chimeric", homology-built, structure is coloured in grey.

Anyway, as it clearly emerges from the Fig.4.1.12, we did not observe a complete flip of the DFG motif. It is worth noting that the conformation reached by the DFG motif in the first 5 ns of the black "hot simulation" (coloured in blue in the Fig.4.1.12), is quite stable and is maintained throughout the remaining 5 ns. A deeper inspection of the blue *almost DFG-in* conformation showed that the Phe179 (F1045) side chain results buried in a hydrophobic pocket formed by the side chains of the residues Leu73 (L887), Val83 (V897), Val98 (V912) and Val100 (V914), which is different from the hydrophobic pocket observed for the *DFG-in* conformation (Fig.4.1.13). Moreover, this *almost DFG-in* conformation seems to receive a further stabilization from a salt bridge between the CO of the backbone of Phe179 (F1045) and the highly conserved residue Lys52 (K866), which in turn is hydrogen bonded to the residue Glu69 (E883).
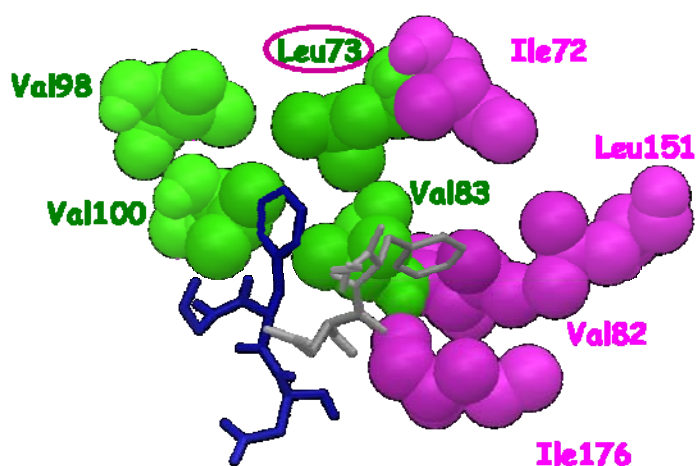
Fig.4.1.13 Superposition of the blue *almost DFG-in* conformation sampled during the black "hot simulation" and the *DFG-in* conformation of the "chimeric" homology-built structure (coloured in grey). Hydrophobic residues which surround the Phe179 (F1045) side chain of the DFG motif in the two conformations are coloured green and purple. The residue Leu73 is common to the two hydrophobic regions.

It is worth noting that even the complete opening movement of the A-loop, and thus the sampling of the open and *DFG-in* limit conformation, is prevented by this new identified network of hydrophobic residues which interferes with the "DFG flip".

Finally, the visual analysis of the black "hot simulation" trajectory highlights only a slight rotation of the αC-helix in the N-terminal lobe with respect to its starting orientation.

For the other three "hot simulations", over the four in which the system seems to evolve towards the open conformation (Fig.4.1.10 lines yellow, blue and green), the visual analysis showed a slightly different behaviour compared to that observed in the black "hot" MD. In particular, we observed a reduced rotation of Phe179 (F1045), whose side chain interacts with the new identified hydrophobic pocket.

4.1.5 Conclusions

In this thesis work the closed and open conformations of the KDR domain have been initially characterized by means of standard MD simulations. The reconstructed A-loop region in the active conformation is structurally similar to known kinase

domains: with respect to the inactive conformation, the αC-helix is rotated towards the ATP-binding site and the A-loop shows smaller oscillations around the active orientation. In order to investigate the opening mechanism of the A-loop, a Hamiltonian REM simulation and a series of targeted "hot simulations" have been performed starting from the inactive and *DFG-out* conformation of the KDR domain. This approach identified the phosphorylated tyrosines (Tyr(P)186 (Yp1052) and Tyr(P)191 (Yp1057)) and the salt bridges involving the residues Asp160 (D1026) - Arg198 (R1064) and Arg164 (R1030) - Asp184 (D1050) as the main responsible of the loop stability in its inactive conformation; weakening these interactions, the A-loop has been able to sample, in a few nanoseconds, conformations that are intermediates between the inactive and the active one. The complete rotation of the A-loop, and thus the sampling of the open *DFG-in* conformation, however, requires a hydrophobic barrier to be overcome. Such a barrier is due to the positioning of the Phe179 (F1045) side chain of the DFG motif into a hydrophobic pocket, located near to that one observed in the open *DFG-in* conformation. In any case, a major result emerging from the computational investigation performed on the KDR domain, is the identification of a stable conformational state for the DFG motif, in which the Phe179 (F1045) side chain is buried in a not previously identified hydrophobic pocket. Indeed, this conformation, that we have called *almost DFG-in*, can be considered a possible intermediate state of the opening mechanism of the KDR domain and it can be very useful in the rational design of inhibitors for this target.

4.2 MODELING INVESTIGATION OF DOUBLE HELIX AND G-QUADRUPLEX DNA STRUCTURES IN COMPLEX WITH ACRIDINE-BASED POLYAMINE LIGANDS

4.2.1 Acridine derivatives binding to double stranded DNA - Docking and MD simulations

The binding mode of the diprotonated species of L1, L2 and L3 towards dsDNA was investigated by docking procedures followed by 10 ns MD simulations, with an explicit treatment of the aqueous environment.

MD simulations allowed an optimization of the complexes resulting from docking calculations. Double-helix oligonucleotides (10-mer) of base composition poly(dG-dC) were investigated considering both CG and GC intercalative binding sites. In this study we decided to investigate only the base composition poly(dG-dC) for oligonucleotides because experimental studies of the binding process of the three ligands to oligonucleotides of defined base composition, namely poly(dA-dT) and poly(dG-dC), showed a lack of sequence selectivity. [98]

For all MD simulations, the root mean square deviation (RMSD) of the complexes was monitored *vs.* the simulation time (Fig.4.2.1a and Fig.4.2.1b).
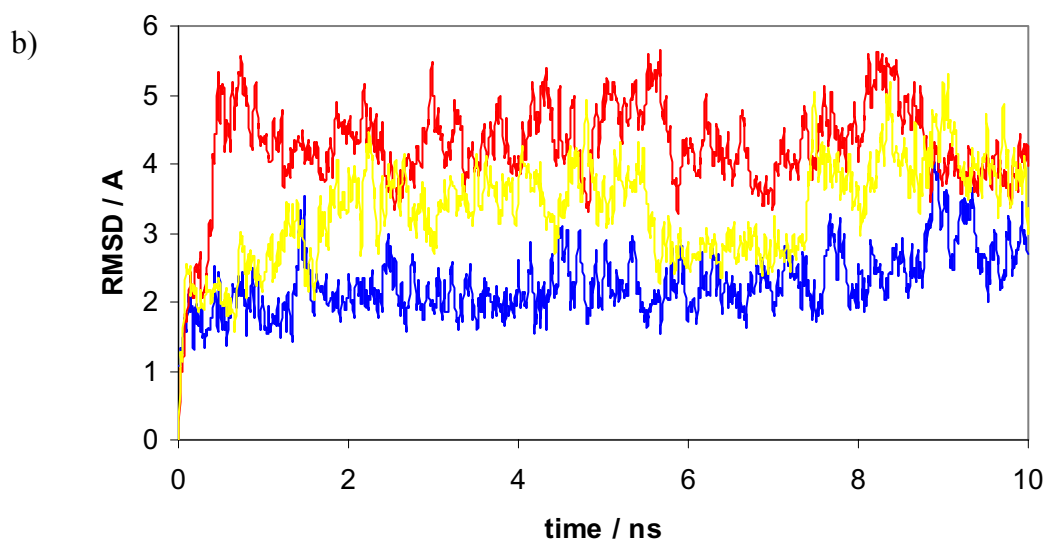
a)
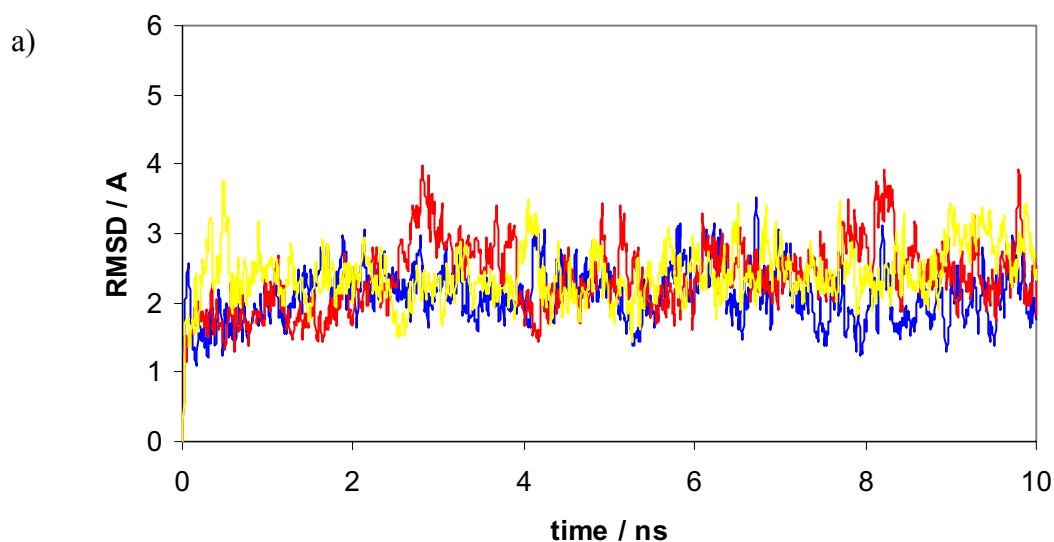
Fig.4.2.1 dsDNA/L RMSD value, evaluated for all non hydrogen atoms, as a function of time during the recorded MD trajectories: (a) ligand molecule lodged in the CG intercalative binding site (blue = L1, yellow = L2 and red = L3), (b) ligand molecule lodged in the GC intercalative binding site (blue = L1, yellow = L2 and red = L3).

RMSD values showed a different behavior for the three ligands in both the intercalative binding sites, CG and GC. The same results have emerged from the structural comparison of the complexes resulting from MD simulations (Fig.4.2.2). All these results largely corroborate the solution studies, as the three ligands feature different binding modes towards the poly-G-poly-C double helix. [98] For the open-chain ligand L1, the interaction can be described as properly intercalative both in the case of the CG site, as well as for the GC site. In all situations the adduct was stabilized by additional H-bond contacts involving the terminal protonated nitrogen atoms and the carbonyl oxygen atom belonging to cytosine and guanine units (Fig.4.2.2a) or the phosphate groups (Fig.4.2.2b). In fact, the planar acridine moiety of L1 can be easily lodged between the base pairs, unhindered by the pendant arms, which, on the contrary, can be set in the grooves. Moreover, the RMSD values evaluated as a function of time during the recorded MD trajectories (Fig.4.2.1a and Fig.4.2.1b) denotes that the L1-poly-G-poly-C adducts are quite stable, and they are not disrupted during the 10 ns simulation time.
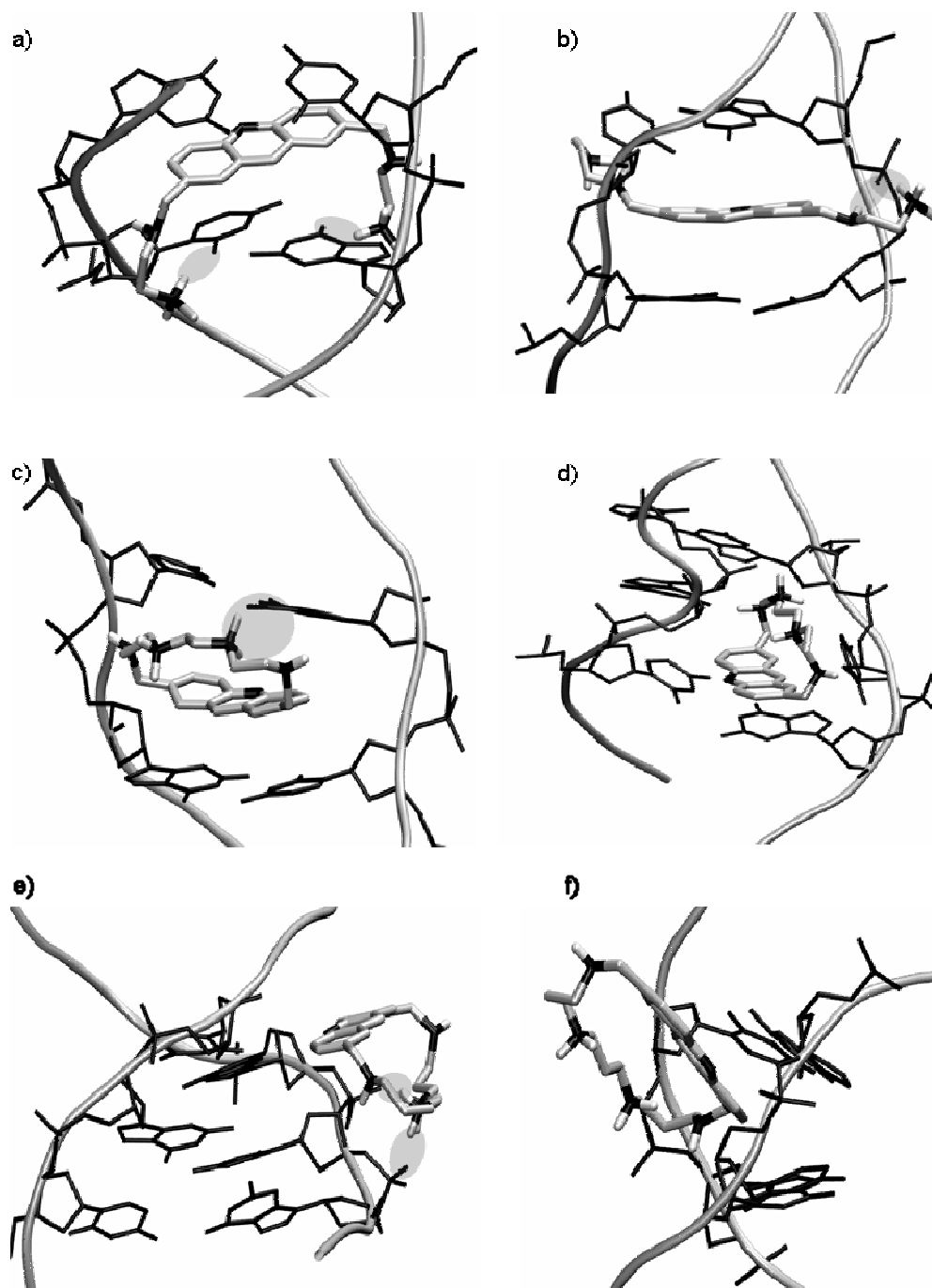
Fig.4.2.2 Results of the modeling procedures (docking and MD) for the adducts formed by diprotonated forms of L1, L2 or L3 and poly-G-poly-C dsDNA. Results were obtained by starting from docking in a CG intercalative binding site - (a) L1, (c) L2, (e) L3 - and in a GC intercalative binding site - (b) L1, (d) L2, (f) L3. Atoms involved in hydrogen bonding are evidenced by light gray circle.

On the other hand L3, most likely hindered by the bent conformation adopted by its macrocyclic structure, seems to fail the intercalation and gives rise to groove binding, interacting mainly through hydrogen bonds between its protonated nitrogen atoms and the polyphosphate DNA backbone (Fig.4.2.2e/f). After 10 ns MD simulation, the DNA double helix is almost completely rebuild, and no trace of the CG or GC intercalative site remains.

As far as L2 is concerned, it shows an intermediate behaviour between L1 and L3. In spite of its rather rigid and strained bent conformation, which is generally considered unsuitable for the stereochemical requirement of a classical DNA intercalator, modeling studies give evidence of stable conformations for the L2-DNA adduct (Fig.4.2.2c), in which the macrocycle is able to give rise to simultaneous π-π and H-bonding interactions. Indeed, L2 behaves as a multifunctional ligand towards DNA by virtue of its small size. Thus, the presence of the shorter ethylene spacers in place of propylene units featured in L3 appears to be essential for obtaining a correct match between the binding sites of L2 and the nucleobases.

4.2.2 Acridine derivatives binding to double stranded DNA - MM-PBSA investigation
A further characterization of the different binding modes has been achieved through the estimation of the relative free energies for the three ligand-dsDNA complexes by using the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) approach. This postprocessing method, employed in the evaluation of the Gibbs free energies of binding, works on a set of structures collected from MD simulations (some representative snapshots, used in the free energy estimation, are shown in Fig.4.2.3).

In this study, for the evaluation of the electrostatic contribution to the solvation free energy, we used a methodology able to solve the Poisson Boltzmann equation in the non-linearized form, which is more suitable to treat highly charged molecules such as DNA. It is noteworthy to underline that the MM-PBSA results, reported in Table.4.1, are in agreement with the thermodynamic parameters obtained by solution-phase studies, [98] showing the higher intercalative ability of L1 and L2 with respect to L3.
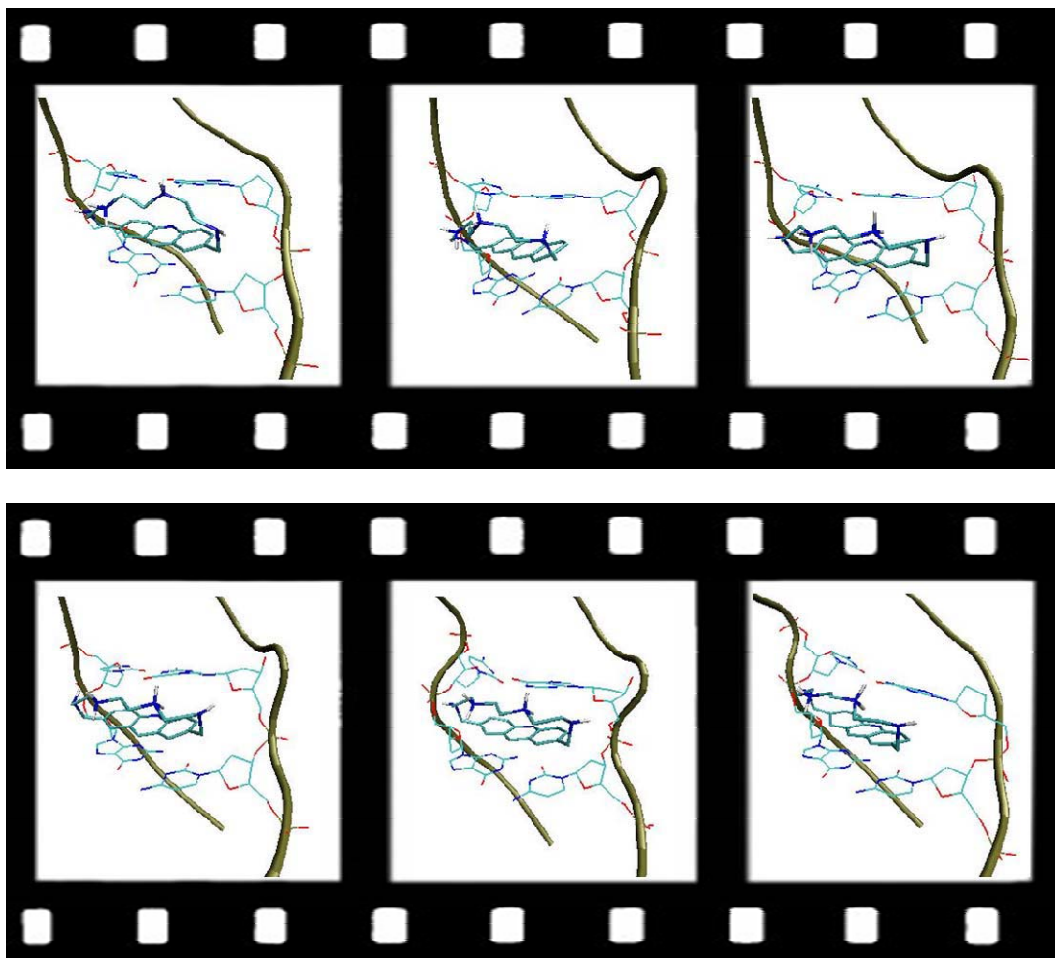
Fig.4.2.3 Representative snapshots extracted from the MD simulation, performed on the adduct formed by the diprotonated form of L2 and poly-G-poly-C dsDNA in a CG intercalative binding site, and employed in the free energy estimation by using the MM-PBSA approach (for detail, paragraph 3.2.2.3).

