UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI SCIENZE MATEMATICHE,
FISICHE E NATURALI

MODELLING INTERMOLECULAR FORCES IN BIOMOLECULES: FROM PROTEIN-PROTEIN INTERACTIONS TO HALOGEN BONDS

Tutor: Prof. Maurizio SIRONI
Co-tutor: Dott. Stefano PIERACCINI

Tesi di Dottorato di Ricerca di**:**
Stefano RENDINE
Matr. R08324

A.A. 2010-2011

Contents

PART II
DESIGN OF PEPTIDES INHIBITING FTSZ POLYMERIZATION

PART III
COMPUTATIONAL DESCRIPTION OF
HALOGEN BONDING

Thesis Overview

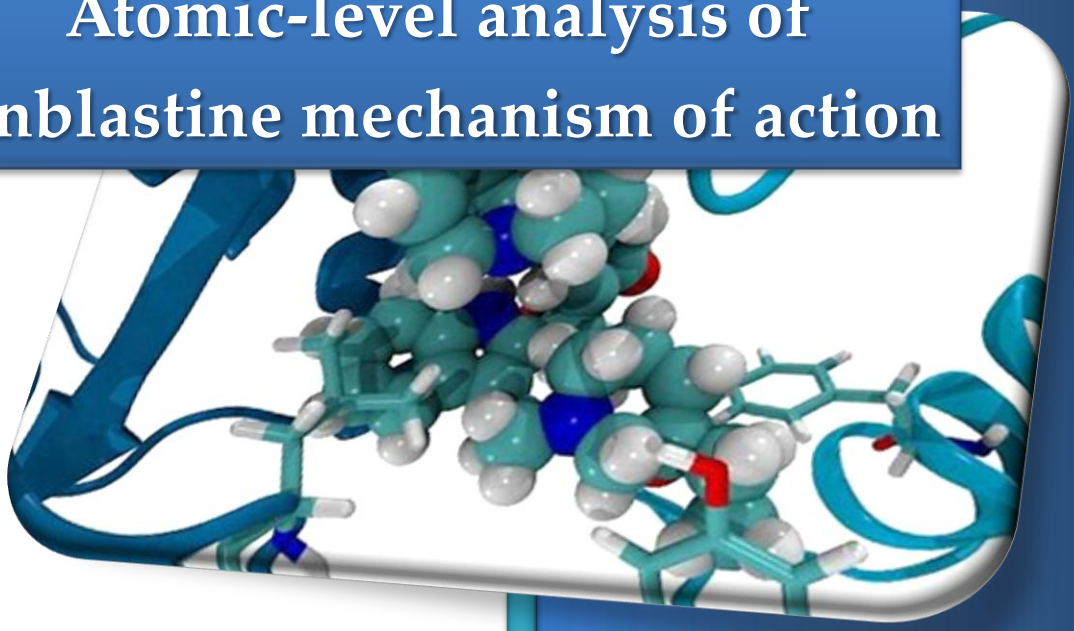
The accurate description and evaluation of the intermolecular interactions has a great importance in the molecular modelling of biological systems.

Protein-protein interactions, in particular, being involved in virtually every cellular process, are nowadays the object of thorough studies aimed at the understanding and modulation of the underlying mechanism. In the Part I of the present work, it will be shown how, through a computational approach, it is possible to get an in-depth analysis of the network of interactions occurring at the interface between tubulin subunits and how vinblastine, a commonly used anticancer agents, is able to interfere with the correct protein association, so having a therapeutic effect. In the Part II, the self-association of the bacterial protein FtsZ will be studied, showing that the detailed description of the protein-protein interactions can provide key information for the *de-novo* design of inhibiting molecules.

Finally, in the Part II, the attention will be focused on the computational study of halogen bonding, which is found to have a great relevance in the recognition process between biological macromolecules and halogenated agents. In particular, it will be shown that a specific approach is mandatory for its correct description in the framework of the classical force-fields.

Part I

Atomic-level analysis of vinblastine mechanism of action



Chapter 1

Protein-protein interactions in computational chemistry

1.1 Protein-protein interactions

Interactions between proteins are essential to a great number of biological processes, ranging from signal transduction to cell duplication. Disease-relevant protein-protein interactions occurring at defined cellular sites represent therefore a viable target in drug design which has lately gained an increasing interest as a therapeutic strategy and is presently one of the main goals of medicinal chemistry, as well as an active research field [1] [2] [3] .

The leading force in protein-protein non-covalent association is hydrophobicity [4] [5] [6]. In fact, hydrophobic interactions occur between non-polar side chains through van der Waals contacts, with a free energy gain due to their movement from the polar aqueous environment to the large nonpolar surface area that usually characterizes protein-protein interfaces. As a result, hydrophobic side chains from facing protein subunit form a tight packing that

highly stabilizes the protein-protein complex [7]. Besides hydrophobic contacts, also electrostatic forces play a key role in the protein-protein association [8], acting as a steering force driving the complex formation according to electrostatic complementarity between the opposite subunits [9].

Nonetheless, the design of molecules inhibiting protein-protein interactions is a difficult task to achieve [10], mainly due to the fact that protein-protein interfaces are large (1200-2000 $Å^2$) [11] and mostly flat, thus not showing cavities where ligands could bind.

It has been pointed out that, in general, only a minor part of the residues making up the interface actually contribute to the protein-protein binding energy [12] [13]. These critical amino acids, called hot-spots [14] [15], are thus to be targeted in order to interfere with the protein assembly process, so it might not be necessary for an inhibitor to cover the whole interface. Hot-spots are usually grouped into small clusters and are found at the centre of a protein-protein interface, excluded by the solvent. Hot-spots are usually surrounded by a set of residues, called O-ring [15] [16], which are not energetically important for the protein-protein association, but have a key role in ensuring a complete exclusion from water.

The prediction of the hot-spots location at a protein-protein interface is of utmost importance in protein engineering and in the structure-based drug design [17] [18] [19]. Experimentally, the identification of the hot-spots at a protein-protein interface can be performed through alanine scanning mutagenesis analysis [20]. Through alanine substitution, it is in fact possible to evaluate the role of the native side-chain functional groups and their energetic contribution to protein association, by estimating the difference in the binding energy of the complex before and after the mutation [21]. Glycine could replace

alanine, in principle, to nullify the side chains, but its use is limited by the conformational flexibility it could introduce into the protein backbone [22].

1.2 Computational evaluation of the binding free energy

From an energetic point of view, residues are usually considered hot-spots if the change in the binding free energy (ΔΔG) upon mutation into alanine is higher than 2 kcal/mol [23]. Nonetheless, the experimental estimation of the binding free energy may be a difficult task to achieve. Moreover, a mutagenesis study is time consuming and has a high cost. A computational approach in the calculation of the binding free energy between proteins, aimed at an *in silico* evaluation of the per-residue ΔΔG contributions, can be valuable in the understanding of the energetics of protein-protein assemblies [24] [25].

Different computational methods have been developed to estimate the binding free energy. Free energy perturbation [26] and thermodynamic integration [27] allow to obtain accurate estimates, but are computationally time-consuming and are not suitable for the study of large protein complexes. Other approximate methodologies have thus been proposed, such as Monte Carlo [28] and linear interactions energy methods [29].

In the present study, the MM-GBSA approach [30] was used for the evaluation of the binding free energy, which combines the molecular mechanical gas energies, the Generalized Born continuum solvation approach and the solvent accessibility method to treat the solvent effect to evaluate the

total free energies. The calculation of the binding free energy is performed *a posteriori* on snapshots extracted from a molecular dynamics simulation of the protein complex under examination.

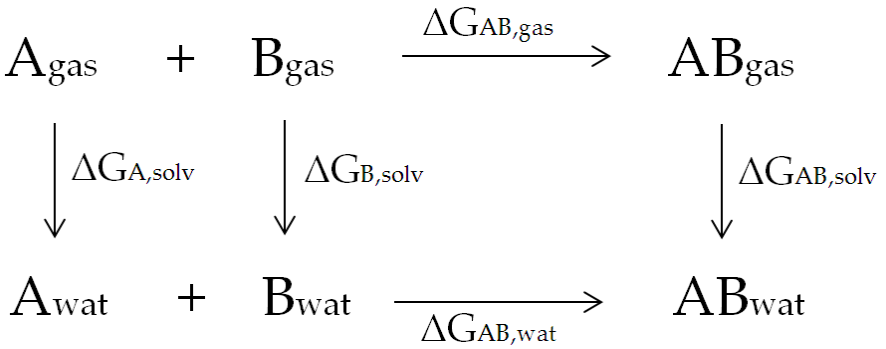


Figure 1 Thermodynamic cycle used in the MM-GBSA approach.

The MM-GBSA approach uses a thermodynamic cycle (**Figure 1**) to calculate the binding free energy ($\Delta G_{AB,wat}$) between two protein subunits A and B in solution. Indicating with ΔG_{solv} the free energy of solvation as the difference between the free energy in solution (G_{wat}) and the free energy in vacuum (G_{gas})

$$\Delta G_{solv} = G_{wat} - G_{gas}$$

the free energy in solvent can be written as

$$G_{wat} = G_{gas} + \Delta G_{solv}$$

As a result, the binding free energy for the complex AB in solution is

$$\Delta G_{\mathrm{AB,wat}} = \Delta \mathrm{G}_{\mathrm{AB,gas}} + \Delta\Delta \mathrm{G}_{\mathrm{solv}}$$

which corresponds to

$$\Delta \mathrm{G}_{\mathrm{AB,wat}} = \Delta \mathrm{G}_{\mathrm{AB,gas}} + \Delta \mathrm{G}_{\mathrm{AB,solv}} - \Delta \mathrm{G}_{\mathrm{A,solv}} - \Delta \mathrm{G}_{\mathrm{B,solv}}$$

so decomposing the binding free energy in solvent into a vacuum term ($\Delta G_{AB,gas}$) and a solvation term ($\Delta\Delta G_{solv}$).

Dealing with snapshots extracted from a molecular dynamics trajectory, the equation above can be written as

$$\Delta \mathrm{G}_{\mathrm{AB,wat}} = \langle \Delta \mathrm{E}_{MM} \rangle + \Delta\Delta \mathrm{G}_{\mathrm{solv}} - \mathrm{T}\Delta \mathrm{S}$$

where $\langle \Delta \mathrm{E}_{MM} \rangle$ is the average molecular mechanics energy calculated during the simulation as the sum of intramolecular and intermolecular terms

$$\Delta \mathrm{E}_{\mathrm{MM}} = \Delta \mathrm{E}_{\mathrm{bonds}} + \Delta \mathrm{E}_{\mathrm{anlges}} + \Delta \mathrm{E}_{\mathrm{torsions}} + \Delta \mathrm{E}_{\mathrm{elec}} + \Delta \mathrm{E}_{\mathrm{vdw}}$$

whereas $\mathrm{T}\Delta \mathrm{S}$ is the entropic term.

While the molecular mechanics energy term can be easily obtained from the results of a molecular dynamics simulation, the entropic term is often

difficult to achieve. The entropic term can thus be approximated to ΔS_{vib}, i.e. the contribution due to the internal vibration, whose calculation is nonetheless usually time-consuming and can be performed only on part of the protein-protein complex (see paragraph 3.2). In other cases, such as in the computational alanine scanning analyses, the entropic term can be neglected, as explained in the next paragraph.

The solvation term $\Delta\Delta G_{solv}$ is calculated *a posteriori* on the snapshots extracted from the molecular dynamics trajectory using an implicit solvent and is decomposed into an electrostatic and a non-electrostatic term.

The electrostatic term takes into account the contribution to the free energy due to the creation of a charge q_i in a spherical cavity of radius a_i in a solvent of dielectric constant ε. For N charges, located at an r_{ij} distance, the term can be calculated through the Generalized-Born equation:

$$\Delta G_{GB} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon}\right)\sum_{i,j\ (i\neq j)}^{N}\frac{q_i q_j}{f(r_{ij}, a_{ij})}$$

with $f(r_{ij}, a_{ij}) = \sqrt{r_{ij}^2 + a_{ij}^2 e^{-D}}$

$$a_{ij} = \sqrt{a_i a_j}$$

$$D = \frac{r_{ij}^2}{(2a_{ij})^2}$$

The non-electrostatic term is calculated according to the Solvent Accessible Surface Area approach and is obtained through

$$\Delta G_{SA} = \gamma A + b$$

where γ and B are constants and A is the total solvent accessible surface.

1.3 Computational alanine scanning

The calculation of the binding free energy allows to perform a computational alanine scanning of the protein-protein interface, in order to highlight the hot-spots, i.e. those residues which are particularly important for the subunits association [31] . The calculation is performed by evaluating the binding free energy between the protein subunits making up the complex (ΔG_{wild}) and the binding energy upon mutation of each interfacial amino acid into an alanine ($\Delta G_{mut,X}$), so obtaining, for each residue, a $\Delta\Delta G$ value defined as

$$\Delta\Delta G_X = \Delta G_{mut,X} - \Delta G_{wild}$$

In this way, it is possible to evaluate the contribution of each amino acid to the total binding free energy, highlighting the hot-spots for which a $\Delta\Delta G$ higher than 2 kcal/mol is found [32] [33].

The relative contribution of the change in conformational entropy to the $\Delta\Delta G$ is considered to be negligible for the mutations studied, since it is supposed to cancel up when calculating it in the native and in the mutated complex. [34]

In the present work, the so-called single-trajectory approach was used when performing the computational alanine scanning. According to this procedure, a single molecular dynamics simulation is performed on the protein complex, using explicit or implicit solvent. A set of snapshots is then extracted from the trajectory for the ΔG evaluation. As to the monomers, it is not necessary to perform further molecular dynamics simulations, but snapshots are obtained from the same complex trajectory, by removing the partner subunit from the dimer structures taken into account. Such approach was proven to provide more reliable results with respect to other approaches where further molecular dynamics simulations are performed on the single monomers [35].

In the Part I, the computational alanine scanning of the interface between tubulin units allowed an in-depth analysis of vinblastine antimitotic action. The same approach allowed, in the Part II, to highlight the main interactions between protein FtsZ monomers.

Chapter 2

Tubulin and microtubules

2.1 Microtubules structure and dynamics

Microtubules [36] are hollow cylindrical structures which play an essential role in a great number of biological processes in eukaryotic cells. In fact, they act as a scaffold to determine the cell shape, allow the organelles transport through the cytoplasm and are responsible for the separation of the chromosomes during mitosis. Microtubules are made up of α-β tubulin heterodimers which assemble in a head-to-tail way to form protofilaments which later on associate laterally to give rise to cylindrical structures (**Figure 2**) [37] [38].

Protofilament elongation, from the so-called minus-end to the plus-end, is closely related to the binding and hydrolysis of guanosine 5'-triphosphate

(GTP) by tubulin [39]. Both α and β tubulin bind GTP, although GTP bound to α tubulin, at the interface between two monomers in the same dimer, is nonexchangeable, whereas the nucleotide bound to β tubulin, at the interface between two adjoining dimers, is exchangeable.

The association between two tubulin dimers leads, after a short lag, to the hydrolysis of the β tubulin-bound GTP, which is buried at dimer-dimer interface and becomes nonexchangeable too. As a result, the microtubule body is made up of GDP-bound tubulin units, whereas, at the growing end, a layer of GTP-bound tubulins (the so-called GTP-cap) is present. A polymer made up of just GDP bound is unstable and would lead to the formation of curved protofilaments which eventually depolymerize [40]. The GTP-cap has a great importance in stabilizing a straight protofilament, thus promoting a regular microtubule growth.

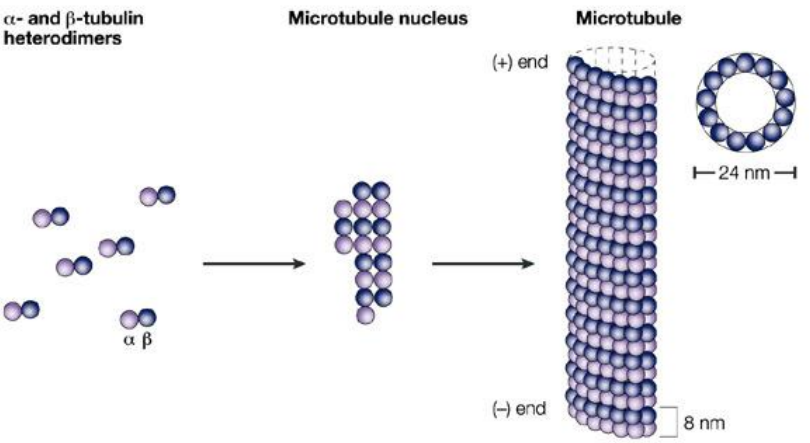


Figure 2 Microtubules polymerization occurs through α-β tubulin head-to-tail association.

Tubulin polymerization is characterized by a complex dynamics, with microtubules elongation or disruption taking place according to the cell needs. The alternation of microtubules growing and shrinking phases is called dynamic instability (**Figure 3**) [41]. According to such process, tubulin polymerization occurs through periodic slow growing phases as far as the GTP cap is present at the plus-end, followed by quick disruption stages (the so-called catastrophes) when the GTP-cap is lost. The recovering of a growing phase is called rescue and takes place when the GTP-cap forms again at the plus-end.

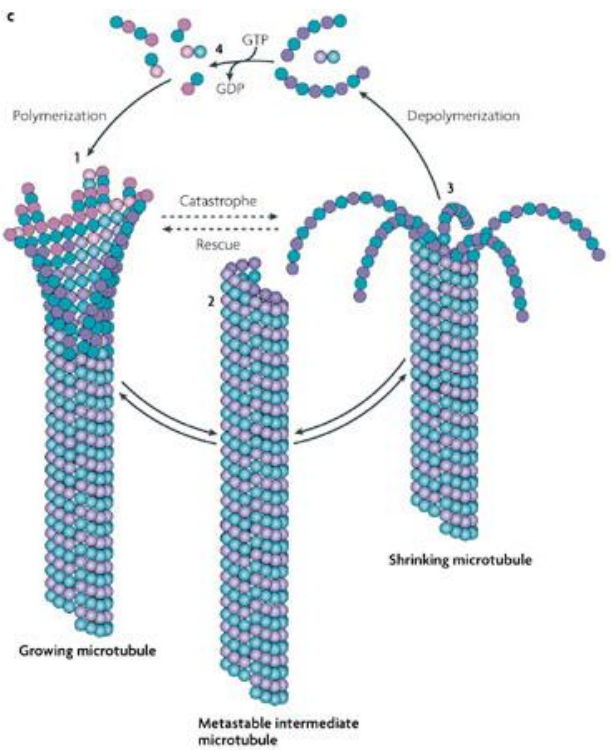


Figure 3 Microtubules dynamics is characterized by alternative growing and shrinking phases, depending on the presence of a GTP-cap at the plus-end. (adapted from [42])

Protofilament polymerization is made possible by a delay between the head-to-tail association and the GTP hydrolysis. In this way, during the growing phase, a GTP molecule can always be present at the plus-end of the protofilaments, thus stabilizing the GDP-rich body and allowing the formation of straight microtubules. On the other hand, when the GTP cap is lost, protofilaments can't grow in a straight conformation any longer and quickly disassemble.

2.2 Microtubules biological role

The high microtubules dynamicity is of fundamental importance for their biological role. In fact, being able to extend throughout the cytoplasm, they're not only allowed to transport organelles and vesicles essential for a cell life, but, during mitosis, can also search and capture chromosomes to make the cell division possible (**Figure 4**) [43] [44]. Microtubules origin from cellular structures called centrosomes, which are composed of two units (the so-called centrioles) made up of tubulin filaments. In the early stage of mitosis (the prophase), the two centrioles of a single centrosome separate and each of them becomes the future centrosome of the daughter cell. A new network of microtubules (the so-called mitotic spindle) can thus grow from the newly generated organelles, extending throughout the cytoplasm. During the

metaphase, the centrosomes migrate to opposite sides of the cell, with the microtubules being able to scan the whole cytoplasm and capture the chromosomes, aligning them in the middle region of the cell. Once the chromosomes are all aligned, the migration of the chromatides toward the opposite pole of the cell takes place. Finally, during the telophase, the division into two daughter cells occurs, once the chromosomes are completely pulled to opposite parts of the mother cell.

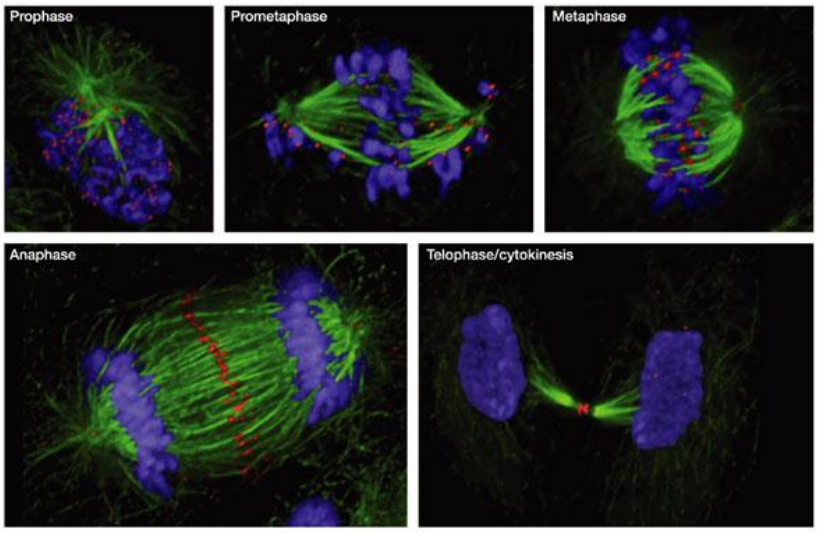


Figure 4 During mitosis, microtubules are responsible for the separation of the genetic material between the daughter cells. Being able to grow throughout the cytoplasm, microtubules can capture the chromosomes and attract them towards opposite poles in the mother cell. [45]

The role of microtubules dynamics during mitosis is thus straightforward. Microtubules plus-end can grow freely in the cytoplasm (reaching a length of typically 5-10μm [46]) and allow the capture of the chromosomes [47], whereas the minus-end is anchored to the centrosomes.

Once a chromosome is bound to one microtubule, it then connects to another one coming from the opposite pole. The strain generated by the microtubules coming from the two poles forces the chromosomes to align along the cell equatorial plane. At that stage, microtubules depolymerization pulls the chromatides to opposite sides of the cell, so allowing the anaphase to take place.

2.3 Microtubules as anticancer target

It is thus clear that microtubules have a fundamental role in the division of the genetic material between the daughter cells. Therefore, tubulin and microtubules can be seen as important targets in the development of anticancer agents [48] [46] [49]. In fact, by developing molecules able to target tubulin, one could interfere with the regular microtubules dynamics, so avoiding tumour cells duplication [50].

Current chemotherapeutic agents targeting microtubules can be classified into two main categories [51] [52]:

- **microtubule destabilizing agents**, inhibiting microtubules polymerization;
- **microtubule stabilizing agents**, promoting microtubules polymerization.

The first class of molecules, which includes vinca alkaloids and colchicinoids, acts, at high concentrations, by blocking microtubules dynamics and thus mitosis, leading to cell death by apoptosis. The second class of molecules, which includes paclitaxel, on the other hand, acts by stabilizing the microtubules, so avoiding the transport of the chromosomes to the daughter cells. In both cases, the antimitotic action is due to the alteration of the correct microtubules dynamics, which inhibits the mitosis and avoid the correct distribution of the genetic material.

The main drawback in the use of chemotherapeutics is that they obviously target every kind of cell, not only tumour ones. Although cancer cells are characterized by a higher replication rate, with respect to non-cancer cells, so that they're more affected by the action of a chemotherapeutic, side effects usually occur [53] [54], such as the decrease in the production of blood cells by the bone marrow, the insurgence of ulcers or the reduction of the immune defences. Moreover, the effectiveness of such agents is made troublesome by the insurgence of mechanisms of resistance by tumour cells, linked to the alteration of the microtubules structure or the mutation of tubulin amino acidic sequence.

Current antimitotic agents are found to target three different sites in microtubules structure (**Figure 5**):

- the **vinca domain**, located at the interface between two tubulin dimers;
- the **taxol site**, located on the β-tubulin surface;
- the **colchicine site**, located at the interface between two monomers.

The vinca domain [55] [56] [57] is targeted by a great number of compounds having a proved antimitotic action, such as vinblastine, rhizoxin, spongistatin and dolastatin, that bind different regions of the large interdimer surface [58]. At the taxol site binds paclitaxel [59] [60], which has a widespread clinical use against ovarian, breast and lung cancer. Colchicine site is the target for colchicinoids [61] [62] [63], whose use as anticancer agents is nonetheless limited by the high number of side effects.

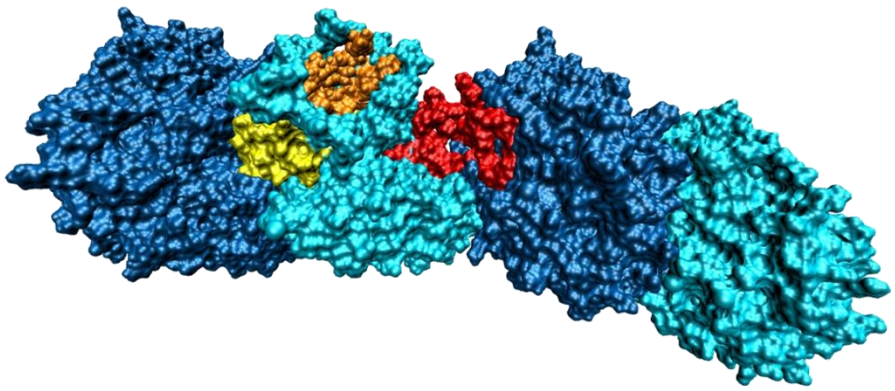


Figure 5 Representation of a tubulin tetramer made up of adjoining α-β tubulin dimers. The site of actions of currently used antimitotic agents are highlighted (the vinca domain in red, the taxol site in orange and the colchicine site in yellow)

2.4 The role of vinblastine as antimitotic agent

Vinblastine (**Figure 6**) [64] is a potent antitumor drug which has been used for a long time in the treatment of leukaemias, lymphomas, lung and breast cancers and was originally discovered in the extracts of leaves of the *Catharanthus rosea (Vinca rosea)* plant independently at the University of Western Ontario in 1958 [65] and at the Lilly Research Laboratories [66].

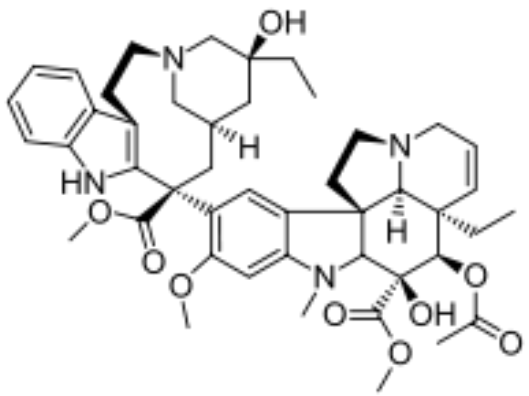


Figure 6 Vinblastine chemical structure.

Vinblastine belongs to the Vinca alkaloids class, so it inhibits the correct microtubules elongation by targeting tubulin [67] [68]. In particular, it is well known that vinblastine favours a curved tubulin protofilaments conformation [69]. Tubulin is thus not able to polymerize in a straight conformation anymore, but form spirals and rings [70] [71] [56] which cannot assemble to form microtubules correctly (**Figure 7**). As a consequence, microtubules biological action during mitosis is inhibited and tumour cells undergo apoptosis. In particular, cells that are blocked in mitosis exhibit a series of vinblastine concentration-dependent changes in microtubules and chromosomes organization [72] [49] [73], with vinblastine-induced abnormalities in

microtubules structure and dynamics becoming increasingly pronounced as vinblastine concentration raises.

Vinblastine binding site has been elusive until the definition of its crystal structure in complex with tubulin [74]. The crystal structure shows that vinblastine is bound at the interface with two tubulin dimers associating longitudinally, in such a way to act as a wedge in between them, thus forcing the tetrameric structure to a bent conformation which differs from the straight one occurring when vinblastine is not present [75].

In particular, it is interesting to observe that vinblastine binding pocket is composed of tubulin residues which are involved in the longitudinal interactions when the straight protofilament polymerizes, i.e. loop T7, helix H10 and strand S9 in the α subunit and loops T5 and H6-H7 in the β subunit.

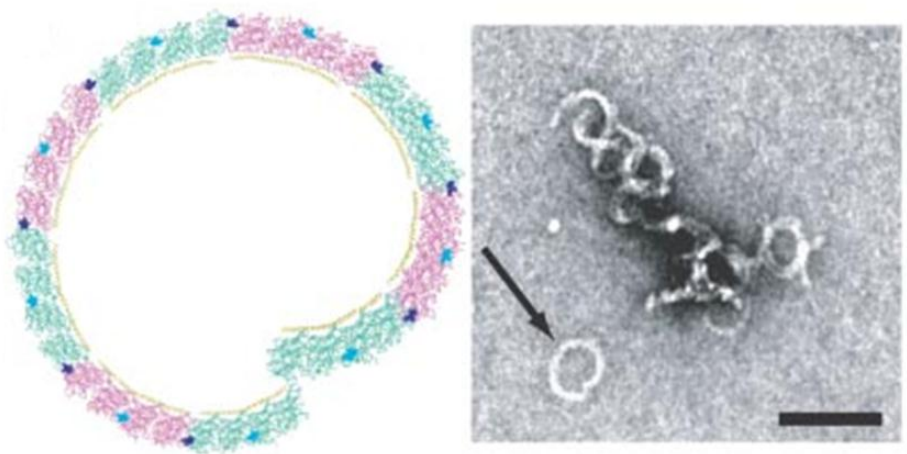


Figure 7 Vinblastine induces a bending in the protofilament structure, resulting in the formation of tubulin spirals and rings [74].

When vinblastine is bound at the interface, 80% of its surface area is buried between the two dimers, so forming a tight packing with α and β tubulin. A deeper understanding of the nature and energetics of the interactions between tubulin and vinblastine is essential to characterize the mechanism through which the bent tubulin conformation is stabilized, providing useful details for the development of new drugs with improved properties targeting the vinca site [76].

Chapter 3

Vinblastine mechanism of action

3.1 Molecular dynamics simulations

In order to get into more depth in the mechanism of stabilization of a curved tubulin protofilament by vinblastine, two molecular dynamics simulations were performed on two different systems. The first system taken into account consisted of two tubulin dimers interacting longitudinally to form a straight tetramer. The second system included a tubulin tetramer-vinblastine complex, which had thus a bent conformation, as mentioned above, due to the presence of vinblastine wedging at the interface between the two protein subunits (

Figure 8). The structures for the molecular dynamics simulations were obtained from the Protein Data Bank (PDB). Nonetheless, particular precautions had to be taken in order to get suitable starting points for the computational analysis, as explained in more details in the Materials and Methods box.

