UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

INSTITUT CATALÀ D' INVESTIGACIÓ QUÍMICA

Computational studies on supramolecular hydrogen-bonded structures: from nanocapsules to proteins

Eva Santos Garcia

Dissertation presented to receive the degree of Doctor of the Universitat Rovira i Virgili

INSTITUT CATALÀ D' INVESTIGACIÓ QUÍMICA (ICIQ)
Tarragona, June 2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

INSTITUT CATALÀ
D'INVESTIGACIÓ
QUÍMICA

Avinguda Països Catalans 16
43007 Tarragona
Tel. 34 977 920 200
Fax 34 977 920 224
e-mail: iciq@iciq.es

Carles Bo Jané, cap de grup de l'Institut Català d' Investigació Química (ICIQ) i Professor Titular de Química Física de la Universitat Rovira i Virgili,

FA CONSTAR:

Que el present treball titulat: "Computational studies on supramolecular hydrogen-bonded structures: from nanocapsules to proteins", que presenta Eva Santos Garcia per a optar al grau de Doctor, ha estat realitzat sota la meva immediata direcció als laboratoris de l'Institut Català d´Investigació Química a Tarragona.

Tarragona, Juny de 2008

Carles Bo Jané

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

El trabajo desarrollado en esta tesis doctoral ha sido posible gracias a la financiación del Institut Català d'Investigació Química (ICIQ) y se ha desarrollado en el marco del proyecto *Diseño de Catalizadores para una Química Sostenible: Una Aproximación Integrada (INTECAT)* (CSD2006-0003) perteneciente al Programa Consolider-Ingenio 2010, y al proyecto CTQ2005-06909-C02-02 del Ministerio de Educación y Ciencia.

Consolider

ICIQ Institut Català d'Investigació Química

intecat

Consolider Ingenio 2010
CSD2006-0003
Diseño de Catalizadores para una Química Sostenible: una Aproximación Integrada

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Agradecimientos

En primer lugar deseo expresar mi sincero agradecimiento al Prof. Carles Bo Jané por haberme brindado la oportunidad de realizar el Doctorado bajo su dirección en el Institut Català d' Investigació Química. Quiero agradecerle su familiaridad, su ayuda y la calma con que ha lidiado mis momentos de ofuscación.

Agradezco al Prof. Javier de Mendoza que haya contado con nosotros para establecer colaboración en varios proyectos que en definitiva constituyen el presente trabajo. Por supuesto les debo también mi agradecimiento a Vera, a Julián, a Elisa, al Prof. Ernest Giralt y a Susana, que han llevado a cabo los estudios experimentales.

También quiero dar las gracias al Dr. Jordi Villà i Freixa y a la Dra. Mireia Garcia Viloca por haber confiado anteriormente en mí para la realización de una Tesis que finalmente no pudo llevarse a cabo.

Gracias a los muchos compañeros con los que he ido coincidiendo a lo largo de cuatro años en el laboratorio, por su ayuda, por los buenos ratos compartidos, por las risas y por las estampas irrepetibles (algunas de ellas inmortalizadas). En especial quiero dar las gracias a Pere Miró por sus innumerables y gratas ayudas desde que llegó hasta ahora, y a Alfons Nonell, con quien he cruzado trayectoria en varias ocasiones, por su apoyo en la búsqueda de nuevos horizontes.

A todas las personas que han alegrado, y espero que sigan alegrando, mi estancia en Tarragona.

A mi familia y a mis amigos. A Cyril y a Rosana, Marta, Pitu, Natalia y Pili por conocerme tanto, por su intensidad, su sentido del humor y su manera de ser felices.

A mis padres y a mis hermanos por estar siempre conmigo.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

A mis padres y a mis hermanos

De este árbol de Oriente,
a mi jardín venido,
un secreto sentido
su hoja guarda latente.

¿De un ser vivo se trata,
partido en dos mitades?
¿O son dos unidades
juntas de forma grata?

Pienso que es lo más noble
aunar dos universos:
¿no sientes en mis versos
que soy sencillo y doble?

("Ginkgo Biloba", Goethe)

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Table of contents

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Chapter I

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-0850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

I. Introduction

I.1. General concepts of hydrogen bonding

The first references to the hydrogen bond interaction appeared at the beginning of the XXth century. Although the concept of hydrogen bonding had been already used in some publications, it was Pauling who finally introduced the concept in a chapter of a book.[1] In 1943 Huggins proposed the existence of hydrogen bonding interactions in polipeptidic chains forming helices and sheets.[2] Since then a large amount of reports on this type of interaction in proteins as well as in many other systems increased substantially. Of remarkable importance was the Watson-Crick report on the DNA structure.[3] During the last decade of the XXth century the research on hydrogen bonding has experienced large progress due to the development of experimental techniques and computational studies.

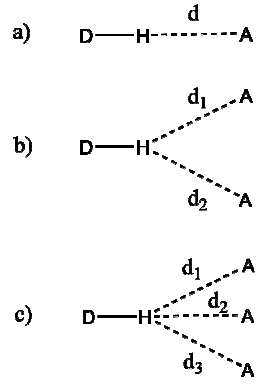


Figure 1. Hydrogen bonds a) bicentrate, b) bifurcate donor, c) trifurcate donor.

Along the time the definition of the hydrogen bond has been continuously changing, as its nature is even nowadays not completely characterized. The most up-to-date definition is the following: when a hydrogen atom is between two more electronegative atoms, D-H···A, the interaction is identified as hydrogen bond if there is evidence of a local bond between the D-H specie and the A specie, acting D-H as a proton donor and A as an acceptor. The interaction could be seen as an incipient proton transfer, as a H···A bond is established while the covalent bond D-H is weakened. H···A distance varies depending on D and A nature and can reach the 4Å. The angle can have a value between 90º and 180º, being this latter case the most favourable. Due to the long-range nature of the interaction the D-H group is able to interact with more than one acceptor (see Figure 1) and viceversa.

I.1.1. Classification of hydrogen bonds

Hydrogen bonds use to be classified in weak, intermediate and strong hydrogen bonds depending on certain properties (see Table 1). Obviously this classification serves only as guide, as the limits between these categories are diffuse. Many hydrogen bonds have properties in between.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Table 1. Properties of the Strong, Intermediate and Weak hydrogen bonds (D-H···A).

	Strong	Intermediate	Weak
Type of interaction	Strongly covalent	Very electrostatic	Electrostatic/Dispersion
Bonding distance	D-H ~ H···A	D-H< H···A	D-H<< H···A
H···A(Å)	1.2-1.5	1.5-2.2	2.2-3.2
D···A(Å)	2.2-2.5	2.5-3.2	3.2-4.0
Angle D-H···A(º)	175-180	130-180	90-150
Bonding energy(kcal·mol^{-1})	14-40	4-15	< 4
Relative displacement IR(cm^{-1})	25%	10-25%	< 10%
Chemical displacement H^1 (ppm)	14-22	< 14	---
Examples	-Dimers in gas phase of strong acids with strong bases. -Acid salts. -HF complexes	-Acids. -Alcohols. -Phenols. -Most of the HB existing in biological molecules.	-Dimers in gas phase of weak acids with weak bases. -C-H···O/N -O/N-H··π

Some hydrogen bonds are catalogued as non-conventional[4], as for instance the weak interaction of C-H···A type, the interaction with delocalized π systems D-H··π and the dihidrogen bonds such as BeH_2 ··· HCN, LiH ··· HF, AlH_3 ··· H_2O.

Besides the always existing changes in the D···A and the H··A distances, and the D-H-A angle, an elongation of the distance of the covalent bond D-H can also take place depending on the system.[5, 6] In very special cases the covalent angle Y-D-H is altered depending on the coordination angle.[7] In certain cases, for very strong hydrogen bonds, the covalent geometry of the molecular backbone no directly bound to the H atom is influenced by the hydrogen bond.

I.1.2. Methods for the study of hydrogen bonds

The experimental methods used in the study of hydrogen bonds are spectroscopy (Infrared (IR), Raman, Microwave and Nuclear Magnetic Resonance), diffraction (X-ray, neutron) and thermochemical methods.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Vibrational spectroscopy is the usual method for the study of hydrogen bonds in condensed phases. The method is based in establishing a relationship between the vibrational parameters obtained and the structure.[8] Also a relationship between vibrational parameters and bonding energy can be established.[9] However the main problem of vibrational spectroscopy is that it is notably more difficult to establish this kind of relationships in the case of weak hydrogen bonds. Microwave spectroscopy provides a measure of rotational constants, centrifugal distortion constants, nuclear quadrupoles and nuclear spin coupling.[10] From these values one can deduce the molecular geometry, the bonding energy, the force constants, the electric dipolar moments, the electric charge distribution and the electric quadrupole moment.

X-ray crystallography and neutron diffraction are very important and complementary in the determination of structures with hydrogen bonds.[11, 12] X-ray determines the position of the centroid of the electron density, which is not the position of the nucleus but it is located closer to the more electronegative atom. Neutron diffraction does determine the position of the nucleus. There are four important crystallographic databases where the information about the analyzed structures is registered: CSD[13](Cambridge Structural Database), PDB[14](Protein Data Bank), NDB[15](Nucleic Acids Data Bank) and ICSD[16](Inorganic Crystal Structure Database).

Thermochemical methods are based on calorimetric measurements or in the determination of equilibrium constants by means of a property sensitive to the formation of the hydrogen bond. By measuring the variations of the constant of formation of the hydrogen bond with the temperature, one can obtain the enthalpy (ΔH^o) and the entropy (ΔS^o) of the system.[17]

At the theoretical level, deformation density studies ($\Delta\rho = \rho(r) - \sum \rho_i(r)$) allow determining hydrogen bonding beyond mere geometry. The theory of "atoms in molecules" (AIM) can be used to characterise hydrogen bonding from the topology of the total charge density of a system.[18] The electron density $\rho(r)$ defines a scalar field which topology can be characterized by its number of critical points (points where the first derivative has a value of zero). As a consequence of the attractive capacity of the nucleus, $\rho(r)$ presents local maxima at these positions. As the distance from the nucleus is larger, the electron density decreases exponentially and goes to zero. If there is a bond between two atoms the electron density tends to accumulate in the internuclear region forming a maximum electron density path. Along this path there is a point in which the density presents its lowest value. This is a second order saddle point called "bond critical point". The presence of such a critical point gives evidence of the existence of a chemical bond and its position allows identifying the nature of the bonding.

Concerning the study of the energy parameters, several theoretical schemes have been reported to better rationalize the properties of hydrogen bonding. These methods partition the total interaction energy in contributions of diverse nature. The contributions considered by Morokuma's scheme[19] are electrostatic energy, polarization energy, exchange repulsion energy, charge transfer and dispersion. All of them are attractive except the exchange repulsion. The larger or slighter influence of each term depends on the donor, the acceptor, the geometry and the environment.

The application of density functional methodologies to hydrogen bonded systems can provide reference values for interaction energy excluding the effects of the solvent or the solid state. Moreover it permits to study not only the ground state systems but other geometries on the potential surface. It is also possible to study interconversion between different geometries and to observe the electron redistribution. Further analysis on the use of computational methods to study hydrogen bonds is made in Chapter II, in particular for the systems concerning the present thesis.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

I.1.3. Chemical influence of the hydrogen bond

The presence of hydrogen bonds affects the physical properties of gases, liquids and solids. For example, in organic compounds it tends to reduce the molar volume due to the larger strength of the hydrogen bonds compared to other intermolecular interactions. The melting points and boiling points change the tendency when hydrogen bonds are present.

In some systems the existence of hydrogen bonds originates cooperativity effects, which affects thermodynamical properties as the formation enthalpy. The effect is due to the formation of chains of the type D-H···D-H···A. In this case the strength is not additive but both of the hydrogen bonds turn into stronger due to the existence of larger polarization in comparison with the D-H···A case. The charge flows through the σ bonds D-H. Cooperativity can lead to long chains, clusters and complicated networks. As an example, a system composed of a chain of 4-pyridones[20] has been studied recently by DFT methods. The interaction enthalpy for a 4-pyridone dimer is of -9.9kcal·mol^{-1}. The interaction enthalpies for the hydrogen bonds rise as there are more 4-pyridones forming the chain, being the largest interaction energy that of the molecule in the centre of the chain (-22.2kcal·mol^{-1}). This illustrates how the coupling of a NH group with the C=O of an amide through a π-polarizable system can produce extraordinarily strong hydrogen bonds between neutral molecules. The polarization effect can be intra- or intermolecular.

I.2. The hydrogen bond in Nature

Non-covalent molecular interactions are key in Biochemistry, playing essential roles in the formation and folding of fundamental supramolecules such as proteins, nucleic acids and other numerous complexes with functions of vital importance in organisms. Furthermore, they are involved in processes such as the replication of DNA, changes in the three-dimensional conformations of proteins, specific recognition of substrates by enzymes and detection of molecular signals.

Indeed the structure of proteins plays a fundamental role in their capability to carry out biological functions. After the polypeptidic chains are formed from the information carried by the DNA, they fold into specific three-dimensional conformations which confer them their functionality. In numerous times the folded polypeptide needs to bind to other polypeptidic chains adapting its conformation to the final complex which is that able to perform certain functions. In the processes of folding, non-covalent interactions such as hydrogen bonding, ionic interactions, Van der Waals forces and hydrophobic packing are of crucial importance.

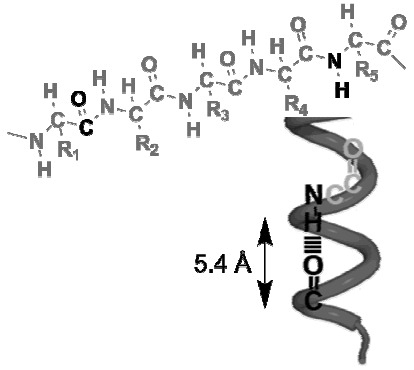


Figure 2. α-helix has 3.6 amino acids per turn of the helix, which places the C=O group of aminoacid 1 exactly directed towards the NH group of aminoacid 5.

The general three-dimensional form of local segments of a protein, or secondary structure, is determined by the possibility of establishing hydrogen bonds between backbone amide groups. A common motif in the secondary structure of proteins is the α-helix, a right-handed helix in which every C=O group in the backbone establishes a hydrogen bond with a N-H group in the backbone of an amino acid 4 residues next (see Figure 2). The also very common ß-sheet consists of polypeptidic chains connected laterally also by –NH···O=C- hydrogen bonds. The binding results in a conformation in a slightly twisted, pleated sheet (see Figure 3).

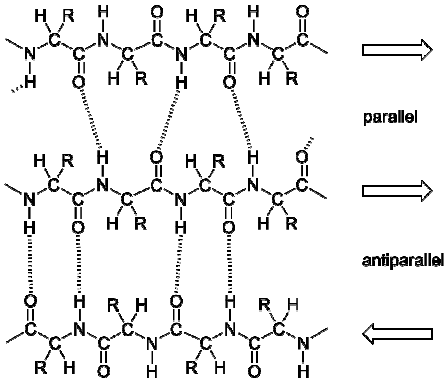


Figure 3. ß-sheets formed by establishment of hydrogen bonds between the NH groups in the backbone of a strand and the C=O groups in the backbone of the adjacent strand.

A clear example of the importance of protein structure concerning functionality is the partial or total loss of activity detected in some proteins with punctual mutations of one amino acid. The change of one residue for another can affect the secondary, the tertiary and the quaternary structures leading to a complex with different properties from the original one.

An example of mutation affecting protein functionality as a consequence of changes in the secondary structure is V382P mutation in the Serotonin transporter[21] (SERT). The SERT is responsible for the removal of serotonin from the synapse to the interior of the cell. It is composed of 12 transmembrane helices connected by intracellular and extracellular loops. A change of the Val 382 to a Proline in the 7th α-helix causes a disruption of the helix around this region due to the Proline being unable to form the typical α-helix hydrogen bonds due to steric hindering caused by the disposition of its sidechain. The helix distortion leads to loss of transport activity.

p53 transcription factor, formed by the union of 4 equal protein chains by means of hydrogen bonds and hydrophobic interactions, loses its activity as tumour suppressor under certain conditions if the residue Arg337, in the tetramerization domain, is mutated for a His.[22] The mutant protein is unable to establish intermolecular hydrogen bonds and hydrophobic interactions involving this particular residue. Although the rest of interactions remain untouched, they are not able to maintain the four chains bound with the same tertiary structure as in the wild type protein. Therefore the mutation affects p53 abilities and favours the appearance of tumours. This particular case will be described in detail in Chapter V and Chapter VI.

The secondary structure of nucleic acids is also determined by hydrogen bonds between the nitrogenated bases composing the strands. Not only this happens in double-stranded polynucleotides but also in those single stranded. For example, the stem-loop structure formed by bending of one part of an RNA strand to find a second part of the strand by forming hydrogen bonds is very

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

common. Furthermore this motif is a building block of larger structural RNA secondary structures as for instance the transfer RNA structure. Most of the biological processes in which DNA takes part involve hydrogen bonds between complementary nitrogenated bases (replication[23], transcription[24, 25], translation[26], etc) or between nitrogenated bases and other groups (DNA-protein/ligand interactions).

There are some examples of enzymes which mediate catalysis by means of hydrogen bonds, such as cholinesterases. Cholinesterases catalyse the hydrolysis of acetylcholine. They use a Glu-His-Ser triad to enhance the nucleophilicity of the catalytic serine, carrying a reaction like the schematized in Figure 4.

Transport of essential nutrients is also mediated in many cases by proteins which bind them by means of hydrogen bonding. As an example, phosphate-binding protein[27] (PBP) consists of a single polypeptide chain with one phosphate-binding site which holds the substrate tightly in place by establishing 12 strong hydrogen bonds between NH and OH groups of different residues with the phosphate oxygens. The receptor is highly specific. Indeed hydrogen bonds enable the protein to distinguish between phosphate and sulphate, this latter showing ineffective as a substrate or inhibitor.

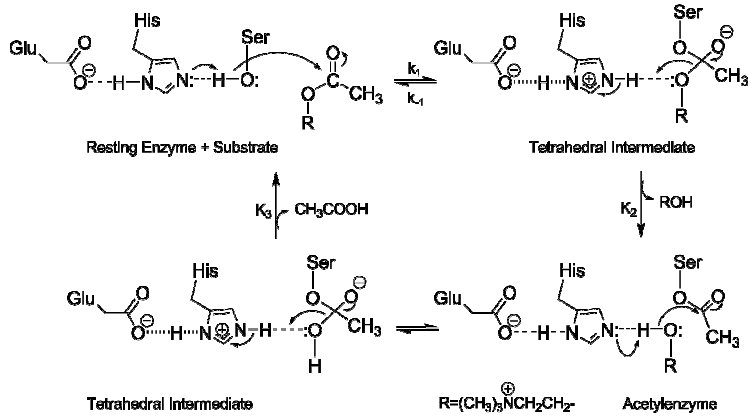


Figure 4. Scheme of action of acetylcholinesterase. Reproduced from ref [28].

I.3. The hydrogen bond in Supramolecular Chemistry

I.3.1. Introduction

Supramolecular Chemistry[29] mimics the cooperative non-covalent forces leading to highly specific molecular recognition processes in numerous natural phenomena, to construct large architectures by assembling smaller molecules (building blocks). It is possible to achieve complexes of nanometric dimensions without forming covalent bonds. The use of weak interactions permits the existence of certain reversibility thus leading to a dynamic assembling process: when wrong connexions take place, they can be broken and new different connexions can be established until the structure with lowest energy is reached. Moreover the final product is easily obtained as the formation of non-covalent interactions is fast. The protocol is highly convergent and has few steps comparing to a lineal synthesis. The final product is the thermodynamically favoured but it is still in equilibrium

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

with its building blocks. The reversibility of the interactions makes possible that small changes in the environment cause large effects in the multiple equilibrium existing in solution. The associations constructed can also be regulated in order to create intermolecular movement. This dynamic nature of the complexes is fundamental in Nanotechnology and for this reason Supramolecular Chemistry has become a fundamental basis in the development of Nanoscience.

I.3.2. Examples of supramolecular self-assembling

Some examples of complexes formed in Supramolecular Chemistry are briefly commented as follows.

Cyclical associations

In the cases in which the building blocks are able to form linear and cyclic aggregations it is crucial to take into account the enthalpic and the entropic contributions. The formation of cyclic complexes is entropically disfavoured but enthalpically favoured if a large number of hydrogen bonds is formed. In other cases the cyclic structures are preferred due to steric factors intervening although the number of hydrogen bonds formed is the same in both linear and cyclic complexes (see Figure 5).

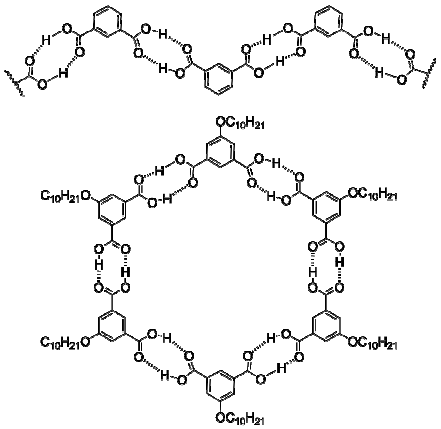


Figure 5. A slight modification in the isophthalic acid leads from a linear supramolecular structure to a cyclic structure. Reproduced from ref [30].

Infinite bidimensional associations

The synthesis of infinite bidimensional associations can be directed by means of steric control.

Figure 6 shows two different products both obtained by self-assembling of the barbituric acid and melamine derivative. In a) the melamine has a fluor atom as substituent in para position and in b) it has an etil ester as substituent in the same position.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

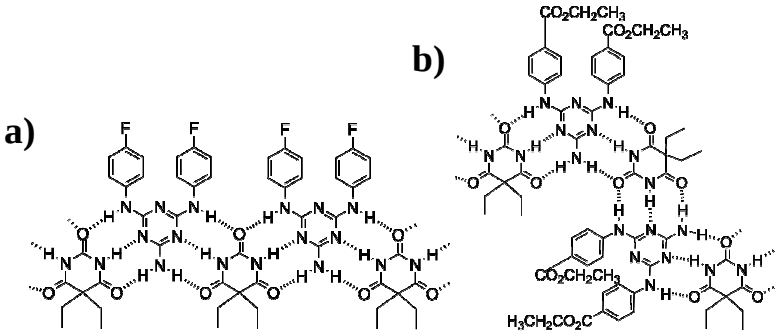


Figure 6. Complexes formed by a) barbituric acid (reproduced from ref [31]) and b) melamine derivatives (reproduced from ref [32]).