Table.4.1 Gibbs free energy values (in kcal/mol) calculated by using the MM-PBSA approach for adducts formed between ligands (L1, L2, L3) and dsDNA.

Ligands	ΔH_{calc}[a]	$T\Delta S_{calc}$[a]	ΔG_{calc}[a]	ΔH_{calc}[b]	$T\Delta S_{calc}$[b]	ΔG_{calc}[b]
	kcal/mol	kcal/mol	kcal/mol	kcal/mol	kcal/mol	kcal/mol
L1	-33.10	-27.90	*- 5.20*	-35.53	-30.47	*- 5.06*
L2	-32.00	-23.00	*-9.00*	-19.98	-22.06	*2.08*
L3	-16.33	-25.18	*8.85*	-20.12	-25.77	*5.65*

[a] CG intercalative site

[b] GC intercalative site

4.2.3 Acridine derivatives binding to G-quadruplex forming sequences - Docking and MD simulations

The orientations resulting from MD simulations carried out on selected poses of L1, L2 and L3 in complex with the G-quadruplex obtained from docking calculations using as target the 3D coordinates of the crystal structure 3CE5 [54] are shown in Fig.4.2.4. The root mean square deviation (RMSD) of the G-quadruplex/L complexes was monitored *vs.* the simulation time for all the three adducts (Fig.4.2.5). All these adducts remained quite stable during the MD.
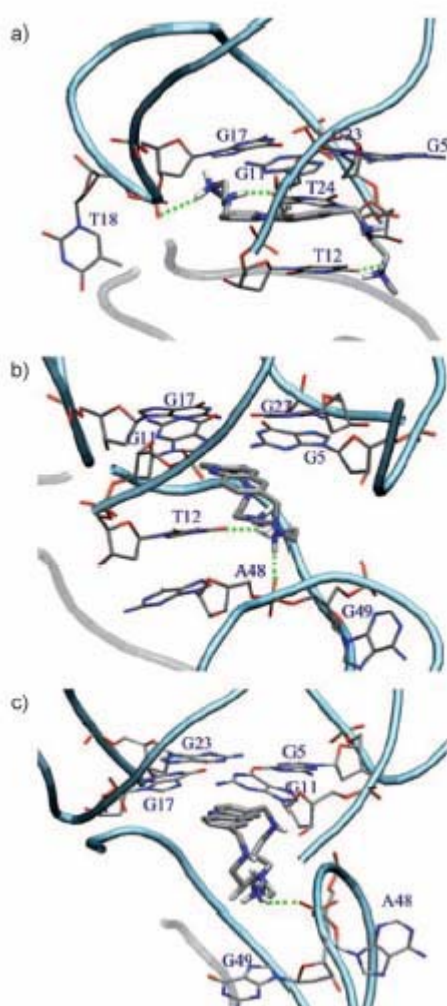

Fig.4.2.4 Results of MD simulations of the adducts formed by the diprotonated species a) L1, b) L2, and c) L3 with G-quadruplex structure formed by the 2GGG oligonucleotide. Hydrogen bonds are evidenced by green lines; interacting and no interacting DNA monomers (tube representation) are coloured in light blue and grey, respectively.
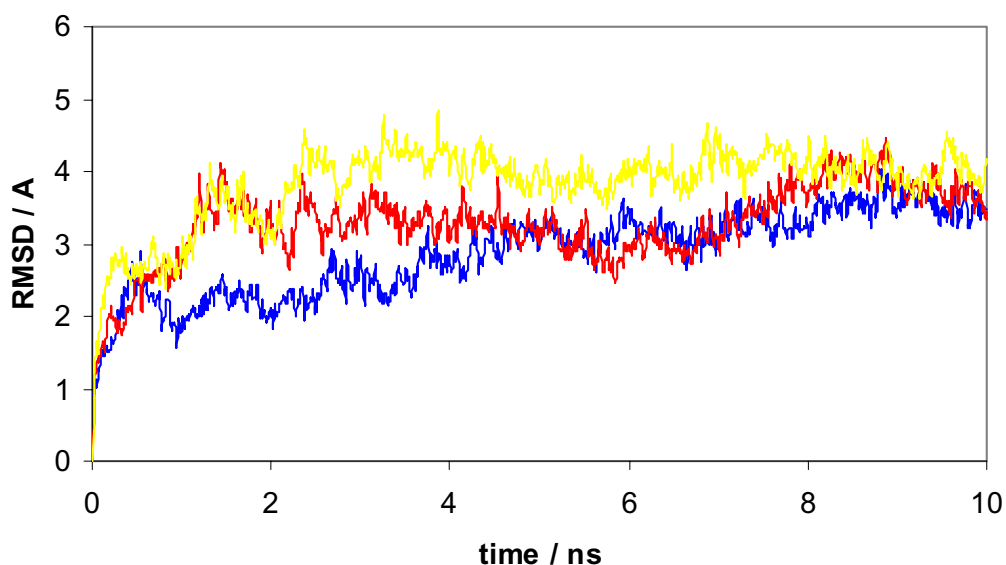
Fig.4.2.5 G-quadruplex/L RMSD value, evaluated for all non hydrogen atoms, as a function of time during the recorded MD trajectory (blue = L1, yellow = L2 and red = L3).

All ligands insert into the tetrameric form of the G-quadruplex, between the 3'-end G-tetrad belonging to one G-quadruplex dimer and the 5'-end of the other, in a manner similar to that shown by the very effective G-quadruplex ligand BRACO-19 present in the X-ray crystallographic complex structure. [54] However, the acridine moieties of L2 and L3 give rise to stronger π-π interactions with the 3'-end G-tetrad than in the case of L1, the acridine system of which shows the most significant deviation from the expected parallel position (16° between the involved planes versus 3° and 3.5° for L2 and L3, respectively). Moreover, all ligands cause, to varying degrees, a rotation of one dimeric G-quadruplex unit relative to the other. Using the BRACO-19-G-quadruplex complex as a reference, and considering the dihedral angle formed between 3'- and 5'-end G-tetrads, a 1°, 3° and 9° twist (Fig.4.2.6, $\Delta\phi = \phi_{ligand} - \phi_{BRACO-19}$) is observed for L2, L3 and L1, respectively.
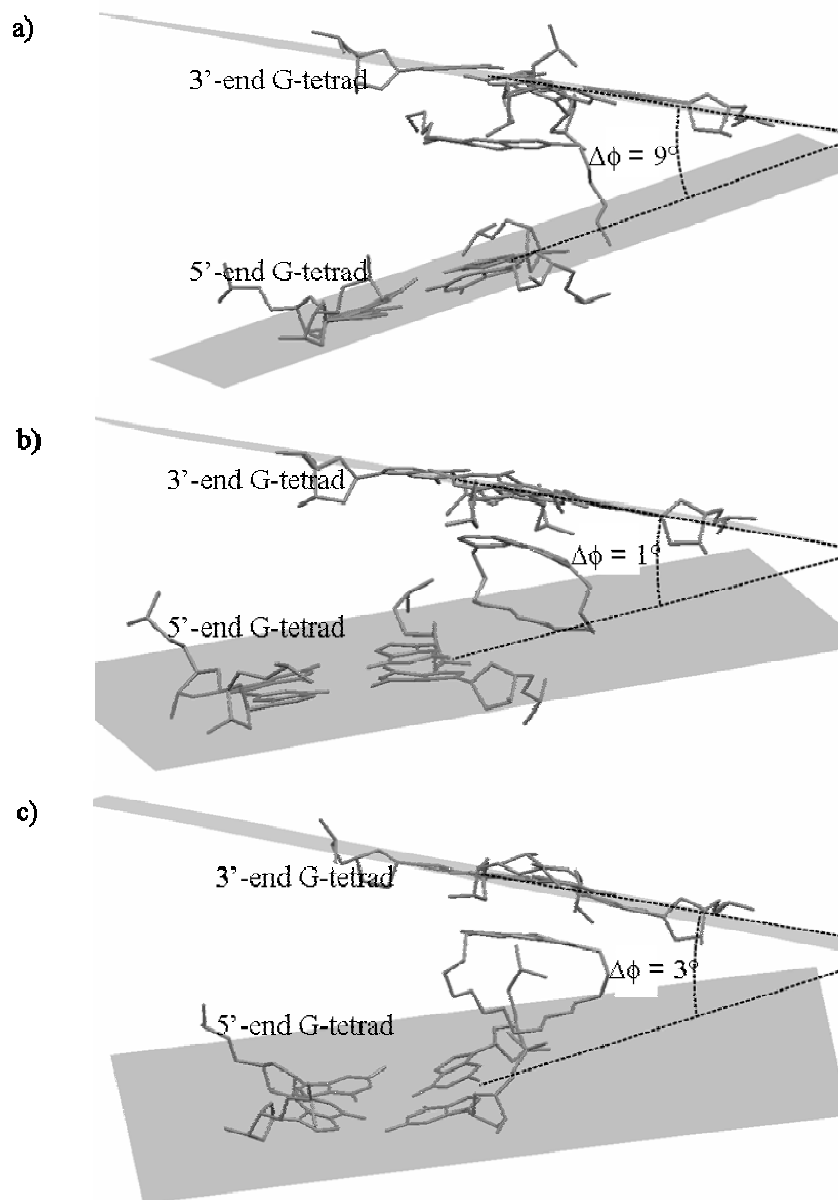
Fig.4.2.6 Dihedral angle formed between 3'- and 5'-end G-tetrads of a) L1, b) L2 and c) L3 G-quadruplex complexes and $\Delta\phi$ values evaluated using the BRACO-19/G-quadruplex complex [54] as a reference ($\Delta\phi = \phi_{ligand} - \phi_{BRACO-19}$).

The analysis of the L1-DNA complex shows that both π-stacking and H-bonding interactions involve only one dimer, whereas both the G-quadruplex units are simultaneously bound by L2 and L3 (Fig.4.2.4). As far as L2 and L3 are concerned, their binding modes seem similar, with both ligands interacting by π-stacking with a guanidine residue of the 5'-end G-tetrad, and by H-bonding with a phosphate group

of the other dimer (Fig.4.2.4 b and c). However, compared with L3, L2 causes a less significant disruption of the TATA tetrad, also present at the 3'- and 5'-end interface in the biological unit of the BRACO-19-DNA complex. In particular, only one adenine residue is considerably shifted from its original position, and, at the same time, the carbonyl group of a thymine residue gives rise to an additional H-bond interaction with a protonated nitrogen of the ligand, thus exerting a significant role in the stabilization of the adduct (Fig.4.2.4 b). However, in the case of the L3 complex, both the aforementioned adenine and thymine residues are in a different position, and no additional H-bond contacts are established.

Altogether, the observations derived from the modeling procedures support the quite surprising results obtained from the solution studies, [98] which show evidence that the bent conformations of L2 and L3 do not prevent G-quadruplex recognition.

4.2.4 Acridine derivatives binding to G-quadruplex forming sequences - MM-PBSA investigation

Following the same procedure used for the adducts formed between the three ligands (L1, L2 and L3) and dsDNA, the binding free energy values for the complexes with G-quadruplex were obtained by MM-PBSA calculations (Table.4.2). The obtained results indicate that the binding affinity decreases in the order L2 > L3 > L1, in agreement with the reported solution studies. [98] In the evaluation of the Gibbs free energies we employed a set of structures collected from MD simulations (some representative snapshots, used in the free energy estimation, are shown in Fig.4.2.7).

Table.4.2 Gibbs free energy values (in kcal/mol) calculated by using the MMPBSA approach for adducts formed between ligands (L1, L2, L3) and G-quadruplex.

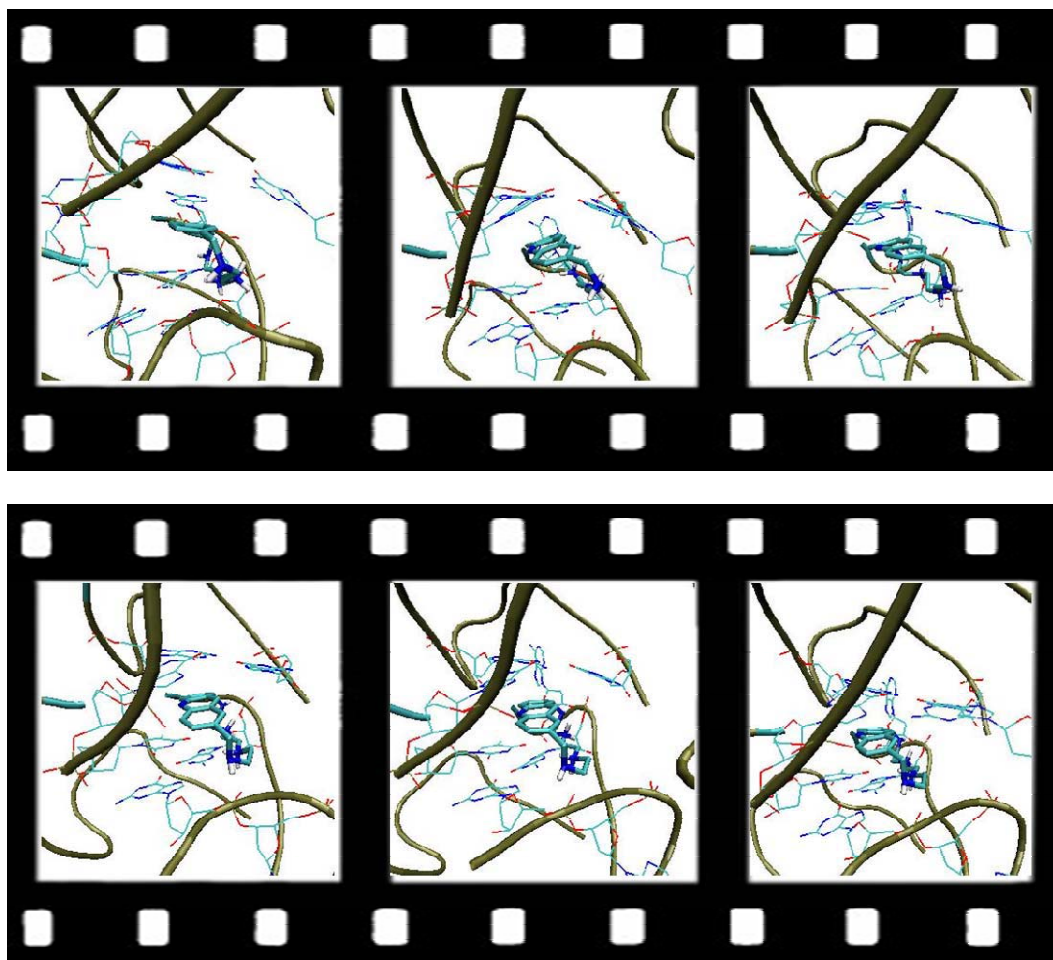| Ligands	ΔH_{calc} kcal/mol	$T\Delta S_{calc}$ kcal/mol	ΔG_{calc} kcal/mol
L1	-39.11	-29.51	*- 9.60*
L2	-40.15	-23.61	*-16.54*
L3	-35.58	-22.11	*-13.47*

Fig.4.2.7 Representative snapshots extracted from the MD simulation, performed on adduct formed by the diprotonated form of L2 and G-quadruplex DNA, and employed in the free energy estimation by using the MM-PBSA approach (for detail, paragraph 3.2.2.3).

4.2.5 Conclusions

The new 2,7-substituted acridine derivatives L1, L2, and L3 show an unusual trend in their binding patterns toward canonical double-helical DNA and non canonical G-quadruplex structures. The DNA recognition profile of L1 is in line with the expectation: it binds to dsDNA through a pure intercalative process and poorly recognizes G-quadruplex structures. Ligand L2 features a small macrocyclic framework which is generally expected to adopt a rather bent and rigid conformation, usually not prone to intercalative interactions with DNA. However, in spite of this, L2 binds to dsDNA almost as well as the open-chain ligand L1, and it shows the best performance toward a G-quadruplex. Distinctly, ligand L3, characterized by the same

81

cyclic structure of L2, but with longer and more flexible propylene spacers, should, in principle, be able to adopt a planar conformation more readily. Nevertheless, it is unable to intercalate into dsDNA, nor can it induce G-quadruplex structures. As pointed out by our molecular modeling studies, the better performance of L2 relative to that of L3 and the open-chain ligand L1 can be attributed entirely to its particular conformation and dimensions, which establish an optimal match with both dsDNA and the tetrameric G-quadruplex structure. These results indicate how important it is to set up simultaneous combinations of interactions of various kinds between DNA and ligand in order to reach an optimal binding efficiency, thus effectively opening new possibilities for the design of novel selective DNA binders. [98]

5. References

5.1 REFERENCES

1. Hubbard, S. R.; Till, J. H. *Annu. Rev. Biochem.*, **2000**, *69*, 373-398.

2. Blume-Jensen, P.; Hunter, T. *Nature*, **2001**, *411*, 355-365.

3. Nolen, B.; Taylor, S.; Ghosh, G. *Mol. Cell*, **2004**, *15*, 661–675.

4. Folkman, J. *Exp. Cell Res.*, **2006**, *312*, 594–607.

5. Maruotti, N.; Cantatore, F. P.; Crivellato, E.; Vacca, A.; Ribatti, D. *Histol. Histopathol.*, **2006**, *21*, 557–566.

6. Wilkinson-Berka, J. L. *Curr. Pharm. Des.*, **2004**, *10*, 3331–3348.

7. Cherrington, J. M.; Strawn, L. M.; Shawver, L. K. *Adv. Cancer Res.*, **2000**, *79*, 1-38.

8. Roskoski Jr., R. *Biochem. Biophys. Res. Commun.*, **2008**, *375*, 287-292.

9. a) Shibuya, M. *Endothelium*, **2006**, *13*, 63-69. b) Holmes, K.; Roberts, O. L.; Thomas, A. M.; Cross, M. J. *Cell. Signal.*, **2007**, *19*, 2003–2012.