Once the linear and the bent tubulin tetramers were correctly set up, they both underwent a 5 ns molecular dynamics simulation in explicit water solvent and counterions. In order to validate the results of the molecular dynamics simulation, the Root Mean Square Deviation (RMSD) from the initial structures was calculated on α carbons (**Figure 9**). The plots clearly showed that both structures evolved to a stable conformation.

3.2 Computational alanine scanning

In order to investigate how the interaction between tubulin dimers changes in the straight or bent conformation, the attention was focused on the interfacial residues, which were identified in the linear structure as the ones showing a non-zero difference in the solvent accessible surface area (SASA), calculated with the software NACCESS [77], in the tetrameric vs. the dimeric unit. In other words, the per-residue SASA was calculated for the two single dimers and for the tetramer and interfacial residues were identified as those having a different value in the dimer and in the tetramer. A total of 176 interfacial residues was identified in this way. [78] A computational alanine scanning was then performed on the identified interfacial residues, which allowed the classification of the amino acids according to the difference in the binding free energy ($\Delta\Delta G$) upon mutation into alanine.

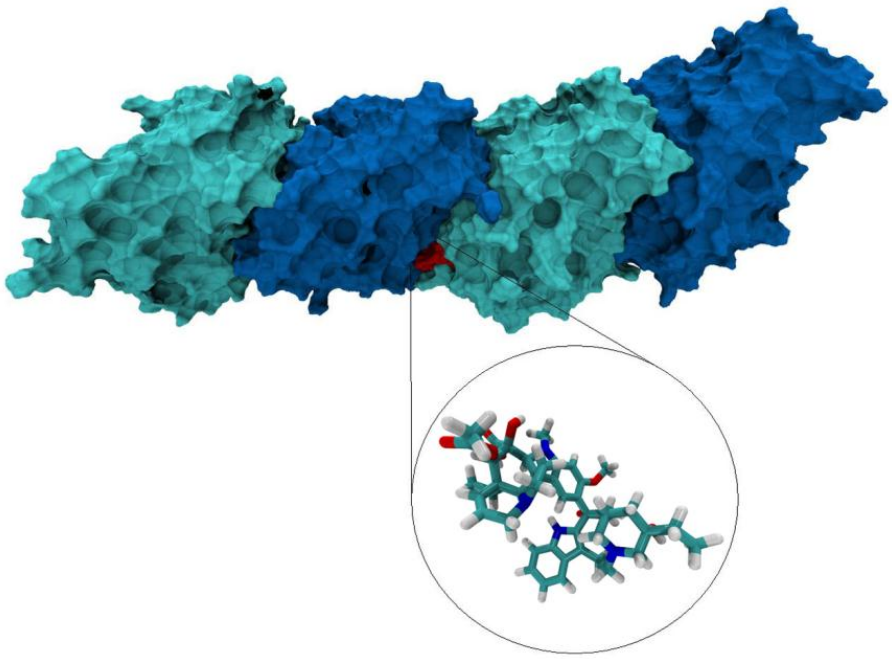


Figure 8 Vinblastine wedging at tubulin-tubulin interface stabilizes a bent tetramer conformation.

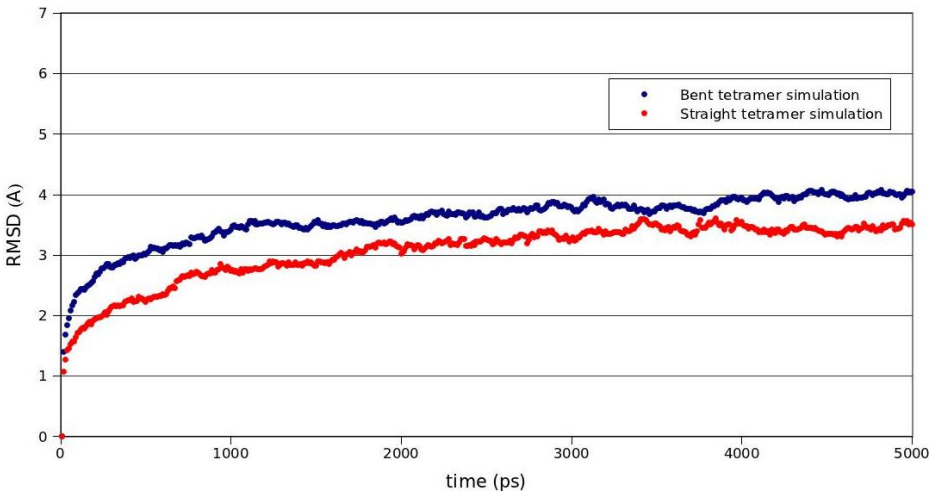


Figure 9 RMSD calculated on α-carbons for the MD simulations of the bent and the linear tetramer.

Materials and methods

The initial structure of the linear tetramer was obtained from the Protein Data Bank (PDB id 1TUB [78] corresponding to a structure of the pig tubulin α/β heterodimer determined in a polymerized state composed of antiparallel protofilaments forming a planar sheet, which results in a model particularly suitable for the study of the longitudinal interactions. The tetrameric system was built by adding to the crystal structure a second dimer with an 80 Å translation (according to crystallographic information) of the tubulin unit along the protofilament axis, which corresponds to the *x*-axis in the PDB file.

Since no crystal structure of pig tubulin in its complex with vinblastine was available, the initial structure of the curved tubulin tetramer was obtained by superimposing the pig tubulin dimer corresponding to the 1TUB pdb file onto each of the two cow tubulin dimers in the 1Z2B pdb file [74] thus obtaining a pig tubulin tetramer which matched the overall bent topology of the vinblastine-bound tubulin. The vinblastine molecule was then manually docked in its natural binding pocket corresponding to the crystal structure position and the absence of overlapping atoms was verified, as well as the conservation of the contact found in the crystal structure. Molecular dynamics simulations were carried out with the AMBER 9 package [79] using explicit solvent and periodic boundary conditions. The 2003 force field [80] was used for the protein and the Generalized Amber Force Field (GAFF) [81] for the vinblastine molecule. Missing vinblastine parameters were calculated according to Wang et al. [82]. The system was solvated with TIP3P waters [83] and neutralized with sodium ions. The SHAKE algorithm [84] was employed to constrain all bonds involving hydrogen to their equilibrium bond lengths, which allowed an integration step of 1 fs. Both the linear and the curved tetramer were first submitted to 10000 steps of minimization (1000 steps using the steepest descent and 9000 using the conjugated gradient method). The systems were then equilibrated for 100 ps with the number of particles, system volume and temperature (300 K) constant (NVT conditions) in order to equilibrate the temperature of the system, and subsequently for 100 ps with number of particles, system pressure (1 atm) and temperature (300 K) constant (NPT conditions) in order to equilibrate the density of the system, with a 500 kcal mol^{-1} $Å^{-2}$ restraint on the positions of backbone atoms.

A 5 ns productive-phase molecular dynamics simulation at NPT conditions (1 atm, 300 K) of each system followed equilibration. Positional restraints on backbone atoms were not imposed in the productive phase. A 9 Å cutoff for non-bonded interactions was applied and Particle Mesh Ewald algorithm [85] was employed to calculate long-range electrostatic interactions.

The length of the molecular dynamics run was adequate to sample the conformational variability of the protein surface and was consistent with the simulation time used by other authors in previous articles on tubulin [86] or on systems of comparable size [87]. Free energy of binding between tubulin dimers in the linear and bent tetramer, as well as that between vinblastine and tubulin was evaluated using the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) approach [30]. The entropic contribution to binding free energy was evaluated with the NMODE module of the software package AMBER 9. Due to the huge dimension of the systems under consideration, only residues forming the contact surface between the two dimers were considered, adopting a strategy that proved to be viable and reliable in dealing with large systems [87]. Calculating the entropic contribution to vinblastine binding free energy, the vibrational entropy term was evaluated for the ligand and for the protein residues laying within a 8 Å radius from vinblastine, as suggested by Kuhn and Kollman [88]. The entropic contribution was evaluated on 50 snapshots extracted from the last 2 ns of the simulation. The computational alanine scanning was carried out on snapshots taken at 10 ps intervals from the last 2 ns of the trajectories (200 snapshots in total). For each amino acid located at the interface, all side-chain atoms beyond C_β were removed and the missing hydrogen was added, so obtaining the alanine sidechain. The ΔG of binding was calculated as described above both for the wild type and the mutated complex. In this instance it is not needed to calculate the entropic contribution, since it is considered to be the same in the two cases, thus cancelling out when computing the $\Delta\Delta G$ value [31]. A 150 nM physiological saline concentration was also used for the alanine scanning calculations.

Residues showing a $\Delta\Delta G > 4$ kcal/mol were considered to give the highest contribution to the binding between tubulin subunits and were labelled as hot-spots. Residues having $\Delta\Delta G$ between 2 kcal/mol and 4 kcal/mol or less

than 2 kcal/mol were labelled as warm-spots or cold-spots, respectively. Focusing on the linear tetramer, a numerous set of hot- and warm-spots was identified, clearly displaying that the binding energy is not evenly distributed at the surface, but is mostly concentrated on some key residues. In particular, it was observed, as it usually happens at protein-protein interfaces (see Chapter 1), that such residues tended to be grouped in small clusters close to one another in the three-dimensional structure and often in the amino acidic sequence, so as to form "hot stretches" of residues of key importance for protein association. Hot- and warm-spots in the linear tetramer were found to be evenly distributed between the α and β subunit and were involved in a complex network of interactions involving salt-bridges, hydrophobic contacts and hydrogen bonds.

Computational alanine scannings on both the linear and the bent structure were performed over the last 2ns of the MD simulations, once the systems reached equilibrium (**Figure 9**). As to the vinblastine-bound system, two separate sets of alanine scanning had to be performed, the first one including vinblastine as part of α tubulin and the second one considering vinblastine as part of the β subunit. This approach was necessary to monitor not only the change in the $\Delta\Delta G$ values upon vinblastine insertion at the interface, but also to highlight interactions of any of tubulin residues with vinblastine itself, taking into account the fact that the mutual interactions between residues belonging to the α and β tubulin subunit, respectively, may be replaced by interactions with the bound vinblastine wedged in between. So, if one includes

vinblastine as part of α tubulin, its interactions with β tubulin can emerge from the results, and the same happens the other way round. This can be better understood by getting into more depth into the results of the computational alanine scanning.

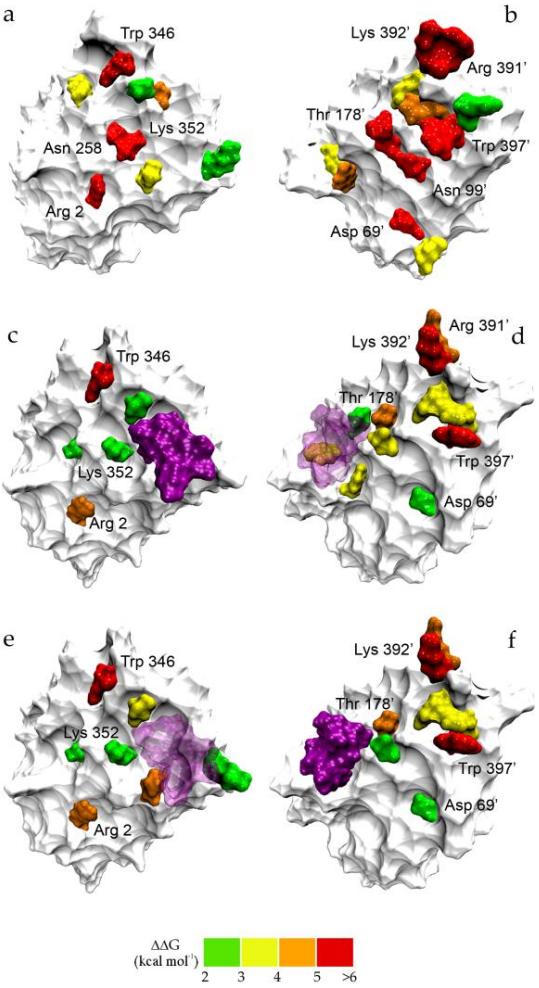


Figure 10 Hot and warm spots distribution at the dimer–dimer interface. Different colors, from green to red, refer to the variation in the binding energy of the complex on mutation of each of the hot and warm spots identified. The α and β subunits contacting surfaces are represented for the linear tetramer (a, b), and for the bent tetramer (c–f). Analyzing the bent tetramer, two distinct alanine scanning calculations were performed, considering vinblastine either as part of the α (c, d) or of the β subunit (e, f). Vinblastine is colored in purple on the subunit to which it belongs, and shaded on the opposite one.

At a first glance, the comparison between the results obtained for the straight and the bent tetramer showed a general reduction in the ΔΔG values upon vinblastine binding, as it would be expected given the conformational change induced in the tetrameric structure (**Figure 10**). The analysis performed on the straight structure led to the identification of 22 hot- or warm-spots. On the other hand, the number decreased to 15 or 17 in the bent conformation, depending on which subunit vinblastine belongs to. In particular, interfacial residues in the vinblastine-bound system were found to exhibit different behaviours of the ΔΔG values.

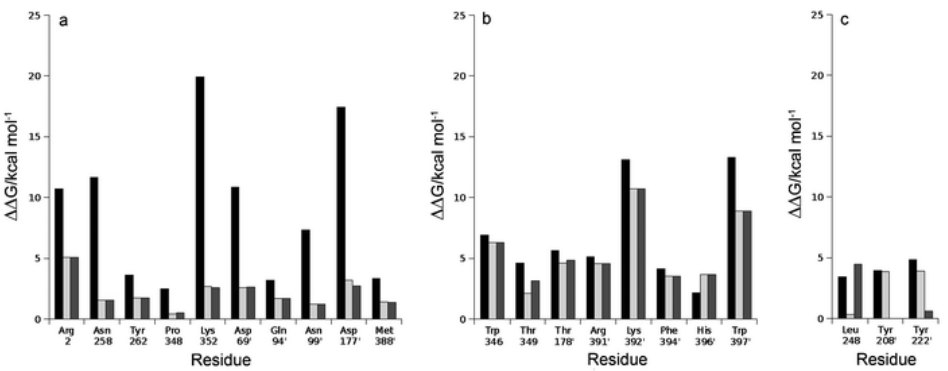


Figure 11 ΔΔ*G* values obtained from alanine scanning for the linear tetramer (black bars), and the bent tetramer, considering vinblastine either as part of the α subunit (pale grey) or the β subunit (dark grey). In panel a, residues exhibiting ΔΔ*G* reduction greater than 40% were grouped. In panel b, residues with smaller ΔΔ*G* reductions are shown. Panel c gathers those residues in the vinca site exhibiting a strong dependence of ΔΔ*G* on the subunit in which vinblastine was included during the calculation.

A group of residues showed a reduction of their ΔΔG values greater than 40% with respect to the straight tetramer, regardless of the subunit vinblastine was included in (**Figure 11a**). In this case, the reduction is due to the perturbative action of vinblastine wedging at the interface, weakening the

interactions that interfacial residues could establish in the linear conformation. This can be observed taking into account the most prominent hot-spots found in the vinblastine-free system Lys352 and Asp177′ (residues on the β subunit at the protein-protein interface are labeled with a prime on their sequence number), which took part in a strong salt bridge and were characterized by the highest ΔΔG values (19.9 and 17.43 kcal/mol, respectively). Lys352 was also involved in a second salt bridge (**Figure 12**) with the γ-phosphate of the GTP molecule bound to the facing β tubulin subunit, whereas Asp177′ established also a hydrogen bond with the Asn258 side chain. The binding of vinblastine at the interface led to a rearrangement of these residues and, as a consequence, the salt bridge between Lys352 and Asp177′ got weakened, with the average distance between the lysine amine nitrogen ant the aspartate carboxylic carbon increasing from 3.79 ± 0.17 Å in the linear system simulation to 4.06 ± 0.65 Å in the bent system one. The interaction of Lys352 with GTP was also affected, with the distance between the lysine amine nitrogen and the GTP γ-phosphorus raising from 3.52 ± 0.15 Å to 7.36 ± 1.0 Å. The hydrogen bond between Asp177′ and Asn258 was even no longer detected in the vinblastine bound complex. It is interesting to notice that Lys352 and Asp177′ are located at the edge of vinblastine binding site, so a dramatic effect is stressed when vinblastine is bound at the interface.

Vinblastine can also affect indirectly the interactions between facing residues, as can be seen taking into account the behaviour of Arg2, which had its interactions weakened though located 15 Å apart vinblastine binding site

(**Figure 12**). In this case, Glu69′ was involved in electrostatic interactions with Arg2 and the amine charged group of Met1 in the linear tetramer. In the vinblastine-bound system, the bending of the structure caused a rearrangement of the N-terminus resulting in a reduction of Arg2 burial, with its solvent accessible surface area rising from an average of 15 ± 4 Å^2 in the linear tetramer to 62 ± 10 Å^2 in the bent one, together with a departure of Met1 from Glu69′. As are result, ΔΔG values for these residues are indirectly lowered by vinblastine binding at the interface.

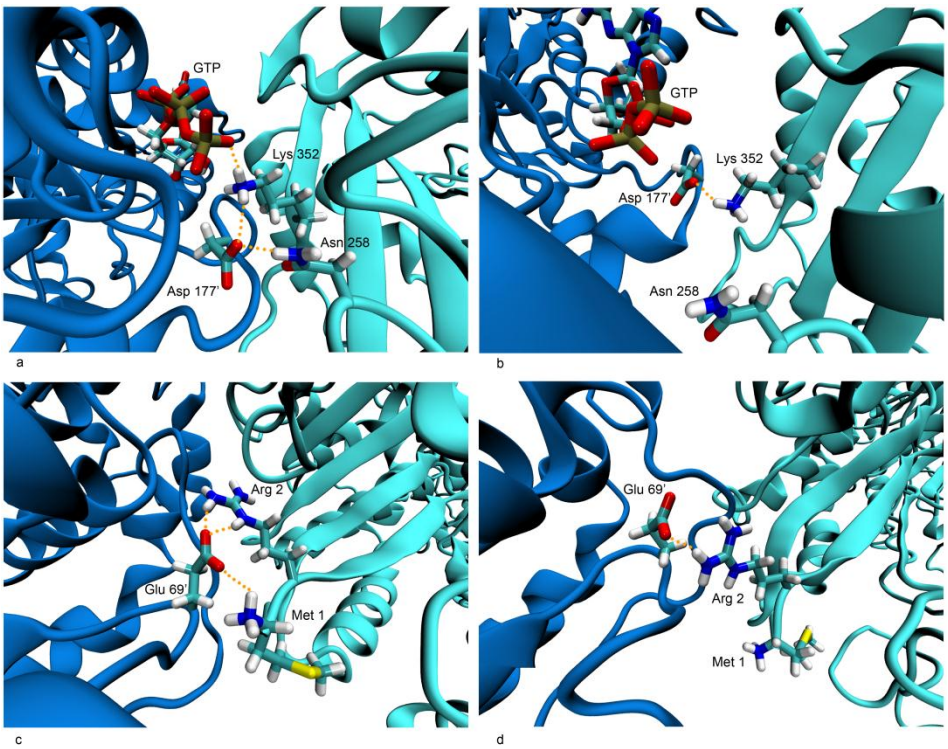


Figure 12 Details of the network of interactions between tubulin dimers in the linear (a, c) and bent (b, d) tetramer are shown. Key salt bridges and hydrogen bonds that are preserved or lost upon vinblastine binding are highlighted.

Another group of amino acids wasn't so much affected by the perturbating action of vinblastine and underwent only small changes in their ΔΔG values with respect to the straight tetramer (**Figure 11b**). This happened mainly for those residues located at the C-terminal of β tubulin, which are located far from vinblastine binding site and whose structure in the linear and the bent tetramer is superimposable [89]. In fact, this region is the hinge around which dimers rotate to reach the bent conformation from the linear one.
These residues thus are able to keep the interactions with the facing subunit, although vinblastine is inserted in between and their ΔΔG values are only slightly changed.

The most interesting behaviour was exhibited by three residues belonging to vinblastine binding site, Leu248 on the α subunit and Tyr208' and Tyr222' on the β subunit (**Figure 11c**). In these case, the ΔΔG resulted to be very different depending on the subunit in which vinblastine was included for the computational alanine scanning. In particular, Leu248, belonging to the α subunit, showed a ΔΔG value of 3.43 kcal/mol in the straight conformation, turning to zero when vinblastine was considered as part of the α subunit itself, but, on the other hand, slightly increased to 4.46 kcal/mol when vinblastine was considered as part of the β subunit. One can thus infer that in this case it is vinblastine itself which act as an interacting counterpart of Leu248 and such interaction can be detected only when vinblastine is considered as part of the facing subunit. For the same reason, Tyr208' and Tyr222' conserved their ΔΔG values when vinblastine was considered as part of the opposite α subunit.

Therefore, in the bent conformation, these three residues did not interact anymore with the facing tubulin residues, but with vinblastine, which thus acted by crosslinking the interactions between α and β subunits. Getting into more detail, Tyr222′ was involved in hydrophobic contacts with Leu248 and Pro325, whereas Tyr208′ established hydrophobic interactions with Lys326, Pro325 and Asp329. On the other hand, in the bent structure Tyr208′ and Tyr222′ turned to interact with vinblastine catharantine domain, whereas Leu248 established hydrophobic contacts with both the catharantine and the vindoline domain (**Figure 13**).

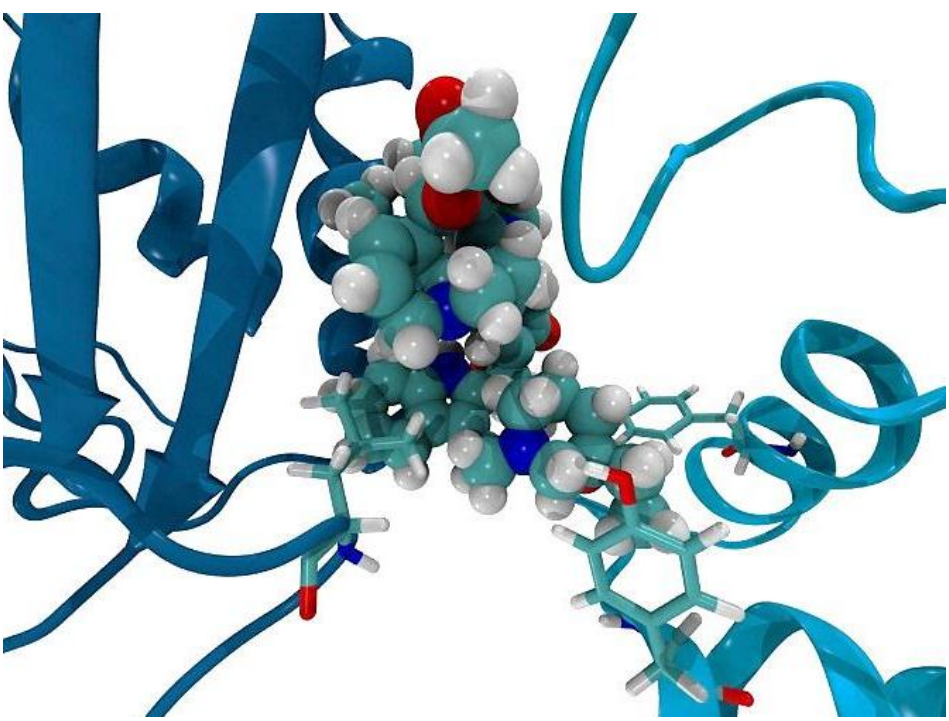


Figure 13 Detail of the hydrophobic interactions contacts established by vinblastine (represented in Van der Waals spheres) with Leu248 on α tubulin and Tyr208′ and Tyr 222′ on β tubulin.

The analyses showed that, although potential hydrogen bond donors and acceptors were present, the interactions stabilizing the tubulin-vinblastine complex were essentially hydrophobic.

The observed vinblastine crosslinking behaviour allowed an explanation for the stabilization of a tubulin bent conformation: whereas the disruption of a number of key interactions for tubulin self-association was induced by vinblastine, on the other hand the molecule itself introduced new interactions with both α and β subunits, acting as a glue between them. As a consequence, the tetramer is obviously forced to a curved conformation, with a ~18° angle between the dimers observed throughout the simulation, in agreement with the value of the crystal structure. The conformational change also led to an increase of the distance between the centres of mass of the interacting tubulin dimers, which turned from 80.49 ± 0.23 Å in the linear tetramer to 82.66 ± 0.48 Å in the bent one.

3.3 Free energy change calculations

Computational alanine scanning only provided and evaluation of the role of the single amino acids in vinblastine stabilizing action. In order to investigate the compensation of energy loss due to the bending upon vinblastine insertion at dimers interface, the free energy change as a whole from

the linear to the bent conformation needed to be calculated. Due to the large dimensions of the system, a molecular dynamics simulation of the bending process was impossible to perform. It was thus necessary to decompose the process into elementary reactions whose ΔG could be more easily estimated computationally. For this purpose, the thermodynamic cycle showed in **Figure 14** was used, so that it was necessary to calculate the binding free energy between two dimers forming the linear tetramer (ΔG_3), the binding free energy between two dimers in the presence of vinblastine to form the bent tetramer (ΔG_4) and the binding free energy of vinblastine bound to only one tubulin dimer (ΔG_2). Also in this case, as it was necessary to do for the computational alanine scanning, vinblastine had to be considered as part of either α or β subunit for the calculation of ΔG_2 and ΔG_4. As a result, two values for the free energy of addition of vinblastine to a linear tetramer to get the bent structure (ΔG_1) were obtained.

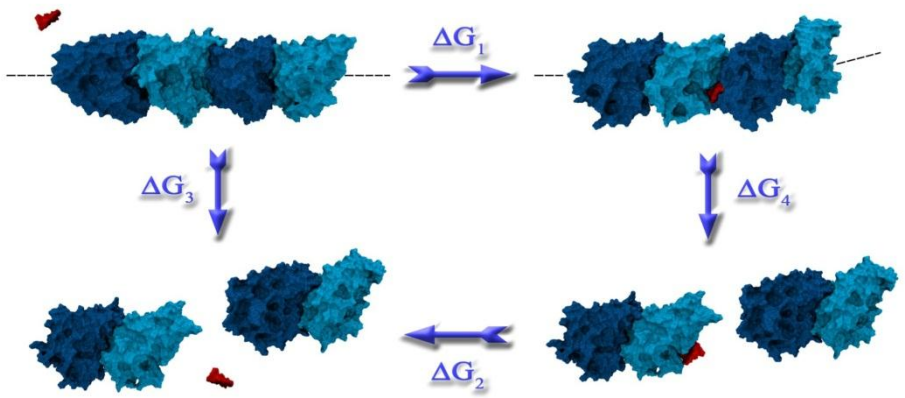


Figure 14 Thermodynamic cycle used to calculate the free energy change as a whole from the linear to the curved tetramer.

The entropic contribution was also taken into account for the binding free energy calculations. Nonetheless, due to the large size of the system introducing memory and CPU limits, it was not possible to evaluate it considering the whole complex, but it was calculated only on a subset of it, as explained in the Materials and methods box. The calculations led to a value of binding free energy of -16.51 kcal/mol (with a standard error of 3.18 kcal/mol) when vinblastine was considered as part of the α subunit and of -11.39 kcal/mol (with a standard error of 3.19 kcal/mol) when vinblastine was included in the β subunit. The results thus showed that the formation of the bent tetramer is favoured in the presence of vinblastine wedging between tubulin subunits and that vinblastine binding free energy actually compensates the energy loss following the tetramer bending. It is clearly deductible that the association of tubulin dimers is still possible in the presence of vinblastine, although the formation of curved protofilaments (and not straight) is favoured, resulting in an incorrect microtubules formation and the subsequent inhibition of tubulin biological activity.

3.4 Vinblastine effect over GTP hydrolysis

A further aspect which is worth noticing, is the influence of vinblastine on the hydrolysis of the tubulin-bound GTP, whose rate is experimentally

observed to decrease [90]. Comparing the MD simulation of the bent and the linear tetramer, it was possible to highlight a change in the positioning of the α tubulin H8 helix [91] [92], which includes the residue Glu254 that was proven to play a fundamental role in the GTP hydrolysis [93]. In fact, an increase of the distance between Glu254 and the GTP γ-phosphate was observed, with its value raising from 3.54 ± 0.26 Å in the linear tetramer to 5.24 ± 0.63 Å in the vinblastine-bound one. As a result, water molecules were allowed to slip into the gap between Glu254 and GTP, so disrupting the correct residues orientation for nucleotide hydrolysis.

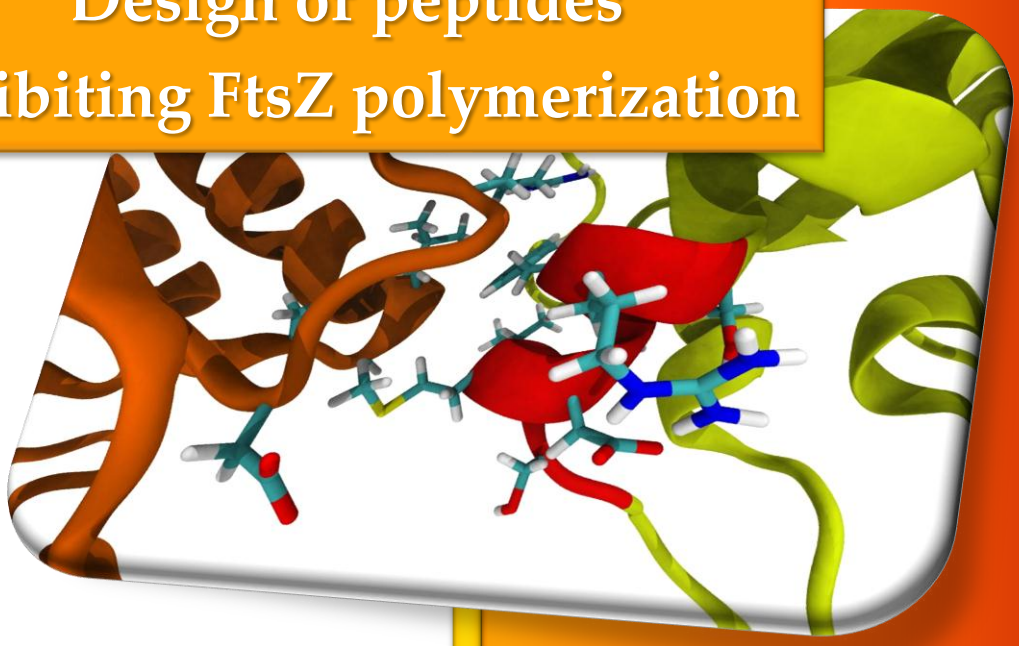
Part I

Conclusions

The changes in the network of interactions at the protein-protein interface induced by vinblastine binding to tubulin were explored through a computational approach comprising molecular dynamics simulations and computational alanine scannings. The analyses allowed the identification of the key residues involved in the interactions between tubulin and vinblastine, pointing out that the molecule effectively acts by crosslinking the interactions between the dimers. As a matter of fact, vinblastine, although binding at the interface between tubulin subunits, does not screen the interdimer interactions, but, on the contrary, bridges them, thus leading to the stabilization of a curved tubulin–vinblastine complex. Moreover, the structural changes induced by vinblastine are consistent with the decrease of the rate of GTP hydrolysis observed experimentally. The stabilization of a curved tetramer conformation can be seen as the first step of the formation of bent tubulin polymers, which is the primary cause of microtubules disassembly. The deep insight into the nature of the interactions between vinblastine and tubulin achieved could be valuable in the better understanding of vinblastine mechanism of action to contribute to the development of new antimitotic agents targeting the vinca site.