Cylindrical associations

These kinds of associations are used as ion carriers or molecular carriers. Figure 7 shows one of the earliest examples of a self-assembling ion channel.[33] It is composed of two structures formed by a cyclodextrin and four lipophilic arms. Embedded in a membrane, the structure is able to transport Co^{2+} trough the channels formed.

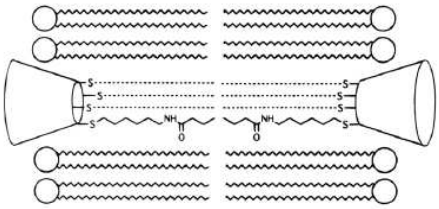


Figure 7. Supramolecular structure embedded in a membrane allowing Co^{2+} transport.

Helical associations

The DNA double helix has inspired the design of this type of structures. In this case hydrophobic interactions and metallic ions are needed in addition to hydrogen bonds in order to support the complexes. Figure 8 shows building blocks based on tartaric acid that self-assemble in helicoidal complexes. When the building blocks are of the L form a left-handed helix is formed while when they are of the D form a right-handed helix is formed instead.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

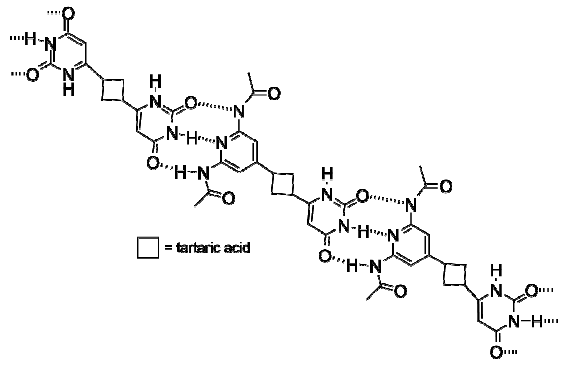


Figure 8. Helicoidal self-assembling of structures based on tartaric acid. Reproduced from ref [34].

Receptors

The so-called receptors form a cavity in which other molecules can be trapped (see Figure 9). These structures are used as sensors, as containers-catalysts in certain reactions or as molecular electronic devices. The major applications pursued for the future are related to drug transport through membranes and catalysis with prevention of the inhibition of the product.

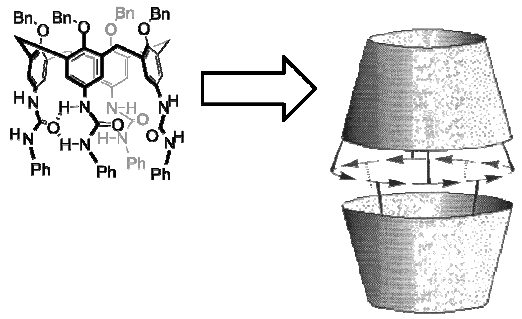


Figure 9. Cavity formed by two calixarenes. The complex is shown schematically with arrows representing ureas hydrogen-bonded in a head to tail topology. Reproduced from ref [35].

In this thesis a new capsule which is able to trap fullerenes inside will be described in Chapter IV. The capsule is formed by the self-assembly of two molecules composed, each of them, of a cyclotriveratrylene moiety and three ureidopyrimidinone units. The structure is supported by 12 hydrogen bonds.

I.4. Importance of theoretical approaches for a correct treatment of hydrogen bonds

As previously mentioned, hydrogen bonds are present in many biological and synthetic macromolecular complexes of interest. At the experimental level it is not obvious to understand the

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

nature of the interaction. Therefore the development and application of theoretical approaches to this kind of systems becomes necessary for their fair simulation and good understanding in combination with information obtained from experiment. Quantum Mechanics based methods provide a quite adequate treatment of hydrogen bonding. However, the correct choice of a method depends on the size and nature of the structures treated. For large macromolecules the application of any Quantum Mechanics methods becomes excessively time-spending to be feasible and Molecular Mechanics have to be applied. Comparison of available experimental results with theoretical results is always recommended. In the present thesis different methods were used for the study of diverse hydrogen-bonded systems, pursuing agreement with experimental evidences.

I.5. Thesis plan and objectives

The results presented in this thesis describe computational studies on various systems where hydrogen bonding is of critical importance. The diversity of the systems studied here was thought to be of importance to probe the accuracy of the methods, and the results were constantly compared with the experimental data that were procured by the group of Prof. Javier de Mendoza, in the context of a close collaboration within the ICIQ centre. In some cases the theoretical results were useful to explain experimental evidences and in other cases, to make predictions.

In Chapter II, a summary of the general concepts of Density Functional Theory and Molecular Mechanics is presented, mainly focused on the methods used in this work. The common problems observed when treating hydrogen bonds by traditional quantum chemistry methods and DFT methods are described, and the approaches available to overcome these problems are discussed.

Chapter III presents a study on the tautomeric equilibrium established from 2-ureidopyrimidinone (UPy). In this equilibrium four species coexist, two of them able to form dimers by self-assembly. The dimers are formed by arrays of four hydrogen bonds of the Donor-Donor-Acceptor-Acceptor and Acceptor-Donor-Acceptor-Donor types. The two dimers are compared in strength in the hypothetical case they were isolated and also in the case they were forming part of the entire tautomeric equilibrium. The effect of two different substituents (CH_3 and CF_3) in a particular position of the pyrimidone ring is studied to explain the equilibrium shifts caused by them, towards one dimer or the other. Some particular DFT methods are tested for the treatment of this kind of systems and the methods of choice are applied to obtain dimerization energies and thermodynamic parameters.

Chapter IV deals with systems also based in dimers of ureidopyrimidinone. A capsule is described formed by self-assembly of a molecule composed of one cyclotriveratrylene moiety and three UPys bound to it, leading to a structure containing three dimers of the Donor-Donor-Acceptor-Acceptor type. It was experimentally proved that this capsule served as a host for fullerenes, showing higher affinity for some of them and thus allowing an easy separation of fullerenes of different size. The theoretical study was devoted to the identification of the most stable configuration and conformation of the capsule among several possible and to the explanation and prediction of its preferences. The same DFT methods used in the previous chapter are applied in this case in spite of the much larger size of the complexes studied. However, additional empirical corrections to fairly take into account the new van der Waals interactions between the capsule and the fullerene are introduced.

Chapter V presents the study of a punctual mutation in the p53 protein tetramerization domain (p53TD). The larger size of this system makes its treatment by DFT methods unattainable. The change of an arginine by a histidine in the position 337 of the protein was experimentally proved to cause a destabilization of the tetrameric structure under certain conditions, thus preventing its action

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

as tumour suppressor. The process of disruption caused by the lost of hydrogen bonding and hydrophobic interactions is studied by means of Molecular Dynamics. Thermal denaturing conditions are applied to see an evident difference between the wild type protein and the mutant protein, since at standard conditions the differences are subtle.

In Chapter VI the behaviour of several synthetic ligands binding the wild type and the mutant p53TD surface is described. A tetraguanidinium, a hexaguanidinium and two different calix[4]arene ligands are tested in order to see if they are able to stabilize the mutant protein in a conformation similar to that of the wild type protein. The methods used are the same as in the previous chapter.

Chapter VII describes the dynamic unspecific interaction between the phosphate backbone of a DNA molecule and an undecaguanidinium ligand, also with a predictive aim. Molecular Dynamics are applied in this case too.

In Chapter VIII the conclusions of the overall thesis are summarized.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Chapter II

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

II. Theoretical treatment of systems with hydrogen bonds

Computational Chemistry has become a very important tool for understanding hydrogen bonding. For simple molecules containing elements from the first and second row of the periodic table the most used computational methods are based on ab-initio molecular orbital calculations. Ab-initio calculations consist of pursuing an approximate solution to the exact solution of the Schrödinger equation. The closer the result is to the exact solution the more time-consuming the calculations are. Thus the choice of a particular method compromises the quality of the results and the time spent for doing the calculation. To simulate molecules containing thousands of atoms, as for instance proteins or DNA, methods based on classical physics (Molecular Mechanics) are primarily used, as ab-initio calculations are out of the computational possibilities for such large systems.

The Born-Oppenheimer approximation states that as nuclei have much larger mass than electrons, so their velocity is much smaller. Thus the nuclei can be considered static and the electrons can be considered like charge clouds that adapt instantaneously to whatever position of the nuclei. In Quantum Chemistry, systems are often studied by fixing the position of the nuclei and studying the electronic structure in these conditions. In Molecular Mechanics the Born-Oppenheimer approximation is also applied. In this case the motions of the nuclei are studied and the electrons are not explicitly examined, but are assumed to be optimally distributed around the nuclei.

II.1. Quantum Mechanics

Quantum Mechanics[36-38] (QM) is the correct mathematical description of the behaviour of electrons. In theory, it can predict any property of an individual atom or molecule exactly. However, QM equations have only been solved in exact form for one electron systems. For multiple electron systems there are several approaches which approximate the solution. Most of the approaches pursue the solution of the time-independent Schrödinger equation (Eq. 1). In Eq. 1, $\hat{H}$ is the Hamiltonian operator for a molecular system consisting of M nuclei and N electrons in the absence of magnetic and electric fields.

$$\hat{H}\,\Psi_i(\vec{x}_1,\vec{x}_2,...,\vec{x}_N,\vec{R}_1,\vec{R}_2,...\vec{R}_M) = E_i\Psi_i(\vec{x}_1,\vec{x}_2,...,\vec{x}_N,\vec{R}_1,\vec{R}_2,...\vec{R}_M)$$

Eq. 1

II.2. Density Functional Theory

II.2.1. Principles

The Density Functional Theory[38] (DFT) represents an alternative way to the conventional ab-initio methods of considering a polielectronic system furthermore introducing the effects of the electron correlation in the solution to the Schrödinger equation. DFT is based on two theorems formulated by Hohenberg and Kohn.

The first Hohenberg-Kohn theorem[39] states the existence of a one-to-one mapping between the ground state electron density ρ_0 and the ground state wave function of a many-particle system. The second theorem[39] proves that ρ_0 minimizes the total electronic energy of the system. Thus, DFT

uses the electron density instead of the wavefunction to obtain the energy. However, the exact mathematical formula that relates the energy with the electron density is not known so that approximations are needed.

II.2.2. The Kohn-Sham approach

The Kohn-Sham approach[40, 41] allowed bringing the Hohenberg-Kohn theorems to practice. The approach consisted in reducing the intractable many-body problem of interacting electrons in a static external potential to a tractable problem of non-interacting electrons moving in an effective potential.

Thus, Kohn and Sham defined the functional as shown in Eq. 2, in which Ts[ρ] is the kinetic energy in a non-interacting system, J(ρ) is the classical Coulomb interaction and E_{XC} is the exchange-correlation energy, which contains all the unknown terms.

$$F[\rho(\vec{r})] = Ts[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{XC}[\rho(\vec{r})]$$

Eq. 2

E_{XC}, the Exchange-correlation energy, is defined through Eq. 3. The residual part of the true kinetic energy Tc, which is not covered by Ts, is simply added to the non-classical electrostatic contributions. Thus E_{xc} is the functional that contains everything that is unknown.

$$E_{XC}[\rho] = T_c[\rho] + E_{ncl}[\rho]$$

Eq. 3

II.2.3. Approximations to the E_{XC} term

As follows, the different strategies existing to obtain $E_{XC}[\rho]$ are briefly commented.

Local Density Approximation (LDA)

The central idea of LDA[40, 42] is a hypothetical uniform electron gas.[39] In this system the electrons move around a distribution of positive charges that make the system electrically neutral. The number of electrons N and the volume V of the gas are considered to approach infinity, while the electron density N/V remains finite at every point in space. The E_{XC} energy is expressed as a functional of the density, considering the exchange-correlation effects as local, i.e., only dependent on the electron density value in each point in space (Eq. 4).

$$E_{XC}^{LDA}[\rho(\vec{r})] = \int \rho(\vec{r}) \varepsilon_{XC}(\rho(\vec{r})) d\vec{r}$$

Eq. 4

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

ε_{XC} is usually divided into the exchange contribution (ε_X) and the correlation contribution (ε_C). The ε_X term representing the exchange energy of one electron in a uniform electron gas at one particular density, is known. For the correlation term ε_C there is no known analytic expression but very accurate Monte-Carlo numerical simulations[43] exist over this ideal system. Based on these simulations several authors presented approximate expressions. One of the most popular is the one developed by Vosko, Wilk and Nusair, also known as VWN.[44]

Generalized Gradient Approximation (GGA)

The GGA methods introduce the density gradient as well as the electron density at each point in space. The corrections from the gradient are added to the local definition of exchange and correlation by redefining E_{XC}. The expression for this approach is shown in Eq. 5.

$$E_{XC}^{GGA}[\rho(\vec{r})] = \int \rho(\vec{r})\varepsilon_{XC} d\vec{r}$$
$$\varepsilon_{XC} = f(\rho, \nabla\rho)$$

Eq. 5

Again the exchange and correlation terms can be considered as independent and this gives rise to the different exchange and correlation functionals.

Nowadays one of the most successful combinations of GGA functionals is the BP86 functional, applied herein in Chapter III and Chapter IV. It is composed by the Becke88 exchange functional[45] and Perdew86 correlation functional.[46, 47] The exchange functional introduces a parameter that was determined in order to reproduce the exchange energies of the noble gases, which are known. The correlation functional proposed by Perdew introduces an empirical parameter fitted to reproduce the correlation energy of the Ne atom and is a correction of the VWN functional.

Hybrid functionals

The hybrid functionals were formulated[48] because the principal error of the LDA and GGA functionals is the exchange description. Trying to solve this problem hybrid functionals incorporate a part of the Hartree-Fock (HF) exact energy of exchange. Depending on the quantity of exact exchange introduced, on the introduction of empiric parameters, on the correlation functional and on the exchange functional there are around 20 different hybrid functionals in use at present (see below).

II.3. Hydrogen bonds and the dispersion energy problem

In hydrogen bonding interactions the donor group A-H and the acceptor group B are generally closed-shell and in their electronic ground state. Thus it is not obvious how the hydrogen atom could be involved in more than one bond with its single 1s valence orbital. Many difficulties exist even at the experimental level to understand the nature of the interaction. The development of energy partitioning schemes, as the scheme of Morokuma[19] provides theoretical means to understand the peculiarities of hydrogen bonding. In Morokuma's scheme the contributions considered for the total interaction energy are: an electrostatic component, a polarization component, a exchange-repulsion component, a charge-transfer component and a dispersion term. A

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

theoretical method to treat hydrogen bonds would be that providing an appropriate description of all these contributions.

The most important problem when treating hydrogen bonds with HF theory is that attractive dispersion interaction is not considered as it is a pure correlation effect. Moreover the use of highly flexible basis sets is needed including diffuse functions. In correlated calculations very large basis sets are needed in order to avoid an underestimation of the interaction energy due to the correlation effects being larger in the interacting complex than in the fragments.

In the context of the traditional quantum chemistry based on the wavefunction a large variety of methods has been developed in order to take the electron correlation into account. Second (MP2) and fourth (MP4) order Moller Plesset Perturbation theory[49] are used in calculating small systems. Higher level MP calculations are rarely used because of their computational cost. Other popular methods are based on the Configuration Interaction (CI), on the Quadratic Configuration Interaction[50] (QCI) and on the Coupled Cluster[51] (CC) approximations. DFT methods represent much cheaper methodology to treat hydrogen bonding and therefore make possible the study of larger systems. With DFT, thermodynamical, structural and vibrational properties of a large variety of systems containing hydrogen bonds can be reasonably described with gradient-corrected and hybrid density functionals.

Dispersion forces (or London forces) are long-range attractive forces. They are the weakest intermolecular forces, caused by the formation of temporary dipoles in two adjacent atoms because of the electrons being unsymmetrically distributed around the nucleus. The formation of an induced dipole in an atom induces in turn the formation of a dipole in a nearby atom. Dispersion forces are present between all molecules, polar or non-polar. In the asymptotic limit this induced dipole-induced dipole attraction decays with the inverse sixth power of the intermolecular distance. The effect is exclusively due to electron correlation so, as commented, the Hartree-Fock model does not take it into account.

Some new hybrid functionals were fitted to reproduce properties of H-bonded complexes and π-stacking interactions. B3LYP[52, 53], X3LYP[54], BH&H[55, 56], MPWB1K[57] and M06[58] are some of them, giving results comparable to those obtained with MP2. However they still have some difficulties, more or less evident depending on the properties of the system treated.

The functional B3LYP is comprised of Becke's three-parameter exchange functional[52] and the Lee-Yang-Parr correlation functional[59]. It seems to be able to succeed in predicting geometries and in describing the electrostatic, exchange and induction components of the hydrogen bonding interactions. However, it still fails for some systems in giving accurate interaction energies due to the neglecting of dispersion forces[60]. X3LYP is an extended hybrid functional combined with Lee-Yang-Parr correlation functional. Firstly it was introduced as a functional to improve accuracy for H-bonds and van der Waals complexes[61] but later was found not to be as good for the latter interactions when studying stacked complexes of nucleic acid base pairs[62]. So far, the most promising results for such π-stacking interactions have been obtained with the BH&H functional[48, 55], which contains an equal mixture of the exact Hartree-Fock exchange and local-density approximation for the description of exchange energy, coupled with Lee, Yang and Parr' s expression for the correlation energy.

Anyway as the suitability of the different methods depends on the system under study it is always essential to carry out some previous calculations on that particular system or similar that can be contrasted to available results obtained experimentally or by means of high-level calculations.

The particular systems studied in the present work are systems with arrays of quadruple hydrogen bonds of the same nature as those present in the acid base pairs adenine-thymine and guanine-

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

cytosine (see Figure 10a and Figure 10b), with two and three H-bonds respectively. For the latter A-T and G-C systems, geometries and bond type enthalpies are obtained in excellent agreement with experiment when using BP86 functional, especially in combination with TZ2P basis set but also with DZP. Thus, in principle, BP86/DZP, BP86/TZP or BP86/TZ2P should be good choices for some of the systems studied in the present work, with arrays of quadruple H-bonds as the shown in Figure 10c. Concretely BP86/TZP was the choice but also B3LYP with 6-311G** basis set, in principle equivalent to TZP, was applied giving results in nice agreement with the former method. However for complexes as the presented Chapter IV (Figure 10d) these methods would not be accurate enough. In this case, besides the existence of hydrogen bonds, there exist van der Waals interactions established between large host and guest. In such a big system the dispersion interaction is considerable and not well reproduced by any of the methods mentioned above.

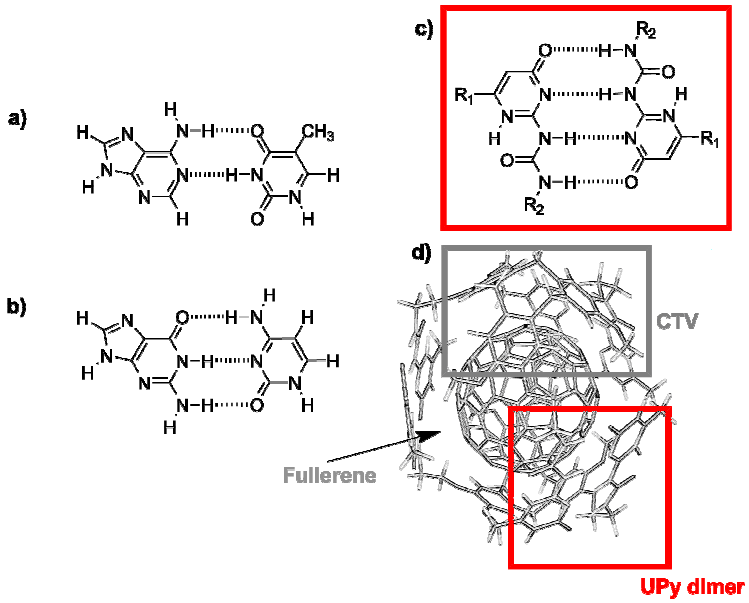


Figure 10. Systems with multiple H-bonds. a) Adenine-Thymine; b) Guanine-Cytosine; c) UPy dimer; d) Supramolecular complex with three UPy dimers (12 H-bonds) and van der Waals host-guest interactions.

The most practical and successful approaches that take dispersion forces into account are not based in hybrid functionals empirically fitted, but by adding an empirical correction term.

In 2004 Grimme[63] reported on an empirical method to account for van der Waals interactions in DFT calculations using BLYP and PBE functionals. Eq. 6 shows the description used for the dispersive energy E_{disp}. It is expressed as a sum of diatomic ***ij*** contributions. $\frac{C_6^{ij}}{R_{ij}^6}$ is the classical London-type term which allows to correct the long-range behaviour of the functionals. $f_d(R_{ij})$ is a damping function needed to switch off the interaction at short distances (see Figure 11), since the functionals work well in this case.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

$$E_{disp} = -\sum_i \sum_j f_d(R_{ij}) \times \frac{C_6^{ij}}{R_{ij}^6}$$

Eq. 6

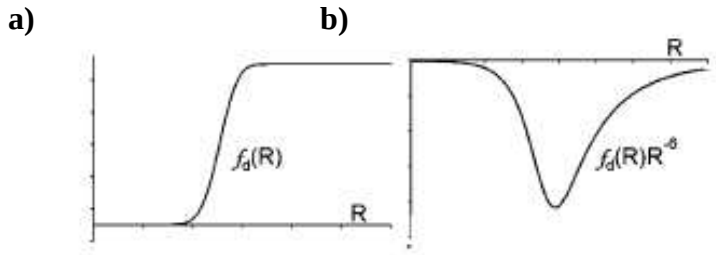


Figure 11. Shape of the damping function $f_d(R)$ (left) and of damped R^{-6} (right). Figure reproduced from ref [64].

Grimme used global C_6 scaling factors for some DFT functionals. The method was tested for a variety of 29 complexes and, for stacked aromatic systems and base pairs, and it seemed to be even superior to standard MP2. In 2007 Grimme et al. reported on a study about onion-like structures of fullerenes[65], in which they compared the computed interaction energy between two sheets with experimental values. All-electron GGA density functional methods together with the empirical correction for dispersion forces (DFT-D) gave very accurate results. The methods uncorrected gave reliable results concerning exchange-repulsion, electrostatic and induction contributions. However, only when the dispersion correction was added, attractive interaction energy between fullerenes was obtained.