10. Mac Gabhann, F.; Popel, A. S. *Biophys. Chem.*, **2007**, *128*, 125–139.

11. Kuriyan, J.; Cowburn, D. *Annu. Rev. Biophys. Biomol. Struct.*, **1997**, 26, 259-288.

12. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer Jr., E. E.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.*, **1977**, *112*, 535-542. http://www.rcsb.org/pdb/.

13. Leppanen, V. M.; Prota, A. E.; Jeltsch, M.; Anisimov, A.; Kalkkinen, N.; Strandin, T.; Lankinen, H.; Goldman, A.; Ballmer-Hofer, K.; Alitalo, K *Proc. Natl. Acad. Sci. U.S.A.*, **2010**, *107*, 2425-2430.

14. Yang, Y.; Xie, P.; Opatowsky, Y.; Schlessinger, J. *Proc. Natl. Acad. Sci. U.S.A.*, **2010**, *107*, 1906-1911.

15. McTigue, M. A.; Wickersham, J. A.; Pinko, C.; Showalter, R. E.; Parast, C. V.; Tempczyk-Russell, A.; Gehring, M. R.; Mroczkowski, B.; Kan, C.; Villafranca, J. E.; Appelt, K. *Struct. Fold. Des.*, **1999**, *7*, 319-330.

16. Hasegawa, M.; Nishigaki, N.; Washio, Y.; Kano, K.; Harris, P. A.; Sato, H.; Mori, I.; West, R. I.; Shibahara, M.; Toyoda, H.; Wang, L.; Nolte, R. T.; Veal, J. M.; Cheung, M. *J. Med. Chem.*, **2007**, *50*, 4453-4470.

17. Tasker, A. S.; Patel, V. F. **2009**, DOI:10.2210/pdb3efl/pdb.

18. Cee, V. J.; Cheng, A. C.; Romero, K.; Bellon, S.; Mohr, C.; Whittington, D. A.; Bak, A.; Bready, J.; Caenepeel, S.; Coxon, A.; Deak, H. L.; Fretland, J.; Gu, Y.; Hodous, B. L.; Huang, X.; Kim, J. L.; Lin, J.; Long, A. M.; Nguyen, H.; Olivieri, P. R.; Patel, V. F.; Wang, L.; Zhou, Y.; Hughes, P.; Geuns-Meyer, S. *Bioorg. Med. Chem. Lett.*, **2009**, *19*, 424-427.

19. Hilberg, F.; Gerald, J. R.; Krssak, M.; Kautschitsch, S.; Sommergruber, W.; Tontsch-Grunt, U.; Garin-Chesa, P.; Bader, G.; Zoephel, A.; Quant, J.; Heckel, A.; Rettig, W. J. *Cancer Res.*, **2008**, *68*, 4774-4782.

20. Harris, P. A.; Boloor, A.; Cheung, M.; Kumar, R.; Crosby, R. M.; Davis-Ward, R. G.; Epperly, A. H.; Hinkle, K. W.; Hunter, R. N.; Johnson, J. H.; Knick, V. B.; Laudeman, C. P.; Luttrell, D. K.; Mook, R. A.; Nolte, R. T.; Rudolph, S. K.; Szewczyk, J. R. Truesdale, A. T.; Veal, J. M.; Wang, L.; Stafford, J. A. *J. Med. Chem.*, **2008**, *51*, 4632-4640.

21. Kunz, R. K.; Rumfelt, S.; Chen, N.; Zhang, D.; Tasker, A. S.; Bürli, R.; Hungate, R.; Yu, V.; Nguyen, Y.; Whittington, D. A.; Meagher, K. L.; Plant, M.; Tudor, Y.; Schrag, M.; Xu, Y.; Ng, G. Y.; Hu, E. *Bioorg. Med. Chem. Lett.*, **2008**, *18*, 5115-5117.

22. Hu, E.; Tasker, A.; White, R. D.; Kunz, R. K.; Human, J.; Chen, N.; Bürli, R.; Hungate, R.; Novak, P.; Itano, A.; Zhang, X.; Yu, V.; Nguyen, Y.; Tudor, Y.; Plant, M.; Flynn, S.; Xu, Y.; Meagher, K. L.; Whittington, D. A.; Ng, G. Y.; *J. Med. Chem.*, **2008**, *51*, 3065-3068.

23. La, D. S.; Belzile, J.; Bready, J. V.; Coxon, A.; DeMelfi, T.; Doerr, N.; Estrada, J.; Flynn, J. C.; Flynn, S. R.; Graceffa, R. F.; Harriman, S. P.; Larrow, J. F.; Long, A. M.; Martin, M. W.; Morrison, M. J.; Patel, V. F.; Roveto, P. M.; Wang, L.; Weiss, M. M.; Whittington, D. A.; Teffera, Y.; Zhao, Z.; Polverino, A. J.; Harmange, J. C. *J. Med. Chem.*, **2008**, *51*, 1695-1705.

24. Harmange J. C.; Weiss, M. M.; Germain, J.; Polverino, A. J.; Borg, G.; Bready, J.; Chen, D.; Choquette, D.; Coxon, A.; DeMelfi, T.; DiPietro, L.; Doerr, N.; Estrada, J.; Flynn, J.; Graceffa, R. F.; Harriman, S. P.; Kaufman, S.; La, D. S.; Long, A.; Martin, M. W.; Neervannan, S.; Patel, V. F.; Potashman, M.; Regal, K.; Roveto, P. M.; Schrag, M. L.; Starnes, C.; Tasker, A.; Teffera, Y.; Wang, L.; White, R. D.; Whittington, D. A.; Zanon, R. *J. Med. Chem.*, **2008**, *51*, 1649-1667.

25. Weiss, M. M.; Harmange, J. C.; Polverino, A. J.; Bauer, D.; Berry, L.; Berry, V.; Borg, G.; Bready, J.; Chen, D.; Choquette, D.; Coxon, A.; DeMelfi, T.; Doerr, N.; Estrada, J.; Flynn, J.; Graceffa, R. F.; Harriman, S. P.; Kaufman, S.; La, D. S.; Long, A.; Neervannan, S.; Patel, V. F.; Potashman, M.; Regal, K.; Roveto, P. M.; Schrag, M. L.; Starnes, C.; Tasker, A.; Teffera, Y.; Whittington, D. A.; Zanon, R. *J. Med. Chem.*, **2008**, *51*, 1668-1680.

26. Potashman, M. H.; Bready, J.; Coxon, A.; DeMelfi Jr., T. M.; DiPietro, L.; Doerr, N.; Elbaum, D.; Estrada, J.; Gallant, P.; Germain, J.; Gu, Y.; Harmange, J. C.; Kaufman, S. A.; Kendall, R.; Kim, J. L.; Kumar, G. N.; Long, A. M.; Neervannan, S.; Patel, V. F.;

Polverino, A.; Rose, P.; van der Plas, S.; Whittington, D.; Zanon, R.; Zhao, H. *J. Med. Chem.*, **2007**, *50*, 4351-4373.

27. Hodous, B. L.; Geuns-Meyer, S. D.; Hughes, P. E.; Albrecht, B. K.; Bellon, S.; Bready, J.; Caenepeel, S.; Cee, V. J.; Chaffee, S. C.; Coxon, A.; Emery, M.; Fretland, J.; Gallant, Paul; Gu, Y.; Hoffman, D.; Johnson, R. E.; Kendall, R.; Kim, J. L.; Long, A. M.; Morrison, M.; Olivieri, P. R.; Patel, V. F.; Polverino, A.; Rose, P.; Tempest, P.; Wang, L.; Whittington, D. A.; Zhao, H. *J. Med. Chem.*, **2007**, *50*, 611-626.

28. Miyazaki, Y.; Matsunaga, S.; Tang, J.; Maeda, Y.; Nakano, M.; Philippe, R. J.; Shibahara, M.; Liu, W.; Sato, H.; Wang, L.;. Nolteb, R. T. *Bioorg. Med. Chem. Lett.*, **2005**, *15*, 2203-2207.

29. Harris, P. A.; Cheung, Mui.; Hunter III, R. N.; Brown, M. L.; Veal, J. M.; Nolte, R. T.; Wang, L.; Liu, W.; Crosby, R. M.; Johnson, J. H.; Epperly, A. H.; Kumar, R.; Luttrell, D. K.; Stafford, J. A. *J. Med. Chem.*, **2005**, *48*, 1610-1619.

30. Johnson, L. N.; Noble, M. E.; Owen D. J. *Cell*, **1996**, *85*, 149-158.

31. Meyer, R. D.; Singh, A. J.; Rahimi, N. *J. Biol. Chem.*, **2004**, *279*, 735-742.

32. Kornev, A. P.; Haste, N. M.; Taylor, S. S.; Eyck, L. F. *Proc. Natl. Acad. Sci. U.S.A.*, **2006**, *103*, 17783-17788 and references therein.

33. Kornev, A. P.; Taylor, S. S. *Biochim. Biophys. Acta*, **2010**, *1804*, 440–444 and references therein.

34. Taylor, S. S.; Kornev, A. P. *Trends Biochem. Sci.*, **2010**, *DOI:10.1016/ j.tibs.2010.09.006* and references therein.

35. Huse, M.; Kuriyan, J. *Cell*, **2002**, *109*, 275-82.

36. Jacobs, M. D.; Caron, P. R.; Hare, B. J. *Proteins*, **2008**, *70*, 1451-1460.

37. Hubbard, S. R.; Wei, L.; Ellis, L.; Hendrickson, W. A. *Nature*, **1994**, *372*, 746-754.

38. Hubbard, S. R. *EMBO J.*, **1997**, *16*, 5572-5581.

39. Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. *J. Med. Chem.*, **2010**, *53*, 2681-2694.

40. Garuti, L.; Roberti, M.; Bottegoni, G. *Curr. Med. Chem.*, **2010**, *17*, 2804-2821.

41. Simard, J. R.; Getlik, M.; Grtter, C.; Schneider, R.; Wulfert, S.; Rauh, D. *J. Am. Chem. Soc.*, **2010**, *132*, 4152-4160.

42. Neidle, S. *Prog. Med. Chem.*, **1979**, *16*, 151-221.

43. Berman, H. M.; Yuong, P. R. *Annu. Rev. Biophys. Bioeng.*, **1981**, *10*, 87-114.

44. Wilson, W. P.; Jones, R. L.; *Adv. Pharmcol. Chemother.*, **1981**, *18*, 177-222.

45. Watson, J. D.; Crick, F. H. C. *Nature*, **1953**, *171*, 737-738.

46. Bessi, I. *Tesi di Laurea Specialistica in Chimica, Università di Firenze*, **2010**.

47. Bansal, M. *Curr. Sci.*, **1999**, *76*, 1178-1181.

48. Brown, T.; Hunter, W. N.; Kneale, G.; Kennard, O. *Proc. Natl. Acad. Sci. U.S.A.*, **1986**, *83*, 2402-2406.

49. Leonard, G. A.; Booth, E. D.; Brown, T. *Nucleic Acids Res.*, **1990**, *18*, 5617-5623.

50. Skelly, J. V.; Edwards, K. J.; Jenkins, T. C.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.*, **1993**, *90*, 804-808.

51. Haider, S.; Parkinson, G. N.; Read, M. A.; Neidle, S. *DNA and RNA Binders. From Small Molecules to Drugs (Eds.: M. Demeunynck, C. Bailly, W. D. Wilson), Wiley-VCH, Weinheim*, **2003**, pp. 337-359.

52. Reed, J. E.; Neidle, S.; Vilar, R. *Chem. Commun.*, **2007**, 4366-4368.

53. Parkinson, G. N.; Ghosh, R.; Neidle, S. *Biochemistry* **2007**, *46*, 2390-2397.

54. Campbell, N. H.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. *J. Am. Chem. Soc.*, **2008**, *130*, 6722-6724.

55. Dai, J.; Carver, M.; Yang, D. *Biochimie*, **2008**, *90*, 1172-1183, and references therein.

56. Neidle, S. *Curr. Opin. Struct. Biol.*, **2009**, *19*, 239-250.

57. Hazel, P.; Parkinson, G. N.; Neidle, S. *Nucleic Acids Res.*, **2006**, *34*, 2117-2127.

58. Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.*, **2005**, *33*, 2908-2916.

59. Mergny, J. L.; Helene, C. *Nat. Med.*, **1998**, *4*, 1366-1367.

60. Sun, D. Y.; Hurley, L. H. Methods *in Enzymology, Drug–Nucleic Acid Interactions, Vol. 340 (Eds.: J. B. Chaires, M. J. Waring), Academic Press Inc., San Diego*, **2001**, pp. 573-592.

61. Hurley, L. H. *Nat. Rev. Cancer*, **2002**, *2*, 188-200.

62. Neidle, S.; Parkinson, G. N. *Nat. Rev. Drug Discov.*, **2002**, *1*, 383-393.

63. Bodnar, A. G.; Ouellette, M.; Frolkis, M.; Holt, S. E.; Chiu, C. P.; Morin, G. B.; Harley, C. B.; Shay, J. W.; Lichtsteiner, S.; Wright, W. E. *Science*, **1998**, *279*, 349-352.

64. Harley, C. B. *Mutat. Res.*, **1991**, *256*, 271-282.

65. Hackett, J. A.; Feldser, D. M.; Greider, C. W. *Cell*, **2001**, *106*, 275-286.

66. Sun, H.; Karow, J. K.; Hickson, I. D.; Maizels, N. *J. Biol. Chem.*, **1998**, *273*, 27587-27592.

67. Sun, H.; Bennett, R. J.; Maizels, N. *Nucleic Acids Res.*, **1999**, *27*, 1978-1984.

68. Harley, C. B.; Futcher, A. B.; Greider, C. W. *Nature*, **1990**, *345,* 458-460.

69. Greider, C. W.; Blackburn, E. H. *Cell*, **1985**, *43*, 405-413.

70. Morin, G. B. *Cell*, **1989**, *59*, 521-529.

71. Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L. C.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. *Science*, **1994**, *266*, 2011-2015.

72. De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J.-F.; Mergny, J.-L. *Biochimie*, **2008**, *90*, 131-155.

73. Kelland, L. *Clin. Cancer Res.*, **2007**, *13*, 4960-4963.

74. Oganesian, L.; Bryan, T. M. *Bioessays*, **2007**, *29*, 155-165.

75. Arora, A.; Kumar, N.; Agarwal, T.; Maiti, S. *Febs J.*, **2010**, *277*, 1345.

76. Neidle, S. *Febs J.*, **2010**, *277*, 1118-1125.

77. Wang, Y.; Patel, D. J. *Structure*, **1993**, *1*, 263-282.

78. Ambrus, A.; Chen, D.; Dai, J. X.; Bialis, T.; Jones, R. A.; Yang, D. Z. *Nucleic Acids Res.*, **2006**, *34*, 2723-2735.

79. Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. Z. *Nucleic Acids Res.*, **2007**, *35*, 4927-4940.

80. Xu, Y.; Noguchi, Y.; Sugiyama, H. *Bioorg. Med. Chem.*, **2006**, *14*, 5584-5591.

81. Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature*, **2002**, *417*, 876-880.

82. Balasubramanian, S.; Neidle, S. *Quadruplex Nucleic Acids, RSC: Cambridge*, **2006**.

83. Haider, S. M.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.*, **2003**, *326*, 117-125.

84. Schouten, J. A.; Ladame, S.; Mason, S. J.; Cooper, M. A.; Balasubramanian, S. *J. Am. Chem. Soc.*, **2003**, *125*, 5594-5595.

85. Harrison, R. J.; Gowan, S. M.; Kelland, L. R.; Neidle, S. *Bioorg. Med. Chem. Lett.*, **1999**, *9*, 2463-2468.

86. Read, M. A.; Wood, A. A.; Harrison, J. R.; Gowan, S. M.; Kelland, L. R.; Dosanjh, H. S.; Neidle, S. *J. Med. Chem.*, **1999**, *42*, 4538-4546.

87. Schultes, C. M.; Guyen, W.; Cuesta, J.; Neidle, S. *Bioorg. Med. Chem. Lett.*, **2004**, *14*, 4347-4351.

88. Read, M. A.; Harrison, R. J.; Romagnoli, B.; Tanious, F. A.; Gowan, S. H.; Reszka, A. P.; Wilson, W. D.; Kelland, L. R.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.*, **2001**, *98*, 4844-4849.

89. Gowan, S. M., Harrison, J. R.; Patterson, L.; Valenti, M.; Read, M. A.; Neidle, S.; Kelland, L. R. *Mol. Pharmacol.*, **2002**, *61*, 1154-1162.

90. Harrison, R. J.; Cuesta, J.; Chessari, G.; Read, M. A.; Basra, S. K.; Reszka, A. P.; Morrell, J.; Gowan, S. M.; Incles, C. M.; Tanious, F. A.; Wilson, W. D.; Kelland, L. R.; Neidle, S. *J. Med. Chem.*, **2003**, *46*, 4463-4476.

91. Martins, C.; Gunaratnam, M.; Stuart, J.; Makwana, V.; Greciano, O.; Reska, A. P.; Kelland, L. R.; Neidle, S. *Bioorg. Med. Chem. Lett.*, **2007**, *17*, 2293-2298.

92. Alberti, P.; Ren, J.; Teulade-Fichou, M. P.; Guittat, L.; Riou, J. F.; Chaires, J. B.; Hélène, C.; Vigneron, J. P.; Lehn, J. M.; Mergny, J. L. *J. Biomol. Struct. Dyn.*, **2001**, *19*, 505-513.

93. Kaiser, M.; De Cian, A.; Sainlos, M.; Renner, C.; Mergny, J. L.; Teulade- Fichou, M. P. *Org. Biomol. Chem.*, **2006**, *4*, 1049-1057.

94. De Cian, A.; Cristofari, G.; Reichenbach, P.; De Lemos, E.; Monchaud, D.; Teulade-Fichou, M. P.; Shin-Ya, K.; Lacroix, L.; Lingner, J.; Mergny, J. L. *Proc. Natl. Acad. Sci. U.S.A.*, **2007**, *104*, 17347-17352.

95. Burger, A. M.; Dai, F. P.; Schultes, C. M.; Reszka, A. P.; Moore, M. J.; Double, J. A.; Neidle, S. *Cancer Res.*, **2005**, *65*, 1489-1496.

96. Harrison, R. J.; Reszka, A. P.; Haider, S. M.; Romagnoli, B.; Morrell, J.; Read, M. A.; Gowan, S. M.; Incles, C. M.; Kelland, L. R.; Neidle, S. *Bioorg. Med. Chem. Lett.*, **2004**, *14*, 5845-5849.

97. Haider, S.; Neidle, S. *Methods Mol. Biol.*, **2010**, *608*, 17-37.

98. Bazzicalupi, C.; Chioccioli, M.; Sissi, C.; Porcù, E.; Bonaccini, C.; Pivetta, C.; Bencini, A.; Giorgi, C.; Valtancoli, B.; Melani, F.; Gratteri, P. *ChemMedChem.*, **2010**, *5*, 1995-2005.

99. a) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Struc. Funct. Genetics,* **2002**, *47*, 409-443. b) Brooijmans, N.; Kuntz, I. D. *Annu. Rev. Biophys. Biomol. Struct.*, **2003**, *32*, 335-373.