Part II

Design of peptides inhibiting FtsZ polymerization



Chapter 4

FtsZ and z-rings

4.1 FtsZ role in bacterial cells

In prokaryotes, such as bacteria, cell division occurs through a protein homologue to tubulin known as FtsZ [94]. Like tubulin, FtsZ polymerizes in a GTP-dependent way [95] [96] to form a circular structure called z-ring [97], whose contraction during the division process allows the septation and the separation of daughter cells (**Figure 15**) [98]. Nonetheless, details about how z-rings assemble and exert their action during cell division are still not clear. What is known is that, *in vivo*, GTP binding to FtsZ promotes a head-to-tail protein association to form linear protofilaments, like in microtubules assembly [99]. Protofilaments later on establish lateral interactions to form protein suprastructures constituting the z-ring [100] [101], but nothing is known yet

about its set-up, i.e. whether it's made up of single protofilaments, bundle or helices.

Nonetheless, recent analyses showed that FtsZ polymers, unlike microtubules, are mainly constituted by GTP-bound monomers, whereas nucleotide hydrolysis upon FtsZ self-association has a destabilizing effect [102].

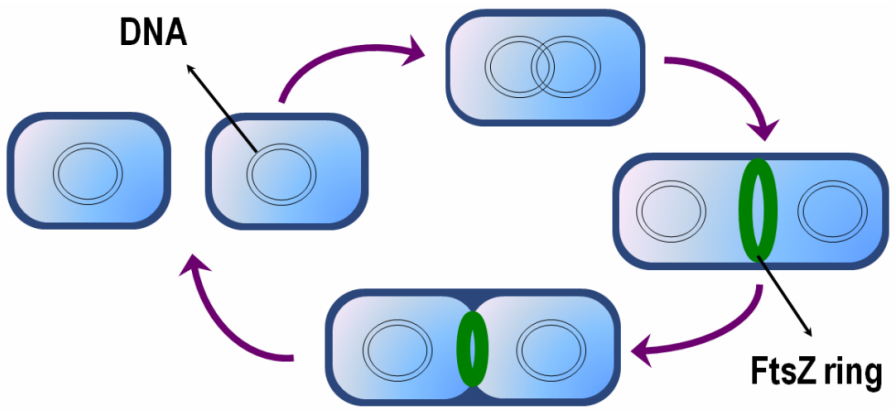


Figure 15 Z-rings are FtsZ polymers whose contraction has a key role in bacterial cells duplication.

FtsZ is thus allowed to polymerize until GTP is present inside the cytoplasm, allowing a continuous nucleotide exchange with hydrolysed FtsZ-bound GTP, whereas protofilaments disassemble when GTP is not available anymore [103]. The main explanation for this dynamics, which is complementary to microtubules' one, is that GTPase reaction kinetics are different for tubulin and FtsZ. In microtubules, GTP hydrolysis occurs very quickly upon tubulin self-association, but the nucleotide binding site is buried at protein interface, thus avoiding nucleotide exchange [104]. Therefore, microtubules are mostly made up of GDP-bound tubulins. On the other hand,

the rate-limiting step for FtsZ polymerization is GTP hydrolysis, since the nucleotide binding site has an opening to the solvent of approximately 8 x 9 Å [105], which allows the GTP to diffuse freely right after its hydrolysis and nucleotide exchange can rapidly occur.

4.2 FtsZ homology to tubulin

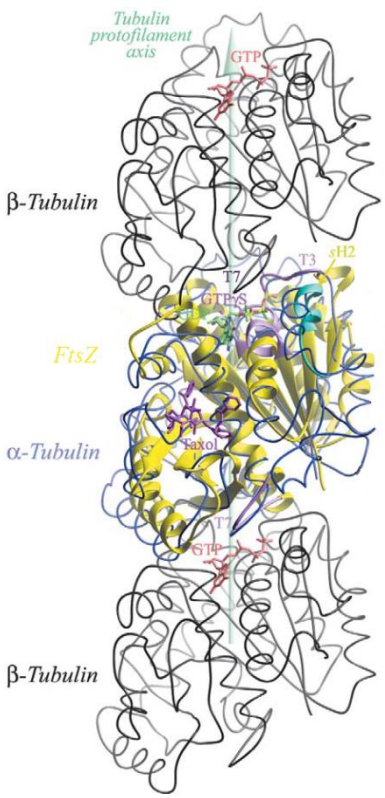


Figure 16 FtsZ has a high structural homology to tubulin.

Tubulin and FtsZ have a high structural homology [106] [107] (**Figure 16**). The cores of the two proteins are in fact structurally identical, whereas discrepancies occur at the C-terminal (which is composed of a pair of long helices for tubulin and of a small β-hairpin for FtsZ) and the N-terminal section (where FtsZ has an extra helix). The similarity between the two proteins is confirmed by the low RMSD value (4.3 Å) calculated over the α-carbon atoms within the common core residues [108].

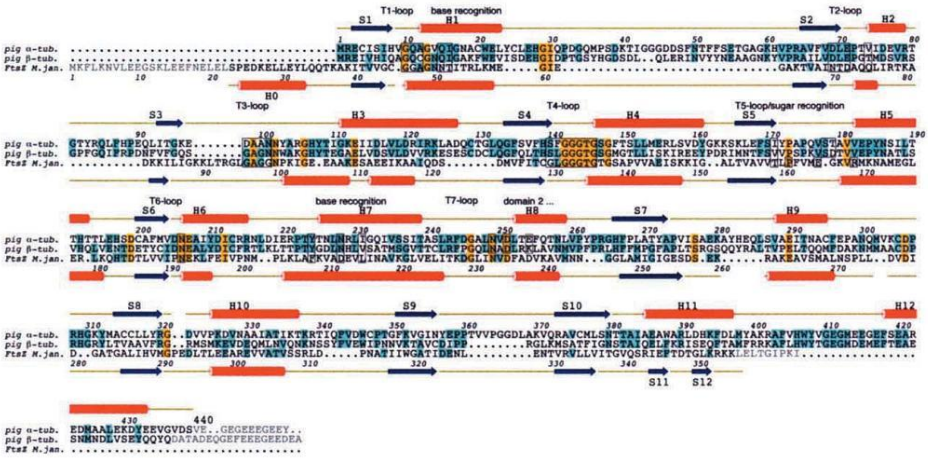


Figure 17 Sequence alignment between tubulin and FtsZ shows a matching only in biologically relevant regions.

Despite the high structural homology, the structure-based sequence alignment shows a weak sequence match (**Figure 17**). The regions of highest homology correspond to residues involved in nucleotide binding site (loops T1, T2, T3, T4, T6). A further region of common residues corresponds to the T7 loop, which is involved in the longitudinal association in tubulin polymerization, thus leading to the assumption that FtsZ polymerizes in a similar way [109]. Mutagenesis studies of *E.coli* FtsZ showed that T7 loop is also important for GTPase activity, indicating that GTP hydrolysis involves residues from two different subunits and can occur only when they come into vicinity as the head-to-tail associations takes place [110].

Regarding the lateral interactions, as already pointed out, only few information is known. Nonetheless, in 2004, Leung and colleagues were able to crystallize *Mycobacterium Tuberculosis* FtsZ arranged in a laterally oriented

dimer showing remarkable differences with respect to tubulin lateral interactions [111].

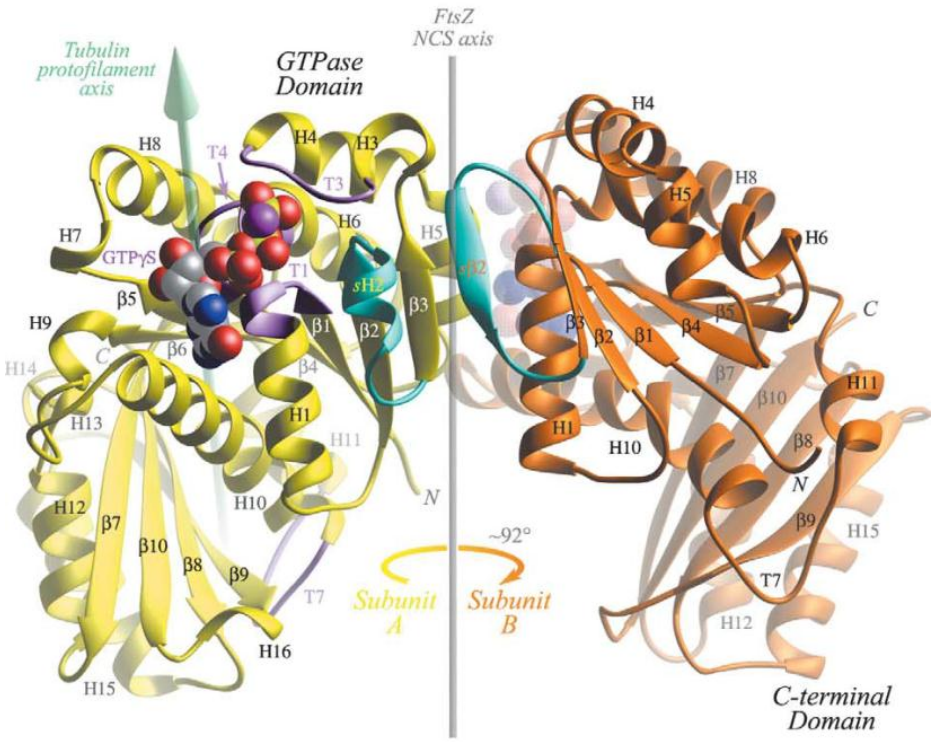


Figure 18 Structure of two FtsZ monomers interacting laterally.

To achieve such dimer conformation, subunit A in **Figure 18** has to undergo unexpected secondary structure rearrangements to establish the proper lateral interactions [112]. In fact, whereas helix H2 adopts the expected α-helix conformation in subunit A, it extends to a β-strand conformation in subunit B. Moreover, it shows that loop T3 is fully ordered around GTP when it's bound to subunit A, whereas the loop becomes disordered in the absence of

the nucleotide in subunit B [113]. Similarly, loop T4 is wrapped around the nucleotide in the GTP-bound FtsZ, but collapses inside the empty active site when nucleotide is not present. It was found that GTP binding is closely linked to the switching between the α-helical and the β-strand conformation of H2. In fact, the only way for H2 to avoid an overlap with the ordered T3 loop is to adopt an α-helix conformation. On the other hand, when the nucleotide is not bound and the T3 loop can open, H2 is free to protrude towards the nucleotide binding site and extend in a β-strand conformation. All these rearrangements make the lateral dimer conformation possible and allow the proper orienting of interfacial residues for binding. Mutagenesis studies pointed out that the formation of such lateral dimers has an important relevance in the biological activity of the FtsZ. In fact, the mutation of the residues Asp84 and Asp94, which are located at the lateral interface, destabilize FtsZ polymerization or even prevent z-ring assembly [114]. Nonetheless,

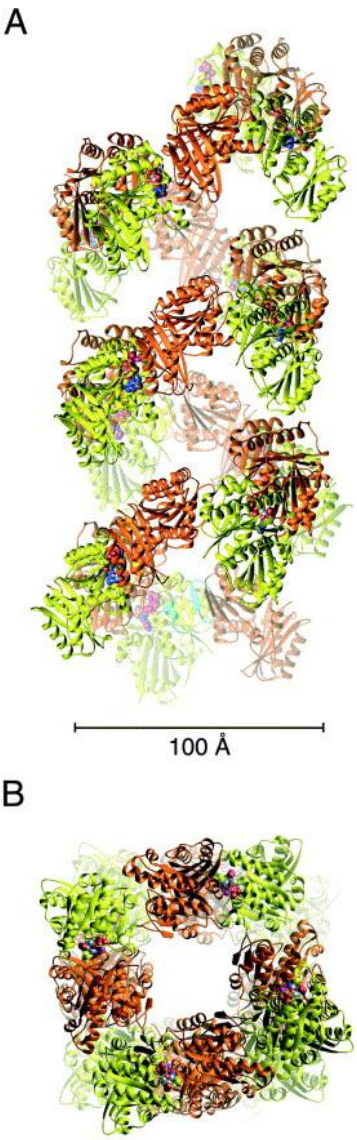


Figure 19 Hypothetical z-ring helical structure based on known FtsZ structural data.

although the physiological relevancy of such lateral contacts is clear, it is not yet known how the longitudinal polymerization combines with lateral FtsZ association to give rise to those suprastructures the z-ring is made up of. Based on the evidence that FtsZ can assemble in spiral filaments, as seen by electron microscopy assays [115], a possible spiral assembly model can be built using the lateral dimer as a base [111] (**Figure 19**).

4.3 FtsZ as innovative target in drug design

Given its importance in bacteria replication, FtsZ is gaining an ever-growing attention in the development of antibiotic agents, since the alteration of the regular FtsZ assembly could eventually prevent bacterial cells replication, similarly to what happens for antimitotic agents and tubulin [116] [117] [118]. The discovery of a new class of antibiotic agents has nowadays an important relevance in the medical field. In fact, since the discovery of penicillin in the first decades of the XX century, bacteria developed in the last years an ever growing resistance to commonly used antibiotics [119] [120]. This was initially due to a misuse and overuse of antibiotics in hospitalized patients, leading to the appearance of resistant bacteria which later on spread to the community, causing severe illnesses in previously non-vulnerable patients.

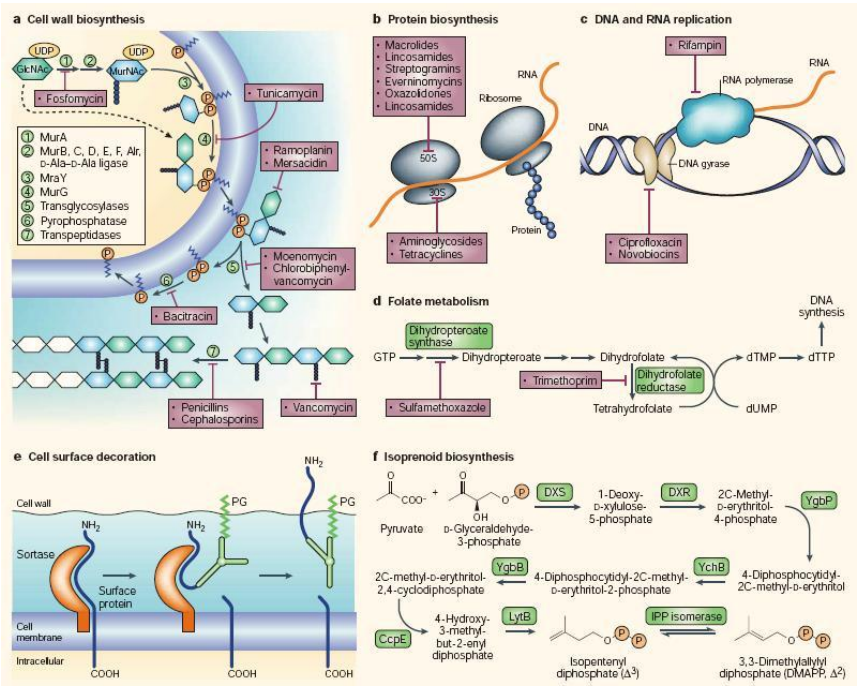


Figure 20 Currently used antibiotics target a large number of sites in bacterial cells [121].

Resistance to antibiotics occurs through mutations of bacteria themselves, which are thus becoming not targetable anymore by the same agents that were originally active. Current antimicrobials can be classified in 17 classes [121] acting against a wide range of targets, i.e. inhibiting cell-wall synthesis or DNA replication, so the way bacteria respond is different for each case (**Figure 20**). In order to overcome resistance, antibiotics have been modified in order to preserve their action. Nonetheless, the discovery of new targets would provide a significant improvement, due to the difficulty in developing active modified agents. Targeting FtsZ could thus be a valuable strategy in the discovery of a new class of effective antimicrobials with low

resistance issues. In fact, interfering with FtsZ self-assembly is an eligible way to stop bacterial proliferation [122]. Numerous small molecules targeting FtsZ have been described so far, mainly derived from natural sources and identified by high throughput screening. A few peptidic inhibitors have also been studied, i.e. the antimicrobial peptides Edeine B1 [123] and MCiZ [124], as well as stathmine-derived peptide [117].

4.4 A new strategy in drug design

A viable approach towards the alteration of z-ring formation is to interfere with protein-protein interactions responsible for FtsZ self association. The design of molecules targeting protein-protein interactions, albeit proven to be a very complex task, due to the peculiar features of protein interfaces and their inherent variability (see Chapter 1), has lately gained an increasing interest as a therapeutic strategy and is presently one of the main goals of medicinal chemistry [125]. As already mentioned, in general only a minor part of the residues at protein-protein interfaces actually contribute to the binding energy. Therefore, molecules inhibiting protein-protein association should be targeted to those particular "hot-regions" critical for proteins binding. In the case of FtsZ, an in-depth analysis of the network of interactions between single monomers (known the dimer structure described above) would lead to a

thorough understanding of how the subunits associate and can thus give a hint about where inhibiting agents should bind. Similarly to the approach used in the Part I to analyse the tubulin system, in the present part a computational alanine scanning can thus be performed over the interfacial FtsZ residues, so identifying those contributing significantly to the binding energy (i.e. warm- and hot-spots). If a potential binding epitope is found, and if it is made up by residues that are next to one another not only in the topology but also in the amino acidic sequence of the protein, then a set of peptides can be synthesised to match the identified motif (**Figure 21**). Such peptides could therefore competitively inhibit FtsZ polymerisation by selectively binding to protein sites critical for the self-assembly process. A similar approach was used in a recent paper by Pieraccini et al. [86] on the inhibition of tubulin-tubulin interactions, providing reliable results and leading to the design of effective antimitotic peptides [126] [127].

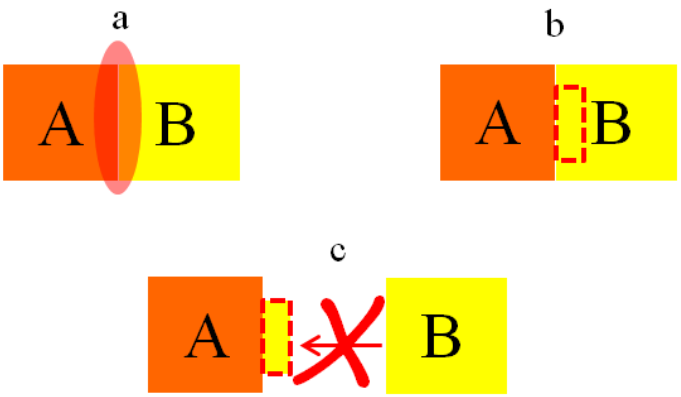


Figure 21 Schematic representation of the approach used. The computational analysis of the protein-protein interface (panel a) can lead to the identification of a set of contiguous hot-spots (panel b). By cutting out the set of hot-spots from the mother subunit, one could obtain a small peptide able to inhibit protein-protein association.

Chapter 5

FtsZ dimer analysis

5.1 Molecular dynamics simulation

As a starting point for the analysis of the interactions between FtsZ monomers, the structure of the dimer obtained crystallographically by Leung and colleagues was taken into account. Nonetheless, it was necessary to improve it so to make it suitable for a molecular dynamics simulation (see Materials and Methods box). Once an optimal starting structure was obtained, the FtsZ dimer underwent a 5 ns molecular dynamics simulation. As it happened in the case of the tubulin systems, the complex reached equilibrium during the simulation, as it can be pointed out by evaluating the RMSD plot calculated on the α-carbons (**Figure 22**). The binding free energy between FtsZ monomers was also calculated throughout the simulation, resulting in an average ΔG value of -41.75 ± 5.70 kcal/mol (**Figure 23**).

Materials and methods

The initial structure of the FtsZ dimer was obtained from the Protein Data Bank (PDB id 1RLU [111]). Nonetheless, the lack of some residues in the crystal structure, which couldn't be detected experimentally, made it in principle impossible to perform a molecular dynamics simulation of the system. Whereas PDB's chain A didn't have missing residues, chain B lacked residues from Arg60 to Gly69 and from Gly170 to Ala173. In particular, residues from Arg60′ to Gly69′ corresponded to loop T3, which, as mentioned in Chapter 4, is disordered and thus not detectable on the subunit B, whereas it is fully ordered around GTP molecule on the subunit A. The missing loops where refined using the software MODELLER 9.6 [128]. The stereochemical quality of the resulting structure was checked using the PROCHECK software [129]. Molecular dynamics simulations were carried out with the AMBER 9 package [79] (2003 force field [80]), using explicit solvent and periodic boundary conditions. The system was solvated with TIP3P waters [83] and neutralized with sodium ions. The SHAKE algorithm [84] was employed to constrain all bonds involving hydrogen to their equilibrium bond lengths, which allowed an integration step of 1 fs. The dimer was first submitted to 10000 steps of minimization (1000 steps using the steepest descent and 9000 using the conjugated gradient method). The system was then equilibrated for 100 ps with the number of particles, system volume and temperature (300 K) constant (NVT conditions) in order to equilibrate the temperature of the system, and subsequently for 100 ps with number of particles, system pressure (1 atm) and temperature (300 K) constant (NPT conditions) in order to equilibrate the density of the system, with a 500 kcal mol^{-1} $Å^{-2}$ restraint on the positions of backbone atoms. A 5 ns productive-phase molecular dynamics simulation at NPT conditions (1 atm, 300 K) of each system followed equilibration. Positional restraints on backbone atoms were not imposed in the productive phase. A 9 Å cutoff for non-bonded interactions was applied and Particle Mesh Ewald algorithm [85] was employed to calculate long-range electrostatic interactions. The free energy of binding between the subunits was evaluated using the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) approach [30]. Computational alanine scanning was carried out on snapshots taken at 10 ps intervals from the last 2 ns of the trajectories (200 snapshots in total) using a 150 nM physiological saline concentration.

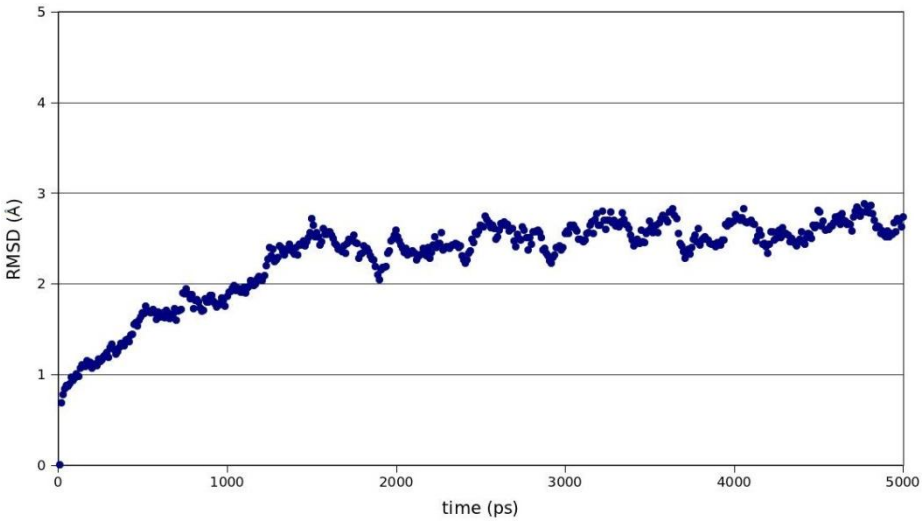


Figure 22 RMSD calculated on α-carbons throughout the simulation of the FtsZ dimer.

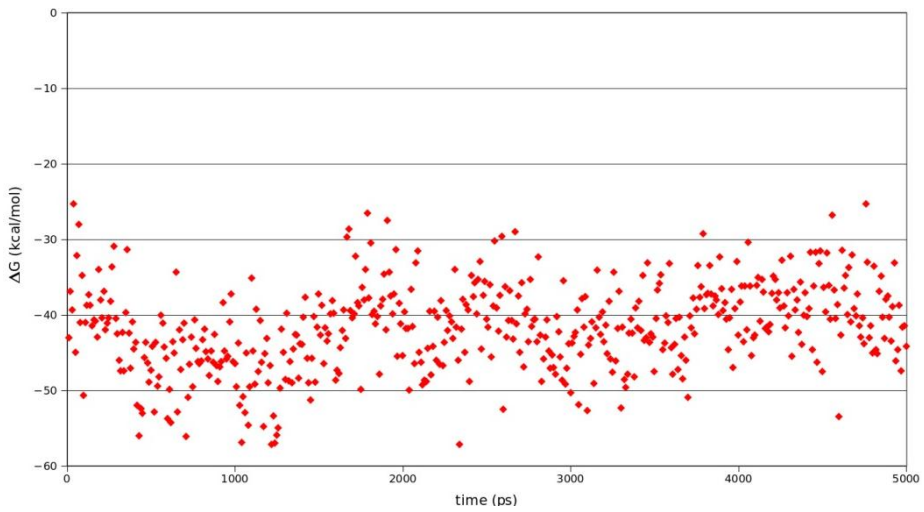


Figure 23 Plot of the binding free energy between FtsZ monomers throughout the simulation.

5.2 Computational alanine scanning

In order to get into more depth in the network of interactions stabilizing the FtsZ dimer and to understand on which residues the free energy of binding is concentrated, a computational alanine scanning was performed on the interfacial amino acids. As in the analysis of the tubulin complexes, the residues were identified as those showing a non-zero difference in the solvent accessible surface area in the dimeric vs. the monomeric unit using the software NACCESS [77]. In total, 71 interfacial residues were identified. A computational alanine scanning (**Figure 24**) was then performed on every interfacial amino acid, taking into account only a subset of snapshots extracted from the last 2 ns of the simulation, once the system reached equilibrium.

As in the case of tubulin, the results pointed out that the binding energy is concentrated only on a subset of residues, classified as cold-spots ($\Delta\Delta G < 1$ kcal/mol), warm-spots (1 kcal/mol $< \Delta\Delta G < 2$ kcal/mol) and hot-spots ($\Delta\Delta G > 2$ kcal/mol). On the whole, 10 hot-spots and 6 warm-spots were detected at the interface (**Table 1**). Residues belonging to subunit A are labelled with a prime. Arg60′ and Asp57 were found to be the most prominent hot-spots, with a $\Delta\Delta G$ value of 16.63 ± 1.92 kcal/mol and 11.76 ± 1.98 kcal/mol, respectively, with the two residues establishing a strong salt bridge. Nonetheless, Arg60′ and Asp57 were isolated hot-spots, since no other warm- or hot-spots were identified in their vicinity, so making them not interesting for the purpose of our research, i.e. the identification of "hot-stretches" to be used as a starting point for the

design of peptides able to bind FtsZ at the intermonomer region. Therefore, the attention was focused on the identification of clusters of contiguous warm- or hot-spots. Three groups, whose location at the interface is shown in **Figure 25** were detected:

- Group 1 on subunit B, from Ser175 to Ser182;
- Group 2 on subunit B, from Gln45 to Ser50;
- Group 3 on subunit A, from Glu88′ to Met95′.

Table 1 Warm- and hot-spots detected at the FtsZ dimer interface. In the last column, the group they belong to is shown.

Residue	ΔΔG value (kcal/mol)			Group
Leu56′	2.78	±	0.62	
Arg60′	16.63	±	1.92	
Leu89′	2.13	±	0.57	3
Arg91′	10.33	±	1.97	3
Asp94′	1.20	±	2.68	3
Lys120′	2.51	±	1.68	
Leu122′	1.30	±	3.37	
Leu47	1.68	±	0.38	2
Leu48	3.89	±	0.60	2
Met49	3.61	±	0.95	2
Asp57	11.76	±	1.98	2
Glu102	4.44	±	1.68	
Arg139	1.20	±	0.62	
Met177	1.86	±	1.19	1
Phe180	4.13	±	0.73	1
Arg181	1.76	±	2.56	1

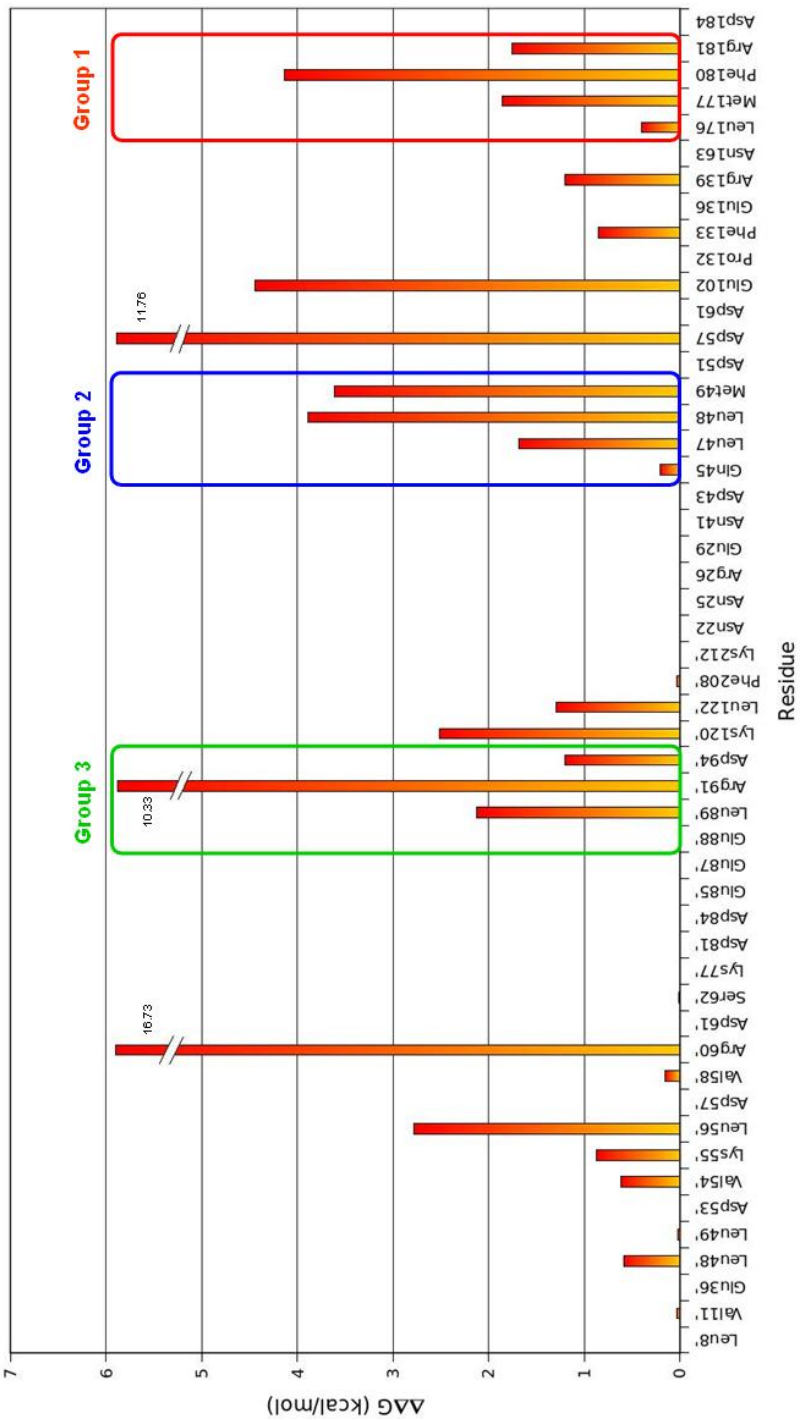


Figure 24 Results of the computational alanine scanning performed on all residues at the FtsZ dimer interface. Three "hotstretches" including contiguous warm- or hot-spots were identified: Group 1 (in red), Group 2 (in blue) and Group 3 (in green).