More recently Cavallo[64] reported on a method which followed Grimme's approach but which introduced new parameters in the damping function in order to gain versatility. Thus, the expression adopted for the damping function was the one in Eq. 7.

$$f_d(R_{ij}) = \frac{c}{1 + \exp(-a(R_{ij} - bR_0))}$$

Eq. 7

R_0 is the sum of the Bondii's radii[66] of atoms ***i*** and ***j***. The parameters added with respect to Grimme approach were ***b*** and ***c***. Whereas Grimme derived the radii from HF calculations, Cavallo scaled the values by an additional parameter to allow the use of any kind of radii. ***c*** is a parameter to be fitted depending on the specific functional and basis set used. ***a*** is the steepness of the function. The parameter values are determined empirically. The correction for the BSSE is included into E_{disp}.

The C_6 parameters were obtained using the approximation shown in Eq. 8, where α_i is the polarizability of molecule ***i***. The polarizabilities proposed by Miller[67] were used.

$$C_6^{ij} = \frac{2C_6^{ii}C_6^{jj}\alpha_i\alpha_j}{C_6^{ii}\alpha_j^2 + C_6^{jj}\alpha_i^2}$$

Eq. 8

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Cavallo and co-workers parametrized a set of damping functions to correct the treatment of dispersion interactions of the Perdew-Burke-Ernzerhof[45, 68] (PBE), the Becke-Perdew[45-47] (BP86) and the Becke-Lee-Yang-Parr[45, 48, 59, 69] (BLYP) functionals. They worked, among others, on the ADF[70-73] implementation of the BP86 functional in connection with the TZP basis set. Then BP/TZP method is thus suitable for treating complexes such as the shown in Figure 10d by adding the calculated term $\boldsymbol{E_{disp}}$ to the energy. The approach was tested to reproduce the interaction energies of the JSCH-2005 benchmark[74] and CCSD(T) complete basis set limit stacking energy between some nucleic base pairs reported by Sponer et al.[75]

II.4. Molecular Mechanics

II.4.1. Introduction

Molecular mechanics[76-79], also known as Force Field methods, is based on the following principles:

1- Nuclei and electrons are grouped into atom-like particles.
2- Atom-like particles are spherical and have a net charge obtained from theory. The radii are obtained from measurements or theory.
3- Interactions are based on springs and classical potentials.
4- Interactions must be pre-assigned to the specific sets of atoms.
5- Interactions determine the spatial distribution of atom-like particles and their energies.

The advantage of Molecular Mechanics is that it allows modelling of huge molecules.[80] The disadvantage is that there are many chemical properties that cannot be studied, as for example electronic excited states or the evolution of a reaction, in which chemical bonds can not be broken.

The mathematics of spring deformation is used to describe ability of bonds to stretch, bend, and twist. The total energy associated to a given conformation of a molecule is calculated by summing different energy terms: stretching energy, bending energy, torsion energy and non-bonded interaction energy (see Figure 12). Non-bonded interactions consist of van der Waals attraction, steric repulsion and electrostatic attraction/repulsion.

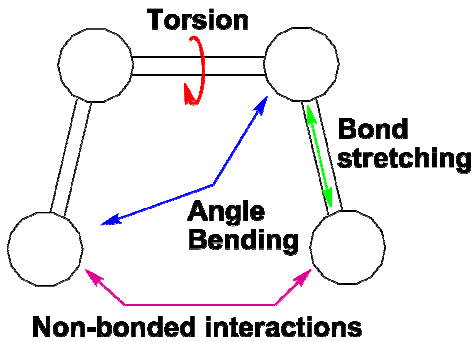


Figure 12. Energy terms contributing to the total energy associated to a given conformation of a molecule.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

A specific Force Field is composed of particular equations to calculate the different energy terms and particular parameters used to describe the behaviour of the different kinds of atomtypes and bonds. In order to get parameters transferable from one molecule to another Force Fields use atom types. For instance a sp^3 carbon is described by different parameters than sp^2 carbon. In some force fields, the united atoms approach simplifies the complexity of the calculations by omitting most of the hydrogen atoms. They do not have parameters to describe the backbone atoms alone but together with the corresponding attached hydrogens. For example, a CH_2 is described as a whole group instead of a carbon bound to two hydrogen atoms.

The equations to calculate the independent energy terms slightly differ for each Force Field. As follows, those in AMBER[81] force field, which is the all-atom used in the present thesis, are described. AMBER is the acronym for "Assisted Model Building with Energy Refinement". Parameters for this force field were extensively optimized with particular emphasis on the treatment of proteins.

The functional form for the potential energy in AMBER is the shown in Eq. 9.

$$E(r^N) = \sum_{bonds} \frac{1}{2} K_b (l - l_0)^2 + \sum_{angles} \frac{1}{2} K_b (\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)]$$
$$+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ 4 \in_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi \in_0 r_{ij}} \right\}$$

Eq. 9

The potential energy is a function of the position of the atoms (***r***) of ***N*** particles.

II.4.2. The partial atomic charges

For some types of molecules experimental data can be poor or non-existing. Quantum mechanics are increasingly used in this sense providing parameter values otherwise unavailable. The parameters composing the force field influence each other so that strategies for gradually refine them in an iterative procedure are essential to get a good fit with experimental data.

The partial atomic charge is not an experimentally observable quantity and cannot be definitely established from the wave function. That is why force fields use charges determined by so different methods. For Amber this is done using the Restrained Electrostatic Potential Method (RESP[82]). The protocol consists of, first of all, making out a quantum mechanics calculation at level HF/6-31G (d) to obtain a grid of multiple point charges from calculated electrostatic potential (ESP) fit. Once the grid obtained, restraints are applied to redistribute the multiple point charges to few points that correspond to the atoms, taking into account the symmetry of the molecule.

Like van der Waals terms, electrostatic terms are long range so that they are computed for non-bonded atoms in a 1-4 relationship or further apart. Thus they dominate the computation time. The number of non-bonded interactions grows quadratically with molecule size. The computation time can be reduced by cutting off the interactions after a certain distance. In the present work a 10Å cutoff was chosen for van der Waals interactions. The value is small enough to not consider the

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

interaction of a particle with its own image and sufficiently large to take into account all the relevant interactions. This approximation is suitable because van der Waals terms die off relatively fast (with r^{-6}). However, the electrostatic terms die off slower (r^{-1}) and therefore are difficult to fairly treat with cutoffs. Using this method, interactions are underestimated, especially when charged groups are present. One possible solution would be to increase the size of the simulation cell but this would also mean a large increase of simulation time. Multiple methods have been developed in order to solve the problem.

The idea of Ewald Summation Method[83, 84] is to transform the calculation of the sum of all the possible electrostatic interactions in two other terms of faster convergence.

$$E_{el} = Ee_{dir} + Ee_{rec} + E_0$$

Eq. 10

$\boldsymbol{Ee_{dir}}$ refers to the *direct space*, $\boldsymbol{Ee_{rec}}$ to the *reciprocal space* and $\boldsymbol{E_0}$ is a constant. The short-range interactions (direct space) are calculated by means of Coulombic potential modified whereas the rest are calculated via vector summation. In the calculation of the direct space gaussian distributions are used to represent the original charges in order to get a faster convergence. These distributions are afterwards compensated by applying gaussian distributions of opposite charge in the reciprocal space. The Particle Mesh Ewald[85] (PME) is a modified version of the Ewald summation, in which the calculation in the reciprocal space is optimized by means of the use of Fourier transforms.

II.4.3. Water models

Many hypothetical models for water have been developed in order to reproduce the physical properties of liquid water (see http://www.lsbu.ac.uk/water/models.html as reference). Each model is developed to fit well some particular physical parameter (for example the density, the radial distribution function or the critical parameters). The more fitting parameters are required by the model, the better the fit.

In the present thesis Simple Point Charge model[86, 87] (SPC) has been used to simulate explicit water. SPC water is rigid and has three centres of concentrated charge: two positive charges on the H atoms and a negative charge on the O atom.

The van der Waals interaction between two water molecules is calculated using a Lennard-Jones function. The interaction is measured considering two points, each of them centred on an O atom.

Because of the use of point charges the value for the permanent dipole moment when the canonical geometry is used, does not agree with the experimental value. Thus, in order to obtain the correct value the angle is changed to 109.42º instead of the experimentally established 104.45º.

II.4.4. Classical Molecular Dynamics

Molecular dynamics[76-79],[88, 89] refers to techniques that enable simulating the behaviour of a system along the time. Being impossible to identify all the minima and saddle points on the potential surface of a large molecule, Molecular dynamics (MD) aims to obtain the information reachable from a determined starting point to apply statistics, as the possible energy levels and conformations.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Although nowadays some quantum molecular dynamics methods are available (i. e. car-Parrinello) the size of the molecules studied here forces us to use cheap energy evaluation methods such as force fields. The evolution of the classical MD as described in Gromacs[90] is the following:

1-Input initial conditions
The initial conditions are defined: potential interaction V, positions r of all atoms in the system and velocities of all atoms in the system.

2-Compute forces
The force of any atom is computed by calculating the force between non-bonded atom pairs plus the forces due to bonded interactions, plus restraining and/or external forces.
The potential and kinetic energies and the pressure tensor are computed.

3-Update configuration
The movement of the atoms is simulated by numerically solving Newton's equation of motion.

4- If required: output step
Positions, velocities, energies, temperature, pressure, etc. are written.

Stages from 2 to 4 are repeated while necessary. Finally the trajectory is analyzed in order to obtain information about the system.

Periodic boundary conditions

Molecular dynamics requires a correct treatment of the boundaries in order to get valid macroscopic properties avoiding the use of a huge number of particles. If for instance a protein is embedded in a box of waters, atoms near the limit of the box have a smaller number of neighbours and, as a consequence, a smaller number of interactions than the rest of the atoms. The situation is not realistic.

The solution is to use periodic boundary conditions i.e. to replicate the cubic box in all directions to give a periodic array. So the box studied is surrounded by 26 neighbour boxes. If one particle leaves the box by one side, its image enters from the opposite one. Thus the number of particles remains constant.

Ensembles

In the theory of Thermodynamics several idealized situations, ensembles, are defined which are particular important or easy to calculate. Thus, demanding some environmental constraints particular statistics can be applied in order to take values out for the desired properties.

Microcanonical ensemble (NVE)

In the NVE ensemble the number of particles (N), volume (V) and energy (E) remains constant along the simulation. It corresponds to an adiabatic process with no heat exchange. The trajectory represents an exchange of potential and kinetic energy, with total energy being conserved. Just solving Newton's equations with a fixed volume cell will fix NVE.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Canonical ensemble (NVT)

In the NVT ensemble the number of particles (N), volume (V) and temperature (T) are fixed. In NVT, the energy of endothermic and exothermic processes is exchanged with a thermostat. Thus thermostats make possible to modify the Newtonian MD so that experimental conditions can be reproduced. Moreover they permit to study processes that depend on temperature: thermal coefficients, conformational changes, etc. and to improve the efficiency in the space exploration (MD at high temperatures[91-93], Simulated Annealing[94]). This is the ensemble used in the present work with Berendsen temperature coupling algorithm[95], which mimics coupling to an external bath with a temperature T_0. The algorithm slowly corrects the deviations of the system temperature from T_0. The change in temperature between successive time steps is the shown in Eq. 11. The time constant τ determines how tightly the bath and the system are coupled together.

$$\Delta T = \frac{\delta t}{\tau}(T_0 - T(t))$$

Eq. 11

The heat flow into or out of the system is carried out by scaling the velocities of each particle every step with a time-dependent factor λ, as shown in Eq. 12.

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left\{\frac{T_0}{T(t - \frac{\Delta t}{2})} - 1\right\}\right]^{\frac{1}{2}}$$

Eq. 12

The parameter τ_T is related to τ according to Eq. 13.

$$\tau = 2C_v\tau_T / N_{df}k$$

Eq. 13

C_v is the total heat capacity of the system, k is Boltzmann's constant and N_{df} is the total number of degrees of freedom. τ and τ_T are not exactly equal because the kinetic energy change caused by scaling the velocities is partly distributed between kinetic and potential energy so that the change in temperature is less than the scaling energy.

Isothermal-Isobaric (NPT) ensemble

In the NPT ensemble particles (N), pressure (P) and temperature (T) are conserved. In addition to a thermostat a barostat is needed. There are several algorithms for the pressure coupling. The one used in the present work to carry out equilibrations was Berendsen pressure coupling[95]. In an analogous way to the Berendsen temperature coupling algorithm the effect is of a first-order kinetic relaxation of the pressure towards a given reference pressure P_0. The algorithm rescales the volume of the box every step by scaling coordinates and box vectors.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Equilibration phase

The equilibration phase is aimed at leading the initial system to an equilibrium state in which macroscopic properties as energy, temperature or pressure, slightly fluctuate around an average value without changing substantially. This step allows water molecules to find more natural positions with respect to one another and to the solutes, like proteins. The MD is initiated afterwards, from the system obtained in the equilibration phase. The equilibration protocol used in this thesis was three-stage: 100ps NVT, 100ps NPT and 100ps NVT.

Thermal denaturation method

Even for not very large proteins the experimental time scale of protein unfolding is in the order of microseconds to milliseconds. Therefore a computational simulation of the process by molecular dynamics would take many months or even years. One way of overcoming this problem is to utilize the thermal denaturation method[91-93] at high temperature such as 400K or greater to promote the folding/unfolding processes. This technique could cause a distorted free energy landscape due to the larger entropic contributions at higher temperatures. However, it permits to study in a straightforward way the mechanism behind the protein stability and misfolding. Thus molecular dynamics complements experimental results, as detailed experimental structure characterization of folding intermediates has some limitations.

In 1998, it was reported on interesting series of small peptide in methanol simulated as a function of temperature. It was found that the pathways of folding and unfolding were independent of temperature. [96] In 2006, molecular dynamics simulations at 400K-500K on wild-type and mutant lysozyme (TRP62GLY) aided to understand the mechanism behind amyloid formation triggered by a single residue mutation.[97] The results were consistent with urea denaturing experiments.[98, 99] Moreover, the results were reproducible with two different force fields (CHARMM and OPLSAA). In the present thesis the technique is used in Chapter V to study the loss of stability caused by a punctual mutation in the tetramerization domain of p53 protein.

Besides MD at high temperature, folding/unfolding processes[100] have also been studied by other approximations or simplifications in protein models. Only small peptides[101] and proteins[102] are fully tractable by classical MD. In case of larger proteins pseudo-atoms representing groups of atoms can be defined and statistical potential (Monte Carlo) can be applied to reproduce the folding pathways.[103]

Analysis

The different trajectory analysis used in Chapter V and Chapter VI are described as follows.

Root mean square deviations in structure

The root mean square deviation (RMSD) of certain atoms in a molecule with respect to a reference structure is measured by least-square fitting the structure to this reference structure and subsequently calculating the RMSD (Eq. 14).

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

Eq. 14

In the equation $M = \sum_{i=1}^{N} m_i$ and $r_i(t)$ is the position of atom i at time t. Proteins are usually fitted on the backbone atoms (N, Cα, C) but the RMSD can be computed for the backbone or for the whole protein. When calculating RMSD along the whole trajectory graphs like those in Figure 13 are obtained where RMSD is represented versus time. In the present work this kind of measure was carried out with the *g_rms* tool in Gromacs.[90]

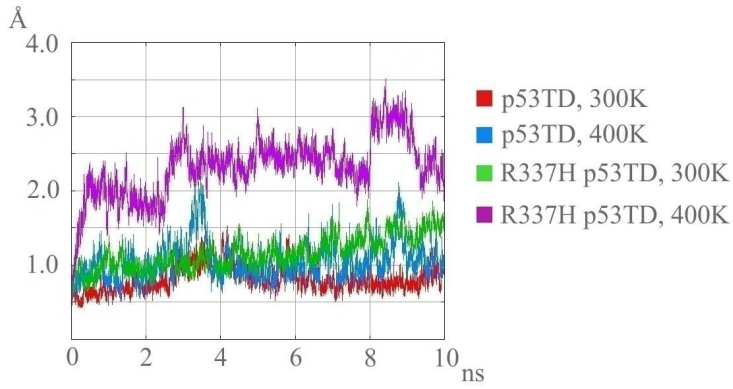


Figure 13. Example of representation of RMSD of the backbone versus time in 4 different trajectories.

Minimum distance

For the study of hydrogen bond interactions and hydrophobic interactions the Gromacs tool *g_mindist* was used to compute the minimum distance between one group of atoms and another. In the case of the hydrogen bonds the measure was carried out between the groups of atoms directly involved in the interaction. In the case of hydrophobic interactions it was carried out between the sidechains of the residues intervening.

X-CSCORE

The program X-Score[104] computes the binding score between a protein and a ligand from the structures of the molecules. It has its major applications to structure-based drug design studies. It has been demonstrated that this binding score correlates well to experimental binding constants. The program includes three empirical scoring functions (see Eq. 15, Eq. 16 and Eq. 17).

$$\begin{aligned} HPSCORE &= C_{0,1} + C_{VDW,1} \times (VDW) \\ &+ C_{HB,1} \times (H-bond) \\ &+ C_{HP} \times (HydrophobicPair) \\ &+ C_{RT,1} \times (Rotor) \end{aligned}$$

Eq. 15

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

$$HMSCORE = C_{0,2} + C_{VDW,2} \times (VDW) + C_{HB,2} \times (H-bond) + C_{HM} \times (HydrophobicMatch) + C_{RT,2} \times (Rotor)$$

Eq. 16

$$HSSCORE = C_{0,3} + C_{VDW,3} \times (VDW) + C_{HB,3} \times (H-bond) + C_{HS} \times (HydrophobicSurface) + C_{RT,3} \times (Rotor)$$

Eq. 17

$C_{0,1}$, $C_{0,2}$, $C_{0,3}$, C_{HB}, C_{HP}, C_{HM}, C_{HS} and C_{RT} are coefficients adjusted empirically. ***VDW*** is the van der Waals interaction energy, calculated by considering all the atom pairs between the ligand and the protein, ***H-bond*** is the sum of all hydrogen bonding interactions between protein and ligand, ***Rotor*** gives an idea of the loss of flexibility in the protein and the ligand upon binding. This term acquires different values depending on the number of atoms involved in rotatable bonds. ***HydrophobicPair*** takes into account hydrophobic interactions by summing up the hydrophobic atom pairs formed between the ligand and the protein. ***HydrophobicMatch*** takes into account hydrophobic interactions by assigning a value to the different atoms of the ligand depending on their binding environment. ***HydrophobicSurface*** accounts for the hydrophobic interactions assuming that they are proportional to the buried part of the ligand.

These three scoring functions give absolute values of dissociation constants for the protein-ligand complex in negative logarithm (pK_d). Except for the hydrophobic effect term, all the other terms in these three scoring functions are calculated using identical algorithms. The scoring functions were calibrated through multivariate regression analysis of a set of 200 protein-ligand complexes[105] and they reproduced the binding free energies of the entire training set with standard deviations of 2.2 kcal·mol^{-1}, 2.1 kcal·mol^{-1} and 2.0 kcal·mol^{-1} respectively. The final result, X-CSCORE (Eq. 18), is an average value between these three.

$$X-CSCORE = (HPSCORE + HMSCORE + HSSCORE)/3$$

Eq. 18

The results reported in 2002[105] showed that this scoring function improves the docking accuracy considerably when compared to conventional force fields used for molecular docking. In 2007, new results were obtained that state the suitability of X-Score to estimate the binding free energy of seven protein-peptide complexes with biological relevance.[106]

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Time-evolution of the secondary structure

The Timeline plugin in the analysis software VMD[107] permits to observe the structural changes suffered by each residue along the time. A 2D graph of the secondary structure vs. time is displayed as a colour-coded map. The horizontal axis is time and the vertical axis is residue number. Thus each residue at each time is represented by a colour which indicates its secondary structure. The assignment of colours is pink for α-helix, yellow for extended conformation (ß-sheets), green for turns and white for undefined secondary structure.

Figure 14 shows an example of 10ns Time-evolution plot for a particular chain of aminoacids with a first part, from the residue 326 to the residue 333, having secondary structure of ß-sheets (coloured yellow) and a second part, from the residue 335 to the 356, having secondary structure of α-helix (coloured pink). The two regions are linked by a residue, 334, with no defined secondary structure (coloured white).

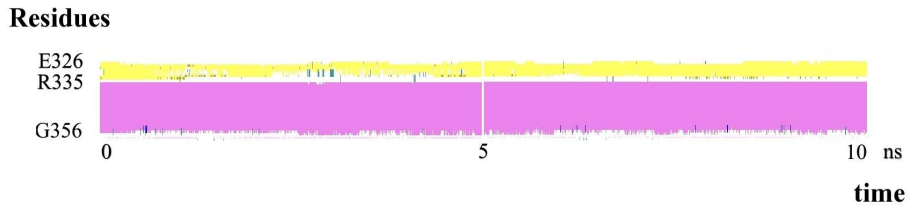


Figure 14. Example of Time-evolution plot in VMD.

Clusterizing structures

The structures obtained from trajectories can be grouped into classes of similar structures. Distances between two structures obtained at different times are determined by measuring RMSD between them after fitting. Each structure is added to a specific cluster when its distance to any element of the cluster is less than a determined value (cutoff). Each cluster can be represented by an average structure. In the present work the analysis was carried out with the tool *g_cluster* in Gromacs.[90]

Contact maps

Contact maps plugin in VMD[107] provides a 2D plot of the residue-residue contacts between two sets of selected atoms from molecules loaded in VMD. The contact distances are displayed as a colour-coded matrix where darker colours indicate residues which are close to each other and lighter colours indicate residue pairs which are distant from each other. In the present work the aminoacids sequence of a given protein was represented in the horizontal axis as well as in the vertical axis. This is a distinctive signature of its folded structure so that it is used to describe the similarity between protein secondary and tertiary structures. Figure 15 shows an example of contact map for a protein composed of four chains A, B, C and D bound.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-6850-7/DL:T-1276-2008

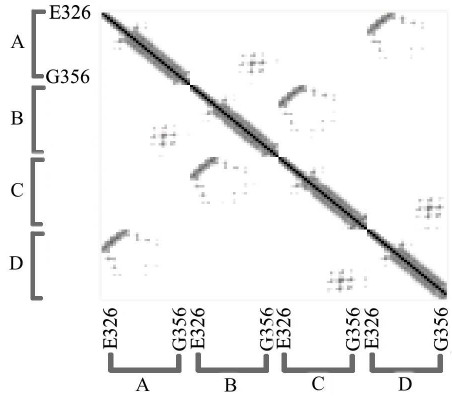


Figure 15. Example of contact map.

RDF

The radial distribution function (or RDF) is an example of a pair correlation function, which describes how, on average, the atoms in a system are radially packed around each other. This proves to be a particularly effective way of describing the average structure of disordered molecular systems such as liquids. Also in systems like liquids, where there is continual movement of the atoms and a single snapshot of the system shows only the instantaneous disorder, it is extremely useful to be able to deal with the average structure.