100. Russell, R. B.; Alber, F.; Aloy, P.; Davis, F. P.; Korkin, D.; Pichaud, M.;   Topf, M.; Sali, A. *Curr. Opin. Struct. Biol.*, **2004**, *14*, 313-324.

101. van Dijk, A. D.; Boelens, R.; Bonvin, A. M. *FEBS J.*, **2005**, *272*, 293-312.

102. May, A.; Zacharias, M. *Biochem. Biophys. Acta*, **2005**, *1754*, 225-231.

103. Huang, S. Y.; Zou, X. *Proteins*, **2007**, *66*, 399-421.

104. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.*, **2004**, *47*, 1739-1749.

105. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J. Med. Chem.*, **2004**, *47*, 1750-1759.

106. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. *J. Med. Chem.*, **2006**, *49*, 6177-6196.

107. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications,* Academic Press, London, **2002**.

108. Rapaport, D. C. *The Art of Molecular Dynamics Simulation*, Cambridge Univ. Press., **2004**.

109. Procacci, P.; Paci, E.; Darden, T.; Marchi, M. *J. Comp. Chem.*, **1997**, *18*, 1848-1862.

110. Marchi, M.; Procacci, P. *J. Chem. Phys.*, **1998**, *109*, 5194-5202.

111. Procacci, P.; Marchi, M.; Martyna, G. J. *J. Chem. Phys.*, **1998**, *108*, 8799-8803.

112. Marsili, S.; Signorini, G. F.; Chelli, R.; Marchi, M.; Procacci, P. *J. Comp. Chem.*, **2010**, *31*, 1106-1116.

113. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. AMBER9, University of California, San Francisco, CA (USA), **2006**.

114. Ewald, P. P. *Ann. Phys.*, **1921**, *64*, 253-287.

115. deLeeuw, S. W.; Perram, J. W.; Smith, E. R. *Proc. R. Soc. London*, **1980**, *A373*, 27-56.

116. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.*, **1993**, *98*, 10089-10092.

117. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.*, **1995**, *103*, 8577-8593.

118. Procacci, P.; Darden, T.; Marchi, M. *J. Phys. Chem.*, **1996**, *100*, 10464-10468.

119. Tuckerman, M. E.; Berne, B. J.; Martyna, G. J. *J. Chem. Phys.*, **1992**, *97***,** 1990-2001.

120. Humphreys, D. D.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem.*, **1994**, *98***,** 6885-6892.

121. Procacci, P.; Berne, B. J. *Mol. Phys.,* **1994***, 83, 255–272.*

122. Procacci, P.; Marchi, M. *J. Chem. Phys.*, **1996**, *104***,** 3003-3012.

123. Andersen, H. C. *J. Chem. Phys.*, **1980**, *72***,** 2384- 2393.

124. Parrinello, M.; Rahman, A. *Phys. Rev. Lett.*, **1980**, *45***,** 1196-1199.

125. Nose, S. *J. Chem. Phys.*, **1984**, *81***,** 511-519.

126. Duan, Y.; Kollman, P. A. *Science*, **1998**, *282*, 740-744

127. Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.*, **1986**, *57*, 2607-2609.

128. Geyer, C. G. In *Computing Science and Statistics, Proceedings of the 23rd Symposium on the Interface, Keramidis E. M., Ed.; Interface Foundation: Fairfax, p. 156*, **1991**.

129. Marinari, E.; Parisi, G. *Europhys. Lett.*, **1992**, *19*, 451-458.

130. Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.*, **1996**, *65*, 1604-1608.

131. Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.*, **2002**, *116*, 9058-9067.

132. Huang, N.; Kalyanaraman, C.; Bernacki, K.; Jacobson, M. P. *Phys. Chem. Chem. Phys.*, **2006**, *8*, 5166-5177.

133. Gilson, M. K.; Zhou, H. X. *Annu. Rev. Biophys. Biomol. Struct.*, **2007**, *36*, 21-42.

134. Gohlke, H.; Case, D. A. *J. Comp. Chem.,* **2004**, *25*, 238-250.

135. Kottalam, J.; Case, D. A. *Biopolymers*, **1990**, *29*, 1409-1421.

136. Protein Preaparation Wizard, Schrödinger, LLC, New York, NY (USA), **2007**.

137. Guex, N.; Peitsch, M. C. *Electrophoresis*, **1997**, *18*, 2714-2723. http://www.expasy.org/spdbv/

138. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. In *Biomolecular Simulation: The GROMOS96 Manual and User Guide; vdf Hochschulverlag AG an der ETH Zürich and BIOMOS b.v.: Zürich, Groningen*, **1996**.

139. Sippl, J. M. *J. Mol. Biol.*, **1990**, *213*, 859-883.

140. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.*, **1994**, *22*, 4673-4680.

141. Hall, T. A. *Nucleic Acids Symp. Ser.*, **1999**, *41*, 95-98.

142. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.*, **2003**, *24*, 1999-2012.

143. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, L. M. *J. Chem. Phys.*, **1983**, *79*, 926-935.

144. Homeyer, N.; Horn, A. H. C.; Lanig, H.; Sticht, H. *J. Mol. Model.*, **2006**, *12*, 281-289.

145. Frembgen-Kesner, T.; Elcock, A. H. *J. Mol. Biol.*, **2006**, *359*, 202-214.

146. Maestro v. 8.5, Schrödinger, LLC, New York, NY (USA), **2007**.

147. Glide, v. 4.5, Schrödinger, LLC, New York, NY (USA), **2007**.

148. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A.; *J. Phys. Chem.*, **1993**, *97*, 10269-10280.

149. Gaussian 09 (Revision A.01), Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.;

Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian Inc., Wallingford, CT (USA), **2009**.

150. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.*, **1977**, *23*, 327-341.

151. Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.*, **1984**, *81*, 3684-3690.

152. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.*, **2001**, *98*, 10037-10041.

153. Sitkoff, D.; Sharp, K. A.; Honing, B. *J. Phys. Chem.*, **1994**, *98*, 1978-1988.

154. Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. *J. Appl. Crystallogr.*, **2008**, *41*, 466- 470.

155. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics*, **1996**, *14*, 27-28.

Appendix

FULL PAPER

Synthesis, Biological Evaluation and Docking Studies of Casuarine Analogues: Effects of Structural Modifications at Ring B on Inhibitory Activity Towards Glucoamylase

Claudia Bonaccini,*[a] Matteo Chioccioli,[a] Camilla Parmeggiani,[b] Francesca Cardona,*[b] Daniele Lo Re,[b] Gianluca Soldaini,[b] Pierre Vogel,[c] Claudia Bello,[c] Andrea Goti,[b] and Paola Gratteri[a]

We report the total synthesis of a series of pyrrolizidine analogues of casuarine (1) and their 6-O-α-glucoside derivatives. The synthetic strategy is based on a totally regio- and stereoselective 1,3-dipolar cycloaddition of suitably substituted alkenes and a carbohydrate-based nitrone. We also report the evaluation of the biological activity of casuarine and its derivatives towards a wide range of glycosidases and a molecular modeling study focused on glucoamylase (GA) in which the binding modes of the newly synthesized compounds within the enzyme cavity are investigated. The results highlight the prominent structural features of casuarine and its derivatives that make them selective glucoamylase inhibitors.

Introduction

Iminosugars are very attractive carbohydrate mimics in which the endocyclic oxygen atom is replaced by the more basic, trivalent nitrogen atom.[1] In their protonated form, iminosugars resemble the transition state or intermediate generated during the hydrolysis reaction catalysed by glycosidases, key hydrolytic enzymes involved in many physiological functions. Since the discovery of the inhibitory properties of iminoalditols towards glycosidases, they have received increasing attention as diagnostic compounds as well as tools for the investigation of the structures, functions and catalytic mechanisms of carbohydrate-processing enzymes.[2,3] Furthermore, given the important role of glycosidases and glycosyltransferases in controlling the structures and functions of carbohydrates at the cell surface, competitive inhibitors of these classes of enzymes are potential anti-diabetes, anti-viral and anti-cancer agents.[1,2] Recently, interesting immunosuppressive activities have been discovered for this class of compounds.[4] In the past 40 years, more than 100 polyhydroxylated alkaloids have been isolated from plants and microorganisms[5] with structures that include polyhydroxylated piperidines, pyrrolidines, indolizidines, pyrrolizidines and nortropanes. For instance, the piperidine alkaloid 1-deoxynojirimycin (DNJ, Scheme 1), prepared first by Paulsen et al. in 1967[6a] and then isolated from a species of *Moris* (Moraceae),[6b] was found to strongly inhibit α-glucosidases.[5a] *N*-Alkylated derivatives of DNJ have found applications as anti-diabetic drugs (i.e., Miglitol, Glyset) or anti-HIV agents (Glycovir, SC 49483).[2a] The indolizidine alkaloid (+)-lentiginosine (Scheme 1) was isolated in 1990 from the leaves of *Astragalus lentiginosus* and was found to inhibit amyloglucosidases.[7] Its non-natural enantiomer (–)-lentiginosine was recently discovered to possess proapoptotic activity towards tumoral cells.[8] Castanospermine (Scheme 1), isolated in 1981 from the seeds, leaves and barks of *Castanospermum australe* and in 1988 from the seeds, leaves and barks of *Alexa sp.*,[5a] and its ester and salt derivatives are able to inhibit tumour growth and metastasis.[9]

Casuarine (1, Scheme 2) and its 6-O-α-glucoside, casuarine-6-O-α-glucopyranoside (2, Scheme 2), have been isolated from the bark of *Casuarina equisetifolia* L. (Casuarinaceae) and from the leaves of *Eugenia jambolana* Lam. (Myrtaceae).[10] We recently reported that casuarine (1) is able to inhibit a human maltase-glucoamylase (MGAM,

[a] Laboratory of Molecular Modeling Cheminformatics & QSAR, Department of Pharmaceutical Sciences, Laboratory of Design, Synthesis and Study of Biologically Active Heterocycles (HeteroBioLab), University of Florence,
Via Ugo Schiff 6, 50019 Sesto Fiorentino, Firenze, Italy
Fax: +39-055-4573780
E-mail: claudia.bonaccini@unifi.it
paola.gratteri@unifi.it
[b] Department of Chemistry "Ugo Schiff", Laboratory of Design, Synthesis and Study of Biologically Active Heterocycles (HeteroBioLab), University of Florence,
Via della Lastruccia 3-13, 50019 Sesto Fiorentino, Firenze, Italy
Fax: +39-055-4573531
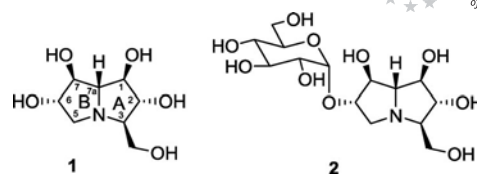E-mail: francesca.cardona@unifi.it
[c] Laboratory of Glycochemistry and Asymmetric Synthesis (LGSA), Swiss Institute of Technology (EPFL),
Batochime, 1015 Lausanne, Switzerland
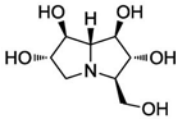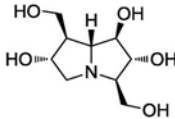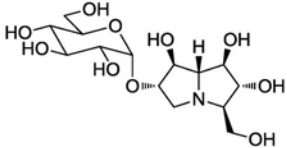Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/ejoc.201000632.
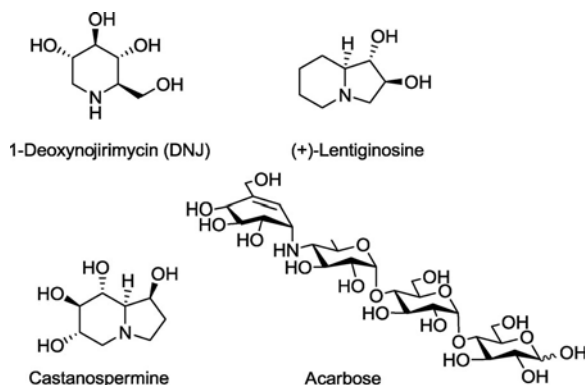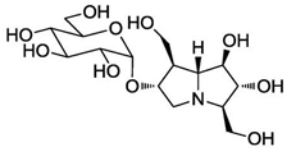
Scheme 1. Glycosidase inhibitors.

Scheme 2. Casuarine and its 6-O-α-D-glucoside.
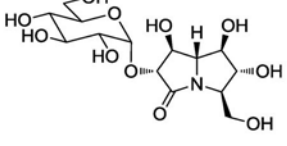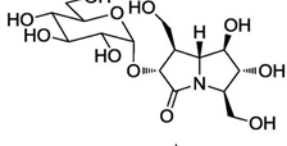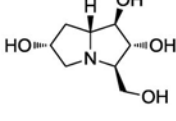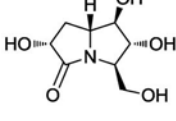
EC 3.2.1.20) more strongly than the pseudo-tetrasaccharide acarbose (Scheme 1) currently on the market as an antidiabetic drug (Glucobay, Precose) and thus has promise for the development of novel anti-diabetic drugs.[11]

Our total synthesis of casuarine (1) and its 6-O-α-gluco-side (2) took advantage of a complete stereoselective nitrone cycloaddition strategy with Tamao–Fleming oxidation and selective α-glucosylation as the key steps.[11]

Glucoamylase (1,4-α-D-glucan glucohydrolase, GA; EC 3.2.1.3; glycoside hydrolase family GH15, www.cazy.org) is an *exo*-hydrolase that catalyses the removal of glucose units from the non-reducing end of starch and related oligosaccharides. The hydrolytic reaction, which preferentially occurs at α-1,4 linkages, proceeds with inversion of configuration at the anomeric carbon atom. Glucoamylases are also able to hydrolyse α-1,6 linkages, but the specific activity is

Table 1. Structures and inhibition activities (IC$_{50}$) of compounds **1–11** towards glucoamylase from *aspergillus niger*.
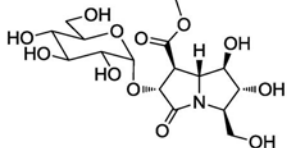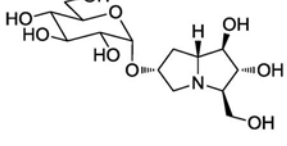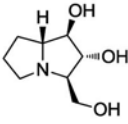
Molecule	Structure	IC$_{50}$ [μM]	Molecule	Structure	IC$_{50}$ [μM]
1 casuarine		1.9	**7** 7-homocasuarine		8.1
2		4.4	**8**		23.7
3		0%[a]	**9**		25%[a]
4 7-deoxycasuarine		4.5	**10**		76%[a]
5		7.7	**11** hyacinthacine A$_2$ (6,7-dideoxycasuarine)		1.9
6		214			

[a] For compounds with an inhibition percentage less than 90% at 1 mM concentration, the IC$_{50}$ values were not determined, and the percentage inhibition is reported.

only 0.2% with respect to α-1,4 hydrolysis.[12] The interest in glucoamylase is related to its use in the industrial production of bioethanol, glucose and fructose syrups.[13] Furthermore, given the presence of these enzymes in a wide variety of organisms and their quite simple obtainment in a pure form, GA has been extensively studied as a model for other members of the glycosyl hydrolase family.[14]

In this paper we report the total synthesis of the pyrrolizidine analogues of casuarine **4**, **6**, **7** and **11** (see Table 1), all bearing the same stereochemical pattern at the more substituted five-membered ring A as that of casuarine, and their 6-O-α-glucoside derivatives **3**, **5** and **8**–**10**. We also report the evaluation of the inhibitory activities of compounds **1**–**11** towards a wide range of commercially available glycosidases and a molecular modelling study on glucoamylase (GA) from *Aspergillus awamori* as a result of the selective inhibitory activity towards GA shown by the tested compounds.

## Results and Discussion

Synthesis

The general strategy followed for the synthesis of the pyrrolizidine alkaloids is outlined in Scheme 3. We took advantage of a stereocontrolled cyclic nitrone cycloaddition strategy[15] employing polyfunctionalized nitrone **12** and suitable dipolarophiles **13**, which afforded regio- and stereoselectively isoxazolidines **14**. These were then converted into pyrrolizidinone derivatives **15** by reductive ring-opening/cyclization. Intermediates **15** bear a free hydroxy group at C6 of the pyrrolizidine ring, which allows selective glucosylation at this position. A good choice of dipolarophile is crucial for the success of the strategy. For instance, for the synthesis of casuarine (**1**) a good regioselectivity of the cycloaddition was assured by using dipolarophile **13** with Y = SiMe₂Ph and X = OEt.[11]

a: X = NMe₂, Y = H
b: X = OMe, Y = CO₂Me
c: X = OEt, Y = SiMe₂Ph
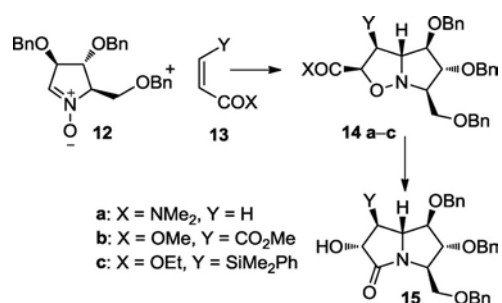
Scheme 3. General procedure for the synthesis of the pyrrolizidine alkaloids.

Nitrone **12** was conveniently prepared on a multigram scale by starting from commercially available tribenzyl D-arabinose.[16] It has the absolute configuration of the stereogenic centres at C1, C2 and C3 required for casuarine and its analogues such as non-natural 7-deoxycasuarine (**4**),[16a,17] its lactam derivative **6** and hyacinthacine A₂ (**11**).[16a,18]

The first key step in the synthesis of 7-deoxycasuarine (**4**), its lactam derivative **6** and hyacinthacine A₂ (**11**) is the 1,3-dipolar cycloaddition of nitrone **12** to dimethylacrylamide, which, after N–O bond cleavage of the isoxazolidine **14a** with Zn in acetic acid gave lactam **16** in 68% yield over two steps (Scheme 4). Compound **16** is the key intermediate for the total synthesis of all three target molecules. Hydrogenation in EtOH catalysed by Pd/C afforded lactam **6** in 88% yield. Reduction of the C=O bond with LiAlH₄ in THF at reflux gave compound **17** in 75% yield, which, after catalytic hydrogenation in EtOH, afforded 7-deoxycasuarine (**4**) in 88% yield (Scheme 4). Deoxygenation at C6 was achieved through the mesylation of **16** followed by reduction with LiAlH₄ in THF at reflux. This gave **18** in 80% yield over two steps. Finally, catalytic hydrogenation gave hyacinthacine A₂ (**11**) in 72% yield (Scheme 4).
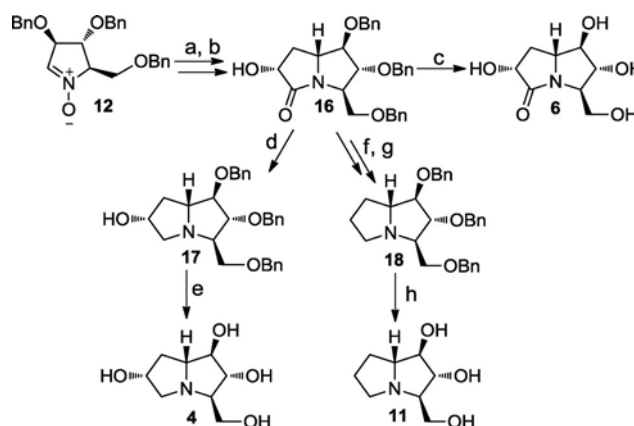
Scheme 4. Syntheses of 7-deoxycasuarine (**4**), its lactam derivative **6** and hyacinthacine A₂ (**11**): Reagents and conditions: (a) dimethylacrylamide, CH₂Cl₂, room temp., 3 d, 85%; (b) Zn, AcOH/H₂O, 50 °C, 4 h, 80%; (c) H₂, Pd/C, EtOH, 3 d, 88%; (d) LiAlH₄, THF, reflux, 3 h, 75%; (e) H₂, Pd/C, HCl, EtOH, 3 d, 88%; (f) MsCl, NEt₃, CH₂Cl₂, room temp., 2 h, 100%; (g) LiAlH₄, THF, reflux, 1.5 h, 80%; (h) H₂, Pd/C, HCl, MeOH, 3 d, 72%.