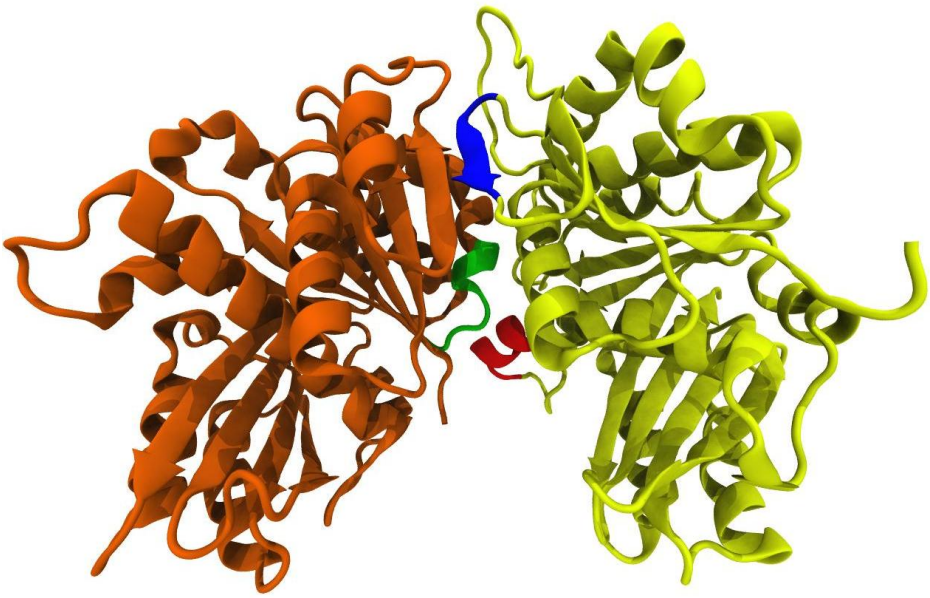


Figure 25 Representation of the simulated dimer. The "hot-stretches" identified by computational alanine scanning are highlighted in red (Group 1), blue (Group 2) and green (Group 3)

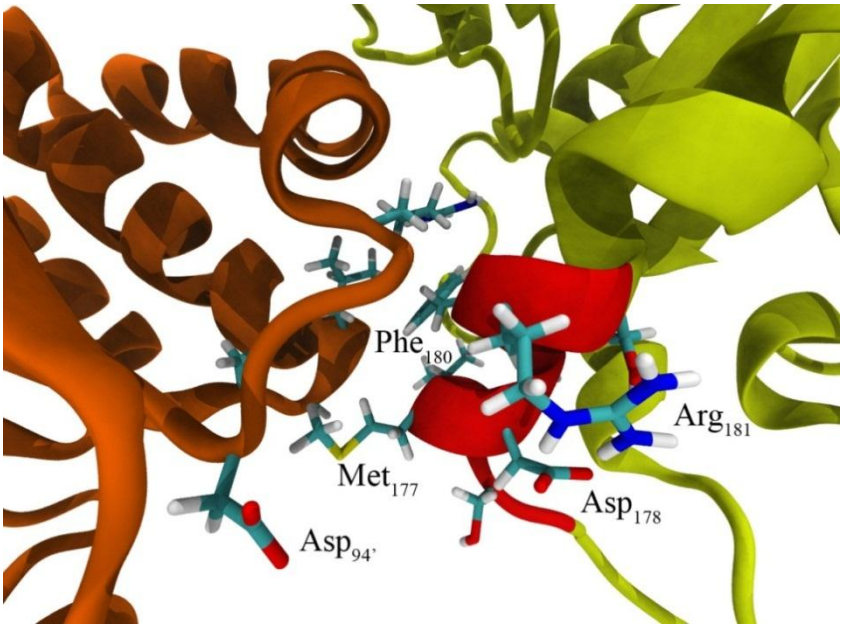


Figure 26 Detail of Group 1 residues at the FtsZ interface.

Taking into accout the Group 1, the main contributions to the binding are given by the hydrophobic interactions established by Phe180, with a ΔΔG of 4.13 ± 0.76 kcal/mol, and Met177, with a ΔΔG of 1.86 ± 1.19 kcal/mol. Arg181 is involved in a salt bridge with Asp94', although, as shown in **Figure 26**, arginine is also free to orient towards subunit B itself and interact with Asp178, so reducing its contribution to the binding with the facing protein.

As to the Group 2, the binding to subunit B is ensured by a stable hydrophobic cluster composed by Leu47, Leu48 and Met49, whose ΔΔG values are 1.68 ± 0.38 kcal/mol, 3.90 ± 0.59 kcal/mol and 6.61 ± 0.95 kcal/mol, respectively.

In the Group 3 (**Figure 28**), the most prominent hot-spot was Arg91', with a ΔΔG value of 10.33 ± 1.97 kcal/mol, establishing strong salt bridges with Glu102 and Glu136 with the facing subunit B. The binding is further stabilized by a hydrophobic contact established by Leu89', with a ΔΔG of 2.12 ± 0.56 kcal/mol. Asp94', as already mentioned, estabishes a weak salt bridge with Arg181 belonging to Group 1, so its ΔΔG of 1.20 ± 2.68 kcal/mol is charactherized by a high standard deviation.

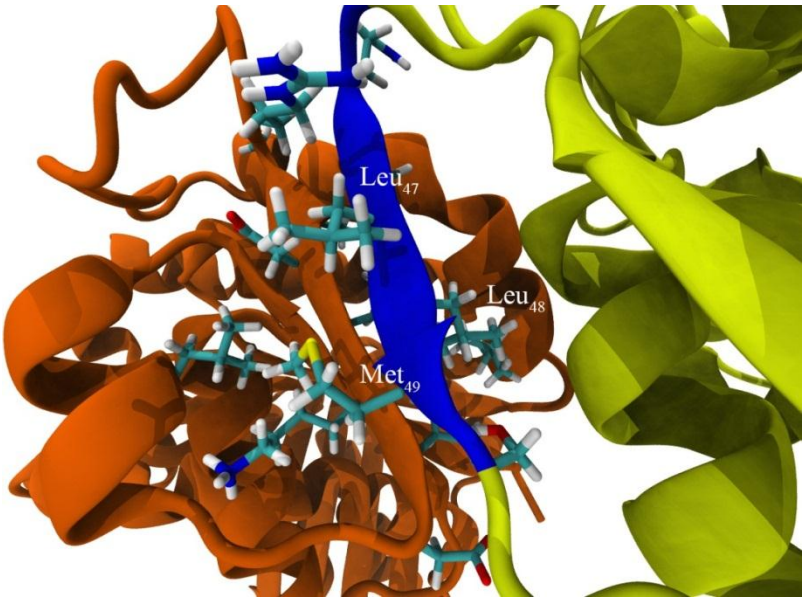


Figure 27 Detail of the interaction established by residues belonging to Group2.

.

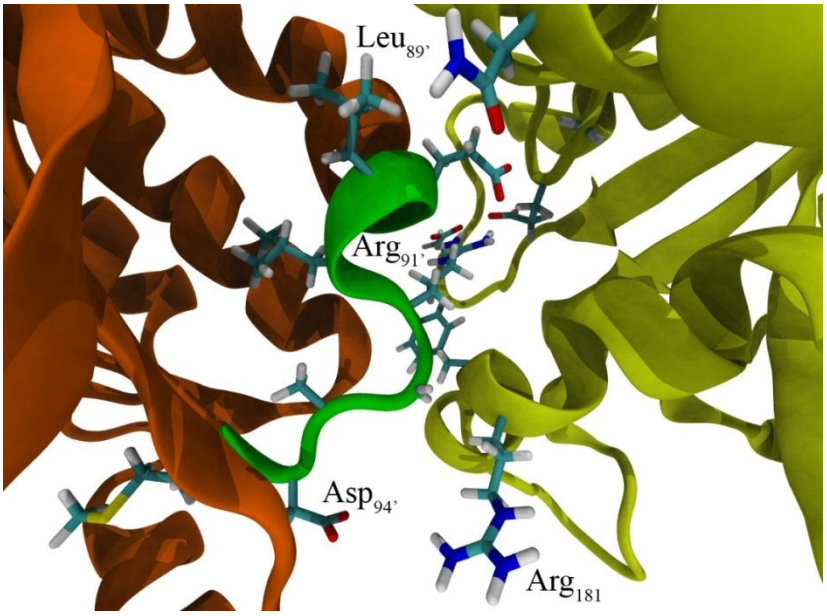


Figure 28 Close-up of the Group 3 residues at FtsZ interface.

Chapter 6

Design of peptides

As already mentioned, the aim of the work was the design of peptides able to bind FtsZ at the interface between monomers, in order to inhibit its biological activity. The "hot-stretches" found by computational alanine scanning could be used for this purpose, known their importance in FtsZ self-association. In fact, by cutting them out of the mother subunit, one could obtain, in principle, peptides able to retain their binding capability to the facing protein. Nonetheless, the extraction of the peptides from their native protein environment could affect their binding ability, due to possible structural rearrangements.

Therefore, three peptides were obtained from the three groups identified (Peptide 1, Peptide 2 and Peptide 3, according to the respective group names) and each of them underwent a further molecular dynamics simulation

in complex with the partner FtsZ subunit, in order to verify that they could still retain a good binding even when extracted from the mother protein. Peptides' termini were capped with acetyl and amine groups (at N-terminal and C-terminal, respectively), in order to avoid that their charged zwitterionic form could perturb their structure.

6.1 Peptide 1

Peptide 1 was obtained from the Group 1 located on subunit B, ranging from Ser175 to Ser182, and was simulated for 5 ns in complex with subunit A. Peptide 1 was characterized by a helical secondary structure, which was preserved throughout the whole dynamics (**Figure 29**) and which had a significant relevance in view of possible further optimizations for the development of non-peptidic mimetics [130] [131] [132].

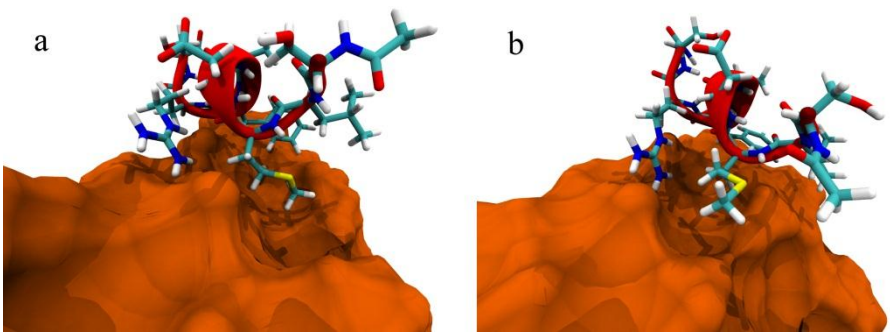


Figure 29 Starting and final snapshots extracted from the simulation of Peptide 1 in complex with FtsZ.

As a consequence, the RMSD calculated on peptide's α-carbons (**Figure 30**) did not show noteworthy rearrangements of its secondary structure throughout the simulation.

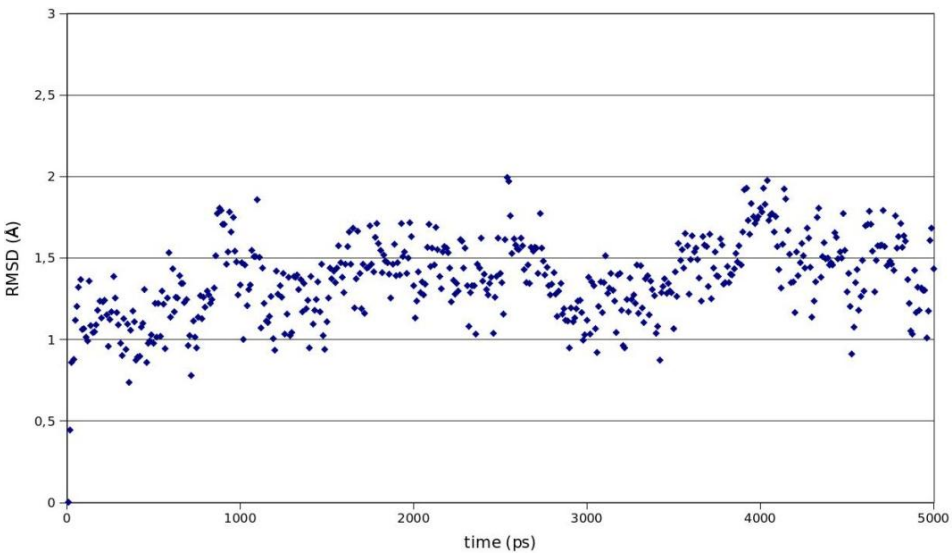


Figure 30 RMSD calculated on Peptide 1 α-carbons during its simulation in complex with FtsZ.

Also the binding with the facing FtsZ subunit is preserved for the whole dynamics, as can be observed by calculating the binding free energy between Peptide 1 and FtsZ through time. As can be observed plotting the binding free energy vs. time (**Figure 31**), the peptide retained its association with the facing protein subunit, even when extracted from its mother protein, with an average ΔG value of -18.06 ± 2.90 kcal/mol.

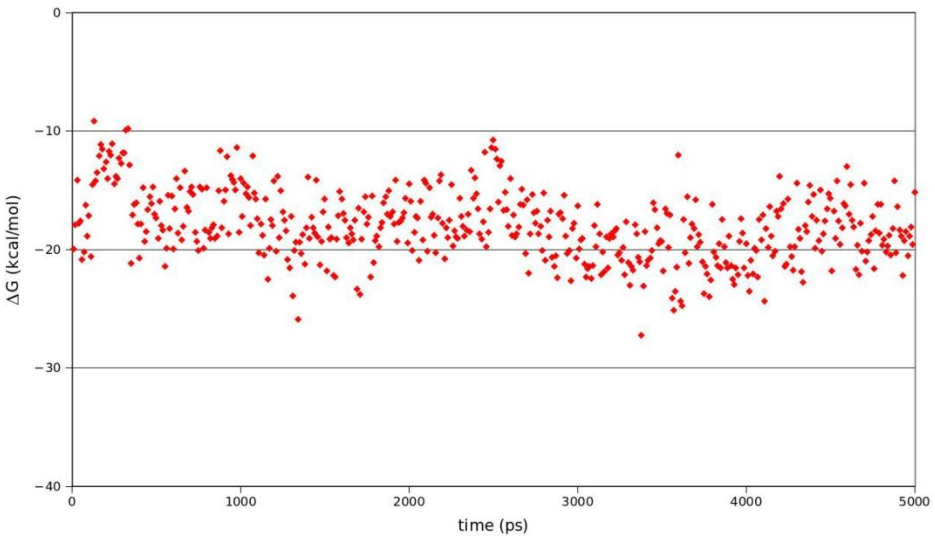


Figure 31 Plot of the ΔG between Peptide 1 and FtsZ throughout the simulation.

A computational alanine scanning was then performed on the peptide residues, in order to evaluate how the results changed with respect to the same amino acids in the dimer (**Figure 32**). It resulted that the ΔΔG values were mostly preserved in the peptide (so confirming that the peptide actually retained its binding capability), with a noticeable increase for the Arg181 value, which turned from 1.76 ± 2.56 kcal/mol in the dimer to 6.50 ± 1.83 kcal/mol in the peptide. The reason is that arginine, when at the monomer-monomer interface, predominantly orients its side chain towards the subunit B it belongs to (as already seen in **Figure 26**). On the other hand, when the peptide is extracted from the protein, arginine turns to orient more favourably towards the facing subunit, so establishing a strong salt bridge with Asp94'. Besides the

electrostatic interaction due to arginine, the association between Peptide 1 and FtsZ is also ensured by hydrophobic interactions established by Met177 and Phe180, which preserve their relevance for binding.

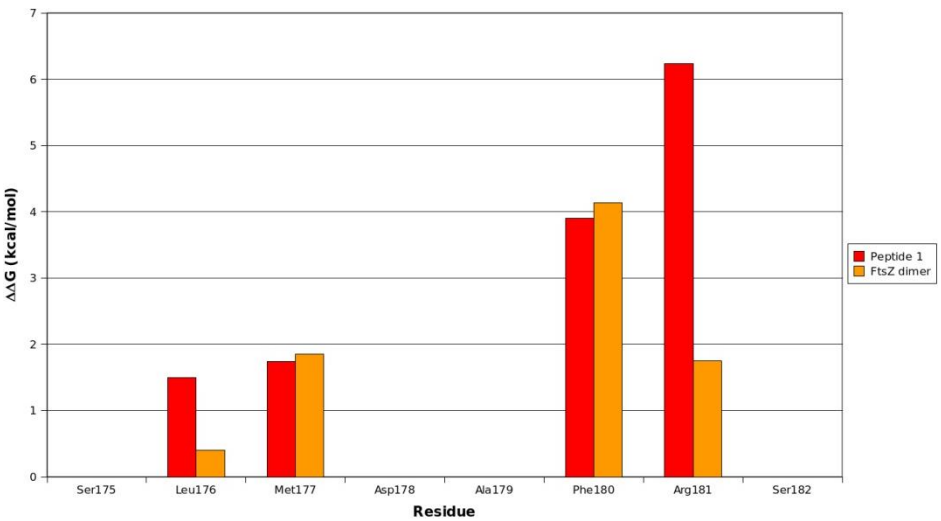


Figure 32 Results of the computational alanine scanning for the Peptide 1.

6.2 Peptide 2

Peptide 2 was obtained from the hot-stretch ranging from Gln45 to Ser50 and was simulated for 5 ns in complex with subunit A. Also in this case, the peptide did not undergo evident structural changes, with respect to the starting structure obtained from the dimer (**Figure 33**), although it slightly shifts

over the protein surface throughout the simulation. The preservation of the original peptide structure was confirmed by the results of the RMSD calculated on α-carbons throughout the simulation (**Figure 34**).

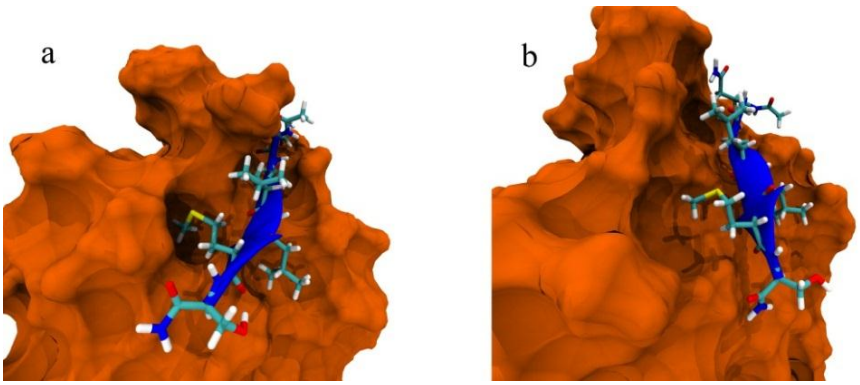


Figure 33 Starting and final snapshots of the simulation of Peptide 2 in complex with FtsZ.

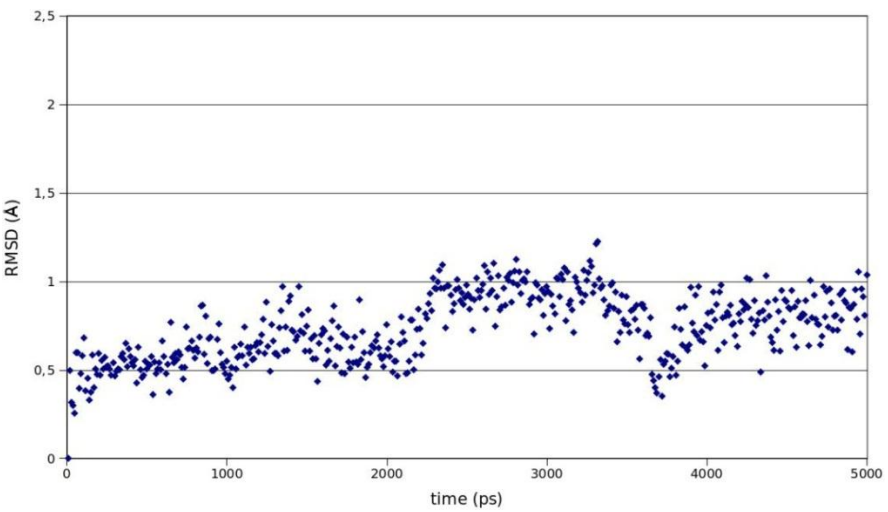


Figure 34 RMSD calculated on Peptide 2 α-carbons throughout the simulation.

Also the free energy of binding was essentially stable during the simulation (**Figure 35**), with an average value of -35.45 ± 5.44 kcal/mol.

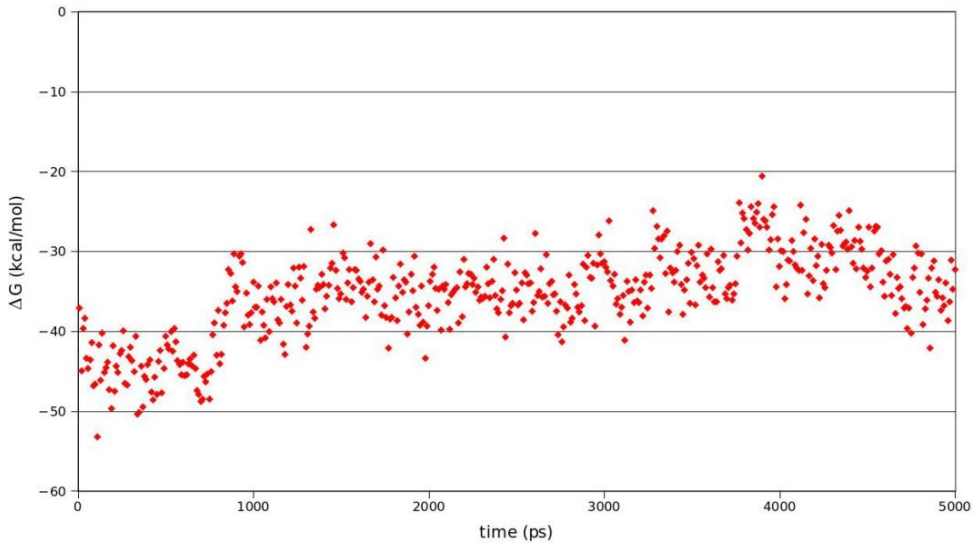


Figure 35 Plot of the binding free energy vs. time calculated for the complex Peptide 2 - FtsZ.

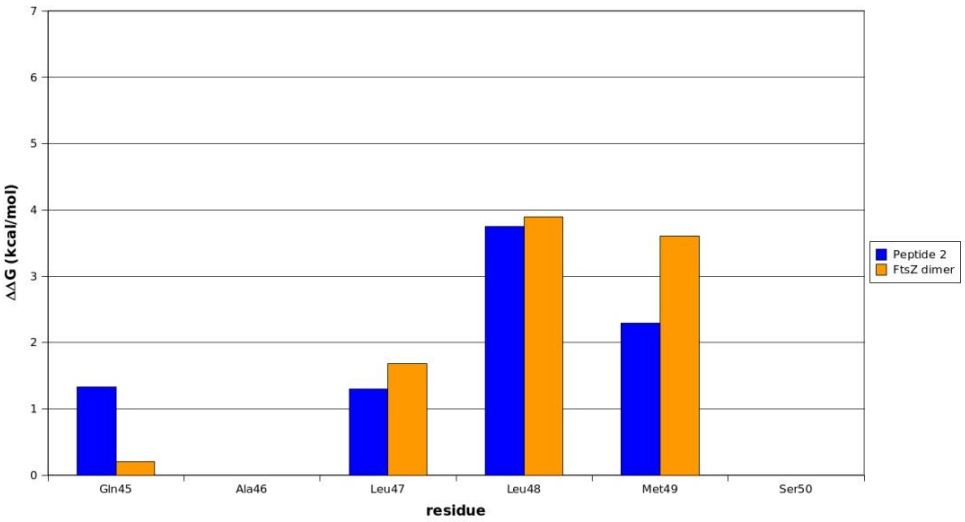


Figure 36 Results of the computational alanine scanning for Peptide 2.

As it happened for Peptide 1, the comparison of the computational alanine scanning with the results obtained in the dimer didn't show particular differences, with ΔΔG values being preserved in the peptide (**Figure 36**). The biggest discrepancy is related to the residue Met49, whose ΔΔG decreased from 3.61 ± 0.95 kcal/mol to 2.30 ± 1.12 kcal/mol. In this case, Met49 was involved both in a hydrophobic contact with Lys55' through its side chain, both in a hydrogen bond with Lys55' itself through its backbone. Both oxygen and nitrogen atoms belonging to Met49 are involved in stable hydrogen bonds with nitrogen and oxygen, respectively, belonging to Lys55' backbone. In the peptide, the hydrophobic interaction is weakened, so leading to the observed decrease in the ΔΔG value. The hydrogen bonds are instead preserved, though it does not contribute to the alanine scanning result, which only takes into account side chain interactions. The other main interactions between Peptide 2 and FtsZ include hydrophobic contacts between the two leucines and the side chains of the residues Leu56' and Asp57', as well as a weak salt bridge between Gln45 and Asn81'. Leu47 is also involved in a double hydrogen bond through its oxygen and nitrogen backbone atoms, as it happened for Met349, with the backbone of Asp57'. Also in this case, this interaction couldn't be detected by the computational alanine scanning, although it greatly contributes to the association between the peptide and the protein.

6.3 Peptide 3

Peptide 3 was designed starting from the hot-stretch ranging from Glu88′ to Met95′ and was simulated for 5 ns in complex with subunit B. Unlike Peptide 1 and Peptide 2, a qualitative analysis of the trajectory clearly showed that Peptide 3 underwent major structural rearrangements once extracted from the mother subunit (**Figure 37**).

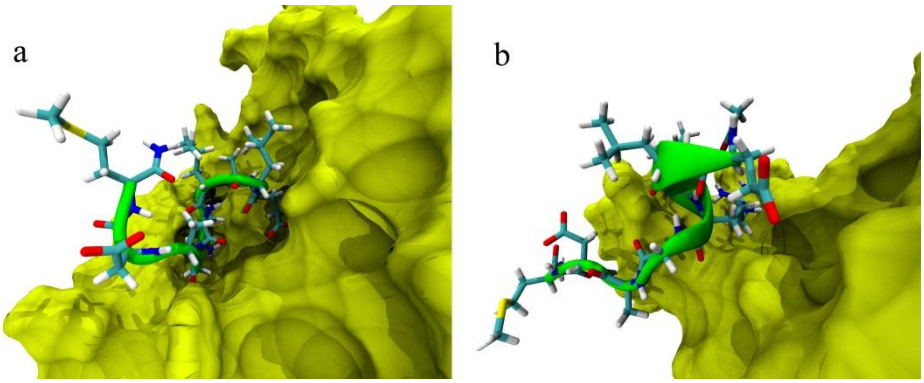


Figure 37 First and final snapshots of the simulation of Peptide 3 in complex with FtsZ.

By plotting the RMSD calculated on α-carbons vs. time (**Figure 38**), one can clearly observe that the peptide progressively lost its initial structure. As a consequence, the observed rearrangement led to a progressive weakening of peptide's interactions with FtsZ, as can be pointed out by calculating how the free energy of binding varies through the simulation time (**Figure 39**).

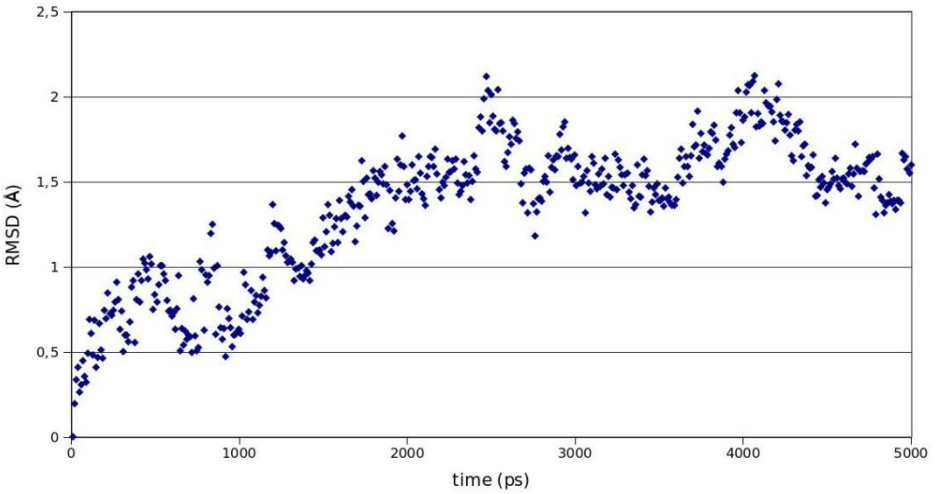


Figure 38 RMSD calculated on Peptide 3 α-carbons throughout the simulation.