The RDF is useful in other ways. For example, it is something that can be deduced experimentally from x-ray or neutron diffraction studies, thus providing a direct comparison between experiment and simulation. It can also be used in conjunction with the interatomic pair potential function to calculate the internal energy of the system, usually quite accurately.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Chapter III

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

III. The ureidopyrimidinone dimer

III.1. Introduction

Hydrogen bonds can be used in an efficient way to build large architectures from smaller molecules. A large number of self-assembled systems are constructed by means of hydrogen bond arrays composed of two[108, 109], three[110], four[111-116] or even six[117] donor/acceptor couples. Thus the cooperative action of the hydrogen bonds leads to larger interaction energy and higher molecular recognition specificity. In case of 4 and 6 arrays, autocomplementarity is also possible so that the stability of dimeric forms is strengthened.

Figure 16 shows one of the systems studied in this thesis: it is a molecule composed by urea and pyrimidone (ureidopyrimidinone) that self-assembles forming a Donor-Donor-Acceptor-Acceptor (DDAA) array of hydrogen bonding sites in the 4[1H]-pyrimidinone tautomer. Intramolecular hydrogen bonds are also present in this particular case, which pre-organize the monomer for the interaction. This pre-organization reduces the entropic cost due to the formation of a rigid dimer with both monomers lying on the same plane. The function fulfilled by intramolecular hydrogen bonds combined with hetero-cycles in the monomers, can be achieved with hetero-bicyclic moieties instead. However, the synthesis is then substantially more complicated.

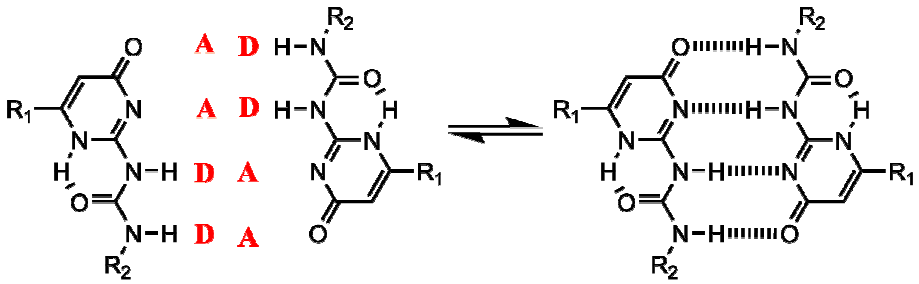


Figure 16. Quadruple hydrogen bonding array based on ureidopyrimidinone (UPy).

The association energy of the ureidopyrimidinones (UPys) is high enough to use the monomers as building blocks to construct larger structures. However, there is a particular complexity in dealing with such systems. This kind of hetero-bicyclic systems with acidic protons often exist in solution as a mixture of tautomeric states in dynamic equilibrium.[118] The shift of the equilibrium towards the non-desired tautomers can reduce considerably the association constant of the supramolecular system aimed to be constructed. Furthermore the displacement is difficult to predict as it depends on multiple factors. It can be influenced by temperature, by substituent effects, by solvent properties and especially by the possibility of formation of intra- and inter-molecular hydrogen bonds, since the enthalpy generated by the formation of such interactions is large.

The ureidopyrimidinone was studied for the first time as an auto-complementary unit by Felix Beijer et al. It was experimentally proved that it strongly dimerize and as a consequence it began to be used in Supramolecular Chemistry.[119-122]

For the ureidopyrimidinone in solution a tautomeric equilibrium is established[123] in which four monomeric species coexist : two forms of the 6[1H]-pyrimidinone (***1*** and ***4*** in Figure 17), the 4[1H]-

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-88507/DL:T-1276-2008

pyrimidinone (***2*** in Figure 17) and its corresponding alcohol pyrimidin-4-ol (***3*** in Figure 17). The two latter have the capability to dimerize by self-assembling, leading to two different kinds of arrays of hydrogen bonds: Donor-Donor-Acceptor-Acceptor (DDAA) and Acceptor-Donor-Acceptor-Donor (ADAD) (see ***5*** and ***6***, respectively, in Figure 17).

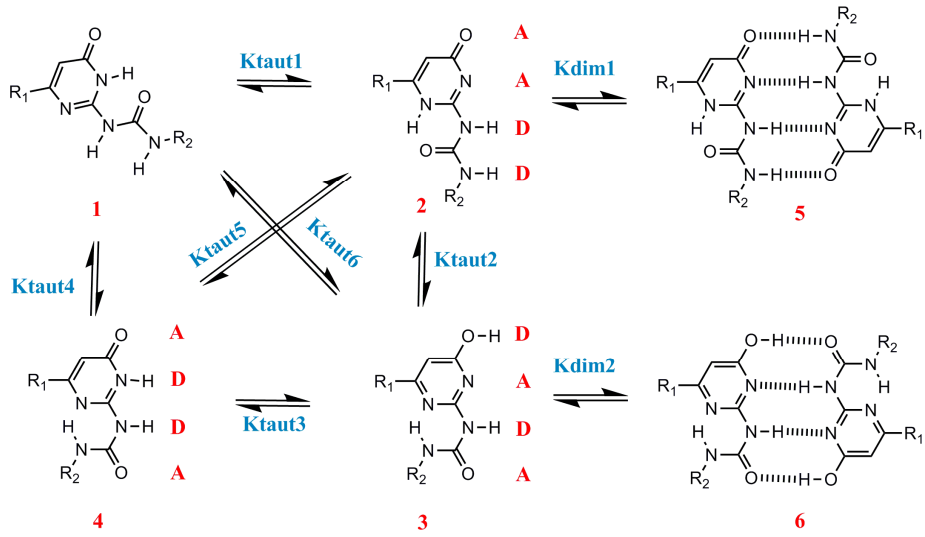


Figure 17. Tautomeric equilibrium established from ureidopyrimidinone.

Both dimers have four similar primary interactions (hydrogen bonds). However, in arrays like these, secondary electrostatic contributions between adjacent donors/acceptors[111, 124] have to be taken into account too. Figure 18 shows a scheme of all the interactions present in dimers of the DDAA type and of the ADAD type. Primary interactions are coloured blue, attractive secondary interactions are coloured green and repulsive secondary interactions are coloured red. It can be observed that in a dimer like ***5*** (Figure 18a) there are four favourable and two unfavourable secondary interactions, whereas in a dimer like ***6*** (Figure 18b) there are six unfavourable secondary interactions. Therefore, at first sight dimer ***5*** would be expected to be stronger than dimer ***6***.

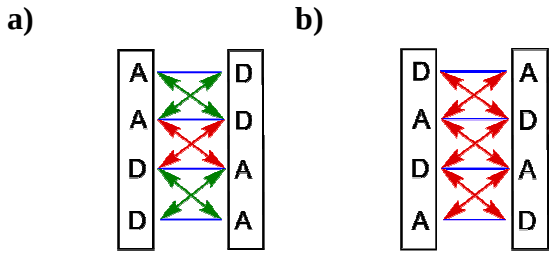


Figure 18. Scheme of the interactions present in dimers of the a) DDAA and b) ADAD type. Attractive and repulsive secondary interactions marked in green and red respectively.

It was experimentally proved that the tautomeric equilibrium in Figure 17 depends on the solvent and the substituents at R_1 and R_2 positions. In general, the forms ***2*** and ***5*** in Figure 17 are very

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

favoured in chloroform, supposedly due to the high energy of dimerization between DDAA motifs. When working in THF or toluene the equilibrium is displaced towards the species ***3*** and ***6***. The reason is attributed to the respective polar and aromatic nature of the solvents (In this later case stacking interactions intervene[125, 126]). Tautomer ***1*** is favoured when there is not the possibility of establishing intermolecular hydrogen bonds. It is only predominant when working in DMSO, which is a very polar solvent. These general tendencies are modified by the nature of the substituent in position R_1.[122] When R_1 is electronegative, as for instance a CF_3 group, the forms ***3*** and ***6*** are more favoured. When it is a weak donor, as CH_3, the displacement takes place towards species ***2*** and ***5***. A good understanding of the influence of the different parameters that govern the equilibrium is necessary to exploit UPy as a fundamental entity for the rational design of larger complexes.

III.1.1. Examples of applications

By means of a convenient functionalization, the UPy monomers can generate polymeric linear structures of high viscosity[127, 128] or helicoidal polymers.[129] Using spacers of the appropriate size, shape and rigidity between UPys, the assembly can be leaded to aggregates with defined structures, such as rosettes and tubes.[130, 131]

In 2006 a ferrocene derivative was designed[122] which dimerized by self-assembling between two UPy units with quadruple DDAA hydrogen bonds. The UPys had been modified in order to avoid the keto-enol tautomerism. The structure (in Figure 19) allowed high electronic communication between the ferrocene units, separated by a long distance of 10Å through the UPy dimer.

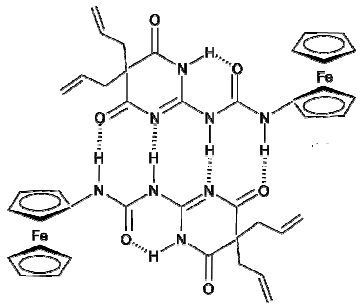


Figure 19. Two ferrocene centres connected by an UPy dimer.

Knowledge of the structural moieties that favour electronic delocalization is very important for the preparation of molecules with potential applications as part of molecular electronic devices. So far multiple types of unsaturated linkers have been used[132] with this capability. In contrast, metal-metal bridges[133] and linkage by means of non-covalent bonds as the herein shown were reported in few occasions. The system in Figure 19 has potential applications in the design of new molecular electronic devices and gives evidence once more of the strength of the binding between ureidopyrimidinone monomers.

Different tautomers of the ureidopyrimidinone (with DDAA and ADAD motifs) were used to form highly stable dimers and polymers of fullerenes[121] like the system shown in Figure 20.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-88507/DL:T-1276-2008

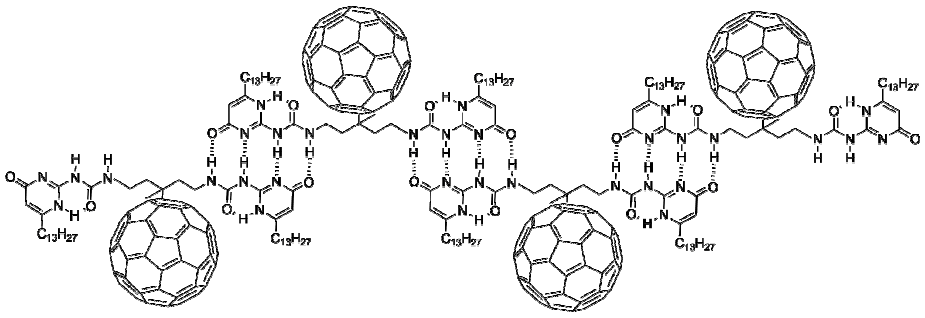


Figure 20. C_{60}-based polymer constructed by means of UPy dimers.

It was proved for a dimer[134] of this kind that there is no mutual interaction in the ground state between C_{60} units. In contrast, photophysical differences were appreciated between the monomer and the dimer when studying excited states, pointing out a strong electronic coupling between the fullerenes, mediated through the hydrogen bond edge. This and other examples of molecular organization of this kind are very promising in the field of molecular materials.

Another illustration of the use of UPys as building blocks, was the reported formation of cavities[119] composed of two calix[4]arenes bounded by two UPy dimers. Previous calixarene dimers with urea substituents had already been reported, able to encapsulate benzene. The use of UPys made possible the encapsulation of larger molecules.

UPys were also applied in chiral polymeric catalyst engineering.[135] In this case the use of arrays of hydrogen bonds was combined with the use of coordination metals in order to construct a polymer in which catalytic centers were immobilized. A scheme of the formation of the polymer is shown in Figure 21: chiral ligands with UPy substituents were synthesized to reach self-assembly by auto-complementarity. These structures spontaneously aggregated leading to immobilized catalytic systems.

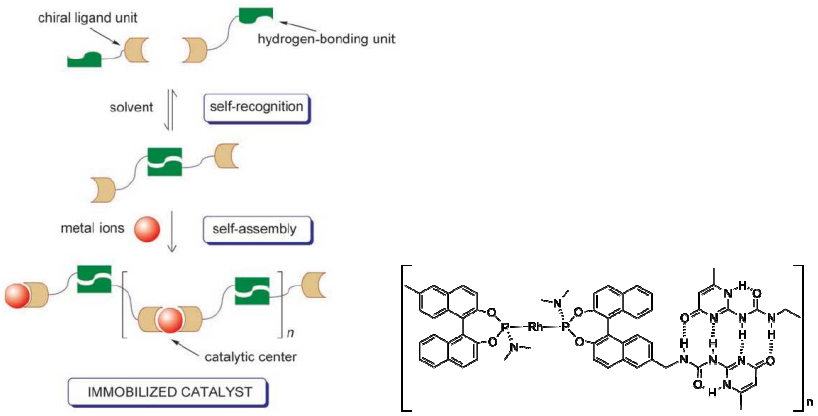


Figure 21. Scheme of the strategy for the construction of a chiral polymeric catalyst based on UPy dimers. Reproduced from ref [135].

This polymer allowed asymmetric reactions. It was used as a catalyst in the asymmetric hydrogenation of de-hydro-α-aminoacids derivatives and of an enamide, obtaining products with a 91-96% of enantiomeric excess.

In Chapter IV a new molecule will be described composed of 3 UPys and a cyclotriveratrylene (CTV) moiety, which is able to dimerize by means of 12 hydrogen bonds. The resulting structure is a capsule able to trap fullerenes of different size and shape inside.

III.2. Goals

The aim of this chapter was to study some aspects of the tautomeric equilibrium established by ureidopyrimidinone. The first goal was to establish theoretically whether the dimer DDAA ***5*** (in Figure 17) was really stronger than the dimer ADAD ***6*** (in Figure 17) due to the influence of secondary interactions. Secondly, an explanation for the effect of a CF_3 group at position R_1, which shifted the equilibrium towards species ***3*** and ***6***, was pursued. Finally, the effect of the chloroform as solvent was analyzed.

III.3. Methodology

The different systems intervening in the equilibrium were optimized with CH_3 as substituent in positions R_1 and R_2 and also with R_1=CF_3 and R_2=CH_3. The deformation energy of the monomers was also calculated. Frequency calculations of the optimised geometries were carried out and thermodynamical parameters for the equilibria were obtained from them.

On the one hand, the calculations were performed in gas phase with ADF program.[70-73] The numerical integration scheme applied for the calculations was developed by te Velde.[136, 137] The geometry optimization procedure was based on the method reported by Versluis and Ziegler.[138] The BP86 functional described as a combination between local VWN exchange-correlation potential with nonlocal Becke's exchange correction[45] and Perdew's correlation correction was applied.[46, 47] A triple-ξ plus polarization basis set was used. The Basis Set Superposition Error (BSSE) was calculated for species 2 by the counterpoise (CP) procedure of Boys and Bernardi.[139]

On the other hand, the same calculations were performed with Gaussian program[140], B3LYP[52, 53] functional and 6-311G** basis set. Analytic vibrational frequencies were computed.[141-143]

III.3.1. Effect of the method on the structure and interaction energy

The choice of the methodology was driven by its success in reproducing experimental data on similar systems and also experimental data available for the dimer ***5***. Also, the computational cost was carefully taken into account since our goal was to study much larger systems. In particular, a nanocapsule that is constructed by self-assembling UPy dimers and that is reported in Chapter IV.

We chose a previous study[144] on adenine-thymine (A-T) and guanine-cytosine (G-C) systems as reference to validate a method. These systems present double and triple arrays of H-bonds. ΔE and ΔH were obtained for both systems with several DFT based methods, taking into account the BSSE.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

ΔH theoretical values were very close to the ΔH experimental values. The calculations were done using BP86/TZP and B3LYP/6-311G** and showed very similar results to reported data. All these data are shown in Table 2.

Table 2. ΔE and ΔH values (kcal·mol^{-1}) for the adenine-thymine (A-T) and guanine-cytosine (G-C) systems. Experimental results, data extracted from literature (ref [144]) and results obtained in this study.

	A-T			G-C		
	ΔE	**ΔE(BSSE)**	**ΔH**	**ΔE**	**ΔE(BSSE)**	**ΔH**
Experimental			-12.1			-21.0
BP86/TZ2P (ref [144])	-13.0	-12.3	-11.8	-26.1	-25.2	-23.8
B3LYP/6-31G** (ref [144])		-12.3	-10.9		-25.5	-24.0
BP86/TZP	-13.2		-9.4	-26.1		-21.8
B3LYP/6-311G**	-15.1		-11.3	-28.5		-24.2

It was observed that ΔE values for the adenine-thymine and guanine-cytosine systems with BP86/TZP (-13.2 and -26.1 respectively) were almost identical to those previously obtained with BP86/TZ2P, which corresponded considering BSSE to ΔH of -11.8 and -23.8 kcal·mol^{-1} respectively. In turn, these values were very close to those obtained with B3LYP/6-31G** and to experimental data. With B3LYP/6-311G** method the ΔE values were slightly higher but comparable.

The second reference to validate our methodology was a report on geometric parameters of the dimer ***5*** (with R_1= CH_3 and R_2= CH_3) obtained by means of experimental data and calculations.[145] Experimental data were obtained from X-ray[123] and from solid-state NMR techniques, capable of providing proton-nitrogen distances of up to about 2.5Å with an accuracy level of ± 0.01-0.05Å. On the quantum chemical side, the equilibrium geometry for the dimer ***5*** had been determined using second-order Møller-Plesset perturbation theory[49] (MP2) and density functional theory (BLYP[45, 59] functional) in combination with a polarised split-valence[146] (SVP) and a polarised triple zeta-valence[147] (TZVP) basis sets. As mentioned in Chapter II, MP2 method takes into account the electron correlation so that it is the most reliable method to treat hydrogen bonds. However, it is excessively computationally expensive.

Table 3 shows these data together with new results obtained from optimization of the UPy dimer with the following methods: LDA/DZP, BP86/DZP, BP86/TZP, BP86/TZ2P and B3LYP/6-311G**. In general, for all the parameters, the LDA/DZP was the method that gave values most far away from the most reliable fully experimental data and X-ray combined with MP2/SVP data.

Distances H3-N1' and H3-N3 obtained with the BP86 and B3LYP functionals were very similar to those obtained by NMR and to those obtained with BLYP/TZVP and MP2/SVP in combination or not with X-ray data. These parameters correspond to a diagonal distance between the UPy monomers and a diagonal distance between two atoms of an ureido group. The fact that they were well reproduced with the computational methods indicated that, in the obtained geometry, the conformation of the monomers as well as their relative disposition in the plane defined by the dimer was correct. H1-N2 values, which depended on a correct treatment of the intramolecular hydrogen bonds, were also very close to the experimental and theoretical results of reference.

The proper description of the intermolecular hydrogen bonds was also guaranteed, as values for H2-N3, N4-O1', N3-N1', H3-O1' and H2-N1' distances were well reproduced except in the LDA/DZP case. H2-N1' was the distance that showed a slightly larger difference. In comparison with the bond

length obtained by NMR, the distances were underestimated in a 15% by LDA/DZP, in a 6-7% by BP86 functional with DZP, TZP and TZ2P basis sets and in a 4% by B3LYP/6-311G**.

Note also from Table 3 that the interaction energy calculated with PB86/TZP and B3LYP/6-311G** was very similar (around -46.0kcal·mol^{-1}). Adding a second polarization function (BP86/TZ2P) changed only very slightly the interaction energy. However, the smallest basis set used, DZP, produced results that deviated a lot. The BSSE was rather small, -1.6 kcal·mol^{-1}, using BP86/TZP and even smaller with the TZ2P basis set.

Therefore, BP86/TZP and B3LYP/6-311G** gave geometric and energetic results close to experimental data for the DDAA dimer ***5***, as well as for the nitrogenised bases discussed previously. So, these were the methods chosen for the study of the tautomeric equilibrium as a balance of the accuracy and the computational cost. It must be borne in mind that future studies on similar but much larger systems were to be performed further on.

Actually, the original idea was to carry out the study using the method BP86/TZP with ADF program due to is efficiency and good parallel scaling. However, the vibrational frequency calculations for some of the entities involved in the equilibrium could not be performed. In the ADF version used (ADF 2005), the force constants are computed firstly and, from them, the normal vibration modes and harmonic frequencies. The force constants are calculated numerically, ie., by evaluating the differences in gradient obtained when producing small displacements of the atomic coordinates. This method presented problems to converge the SCF due to the fact that the displacements of the hydrogen atoms bounded to the heterocycles took them out of the plane formed by the ring, and this originated a very large change in energy. For that reason frequency calculations were carried out analytically by using the Gaussian program, B3LYP functional and 6-311G** basis set. It was checked that in the cases in which ADF calculations converged, the list of frequencies and the thermodynamical parameters obtained were very similar to those obtained with Gaussian. The comparison was possible with species ***2*** (R_1=CF_3), ***3*** (R_1=CH_3 and R_1=CF_3) and ***4*** (R_1=CH_3 and R_1=CF_3).

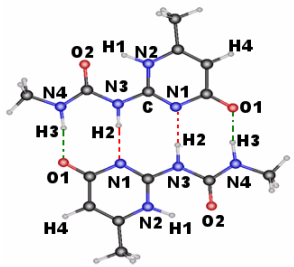


Table 3. Geometric parameters and interaction energies for the UPy dimer (DDAA). Distances in angstroms and angles in degrees. (1) N-H distances from advanced NMR[145] techniques; (2) X-ray data[123]; (3) Fully optimized structure using the method indicated in each case; (4) X-ray structure where hydrogen atoms positions where optimized[145]; (5) Previous calculations [145]; (6) Angle between the lines linking N2 and C atoms of each monomer; (7) Energy of the optimised dimer minus energy of the separately optimised monomers; (8) BSSE computed following the counterpoise method.