The lactam intermediate **16** was also employed in the synthesis of glucoside **5**; the initial selective α-glucosylation provided compound **19**.[19] Reduction of its amide moiety with LiAlH₄ followed by catalytic hydrogenolysis gave **5** in 45% yield over two steps (Scheme 5).
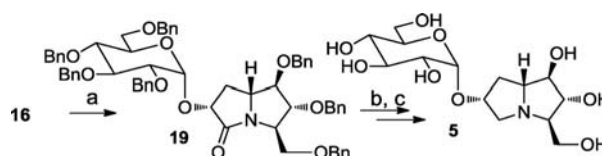
Scheme 5. Synthesis of 7-deoxycasuarine glucosyl derivative **5**: Reagents and conditions: (a) 2,3,4,6-tetra-O-benzylglucopyranosyl trichloroacetimidate, TMSOTf, Et₂O, 1 h, 88%; (b) LiAlH₄, THF, room temp., 1 h, 58%; (c) H₂, Pd/C, HCl, MeOH, room temp., 18 h, 77%.

For the synthesis of the non-natural 7-homocasuarine (**7**) and of its glucosyl derivatives **8**–**10**, dimethyl maleate was chosen as the dipolarophile. Treatment of the isoxazolidine

14b with Zn in acetic acid at 50 °C for 3 h gave lactam **20** in 70% yield over two steps (Scheme 6). Reduction of both the ester and lactam moieties with excess LiAlH$_4$ in THF at reflux provided diol **21** quantitatively. Finally, catalytic hydrogenolysis with Pd/C as catalyst in the presence of HCl gave non-natural 7-homocasuarine (**7**) in 89% yield (Scheme 6).
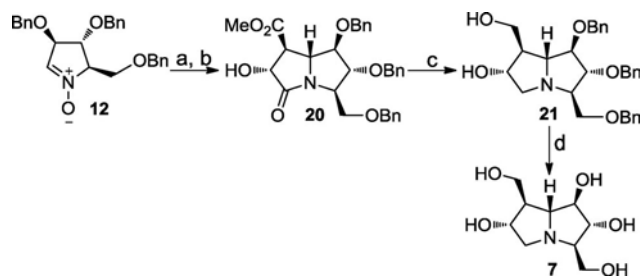
Scheme 6. Synthesis of 7-homocasuarine (**7**). Reagents and conditions: (a) dimethyl maleate, CH$_2$Cl$_2$, room temp., 4 d, 78%; (b) Zn, AcOH/H$_2$O, 50 °C, 3 h, 90%; (c) LiAlH$_4$, THF, reflux, 2 h, 100%; (d) H$_2$, Pd/C, HCl, EtOH, room temp., 4 d, 89%.

Lactam **20** was also selectively α-glucosylated to give α-glucoside **22** in 75% yield.[19] Intermediate **22** could be manipulated in different ways allowing us to obtain the three glucosyl derivatives **8–10**. Catalytic hydrogenolysis of **22** afforded compound **10** (96%), which retained both the ester and lactam moieties. Treatment of **22** with excess LiBH$_4$ and BH$_3$·THF[20] led to the complete reduction of both the ester and lactam moieties, and subsequent catalytic hydrogenolysis afforded the target glucosylated 7-homocasuarine **8** in 77% yield over two step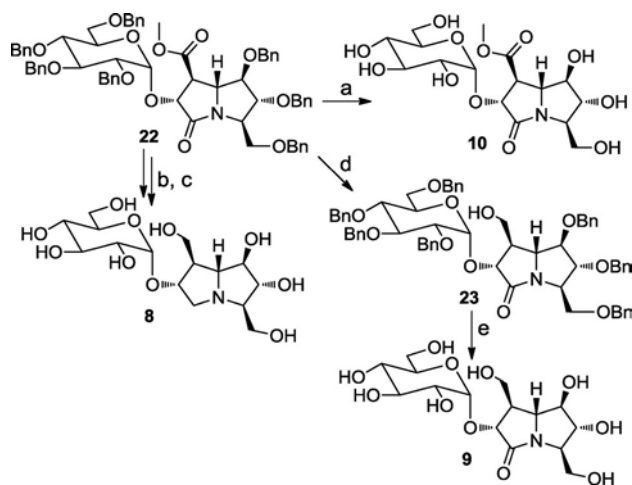s. Selective reduction of the ester moiety was achieved by treatment of **22** with LiBH$_4$ in THF at room temperature for 18 h. This afforded **23** in 62% yield. Finally, catalytic hydrogenolysis of **23** provided glucoside **9** in 72% yield (Scheme 7).

Scheme 7. Synthesis of 7-homocasuarine glucosyl derivatives **8–10**. Reagents and conditions: (a) H$_2$, Pd/C, MeOH/EtOAc, room temp., 4 d, 96%; (b) LiBH$_4$, BH$_3$·THF, THF, room temp., 11 d, 98%; (c) H$_2$, Pd/C, HCl, MeOH, room temp., 12 h, 79%; (d) LiBH$_4$, THF, room temp., 18 h, 62%; (e) H$_2$, Pd/C, MeOH, 24 h, 72%.

We recently presented the total synthesis of casuarine (**1**) and its 6-*O*-α-glucoside **2**,[11,21] as well as its epimer at C6, namely uniflorine A,[22] obtained from a derivative of lactam **15** (Y = SiMe$_2$Ph) by inversion of the configuration at C6. We have now also synthesized the lactam derivative **3** in 77% yield by hydrogenolysis of the fully protected lactam **24** (Scheme 8).[11]
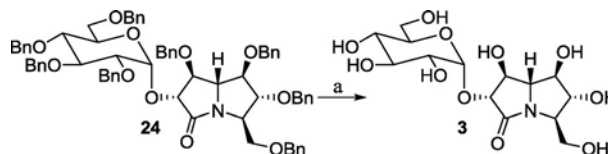
Scheme 8. Synthesis of casuarine 6-*O*-α-glucoside derivative **3**. Reagents and conditions: (a) H$_2$, Pd/C, MeOH/EtOAc, room temp., 24 h, 77%.

The lactams **10**, **9** and **3** were synthesized to investigate the importance of the basic nitrogen atom in glycosidase inhibition. With all these compounds in hand, we investigated their inhibitory activity towards a wide range of commercially available glycosidases.

Glycosidase Inhibitory Activities

Compounds **1–11** were assayed with respect to a panel of 13 commercially available glycosidases (Table 2): α-L-fucosidase (EC 3.2.1.51) from bovine kidney, α-galactosidase (EC 3.2.1.22) from coffee beans, β-galactosidase (EC 3.2.1.23) from *Escherichia coli* and *Aspergillus orizae*, α-glucosidase (EC 3.2.1.20) from yeast and rice, amyloglucosidase (EC 3.2.1.3) from *Aspergillus niger*, β-glucosidase (EC 3.2.1.21) from almonds, α-mannosidase (EC 3.2.1.24) from jack beans, β-mannosidase (EC 3.2.1.25) from snails, β-xylosidase (EC 3.2.1.37) from *Aspergillus niger*, β-*N*-acetylglucosaminidase (EC 3.2.1.30) from jack beans and bovine kidney with appropriate *p*-nitrophenyl glycoside substrates.[23] The errors in the measurements were estimated to be around 20% (statistical study carried out with model compounds) and in the concentrations to be around 10–15% (errors in sample weight). Casuarine (**1**) was found to be a potent and competitive inhibitor of amyloglucosidase from *Aspergillus niger* (IC$_{50}$ = 1.9 ± 0.4 μM, K_i = 2.0 ± 0.4 μM; ref.[24] IC$_{50}$ = 0.7 μM). It also inhibits α-glucosidase from yeast and rice (91 and 94% at 1 mM, respectively) and, to a lesser extent, β-glucosidase from almonds, α-mannosidase from jack beans, β-xylosidase from *Aspergillus niger* and β-*N*-acetylglucosaminidase from jack beans (46, 21, 24 and 16% at 1 mM, respectively). Glucoside **2** was slightly less potent with the IC$_{50}$ and K_i values in the same order of magnitude (IC$_{50}$ = 4.4 ± 0.9 μM, K_i = 3.9 ± 0.8 μM, mixed-type inhibition; ref.[24] IC$_{50}$ = 1.1 μM), but more selective than casuarine: indeed at 1 mM concentration it gave only 20% inhibition towards α-glucosidase from yeast and did not inhibit at all α-glucosidase from rice. Moreover, it showed only 19% inhibition towards α-L-fucosidase from bovine kidney (1 mM). Glucosides **5** and **8** and 7-homocasu-

arine (**7**) were good and selective competitive inhibitors of amyloglucosidase from *Aspergillus niger* (IC$_{50}$ = 7.7 ± 1.5, 8.1 ± 1.6 and 24 ± 4.7 μM and K_i = 7.4 ± 1.2, 11 ± 2.1 and 23 ± 4.6 μM, respectively). They inhibited α-glucosidase from yeast weakly (29, 45 and 55% at 1 mM, respectively) and did not inhibit α-glucosidase from rice. 7-Deoxycasuarine (**4**) was a very potent and selective competitive inhibitor of amyloglucosidase from *Aspergillus niger* (IC$_{50}$ = 4.5 ± 0.9 μM, K_i = 3.5 ± 0.7 μM). Among the other glycosidases assayed, only α-glucosidase from rice was also inhibited weakly (36% at 1 mM). Lactam **6** exhibited weak (IC$_{50}$ = 0.21 ± 0.04 mM) but very selective inhibitory ac-

tivity towards amyloglucosidase from *Aspergillus niger*. Hyacinthacine A$_2$ (6,7-dideoxycasuarine, **11**) was also a very strong inhibitor of this enzyme (IC$_{50}$ = 1.9 μM ± 0.4 μM, K_i = 2.6 μM ± 0.5 μM, non-competitive inhibition; ref.[25] IC$_{50}$ = 8.6 μM). Among the remaining compounds, **3** was not active towards any of the glycosidases assayed, **9** and **10** showed weak and very selective inhibitory activity (25 and 76% inhibition at 1 mM, respectively) towards amyloglucosidase from *Aspergillus niger*. Glucosides **5** and **8** together with the parent compound **2** have also been found to be potent inhibitors of bacterial and insect trehalases.[19]

Table 2. Inhibitory activity of compounds **1**–**11** towards commercially available glycosidases.[a]

Enzyme (pH)	1	2	3	4	5	6
α-L-Fucosidase Bovine kidney (6)	n.i.	19	n.i.	n.i.	n.t.	n.i.
α-Galactosidase Coffee beans (6)	n.i.	n.i.	n.i.	n.i.	n.t.	n.i.
β-Galactosidase *Escherichia coli* (7)	n.i.	n.i.	n.i.	n.i.	n.t.	n.i.
Aspergillus orizae (4)	n.i.	n.i.	n.i.	n.i.	n.t.	n.i.
α-Glucosidase Yeast (7)	91	20	n.i.	n.i.	29	n.i.
Rice (4)	94	n.i.	n.i.	36	n.i.	n.i.
Amyloglucosidase *Aspergillus niger* (5)	98 IC$_{50}$ = 1.9 μM K_i = 2.0 μM	97 IC$_{50}$ = 4.4 M K_i = 3.9 μM	n.i.	100 IC$_{50}$ = 4.5 μM K_i = 3.5 μM	97 IC$_{50}$ = 7.7 μM K_i = 7.4 μM	91 IC$_{50}$ = 0.21 mM
β-Glucosidase Almonds (5)	46	n.i.	n.i.	n.i.	n.t.	42
α-Mannosidase Jack beans (5)	21	n.i.	n.t.	n.i.	n.t.	n.i.
β-Mannosidase Snails (4)	n.i.	n.t.	n.t.	n.i.	n.t.	n.i.
β-Xylosidase *Aspergillus niger* (5)	24	n.t.	n.t.	n.i.	n.t.	n.i.
β-*N*-Acetylglucosaminidase Jack beans (5)	16	n.t.	n.i.	n.i.	n.t.	44
Bovine kidney (4)	n.i.	n.t.	n.i.	n.t.	n.t.	20

Enzyme (pH)	7	8	9	10	11
α-L-Fucosidase Bovine kidney (6)	n.i.	n.t.	n.i.	n.i.	n.t.
α-Galactosidase Coffee beans (6)	n.i.	n.t.	n.i.	n.i.	n.i.
β-Galactosidase *Escherichia coli* (7)	n.i.	n.t.	n.i.	n.i.	n.i.
Aspergillus orizae (4)	n.i.	n.t.	n.i.	n.i.	n.i.
α-Glucosidase Yeast (7)	45	55	n.i.	n.i.	72
Rice (4)	n.i.	n.i.	n.i.	n.i.	n.i.
Amyloglucosidase *Aspergillus niger* (5)	99 IC$_{50}$ = 8.1 μM K_i = 11 μM	92 IC$_{50}$ = 24 μM K_i = 23 μM	25	76	97 IC$_{50}$ = 1.9 μM K_i = 2.6 μM
β-Glucosidase Almonds (5)	n.t.	n.t.	n.i.	n.i.	n.i.
α-Mannosidase Jack beans (5)	n.t.	n.t.	n.i.	n.i.	n.i.
β-Mannosidase Snails (4)	n.t.	n.t.	n.t.	n.i.	n.i.
β-Xylosidase *Aspergillus niger* (5)	n.t.	n.t.	n.t.	n.i.	n.i.
β-*N*-Acetylglucosaminidase Jack beans (5)	n.t.	n.t.	n.t.	n.i.	n.i.
Bovine kidney (4)	n.t.	n.t.	n.t.	n.i.	n.i.

[a] Percentage inhibition at a concentration of 1 mM. n.i. = no inhibition, n.t. = test not performed.

As **2** is a glucoside we verified whether the amyloglucosidase was able to hydrolyse its glucosidic bond or not under our test conditions. Indeed, we had to consider the possibility that the inhibitory activities observed for **2** were due to the casuarine liberated by the hydrolysis reaction catalysed by amyloglucosidase. Thus, we performed a series of mass spectral analyses using HR-ESI-TOF-MS (high-resolution ESI mass spectrometry, positive ionization mode). As shown in Figure S48 (see the Supporting Information) the peak assigned to the glucoside (A), MW = 367.35, was detected in solution when the measurement was performed immediately after the addition of the enzyme (C) as well as after 20 min of incubation at room temperature (D) and after 20 min of incubation at 37 °C (E). We verified that the mixture of enzyme and buffer did not give similar signals (B). The spectra of the different assays performed with the buffered solution (phosphate) of the inhibitor in the presence of the enzyme neither showed peaks corresponding to the aglycon (MW = 205.21) nor to glucose (180.16) alone (see Figure S48 of the Supporting Information). This indicates that glucoside **2** is not hydrolysed significantly by the enzyme (amyloglucosidase from *Aspergillus niger*) under the conditions of our test.

Computational Studies

The data reported in Table 2 show that the casuarine derivatives presented here, with the exception of glycosylated lactams **3**, **9** and **10**, all inhibit more than 90% of *Aspergillus niger* amyloglucosidase activity with IC_{50} values ranging from 1.9 µM for pyrrolizidines **1** and **11** to 24 µM for glucoside **8**, with lactam **6** showing the weakest activity (0.21 mM). Furthermore, the active compounds showed a competitive inhibition profile, thus indicating a similar binding mode within the enzyme. Nevertheless, no significant differences were found in the inhibition activity of glucosides and their parent compounds, which clearly indicates a lack of correlation between inhibition and the ability to occupy a second subsite.

In the past years, several crystallographic structures of the proteolytic fragment of glucoamylase G2 from *Aspergillus awamori* (95% sequence identity with the *A. niger* protein) bound to different inhibitors have been reported,[26] which has made it possible to investigate the nature of the interaction between glucoamylase and its ligands in detail. The enzyme active site is characterized by an excess of negative charge, which has been principally ascribed to residues D55, E179 and E400 from the –1 subsite. Furthermore, E179 and E400 have been shown to be the putative catalytic acid and base, respectively, and the hydrolysis reaction was hypothesized to proceed through the formation of a glucopyranosyl cation intermediate after nucleophilic attack by a water molecule. With this information in mind and with the aim to interpret the biological data in a structural way, we decided to study the docking on the glucoamylase structure to investigate the possible binding mode of the casuarine derivatives presented here.

After completion of docking calculations, ring A of casuarine (**1**) was found to be deeply located within the –1 site where it is involved in an optimal hydrogen-bonding network involving C8–OH and C2–OH of the ligand and active-site residues R54 and D55 together with the nucleophilic water Wat501. In addition, C1–OH is positioned at a hydrogen-bond distance from both the carbonyl oxygen atom of residue L177 and Wat501 (Figure 1). All other pyrrolizidine molecules were oriented in a similar way and conserved these interactions, which have also been observed in the crystallographic complexes with both DNJ and acarbose.[26a,26b]
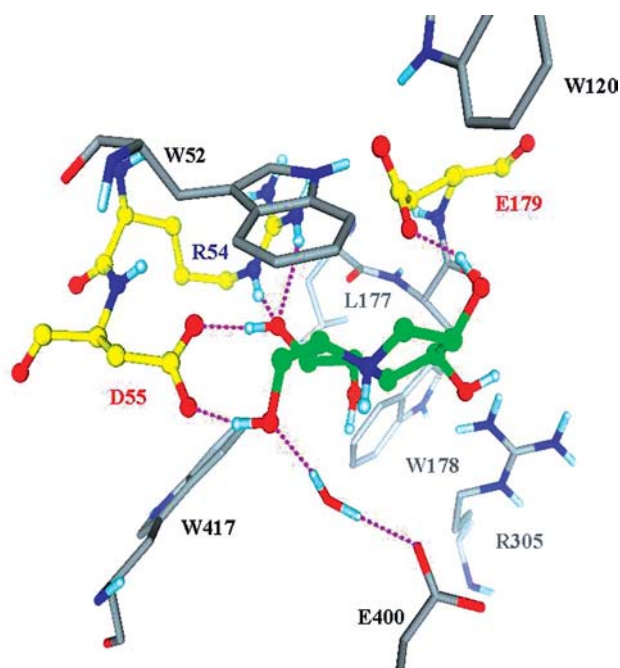
Figure 1. Docked orientation of molecule **1** within the glucoamylase active site. Hydrogen-bonds are depicted as magenta dashed lines.