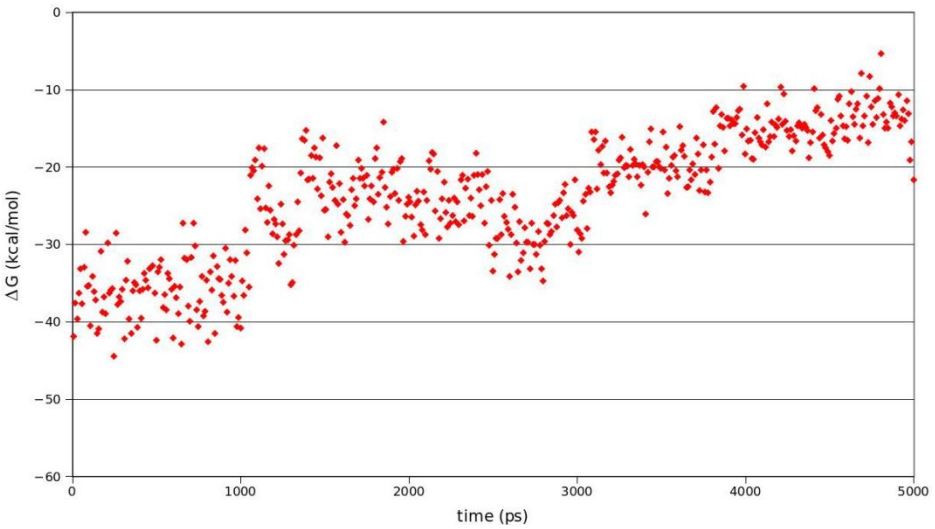


Figure 39 Binding free energy calculated on Peptide 3 - FtsZ complex during the simulation.

Whereas in the initial stages of the simulation, the ΔG value ranges around -40 kcal/mol, at the end of the trajectory it turns to ca. -15 kcal/mol, clearly showing

that only a small part of the interactions which originally made the association possible is left. In order to understand how the contribution of the single residues changed with respect to the dimer, a computational alanine scanning was performed (**Figure 40**).

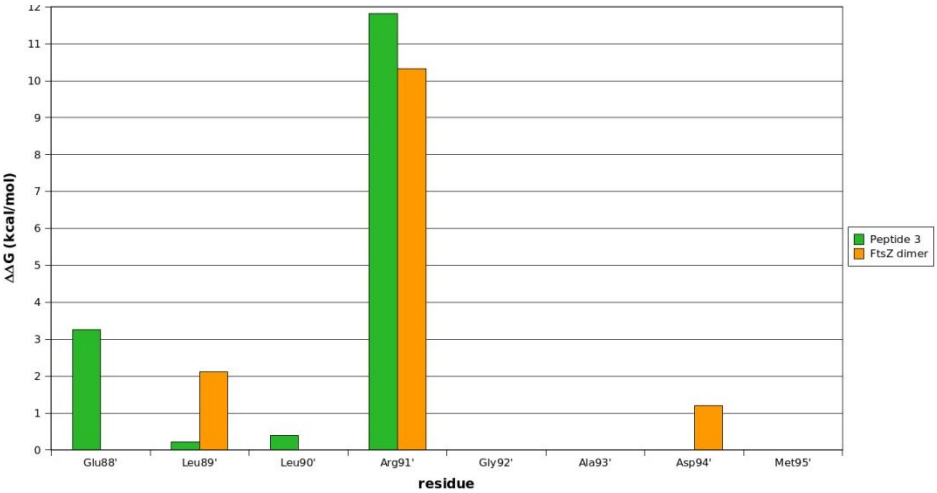


Figure 40 Results of the computational alanine scanning performed on the Peptide 3 - FtsZ complex.

The results showed that Arg91' still kept its hot-spot character, but the ΔG values of all the other amino acids turned to zero. Also Glu88', which apparently turned to hot-spot (with a ΔΔG value of 3.26 ± 4.23 kcal/mol), actually had no significant and stable contribution, if one takes into account the high standard deviation associated to its ΔΔG value.

6.4 Cyclic peptides

Based on the computational analyses, Peptide 1 and Peptide 2 therefore seemed to be the most promising ones in perspective of the design of molecules able to inhibit FtsZ-self association, since they retained the main interactions observed in the analysis of the FtsZ dimer interface. Nonetheless, the attention was subsequently focused on the optimization of Peptide 1, rather than Peptide 2. In fact, Peptide 1 peculiar helical secondary structure made it particularly interesting, since it offered an ordered and stable scaffold for the proper orientation of the side chains involved in the interaction with FtsZ. On the other hand, the linear structure of Peptide 2 suggested that it could undergo large structural rearrangements when in solution, so it was reasonably less likely to adopt the correct geometry for binding. In order to check whether Peptide 1 could preserve its helical structure even in solution, i.e. not bound to FtsZ, a 40 ns molecular dynamics simulation was performed on the peptide in explicit water, monitoring its behaviour throughout the trajectory. It turned out that Peptide 1 had a tendency to lose its helical structure in solvent, as can be pointed out by observing the first and the last snapshot of the simulation (**Figure 41**). As a consequence, the side chains of the warm- and hot-spots lost the proper orientation for a correct binding of Peptide 1 to its putative binding site.

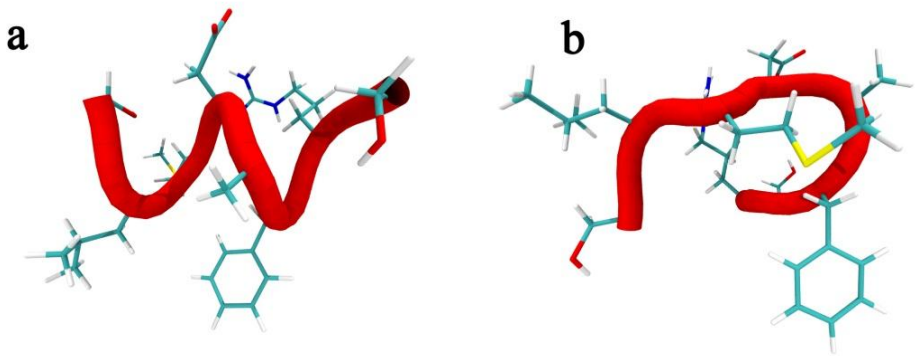


Figure 41 Starting (a) and final (b) snapshots of the simulation performed on Peptide 1 in solution. The secondary structure is clearly lost throughout the simulation.

The attention was thus focused on the identification of a way to constrain Peptide 1 in its native helical structure, so to make it retain the proper conformation for binding. By analysing the results of the simulation performed on Peptide 1 in complex with FtsZ, it was found out that Asp178, belonging to the peptide, could establish a hydrogen bond with Ser182 or, alternatively, by differently orienting its side-chain, with Ser175. These residues did not interact with the facing FtsZ subunit neither in the dimer nor in the protein-peptide complex and could lead to the formation of a virtual ring involving all the amino acids between Asp178 and Ser182 in one case and between Ser175 and Asp178 in the other case, giving a hint on how to constrain the helical structure with simple chemical modifications. In fact, a ring was obtained by mutating *in silico* Ser182 into an ornithine and cross-linking its side-chain with that of Asp178 through an amide bond, so obtaining a cyclic peptide, named Peptide 1a (**Figure 42b**).

In the same way, Ser175 was mutated into ornithine and cross-linked to Asp178, so obtaining a further cyclic peptide called Peptide 1b (**Figure 42c**).

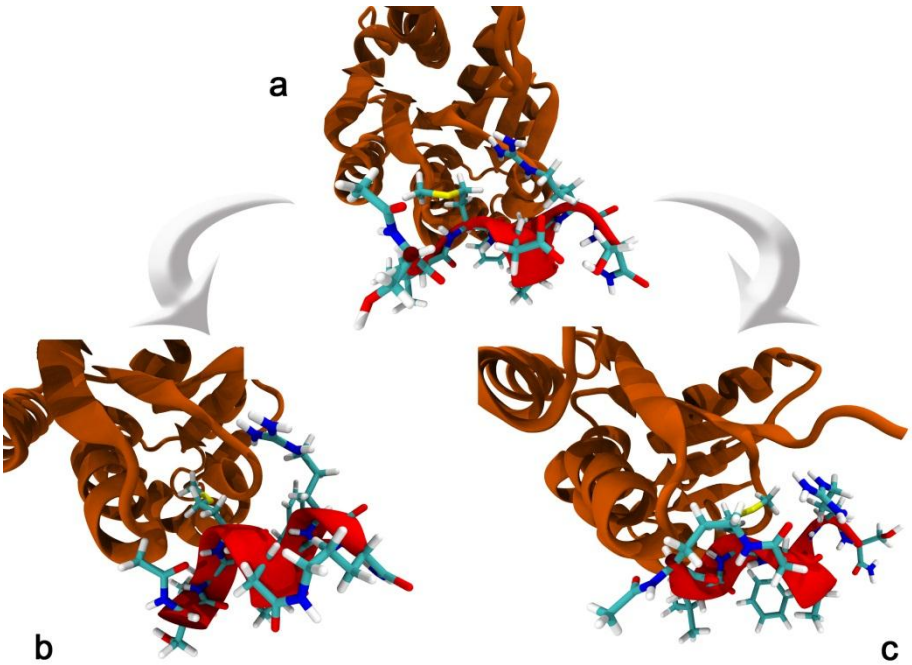


Figure 42 Peptide 1a (panel b) and Peptide 1b (panel c) were obtained from Peptide 1 (panel a) by mutating *in silico* Ser182 or Ser175, respectively, into an ornithine and cross-linking them to Asp178.

The design of cyclic peptides seemed a good approach to optimize and constrain the secondary structure of Peptide 1. In fact, cyclization is typically the first step in the optimisation of a peptide because it allows to restrain its conformational freedom, thus increasing the population of the active conformations, which exhibit the greater binding affinity with the target, and can moreover increase binding affinity reducing the entropic cost of peptide binding.

The so-obtained peptides were simulated for 40 ns in solution, to evaluate if the cyclization actually improved the native helical structure stability. For both peptides, it turned out that the helical conformation was highly stabilized by the proposed modification, as can be pointed out by observing the final structures of the simulations (**Figure 43** and **Figure 44**), with a great improvement with respect to the simulation of Peptide 1, which underwent major structural rearrangements.

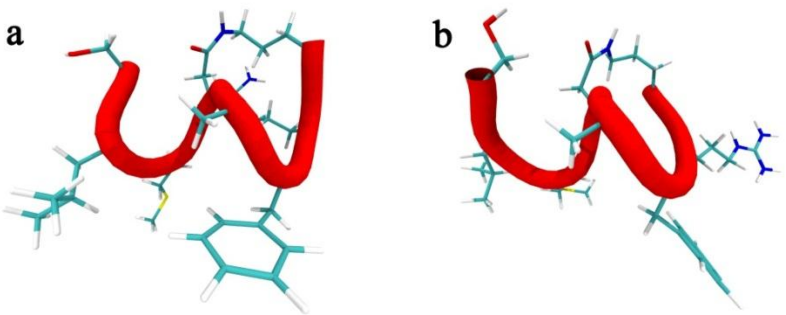


Figure 43 First and final snapshots of the simulation of Pepide 1a in solution.

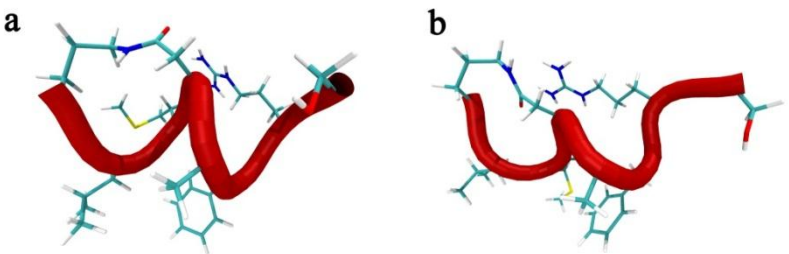


Figure 44 First and final snapshots of the simulation of Pepide 1b in solution.

The calculation of the RMSD on α-carbons calculated for both the linear and the cyclic derived peptides clearly confirmed that the backbone is more stabilized for Peptide 1a and Peptide 1b (**Figure 45**). Given the good results obtained on the peptides free in solvent, Peptide 1a and Peptide 1b were later on simulated for 5 ns in complex with FtsZ, positioning them in the same binding site of Peptide 1, in order to evaluate whether the modifications, though stabilizing the peptides in solution, could affect their binding to the protein. It turned out that the modifications did not weaken the interactions of the peptides to FtsZ, but, on the contrary, the binding resulted to be even slightly improved, with respect to the original Peptide 1 (**Figure 46**). The average ΔG of binding was -26.65 ± 3.19 kcal/mol for Peptide 1b and -22.42 ± 5.44 kcal/mol for Peptide 1a.

A computational alanine scanning was then performed for both the cyclic peptides, confirming that the contribution of the single residues to the binding is on the whole not affected, with respect to Peptide 1 (**Figure 47**). It is important to point out that Met177 in both cases had its $\Delta\Delta G$ value noticeably increased, since in the cyclic peptides it can orient more properly towards the facing residues to establish stronger hydrophobic interactions which contribute to the stabilization of the complex. The stabilized helical structures thus provided the right scaffold for the orientation of the "hot-triad" composed of Met177, Phe180 and Arg181 towards their partner of interactions (**Figure 48**). Peptide 1a and Peptide 1b thus resulted to have the optimal features with regards to both the secondary structure stability in solution and the binding to the FtsZ-FtsZ interface.

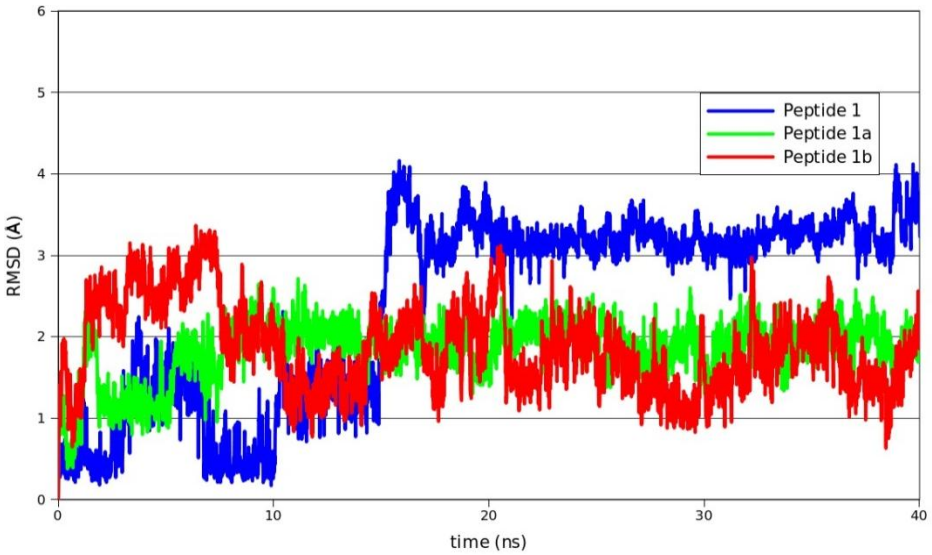


Figure 45 RMSD calculated on α-carbons of Peptide 1, Peptide 1a and Peptide 1b in solution.

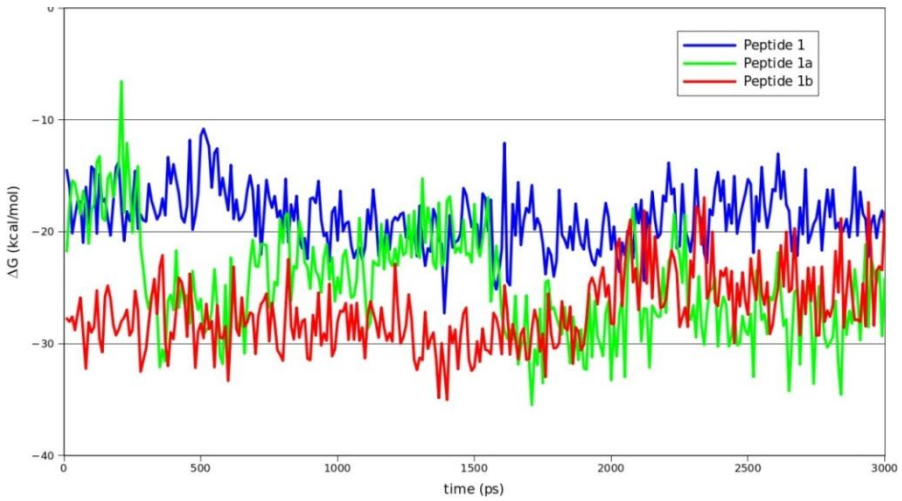


Figure 46 Binding free energy calculated for the simulations of Peptide 1, Peptide 1a and Peptide 1b in complex with FtsZ.

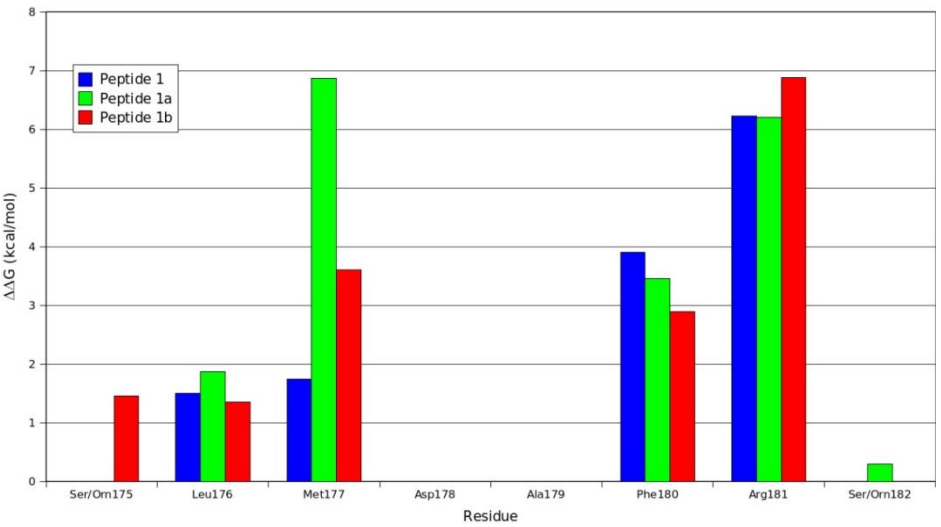


Figure 47 Comparison between the computational alanine scanning performed on the cyclic peptides and Peptide 1.

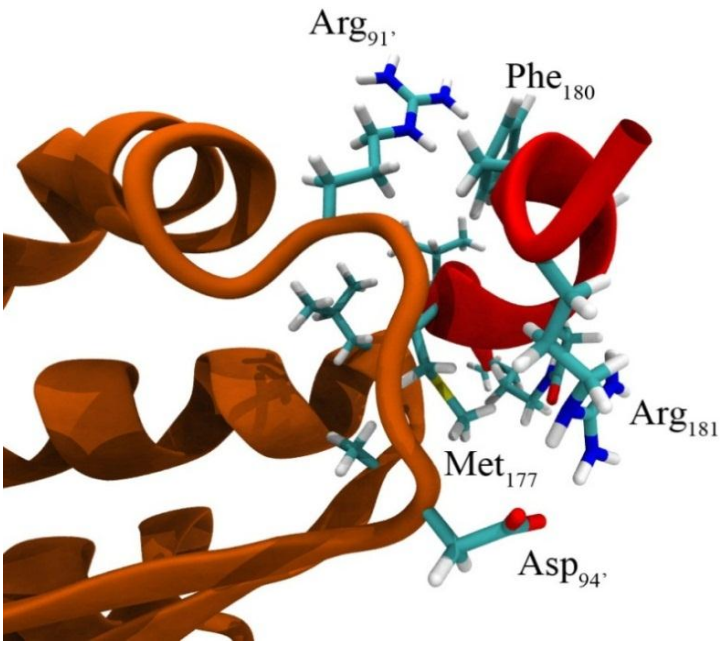


Figure 48 The cyclization allowed the stable and efficient orientation of the key hot-spots towards their partner of interactions on the facing FtsZ subunit.

Chapter 7

Biological assays

In order to confirm the computational results, Peptide 1a and Peptide 1b were synthesized and underwent biological assays to evaluate their activity over FtsZ polymerization *in vitro*. Peptide 1 was also taken into account for comparison. The synthesis of the peptides was performed by Prof. Giovanna Speranza and Dr. Pierangelo Francescato of the Department of Organic and Industrial Chemistry of University of Milan. As to the biological tests, the analyses were carried out by Prof. Jayaraman, Dr. Chacko Jobichen and Dr. Prerna Domadia of the University of Singapore.

7.1 GTPase activity tests

As mentioned in Chapter 3, FtsZ polymerization is strictly GTP dependent and its dynamics is closely linked to the rate of GTPase hydrolysis. Therefore, monitoring the amount of inorganic phosphate (Pi) released upon treatment of FtsZ with peptides can provide a direct monitoring of the rate and extent of inhibition of FtsZ polymerization.

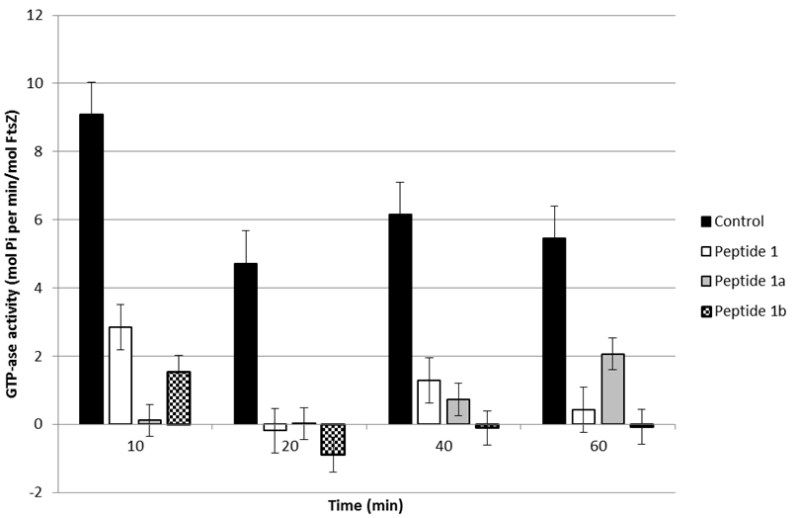


Figure 49 Time course of GTPase activity with or without addition of peptides. Error bars represent the standard error of the mean.

As can be seen in **Figure 49**, upon induction of GTP hydrolysis in control FtsZ, Pi molecules released were seen to initially increase up to 10 min. Thereafter, a steady-state phase up to 60 min was attained as hydrolysis rate plateau due to GTP consumption. GTP exhaustion and disassembly of FtsZ polymers was not observed up to 60 min in assay run in the absence of

peptides. However, upon treatment of FtsZ with both Peptide 1 and the cyclic peptides, a strong perturbation of the GTPase activity of FtsZ assembly was observed. The linear peptide strongly inhibited the GTP hydrolysis from the values of 4.72 ± 0.95 min^{-1} in the absence of the peptide to -0.19 ± 0.42 min^{-1}. Peptide 1a significantly abolished the initial rate of GTP hydrolysis of FtsZ assembly from the value of 9.09 ± 0.72 min^{-1}, calculated in the absence of peptide, to the value of -0.15 ± 0.15 min^{-1} in the presence of peptide. Peptide 1b totally suppressed the rate of GTP hydrolysis by FtsZ assembly from the values of 4.72 ± 0.95 min^{-1}, 6.15 ± 2.0 min^{-1} and 5.45 ± 1.75 min^{-1} in the absence of peptide to -0.89 ± 0.75 min^{-1}, -0.11 ± 0.72 min^{-1} and -0.08 ± 0.59 min^{-1} at 20 min, 40 min and 60 min, respectively. Concomitantly, at 5 μM, all three peptides remarkably inhibited the extent of GTP hydrolysis of FtsZ by 100 %, showing that both the linear and the cyclic peptides were able to impede the FtsZ dynamics.

7.2 Transmission Electron Microscopy

Further analyses were carried out to explore the effect of the peptides on the FtsZ assembly and identify the related changes in polymer morphology. Transmission electron microscopy (TEM) imaging of *E.coli* FtsZ was thus performed. *E.coli* FtsZ protein was incubated with or without peptide and

visualized for EM imaging (**Figure 50**). Control FtsZ readily assembled into abundant, thick, dense polymeric network of protofilaments (**Figure 50a**).

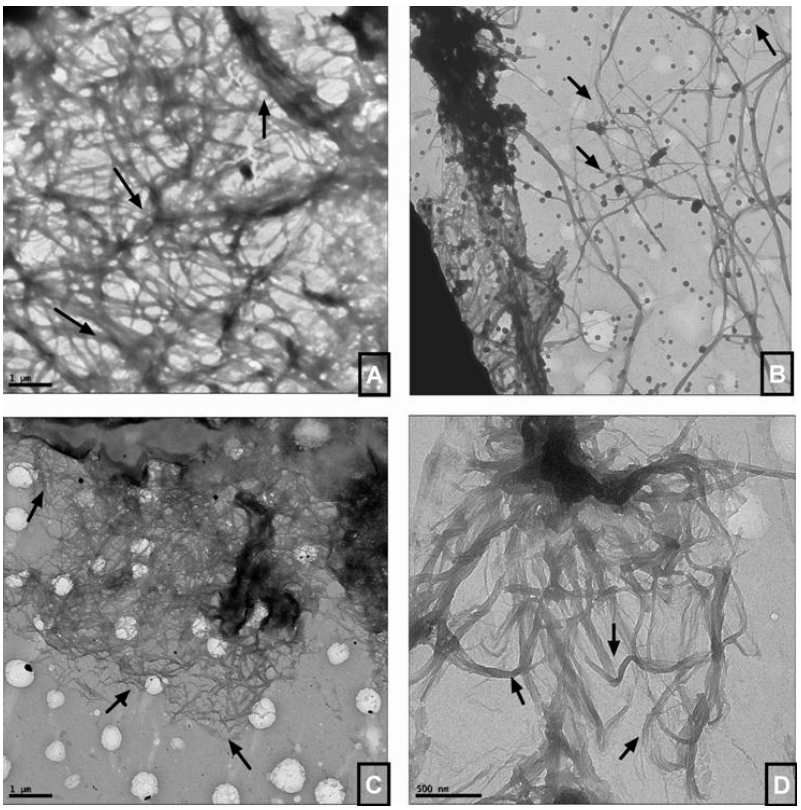


Figure 50 Visualization of FtsZ assembly with or without addition of peptides by TEM. Electron micrographs of negatively stained FtsZ polymers were observed under a transmission electron microscope at ×50000 magnification, shown in the absence (A) and presence of 5 μM linear peptide 1 (B) 5 μM cyclic peptide 1b (C) and 5 μM cyclic peptide 1a (D).

However, 5 μM of all three peptides could readily prevent the formation of thick, dense filamentous crowding. The linear Peptide 1, after 5 min showed very few, less dense, isolated, thin and curved FtsZ polymers. This

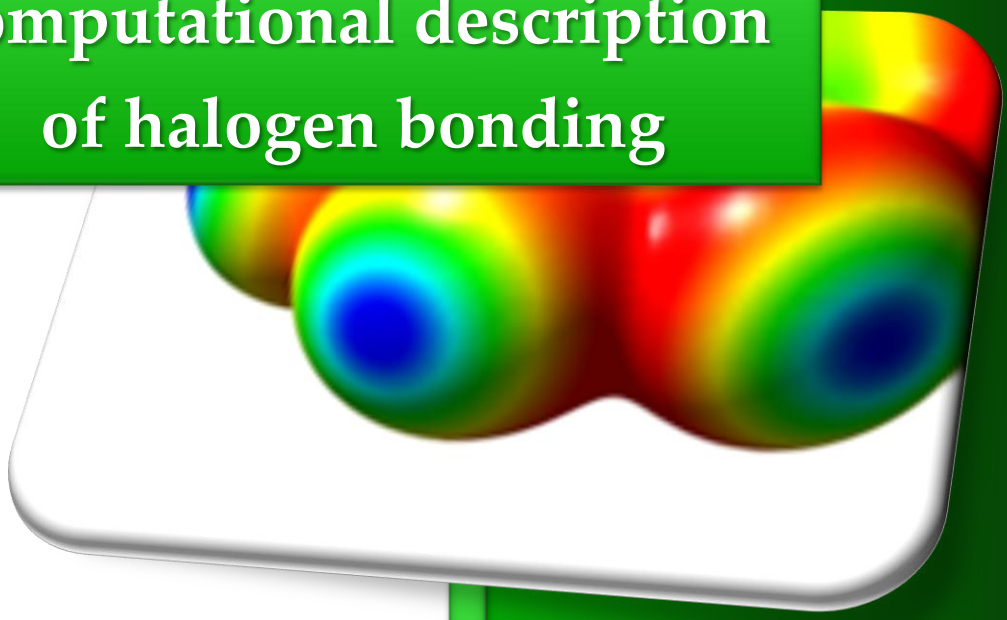
peptide showed substantial reduction in the degree of forming protofilaments (**Figure 50b**). However, in the presence of Peptide 1b, polymer morphology was clearly perturbed. It resulted into formation of less frequent, remarkably thin and curved, shorter FtsZ (**Figure 50c**). Peptide 1a could not form FtsZ higher order assemblies such as bundles or extensive network of protofilaments after 5 min. Few remains of very thin and markedly short, single stranded, strikingly curved polymers were searched and selectively imaged using a bar scale of 500 nm (**Figure 50d**) instead of scale bar of 1 μm as used for the other images.

Part II

Conclusions

As in the study performed in the Part I, a computational approach allowed a detailed insight of the network of interactions occurring at the interface between proteins. Nonetheless, in the case of the FtsZ dimer, the study was not only limited to a deeper understanding of the mechanism of protein self-association, but the gathered data were used for the design of effective inhibitor agents. In fact, the analysis allowed the design of a set of peptides, starting from the amino acidic sequence of contiguous hot-spots highlighted at the interface, able to bind FtsZ and modulate its polymerization. In particular, Peptide 1 was the subject of further optimizations aimed at the improvement of its binding ability to FtsZ. Peptides effectiveness was verified by biological assays *in vitro*, so proving the reliability of the computational approach in the study of protein-protein interactions.

Part III
Computational description of halogen bonding



Chapter 8

Halogen bonds

8.1 Halogen bond definition

Halogen bonds (**Figure 51**) are particular interactions established by halogen atoms which present similarities with hydrogen bonds [133] [134] [135] [136] [137]. While the hydrogen bond can be schematically represented as D–H···A, where D and A denote respectively the donor and the acceptor group (behaving as a Lewis base), the halogen bond can be represented as D–X···A where X is an halogen atom, with the force of the interaction growing up with the dimensions of the halogen (F << Cl < Br < I).