			(3)Model							(4)X-Ray	
	(1) NMR	(2)X-ray	LDA DZP	BP86 DZP	BP86 TZP	BP86 TZ2P	B3LYP 6-311G**	(5)BLYP TZVP	(5)MP2 SVP	(5)BLYP TZVP	(5)MP2 SVP
d(H2-N3)	1.07±0.01		1.09	1.06	1.05	1.05	1.04	1.04	1.04	1.04	1.04
d(H1-N2)	1.07±0.01		1.09	1.06	1.05	1.05	1.04	1.04	1.04	1.05	1.04
d(H3-N1)	2.80±0.20		2.70	2.83	2.83	2.85	2.87	2.93	2.86	2.79	2.80
d(H3-N3)	2.40±0.15		2.48	2.48	2.49	2.49	2.47	2.50	2.48	2.48	2.47
d(H3-O1)			1.55	1.67	1.68	1.69	1.70				
d(N4-O1)		2.76	2.62	2.72	2.73	2.73	2.73				
d(H2-N1)	2.01±0.05		1.66	1.83	1.84	1.86	1.88	1.97	1.92	1.92	1.93
d(N3-N1)		2.97	2.75	2.89	2.89	2.91	2.92				
(6)**Angle**			179.9	179.9	179.9	179.9	180.0				
(7) **Eint**			-75.4	-49.8	-46.0	-45.4	-46.8				
(8)**BSSE**			|2.8|	|2.6|	|1.6|	|1.2|					

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

III.4. Results

As follows, Figure 22 and Figure 23 present the geometries obtained from the optimisations at BP86/TZP level for monomers ***1***, ***2***, ***3*** and ***4***, and for dimers ***5*** and ***6***, for R_1=CH_3 and CF_3. The most relevant differences were found in the intra- and inter-molecular hydrogen bond distances.

If we compare the monomer ***2*** to the dimer ***5*** the intramolecular hydrogen bond shortened from 1.836Å to 1.647 Å when R_1=CH_3. This variation was larger than the produced from the monomer ***3*** to the dimer ***6***, in which case the respective values were of 1.880 Å and 1.829 Å. When R_1=CF_3 the intramolecular hydrogen bond shortened more in ***2*** than in ***3***, respect to the same monomers with R_1=CH_3. In dimer ***5*** the distance was also shorter when R_1=CF_3 than when R_1=CH_3 but in dimer ***6*** the opposite happened.

In the case R_1=CH_3 the intramolecular hydrogen bond distance of the monomers ***1*** and ***4*** was always larger than in the rest of the species, with values 1.918 Å and 1.904 Å. In the case R_1=CF_3 it occurred the same with the monomer ***1***, with the only exception of the monomer ***3*** which had a larger distance.

Concerning the intermolecular hydrogen bonds, the substitution of CH_3 by CF_3 at position R_1 caused an elongation of all the distances in the dimer ***5***: those of the N-H···N type as well as those of the N-H···O type. In the dimer ***6*** the N-H···N distances became longer but, in contrast, O-H···O distances became shorter. N-H···N distances showed always larger in the dimer ***6*** than in the dimer ***5***, with R_1=CH_3 and with R_1=CF_3.

In all the cases the distances of N-H, O-H and C=O covalent bonds containing atoms involved in intermolecular hydrogen bonds, showed longer in the dimers than in the monomers.

All the hydrogen bonds were characterised by the presence of bond critical points in the electronic charge density function. Also, the corresponding ring critical points were found. No large differences were found in the value of $\rho(r)$ at the bond critical points when R_1=CH_3 was changed by R_1=CF_3.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

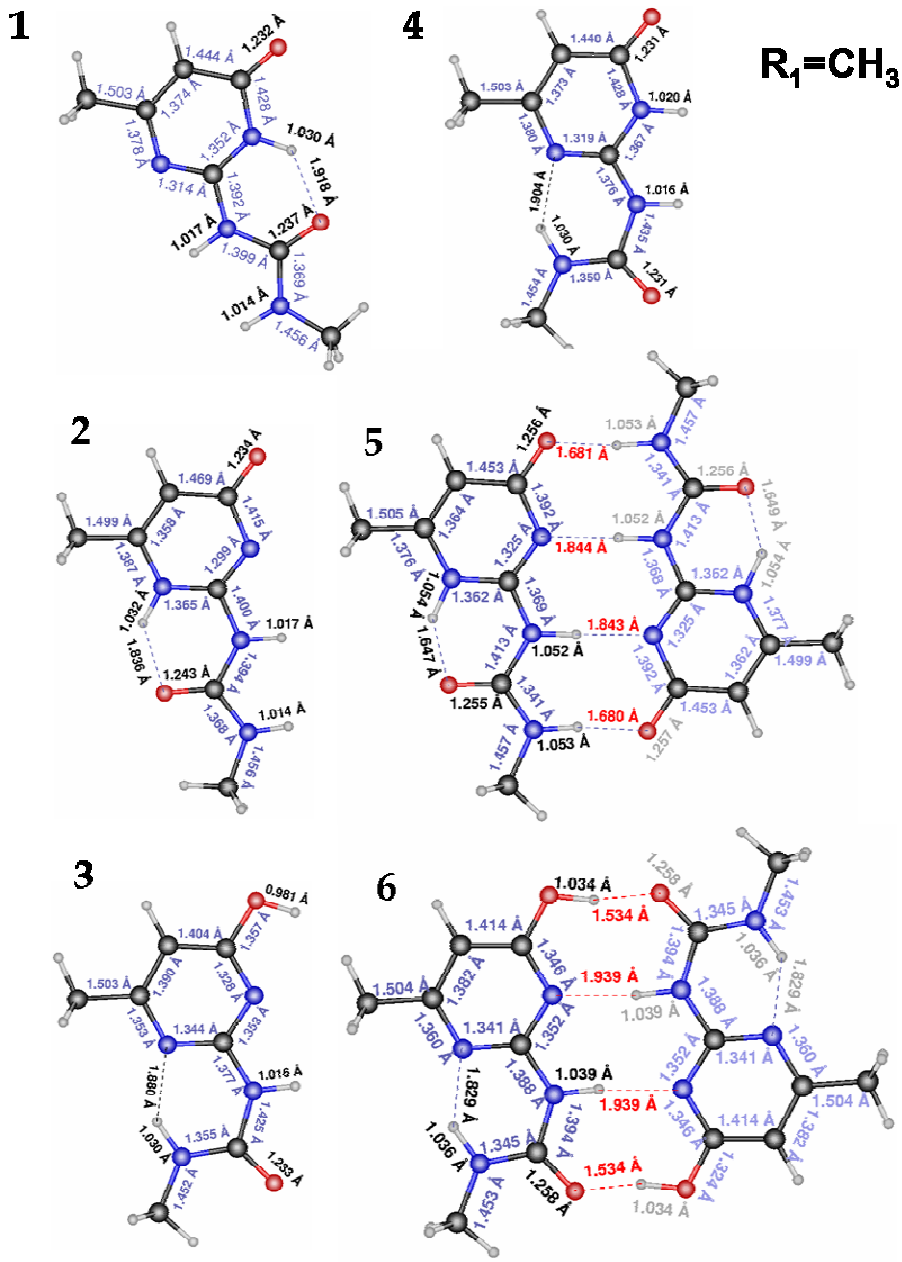


Figure 22. Optimised geometries with R_1=CH_3.

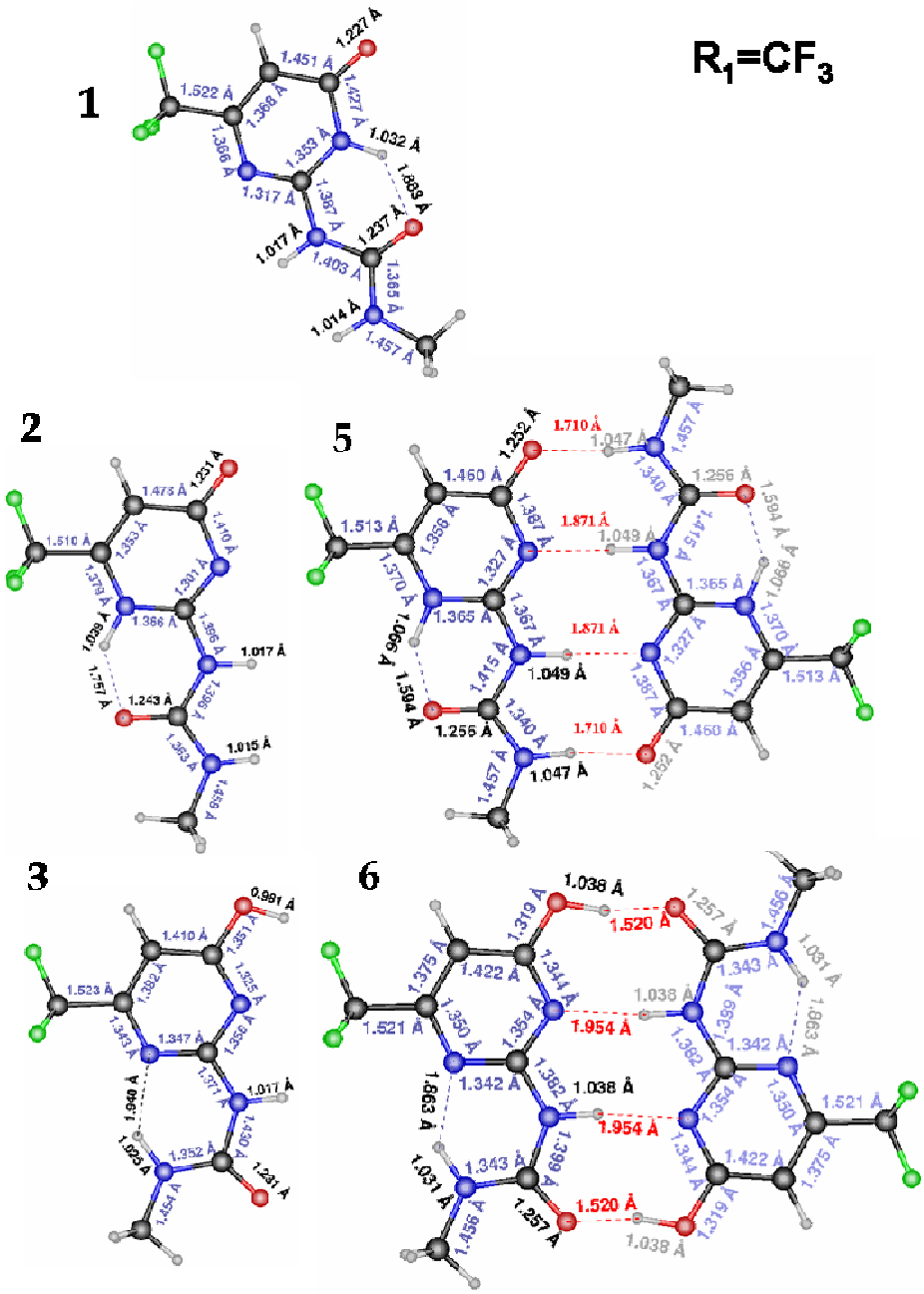


Figure 23. Optimised geometries with R_1=CF_3.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-88507/DL:T-1276-2008

Table 4. Dimerization energy (kcal·mol^{-1}) from 2 to 5 and from 3 to 6 in the two cases R_1= CH_3 and R_1= CF_3.

		BP86/TZP			B3LYP/6-311G**
		(1)EintSp	(2)Edeform	(3)Edim	(3)Edim
2-5					
	R_1=CH_3	-55.3	4.7	-45.9	-46.8
	R_1=CF_3	-52.7	4.8	-43.1	-43.9
3-6					
	R_1=CH_3	-40.5	5.6	-29.3	-31.2
	R_1=CF_3	-42.9	5.9	-31.1	-32.4

(1) Energy of the dimer minus energy of the two monomers in the geometry of the dimer.
(2) Energy involved in the deformation of each monomer to form the dimer.
(3) Dimerization energy calculated as the energy of the optimized dimer minus the energy of the optimized monomers.

Table 4 shows the interaction energy values obtained in the 2-to-5 and 3-to-6 dimerization processes, together with the energy necessary for the deformation of the respective monomers. The term **EintSp** accounts for the establishment of four hydrogen bonds between monomers in the final geometry. It was calculated as the energy of the dimer minus the energy of the two monomers in the geometry of the complex. **Edim** in Table 4 is the dimerization energy taking into account the deformation suffered by the monomers to form the dimers. It was calculated as the energy of the dimer minus the energy of the two monomers in their optimized geometry. **Edeform** was the difference between the two terms and accounted for the deformation energy per monomer.

Larger **EintSp** values in the 2-to-5 cases were due to the more favourable secondary interactions. However, it is remarkable that the interaction when R_1=CH_3 was larger than when R_1=CF_3. In contrast, the opposite effect was observed in the 3-to-6 cases. The 2-to-5 cases were not only favoured in terms of the **EintSp** value, but also in terms of monomer deformation. **Edim** trends were the same as **EintSp** tendencies but the difference between the different dimers was stressed.

Thus, the presence of the CF_3 group slightly destabilized the 2-to5 process and slightly stabilized the 3-to-6 process. However, cooperativity resulted in major dimerization energy of the DDAA type in all cases. This indicated that an extra contribution of the electronic effect of CF_3 on the stabilization of the dimer ***6*** must be due to further influence in the tautomeric equilibrium. Therefore, thermodynamic parameters for all the equilibria involved were investigated. Free energy (ΔG), enthalpy (ΔH) and entropy (ΔS) were calculated. Table 5 collects these results, graphically represented in Figure 24, Figure 25 and Figure 26 with the energy of ***1*** taken as reference.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Table 5. Thermodynamic values for the equilibria between the different species in gas phase.

		BP86/TZP				B3LYP/6-311G**		
		ΔE	**ΔH**	**ΔS**	**ΔG**	**ΔE**	**ΔH**	**ΔG**
1-to-2								
	R_1=CH_3	**6.3**	**6.8**	**-0.6**	**6.9**	**7.4**	**7.8**	**8.0**
	R_1=CF_3	**8.6**	**8.6**	**2.8**	**7.8**	**9.4**	**9.4**	**8.6**
2-to-5								
	R_1=CH_3	**-46.0**	**-41.8**	**-48.7**	**-27.3**	**-46.8**	**-42.6**	**-28.1**
	R_1=CF_3	**-42.9**	**-39.9**	**-60.9**	**-21.8**	**-43.9**	**-40.8**	**-22.6**
1-to-3								
	R_1=CH_3	**-1.8**	**-0.7**	**-3.3**	**0.3**	**-1.0**	**0.1**	**1.0**
	R_1=CF_3	**-0.3**	**0.7**	**0.7**	**0.5**	**0.2**	**1.3**	**1.1**
3-to-6								
	R_1=CH_3	**-29.4**	**-25.9**	**-44.5**	**-12.6**	**-31.2**	**-27.8**	**-14.5**
	R_1=CF_3	**-30.9**	**-27.9**	**-47.2**	**-13.8**	**-32.4**	**-29.3**	**-15.2**
2-to-3								
	R_1=CH_3	**-8.1**	**-6.9**	**-2.7**	**-6.1**	**-8.4**	**-7.1**	**-6.3**
	R_1=CF_3	**-8.9**	**-7.3**	**-2.1**	**-6.7**	**-9.1**	**-7.6**	**-6.9**
1-to-4								
	R_1=CH_3	**0.1**	**0.8**	**-2.7**	**1.6**	**0.7**	**1.4**	**2.2**

ΔE, ΔH and ΔG in kcal·mol^{-1}. ΔS in cal·mol^{-1}K^{-1}. ΔH and ΔS calculated with B3LYP/6-311G. ΔE calculated with two methods (BP/TZP and B3LYP/6-311G**).**

Figure 24 and Figure 25 show the ΔG values obtained with BP86/TZP and B3LYP/6-311G** methods, taking ***1*** as reference. As it can be observed, ***1*** was the most stable monomer in all cases, followed by ***3***, ***4*** and ***2***. The ΔG values for ***3*** and ***4*** with respect to ***1*** were very low(less than 3kcal·mol^{-1}).

Between monomer ***1*** and monomer ***2*** there were larger differences of 6.9kcal·mol^{-1} when R_1=CH_3 and of 7.8 kcal·mol^{-1} when R_1=CF_3, using BP86/TZP method. It is remarkable that the presence of the CF_3 group at R_1 in ***2***, caused a certain destabilization of 0.9 kcal·mol^{-1} with respect to the R_1=CH_3 case.

The energy values for ***1*** and ***2*** using B3LYP/6-311G** were 8.0 kcal·mol^{-1} and 8.6 kcal·mol^{-1} in R_1=CH_3 and R_1=CF_3 cases respectively. Thus the destabilizing effect of the CF_3 group over the monomer ***2*** appeared to be less important at this level of theory (only 0.6 kcal·mol^{-1}). As observed, the tendency between the monomers ***1*** and the monomers ***3*** did not depend on the substituent R_1 (green and violet profiles for these species are practically superposed in both Figure 24 and Figure 25).

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-88507/DL:T-1276-2008

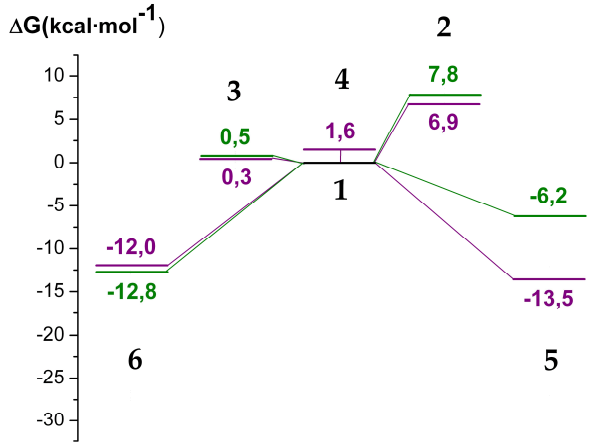


Figure 24. ΔG in kcal·mol^{-1}, using BP86/TZP method. R_1= CH_3 coloured violet; R_1=CF_3, coloured green.

When R_1=CH_3 at BP86/TZP level (violet trace in Figure 24), ***5*** was the thermodynamically most stable dimer, with a value of -13.5kcal·mol^{-1} with respect to ***1***. Indeed the difference with the dimer ***6*** was slight. The ΔG value for ***6*** in this case was of -12.0 kcal·mol^{-1}, only 1.5 kcal·mol^{-1} unfavoured. When R_1=CF_3 at the same level of theory (green trace in Figure 24), the ΔG value for ***6*** remained almost the same, only becoming 0.8 kcal·mol^{-1} more favourable than in R_1=CH_3 case. However, the ΔG value for ***5*** experienced a large destabilizing shift of 7.3 kcal·mol^{-1} with respect to the analogous R_1=CH_3 case, reaching a value of -6.2 kcal·mol^{-1} in Figure 24.

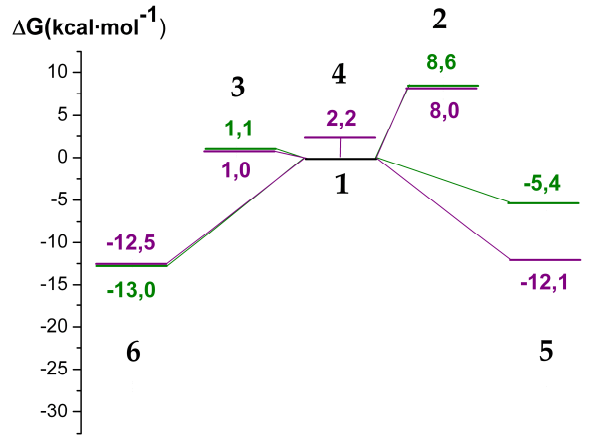


Figure 25. ΔG in kcal·mol^{-1}, using B3LYP/6-311G method. R1=CH_3 coloured violet; R_1=CF_3, coloured green.**

At B3LYP/6-311G** level, the ***5*** and ***6*** were almost equally favoured in R_1=CH_3 case (violet trace in Figure 25). Their respective ΔG values were -12.1 kcal·mol^{-1} and -12.5 kcal·mol^{-1}. The most

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

remarkable effect at this level of theory was again the strong destabilization of ***5*** when changing CH_3 by CF_3 at R_1 position. The shift in energy was of 6.7 kcal·mol^{-1}.

Concerning the ΔH values, at BP86/TZP level as well as at B3LYP/6-311G** level the same trends found in the free energy variations were observed (see the case BP86/TZP in Figure 26). As it was observed in the ΔG profiles in Figure 24 and in Figure 25, when R_1=CH_3 (violet profile in Figure 26) ***5*** was found slightly more favoured than ***6***, while when R_1=CF_3 (green profile in Figure 26) ***5*** was more destabilized than ***6***. Therefore, both ΔG and ΔH values indicated a clear electronic effect of R_1 substituent on the relative stability of ***5*** and ***6***. When a more electronegative substituent is present in position R_1, the equilibrium is displaced towards the formation of ADAD dimer ***6***, as it was found experimentally.

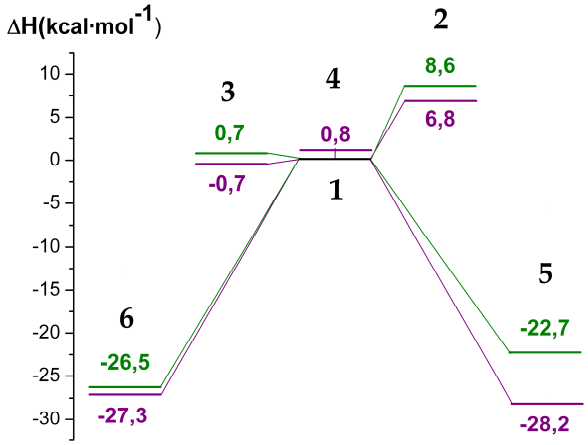


Figure 26. ΔH in kcal·mol^{-1}, using BP86/TZP calculations. R_1=CH_3 coloured violet; R_1=CF_3, coloured green.

Since it is known that the solvent affects the position of this equilibrium[123], we attempted to include somehow solvent effects in our calculations. This was done by using the PCM method as implemented in the Gaussian program. We used B3LYP/6-311G** level together with the standard options in Gaussian for $CHCl_3$. We evaluated the solvation energy for ***1***, ***2*** and ***3***, for R_1=CH_3 and CF_3. Table 6 shows the results. As it can be observed, the solvation energy was always the most favourable for ***2***, followed by ***1*** and ***3***. However, when R_1=CF_3 the difference between the solvation energy of ***2*** and ***1*** was smaller, since the solvation energy of ***2*** got less favourable in 1.65 kcal·mol^{-1}. The solvation energy of ***1*** and ***3*** got more favourable in 0.55 kcal·mol^{-1} and 0.72 kcal·mol^{-1} respectively.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-88507/DL:T-1276-2008

Table 6. Solvation free energy (in kcal·mol^{-1}) of 1, 2 and 3 for R_1=CH_3 and CF_3 in $CHCl_3$.