As a consequence of this orientation, the protonated nitrogen atom of pyrrolizidine derivatives is not involved in the strong hydrogen-bond interaction with E179 observed for the nitrogen atom in the acarbose complex,[26b] but is oriented towards the region in which the nucleophile Wat501 lies, analogous to what was observed for the iminosugar-type inhibitor DNJ.[26a] We could hypothesize that the high affinity of acarbose ($K_i = 10^{-12}$ M)[26b] for glucoamylase is due, at least in part, to the presence of the charged hydrogen-bond interaction with E179 given that it has been suggested that charged hydrogen bonds can be responsible for a change in the binding constant by a factor of 1000.[27] Indeed, a maltoside hetero-analogue carrying a nitrogen atom at the interglycosidic linkage, which enables it to establish a charged hydrogen bond with E179, showed a 1000-fold stronger competitive inhibition than the analogue in which the interglycosidic atom is sulfur.[28] In contrast, a comparison of the crystal complexes of GA with acarbose and its weaker D-*gluco*-dihydro derivative ($K_i = 10^{-8}$ M) shows that the two molecules bind in a very similar

way (including the charged hydrogen bond with E179), and the 6 kcal/mol difference in binding energies could be largely attributed to unfavourable steric interactions between the hydrogen atoms at C7A of D-gluco-dihydroacarbose and the catalytic water,[26b] thus highlighting a comparable effect of electrostatic and hydrophobic interactions in guiding interactions with GA. Molecules **1**, **4**, **6** and **7** are able to bind E179 through a neutral hydrogen bond involving C6–OH, whereas the 6-deoxy derivative **11** clearly lacks this interaction. Anyway, apart from C6–OH when present, pyrrolizidine ligands are oriented towards the negatively charged E179, a very hydrophobic portion of the molecule, which could contribute negatively to the binding.

With respect to the crystallographic orientation of acarbose, the ring B atoms of the pyrrolizidines extend towards the +1 site, with C6 and C7 of **1** almost perfectly overlapped with C5B and N4B of the acarbose ring B, respectively. C7–OH of **1** is thus in close proximity to both R305 and the carbonyl oxygen atom of W178, but it does not present the correct geometry for hydrogen-bond formation, in contrast to what has been observed for C3B–OH of acarbose. The elimination of the C7 substituent (molecule **4**) led to a slight decrease in the IC$_{50}$ value as well as its elongation (molecule **7**), although this latter modification allowed the molecule to hydrogen-bond to both W178 and R305. Molecule **6**, which shows a very weak inhibition, is the only non-glycosylated lactam of this series, and after docking it is oriented like molecule **4**. Anyway, given the limited conformational flexibility imposed on molecule **6** by the presence of the lactam structure, the conformation of ring B is clearly influenced. Indeed, molecule **6** is also able to hydrogen-bond E179 through C6–OH, but the absence of the positive charge on the nitrogen atom eliminates the possibility of molecule **6** compensating the excess negative charge present in the glucoamylase active site, which has been hypothesized as one of the mechanisms involved in complex stabilisation.[26a,26b] Finally, the unfavourable effects due to the positioning of a hydrophobic portion of the ligand close to a charged amino acid (E179) is even more pronounced here given the higher hydrophobicity of **6** relative to **4**. All these considerations are in agreement with the very low IC$_{50}$ value found for **6**.

As far as the glucoside derivatives are concerned, the results of the docking calculations are comparable for all molecules, with ligands showing two possible binding modes in which either the pyrrolizidine or the glucose moiety is oriented in the –1 site; we call these two orientations CAS-IN and GLU-IN, respectively. In the CAS-IN orientation, the position of the pyrrolizidine nucleus is almost coincident with the docked pose of the corresponding unglucosylated compounds, except for the conformation of ring B, which is influenced by the positioning of the glucose moiety, and for the lack of a hydrogen-bond donor for E179 at the C6 atom. The glucose moiety is oriented in the outer part of the +1 site, where it can assume different conformations that allow it to hydrogen-bond E180 and/or Y311. When the molecules adopt the GLU-IN orientation, the glucose in the –1 site overlaps well with DNJ and with

ring A of acarbose. In all the selected poses showing this orientation the pyrrolizidine nucleus orients the N–H$^+$ towards the aromatic ring of Y311 in the +1 site with a geometry compatible with an NH· π interaction. Furthermore, C8–OH is able to donate one hydrogen bond to E180, thus contributing to the stabilization of this binding mode. Anyway, by comparison of the GLU-IN orientation of **2** with the docked orientation of the hydrolysable maltose, which perfectly overlaps the glucose moiety of **2** with its non-reducing end, it is not clear how the casuarine glucoside could resist hydrolysis, as we observed, because the nucleophile Wat501 is perfectly oriented towards the anomeric carbon atom of the ligand. In contrast, the CAS-IN binding mode of glucosides explains the resistance to hydrolysis that we observed for molecule **2** because the glucose moiety is located in the +1 site, far from the nucleophile Wat501. Moreover, given the longer C2–C6 distance of pyrrolizidine (4.9 Å) with respect to the C1–C4 distance of glucose (2.9 Å), the CAS-IN orientation allows ring B of pyrrolizidine to extend to the +1 site such that the glucose moiety is projected towards the third GA subsite (see Figure S49 of the Supporting Information). Interestingly, after docking, the lactam derivatives of glucosides **3**, **9** and **10**, which did not show inhibitory activity, all adopted the same GLU-IN orientation. The interactions of glucose at the –1 site are the same as those observed for active glucosides, and at the level of the +1 site the results of docking converged to a unique solution, which differ from the active glucosides for the aglycon conformation clearly influenced by the presence/absence of the lactam structure. Furthermore, the absence of charge on the nitrogen atom prevents lactam derivatives in the GLU-IN orientation from reinforcing the interaction at the +1 site through NH· π interactions. These observations on lactam molecules highlight some interesting features of the GA interaction: the fact that the CAS-IN orientation was never found for lactam-glucoside, together with the observations we made on the structural features of the interaction with molecule **6**, clearly indicate a poor binding of the lactam moiety to the –1 site due to both the absence of positive charge on the ligand and the steric effects associated with the lactam structure. Given also that at the +1 site the lack of charge on the ligand could be detrimental to the binding affinity and considering that the GLU-IN orientation of glucoside molecules could be associated with the hydrolysis of the molecule, it is not surprising that lactam molecules are not able to inhibit glucoamylase.

In summary, we have analysed the binding features of competitive pyrrolizidine inhibitors of glucoamylase as determined by docking simulations. None of the molecules we considered in this study presented structural variations in ring A, which in casuarine perfectly mimics the stereochemical arrangement of glucose. As far as ring B is concerned, none of the structural variations introduced at the 6- (including glycosylation) and 7-positions seem to significantly influence the inhibitory activity. In contrast, the presence of the lactam structure at the 4- and 5-positions has a very dramatic effect on the activity, and this could be due to

both the lack of a positive charge, whose role in stabilizing the complex has been already highlighted, and the limited conformational flexibility, which could determine unfavourable steric contacts.

Conclusions

We have reported a novel and efficient strategy for the synthesis of casuarine-like pyrrolizidines and their 6-O-α-glucoside derivatives. Our methodology was based on a totally regio- and stereoselective 1,3-dipolar cycloaddition of suitably substituted alkenes with a carbohydrate-based nitrone. After N–O bond cleavage of the cycloadducts thus obtained, the lactams were used as key intermediates in the synthesis of all the target compounds, including the glucosyl derivatives that were obtained by selective α-glucosylation. Evaluation of the inhibitory activity of casuarine and its derivatives towards a wide range of commercially available glycosidases was undertaken, and several new inhibitors of glucoamylase from *Aspergillus niger* were discovered. Docking experiments performed on pyrrolizidine derivatives allowed us to investigate the binding mode of the competitive inhibitors. It is evident that an optimal network of hydrogen-bonding interactions at the inner –1 site has to be achieved for a ligand to bind. The presence of a positive charge on the ligand is helpful for the stabilization of the complex, independently of the possibility that a ligand has to hydrogen-bond E179 through the charged atom. Finally, interactions at the +1 site also seem to have an important role in modulating the affinity of more extended molecules, but the comparison of **2** with acarbose clearly shows that to increase the affinity for GA it is necessary to bind subsites outside of the +1 site.

Experimental Section

General: Commercial reagents were used as received. All reactions were carried with magnetic stirring and were monitored by TLC on 0.25 mm silica gel plates (Merck F_{254}). Column chromatography was carried out on silica gel 60 (32–63 mm). Yields refer to spectroscopically and analytically pure compounds unless otherwise stated. ^{1}H NMR spectra were recorded with a Varian Mercury-400 spectrometer. ^{13}C NMR spectra were recorded with a Varian Gemini-200 spectrometer. Infrared spectra were recorded with a Perkin–Elmer Spectrum BX FT-IR System spectrophotometer. Mass spectra were recorded with a QMD 1000 Carlo Erba instrument by direct inlet injection; relative percentages are shown in parentheses. ESI full mass spectra were recorded with a Thermo LTQ instrument by direct inlet injection; relative percentages are shown in parentheses. HR-ESI-TOF-MS experiments were performed with a Q-Tof Ultima mass spectrometer (Waters) fitted with a standard Z-spray ion source and operated in the positive ionization mode. Elemental analyses were performed with a Perkin–Elmer 2400 analyser. Optical rotation measurements were performed with a JASCO DIP-370 polarimeter.

(1*R*,2*R*,3*R*,6*R*,7a*R*)-1,2-Bis(benzyloxy)-3-[(benzyloxy)methyl]-6-hydroxyhexahydro-5*H*-pyrrolizin-5-one (16): A mixture of **14a** (816 mg, 1.58 mmol) and Zn dust (407 mg) in CH_3COOH/H_2O (9:1, 12.5 mL) was heated at 50 °C for 4 h and then filtered through

cotton. The solution was cooled to 0 °C, and, under vigorous stirring, a saturated aqueous solution of $NaHCO_3$ (100 mL) was added until a basic pH was reached. The aqueous phase was extracted with EtOAc (3×20 mL), and the combined organic phases were dried with Na_2SO_4. After filtration and concentration under reduced pressure, **16** was obtained as a yellow oil, pure enough to be used in the next step (598 mg, 80%). An analytically pure sample was obtained through purification by flash column chromatography on silica gel (eluent: petroleum ether/EtOAc, 1:2, R_f = 0.4). $[a]_D^{20}$ = +3.05 (*c* = 0.9, CHCl$_3$). ^{1}H NMR (400 MHz, CDCl$_3$): δ = 7.38–7.26 (m, 15 H, Ar), 4.60–4.45 (m, 7 H, Bn, 6-H), 4.34 (t, *J* = 4.3 Hz, 1 H, 2-H), 4.11 (q, *J* = 4.5 Hz, 1 H, 3-H), 3.87 (br. s, 1 H, OH), 3.79 (dd, *J* = 6.9, 5.0 Hz 1 H, 1-H), 3.73 (dt, *J* = 8.6, 6.5 Hz, 1 H, 7a-H), 3.63 (dd, *J* = 9.8, 5.5 Hz, 1 H, 8-Ha), 3.52 (dd, *J* = 9.8, 4.1 Hz, 1 H, 8-Hb), 2.70 (ddd, *J* = 12.3, 7.8, 6.1 Hz, 1 H, 7-Ha), 1.80 (ddd, *J* = 12.3, 10.4, 8.6 Hz, 1 H, 7-Hb) ppm. ^{13}C NMR (50 MHz, CDCl$_3$): δ = 174.5 (s, C=O), 137.8, 137.7, 137.4 (s, Ar), 128.5–127.6 (d, 15 C, Ar), 89.1 (d, C-1), 85.7 (d, C-2), 73.3, 72.6, 72.4 (t, Bn), 72.0 (d, C-6), 68.8 (t, C-8), 59.9 (d, C-7a), 58.6 (d, C-3), 37.2 (t, C-7) ppm. IR (CDCl$_3$): $\tilde{v}$ = 3671, 3373, 3012, 2867, 1697, 1454, 1100 cm^{-1}. MS (EI): *m/z* (%) = 381 (12) [M – Bn]$^+$, 336 (20), 275 (32), 180 (98), 153 (100), 88 (100). $C_{29}H_{31}NO_5$ (473.56): calcd. C 73.55, H 6.60, N 2.96; found C 73.33, H 6.80, N 2.87.

(1*R*,2*R*,3*R*,6*R*,7a*R*)-1,2,6-Trihydroxy-3-(hydroxymethyl)hexahydro-5*H*-pyrrolizin-5-one (6): Pd (10% on C, 300 mg) was added to a stirred solution of **16** (150 mg, 0.32 mmol) in EtOH (15 mL) under nitrogen. The suspension was stirred under hydrogen at room temp. for 3 d, then filtered through Celite® and washed with MeOH. Concentration under reduced pressure afforded a viscous oil that was purified by flash column chromatography on silica gel (eluent: MeOH/EtOAc, 1:3, R_f = 0.14) to afford pure **6** as a transparent oil (57 mg, 88% yield). $[a]_D^{20}$ = –1.2 (*c* = 0.25, MeOH). ^{1}H NMR (400 MHz, CD$_3$OD): δ = 4.54 (dd, *J* = 10.5, 7.8 Hz, 1 H, 6-H), 4.16 (t, *J* = 6.3 Hz, 1 H, 2-H), 3.84 (dd, *J* = 3.0, 8.0 Hz, 1 H, 8-Ha), 3.67–3.53 (m, 4 H, 8-Hb, 1-H, 3-H, 7a-H), 2.78 (ddd, *J* = 12.2, 7.8, 5.9 Hz, 1 H, 7-Ha), 1.74 (ddd, *J* = 12.0, 10.7, 8.3 Hz, 1 H, 7-Hb) ppm. ^{13}C NMR (50 MHz, CD$_3$OD): δ = 175.3 (s, C-5), 81.8 (d, C-1), 78.3 (d, C-2), 71.7 (d, C-6), 62.0 (d, C-3), 60.2 (t, C-8), 59.8 (d, C-7a), 36.8 (t, C-7) ppm. MS: *m/z* (%) = 204 (5) [M + H]$^+$, 203 (3) [M]$^+$, 185 (27) [M – H$_2$O]$^+$, 172 (52), 144 (74), 126 (51), 100 (87), 86 (100), 72 (59), 57 (38). $C_8H_{13}NO_5$ (203.19): calcd. C 47.29, H 6.45, N 6.89; found C 47.54, H 6.36, N 6.95.

(1*R*,2*R*,3*R*,6*R*,7a*R*)-1,2-Bis(benzyloxy)-3-[(benzyloxy)methyl]hexahydro-1*H*-pyrrolizin-6-ol (17): A 1 M solution of LiAlH$_4$ in THF (1.6 mL, 1.61 mmol) was added to a cooled (0 °C) solution of **16** (255 mg, 0.54 mmol) in dry THF (6 mL) under nitrogen. The mixture was then heated at reflux for 1.5 h. Then, after cooling to 0 °C, an aqueous saturated solution of Na$_2$SO$_4$ (560 µL) was added dropwise. The suspension was then filtered through Celite® and washed with EtOAc. Concentration under reduced pressure afforded **17** as a yellow oil pure enough for the next step (185 mg, 75% yield). An analytically pure sample was obtained through purification by flash column chromatography on silica gel (eluent: petroleum ether/EtOAc, 1:4, R_f = 0.3). $[a]_D^{20}$ = +9.1 (*c* = 0.83, CHCl$_3$). ^{1}H NMR (400 MHz, CDCl$_3$): δ = 7.36-7.24 (m, 15 H, Ar), 4.67–4.46 (m, 6 H, Bn), 4.35–4.31 (m, 1 H, 6-H), 4.11 (t, *J* = 4.6 Hz, 1 H, 1-H), 4.07 (t, *J* = 4.7 Hz, 1 H, 2-H), 3.60-3.44 (m, 4 H, 7a-H, 3-H, 8-Ha, 8-Hb), 3.21 (dd, *J* = 12.3, 4.5 Hz 1 H, 5-Ha), 2.98 (dm, *J* = 12.3 Hz, 1 H, 5-Hb), 2.21 (ddd, *J* = 13.8, 9.0, 5.7 Hz, 1 H, 7-Ha), 1.84 (dm, *J* = 13.8 Hz, 1 H, 7-Hb) ppm. ^{13}C NMR (50 MHz, CDCl$_3$): δ = 138.1, 137.8, 137.4 (s, Ar), 128.1-127.2 (d, 15 C, Ar), 88.9 (d, C-1), 85.2 (d, C-2), 73.6 (d, C-6), 72.9, 72.1, 71.8, 71.6 (t, Bn, C-8), 69.9 (d, C-3), 67.4 (d, C-7a), 63.3 (t, C-5), 40.0 (t, C-7)

ppm. IR (CDCl$_3$): ṽ = 3392, 3010, 2927, 2858, 1748, 1710, 1454, 1262 cm^{-1}. MS (EI): m/z (%) = 366 (6), 336 (6), 216 (61), 160 (100), 90 (100). C$_{29}$H$_{33}$NO$_4$ (459.58): calcd. C 75.79, H 7.24, N 3.05; found C 75.99, H 7.16, N 3.02.

(1R,2R,3R,6R,7aR)-3-(Hydroxymethyl)hexahydro-1H-pyrrolizine-1,2,6-triol (7-Deoxycasuarine, 4): Concentrated HCl (4–5 drops) and Pd (10% on C, 250 mg) were added to a stirred solution of **17** (120 mg, 0.26 mmol) in EtOH (10.5 mL) under nitrogen. The suspension was stirred at room temp. under hydrogen for 3 d, then filtered through Celite® and washed with MeOH. Concentration under reduced pressure afforded a viscous yellow oil (66 mg) that was transferred to a column of DOWEX 50WX8 and then washed with MeOH (10 mL), H$_2$O (10 mL) to remove non-amine-containing products and then with 6% NH$_4$OH (15 mL) to elute 7-deoxycasuarine (**4**) as a white solid (43 mg, 88% yield), m.p. 205–208 °C. $[a]_D^{20}$ = +19.8 (c = 0.4, H$_2$O) {ref.[17a] $[a]_D^{20}$ = +10.9 (c = 0.11, H$_2$O); ref.[16b] $[a]_D^{25}$ = +23 (c = 0.3, MeOH)}. ^{1}H NMR (400 MHz, D$_2$O): δ = 4.44–4.38 (m, 1 H, 6-H), 4.05 (t, J = 8.2 Hz, 1 H, 1-H), 3.74–3.69 (m, 2 H), 3.56 (dd, J = 11.9, 6.5 Hz, 1 H), 3.24 (ddd, J = 12.6, 8.2, 4.2 Hz, 1 H), 3.09–3.02 (m, 2 H), 2.86 (m, 1 H, 5-Hb), 2.12 (ddd, J = 13.6, 8.5, 3.4 Hz, 1 H, 7-Ha), 1.89 (dm, J = 13.6 Hz, 1 H, 7-Hb) ppm. ^{13}C NMR (50 MHz, D$_2$O): δ = 82.6 (d, C-1), 79.0 (d, C-2), 75.1 (d, C-6), 72.5 (d, C-7a), 67.9 (d, C-3), 64.6 (t, C-8), 63.3 (t, C-5), 39.3 (t, C-7) ppm. MS (EI): m/z (%) = 190 (3) [M + H]$^+$, 189 (1) [M]$^+$, 176 (66), 158 (65) [M – CH$_2$OH]$^+$, 132 (64), 112 (23), 85 (62), 58 (100). C$_8$H$_{15}$NO$_4$ (189.21): calcd. C 50.78, H 7.99, N 7.40; found C 50.37, H 7.63, N 7.74.