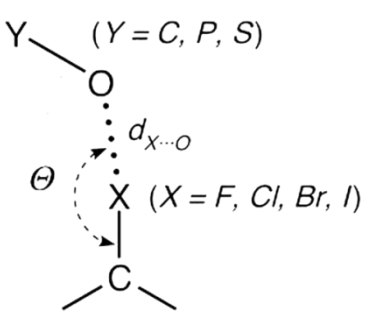


Figure 51 Schematic representation of a halogen bond. The angle Θ is found to be close to 180°, whereas the distance between the halogen and the acceptor group (an oxygen atom in the example shown) is below the sum of the van der Waals radii of the atoms involved.

It is well known that hydrogen bonds are of fundamental importance in determining the three-dimensional structure of biomolecules and in molecular recognition processes. For example, α-helices and β-sheets motifs in proteins, and helices in DNA and RNA are due to a well-defined network of hydrogen bonds. The formation of ligand-protein complexes is also frequently ascribable to the possibility to establish hydrogen bonds between the two partners of the complex. More recently, the importance of halogen bonding in biological systems has been as well evidenced [138] [139] [140] , due in part to the increasing number of resolved protein structures with halogenated ligands. Voth *et al.* have also demonstrated that halogen bonds can compete with hydrogen bonds in stabilizing DNA junctions using brominated uracil [141]. It should also be noted that a large number of currently used drugs are halogen-substituted. Halogen atoms are many times introduced in a drug to increase its half-life, so as to prolong its activity in the organism, or to facilitate the membrane permeability so as to permit the drug to reach its biological target.

The growing attention towards halogen bonding and a correct interpretation of its nature are now being aimed at increasing the activity of the ligands. For example, a systematic study of the importance of halogen bonding in novel inhibitors of human Cathepsin L has been recently reported [142], showing that it was possible to increase the activity of a ligand of almost two orders of magnitude, by substitution of a hydrogen with an iodine atom.

Besides the biological aspects, halogen bonding was also exploited by several research groups as a tool to direct intermolecular recognition processes

with applications in crystal engineering [143], in the development of solid state materials with electronic properties, ranging from organic semiconductors to superconductors [144] [145] [146] and also in the development of new materials with non-linear optical properties [147].

8.2 Electrostatic potential anisotropy around halogen atoms

A simple interpretation for the formation of halogen bonds was given by Politzer *et al.* [148] [149]. The electrostatic potential of any neutral spherical atom is always positive at any distance from the nucleus, and this can be seen as the predominance of the concentrated nuclear charge over the dispersed electron cloud. Taking into account the valence electronic structure of a halogen atom covalently bonded to another atom along the z-axis, sketched as $s^2p_x^2p_y^2p_z^1$, it is found that, with respect to the spherical isolated halogen atom (with a 5/3 mean population of p electrons in each direction), there is a depletion of electronic charge along the bond axis, in particular in the region outwards the covalent bond. Such a depletion is compensated by an increase of electronic charge in the directions perpendicular to the bond axis. Based on this interpretation, Politzer was able to explain the strong anisotropy of the electrostatic potential around a covalently bonded halogen atom, as obtained by

ab initio methods. A region of positive electrostatic potential, called σ-hole (**Figure 52**), is in fact built up on the halogen in the region outwards the covalent bond and a crown of negative electrostatic potential is built up around the direction of the D–X bond.

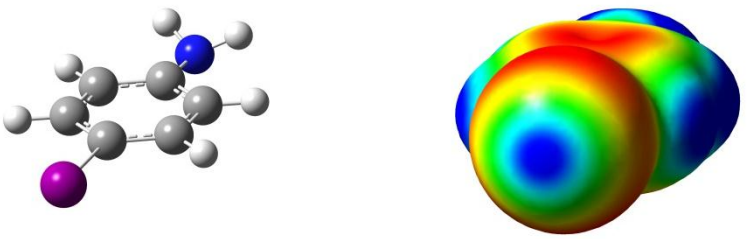


Figure 52 In halogenated ligands, a positive cap (σ- hole) on the electrostatic potential is found around the halogen atoms, outwards the C-X bond.

Halogen atoms are therefore able to accept electron density from a Lewis base located along the extension of the D–X bond, giving rise to halogen bonding. The size of the positive electrostatic area is found to depend on the electron withdrawing capabilities of the D group and on the nature of halogen atoms (this feature will be discussed in more detail in Chapter 10). The extension of the σ-hole is in fact more pronounced for the iodine atom, owing to its larger polarizability, decreases for bromine and chlorine atoms, and then disappears for the fluorine atom, except in the presence of strongly electronegative D groups. Moreover, a positive correlation has been reported between the value of the electrostatic potential on the σ-hole and the strength of the halogen bond [148] [149] [150].

8.3 Halogen bonds in molecular mechanics

A rigorous theoretical description of halogen bonding can be obtained using quantum mechanical methods like MP2 or DFT. However, these methods cannot be applied to the study of very large molecules such as halogen-containing biological systems. In these cases, it is mandatory to use the so-called hybrid methods, based on the partition of the entire system into a small region which contains the halogen bonds and that can be described by quantum mechanics methods (QM), and the remaining atoms which constitute the majority of the system, that can be described by molecular mechanics methods (MM). For example, QM/MM methods in the ONIOM [151] [152] [153] implementation have been recently successfully applied to the description of different protein-halogenated ligand complexes whose structure had been determined by X-Ray methods [154].

In order to describe the dynamical behaviour of such complexes and to evaluate different docking poses, it would be desirable to have the possibility to model the full system by resorting to only molecular mechanics methods. Up to now, the problem to describe the electrostatic potential around the molecule in the framework of most MM methods used to describe large molecules of biological interest has been solved by using single point charges cantered on atomic positions. Such an approach, however, does not allow to describe the anisotropy of the potential around the halogen atom. More sophisticated methods based on multipolar approaches could overcome the problem, but

they are evidently more demanding from a computational point of view. The development of simpler approaches then appears to be highly desirable.

In this part of the work, a simple strategy to correctly describe, by MM methods, halogen bonding in biological systems was searched. Attention was focused on interactions where iodine or bromine atoms are involved, as they give rise to the stronger halogen bonds. The approach consisted in introducing a pseudo-atom located along the D–X bond, outwards the bond itself, and to use this extra point to correctly fit the electrostatic potential around the molecule (**Figure 53**).

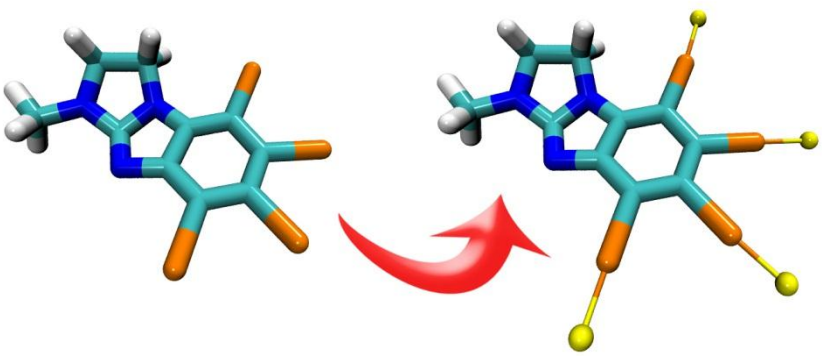


Figure 53 Using the proposed approach to describe the σ-hole in molecular mechanics, a point charge (represented in yellow) is added on the extension of the carbon-halogen bond, in order to better fit the electrostatic potential in that particular region.

It is to be noted that the introduction of extra-points was already used in the force field methods to describe anisotropies, for example those due to the presence of lone pairs of electrons [155] [156]. Following strategies already implemented for taking into account such 'extra-points' [155] [156], the pseudo-

atom can be linked with an harmonic potential to the halogen atom. In this way, it could be possible to reproduce the σ-hole and the full anisotropy of the potential around the halogen atoms, that was pointed out by Politzer to be responsible of the formation of halogen bonding [148] [149] .Therefore, by simply introducing a little variation in the traditional force field, halogen bonds in biomolecules could be adequately reproduced by molecular dynamics simulations. In order to validate the results obtained through this approach, a comparison with both crystallographic structures and QM/MM optimization had to be performed.

Chapter 9

Describing halogen bonds in molecular dynamics

9.1 Pseudo atom parametrization

As mentioned before, halogen bonds are mainly due to an electrostatic interaction between the positive σ-hole, i.e., a region of positive electrostatic potential located on the halogen atom X in the direction of the D–X bond, and an acceptor atom acting as a Lewis base. This interaction brings the distances between the halogen and the acceptor atoms below the sum of their van der Waals radii. The electrostatic interaction in a force field is usually described attributing a point charge on each atom. While charges for the atoms of each amino acid are provided by the developers of the force field, charges of the ligands can be determined by means of a fitting of the electrostatic potential determined at the quantum mechanical level, following the so-called Restrained ElectroStatic Potential (RESP) procedure [157].

In **Figure 54**, the electrostatic potential obtained at different levels for halogenated ligands, e.g. K44 and I3C, respectively a bromo- and an iodo-derivative (panels a and e; see **Figure 55** for the chemical structure of the ligands) is shown. The QM electrostatic potential of the two ligands (panels b and f, respectively) clearly showed a strong anisotropy around each halogen atom. The σ-hole was clearly visible and appeared to be more pronounced for the iodine with respect to the bromine atom, in agreement with the above-mentioned trend observed for the strength of halogen bonding. A fitting of the electrostatic potential by the RESP procedure simply using a single point charge on each halogen atom was carried out, and the electrostatic potential was recomputed using the set of derived RESP charges (panels c and g).

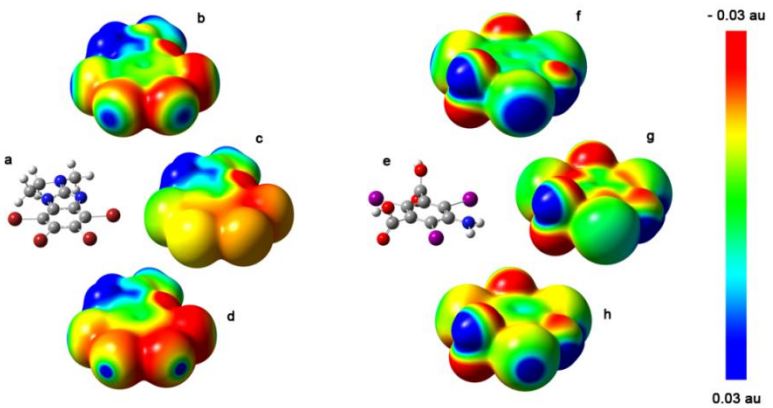


Figure 54 Electrostatic potential mapped onto the surface of electron density (0.002 electrons/au^3) of ligands K44 (a) and I3C (e), calculated at the QM level (b, f) or using the fitting by the RESP procedure without pseudo-atom (c, g) and after pseudo-atom addition (d, h).

This is the electrostatic potential that would be traditionally used in a molecular dynamics simulation. It is evident that the σ-hole was not reproduced and that the electrostatic potential is poorly described around the halogen atoms.

A pseudo-atom was thus added along the D–X bond located at a fixed distance from the halogen atom, to better fit the electrostatic potential around the halogen atoms and correctly describe anisotropic features [158]. The RESP procedure was then applied to determine the electrostatic charges also on the pseudo-atoms. The electrostatic potential computed from the new set of charges (panels d and h) was noticeably improved and showed the correct anisotropy, with a σ-hole well evident on the halogen atoms. The addition of a pseudo-atom thus seemed to be a simple though efficient method to improve the description of the electrostatic potential and in particular to introduce those features which are responsible for the halogen bond formation The full set of parameters used to describe the pseudo- atoms are reported in **Table 2**. Their distances from the halogen atoms were chosen to give the best fit with the QM electrostatic potential.

9.2 Molecular dynamics simulations

Proven the effectiveness of the pseudo-atom approach in correctly describing the electrostatic potential for halogenated molecules, in order to investigate their role in describing halogen bonding arising in ligands bonded to a protein, a number of molecular dynamics simulations was performed on a set of halogenated ligand-protein complexes.

Table 2 Pseudo-atom parameters used for the MM force field.

Parameter[a]	Value
Mass	3.0 amu
R_{eq} (X-PA)[b]	1.4 Å – 1.8 Å
K_r (X-PA)[c]	600.0 kcal mol^{-1} Å^{-2}
Θ_{eq} (C-X-PA)[c]	180.0°
K^{Θ} (C-X-PA)[c]	150.0 kcal mol^{-1} Å^{-2}
r*[c]	0.0 Å
ε[c]	0.0 kcal mol^{-1}

[a]X = bromine or iodine atom, PA = pseudo-atom.
[b]The equilibrium distance between halogen and pseudo-atom, R_{eq}, varies according to the best fitting with the electrostatic potential calculated at the QM level: 1.4 Å for the molecule BML, 1.6 Å for I3C and BRT, 1.7 Å for T44, T3 and 4HY and 1.8 Å for TBB and K44.
[c]The bond force constant, K_r, the equilibrium angle, Θ_{eq}, and the angle force constant, K^{Θ}, are kept constant regardless of the ligand. The Van der Waals parameters of the pseudo-atom (r* , ε) are set to zero.

The structures were retrieved from a recently published work on the importance of halogen bonding in the field of rational drug design [154] in which the authors analysed the entire Protein Data Bank (December 2008 release) and collected a total of 248 structures where the distances between halogens and oxygen atoms are less than the sum of their respective van der Waals radii (which amount to 3.27 Å, 3.37 Å, 3.50 Å respectively for Cl···O, Br···O and I···O contacts) and with a D–X···O angle greater than 120°. It is to be noted that halogen bonds where the acceptor Y is an oxygen atom are particularly important in biochemistry, as they represent one of the most

common halogen bonds in protein-ligand interactions. Usually the oxygen involved in such interactions belongs to the protein backbone amidic group.

From this set, four X-ray structures containing bromine atoms (PDB IDs: 1P5E [159], 1GXZ [160], 1T0S [161], 1ZOH [162]) were selected, two of which (1P5E and 1ZOH) established multiple halogen bonds. Five structures containing iodine atoms (PDB IDs: 1Y0X [163], 2H77 [164], 3E79 [165], 3E3T [166], 1NQ2 [167]) were also chosen. The selected structures contained halogenated ligands (**Figure 55**) in complex with particularly important proteins, like, for example, thyroid hormone receptors, kinase and cell division kinase proteins.

For all these structures, molecular dynamics simulations were performed where the ligands were described with the addition of a pseudo-atom for each halogen atom. For comparison, the same simulations were repeated without introducing the pseudo-atom. The mean values of the geometrical parameters of the halogen bonds are reported in **Table 3**.

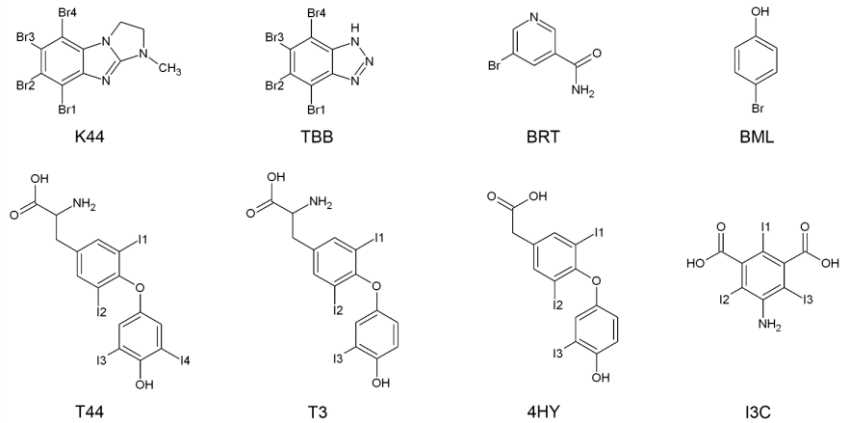


Figure 55 Chemical structures of the ligands considered in the molecular dynamics simulations.

Table 3 Geometrical parameters of C–X···O halogen bonds in protein–ligand complexes from molecular dynamics simulations in the absence and in the presence of pseudo-atom on the halogen.

			no pseudo-atom		pseudo-atom	
PDB Id	Ligand	Halogen bond[a]	*d*(X···O) (Å)	∠C–X···O (°)	*d*(X···O) (Å)	∠C–X···O (°)
1P5E	TBB	Br1 ··· Phe80[b]	4.18 ± 0.49	142.9 ± 15.5	3.64 ± 0.32	148.7 ± 15.7
		Br2 ··· Glu81O	4.11 ± 0.47	139.1 ± 14.3	3.54 ± 0.50	162.5 ± 11.5
		Br3 ··· Leu83O	4.24 ± 0.91	129.7 ± 16.9	3.58 ± 0.48	150.9 ± 15.7
1GXZ	BRT	Br ··· Gln187O	3.86 ± 0.36	159.2 ± 7.5	3.43 ± 0.29	163.8 ± 7.4
1T0S	BML	Br ··· Glu231O	3.79 ± 0.64	150.6 ±10.0	3.33 ± 0.22	170.0 ± 6.0
1ZOH	K44	Br3 ···Glu114O	5.31 ± 1.21	138.6 ±19.0	3.99 ± 0.45	142.7 ± 20.5
		Br4 ···Val116O	4.43 ± 0.91	153.2 ±19.8	3.51 ± 0.35	158.0 ± 5.2
IY0X	T44	I2 ···Phe272O	3.62 ± 0.25	154.2 ± 8.6	3.48 ± 0.34	165.4 ± 7.9
2H77	T3	I1 ···Phe218O	3.64 ± 0.26	162.7 ± 8.5	3.43 ± 0.24	170.5 ± 5.4
3E79	I3C	I3 ··· Ala334O	3.79 ± 0.32	117.7 ± 12.9	3.33 ± 0.25	124.3 ± 26.4
3E3T	I3C	I3 ··· Leu141O	12.18 ±2.96	32.7 ± 18.2	3.74 ± 0.62	151.5 ±16.9
1NQ2	4HY	I1 ··· Phe272O	3.86 ± 0.29	150.1 ±7.8	3.46 ± 0.30	163.8 ± 7.8

Molecular dynamics details

The ligand-protein complexes structures were obtained from the Protein Data Bank (PDB codes: 1P5E [159], 1GXZ [160], 1T0S [161], 1ZOH [162], 1Y0X [163], 2H77 [164], 3E79 [165], 3E3T [166], 1NQ2 [167]). In the case of 1P5E, where two almost identical protein complexes were present in the crystal structure, the A chain was removed and only the B chain has been considered in calculations.The partial atomic charges for the ligands were derived using the Restrained ElectroStatic Potential (RESP) method [157], as implemented in AMBER11 [168]. For each halogen atom X, a pseudo-atom has been added along the direction of the D–X bond, in the region outward the bond itself. Its distance from the halogen has been determined to give the best fit with the quantum mechanical electrostatic potential. The parameters used for the pseudo-atom are reported in **Table 2**. Atomic charges after pseudo-atom addition were recalculated using the RESP module.

Molecular dynamics simulations were performed with the AMBER11 package on all complexes both in the native form and after pseudo-atom insertion. The dynamics were run with periodic boundary conditions using the 2003 force field for the protein and the Generalized Amber Force Field (GAFF) for the ligands. Missing parameters were calculated according to Wang *et al.* [82] The systems were simulated in TIP3P explicit water solvent, adding sodium or chloride ions to get neutrality. The SHAKE algorithm was employed to constrain all bonds involving hydrogen to their equilibrium bond lengths, which allowed an integration step of 1 fs. All complexes underwent 1000 steps of steepest descent minimization, followed by 9000 steps of conjugate gradient minimization with a 500 kcal mol^{-1} A^{-2} restraint on the position of protein backbone atoms and of all molecule atoms. A 100 ps NVT equilibration at 300K and a subsequent 100 ps NPT equilibration at 1 atm and 300K were performed to ensure temperature and density equilibration. A 500 kcal mol^{-1} $Å^{-2}$ restraint on the position of protein backbone atoms was applied during the equilibration phases. A 2 ns productive phase under NPT conditions (1 atm, 300K) was then performed using a 10 Å cutoff for non-bonded interactions and the Particle Mesh Ewald algorithm to calculate long-range electrostatic interactions. The systems were then cooled down to 0K at a 4 K ps^{-1} constant cooling rate to get a structure devoid of fluctuations, to be used for comparison with Quantum Mechanics/Molecular Mechanics (QM/MM) results and crystallographic structures.

It resulted that, in the presence of the pseudo-atom, shorter halogen bond distances and larger (i.e., closer to 180°) D–X···O angles were achieved with respect to the corresponding values obtained by simulations excluding the use of pseudo-atoms. Both geometrical aspects can be easily ascribed to a proper description of the σ-hole, as allowed by the use of pseudo-atoms. It is to be noted that, also in the presence of the pseudo-atom, the X···O distances happened to be significantly longer than the sum of their van der Waals radii, during the simulation, because the trajectories of molecular dynamics simulations are of course characterised by fluctuating distances. In order to reduce such fluctuations and to identify a value unaffected by a kinetic contribution, all the systems, after the production phase, were gradually cooled to a temperature T=0 K. The corresponding geometrical parameters are reported in **Table 4** together with the crystallographic values. The results show that the simulation protocol with pseudo-atoms always provided smaller final halogen bond distances with respect to the corresponding simulations without pseudo-atoms and the values were always less than the sum of the van der Waals radii. The obtained results compared also very well with the crystallographic values.

9.3 QM/MM optimizations

For comparison purposes, for each system, an extended series of QM/MM geometry optimizations was performed, where the QM part, treated at DFT level, was constituted by the ligand and the halogen-bonded residue(s). Different protocols were used in the QM/MM methodology, using different functionals and coupling schemes between the QM and the MM regions and including or not water molecules located near the ligands. Dr. Alessandra Forni (ISTM, CNR) is acknowledged for the support in the set-up of the calculations.

QM/MM results, reported in **Table 4**, were quite scattered, depending on the protocol adopted in the calculations, but were overall closer to the results obtained in the presence of pseudo-atoms. The agreement was generally better for QM/MM results obtained with the electrostatic embedding approach, where the quantum mechanical region can be polarized by the MM region. Unfortunately, calculations in this case were more expensive with respect to the simpler mechanical embedding scheme and the geometry optimization was more problematic, resulting in some cases into convergence failure. Simulations performed in the absence of pseudo-atoms caused in several cases a complete disruption of the halogen bond interactions, while in some other cases the halogen bond distances were only slightly greater with respect to those obtained in the presence of pseudo-atoms. However, it is important to point out that the halogen bond interaction is strongly anisotropic due to the presence of the σ-hole, and a correct geometrical description of this interaction should be

characterized not only by the distance d(X···O), but also by the angle D–X···O. Taking this point into account, it resulted that in the cases where simulations without pseudo-atoms gave acceptable value in the halogen bond distances, generally the D–X···O angles were poorly described.

QM/MM optimization details

QM/MM calculations were performed using the two-layer ONIOM model [151], [152] [153] as implemented in Gaussian09 [170]. The ligands and the protein residues halogen-bonded to the ligands were included in the QM layer, whereas the rest of the protein was included in the MM layer. Calculations were performed by either excluding water molecules or including those falling into a shell of 8 Å around the ligand as part of the MM layer. The QM part was treated at the DFT level with two functionals, B3LYP [171] [81] and MPWLYP [81] [172] , which had been previously proved to give satisfactory results in studying halogen bond properties [173]. The 6-311G* basis set was used for all atoms [174]. The Amber Force Field was chosen to describe the MM region. A well-known critical point in QM/MM calculations [175] [176] [177] is the mutual electrostatic interaction between the mechanical and the quantum part. Calculations were thus performed using both the less expensive 'mechanical embedding' and the more sophisticated, even if more expensive, 'electrostatic embedding'. In the mechanical embedding, the mutual electrostatic interaction between the two layers is treated at the MM level. In this way the QM electron density cannot be polarized by the atoms of the MM layer. To describe this polarization, the 'electrostatic embedding' scheme can be adopted, where the point charges of the MM layer are included in the Hamiltonian of the QM layer, so introducing a polarization of the wave function. This results in a more expensive but more rigorous way to describe the interactions between the two layers. Each QM/MM calculation was carried out using, as starting geometry, the structure obtained at the end of the MD simulation (i.e., the structure at 0 K). Exceptions to this strategy were some 'electrostatic embedding' calculations afflicted by some convergence difficulties. In these cases the corresponding converged structure obtained with the 'mechanical embedding' calculation was used.

In order to appreciate the overall accuracy of the different computational protocols in the halogen bond description, distances and angles for each halogen bond considered were plotted (**Figure 56**). It is evident that the points obtained by molecular dynamics simulations using pseudo-atoms tended to cover the same area of crystallographic structures and QM/MM results, whereas simulations without pseudo-atoms were more scattered and characterized by the presence of some outliers.

The stability of protein-ligand complexes, nevertheless, results from a sum of different contributions; in fact, the protein-ligand binding is almost always due to a network of different interactions [178] ,such as hydrogen and halogen bonds, hydrophobic contacts and van der Waals interactions. In some of the reported cases (e.g. 1T0S, 3E79, 3E3T, 1ZOH), a poor description of halogen bonding caused also a disruption of the other interactions and a consequent loss of the correct docking of the ligand. For example, the BML ligand in the 1T0S system, established, in addition to the halogen bond with Glu231, an hydrogen bond between the oxydrilic group and Glu265, and hydrophobic contacts between the ring of the BML molecule and the atoms of the hydrophobic portion of the chains of residues Arg256 and Lys258. All these interactions were correctly described if the simulation was carried out in the presence of the pseudo-atom. When pseudo-atom was absent, a strong increase of the halogen-oxygen distance (up to 5.2 Å) was observed, together with the disruption of the hydrogen bond and a movement of the residues Arg256 and Lys 258, bringing to a loss of the hydrophobic contacts.

In other cases (e.g. 2H77, 1NQ2, 1P5E), the absence of pseudo-atoms gave rise to less dramatic effects, since other interactions between the ligand and the protein, besides halogen bond, kept the ligand correctly docked in its crystallographic position. For example, the ligand T3 in the 2H77 system established, in addition to hydrogen and halogen bonds, a complicate network of hydrophobic interactions with different residues, namely Ile222, Leu292, Leu276, Met259 and Leu287. In the absence of pseudo-atoms, the distance between I1 and the oxygen of the Phe218 residue was only a little longer with respect to the value obtained by simulation with pseudo-atom (i.e. 3.41 Å instead of 3.28 Å). The hydrophobic contacts were preserved and the ligand was able to rearrange in the pocket. The incorrect description of the halogen bond was still certified by the D–X···O angle value (161.78°), quite different from the crystallographic one (174.36°), which was instead adequately reproduced if pseudo-atoms were used (178.85°). A similar situation occurred for the 1NQ2 system.

It is also worth noting that the use of the pseudo-atom also allowed, in the system 1P5E, the correct description of the interaction between Br1 and the phenyl ring of the residue Phe80. In this particular case, the π-cloud of the aromatic ring could act as an acceptor group in the place of an oxygen atom, leading to the formation of an aromatic halogen bond. In the presence of the pseudo-atom, the distance between the bromine atom and the phenyl ring turned out to be very close to both the crystallographic value and that obtained by QM/MM calculations.

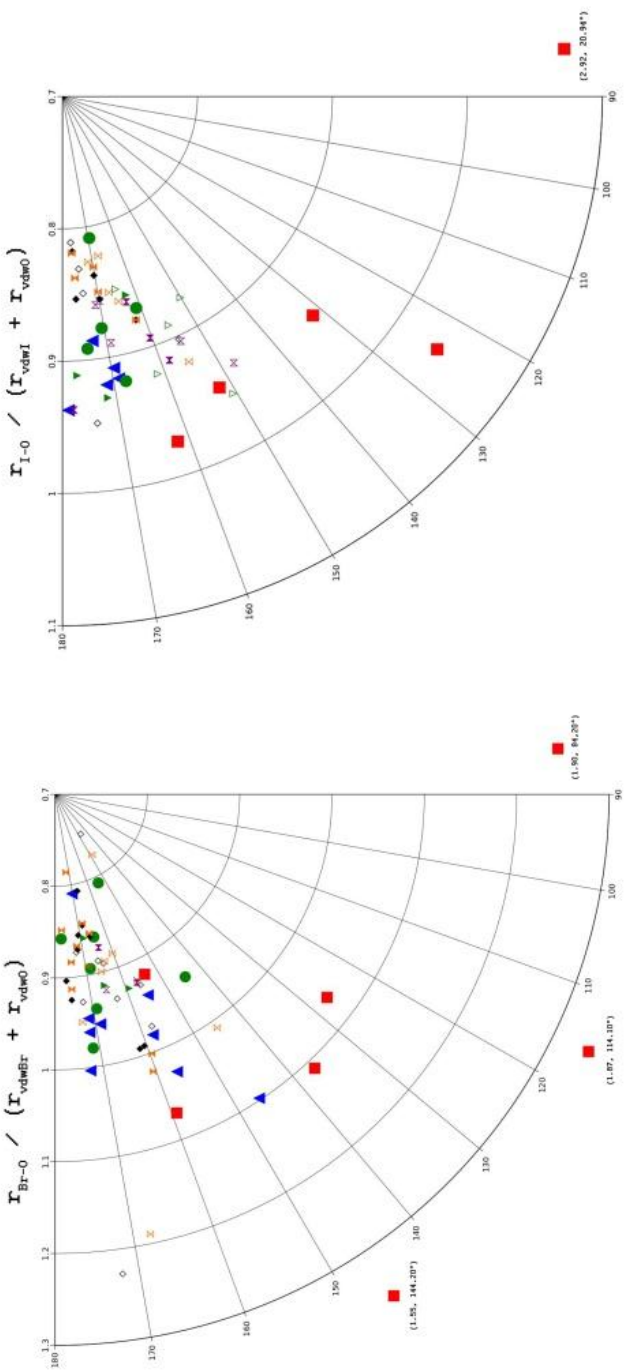


Figure 56 Distribution of the geometrical parameters of the halogen bonds involving bromine (left) and iodine (right) (distances weighted by the sum of the van der Waals radii of the atoms involved and by 4.3 Å in the case of bromine-phenyl interaction; green circles: crystallographic data; blue triangles and red squares: final structures of the MD simulations with or without pseudo-atom, respectively. The results obtained by QM/MM methods are indicated by smaller symbols: empty symbols: calculations without water molecules; filled symbols: corresponding calculations including water molecules).