		ΔGsolv
R_1=CH_3		
	1	-11.48
	2	-15.40
	3	-6.49
R_1=CF_3		
	1	-12.03
	2	-13.75
	3	-7.21

With these results we obtained the difference of solvation energy from ***1*** to ***2*** and from ***1*** to ***3*** (*ΔΔGsolv* in Table 7). Then we added the values obtained to the already calculated ΔG in gas phase (*ΔGgas* in Table 7) for each step of the equilibrium. The resulting *ΔGtot* values are shown in the third column of Table 7.

Table 7. ΔG (in kcal·mol^{-1}) for the processes 1-to-2 and 1-to-3 in $CHCl_3$.

		ΔGgas	ΔΔGsolv	ΔGtot
R_1=CH_3				
	1-to-2	8.0	-3.9	4.1
	1-to-3	1.0	5.0	6.0
R_1=CF_3				
	1-to-2	8.6	-1.7	6.9
	1-to-3	1.1	4.8	5.9

The ΔG profiles obtained showed changes with respect to the results in the gas phase. In both R_1=CH_3 and R_1=CF_3 cases, the most stable monomer was the monomer ***1*** and therefore ΔG values were calculated taking it as reference, exactly in the same way as above.

When R_1=CH_3 the ΔG value for ***3*** increased in 5 kcal·mol^{-1} with respect to the value in gas phase. In contrast, ***2*** decreased in 3.9 kcal·mol^{-1}. This made the DDAA monomer ***2*** be 1.9 kcal·mol^{-1} more favoured than the ADAD monomer ***3***. This could be an explanation for the tendency of the equilibrium towards the DDAA forms when working in $CHCl_3$ solvent.

When R_1=CF_3 this effect was not as stressed. The change in the ΔG value of the monomer ***3*** with respect to the value in gas phase was almost the same as in the R_1=CH_3 case. However, the decrease in the value of the monomer ***2*** was smaller (only 1.7 kcal·mol^{-1}). The final result was that the monomer ***3*** was still 1 kcal·mol^{-1} more stable than the monomer ***2***. The destabilization of the monomer ***2*** caused by the CF_3 group at R_1 position had already been observed in the gas phase. In $CHCl_3$ this effect showed stronger.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

III.5. Conclusions

The interaction energy between two monomers ***2*** to form the DDAA dimer ***5*** is larger than the interaction energy between two monomers ***3*** to form the ADAD dimer ***6***. This was expected, as the dimer ***5*** has an array of hydrogen bonds of the DDAA type, in which there are four favourable secondary interactions and only two unfavourable. Hence, the DDAA dimer is stronger than the ADAD.

When a CF_3 group is present at position R_1, the dimerization energy to the specie ***5*** was lower than when there was a CH_3. The opposite happened with the dimerization energy to the specie ***6***, favoured by the presence of a CF_3 at R_1. However, this effect was not enough to make the dimerization energy of the dimer ***6*** be the most favoured in any case.

When R_1=CF_3, the dimerization energy in the DDAA case drops and the dimerization energy in the ADAD case rises. But, this effect is not large enough, since the DDAA dimer is still stronger than the ADAD.

The study of ΔG in gas phase permitted to see that the change from a CH_3 group to a CF_3 group at R_1, fundamentally causes a thermodynamical destabilization of the dimer ***5*** with respect to the most stable monomer in the equilibrium, which is the monomer ***1***. At the same time a slight destabilization of the monomer ***2*** is observed, which would stress the effect of the electron-withdrawing group. The relative stability of the rest of species taking the monomer ***1*** as reference was the same in both R_1=CH_3 and R_1=CF_3 cases.

When solvent effects were included, it was found that the monomer ***2*** can be more stable than the monomer ***3***, being ***1*** always the most stable tautomer. Therefore, although solvent can favour either ***2*** or ***3***, we think that the dimerization process is the driving force towards the formation of a DDAA or an ADAD dimer.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Chapter IV

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-0857-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

IV. Fullerenes in cages

IV.1. Introduction

Fullerenes[148, 149] are the most stable form of carbon, after diamond and graphite. They were called buckminsterfullerenes in honour to Richard Buckminster Fuller, architect that popularized "domo geodesic", as the molecules present a similar shape to that type of vault. Since their discovery[150, 151] around 1990 much time and effort have been devoted to the exploration of their properties and applications. They present unique and fascinating structural, electronic, electric, mechanic and optical properties and are very versatile.

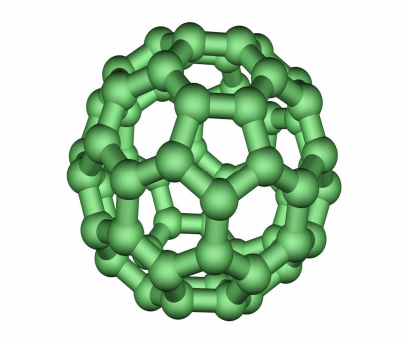


Figure 27. Buckminsterfullerene (C_{60}).

Fullerenes are convex carbon cages (see Figure 27 as an example) comprising hexagons and pentagons. The most stable fullerenes fulfil two chemical conditions: the Isolated Pentagon Rule (IPR)[152] and the Hollow Pentagon Rule (HPR). The IPR states that there are no adjacent pentagons and the HPR that all the pentagons are holes; this means that every pentagon can have only external double bonds.

With 60 carbon atoms the smallest fullerene that fulfils the IPR is the Buckminsterfullerene, C_{60}[153] (Figure 27). It has icosahedral symmetry, resembling a football ball, with 20 hexagons and 12 pentagons and it is the best known fullerene.

Fullerenes with adjacent pentagons are unstable due to an enhanced steric strain and an electronic effect of resonance destabilization. In fullerenes that cannot follow IPR rule the most stable isomer is the one with least adjacent pentagons.[154] This is stated by the Pentagon Adjacency Penalty Rule (PAPR).[155] However, stabilization of fullerenes larger than C_{70} and that do not accomplish IPR is possible by charging or encapsulating electron-donating metal atoms and clusters.[156] Fullerenes smaller than C_{60}[157] have always adjacent pentagons and higher pentagon-to-hexagon ratio. This leads to a higher curvature of their carbon surface with much steric strain[158, 159] and to a weakened π-conjugation. The consequence is a much higher lability so that they can only be observed in the gas phase[160] although it has been demonstrated that they can be stabilized by functionalization.[161-163]

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8857-7/DL:T-1276-2008

In fullerenes, the geometric demand of the spherical cage is such that all the double bonds deviate from planarity.[164] The formally sp^2-hybridized carbon atom is pyramidalized so that their chemical reactivity resembles that of the cycloalkenes more than benzene. A measure of the pyramidalization degree in each atom can be used to assess the local reactivity of carbon atoms in the molecule. In fact pyramidalization is one of the reasons for which fullerenes are electronegative. The deviation from planarity causes orbital rehybridization[165] of the sp^2 orbitals and pi orbitals to new sp orbitals with a gain in p-character. The p-lobes extend further outside the surface than they do into the interior of the sphere.

When discovered, fullerenes were supposed to be extremely stable aromatic molecules. For example, C_{60} was expected to have 12500 possible resonance structures.[166] However the most stable isomers avoid having double bonds in the pentagonal rings as their presence would shorten bonds in the already strained ring. Thus, for the case of C_{60} there is only one structure that fulfils this requirement. Electron delocalization is then not as rich as was thought a priori and, as a consequence, the fullerene is much more reactive than expected.[153] Indeed the double bonds are not all equal. The bonds connecting two hexagons are shorter than those connecting a hexagon and a pentagon.

Fullerenes react as electrophiles and also driven by a force leading to a relief of strain by saturating double bonds. The most important methods for the derivatization of a fullerene are based on exohedral addition reactions: nucleophilic and radical additions, cycloadditions, alkylation, hydrogenations, transition metal complex formations, oxygenations and halogenations.

It is also possible to obtain fullerenes with additional atoms trapped inside them. It can be metals, in which case the complexes are called endohedral metallofullerenes, or no-metals, in which case the complexes are known as non-metal doped fullerenes. The production of endohedral fullerenes has been based on physical methods such as co-vaporization of carbon and metal atoms and high-pressure/high-temperature treatment with gases. These methods are very laborious and the quantity of product obtained is very low. An alternative method tested recently was their synthesis by opening the fullerene, putting the guest inside and closing again by means of a series of organic reactions. This has been reached recently in the particular case of insertion of a hydrogen molecule inside C_{60}.[167-169]

Concerning fullerene applications in biomedicine, drug delivery is still a hope. Previously, further knowledge is needed on how to bind the appropriate ligands onto the outside of the cage and a better understanding of the interaction of fullerenes with a biological environment is necessary. In relation to this, a curious ability of a derivatized C_{60} to inhibit HIV-1 protease[170-172] (HIVP) was reported in 1993. HIVP has an important role in the life cycle of the HIV virus so that its inhibition in vivo avoids the maturation of new infectious virions and fights AIDS. Models indicated that the inhibition by C_{60} was due to a perfect fit of the fullerene in the active site of the protein, by interacting with a non-polar part of the surface (see Figure 28) and by establishing hydrogen bonds with two aspartates.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

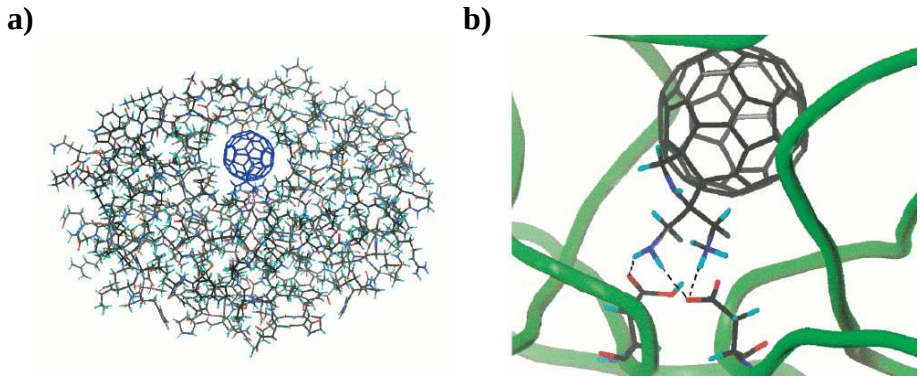


Figure 28. a) Computer-designed accommodation of one water-soluble C_{60} derivative in the cavity of HIV-1 protease. b) Closer view of the complex showing H-bonds between NH_2 or NH_3^+ groups from the fullerene and aspartates 25 and 125. Reproduced from ref [170].

Fullerenes have exhibited other in vitro biological activities as site-selective DNA cleavage and cytotoxicity. The capability of fullerenes to cause DNA-photocleavage[173-175] under visible-light irradiation has been reported by several groups. The process is related to the easy photoexcitation of fullerenes (see Figure 29). By photoirradiation, C_{60} can pass from its singlet ground state to a singlet excited state and afterwards to a triplet excited state. By transferring energy to the O_2 molecule the fullerene can return to its ground state. $^3O_2$ molecule converts to singlet oxygen $^1O_2$, known to be highly damaging. A superoxide radical can be produced as well in the case C_{60} takes an extra electron from an electron donor and transfers it to the O_2. In turn, the superoxide radical facilitates the production of a hydroxyl radical, which can react with the superoxide anion producing singlet oxygen again. Singlet oxygen and superoxide radical anion are well known reactive species towards DNA causing hydrolysis of the phosphate bonds in guanosine residues. Derivatives of C_{60} bearing sequence-specific DNA-binders are pursued in order to reach directed DNA cleavage.

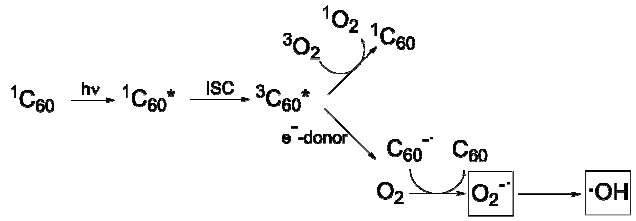


Figure 29. Scheme representing the production of species such as $^1O_2$, $O_2^{\cdot-}$ and ·OH through the photoinduction of C_{60}.

Cytotoxicity of fullerene species is a sensitive function of surface derivatization. Non-derivatized fullerenes can cause cell death by insertion in the phospholipidic bilayers and oxidative damage to the cell membranes. However, fullerenes with highly soluble attached moieties are not toxic.[176]

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8857-7/DL:T-1276-2008

Thus, the toxicity of the molecules could be controlled depending on the use required, existing the possibility of applying them as drug carriers but also as cancer therapeutics or bactericides.

Apart from their potential applicability in medicinal chemistry, fullerenes are promises in the materials science field for the construction of solar cells cheaper than the existing crystalline silicon-based technology.[177-180] Photoactive devices can be construct consisting of a mixture of π-conjugated polymers or porphyrin moieties acting as electron donors and soluble fullerene-derivatives acting as electron acceptors.

Endohedral fullerenes are expected to have a future in several fields such as molecular electronics[181], magnetic resonance imaging[182] and nuclear magnetic resonance[183, 184] (NMR) analysis.

Hydrogen-bonded motifs in the design of supramolecular architectures based on fullerenes have relatively recently been reported. Several donor-acceptor assemblies have been constructed by means of hydrogen bonds.[185-189] Electronic communication in C_{60}-based hydrogen-bonded donor-acceptor ensembles has been proved to be as strong as in covalently connected systems. In the case of new hydrogen-bonded C_{60}·TTF (tetrathiafulvalene) complexes in Figure 30, the lifetimes measured for the radical-ion pair states are in the range of hundred nanoseconds, several orders of magnitude higher than they were in covalently linked C_{60}·TTF dyads[190] reported previously. This point out the fundamental advantages of strong and highly directional hydrogen-bonding networks in assisting electron-transfer processes to the design of efficient photovoltaic devices.

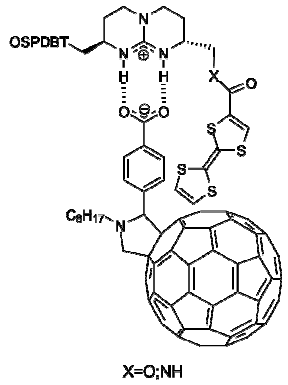


Figure 30. C_{60}·TTF dyads linked by hydrogen bonding and strong electrostatic interactions.

As it can be seen in the examples above, fullerenes are considered as promise for a broad range of applications. However their application has not already become commercial because their production and purification keep on being a great challenge for chemists.

The usual method to produce fullerenes is to generate an electric current between two nearby graphite electrodes in an inert atmosphere (Figure 31). A deposit of soot is generated, from which very different fullerenes can be isolated. The amount of C_{60}, C_{70} and fullerenes with a larger number of carbon atoms generated following this protocol is very low, less than 9%of the soot mass.[149]

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

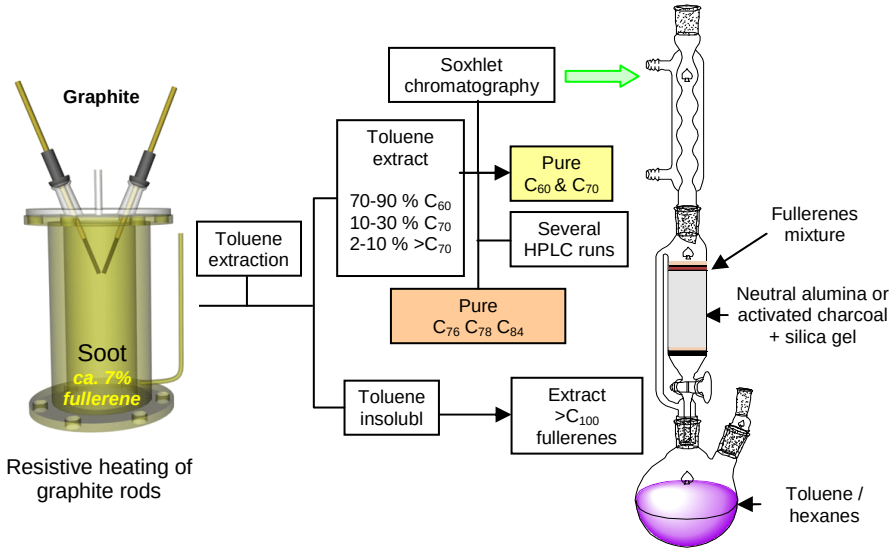


Figure 31. Scheme of the most usual method for the manufacture and purification of fullerenes.

For the separation and purification of C_{60}-C_{70} mixtures, solid-liquid extractions are normally carried out with toluene, evaporations and chromatographic separations that require huge amounts of solvents.[191-195] This determines the exorbitant prices for these compounds. Most of the experimentation on fullerenes has been carried out with C_{60} as it is the easiest to obtain because of its abundance. Experimentation with fullerenes of larger size is restricted. Therefore, new techniques for a cheaper obtaining of the compounds are pursued.

Some alternative methods of separation have been reported based on host-guest encapsulation by means of cyclodextrins[196] or calix[8]arenes.[197] However these capsules are selective for the major compound C_{60} but not for C_{70} or other higher fullerenes. Two examples of calixarenes were reported to have more affinity for C_{70} than for C_{60}, but in these cases the release of the fullerene and the recovery of the host from the complex proved difficult or required the use of chromatography.[198, 199] Macrocyclic dimers of zinc porhyrins[200] were used to extract fullerenes $\geq C_{76}$ from mixtures, allowing the separation of rare fullerenes C_{102}-C_{110}. However, chromatography was still required at some steps of the process.

IV.2. Motivation

On the one hand, in the early 1990´s large affinity between fullerenes and cyclotriveratrylene[201] (CTV) was reported. The C_{60}@CTV complexes formed, showed a 1:1 complexation stoichiometry in solution. However, 2:1 complexes were described only rarely.[202-204] On the other hand, 2-ureido-4-[1H]-pyrimidinone (UPy)[114, 123, 205, 206] had been shown to be one of the strongest hydrogen-bonded self-complementary units with quadruple DDAA hydrogen bonds, as mentioned in the previous chapter.

Based on these previous results, Javier de Mendoza and collaborators synthesized a molecule composed of three UPys bound by means of a central structure of CTV. It was proved that its dimerization with another equal molecule leaded to a cage-shaped structure able to trap a fullerene

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8857-7/DL:T-1276-2008

inside (Figure 32). Moreover it was observed that in the presence of several fullerenes the capsule had higher selectivity for some of them. Indeed, the host could be used to efficiently purify C_{70} from crude soot or fullerite mixtures by simple solid-liquid extractions thanks to much higher selectivity of the host for C_{70} versus C_{60}.

The computational study that we performed pursued two goals. First, to assess the viability of the complex as well as a theoretical explanation for the $(CTV\text{-}3UPy)_2$ capsule preferences. And second, to predict the affinity of the capsule for other higher fullerenes present in fullerite.

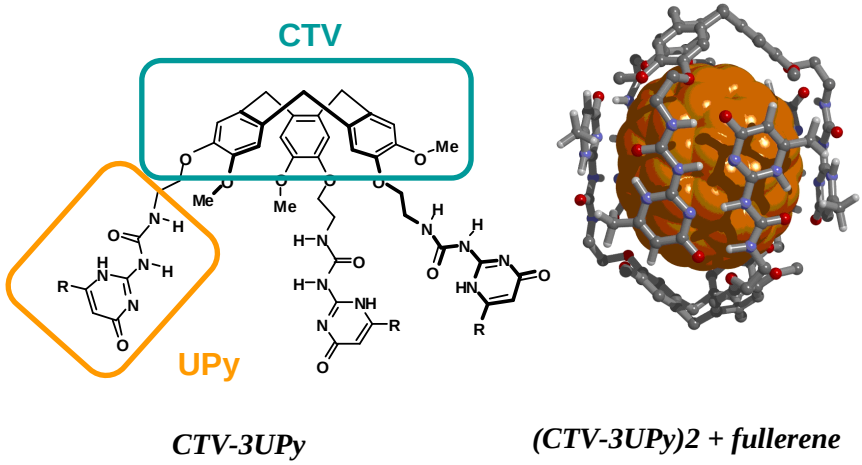


Figure 32. Design of a molecule based on CTV and UPy moieties, able to dimerize trapping fullerenes inside.

IV.3. Building the structures

Two possible configurations for the CTV-3UPy molecule were considered, arbitrarily called herein CTV-3UPy(R) and CTV-3UPy(L) (see Figure 33). One had the UPy fragments 180º turned respect to the other. The consideration of these two configurations was necessary for constructing certain capsules considered in the present work.

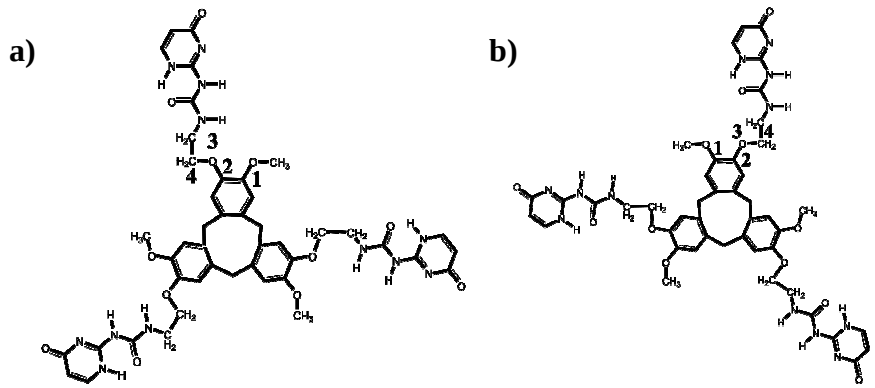


Figure 33. View from below of a) CTV-3UPy(L) and b) CTV-3UPy(R) molecules.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Moreover, for each of both CTV-3UPy configurations two different conformations were taken into account: Conformation **a**, in which the dihedral angle formed by the atoms marked as 1, 2, 3 and 4 in Figure 33 had an approximate value of -77º and conformation **b**, in which it was close to -174º. All the possible combinations of CTV-3UPy configurations and conformations were considered:

-RaRa, in which the cage was formed by two CTV-3UPys with the same configuration R and the same conformation **a**.
-LbLb, composed of two CTV-3UPy(L) molecules with conformation **b**.
-RbLb, in which the CTV-3UPys had different configuration but the same conformation **b**.
-RaLb, with a CTV-3UPy of configuration R and conformation **a** and a CTV-3UPy of configuration L and conformation **b**.
-RbRb, with two CTV-3UPy with configuration R and conformation **b**.
-LaLa, with two CTV-3UPy with configuration L and conformation **a**.
-RaLa, in which the CTV-3UPys had different configuration but the same conformation **a**.
-RbLa, in which a CTV-3UPy with configuration R had conformation **b** and a CTV-3UPy with configuration L had conformation **a**.