(1R,2R,3R,7aR)-1,2-Bis(benzyloxy)-3-(benzyloxymethyl)hexahydro-1H-pyrrolizine (18): NEt$_3$ (75 µL, 0.54 mmol) was added to a stirred solution of **16** (95 mg, 0.20 mmol) in dry CH$_2$Cl$_2$ (0.45 mL) under nitrogen, and, at 0 °C, MsCl (20 µL, 0.26 mmol) was added dropwise. The solution was stirred at 0 °C for 30 min and at room temp. for 2 h. The mixture was filtered through Celite® and washed with EtOAc. Concentration under reduced pressure afforded the mesylated derivative as a white oil (quantitative yield), which was dissolved in dry THF (2.5 mL). A 1 M solution of LiAlH$_4$ in THF (0.8 mL, 0.8 mmol) was added dropwise at 0 °C under nitrogen. The mixture was heated at reflux for 3 h. An aqueous saturated solution of Na$_2$SO$_4$ (280 µL) was added dropwise, and the mixture was stirred at room temp. for 10 min. After filtration through Celite®, a crude residue (109 mg) was obtained that was purified by flash column chromatography on silica gel (petroleum ether/ethyl acetate, 1:2, R_f = 0.28) to afford pure **18** (71 mg, 80% yield) as an oil. $[a]_D^{24}$ = –5.1 (c = 0.6, CHCl$_3$). ^{1}H NMR (400 MHz, CDCl$_3$): δ = 7.37–7.25 (m, 15 H, Ar), 4.73–4.46 (m, 6 H, Bn), 4.08 (dd, J = 7.4, 5.9 Hz, 1 H, 2-H), 3.81 (t, J = 5.9 Hz, 1 H, 1-H), 3.60 (dd, J = 9.6, 4.7 Hz, 1 H, 8-Ha), 3.54 (dd, J = 9.6, 5.7 Hz, 1 H, 7a-H), 3.51–3.46 (m, 1 H, 8-Hb), 3.07 (dt, J = 10.5, 6.1 Hz, 1 H, 5-Ha), 2.98–2.94 (m, 1 H, 3-H), 2.78 (dt, J = 10.5, 6.6 Hz, 1 H, 5-Hb), 2.03–1.95 (m, 1 H, 7-Ha), 1.91–1.82 (m, 1 H, 6-Ha), 1.81–1.74 (m, 1 H, 6-Hb), 1.72–1.62 (m, 1 H, 7-Hb) ppm. ^{13}C NMR (50 MHz, CDCl$_3$): δ = 138.2, 138.1, 137.9 (s, Ar), 128.0–127.1 (d, 15 C, Ar), 88.6 (d, C-1), 85.5 (d, C-2), 73.0, 72.3, 71.8, 71.6 (t, C-8, Bn), 68.0 (d, C-3), 67.2 (d, C-7a), 54.8 (t, C-5), 31.4 (t, C-6), 25.5 (t, C-7) ppm.

(1R,2R,3R,7aR)-3-(Hydroxymethyl)hexahydro-1H-pyrrolizine-1,2-diol (Hyacinthacine A$_2$, 11): Concentrated HCl (3 drops) and Pd (10% on C, 45 mg) were added to a stirred solution of **18** (25 mg, 0.056 mmol) in MeOH (2.5 mL). The mixture was stirred at room temp. under hydrogen for 3 d. The mixture was then filtered through Celite® and washed with MeOH. The solvent was evaporated under reduced pressure to afford a viscous white oil that was

transferred to a column of DOWEX 50WX8 and then washed with MeOH (10 mL), H$_2$O (10 mL) to remove non-amine-containing products and then with 6% NH$_4$OH (15 mL) to elute hyacinthacine A$_2$ (**11**) as a white solid (7 mg, 72% yield). $[a]_D^{24}$ = +12.4 (c = 0.2, H$_2$O); {ref.[16a] $[a]_D^{24}$ = +12.7 (c = 0.13, H$_2$O); ref.[18a] $[a]_D$ = +12.5 (c = 0.4, H$_2$O); ref.[16c] $[a]_D^{20}$ = +19.9 (c = 0.97, MeOH); ref.[18b] $[a]_D^{25}$ = +10.5 (c = 0.6, H$_2$O); ref.[18f] $[a]_D$ = +12.1 (c = 0.3, H$_2$O); ref.[18g] $[a]_D^{26}$ = +12 (c = 0.4, H$_2$O); ref.[25] $[a]_D$ = +20.1 (c = 0.44, H$_2$O)}. ^{1}H NMR (200 MHz, D$_2$O): δ = 3.72–3.61 (m, 3 H), 3.53 (dd, J = 11.7, 6.2 Hz, 1 H), 3.10–3.00 (m, 1 H), 2.86–2.75 (m, 1 H), 2.70–2.56 (m, 2 H), 1.90–1.58 (m, 4 H) ppm. ^{13}C NMR (50 MHz, D$_2$O): δ = 82.6 (d, C-1), 79.6 (d, C-2), 71.7 (d, C-3), 68.6 (d, C-7a), 65.3 (t, C-8), 57.4 (t, C-5), 32.2 (t, C-7), 27.0 (t, C-6) ppm. C$_8$H$_{15}$NO$_3$ (173.21): calcd. C 55.47, H 8.73, N 8.09; found C 55.49, H 8.61, N 8.10.

Methyl (1S,2R,6R,7R,7aR)-6,7-Bis(benzyloxy)-5-[(benzyloxy)methyl]-2-hydroxy-3-oxohexahydro-1H-pyrrolizine-1-carboxylate (20): A mixture of **14b** (870 mg, 1.55 mmol) and Zn dust (400 mg, 6.2 mmol) in CH$_3$COOH/H$_2$O (9:1, 12.5 mL) was heated at 50 °C for 3 h and then filtered through cotton. The solution was cooled to 0 °C, and, under vigorous stirring, a saturated aqueous solution of NaHCO$_3$ (100 mL) was added until a basic pH was reached. The aqueous phase was extracted with EtOAc (3 × 60 mL), and the combined organic phases were dried with Na$_2$SO$_4$. After filtration and concentration under reduced pressure, **20** was obtained pure as a white solid (743 mg, 90% yield), m.p. 111–113 °C. $[a]_D^{20}$ = –28.7 (c = 0.64, CHCl$_3$). ^{1}H NMR (400 MHz, CDCl$_3$): δ = 7.37–7.23 (m, 15 H, Ar), 4.76 (dd, J = 9.8, 2.9 Hz, 1 H, 2-H), 4.57–4.44 (m, 6 H, Bn), 4.31–4.28 (m, 1 H, 5-H), 4.23–4.22 (m, 1 H, 6-H), 3.98–3.95 (m, 1 H, 7a-H), 3.92–3.90 (m, 1 H, 7-H), 3.79 (s, 3 H, Me), 3.58–3.50 (m, 2 H, 8-H), 3.44 (d, J = 3.1 Hz, 1 H, OH), 3.03 (t, J = 9.3 Hz, 1 H, 1-H) ppm. ^{13}C NMR (50 MHz, CDCl$_3$): δ = 173.2, 171.1 (s, C=O), 137.7–137.2 (s, 3 C, Ar), 128.4–127.5 (d, 15 C, Ar), 87.3 (d, C-7), 84.6 (d, C-6), 74.2 (d, C-2), 73.1, 72.1, 71.8 (t, Bn), 68.1 (t, C-8), 62.7 (d, C-7a), 59.3 (d, C-5), 54.8 (d, C-1), 52.5 (q, Me) ppm. IR (CHCl$_3$): ṽ = 3690, 3600–3500 (br), 3027, 2920, 1708, 1601, 1155, 1070 cm^{-1}. MS (EI): m/z (%) = 513 (0.6) [M – H$_2$O]$^+$, 212 (4), 91 (100), 69 (14). C$_{31}$H$_{33}$NO$_7$ (531.6): calcd. C 70.04, H 2.63, N 6.26; found C 70.02, H 2.56, N 6.28.

(1R,2R,3R,6R,7R,7aR)-1,2-Bis(benzyloxy)-3-[(benzyloxy)methyl]-7-(hydroxymethyl)hexahydro-1H-pyrrolizin-6-ol (21): A 1 M solution of LiAlH$_4$ in THF (1.1 mL, 1.1 mmol) was added to a cooled (0 °C) solution of **20** (115 mg, 0.22 mmol) in dry THF (3 mL) under nitrogen. The mixture was then heated at reflux for 2 h. Then, after cooling at 0 °C, an aqueous saturated solution of Na$_2$SO$_4$ (700 µL) was added dropwise. The suspension was then filtered through Celite® and washed with EtOAc. Concentration under reduced pressure afforded solid **21** pure enough to be used in the next step (104 mg, quantitative yield). An analytically pure sample was obtained by filtration through a short pad of silica gel (eluent: EtOAc then EtOAc/MeOH, 5:1), m.p. 85–87 °C. $[a]_D^{20}$ = +3.58 (c = 1.18, CHCl$_3$). ^{1}H NMR (400 MHz, CDCl$_3$): δ = 7.40–7.25 (m, 15 H, Ar), 4.64–4.56 (m, 6 H, Bn), 4.22 (q, J = 5.6 Hz, 1 H, 6-H), 4.13–4.09 (m, 2 H, 1-H, 2-H), 3.68 (d, J = 6.8 Hz, 2 H, 8-Ha, 8-Hb), 3.53 (d, J = 6.4 Hz, 2 H, 9-Ha, 9-Hb) 3.43–3.37 (m, 2 H, 3-H, 5-Ha) 3.27 (dd, J = 7.0, 3.9 Hz, 1 H, 7a-H), 2.95 (dd, J = 10.8, 5.6 Hz, 1 H, 5-Hb), 2.24 (quint., J = 6.6 Hz, 1 H, 7-H) ppm. ^{13}C NMR (50 MHz, CDCl$_3$): δ = 138.2–137.7 (s, 3 C, Ar), 128.5–127.5 (d, 15 C, Ar), 88.0 (d), 86.1 (d), 76.6 (d), 73.3 (t), 72.4 (t), 72.0 (t), 71.5 (t), 70.4 (d), 70.4 (d), 64.2 (t), 62.4 (t), 54.3 (d) ppm. IR (CDCl$_3$): ṽ = 3412, 3031, 3010, 2866, 1496, 1454, 1363, 1216, 1212, 1211, 1097 cm^{-1}. MS (EI): m/z (%) = 398 (11) [M – C$_7$H$_7$]$^+$, 368 (85), 248 (11), 186 (27), 160 (30), 142 (25), 116 (22), 91 (100), 64

(21). $C_{30}H_{35}NO_5$ (489.6): calcd. C 73.59, H 7.21, N 2.86; found C 73.58, H 7.09, N 3.18.

(1R,2R,3R,6R,7R,7aR)-3,7-Bis(hydroxymethyl)hexahydro-1H-pyrrolizine-1,2,6-triol (7-Homocasuarine, 7): Concentrated HCl (4–5 drops) and Pd (10% on C, 230 mg) were added to a stirred solution of **21** (106 mg, 0.22 mmol) in EtOH (14 mL). The suspension was stirred at room temp. under hydrogen for 4 d, then filtered through Celite® and washed with EtOH. Evaporation under reduced pressure afforded a viscous oil that was transferred to a column of DOWEX 50WX8 and then washed with MeOH (10 mL), H_2O (10 mL) to remove non-amine-containing products and then with 6% NH_4OH (15 mL) to elute 7-homocasuarine (**7**). Evaporation of the solvent afforded 7-homocasuarine as a yellow viscous oil (38.5 mg, 89%). $[a]_D^{20}$ = +30.8 (c = 0.7, MeOH). ^{1}H NMR (400 MHz, D_2O): δ = 4.39 (q, J = 10.0 Hz, 1 H, 6-H), 4.04 (t, J = 7.8 Hz, 1 H, 1-H), 3.71 (t, J = 8.5 Hz, 1 H, 2-H), 3.67 (dd, J = 11.7, 3.5 Hz, 1 H, 8-Ha), 3.55-3.44 (m, 3 H, 8-Hb, 9-Ha, 9-Hb), 3.19 (dd, J = 10.0, 5.2 Hz, 1 H, 5-Ha), 3.02–2.95 (m, 2 H, 3-H, 7a-H), 2.80 (dd, J = 10.0, 5.2 Hz, 1 H, 5-Hb), 2.21 (q, J = 5.9 Hz, 1 H, 7-H) ppm. ^{13}C NMR (50 MHz, D_2O): δ = 80.2 (d, C-1), 77.5 (d, C-2), 74.5 (d, C-6), 70.6, 68.5 (d, C-3, C-7), 62.7 (t, C-8), 61.3 (t, C-9), 60.4 (t, C-5), 52.9 (d, C-7a) ppm. MS (EI): m/z (%) = 188 (100), 170 (10), 159 (13), 142 (14), 128 (83), 116 (26), 68 (38), 55 (24). $C_9H_{17}NO_5$ (219.23): calcd. C 49.31, H 7.82, N 6.39; found C 49.06, H 7.43, N 6.54.

7-Deoxy-6-O-(α-D-glucopyranosyl)-7-(methoxycarbonyl)-5-oxocasuarine (10): Pd (10% on C, 150 mg) was added to a stirred solution of **22** (125 mg, 0.118 mmol) in MeOH/AcOEt (3:1, 12 mL). The suspension was stirred at room temp. under hydrogen for 4 d, then filtered through Celite® and washed with MeOH. Concentration under reduced pressure afforded pure **10** as a waxy solid (48 mg, 96%). $[a]_D^{21}$ = +75.3 (c = 0.15, MeOH). ^{1}H NMR (400 MHz, D_2O): δ = 5.18 (d, J = 3.6 Hz, 1 H, 1-H), 4.97 (d, J = 9.6 Hz, 1 H, 6′-H), 4.08 (dd, J = 6.8, 6.4 Hz, 1 H, 2′-H), 3.88 (dd, J = 7.6, 7.2 Hz, 1 H, 1′-H), 3.80–3.55 (m, 11 H, 5-H, OCH$_3$, 4-H, 3-H, 6-Ha, 6-Hb, 8-Ha′, 7a′-H, 3′-H), 3.41 (dd, J = 9.6, 3.7 Hz, 1 H, 2-H), 3.37 (t, J = 9.6 Hz, 1 H, 8-Hb′), 3.28 (dd, J = 9.4, 8.2 Hz, 1 H, 7′-H) ppm. ^{13}C NMR (50 MHz, D_2O): δ = 171.9, 171.5 (s, C=O), 98.8 (d, C-1), 79.2 (d, C-6′), 78.3, 77.3, 72.2, 72.1, 70.9, 68.6, 61.4, 60.9 (d, 1 C), 59.5, 52.2 (t, 1 C), 52.8 (q, OMe), 52.2 (d, 1 C) ppm. IR (KBr): $\tilde{v}$ = 3420 (OH), 1710 (C=O), 1684 (C=O), 1205, 1143, 1024 cm^{-1}. HRMS (ESI): calcd. for $C_{16}H_{25}NO_{12}Na$ [M + Na]$^+$ 446.1269; found 446.1266. $C_{16}H_{25}NO_{12}$ (423.37): calcd. C 45.39, H 5.95, N 3.31; found C 44,93, H 6.27, N 3,37.

6-O-α-D-Tris(benzyloxy)-5-oxo-6-O-(2,3,4,6-tetra-O-benzyl-α-D-glucopyranosyl)-7-homocasuarine (23): A 2 M solution of LiBH$_4$ in THF (0.42 mL) was added dropwise to a cooled (0 °C) solution of **22** (221 mg, 0.21 mmol) in dry THF (1.5 mL). The reaction mixture was stirred at room temp. overnight, and then, after cooling to 0 °C, H_2O was added dropwise. The mixture was then filtered through Celite®, washed with CHCl$_3$ and concentrated under reduced pressure. The residue was purified by flash column chromatography on silica gel (EtOAc/hexane, 1:4) to afford pure **23** (R_f = 0.33, EtOAc/petroleum ether, 1:3) as a colourless oil (134 mg, 62%). $[a]_D^{23}$ = +44.4 (c = 0.4, CHCl$_3$). ^{1}H NMR (400 MHz, CDCl$_3$): δ = 7.47–7.08 (m, 35 H, Ar), 5.68 (d, J = 3.6 Hz, 1 H, 1-H), 5.09 (d, J = 11.7 Hz, 1 H, Bn), 5.02 (d, J = 10.9 Hz, 1 H, Bn), 4.81–4.75 (m, 3 H, Bn), 4.72 (d, J = 7.6 Hz, 1 H, 6′-H), 4.60–4.40 (m, 9 H, Bn), 4.26 (dd, J = 3.6, 4 Hz, 1 H, 2′-H), 4.21 (m, 1 H, 3′-H), 3.94 (t, J = 9.2 Hz, 1 H, 3-H), 3.87–3.66 (m, 6 H, 5-H, 6-Ha, 7a′-H, 8′-Ha,b, 1′-H), 3.60 (dd, J = 9.2, 3.6 Hz, 1 H, 2-H), 3.45 (d, J = 5.2 Hz, 2 H, 9′-Ha,b), 3.37 (dd, J

= 9.8, 7.6 Hz, 1 H, 6-Hb), 3.26 (dd, J = 10.8, 9.2 Hz, 1 H, 4-H), 3.13 (t, 1 H, OH), 2.44 (m, 1 H, 7′-H) ppm. ^{13}C NMR (50 MHz, CDCl$_3$): δ = 171.4 (s, C=O), 138.4–136.7 (s, 7 C, Ar), 128.7–127.3 (d, 35 C, Ar), 94.9 (d, C-1), 87.7 (d, C-1′), 86.1 (d, C-2′), 81.4 (d, C-3), 78.2 (d, C-2), 77.4 (d, C-4), 75.6 (d, C-6′), 75.6, 74.9, 73.3, 73.0, 71.9, 71.7, 71.2 (t, Bn), 70.5 (d, C-5), 68.8 (t, C-6), 68.7 (t, C-9′), 60.5 (d, C-7a′), 58.4 (t, C-8′), 58.4 (d, C-3′), 50.4 (d, C-7′) ppm. IR (CDCl$_3$): $\tilde{v}$ = 3463 (OH), 3032, 2925, 2870, 1703 (C=O), 1454, 1078 cm^{-1}. HRMS (ESI): calcd. for $C_{64}H_{67}NO_{11}Na$ [M + Na]$^+$ 1048.4606; found 1048.4602. $C_{64}H_{67}NO_{11}$ (1026.22): calcd. C 74.90, H 6.58, N 1.36; found C 74.65, H 6.69, N 1.53.