Table 4 Geometrical parameters of C–X···O halogen bonds in protein–ligand complexes from molecular dynamics (MD) simulations at T=0 K in the absence and in the presence of pseudo-atom on the halogen, from X-ray crystal structure and from QM/MM calculations with B3LYP or MPWLYP functional, in the absence and in the presence of water molecules, and with mechanical or electrostatic embedding.[a]

	MD			QM/MM (no water)				QM/MM (water)			
				Mechanical embedding		Electrostatic embedding		Mechanical embedding		Electrostatic embedding	
System	No pseudo-atom	Pseudo-atom	X-Ray	B3LYP	MPWLYP	B3LYP	MPWLYP	B3LYP	MPWLYP	B3LYP	MPWLYP
Br···O (Å)											
1P5E-Br1[b]	3.95	3.48	3.47	3.230	3.340	nc	nc	3.473	3.378	nc	nc
Br2	3.74	3.38	3.01	3.128	3.201			3.115	2.978		
Br3	3.60	3.17	2.90	2.943	3.001			2.883	2.858		
1GXZ	3.61	3.22	3.16	3.003	2.996	3.100	3.082	2.896	2.883	3.110	nc
1T0S	LOST	3.19	3.30	3.125	3.102	nc	3.120	2.849	2.843	2.942	2.897
1ZOH-Br3	LOST	3.29	3.18	3.280	3.399	nc	nc	3.342	3.436	nc	nc
Br4	LOST	3.31	2.89	2.990	2.978			3.045	3.076		
Br2	Not reproduced	3.47	Not present	4.135	4.009			3.337	3.374		
Br3	Not reproduced	3.24	Not present	3.141	3.031			2.934	2.921		
C–Br···O (°)											
1P5E-Br1[b]	153.47	170.22	153.79	146.49	148.12	nc	nc	166.88	172.04	nc	nc
Br2	136.62	172.60	168.55	172.28	173.01			175.22	174.44		
Br3	126.80	155.15	164.70	172.23	169.12			170.58	177.32		
1GXZ	159.11	168.71	169.03	164.08	163.52	165.31	165.55	166.23	166.21	156.50	nc
1T0S	LOST	171.45	171.40	155.83	156.22	nc	159.09	168.12	168.05	164.04	168.74
1ZOH-Br3	LOST	146.18	144.56	157.33	145.30	nc	nc	161.53	160.45	nc	nc
Br4	LOST	157.85	177.45	165.45	160.11			176.46	174.96		
Br2	Not reproduced	156.26	Not present	171.97	167.80			160.47	159.61		
Br3	Not reproduced	171.59	Not present	163.08	165.38			171.54	171.87		
I···O (Å)											
1Y0X	3.32	3.22	3.22	2.907	2.893	3.008	2.977	2.860	2.864	3.127	nc
2H77	3.41	3.28	3.12	2.836	2.982	3.113	3.224	2.987	2.931	3.279	3.188
3E79	3.63	3.21	3.04	3.159	3.225	3.166	3.114	3.071	3.070	3.017	3.000
3E3T	LOST	3.18	2.83	2.973	3.010	2.998	3.063	2.994	2.974	nc	nc
1NK2	3.32	3.10	3.07	3.319	2.881	3.282	3.352	2.930	2.907	3.201	3.255
C–I···O (°)											
1Y0X	152.03	171.35	167.64	174.56	171.28	171.08	164.85	176.35	176.63	160.37	nc
2H77	161.78	178.85	174.36	176.84	166.77	169.01	161.39	176.14	176.07	177.96	176.96
3E79	124.44	168.81	161.08	154.76	154.86	154.61	155.57	162.03	162.04	163.03	162.60
3E3T	LOST	169.43	169.38	174.11	164.90	169.91	150.00	169.66	169.78	nc	nc
1NQ2	131.68	173.09	170.51	173.96	167.50	147.71	150.57	170.21	169.84	158.21	171.52

[a] Non-converged geometry optimizations are indicated with the acronym nc; geometries with X···O distances greater than 5 Å are reported as 'LOST'. [b] The center of mass of the phenyl ring is taken into account for the calculation of distances and angles.

9.4 Complex K44 – protein kinase CK2

Among the systems studied, it is very interesting to observe the dynamical behaviour of the 1ZOH structure, where the ligand K44 possessed four adjacent bromine atoms (Br1 to Br4). During the MD simulation of 2ns, the ligand K44 kept for almost the whole dynamics the conformation observed in the crystal structure, where the molecule established hydrophobic interactions (with the residues Ile66 and Ile174) and two halogen bonds involving Br3 and Br4 with, respectively, the carbonyl oxygens of the peptidic bond of Glu114 and Val116 (**Figure 57a**). Occasionally, the ligand underwent a small rotation, so as to give rise to new halogen bonds, involving Br3 and Br2 and the same residues (**Figure 57b**) still preserving the hydrophobic interactions with Ile66 and Ile174.

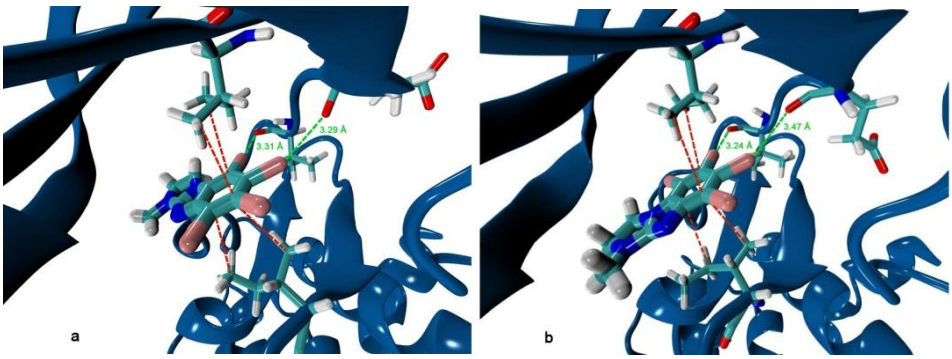


Figure 57 Comparison between the conformation A (a) and B (b) adopted by the ligand K44 during the molecular dynamics simulation (green dashed lines indicate the halogen bonds with the residues Glu114 and Val116; red dashed lines indicate the hydrophobic interactions with the residues Ile66 and Ile174).

The cooling to 0 K was conducted for both conformations, giving rise to two different final conformations denoted as A and B in **Table 4**. It is to be noted that the protein of the 1ZOH system is kinase2 (CK2), which has been crystallized with different ligands besides K44. In one of these structures (PDB id: 1ZOG [162] [179]), the ligand is a tetrabromo-benzimidazole derivative, like K44, which is found to crystallize into two different poses, corresponding to the two conformations obtained for the ligand K44. Interestingly, the maximum values assumed by the electrostatic potential on the electron density isosurface in the σ-hole regions had the following ratios: 1:0.74:0.66:0.60 (Br4:Br3:Br2:Br1). Accordingly to Politzer [149], a similar relationship in the strength of the interactions that each halogen atom can establish with the protein residues could be expected. This explains why the dynamics well described the crystallographic conformation involving the two halogens with the most prominent σ-holes, Br4 and Br3, but occasionally the molecule could shift toward another conformation involving Br3 and Br2. This supported the existence of two different energy minima which were both explored by the simulation with pseudo-atoms. QM/MM calculations confirmed the presence of such two minima, with the second conformation characterized by longer halogen bond distances. On the other hand, the simulation of the 1ZOH system without pseudo-atoms was not able to describe any of the two conformations, since the distances between bromine and oxygen atoms were always found much longer than the sum of their van der Waals radii. These results provided a further proof of the importance of including the pseudo-atom on the halogen

atoms to get a proper description of the conformations of the ligands and of their relative stability. All these results indicate that the use of pseudo-atoms located along the D–X direction, in such a way to optimize the fitting with the electrostatic potential obtained at the quantum mechanical level, is a simple and viable strategy to perform molecular dynamics simulations of halogen bonded systems, using traditional force fields based on atom-centred charges.

Chapter 10

Solvent effect on halogen bonding

In order to deepen the knowledge on the geometrical and energetic features of halogen bonding, a systematic theoretical study of the solvation effects on this interaction was also performed. In fact, up to now, only few theoretical investigations of the solvent effect on the halogen bond properties can be found in literature [180] [181] [182].

As model system, dimers constituted by differently substituted iodobenzene with formaldehyde in two solvents (diethylether and water) were taken into account for QM calculations, in order to evaluate differences in geometry and energy moving from vacuo to solvent. The choice of formaldehyde as acceptor group was due to the fact that a survey [154] of protein-halogenated ligand complexes evidenced the frequent occurrence of halogen bonding with oxygen atoms of a peptidic bond. In our case, the attention was focused on iodine-based halogen bonds, since they are

characterized by the stronger interaction energies, allowing to better elucidate the variations observed when moving from vacuo to solvent. Different schemes of substitution were used, aimed at modulating the force of the halogen bond, from weak to moderate interaction.

Calculations were carried out at both the MP2 level, in order to adequately describe the electronic correlation effects, and the DFT level using two different density functionals, in order to evaluate their ability to reproduce the effects of the solvents in the description of the halogen bond (see the Materials and Methods box for more details).

10.1. Gas-phase calculations

A first set of calculations was performed in vacuo on a series of halogen-bonded complexes of variously substituted iodobenzene with formaldehyde (see **Table 5** for the list of compounds taken into account). Different substitution schemes on iodobenzene were used as a simple way to modulate the strength of the interaction with formaldehyde. The same approach was followed by Riley and Merz [180] in their study on the strength and origin of halogen bonding. The halogen bond interaction energies and geometrical parameters of the optimized geometries at different theoretical levels are listed in **Table 5**.

Table 5 Interaction energies and geometrical parameters of the optimized structures obtained in vacuo at MP2 and DFT level.

		Interaction energy (kcal/mol)						Distance I···O (Å)			Angle C-I···O (°)			Distance C-I (Å)		
		MP2		PBE		B3LYP										
	Molecule	ΔE	ΔE_{cp}	ΔE	ΔE_{cp}	ΔE	ΔE_{cp}	MP2	PBE	B3LYP	MP2	PBE	B3LYP	MP2	PBE	B3LYP
1	*p*-amino-iodobenzene	-3.97	-2.25	-2.17	-1.82	-1.33	-0.99	3.131	3.183	3.296	169.6	167.9	168.2	2.107	2.132	2.133
2	iodobenzene	-4.07	-2.34	-2.33	-2.01	-1.49	-1.17	3.114	3.146	3.220	170.2	169.4	172.6	2.109	2.134	2.134
3	*o*-amino-iodobenzene	-4.29	-2.54	-2.36	-2.07	-1.52	-1.23	3.114	3.165	3.254	169.7	169.3	170.3	2.111	2.139	2.140
4	*p*-chloro-iodobenzene	-4.31	-2.57	-2.64	-2.28	-1.78	-1.44	3.083	3.103	3.179	171.0	170.5	173.4	2.106	2.131	2.131
5	*m*-chloro-iodobenzene	-4.35	-2.62	-2.69	-2.35	-1.78	-1.49	3.079	3.085	3.180	171.0	173.3	173.2	2.107	2.132	2.132
6	3,5-difluoro-iodobenzene	-4.57	-2.82	-2.97	-2.62	-2.10	-1.77	3.054	3.046	3.141	171.6	173.6	173.8	2.105	2.129	2.129
7	*p*-cyano-iodobenzene	-4.55	-2.83	-3.03	-2.68	-2.15	-1.81	3.056	3.049	3.144	171.9	173.8	174.0	2.104	2.127	2.128
8	3,4,5-trifluoro-iodobenzene	-4.74	-2.98	-3.15	-2.79	-2.29	-1.94	3.037	3.031	3.122	172.1	173.8	174.1	2.104	2.129	2.128
9	3,5-dicyano-iodobenzene	-5.23	-3.48	-3.79	-3.42	-2.86	-2.51	2.994	2.984	3.077	173.5	174.5	175.0	2.103	2.128	2.127
10	pentafluoroiodobenzene	-5.81	-3.93	-4.35	-3.94	-3.36	-2.97	2.946	2.922	3.006	173.2	174.4	174.6	2.090	2.116	2.116

Materials and methods

All ab initio calculations were carried out using the Gaussian '09 program [170]. Full geometry optimization were performed by the ab-initio second order Möller-Plesset perturbation (MP2) method [183] and density functional theory (DFT) using two widely used functionals, namely B3LYP [171] [81] and PBE [184]. The aug-cc-pVDZ-PP basis set [185], which uses pseudopotential to describe the inner core orbitals, was used for iodine, whereas for all other atoms the aug-cc-pVDZ basis set [186] was adopted. These basis set have already shown to be of good quality to describe various halogen-bonded systems [150] [173] [187]. All the minimum geometries were confirmed by verifying that all vibrational frequencies at the MP2 level are real. Solvent effects have been taken into account using the CPCM polarisable conductor calculation model [188].

The interaction energy ΔE was computed as the difference between the energy of the dimer and the sum of the energies of the monomers at the geometry assumed in the dimer. In order to take into account the Basis Set Superpostion error (BSSE), the values of interaction energy obtained in vacuo were corrected with the Boys and Bernardi procedure [189], to give the counterpoise corrected interaction energies ΔE_{CP}. The BSSE correction term evaluated in vacuo was also used to correct the interaction energies evaluated in solvent calculations as reported in [180]

In **Figure 58**, the electrostatic potential computed on an isosurface of electron density of 0.001 au, according to the Bader's suggestion [190], is reported for a subset of iodobenzene derivatives, *p*-amino-iodobenzene (top), *m*-chloro-iodobenzene (middle) and pentafluoro-iodobenzene (bottom). In all the cases, it is well evident the anisotropy of the electrostatic potential around the halogen atom, responsible for the formation of halogen bonding. In particular, it is well represented the σ-hole, located in the direction of the D–X bond, responsible for the interaction with an oxygen's lone pair of formaldehyde.

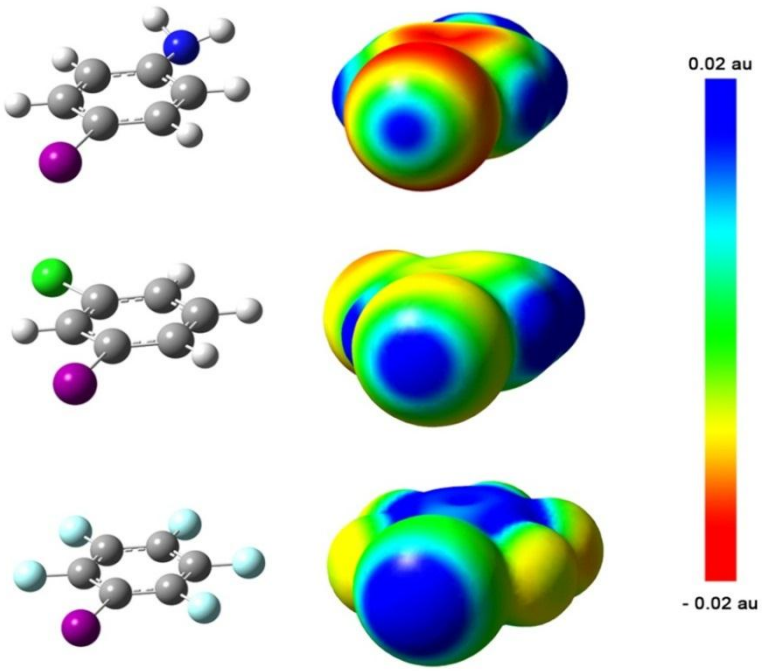


Figure 58 Electrostatic potential mapped onto the surface of electron density (0.001 electrons/bohr^3) for *p*-amino-iodobenzene, *m*-chloro-iodobenzene and pentafluoroiodobenzene: the extension of the σ-hole increases with the introduction of electron withdrawing groups on the benzene ring.

Moreover, from **Figure 58** it can be qualitatively observed that the introduction of electron-withdrawing groups on iodobenzene increased the extension of the σ-hole on the iodine atom. This corresponds to increasing (i.e., more negative) interaction energies, from -2.25 to -2.57 and then to -3.93 kcal/mol (MP2 values corrected for BSSE) on going from *p*-amino- to *m*-chloro- and pentafluoro-iodobenzene. In general, it can be observed from **Table 5** that the interaction energies depended on the nature, number and position of the substituent(s) on the ring: stronger halogen bonds were in fact observed with increasing the strength of the electron-withdrawing groups (compare e.g. *p*-

chloro-iodobenzene with *p*-cyano-iodobenzene, ΔE_{CP}=-2.57 and -2.83 kcal/mol, respectively, at MP2 level), the number of substituents (compare e.g. 3,5-difluoro- with 3,4,5-trifluoro and pentafluoro-iodobenzene, ΔE_{CP}=-2.82, -2.98 and -3.93 kcal/mol, respectively), and with changing the position of the substituents (compare e.g. *p*-amino- with *o*-amino-iodobenzene, ΔE_{CP}=-2.25 and -2.54 kcal/mol, respectively). Such dependencies, as already qualitatively observed in **Figure 58**, are ascribable to the different anisotropy of the electrostatic potential around the iodine atom, and in particular on the nature of the σ-hole, as originated by the different substituents on the ring.

By comparing the different methods of computation, it turns out from **Table 5** that, taking the MP2 results as a reference, the interaction energies computed at DFT level were in all cases underestimated. Moreover, the PBE functional appeared to better reproduce the MP2 results with respect to B3LYP. Average absolute deviations interactions energies were in fact 0.24 and 1.10 kcal/mol for the PBE and the B3LYP functionals, respectively.

Moreover, it is to be pointed out the entity of the CP correction was in all cases very large despite the quite extended basis set used in these calculations. This was in particular true as far as the MP2 results were concerned. For example, for the weaker interaction considered in this study (dimer 1, see **Table 5** for the numbering of the compounds) the error due to BSSE was 43%, 16% and 26% of the uncorrected energy at MP2, PBE and B3LYP levels, respectively. It is however to be noted that such correction was quite independent on the system studied: it ranged from 1.72 to 1.76 kcal/mol for

dimers 1-9 (1.88 kcal/mol for 10) at MP2 level, and from 0.29 to 0.41 kcal/mol for 1-10 at both DFT levels.

In **Table 5**, the geometrical parameters of halogen bonding in the different dimers examined are also reported. In all cases, the I···O distances were well below the sum of the van der Waals radii of iodine and oxygen [191], amounting to 3.50 Å, and, as expected, decreased regularly with increasing the strength of the interaction. Taking again the MP2 results as a reference, the B3LYP functional tended to overestimate the I···O distances, especially for the weaker interactions. The PBE functional instead provided both larger and smaller I···O distances with respect to the MP2 ones, but with a better overall agreement with the latter values with respect to B3LYP. In fact, the average absolute deviations of the distances amounted to 0.101 and 0.021 Å for the B3LYP and the PBE functionals, respectively. The angle C–I···O was always close to linearity and the deviations with respect to 180° tended to reduce with increasing the halogen bond strength at all levels of theory used. DFT calculations gave generally wider angles with respect to MP2. Once again, the PBE functional appeared slightly better with respect to the B3LYP one. Average absolute deviations for the C–I···O angle were 1.8° and 1.3° for B3LYP and PBE, respectively. Finally, a slight, though not systematic, shortening of the C–I bond length was observed with increasing the strength of halogen bonding, which was reproduced by all computation methods. It amounted to 0.017 Å at all levels of theory on going from dimer 1 to dimer 10.

10.2 Solvent calculations

Owing to the importance of the electrostatic contribution in halogen bonding, it is to be expected that solvent would have a non-negligible influence on the physical chemical properties of this interaction. To describe such solvation effects, the PCM model was used, where the solvent is treated as a uniform medium characterized by its dielectric constant. All the systems were optimized in two solvents, namely diethylether, whose dielectric constant ε=4.24 is very close to the value existing at the interior of a protein [192], and water (ε=78.36). The interaction energies and the geometrical parameters of the optimized structures obtained using the PCM model are reported in **Table 6** and **Table 7** for calculations in diethylether and in water, respectively.

From the energetic point of view, the solvent systematically reduced the strength of the interaction, and the reduction was more pronounced with the increase of the dielectric constant. Moreover, the strength order observed in vacuo remained unaltered when moving to solvent. Though in absolute value the stronger halogen bonds underwent a greater decrease in the interaction energy with respect to the weaker ones, the opposite trend was observed when considering the relative variations. For example, at MP2 level, ΔE_{CP} of dimer 1 decreased by 0.64 and 0.83 kcal/mol (i.e., 28% and 37%) in diethylether and in water, respectively, while for dimer 10 it decreased by 0.85 and 1.14 kcal/mol (i.e., 22% and 29%) in diethylether and in water, respectively. This tendency was as well observed at both DFT levels of theory.

Table 7 Interaction energies and geometrical parameters of the optimized structures obtained in diethylether at MP2 and DFT level.

		Interaction energy (kcal/mol)						Distance I···O (Å)			Angle C-I···O (°)			Distance C-I (Å)		
		MP2		PBE		B3LYP										
	Molecule	ΔE	ΔE_{cp}	ΔE	ΔE_{cp}	ΔE	ΔE_{cp}	MP2	PBE	B3LYP	MP2	PBE	B3LYP	MP2	PBE	B3LYP
1	*p*-amino-iodobenzene	-3.33	-1.61	-1.73	-1.38	-0.83	-0.49	3.111	3.133	3.252	174.0	175.9	176.6	2.110	2.136	2.137
2	iodobenzene	-3.42	-1.70	-1.91	-1.60	-0.98	-0.66	3.097	3.105	3.220	174.0	175.9	176.6	2.111	2.138	2.138
3	*o*-amino-iodobenzene	-3.57	-1.82	-1.92	-1.62	-1.00	-0.71	3.090	3.116	3.232	173.9	175.5	175.8	2.112	2.142	2.142
4	*p*-chloro-iodobenzene	-3.61	-1.87	-2.15	-1.79	-1.19	-0.85	3.069	3.066	3.177	174.6	176.1	176.7	2.108	2.134	2.134
5	*m*-chloro-iodobenzene	-3.64	-1.91	-2.29	-1.96	-1.31	-1.02	3.062	3.053	3.160	174.5	177.3	178.2	2.109	2.136	2.135
6	3,5-difluoro-iodobenzene	-3.83	-2.08	-2.45	-2.10	-1.45	-1.11	3.037	3.026	3.126	174.9	176.2	177.0	2.106	2.132	2.131
7	*p*-cyano-iodobenzene	-3.75	-2.03	-2.42	-2.06	-1.42	-1.08	3.049	3.034	3.142	175.0	176.2	177.0	2.106	2.129	2.130
8	3,4,5-trifluoro-iodobenzene	-3.96	-2.20	-2.59	-2.23	-1.58	-1.24	3.021	3.007	3.111	175.1	176.2	177.0	2.106	2.131	2.130
9	3,5-dicyano-iodobenzene	-4.26	-2.51	-3.00	-2.63	-1.90	-1.55	2.987	2.966	3.070	176.0	176.5	177.8	2.104	2.130	2.129
10	pentafluoroiodobenzene	-4.97	-3.08	-3.75	-3.35	-2.57	-2.18	2.916	2.878	2.964	176.6	177.4	179.9	2.092	2.120	2.118

Table 6 Interaction energies and geometrical parameters of the optimized structures obtained in water at MP2 and DFT level.

		Interaction energy (kcal/mol)						Distance I···O (Å)			Angle C-I···O (°)			Distance C-I (Å)		
		MP2		PBE		B3LYP										
	Molecule	ΔE	ΔE_{cp}	ΔE	ΔE_{cp}	ΔE	ΔE_{cp}	MP2	PBE	B3LYP	MP2	PBE	B3LYP	MP2	PBE	B3LYP
1	*p*-amino-iodobenzene	-3.14	-1.42	-1.62	-1.27	-0.70	-0.36	3.105	3.132	3.261	175.6	177.1	178.1	2.111	2.138	2.138
2	iodobenzene	-3.22	-1.5	-1.79	-1.48	-0.83	-0.51	3.093	3.101	3.223	175.4	177.0	178.0	2.112	2.140	2.139
3	*o*-amino-iodobenzene	-3.33	-1.58	-1.76	-1.46	-0.81	-0.52	3.083	3.111	3.232	175.5	174.7	177.2	2.113	2.143	2.143
4	*p*-chloro-iodobenzene	-3.39	-1.65	-1.99	-1.63	-1.00	-0.66	3.066	3.063	3.182	175.8	177.1	177.8	2.109	2.135	2.135
5	*m*-chloro-iodobenzene	-3.42	-1.69	-2.13	-1.80	-1.11	-0.82	3.057	3.045	3.157	175.9	178.0	179.3	2.110	2.137	2.136
6	3,5-difluoro-iodobenzene	-3.59	-1.84	-2.27	-1.92	-1.23	-0.89	3.033	3.019	3.127	176.1	177.1	178.1	2.107	2.133	2.131
7	*p*-cyano-iodobenzene	-3.48	-1.76	-2.25	-1.89	-1.20	-0.86	3.050	3.028	3.140	176.1	176.8	177.7	2.107	2.130	2.131
8	3,4,5-trifluoro-iodobenzene	-3.70	-1.94	-2.39	-2.03	-1.33	-0.99	3.019	3.001	3.107	176.2	177.2	178.2	2.106	2.132	2.130
9	3,5-dicyano-iodobenzene	-3.92	-2.17	-2.75	-2.38	-1.59	-1.24	2.989	2.962	3.062	176.7	177.2	179.8	2.104	2.130	2.129
10	pentafluoroiodobenzene	-4.68	-2.79	-3.60	-3.20	-2.30	-1.91	2.906	2.852	2.949	177.7	179.7	178.7	2.093	2.122	2.119

From the geometrical point of view, on the other hand, the presence of the solvent implied a systematic decrease in the I···O distances at all theoretical levels, as previously reported for other halogen bonded dimers [182]. Moreover, an increase (towards 180°) of the C–I···O angle was observed on going from vacuo to diethylether and then to water. The mean reductions of the I···O distances from vacuo to diethylether were 0.017 Å for MP2 and B3LYP calculations, and 0.033 Å for PBE calculations. The corresponding mean increases for the angle values were 3.5° for MP2 and 4.3° for both DFT calculations, leading to halogen bonds very close to the linear arrangement. The corresponding variations from vacuo to water were 0.021, 0.040, 0.018 Å and 4.7, 5.1 and 5.4° for MP2, PBE and B3LYP calculations, respectively. Interestingly, the reduction in the distances and the greater tendency to linearity observed when going from vacuo to solvent were then not correlated to an increase of the strength of the interaction, as could be naively expected. In other words, for a same halogen bond distance and angle, a much weaker interaction should be expected in solvent with respect to the gas phase.

The decrease of the I···O distances in solvent with respect to vacuum was always accompanied by a slight increase of the C–I bond lengths. The relationships between the variations of the I···O and the C–I distances has been already observed for hydrogen bonded systems [193], for which it has been interpreted as a nonadditivity/polarization effect in PCM calculations [181]. It has also been reported that the distance I···O in the dimer iodobenzene/water is further reduced if one or two explicit water molecules, hydrogen-bonded to the

water molecule acting as halogen bond acceptor, are added to the dimer [182]. This further confirms the solvent effect on the halogen bonding geometry. It also resulted that, while the PBE functional appeared superior to the B3LYP one in reproducing absolute values of energy or geometrical parameters, their variations were well reproduced by both functionals, evidently for a cancellation of errors.

In **Figure 59** (top) is reported a plot of the strength of halogen bonding against the maximum value assumed by the electrostatic potential, computed at the MP2 level, on the isosurface of electron density (0.001 au) in the region of the σ-hole, denoted as $V_{S,max}$. Results obtained in vacuo, diethylether and water are shown (**Figure 59** left, centre and right, respectively). In all cases, it is evident a negative correlation, according to results already reported by Politzer for a series of halogen bonded dimers [149] [148] [150], including in particular brominated benzenes and pyrimidines [193], indicating explicitly that a more pronounced σ-hole determines a stronger halogen bond. This is also confirmed by the I···O distances (**Figure 59**, bottom), which tended as well to decrease with increasing of $V_{S,max}$. The presence of the solvent gave rise to a shifting of the curves towards lower interaction energies and shorter I···O distances. The same behaviour was obtained at the DFT level for both functionals.

The previous considerations have been made by considering the variations of the electrostatic potential on the 0.001 au isosurface of electron density. It can be interesting also to analyze how the electrostatic potential changes moving away from this isosurface.

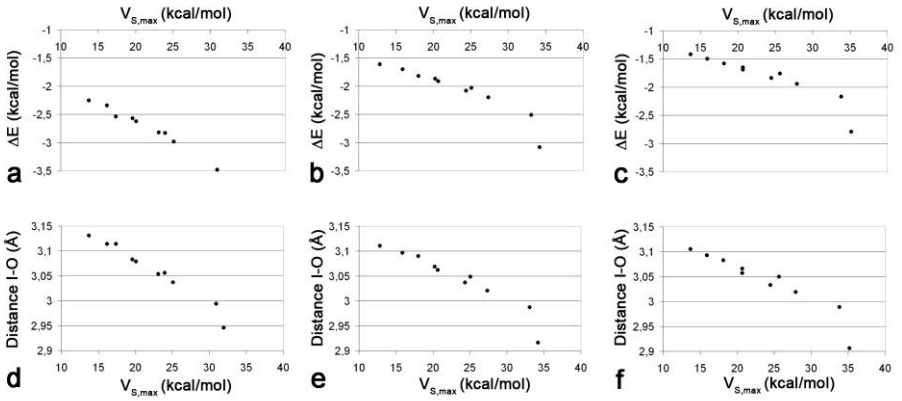


Figure 59 Plots of the I…O distance and interaction energy ⌾E versus $V_{S,max}$, the maximum value assumed by the electrostatic potential on the isosurface of electron density (0.001 au) in the region of the sigma hole, calculated in vacuo (a, d), diethylether (b, e) and water (c, f) at MP2 level.

In **Figure 60**, some positive contour lines of the electrostatic potential are reported superimposed with the 0.001 au contour line of the electron density, on the plane of the phenyl group of two compounds, namely *p*-amino-iodobenzene and pentafluorobenzene, in vacuo (**Figure 60a** and **Figure 60b**) and in water (**Figure 60c** and **Figure 60d**). From both vacuum and solvent results, it is well evident the different extension of the protuberance in the C–I region outside the space delimited by the 0.001 au electron density isosurface: it is quite limited for *p*-amino-iodobenzene, which gives rise to the weaker halogen-bonded dimer (**Figure 60a** and **Figure 60c**), while increases considerably along the C–I bond for pentafluorobenzene, which is associated to the stronger interaction (**Figure 60b** and **Figure 60d**). Clearly, the extension of such protuberance is strictly connected to the extension of the σ-hole. Such representation makes however more evident why larger σ- holes give rise not

only to stronger interactions, but also to C–I···O angles closer to linearity. Stronger halogen bond donors are in fact able to 'capture' Lewis basis at longer distances with respect to weaker donor, making more favourable the linear approach between the interacting species.