Thus, the systems studied consisted of fullerenes [60-I_h]-, [70-D_{5h}]-, [76-D_2]-, [78-C_{2V}], [84-D_2] and [90-D_{5h}] trapped inside $(CTV\text{-}3UPy)_2$ capsules, formed by two CTV-3UPy structures bounded by hydrogen bonds. The nomenclature used for the complexes was the following: (type of the first CTV-3UPy molecule)(type of the second CTV-3UPy molecule)(fullerene). For instance, RbRb60 refers to a capsule composed of two CTV-3UPy molecules of the Rb type containing a fullerene C_{60}.

IV.4. Methods

The different capsule-fullerene systems were constructed using Maestro (Schrodinger Inc., New York, NY, USA) molecular modelling environment. Optimization of the geometries of these systems and of its components was carried out with the ADF program.[70-73] The numerical integration scheme applied for the calculations was developed by te Velde.[136, 137] The geometry optimization procedure was based on the method reported by Versluis and Ziegler.[138] The BP86 functional, described as a combination of local VWN exchange-correlation potential with nonlocal Becke's exchange correction[45] and Perdew's correlation correction, was applied.[46, 47] A double-ξ plus polarization basis set was used. From the optimized geometry, single point calculations with TZP basis were performed for the whole systems, for the capsule in the geometry of the system, for the CTV-3UPys and for the fullerenes. The energy of formation of the host and the host-guest interaction energy were obtained from these data. As commented in the previous chapter, this methodology was chosen as a balance between accuracy and computational cost, taking into account that it gave results very close to experimental data for the description of multiple hydrogen bonds between nitrogenated bases and between UPy monomers. In order to properly consider the π-stacking interactions between host and guest, an empirical correction term[64] was added to the BP86/TZP results (for more details see Chapter II). Furthermore, single point calculations were also done for some of the systems with the program NWChem.[207, 208] X3LYP and/or BH&H were applied with 6-311G** basis set. These calculations were carried out in Marenostrum supercomputer at BSC-CNS (Barcelona Supercomputing Center- Centro Nacional de Supercomputación).

A calculation of the molecular electrostatic potential distribution was done for the CTV-3UPy molecule and for the C_{60} fullerene in order to prove host-guest complementarity. Electrostatic

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8857-7/DL:T-1276-2008

potential correlates with dipole moment, electronegativity and partial charges so that its mapping provided a visual method to understand the relative polarity of the molecules.

IV.5. Results

IV.5.1. Electrostatic Potential

Figure 34 and Figure 35 show the electrostatic potential distribution of the CTV-3UPy molecule and of the fullerene, i.e. the potential energy of a proton at a particular location near the molecules. Negative electrostatic potential corresponds to an attraction of the proton by the concentrated electron density in certain parts of the molecule. Positive electrostatic potential corresponds to repulsion of the proton in regions were low electron density exists. An electronic charge density isosurface of the molecules is shown, which is coloured with potential values represented by colours in a continuous range from red to blue depending on the more/less positive/negative values of its different regions. Thus from most negative values to most positive values the range of colours is red, orange, yellow, green, light blue, dark blue. Green regions are the most neutral. Figure 34a shows a side-view of the electrostatic potential mapping for a CTV-3UPy molecule with a superposed scheme in white showing the location of the atoms. The most negative values are found surrounding the carbonyl groups of the pyrimidones (in red), followed by the N atoms of the same ring (in orange). This states the power of these groups to act as hydrogen bonding acceptors. The oxygen atoms in the ether groups and in the ureido groups are to some extent less positive (in yellow). The NH groups in the ureido chains and the NH groups in the pyrimidones are the most positive regions (in dark blue), stating their power as hydrogen bonding donors. This confirms the complementarity of two CTV-3UPy molecules bound in a capsule fashion, as the regions of most positive potential perfectly match the regions of most negative potential. Figure 34b represents the mapping for a C_{60} fullerene. As it can be seen, the whole isosurface is mostly positive, especially in the centre of the pentagonal and hexagonal rings. Figure 35 shows the outer (a) and inner (b) surfaces of the CTV-3UPy molecule. The map of the outer face (a) shows mostly green-blue colours, indicating neutral to positive values of the electrostatic potential. CTV moieties are clearly neutral. In contrast, the inner face (b) has several yellow regions at the accessible “ether” oxygens and at the aromatic rings in CTV. These different outside/inside properties displayed by the rings in the cyclotriveratrylene explains the affinity CTV-fullerene already reported[199], as the more negative electrostatic potential inside the cavity attracts the positive electrostatic potential of the fullerene.

A recent example in which another concave aromatic surface was exploited in the molecular recognition of C_{60} was the reported by Ortí, Martín and collaborators in 2007.[209] In this case π-extended derivatives of the tetrathiafulvalene (TTF) were proved to meet the requirements for the electron-donor partner to be used in the construction of optoelectronic devices.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

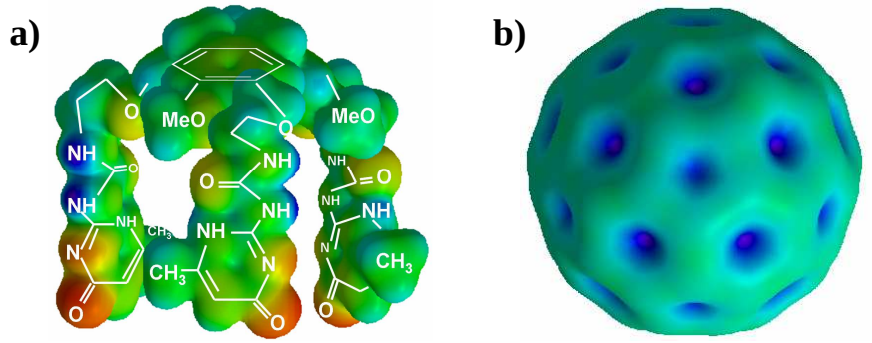


Figure 34. Electrostatic potential surfaces of a) CTV-3UPy molecule in a side-view and of b) fullerene C_{60}.

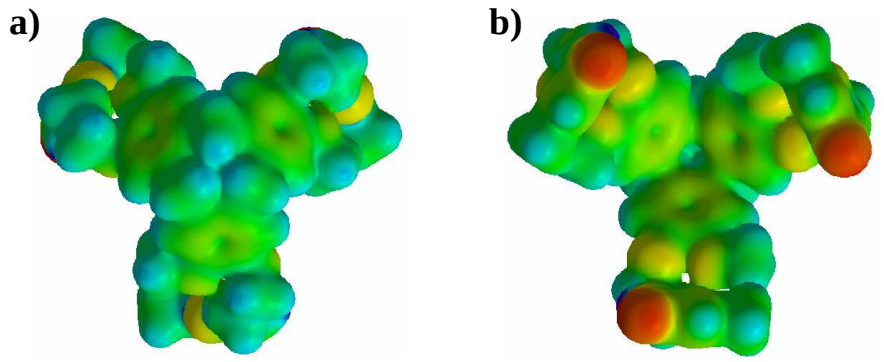


Figure 35. Electrostatic potential surfaces of the a) inner and b) outer faces of the CTV-3UPy molecule.

IV.5.2. C_n @ $(CTV\text{-}3UPy)_2$ (n=60,70)

The geometries of all the capsules $(CTV\text{-}3UPy)_2$ described above encapsulating C_{60} and C_{70} were investigated first. On the basis of those structures, the energy released when the complex host(H)-guest(G) is formed was dissected in several terms, which were evaluated separately by means of a series of additional calculations. Thus, we considered the interaction energy between two CTV-3UPy molecules to form the final void capsule, $\Delta E_{H\ form(sp)}$. This value was calculated as the energy of the entire $(CTV\text{-}3UPy)_2$ capsule minus the energy of the two corresponding CTV-3UPy units at the geometry of the complex, and it accounts for the arrangement of the twelve hydrogen bonds between the molecules. Then, the term $\Delta E_{H\ form(op)}$ includes the deformation suffered by each unit in order to adopt the geometry in the complex. So, this term was calculated as the energy of the entire $(CTV\text{-}3UPy)_2$ capsule minus the energy of the two CTV-3UPy molecules at their optimal geometry.

The host-guest interaction energy ($\Delta E_{H\text{-}G}$) was evaluated as the sum of two terms, namely $\Delta E_{H\text{-}G(e)}$ and $\Delta E_{H\text{-}G(d)}$. The former is the electronic interaction energy, calculated by a DFT method as the energy of the whole complex minus the energy of the host and the energy of the guest. $\Delta E_{H\text{-}G(d)}$ was

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8857-7/DL:T-1276-2008

the empirical correction term added for dealing with these weak interactions. The method used was described in detail in Chapter II. Finally the total interaction energy ΔE_{tot} was calculated as the sum of $\Delta E_{H\ form(op)}$ and $\Delta E_{H\text{-}G}$. The values for all the energetic terms just defined are collected in Table 8.

Regarding the formation of the capsule, let us remind the reader that the interaction energy for forming four hydrogen bonds in the UPy dimer was -55.3 kcal·mol^{-1}, as discussed in the previous Chapter. This corresponds to approximately -13.8 kcal·mol^{-1} for every interaction if the energy is divided equitable, which represents intermediate-to-strong hydrogen bonds. An ideal disposition of the three UPy dimers in the capsule similar to that of the isolated dimer would reach an upper value for $\Delta E_{H\ form(sp)}$ around -165.9 kcal·mol^{-1}. As it can be seen in Table 8, RbRb was the only conformer that reached a value close to the upper limit. In the rest of the hosts, the $\Delta E_{H\ form(sp)}$ term acquired lower values, being RbLa the second more strongly interacting capsule. The high energy of formation of the capsule RbRb is related to an almost perfect orientation of the UPy monomers. As Figure 36 shows for RaRa and RbRb capsules trapping C_{60}, the UPy dimers were forced to adopt a bent shape in both cases, but the angle between two UPy dimers were closer to planarity in RbRb than in RaRa.

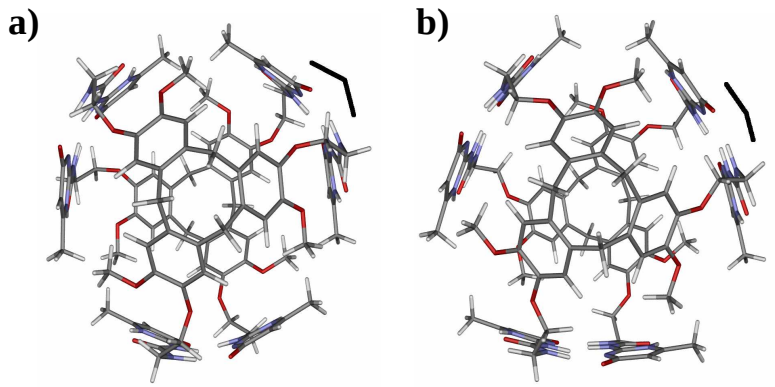


Figure 36. Top view of the capsules obtained by optimization of the a) RaRa60 and b) RbRb60 systems.

When the deformation suffered by the two CTV-3UPy units to form the dimer was taken into account, the $\Delta E_{H\ form(op)}$ values in Table 8 indicated that RbRb kept on being the capsule that gave highest interaction values, followed by RbLa and LbLb. Although not surprising, it is worth mentioning that the fullerenes did not suffer any appreciable deformation when forming the complexes.

After the formation of the host was studied, the other important contribution to the total energy was the interaction between the capsule and the fullerene. Note that the electronic interaction energy ($\Delta E_{H\text{-}G(e)}$) was always positive in all cases, i.e. repulsive, at this level of theory. This result was expected since DFT method used does not nicely describe the dispersion forces existing in the system (see Chapter II). However, the $\Delta E_{H\text{-}G(d)}$ empirical term was negative and rather large. Note that, for instance, a value of -187.3 kcal·mol^{-1} for C_{70}@RaRa complex would correspond to an interaction of 2.7 kcal·mol^{-1} per guest atom if divided equitable. It can be observed that the most favourable values of $\Delta E_{H\text{-}G}$ were obtained for RaRa by far, followed by RaLb and RbRb.

Summing up all the terms, i.e., the formation energy of the capsule, $\Delta E_{H\ form(op)}$, and the interaction energy host-guest, $\Delta E_{H\text{-}G}$, the total interaction energy ΔE_{tot} for RbRb was the largest for both C_{60} and C_{70} as a consequence of the balance between the two energetic contributions.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Table 8. Interaction energies (kcal·mol^{-1}) calculated with BP86/TZP method for C_{60} and C_{70} in RbRb, RaRa,RbLa, LaLa, LbLb, RaLa and RaLb.

	RbRb		RaRa	
	C_{60}	C_{70}	C_{60}	C_{70}
$\Delta E_{H\ form}(sp)^{(1)}$	-166.2	-166.9	-135.9	-144.3
$\Delta E_{H\ form}(op)^{(2)}$	-109.3	-110.7	-75.3	-88.2
$\Delta E_{H\text{-}G(e)}^{(3)}$	16.1	28.0	20.4	31.2
$\Delta E_{H\text{-}G(d)}^{(4)}$	-133.9	-174.1	-157.9	-187.3
$\Delta E_{H\text{-}G}^{(5)}$	-117.8	-146.1	-137.5	-156.1
$\Delta E\ tot^{(6)}$	-227.1	-256.8	-212.8	-244.3

	RbLa		LaLa	
	C_{60}	C_{70}	C_{60}	C_{70}
$\Delta E_{H\ form}(sp)^{(1)}$	-157.5	-158.9	-137.8	-142.1
$\Delta E_{H\ form}(op)^{(2)}$	-101.7	-104.5	-84.9	-92.1
$\Delta E_{H\text{-}G(e)}^{(3)}$	14.6	25.8	11.5	20.3
$\Delta E_{H\text{-}G(d)}^{(4)}$	-128.3	-163.5	-123.7	-155.2
$\Delta E_{H\text{-}G}^{(5)}$	-113.7	-137.7	-112.2	-134.9
$\Delta E\ tot^{(6)}$	-215.4	-242.2	-197.1	-227.0

	LbLb		RaLa	
	C_{60}	C_{70}	C_{60}	C_{70}
$\Delta E_{H\ form}(sp)^{(1)}$	-134.9	-134.9	-145.9	-144.9
$\Delta E_{H\ form}(op)^{(2)}$	-102.6	-104.4	-88.6	-88.0
$\Delta E_{H\text{-}G(e)}^{(3)}$	18.6	24.7	12.6	23.4
$\Delta E_{H\text{-}G(d)}^{(4)}$	-132.4	-163.1	-125.5	-165.7
$\Delta E_{H\text{-}G}^{(5)}$	-113.8	-138.4	-112.9	-142.3
$\Delta E\ tot^{(6)}$	-216.4	-242.8	-201.5	-230.3

	RaLb	
	C_{60}	C_{70}
$\Delta E_{H\ form}(sp)^{(1)}$	-132.8	-131.2
$\Delta E_{H\ form}(op)^{(2)}$	-88.0	-87.8
$\Delta E_{H\text{-}G(e)}^{(3)}$	15.3	24.8
$\Delta E_{H\text{-}G(d)}^{(4)}$	-134.1	-173.0
$\Delta E_{H\text{-}G}^{(5)}$	-118.8	-148.2
$\Delta E\ tot^{(6)}$	-206.8	-236.0

(1), (2) Interaction energy for the host formation
(1) Energy of the $(CTV\text{-}3UPy)_2$ minus energy of the twoCTV-3UPy measured as single points in the geometry of the dimer.
(2) Energy of the $(CTV\text{-}3UPy)_2$ minus energy of the twoCTV-3UPy optimized.
(3) Electronic host-guest interaction.
(4) Dispersion host-guest interaction.
(5) Sum of (3) and (4).
(6) Total interaction energy as sum of (2) and (5).

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8857-7/DL:T-1276-2008

If we compare C_{60} and C_{70} in the same host, the energy of formation of the respective capsules was very similar in all cases except in RaRa and LaLa, in which C_{70} was more favoured by 12.9 kcal·mol^{-1} and 7.2 kcal·mol^{-1}, respectively. In all the capsules the main difference between the two fullerenes was mainly due to the host-guest interaction energy term, $\Delta E_{H\text{-}G}$, which was always much more favourable for C_{70} than for C_{60} (in 20-40 kcal·mol^{-1}), what made the total interaction energy be larger for the C_{70}.

Consequently, the results in Table 8 indicated that RbRb capsule showed the best total energy of formation compared to RaRa, RbLa, LaLa, LbLb, RaLa and RaLb complexes. C_{70} was always more favoured than C_{60} mainly because of the host-guest interaction although in some cases also a better stabilization of the host contributed.

Some calculations were also carried out with X3LYP and BH&H DFT functionals. With them, the tendency in the results of the energy for the host formation was the same as with BP86/TZP method, although the absolute values obtained were larger. For instance, when the method BP86/TZP was used, the term $\Delta E_{\ H\ form}$(sp) was -166.2 kcal·mol^{-1} for the complex RbRb60. When X3LYP and BH&H were used, the respective values were -192.4 kcal·mol^{-1} and -194.7 kcal·mol^{-1}.

Slightly attractive values were obtained for the host-guest interactions. For the complex RaRa60, the term $\Delta E_{\ H\text{-}G(e)}$ had values of -1.4 kcal·mol^{-1} and -4.6 kcal·mol^{-1} with X3LYP and BH&H respectively. However, for the complex RaRa70, the values obtained were 4.9 kcal·mol^{-1} and 0.2 kcal·mol^{-1}. This pointed out that none of the two methods was describing well the weak interactions in the system. Hence, the applying of a dispersion correction term, like the used with BP86/TZP method, would be still necessary using these functionals.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

IV.5.3. Higher fullerenes C_{76}, C_{78}, C_{84} and C_{90}

The results obtained in Mendoza's laboratories indicated that some larger fullerenes might be also trapped inside the capsule, and we were inquired about such a possibility. Therefore, higher fullerenes were considered and we investigated whether they could be encapsulated or not, and evaluated possible geometries. The best host for C_{60} and C_{70}, RbRb, and two other hosts, RaRa and RbLa, were chosen and the corresponding host-guest complexes with C_{76}, C_{78}, C_{84} and C_{90} were built, and studied following the computational protocol described above. Table 9 shows the decomposition of the interaction energy in typical terms for the new complexes together with the previous results for C_{60} and C_{70}, for an easy comparison.

Table 9. Interaction energies (kcal·mol^{-1}) calculated with BP86/TZP method for several fullerenes in RbRb, RaRa and RbLa capsules.

	RbRb					
	C_{60}	**C_{70}**	**C_{76}**	**C_{78}**	**C_{84}**	**C_{90}**
$\Delta E_{H\ form}$(sp) (1)	-166.2	-166.9	-168.3	-170.2	-166.3	-166.7
$\Delta E_{H\ form}$(op) (2)	-109.3	-110.7	-113.1	-113.4	-110.3	-104.6
$\Delta E_{H\text{-}G(e)}$ (3)	16.1	28.0	33.7	33.6	40.0	56.9
$\Delta E_{H\text{-}G(d)}$ (4)	-133.9	-174.1	-190.2	-194.6	-209.8	-228.9
$\Delta E_{H\text{-}G}$ (5)	-117.8	-146.1	-156.5	-161.0	-169.8	-172.0
ΔE tot (6)	-227.1	-256.8	-269.6	-274.4	-280.1	-276.6
	RaRa					
	C_{60}	**C_{70}**	**C_{76}**	**C_{78}**	**C_{84}**	**C_{90}**
$\Delta E_{H\ form}$(sp) (1)	-135.9	-144.3	-146.7	-148.7	-148.5	-149.2
$\Delta E_{H\ form}$(op) (2)	-75.3	-88.2	-90.8	-92.5	-91.9	-84.6
$\Delta E_{H\text{-}G(e)}$ (3)	20.4	31.2	36.5	35.7	41.0	56.8
$\Delta E_{H\text{-}G(d)}$ (4)	-157.9	-187.3	-199.3	-200.2	-213.3	-227.8
$\Delta E_{H\text{-}G}$ (5)	-137.5	-156.1	-162.8	-164.5	-172.3	-171.0
ΔE tot (6)	-212.8	-244.3	-253.6	-257.0	-264.2	-255.6
	RbLa					
	C_{60}	**C_{70}**	**C_{76}**	**C_{78}**	**C_{84}**	**C_{90}**
$\Delta E_{H\ form}$(sp) (1)	-132.8	-131.2	-159.7	-160.7	-161.0	-162.4
$\Delta E_{H\ form}$(op) (2)	-88.0	-87.8	-104.7	-106.1	-106.8	-101.4
$\Delta E_{H\text{-}G(e)}$ (3)	15.3	24.8	33.3	31.8	35.2	48.0
$\Delta E_{H\text{-}G(d)}$ (4)	-134.1	-173.0	-179.2	-181.7	-198.7	-211.9
$\Delta E_{H\text{-}G}$ (5)	-118.8	-148.2	-145.9	-149.9	-163.5	-163.9
ΔE tot (6)	-206.8	-236.0	-250.6	-256.0	-270.3	-265.4

(1), (2) Interaction energy for the host formation
(1) Energy of the (CTV-3UPy)$_2$ minus energy of the two CTV-3UPy measured as single points in the geometry of the dimer.
(2) Energy of the (CTV-3UPy)$_2$ minus energy of the two CTV-3UPy optimized.
(3) Electronic host-guest interaction.
(4) Dispersion host-guest interaction
(5) Sum of (3) and (4).
(6) Total interaction energy as sum of (2) and (5).

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8857-7/DL:T-1276-2008

If we analyze the interaction of each fullerene with the different hosts, it can be seen that the RbRb capsule kept on being the most favoured regarding the formation of the host. $\Delta E_{H\text{-}form(sp)}$ and $\Delta E_{H\text{-}form(op)}$ reached the highest values in these complexes, followed by RbLa capsule. The host-guest interaction energy was in all cases slightly higher in RaRa, followed by RbRb complexes. For all the fullerenes, the most favourable total interaction energy was obtained with the RbRb host, as a result of the better balance between the two energetic contributions. From the values of ΔE_{tot} in Table 9, it can be concluded that the affinity of the capsule towards C_{84} is much higher than to C_{60} and C_{70}. As it will be shown afterwards, this computational prediction was subsequently verified experimentally. Moreover, the results collected in Table 9 indicate that C_{76}, C_{78} and C_{90} can be also trapped as well.