6-O-(α-D-Glucopyranosyl)-5-oxo-7-homocasuarine (9): Pd (10% on C, 230 mg) was added to a stirred solution of **23** (134 mg, 0.13 mmol) in MeOH (13 mL). The suspension was stirred at room temp. under hydrogen for 24 h, then filtered through Celite® and washed with MeOH. Concentration under reduced pressure afforded pure **9** as a waxy solid (37 mg, 72%). $[a]_D^{26}$ = +59.3 (c = 0.75, MeOH). ^{1}H NMR (400 MHz, D_2O): δ = 5.18 (d, J = 4 Hz, 1 H, 1-H), 4.57 (d, J = 9.6 Hz, 1 H, 6′-H), 4.09 (dd, J = 6.8, 6.4 Hz, 1 H, 2′-H), 3.82–3.58 (m, 10 H, 1′-H, 3′-H, 8′-Ha,b, 9′-Ha, 3-H, 4-H, 5-H, 6-Ha,b), 3.50–3.45 (m, 2 H, 7a′-H, 2-H), 3.34 (dd, J = 9.6, 9.2 Hz, 1 H, 9′-Hb), 2.50–2.44 (m, 1 H, 7′-H) ppm. ^{13}C NMR (50 MHz, D_2O): δ = 173.9 (s, C=O), 98.6 (d, C-1), 79.9 (d, C-6′), 77.9, 77.3, 72.7, 72.5, 71.2, 69.3, 62.1, 61.2 (d, 1 C), 60.3, 59.8, 59.7 (t, 1 C), 50.0 (d, 1 C) ppm. IR (KBr): $\tilde{v}$ = 3378 (OH), 1689 (C=O) cm^{-1}. HRMS (ESI): calcd. for $C_{65}H_{67}NO_{12}Na$ [M + Na]$^+$ 418.1320; found 418.1312. $C_{15}H_{25}NO_{11}$ (395.36): calcd. C 45.57, H 6.37, N 3.54; found C 45.25, H 6.28, N 3.04.

6-O-(α-D-Glucopyranosyl)-5-oxocasuarine (3): Pd (10% on C, 180 mg) was added to a stirred solution of **24** (159 mg, 0.144 mmol) in MeOH/EtOAc (7:1, 12 mL). The suspension was stirred at room temp. under hydrogen for 24 h, then filtered through Celite® and washed with MeOH. Concentration under reduced pressure afforded pure **3** (42 mg, 0.110 mmol, 77% yield) as a hygroscopic pale-yellow oil. $[a]_D^{24}$ = +37.6 (c = 0.28, MeOH). ^{1}H NMR (400 MHz, D_2O): δ = 5.14 (d, J = 3.7 Hz, 1 H, 1-H), 4.53 (d, J = 8.4 Hz, 1 H, 6′-H), 4.27 (dd, J = 8.4, 7.0 Hz, 1 H, 7′-H), 4.08 (t, J = 6.1 Hz, 1 H, 2′-H), 3.90 (t, J = 6.8 Hz, 1 H, 1′-H), 3.78–3.58 (m, 7 H, 3-H, 3′-H, 5-H, 6-Ha,b, 8′-Ha,b), 3.50–3.44 (m, 2 H, 2-H, 7a′-H), 3.34 (t, J = 9.5 Hz, 1 H, 4-H) ppm. ^{13}C NMR (50 MHz, D_2O): δ = 170.8 (s, C-5′), 98.0 (d, C-1), 81.6 (d, C-6′), 78.2 (d, 1 C), 77.3 (d, C-7′), 76.9 (d, C-2′), 71.9 (d), 71.5 (d), 70.5 (d), 68.6 (d), 64.7 (d), 60.7 (d), 59.6 (t), 58.8 (t) ppm. MS (ESI): m/z (%) = 404 (100) [M + Na]$^+$. $C_{14}H_{23}NO_{11}$ (381.33): calcd. C 44.10, H 6.08, N 3.67; found C 43.96, H 6.19, N 3.42.

Enzymatic Assays: The experiments were performed essentially as follows: 0.01–0.5 unit/mL of enzyme (1 unit = 1 mol of glycoside hydrolysed/min), preincubated at 20 °C with the inhibitor for 5 min, and increasing concentrations of an aqueous solution of the appropriate p-nitrophenyl glycoside substrates (buffered at the optimum pH of the enzyme) were incubated at 37 °C for 20 min. The reactions were stopped by the addition of 0.3 M sodium borate buffer (100 µL, pH = 9.8). The p-nitrophenolate formed was quantified at 405 nm, and IC$_{50}$ values were calculated. Double-reciprocal (Lineweaver–Burk) plots were used to determine the inhibition characteristics and the K_i values for each compound.

Molecular Modelling: The ligand structures (Table 1) were constructed by using Maestro v8.5.[29] All the molecules were subjected to conformational search and clusterization with Macromodel 9.6[30] in order to sample the most accessible conformations of both the aglyconic and glucose moieties. The bridgehead nitrogen atoms were treated as ionized to better simulate the physiological condi-

tions, except for the lactam intermediates. All the docking calculations were performed by using Glide 5.0.[31] The crystal structure of glucoamylase-471 from *Aspergillus awamori* complexed with 1-deoxynojirimycin (PDB ID: 1DOG)[26a] was downloaded from the PDB and prepared according to the recommended Protein Preparation module in Maestro 8.5 by using default input parameters (no scaling factors for the van der Waals radii of non-polar protein atoms, 0.8 scaling factor for non-polar ligand atoms). This procedure was used to remove water molecules (except for molecule Wat501, which is considered as part of the target structure), to assign missing hydrogen atoms, to optimize hydrogen-bonding interactions and to reduce structural problems. The grids were prepared with the centre of the site defined by the centre of the complexed ligand. All the relevant conformations for the ligands in Table 1 were docked in the binding site by using the SP scoring function to score the ligand poses. The docking calculations were performed in the presence of Wat501. After completion of each docking run, one pose per ligand conformation was saved. Finally, for each ligand, the poses (conformations) with the lowest (best) value of either the model energy score (Emodel)[32] or the Glide score were chosen.

Supporting Information (see footnote on the first page of this article): Syntheses of **5**, **8**, **8a**, **10a**, **16a**, **19**, **19a,b**, **22** and **22a**, copies of the NMR spectra of **3–11**, **8a**, **10a**, **16–23**, **16a**, **19a,b** and **22a**, HR-ESI-TOF-MS spectra and the docking results for **2**.

Acknowledgments

[1] a) P. Compain, O. R. Martin (Eds.), *Iminosugars: from synthesis to therapeutic applications*, Wiley-VCH, Weinheim, **2007**; b) N. Asano, *Curr. Top. Med. Chem.* **2003**, *3*, 471–484, and references cited therein.

[2] a) N. Asano, *Glycobiology* **2003**, *13*, 93R–104R, and references cited therein; b) N. Asano, R. J. Nash, R. J. Molyneux, G. W. J. Fleet, *Tetrahedron: Asymmetry* **2000**, *11*, 1645–1680; c) I. Robina, A. J. Moreno-Vargas, A. T. Carmona, *Curr. Drug Metab.* **2004**, *5*, 329–361.

[3] a) A. Vasella, T. D. Heightman, *Angew. Chem. Int. Ed.* **1999**, *38*, 750–770; b) T. M. Gloster, P. Meloncelli, R. V. Stick, D. Zechel, A. Vasella, G. J. Davies, *J. Am. Chem. Soc.* **2007**, *129*, 2345–2354; c) M. E. C. Caines, S. M. Hancock, C. A. Tarling, T. M. Wrodnigg, R. V. Stick, A. E. Stütz, A. Vasella, S. J. Withers, N. C. J. Strynadka, *Angew. Chem. Int. Ed.* **2007**, *46*, 4474–4476; d) T. M. Gloster, G. J. Davies, *Org. Biomol. Chem.* **2010**, *8*, 305–320.

[4] a) A. A. Watson, R. J. Nash, E. L. Evinson, PCT Int. Appl. WO2004064715, **2004**; b) R. J. Nash, A. A. Watson, E. L. Evinson, H. St. P. Parry, PCT Int. Appl. WO2005070418, **2005**.

[5] a) A. A. Watson, G. W. J. Fleet, N. Asano, R. J. Molyneux, R. J. Nash, *Phytochemistry* **2001**, *56*, 265–295; b) T. Yamashita, K. Yasuda, H. Kizu, Y. Kameda, A. A. Watson, R. J. Nash, G. W. J. Fleet, N. Asano, *J. Nat. Prod.* **2002**, *65*, 1875–1881; c) N. Asano, K. Ikeda, M. Kasahara, Y. Arai, H. Kizu, *J. Nat.*

[6] a) H. Paulsen, I. Sangster, K. Heyns, *Chem. Ber.* **1967**, *100*, 802–815; b) M. Yagi, T. Kouno, Y. Aoyagi, H. Murai, *Nippon Nogei Kagaku Kaishi* **1976**, *50*, 571–572.

[7] a) I. Pastuszak, R. J. Molyneux, L. F. James, A. D. Elbein, *Biochemistry* **1990**, *29*, 1886–1891; b) F. Cardona, A. Goti, A. Brandi, *Eur. J. Org. Chem.* **2007**, 1551–1565; c) F. Cardona, G. Moreno, F. Guarna, P. Vogel, C. Schuetz, P. Merino, A. Goti, *J. Org. Chem.* **2005**, *70*, 6552–6555.

[8] B. Macchi, A. Minutolo, S. Grelli, A. Mastino, F. Cardona, F. M. Cordero, A. Brandi, *Glycobiology* **2010**, *20*, 500–506.

[9] a) M. J. Humphries, K. Matsumoto, L. S. White, K. Olden, *Cancer Res.* **1986**, *46*, 5215–5222; b) G. K. Ostrander, N. K. Scribner, L. R. Rohrschneider, *Cancer Res.* **1988**, *48*, 1091–1094; c) S. P. Sunkara, P. S. Liu, Eur. Pat. Appl. EP 373663, **1990**; d) M. A. Spearman, J. M. Ballon, J. M. Gerrard, A. H. Greenberg, J. A. Wright, *Cancer Lett.* **1991**, *60*, 185–191; e) R. Pili, J. Chang, R. A. Partis, R. A. Mueller, F. J. Chrest, A. Passaniti, *Cancer Res.* **1995**, *55*, 2920–2926; f) C. S. Yee, E. D. Schwab, J. E. Lehr, M. Quigley, K. J. Pienta, *Anticancer Res.* **1997**, *17*, 3659–3663.

[10] a) R. J. Nash, P. I. Thomas, R. D. Waigh, G. W. J. Fleet, M. R. Wormald, P. M. Q. Lilley, D. J. Watkin, *Tetrahedron Lett.* **1994**, *35*, 7849–7852; b) M. R. Wormald, R. J. Nash, A. A. Watson, B. K. Bhadoria, R. Langford, A. Sims, G. W. J. Fleet, *Carbohydr. Lett.* **1996**, *2*, 169–174.

[11] F. Cardona, C. Parmeggiani, E. Faggi, C. Bonaccini, P. Gratteri, L. Sim, T. M. Gloster, S. Roberts, G. J. Davies, D. R. Rose, A. Goti, *Chem. Eur. J.* **2009**, *15*, 1627–1636.

[12] K. Hiromi, Z. I. Hamauzu, K. Takahashi, S. Ono, *J. Biochem.* **1966**, *59*, 411–418.

[13] B. C. Saha, J. G. Zeikus, *Starch/Staerke* **1989**, *41*, 57–64.

[14] J. Sauer, B. W. Sigurskjold, U. Christensen, T. P. Frandsen, E. Mirgrorodskaya, M. Harrison, P. Roepstorff, B. Svensson, *Biochim. Biophys. Acta* **2000**, *1543*, 275–293.

[15] A. Brandi, F. Cardona, S. Cicchi, F. M. Cordero, A. Goti, *Chem. Eur. J.* **2009**, *15*, 7808–7821.

[16] a) F. Cardona, E. Faggi, F. Liguori, M. Cacciarini, A. Goti, *Tetrahedron Lett.* **2003**, *44*, 2315–2318; b) A. T. Carmona, R. H. Wightman, I. Robina, P. Vogel, *Helv. Chim. Acta* **2003**, *86*, 3066–3073; c) S. Desvergnes, S. Py, Y. Vallée, *J. Org. Chem.* **2005**, *70*, 1459–1462; d) J. Revuelta, S. Cicchi, A. Goti, A. Brandi, *Synthesis* **2007**, 485–504; e) E.-L. Tsou, Y.-T. Yeh, P.-H. Liang, W.-C. Cheng, *Tetrahedron* **2009**, *65*, 93–100; f) X.-G. Hu, Y.-M. Jia, J. Xiang, C.-Y. Yu, *Synlett* **2010**, 982–986.

[17] For other total syntheses of 7-deoxycasuarine, see: a) J.-B. Behr, A. Erard, G. A. Guillerm, *Eur. J. Org. Chem.* **2002**, 1256–1262; b) ref.[16b]; c) J.-B. Behr, A. Gainvors-Claisse, A. Belrabe, *Nat. Prod. Res.* **2006**, *20*, 1308–1314.

[18] For other total syntheses of hyacinthacine A₂, see: a) L. Rambaud, P. Compain, O. R. Martin, *Tetrahedron: Asymmetry* **2001**, *12*, 1807–1809; b) I. Izquierdo, M. T. Plaza, F. Franco, *Tetrahedron: Asymmetry* **2003**, *14*, 3933–3935; c) ref. 16c; d) P. Dewi-Wuelfing, S. Blechert, *Eur. J. Org. Chem.* **2006**, 1852–1856; e) J. Calveras, J. Casas, T. Parella, J. Joglar, P. Clapes, *Adv. Synth. Catal.* **2007**, *349*, 1661–1666; f) C. Ribes, E. Falomir, M. Carda, J. A. Marco, *Tetrahedron* **2009**, *65*, 6965–6971; g) I. Delso, T. Tejero, A. Goti, P. Merino, *Tetrahedron* **2010**, *66*, 1220–1227.

[19] F. Cardona, A. Goti, C. Parmeggiani, P. Parenti, M. Forcella, P. Fusi, L. Cipolla, S. M. Roberts, G. J. Davis, T. M. Gloster, *Chem. Commun.* **2010**, *46*, 2629–2631.

[20] S. Nukui, M. Sodeoka, H. Sasai, M. Shibasaki, *J. Org. Chem.* **1995**, *60*, 398–404.

[21] For other total syntheses of casuarine, see: a) S. E. Denmark, A. R. Hurd, *Org. Lett.* **1999**, *1*, 1311–1314; b) S. Denmark, A. R. Hurd, *J. Org. Chem.* **2000**, *65*, 2875–2886; c) I. Izquierdo,

M. T. Plaza, J. A. Tamayo, *Tetrahedron* **2005**, *61*, 6527–6533; d) J. Van Ameijde, G. Horne, M. R. Wormald, R. A. Dwek, R. J. Nash, P. W. Jones, E. L. Evinson, G. W. J. Fleet, *Tetrahedron: Asymmetry* **2006**, *17*, 2702–2712; e) T. Ritthiwigrom, A. C. Willis, S. G. Pyne, *J. Org. Chem.* **2010**, *75*, 815–824.

[22] C. Parmeggiani, D. Martella, F. Cardona, A. Goti, *J. Nat. Prod.* **2009**, *72*, 2058–2060.

[23] A. Brandi, S. Cicchi, F. M. Cordero, R. Frignoli, A. Goti, S. Picasso, P. Vogel, *J. Org. Chem.* **1995**, *60*, 6806–6812.

[24] A. Kato, E. Kano, I. Adachi, R. J. Molyneux, A. A. Watson, R. J. Nash, G. W. J. Fleet, M. R. Wormald, H. Kizu, K. Ikeda, N. Asano, *Tetrahedron: Asymmetry* **2003**, *14*, 325–331.

[25] N. Asano, H. Kuroi, K. Ikeda, H. Kizu, Y. Kameda, A. Kato, I. Adachi, A. A. Watson, R. J. Nash, G. W. J. Fleet, *Tetrahedron: Asymmetry* **2000**, *11*, 1–8.

[26] a) E. M. S. Harris, A. E. Aleshin, L. M. Firsov, R. B. Honzatko, *Biochemistry* **1993**, *32*, 1618–1626; b) A. E. Aleshin, B. Stoffer, L. M. Firsov, B. Svensson, R. B. Honzatko, *Biochemistry* **1996**, *35*, 8319–8328; c) A. Aleshin, A. Golubev, L. M. Firsov, R. B. Honzatko, *J. Biol. Chem.* **1992**, *267*, 19291–19298; d)

A. E. Aleshin, C. Hoffman, L. M. Firsov, R. B. Honzatko, *J. Mol. Biol.* **1994**, *238*, 575–591; e) A. E. Aleshin, L. M. Firsov, R. B. Honzatko, *J. Biol. Chem.* **1994**, *269*, 15631–15639.

[27] A. M. Davis, S. J. Teague, *Angew. Chem. Int. Ed.* **1999**, *38*, 736–749.

[28] T. Weimar, B. Stoffer, B. Svensson, B. M. Pinto, *Biochemistry* **2000**, *39*, 300–306.

[29] *Maestro*, v. 8.5, Schrödinger, L.L.C., New York, **2008**, available at http://www.schrodinger.com.

[30] *Macromodel*, v. 9.6, Schrödinger, L.L.C., New York, **2008**, available at http://www.schrodinger.com.

[31] *Glide*, v. 5.0, Schrödinger, L.L.C., New York, **2008**, available at http://www.schrodinger.com.

[32] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *J. Med. Chem.* **2004**, *47*, 1739–1749.

Ringraziamenti

Ringraziamenti

Gratuitamente avete ricevuto, gratuitamente date (Mt10,8)

*Alla fine di questo percorso di studi il pensiero più grande va alla mia Famiglia, alla quale dedico di cuore questo lavoro. E' grazie ai miei **genitori**, alla mia **nonna Nella** e a **Barbara** che sono arrivato fino a qui. Grazie ai loro incoraggiamenti, ai loro consigli, alla loro pazienza e un po' … anche grazie agli squisiti mangiarini della mia nonna Nella.*

Grazie alla Professoressa Paola Gratteri per aver avuto fiducia in me, per il prezioso aiuto e per il sostegno che mi ha sempre dato. Un sentito grazie va poi a tutte le persone con cui ho condiviso questi anni di dottorato nel gruppo di Molecular Modeling.

Grazie al Professor Piero Procacci per la sua disponibilità e per avermi permesso di svolgere una parte del dottorato in collaborazione con il suo gruppo di ricerca.

Grazie a Simone per l'aiuto che mi ha dato e per la pazienza che ha avuto.

Un grazie, infine, a tutte quelle persone che ho conosciuto in questi anni di dottorato e che mettono tanta passione in quello che fanno.

Matteo.