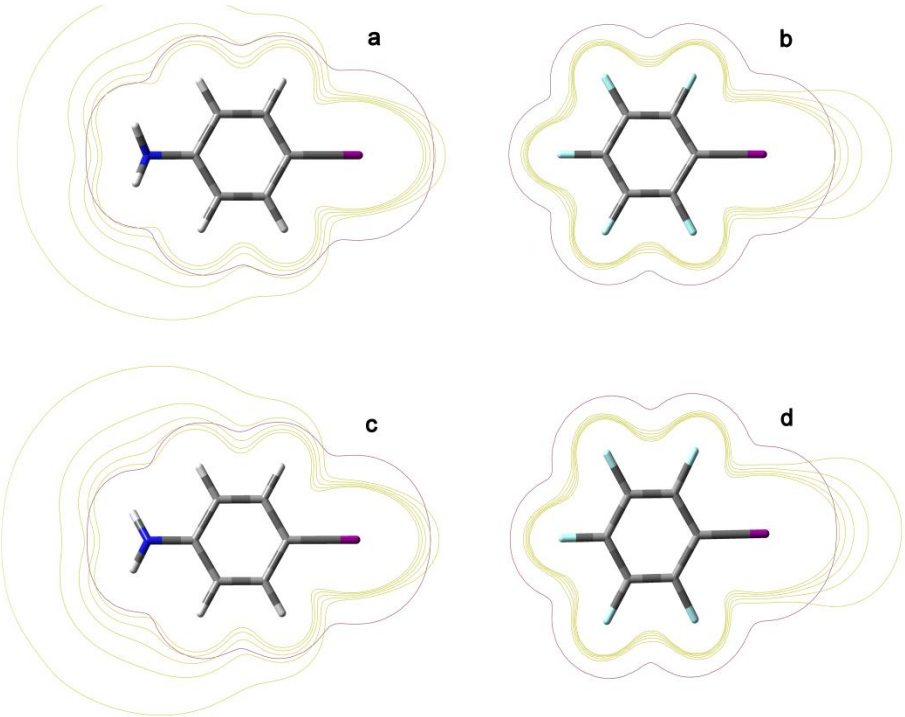


Figure 60 Positive contour lines (from 0.01 to 0.05 au) of the electrostatic potential superimposed with the 0.001 au contour line of the electron density, on the plane of the phenyl group of p-amino-iodobenzene and pentafluorobenzene, in vacuo (a, b) and in water (c, d).

Part III

Conclusions

The problem of representing halogen bonds in molecular mechanics can be solved by simply introducing a pseudo-atom, i.e. an extra point located along the D–X bond, allowing to adequately fit the anisotropy of the electrostatic potential around halogen atoms. By using this approach, it was possible to properly describe the behaviour of halogenated ligands in complex with an extended series of known proteins, providing results in close agreement with crystallographic structures and with results of more expensive QM/MM calculations. The results therefore indicated that traditional force fields can be still used to describe halogen bond interactions, provided that such a simple modification is introduced in the force field itself. The need of introducing pseudo atoms was confirmed by the results obtained by the same simulations performed in the absence of the pseudo atoms. Using the native force field, in fact, halogen bonds either were poorly described, assuming very distorted

geometries, or disappeared completely throughout the simulation. The improvement obtained using the pseudo-atom approach can thus be valuable in the field of rational design of new halogenated drugs.

A systematic study on the solvent effect on halogen bonding was also performed through calculations at the MP2 and DFT level on a series of I···O halogen-bonded complexes of variously substituted iodobenzene with formaldehyde. The effect of two solvents, i.e. diethylether and water, on the energetical and geometrical aspects of the interaction, have been investigated by comparison with results obtained in vacuo. The different substitution schemes on iodobenzene were used as a simple way to modulate the strength of the interaction from weak to moderate halogen bonds. It turned out that solvent acts in a systematic way on both geometry and strength of halogen bonding, implying, from one side, a shortening of the I···O distance and a greater tendency to linearity of the C–I···O angle, and, from the other side, a decrease of the interaction energy. Such effects were found to increase with the polarity of the solvent. Though halogen bonding appears then to be destabilized in solvent with respect to the gas phase, it plays anyway a fundamental role in biological systems, in synergy or even in competition with hydrogen bond.

Bibliography

1 J.A. Wells, C.L. McClendon. *Nature*, 450 (2007), 1001-1009.

2 M.R. Arkin, J.A. Wells. *Nature Rev. Drug Disc.*, 3 (2004), 301-317.

3 T. Berg. *Angew. Chem. Int. Ed.* , 42 (2003), 2462-2481.

4 L. Young, R.L. Jernigan, D.G. Covell. *Protein Sci.*, 3 (1994), 717-729.

5 A.P. Korn, R.M. Burnett. *Proteins*, 9 (1991), 37-55.

6 C.J. Tsai, R. Nussinov. *Protein Sci.*, 6 (1997), 1426-1437.

7 L. Lo Conte, C. Chothia, J. Janin. *J. Mol. Biol.*, 285 (1999), 2177-2198.

8 F.B. Sheinman, R. Norel, B. Honig. *Curr. Opin. Struct. Biol.*, 10 (2000), 153-159.

9 S. Schmitt, D. Kuhn, G. Klebe. *J. Mol. Biol.*, 323 (2002), 387-406.

10 P. Chène. *Chem. Med. Chem.*, 1 (2006), 400-411.

11 N.Horton, M. Lewis. *Prot. Sci.*, 1 (1992), 169-181.

12 I.S. Moreira, P.A. Fernandes, M.J. Ramos. *Proteins*, 68 (2007), 803-812.

13 S. Jones, J.M. Thornton. *Proc. Natl. Acad. Sci. USA*, 93 (1996), 13-20.

14 DeLano, W.L. *Curr. Opin. Struct. Biol.*, 12 (2002), 14-20.

15 A.A. Bogan, K.S. Thorn. *J. Mol. Biol.*, 280 (1999), 1-9.

16 O. Keskin, B. Ma, R. Nussinov. *J. Mol. Biol.*, 345 (2005), 1281-1294.

17 O. Vilacañas, J. Rubio-Martinez. *Proteins*, 63 (2006), 797-810.

18 H. Zhong, H.A. Carlson. *Proteins*, 58 (2005), 222-234.

19 P.H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A.J. Levine, N.P. Pavletich. *Science*, 274 (1996), 948-953.

20 J. Pons, A. Rajpal, J.F. Kirsch. *Prot. Sci.*, 8 (2007), 958-968.

21 J.A. Wells. *Methods Enzymol.*, 202 (1991), 390-411.

22 K.L. Morrison, G.A. Weiss. *Curr. Opin. Chem. Biol.*, 5 (2001), 302-307.

23 K.S. Thorn, A.A. Bogan. *Bioinformatics*, 17 (2001), 284-285.

24 A.H. Elcock, D. Sept, J.A. McCammon. *J. Phys. Chem. B*, 105 (2001), 1504-1518.

25 V. Zoete, M. Meuwly, M. Karplus. *Proteins*, 61 (2005), 79-93.

26 P. Kollman. *Chem. Rev.*, 93 (1993), 2395-2417.

27 H. Gouda, I.D. Kuntz, D.A. Case, P.A. Kollman. *Biopolymers*, 68 (2003), 16-34.

28 G.M. Verkhivker, D. Bouzida, D.K. Gehlhaar, P.A. Rejto, S.T. Freer, P.W. Rose. *Proteins*, 48 (2002), 539-557.

29 J. Åqvist, C. Medina, J.-E. Samuelsson. *Protein Eng.*, 7 (1994), 385-391.

30 I. Massova, P.A. Kollman. *Persp. Drug Disc. Des.*, 18 (2000), 113-135.

31 I. Massova, P.A. Kollman. *J. Am. Chem. Soc.*, 121 (1999), 8133-8143.

32 I.S. Moreira, P.A. Fernandes, M.J. Ramos. *J. Phys. Chem. B*, 110 (2006), 10962-10969.

33 I.S. Moreira, P.A. Fernandes, M.J. Ramos. *Theor. Chem. Acc.*, 117 (2007), 99-113.

34 S. Huo, I. Massova, P.A. Kollman. *J. Comput. Chem.*, 23 (2002), 15-27.

35 I.S. Moreira, P.A. Ferndandes, M.J. Ramos. *J. Comp. Chem.*, 28 (2006), 644-654.

36 E. Nogales, M. Whittaker, R.A. Milligan, K.H. Downing. *Cell*, 96 (1999), 79-88.

37 K.H. Downing, E. Nogales. *Curr. Opin. Cell. Biol.*, 10 (1998), 16-22.

38 E. Nogales, H. Wang. *Curr. Oping. Cell Biol.*, 18 (2006), 179-184.

39 E.M. Mandelkow, E. Mandelkow, R.A. Milligan. *J. Cell Biol.*, 114 (1991), 977-991.

40 Y. Gebremichael, J.W. Chu, G.A. Voth. *Biophys. J.* , 95 (2008), 2487-2499.

41 Nogales, E. *CMLS, Cell. Mol. Life Sci.*, 56 (1999), 133-142.

42 A. Akhmanova, M.O. Steinmetz. *Nat. Rev. Mol. Cell Bio.*, 9 (2008), 309-322.

43 G.C. Rogers, S.L. Rogers, D.J. Sharp. *J. Cell Sci.*, 118 (2005), 1105-1116.

44 E. Nogales. *Annu. Rev. Biochem.*, 69 (2000), 277-302.

45 M. Carmena, W.C. Earnshaw. *Nat. Rev. Mol. Cell Bio.*, 4 (2003), 842-854.

46 M.A. Jordan, L. Wilson. *Nat. Rev.*, 4 (2004), 253-265.

47 J.H. Hayden, S.S. Bowser, C.L. Rieder. *J. Cell. Biol.*, 111 (1990), 1039-1045.

48 E. Hamel. *Med. Res. Rev.*, 16 (1996), 207-231.

49 M.A. Jordan, D. Thrower, L. Wilson. *J. Cell. Sci.*, 102 (1992), 401-416.

50 M.A. Jordan. *Curr. Med. Chem. - Anticancer Agents*, 2 (2002), 1-17.

51 C. Dumontet, B.I. Sikic. *J. Clin. Oncol.*, 17 (1999), 1061-1070.

52 F. Pellegrini, D.R. Budman. *Cancer Investigation*, 23 (2005), 264-273.

53 J. Torin Huzil, K. Chen, L. Kurgan, J.A. Tuszynski. *Cancer Inf.*, 3 (2007), 159-181.

54 G.A. Orr, P. Verdier-Pinard, H. McDaid, S.A. Horwits. *Oncogene*, 22 (2003),

7280-7295.

55 A. Cormier, M. Marchand, R.B.G. Ravelli, M. Knossow, B. Gigant. *EMBO reports*, 9 (2008), 1101-1106.

56 K.M. Haskins, J. Donoso, R.H. Himes. *J. Cell. Sci.*, 47 (1981), 237-247.

57 A. Duflos, A. Kruczynski, J.M. Barret. *Curr. Med. Chem. Anticancer Agents*, 2 (2002), 55-70.

58 S. Gupta, B. Bhattacharyya. *Mol. Cell. Biochem.*, 253 (2003), 41-47.

59 J.J. Manfredi, J. Parness, S.B. Horwitz. *J. Cell. Biol.*, 94 (1982), 688-696.

60 H. Freedman, J.T. Huzil, T. Luchko, R.F. Luduena, J.A. Tuszynski. *J. Chem. Inf. Model.*, 424-436 (2009), 49.

61 S.B. Hastie. *Pharmacol. Ther.*, 512 (1991), 377-401.

62 S. Banerjee, G. Chakrabarti, B. Bhattacharyya. *Biochemistry*, 36 (1997), 5600-5606.

63 A. Lamberi, Y. Engelborghs. *J. Biol. Chem.* , 256 (1981), 3279-3282.

64 R. Bau, K.K. Jin. *J. Chem. Soc., Perkin Trans.*, 1 (2000), 2079-2082.

65 R.L. Noble, C.T. Beer, J.E.T. Curtis. *Biochem. Pharmacol.*, 1 (1958), 347-348.

66 I.S. Johnson, H.F. Wright, G.H. Svoboda. *J. Lab. Clin. Med.*, 54 (1959), 830-837.

67 S.S. Rai, J. Wolff. *J. Biol. Chem.*, 271 (1996), 14707-14711.

68 S.S. Rai, J. Wolff. *Proc. Natl. Acad. Sci. USA*, 95 (1998), 4253-4257.

69 G.C. Na, S.N. Timasheff. *Biochemistry*, 19 (1980), 1347-1354.

70 R.H. Himes. *Pharmacol. Ther.*, 51 (1991), 257-267.

71 R.C. Weisenberg, S.N. Timasheff. *Biochemistry*, 9 (1970), 4110-4116.

72 M.A. Jordan, D. Thrower, L. Wilson. *Canc. Res.*, 51 (1991), 2212-2222.

73 V.K. Ngan, K. Bellman, B.T. Hill, L. Wilson, M.A. Jordan. *Mol. Pharmacol.*, 60 (2001), 225-232.

74 B. Gigant, C. Wang, R.B.G. Ravelli, F. Roussi, M.O. Steinmetz, P.A. Curmi, A. Sobel, M. Knossow. *Nature*, 435 (2005), 519-522.

75 B. Gigant, A. Cormier, A. Dorléans, R.B.G. Ravelli, M. Knossow. *Top. Curr. Chem.*, 286 (2009), 1-20.

76 S. Rendine, S. Pieraccini, M. Sironi. *Phys. Chem. Chem. Phys.*, 12 (2010), 15530-15536.

77 S.J. Hubbard, J.M. Thornton. NACCESS, computer program, Department of

Biochemistry and Molecular Biology, University College London (1993).

78 E. Nogales, S.G. Wolf, K.H. Downing. *Nature*, 391 (1998), 199-203.

79 D. A. Case et al. *AMBER 9*. University of California, San Francisco, 2006.

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

81 C. Lee, W. Yang, R.G. Parr. *Phys. Rev. B*, 37 (1998), 785-789.

82 J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case. *J. Comput. Chem.* , 25 (2004), 1157-1174.

83 W.L. Jorgensen, J. Chandrasekhar, J. Madura, M.L. Klein. *J. Chem. Phys.*, 79 (1983), 926-935.

84 J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen. *J. Comput. Phys.*, 23 (1977), 327-341.

85 T. Darden, D. York, L. Pedersen. *J. Chem. Phys.*, 98 (1993), 10089-10092.

86 S. Pieraccini, G. Saladino, G. Cappelletti, D. Cartelli, P. Francescato, G. Speranza, P. Manitto, M. Sironi. *Nature Chemistry*, 1 (2009), 642-648.

87 V. Zoete, O. Michielin. *Proteins*, 67 (2007), 1026-1047.

88 B. Kuhn, P.A. Kollman. *J. Med. Chem.*, 43 (2000), 3786-3791.

89 A. Mitra, D. Sept. *Biochemistry*, 43 (2004), 13955–13962.

90 T. David-Pfeuty, C. Simon, D. Pantaloni. *J. Biol. Chem.*, 254, 22 (1979), 11696-11702.

91 K.R. Anders, D. Botsein. *Mol. Biol. Cell*, 12 (2001), 3973-3986.

92 K. Dai, A. Mukherjee, Y.Xu, J. Lutkenhaus. *J. Bacteriol.*, 176 (1994), 130-136.

93 C. Wang, A. Cormier, B. Gigant, M. Knossow. *Biochemistry*, 46 (2007), 10595-10602.

94 J. Löwe, L. Amos. *Nature*, 391 (1998), 203-206.

95 A. Mukherjee, J. Lutkenhaus. *J. Bacteriol.*, 176 (1994), 2754-2758.

96 S. Huecas, C. Schaffner-Barbero, W. Garcìa, H. Yébenes, J.M. Palacios, J.F. Dìaz, M. Menéndez, J.M. Andreu. *J. Biol. Chem.*, 282 (2007), 37515-37528.

97 J. Lutkenhaus, S.G. Addinall. *Annu. Rev. Biochem.* , 66 (1997), 93-116.

98 W. Margolin. *Nature Rev. Mol. Cell Biol.*, 6 (2005), 862-871.

99 Y. Chen, D.E. Anderson, M. Rajagopalan, H.P. Erickson. *J. Biol. Chem.*, 282 (2007), 27736-27743.

100 H.P. Erickson, D.W. Taylor, K.A. Taylor, D. Bramhill. *Proc. Natl. Acad. Sci.*

USA, 1 (1996), 519-523.

101 D. Popp, M. Iwasa, H.P. Erickson, A. Narita, Y. Maéda, R.C. Robinson. *J. Biol. Chem.*, 285 (2010), 11281-11289.

102 L. Romberg, M. Simon, H.P. Erickson. *J. Biol. Chem.* , 276 (2001), 11743-11753.

103 L. Romberg, P.A. Levin. *Annu. Rev. Microbiol.*, 57 (2003), 125-154.

104 D.-J. Scheffers, A.J.M. Driessen. *Mol. Microbiol.*, 43 (2002), 1517-1521.

105 M.A. Oliva, S.C. Cordell, J. Löwe. *Nat. Struc. Mol. Biol.*, 11 (2004), 1243-1250.

106 H.P. Erickson. *BioEssays*, 29 (2007), 668-677.

107 J.M. de Pereda, D. Leynadier, J.A. Evangelio, P. Chacòn, J.M. Andreu. *Biochemistry*, 35 (1996), 14203-14215.

108 E. Nogales, K.H. Downing, L.A. Amos, J. Löwe. *Nat. Struc. Bio.*, 5 (1998), 451-458.

109 J. Löwe, L. Amos. *EMBO journal*, 18 (1999), 2364-2371.

110 D.-J. Scheffers, J.G. de Wit, T. den Blaauwen, A.J. Driessen. *Biochemistry*, 15 (2002), 521-529.

111 A.K.W. Leung, E.L. White, L.J. Ross, R.C. Reynolds, J.A. DeVito, D.W. Borhani. *J. Mol. Biol.*, 342 (2004), 953-970.

112 S. Huecas, O. Llorca, J. Boskovic, J. Martìn-Benito, J.M. Valpuesta, J.M. Andreu. *Biophys. J.*, 94 (2008), 1796-1806.

113 J.F. Dìaz, A. Kralicek, J. Mingorance, J.M. Palacios, M. Vicente, J.M. Andreu. *J. Biol. Chem.* , 276 (2001), 17307-17315.

114 J. Stricker, H.P. Erickson. *J. Bacteriol.*, 185 (2003), 4796-4805.

115 M.A. Oliva, S. Huecas, J.M. Palacios, J. Martìn-Benito, J.M. Valpuesta, J.M. Andreu. *J. Biol. Chem.*, 278 (2003), 33562-33570.

116 A. Desai, T. Mitchison. *BioEssays*, 20 (1998), 523-527.

117 M.-J. Clément, B.T. Kuoch, T. Ha-Duong, V. Joshi, L. Hamon, F. Toma, P.A. Curmi, P. Savarin. *Biochemistry*, 48 (2009), 9734-9744.

118 D.J. Haydon et al. *Science*, 321 (2008), 1673-1675.

119 Alanis, A.J. *Archives Med. Res.*, 36 (2005), 697-705.

120 G. Taubes. *Science*, 321 (2008), 356-361.

121 Institute of Medicine (US) Forum on Microbial Threath. Antibiotic Resistance: Implications for Global Health and Novel Intervention Strategies: Workshop Summary. (Washington (DC) 2010), National

Academies Press (US).

122 D.N. Margalit, L. Romberg, R.B. Mets, A.M. Hebert, T.J. Mitchison, M.W. Kirschner, D. RayChaudhuri. *Proc. Natl. Acad. Sci. USA*, 101 (2004), 11821-11826.

123 K.W. Shimotohno, F. Kawamura, Y. Natori, H. Nanamiya, J. Magae, H. Ogata, T. Endo, T. Suzuki, H. Yamaki. *Biol. Pharm. Bull.*, 33 (2010), 568-571.

124 A.A. Handler, J.E. Lim, R. Losick. *Mol. Microbiol.*, 68 (2008), 588-599.

125 Y. Pommier, J. Cherfils. *Trends in pharmacological sciences*, 26 (2005), 138-145.

126 M.-J. Clément, I. Jourdain, S. Lachkar, P. Savarin, B. Gigant, M. Knossow, F. Toma et al. *Biochemistry*, 44 (2005), 14616-14625.

127 M. Clèment et al. *Biochemistry*, 44 (2005), 14616-14625.

128 A. Šali, T.L. Blundell. *J. Mol. Biol.*, 234 (1993), 779-815.

129 R.A. Laskowski, M.W. Macarthur, D.S. Moss, J.M. Thornton. *J. Appl. Cryst.*, 26 (1993), 283-291.

130 J.M. Davis, L.J. Tsou, A.D. Hamilton. *Chem. Soc. Rev.*, 36 (2007), 326-334.

131 Y. Che, B.R. Brooks, G.R. Marshall. *Biopolymers*, 86 (2007), 288-897.

132 R. Fasan, R.L.A. Dias, K. Moehle, O. Zerbe, J.W. Vrijbloed, D. Obrecht, J.A. Robinson. *Angew. Chem. Int. Ed.* , 43 (2004), 2109-2112.

133 P. Metrangolo, G. Resnati. *Halogen bonding: fundamentals and applications*. Springer, Berlin, 2008.

134 P. Metrangolo, G. Resnati. *Science*, 321 (2008), 918-919.

135 J.P.M. Lommerse, A.J. Stone, R. Taylor, F.H. Allen. *J. Am. Chem. Soc.*, 118 (1996), 3108-3116.

136 A.R. Voth, P. Khuu, K. Oishi, P.S. Ho. *Nat. Chem.* , 1 (2009), 74-79.

137 A.C. Legon. *Phys. Chem. Chem. Phys.*, 12 (2010), 7736-7747.

138 Y. Lu, Y. Wang, W. Zhu. *Phys. Chem. Chem. Phys.*, 12 (2010), 4543-4551.

139 P. Auffinger, F.A. Hays, E. Westhof, P.S. Ho. *Proc. Natl. Acad. Sci. USA*, 101 (2004), 16789-16794.

140 E. Parisini, P. Metrangolo, T. Pilati, G. Resnati, G. Terraneo. *Chem. Soc. Rev.*, 40 (2011), 2267-2278.

141 A.R. Voth, F.A. Hays, P.S. Ho. *Proc. Natl. Acad. Sci. USA*, 104 (2007), 6188-6193.

142 L.A. Hardegger, B. Kuhn, B. Spinnler, L. Anselm, R. Ecabert, M. Stihle, B. Gsell, R. Thoma, J. Diez, J. Benz, J.M. Plancher, G. Hartmann, D.W. Banner, W. Haap, F. Dietrich. *Angew. Chem. Int. Ed.*, 50 (2011), 314-318.

143 P. Metrangolo, G. Resnati. *Chem.-Eur- J.*, 7 (2001), 2511.

144 T.Imakubo, H. Sawa, R. Kato. *Synth. Met.*, 73 (1995), 117.

145 T. Imakubo, A. Miyake, H. Sawa, R. Kato. *Synth. Met.*, 120 (2001), 927.

146 R. Kato, T. Imakubo, H. Yamamoto, R. Maeda, M. Fujiwara, H. Sawa. *Mol. Cryst. Liq. Cryst.*, 380 (2002), 61.

147 E. Cariati, A. Forni, S. Biella, P. Metrangolo, F. Meyer, G. Resnati, S. Righetto, E. Tordin, R. Ugo. *Chem. Commun.* (2007), 2590-2592.

148 P. Politzer, P. Lane, M.C. Concha, Y. Ma, S. Murray. *J. Mol. Model.*, 13 (2007), 305-311.

149 P. Politzer, J.S. Murray, T. Clark. *Phys. Chem. Chem. Phys.*, 12 (2010), 7748-7757.

150 K.E. Riley, J.S. Murray, P. Politzer, M.C. Concha, P. Hobza. *J. Chem. Theory Comput.*, 5 (2009), 155-163.

151 S. Dapprich, I. Komaromi, S. Byun, K. Morokuma, M.J. Frisch. *J. Mol. Struct. (THEOCHEM)*, 461 (1999), 1-21.

152 T. Vreven, K.S. Byun, I. Komaromi, S. Dapprich, J.A. Montgomery Jr., K. Morokuma, M.J. Frisch. *J. Chem. Theory Comput.*, 2 (2006), 815-826.

153 T. Vreven, K. Morokuma, O. Frakas, H.B. Schlegel, M.J. Frisch. *J. Comput. Chem.*, 24 (2003), 760-769.

154 Y. Lu, T. Shi, Y. Wang, H. Yang, X. Yan, X. Luo, H. Jiang, W. Zhu. *J. Med. Chem.* , 52 (2009), 2854-2862.

155 R.W. Dixon, P.A. Kollmann. *J. Comput. Chem.*, 18 (1997), 1632-1646.

156 P. Cieplak, J. Caldwell, P.A. Kollmann. *J. Comput. Chem.*, 22 (2001), 1048-1057.

157 W.D. Cornell, P. Cieplak, C.I. Bayly, P.A. Kollmann. *J. Am. Chem. Soc.*, 115 (1993), 9620-9631.

158 S. Rendine, S. Pieraccini, A. Forni, M. Sironi. *Phys. Chem. Chem. Phys.* (2011).

159 E.E. Moliner, N.R. Brown, L.N. Johnson. *Eur. J. Biochem.*, 270 (2003), 3174-3181.

160 C. Mueller-Dieckmann, H. Ritter, F. Haag, F. Koch-Nolte, G.E. Schultz. *J. Mol. Biol.*, 322 (2002), 687-696.

161 M.H. Sazinsky, J. Bard, A. Di Donato, S.J. Lippard. *J. Biol. Chem.*, 279 (2004), 30600-30610.

162 R. Battistutta, M. Mazzorana, S. Sarno, Z. Kazimierczuk, G. Zanotti, L.A.

Pinna. *Chem. Biol.*, 12 (2005), 1211-1219.

163 B. Sandler, P. Webb, J.W. Apriletti, B.R. Huber, M. Togashi, S.T.C. Lima, S. Juric, S. Nilsonn, R. Wagner, R.J. Fletterick, J.D. Baxter. *J. Biol. Chem.*, 279 (2004), 55801-55808.

164 A.S. Nascimento, S.M.G. Dias, F.M. Nunes, R. Aparicio, A.L.B. Ambrosio, L. Bleicher, A.C.M. Figueira, M.A.M. Santos, M.D.O. Neto, H. Fischer, M. Togashi, A.F. Craievich, R.C. Garratt, J.D. Baxter, P. Webb, I. Polikarpov. *J. Mol. Biol.*, 360 (2006), 586-598.

165 K.H. Sippel, A.H. Robbins, R. Reutzel, J. Domsic, S.K. Boehlein, L. Fovindasamy, M. Agbandje-McKenna, C.J. Rosser, R. McKenna. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 64 (2008), 1172-1178.

166 T. Beck, A. Krasauskas, T. Gruene, G.M. Sheldrick. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 64 (2008), 1179-1182.

167 B.R. Huber, B. Sandler, B.L. West, S.T.C. Lima, H.T. Nguyen, J.W. Apriletti, J.D. Baxter, R.J. Fletterick. *Mol. Endocrinol.*, 17 (2003), 643-652.

168 D.A. Case et al. *AMBER 11*. University of California, San Francisco, 2010.

169 J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollmann, D.A. Case. *J. Comput. Chem.*, 25 (2004), 1157-1174.

170 M.J. Frisch et al. *Gaussian 09*. Gaussian Inc., Wallingford CT, 2009.

171 A.D. Becke. *Phys. Rev. A: At., Mol., Opt. Phys.*, 38 (1988), 3098-3100.

172 C. Adamo, V. Barone. *J. Chem. Phys.*, 108 (1998), 664-675.

173 Y.X. Lu, J.W. Zou, J.C. Fan, W.N. Zhao, Y.J. Jiang, Q.S. Yu. *J. Comput. Chem.*, 30 (2009), 725-732.

174 M.N. Glukhovtsev, A. Pross, M.P. McGrath, L. Radon. *J. Chem. Phys.*, 103 (1995), 1878-1885.

175 H.M. Senn, W. Thiel. *Angew. Chem. Int. Ed.*, 48 (2009), 1198-1229.

176 H. Lin, D.G. Truhlar. *Theor. Chem. Acc.* , 117 (2007), 185-199.

177 K. Senhilkumar, J.I. Mujika, K.E. Ranaghan, F.R. Manby, A.K. Mulholland, J.N. Harvey. *J. R. Soc. Interface*, 5 (2008), 207-216.

178 C. Biassantz, B. Kuhn, M. Stahl. *J. Med. Chem.*, 53 (2010), 5061-5084.

179 A.R. Voth, P.S. Ho. *Curr. Top. Med. Chem.*, 7 (2007), 1336-1348.

180 K.E. Riley, K.M. Merz. *J. Phys. Chem. A*, 111 (2007), 1688-1694.

181 Y. Lu, H. Li, X. Zhu, H. Liu. *J. Phys. Chem. A*, 115 (2011), 4467-4475.

182 Y. Lu, H. Li, X. Zhu, H. Liu, W. Zhu. *Int. J. Quant. Chem.* (2011).

183 C. Möller, M.S. Plesset. *Phys. Rev.*, 46 (1934), 618-622.

184 J.P. Perdew, K. Burke, M. Ernzerhof. *Phys. Rev. Lett.*, 77 (1996), 3865.

185 K.A. Peterson, B.C. Shepler, D. Figgen, H. Stoll. *J. Phys. Chem. A*, 110 (2006), 13877.

186 Dunning, T.H. *J. Chem. Phys.*, 90 (1989), 1007-1023.

187 Y.X. Lu, J.W. Zou, H.Q. Wang, Q.S. Yu, H.X. Zhang, Y.J. Jiang. *J. Phys. Chem. A*, 111 (2007), 10781.

188 V. Barone, M. Cossi. *J. Phys. Chem. A*, 102 (1998), 1995-2001.

189 S.F. Boys, F. Bernardi. *Mol. Phys.*, 19 (1970), 553-566.

190 R.F.W. Bader, M.T. Carroll, J.R. Cheeseman, C. Chang. *J. Am. Chem. Soc.*, 109 (1987), 7968-7979.

191 A. Bondi. *J. Phys. Chem.*, 68 (1964), 441-451.

192 S. Scheiner, T. Kar. *J. Phys. Chem. B*, 109 (2005), 16132-16141.

193 A.J.A. Aquino, D. Tunega, G. Haberhauer, M.H. Gerzabek, H. Lischka. *J. Phys. Chem. A*, 106 (2002), 1862-1871.