Regarding the relative energies of formation of the preferred RbRb capsules, the differences between C_{60}, C_{70}, C_{76}, C_{78} and C_{84} complexes were very low, less than 3 kcal·mol^{-1}. However, the capsule for C_{90}@(CTV-UPy)$_2$ showed a value slightly less favourable than the rest. It had an energy 5 kcal·mol^{-1} above that of C_{84} system, probably due to the bigger size of this fullerene that induce steric strain in the capsule. On the other hand, the differences in the host-guest interaction energy among all the systems were more evident. $\Delta E_{H\text{-}G}$ showed more favourable values as the fullerene size grew up. An interval of energy of 54.2 kcal·mol^{-1} existed between the highest and the lowest $\Delta E_{H\text{–}G}$ values, i.e., from C_{60} to C_{90}. The interaction energy increased as the size of the fullerene increased too, due the larger number of atoms in the fullerene interacting with the capsule. But as the cavity got filled by guest, the difference in the host-guest interaction energy between fullerenes of increasing size decreased: 28.3 kcal·mol^{-1} when going from C_{60} to C_{70}; 10.4 kcal·mol^{-1} from C_{70} to C_{76}; 4.5 kcal·mol^{-1} from C_{76} to C_{78}; 8.8 kcal·mol^{-1} from C_{78} to C_{84}; and 2.2 kcal·mol^{-1} from C_{84} to C_{90}. This little difference in the host-guest interaction between C_{84} and C_{90} and the host-formation energy being 5 kcal·mol^{-1} more favourable for C_{84} made the total interaction energy for this fullerene the largest. The systems closest in energy to the preferred C_{84} were formed with C_{90} and C_{78} with differences of only 3.5 kcal·mol^{-1} and 5.7 kcal·mol^{-1} respectively.

In general in RbRb architectures, the formation of the capsule contributed to the total interaction energy in a 42% whereas the host-guest interaction contributed in a 58%. In RaRa type systems these percentages were of 35% and 65%, respectively, and in RbLa complexes they were of 40% and 60%. Therefore, in this aspect no large differences between the capsules were found.

Based on the results shown in Table 9, we can establish an scale of affinity between the (CTV-UPy)$_2$ capsule and the set of fullerenes considered in this study. This scale would follow the order $C_{60}<C_{70}<C_{76}<C_{78}<C_{90}<C_{84}$.

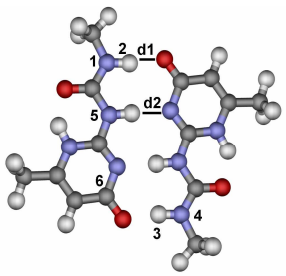


Figure 37. Scheme for the understanding of the geometric parameters measured in the capsules as well as in the isolated UPy dimer.

Several geometric parameters measured in an isolated UPy dimer are collected in Table 10 and are also schematically represented in Figure 37 for clarity. **d1** and **d2** stand for N-H···O and N-H···N hydrogen bonding distances. The former showed shorter than the latter with respective values of 1.680 and 1.843Å.

Table 10. Geometric parameters for the ideal UPy dimer.

	Upy dimer
d1[(1)]	1.680
d2[(2)]	1.843
dih[(3)]	179.7
angle1[(4)]	172.7
angle2[(5)]	179.2
angleCurv[(6)]	180

(1) , (2) Hydrogen bonding distances marked in Figure 37
(3) Torsion angle between atoms 1, 2, 3 and 4.
(4) Angle in the hydrogen bond of N-H···O type.
(5) Angle in the hydrogen bond of N-H···N type.
(6) Angle formed by the atoms 1, 5 and 6.

The parameters **angle1** and **angle2** represent the angles between the atoms involved in the hydrogen bonding: N-H-N and N-H-O. The angle **angle2** adopted a value of 179.2º closer to linearity than **angle1**, with a value of 172.7º. **dih** was the torsion angle between atoms 1, 2, 3 and 4 in Figure 37 and thus informed about the extent to what the two UPy monomers were arranged in the same plane. A value very close to 180º indicated planarity. **angleCurv** was the angle formed by the atoms 1, 5 and 6 so that it gave evidence of the degree of curvature (in length) of the UPy monomers. Its value was also 180º, meaning that the monomers were flat.

All these and some additional geometric parameters were measured for the capsules obtained in the RbRb complexes in order to see to what extent the dimers deviated from the geometry of the isolated UPy dimer, and to identify differences between the systems with different fullerenes. The results are shown in Table 11.

As it can be seen, the hydrogen bond distances (**d1** and **d2** in Table 11 and Figure 37) were very similar between all the capsules and much close to the ideal values of 1.680 and 1.843Å for the dimer. **angle1** and **angle2** adopted values lower than in the ideal dimer indicating a certain loss of linearity concerning the three atoms directly involved in the interaction. It is remarkable that as the size of the fullerene grew up, the values of the angles went closer to 180º.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8857-7/DL:T-1276-2008

Table 11. Geometric parameters measured from the obtained RbRb capsules with the several fullerenes.

RbRb	C_{60}	C_{70}	C_{76}	C_{78}	C_{84}	C_{90}
d1(1)	1.678	1.674	1.673	1.673	1.687	1.680
d2(2)	1.834	1.843	1.853	1.852	1.883	1.854
dih(3)	160.1	163.7	164.3	165.2	159.0	160.7
angle1(4)	164.7	168.7	169.6	170.1	170.6	174.2
angle2(5)	164.1	172.9	174.0	174.0	172.5	173.6
angleCurv(6)	174.5	174.5	175.0	175.0	171.5	175.1
distDomes(7)	15.9	15.9	16.1	16.1	15.9	17.0
angleCTV(8)	109.9	111.6	114.0	115.7	113.7	118.8
distUPy(9)	11.6	11.7	11.8	11.8	12.4	11.6
AvrgMinD(10)	4.561	4.338	4.143	4.105	3.841	3.690
StandardDev(11)	0.767	0.728	0.743	0.712	0.631	0.577

(1) , (2) Hydrogen bonding distances marked in Figure 37.
(3) Torsion angle between atoms 1, 2, 3 and 4.
(4) Angle in the hydrogen bond of N-H···O type.
(5) Angle in the hydrogen bond of N-H···N type.
(6) Angle formed by the atoms 1, 5 and 6.
(7) Distance between the two domes of the CTV-3UPys.
(8) Angle formed by the centroid of consecutive CTV aromatic rings and the carbon linking them.
(9) Average distance from centroids of the bonds linking the pyrimidone ring to the ureido group in consecutive UPys of a CTV-3UPy molecule.
(10) Average minimum host-guest distance.
(11) Standard deviation for (10).

The parameters **dih** and **angleCurv** adopted values that ranged 160º-165º and 171º-175º respectively, thus indicating a certain loss of planarity concerning the UPy dimers and UPy monomers (slightly bent in length). This deformation is needed to adopt a shape that maximizes the interaction of the capsule with the guests. Indeed, the system formed with C_{84} had the lowest values for both terms, indicating a larger adaptation of the capsule to this particular guest. The UPy monomers adopted a slightly bent shape (see Figure 38) without affecting the strength of the hydrogen bonds. The capsules formed with C_{84} were the most symmetrical, i.e., the less deformed if we focus on the alignment of the UPy monomers. In this way, this guest allows the UPy monomers forming each dimer to be closer to an ideal spatial distribution.

Figure 38 shows how the geometry of the capsule changed in the weakest and the strongest interaction with the fullerene inside (C_{60} and C_{84} respectively).

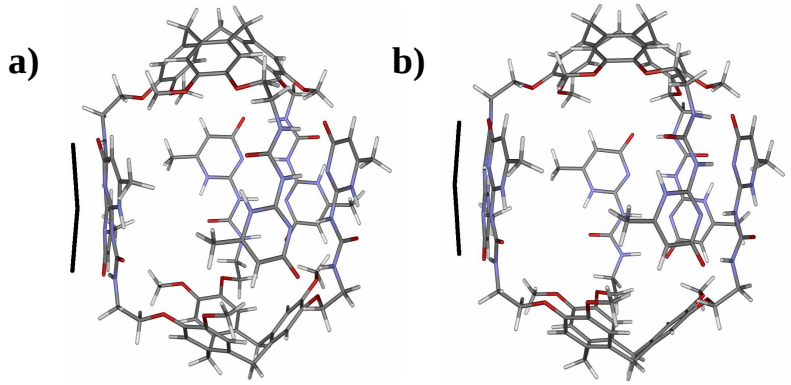


Figure 38. Side view of the capsules obtained from the optimization of a) RbRb60 and b) RbRb84 systems.

Also in Table 11, **distDomes** indicated the distance between two centroids located in the middle of the two CTV moieties. It represented the maximum distance between opposite sides in the capsule. This value was around 16 Å in all the cases. However, in the C_{90} complex it was slightly larger (17.0Å).

The term **angleCTV** indicated the angle existing between the centroids of two consecutive CTV aromatic rings and the carbon atom that linked both rings. In Table 11, it can be seen that that value increased with the size of the fullerene. This stated a certain ability of the CTV molecule to adapt to the guest. As larger guests occupied more space in the cavity, the capsule adopted a shape by enlarging this cavity in order to adapt better to the guest. RbRb84 system is a particular case. C_{84} was one of the largest fullerenes but the **angleCTV** parameter did not reach values as high as in RbRb78 and RbRb90 complexes. The reason was that due to the more spherical shape in C_{84}, the space left in the proximity of the CTV moieties was smallest.

The parameter **distUPy** indicated the distance between consecutive UPys belonging to the same CTV-3UPy molecule. The parameter was measured between centroids of bonds linking the pyrimidone rings and the ureido chains. The values in all the cases were very similar. In RbRb84 this term presented a slightly higher value, indicating a larger width of the capsule.

AvrgMinD is the average host-guest minimum distance. All the distances from each atom in the guest to the atoms in the host were measured, and the minimum value for each was kept. The average of all the minimum values was then calculated. **StandardDev**, also in Table 11, is the corresponding standard deviation. The lowest values for **AvrgMinD** were obtained for C_{90} and C_{84}, which meant that these two fullerenes were those that fulfilled the cavity perfectly. Furthermore, also in these two cases, the **StandardDev** was the smallest, indicating that the host-guest distance, and as a consequence the host-guest interaction, was more homogeneous than in the rest of the systems.

IV.6. Experimental results

In the experiments carried out by Mendoza's group, the affinity of the capsule for C_{60}, C_{70} and C_{84} showed to be increasing in this order.[210],[211] When the molecule CTV-3UPy was mixed with a

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8857-7/DL:T-1276-2008

concentration of 1.5 mM of fullerene, UV/Vis measurements resulted in association constants of $1.93*10^{6}M^{-2}$, $7.4*10^{7}M^{-2}$ and $2.63*10^{10}M^{-2}$ for C_{60}, C_{70} and C_{84}, respectively.

Solid-liquid extractions of fullerite were performed in THF (polar and aprotic), were the fullerenes are poorly soluble. After stirring the fullerene suspension in the THF with CTV-3UPy at room temperature, solids were filtrated and the clean filtrate was analyzed by reverse phase HPLC. After that, release of the fullerenes from the capsule was very simple. Addition of few drops of trifluoroacetic acid (TFA) to the THF solutions of the complexes caused the hydrogen-bonded network to break and the fullerenes to precipitate. The host could then be recycled by evaporation.

Fullerite was determined to be 65% of C_{60}, 22% of C_{70}, 5% of C_{84} and 8% of other materials. Variable concentrations of CTV-3UPy were tested and it was found that high receptor/fullerite ratios (ca. 50-60%) were better suited for a selective extraction of C_{70} whereas the selectivity for C_{84} was maximized for ca. 5-15% receptor/fullerite ratios (Figure 39). The amount of C_{60} extracted remained constant at all ratios of extractant, indicating that this represented the inherent solubility of C_{60} in THF.

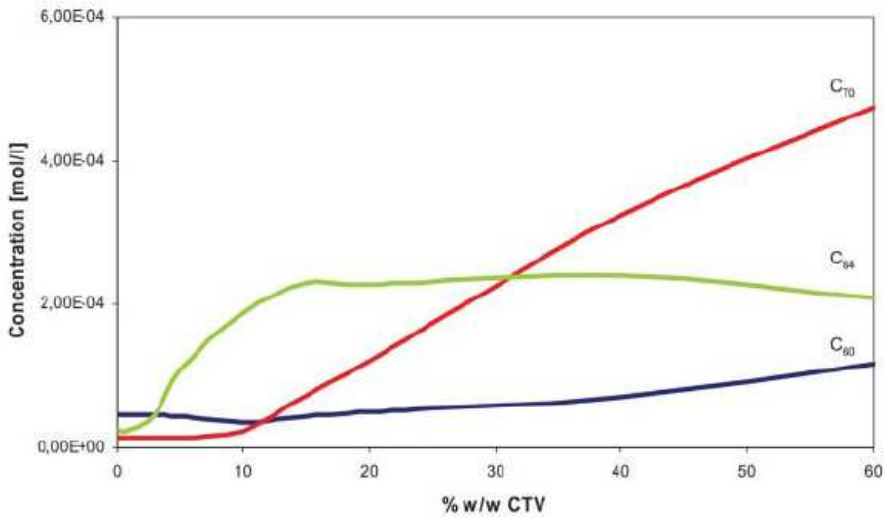


Figure 39. Plot of concentration of fullerenes extracted against amount of receptor CTV-3UPy used (w/w with respect to the solid fullerite). Reproduced from ref [211].

The amount of solvent (concentration) was also relevant. Indeed better values for selectivity were achieved when the amount of solvent was reduced, as the amount of C_{60} solubilised decreased. Thus enrichment in C_{84} up to 76% was achieved after one single extraction. Further extractions of the resulting C_{84}-enriched mixture did not result in any improvement of C_{84} purity due to the inherent solubilities of residual C_{60} and C_{70}. However, wash-out with some THF allowed the purity of C_{84} to rise up to 85%.

Thus, playing with the conditions of the extraction one fullerene or the other could be purified. Other higher fullerenes were also extracted with CTV-3UPY, although the amount of these (<2%) was very low in fullerite.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

IV.7. Conclusions

The electrostatic potential distribution of the capsules and the fullerenes stated for a clear tendency of the CTV-3UPy molecules to form capsules by self-assembly and for a strong attraction between the inner walls of these capsules and the fullerenes.

By combining different configurations and conformations of the CTV-3UPy molecules, which the herein called RbRb was the preferred, seven different hosts could form. This architecture presented the best compromise between the two terms contributing to the total interaction energy: formation of the host and host-guest interaction.

Concerning the systems composed of RbRb capsule and different fullerenes, the interaction energy for the formation of the host was not very different among the complexes but it was slightly more unfavourable for RbRb90 than for the rest. The reason was a larger deformation of the CTV-3UPy molecules to reach the geometry of the final complex.

In general the differences in the host-guest interaction depending on the fullerene were larger than the differences in the formation of the host. The energy became more favourable as the size of the fullerene did, so that RbRb90 complex was the one with the maximum interaction. However, in this particular case the difference in energy respect to RbRb84 system was less than 3 $kcal \cdot mol^{-1}$. As the formation of the capsule in this case was the most unfavourable, this made the total interaction in RbRb84 be better than in RbRb90.

The geometry of the capsule adapt to higher fullerenes by increasing either its width, its high or by modifying the orientation of the UPy units depending on the shape of the guest. In any of these three cases the changes in the geometry are mostly reflected in the mutual arrangement of the UPy monomers when forming the dimer. However, the changes only slightly weakened the hydrogen bonding interactions.

The size and shape of fullerene C_{84} was “the best”, so it enabled a homogenised adaptation of the capsule to the guest. This permitted a nice relative disposition of the atoms involved in the hydrogen bonding although the UPy monomers were slightly bent in shape.

In full agreement with the experiments mentioned above, the theoretical results indicated that the affinity of the capsule for C_{70} was greater than the affinity for C_{60}.[210] Furthermore, our results predicted an even better value for C_{84}, and this fact was later experimentally confirmed.[211] The results showed that other fullerenes could be trapped in the capsules. In a scale of affinity, C_{76}, C_{78} and C_{90} would be located in between C_{70} and C_{84}. However, these fullerenes are in very small quantities in fullerite, which makes performing quantitative titrations very difficult.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-6857-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Chapter V

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

V.p53TD versus R337H p53TD. "A theoretical approach"

V.1. Introduction to the protein p53

p53[212],[213, 214] is a tumour suppressor gene that was discovered in 1979 while studying SV40 virus. It was observed that in infected cells the concentration of p53 was much higher than in normal cells. SV40 T-antigen was found binding the protein. At the same time another research group showed that animals with tumorigenic cells expressed p53.[215] The fact that the protein was present in tumours with no known viral etiology indicated that it was coded by resident cellular genes, although it could also be possible that some viruses coded for it. The first supposition was that p53 was an oncogene that could transform cells. However, successive studies carried out in 1989 and based in transfections on tumorigenic cells demonstrated that it was not the case.[216] When mixing wild type p53cDNA clone and mutant p53cDNA clone in the same cell culture, the only transformed cells that survived to the experiment had selected a mutant clone.[217] Thus p53 turned out to be a tumour suppressor and not an oncogene. This conclusion leaded to the consequent realization that SV40 transforms cells by inactivating the p53 function.[217] Also in 1989 an independent work was presented that proved the important role that p53 plays in human cancer. Two colon cancers were reported to contain point mutations in one copy of the p53 gene while the other allele was lost.[218] Since then on, a large amount of cancers in which the protein has found to be mutated have been reported.[219]

p53's name is related with its mass of 53kDa. The gene that codifies for it is in the short arm of the chromosome 17(17p13).[220] It is an ubiquitous protein in the nucleus of the cells that has a wide range of functions, all of them doomed to safeguard genomic integrity. This is why it is often called "the guardian of the genome".

V.1.1. Main functions

In normal conditions the tumour suppressor protein is present in a latent state and maintained at low levels. When a cell notices any kind of DNA damage it initiates specific signalling pathways that lead to p53 activation. Furthermore an increase of the protein levels is caused. All this is controlled by extensive posttranslational modifications.[221] Figure 40 shows the complexity of the network of proteins influencing p53 known until now. The huge and increasing amount of proteins intervening makes it extremely difficult to reach a deep and concrete knowledge about the processes in every case. Once p53 is active, it triggers new processes in order to stop the cell cycle[222] so that the cell can not divide. Thus it avoids the spread of the DNA defects to next generations.

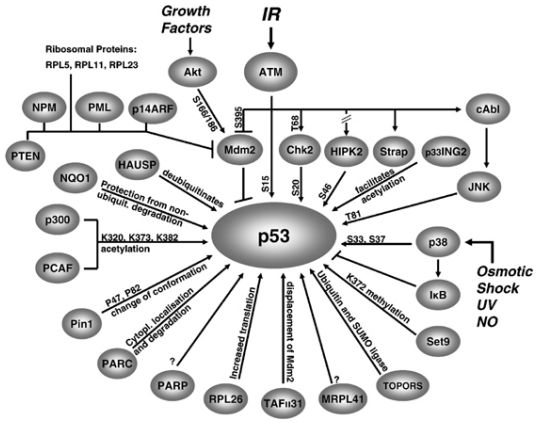


Figure 40. Proteins interacting with p53 for its activation. Figure reproduced from ref [221].

The cell cycle is formed by four different phases (Figure 41): G1 phase, S phase, G2 phase and M phase. In the M phase the cell divides in two new cells. In the following G1 phase the daughter cells restore the normal biosynthetic processes and synthesize enzymes involved in DNA replication which they will need in the next phase. In the S phase the chromosomes replicate each one in two sister chromatids. In G2 phase microtubules are produced, which are essential for the mitosis. After that the cell enters again in the M phase. The cell cycle checkpoints are systems to monitor and control the progress of the cell cycle. As the cell cycle progress some verification tests are carried out in order to see if the evolution is completely normal or not. In the latter case the cell cycle is stopped. The protein p53 is involved in these kind of mechanisms.

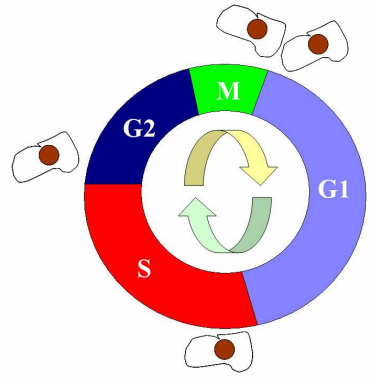


Figure 41. The cell cycle is composed of four stages: mitosis (M) and 3 stages composing the called interphase (G1, S, G2).

The growth arrest by p53 was thought until 2000 to be carried out at the G(1)/S or G(2)/M transitions of the cell cycle.[222] However, some evidences have appeared recently that point out the participation of p53 at multiple checkpoints in the cell cycle.[223] After stopping it, p53 itself triggers new processes for the DNA reparation.[224] It seems that p53 affects one of the five DNA-repair mechanisms[225] ,called NER (nucleotide-excision repair), in a transactivation independent manner,

by affecting the function of proteins involved in the process. In case of irreparable damage the tumour suppressor initiates the signal for the apoptosis[226] (or programmed cell death PCD).
p53 acts primarily as a transcription factor rising the levels of some proteins such as p21[227-230] , GADD45[229, 231]and 14-3-3-σ[232, 233] that establish signalling networks[234] by a multitude of protein-protein interactions. The latest activated proteins in this cascades are from the family of the cyclin-dependent kinases[235] (CDK), as for example cdk2[236] and cdc2.[237] The former stops the cell cycle before the S phase while the latter does it before the M phase. In a similar way p53 can increase the levels of other proteins such as Bax[238] or Apaf1[239, 240] which also initiate a cascade of reactions[234] that finally activates proteins from the family of the caspases[241], essential mediators of the apoptosis processes.

p53 has also been reported as a cellular senescence mediator in response to potentially oncogenic stimuli, although many questions remain regarding how p53 senses senescence-inducing stimuli and how it acts afterwards.[242]. Lately it has been found to be important also in the regulation of maternal reproduction[243] so that lack of the protein in female mice leads to significant decreases in embryonic implantation, pregnancy rate and litter size.

Thus, p53 is one of the critical proteins that regulate cell proliferation and survival. It is part of the mechanisms that give the organisms the capability to recover even when an oncogenic mutation occurs. Understanding of these innate mechanisms is suggesting novel therapeutic strategies for cancer.

It is to say that p53 has many other functions in addition to the mentioned above, all of them basic for the cell. Its versatility and the great amount of pathways in which intervenes are still mostly unclear although the advances since 28 years ago. The reality is that the more are the advances in p53 knowledge the greater is the sum of mysteries to be solved.

V.1.2. Structure

p53 is a multidomain protein.[244, 245] It is a homotetramer of 4x393 residues. Each chain consists of two folded domains (the core or DNA binding domain (94-294) and the tetramerization domain (323-360) that are linked by an intrinsically disordered sequence. The N-terminal transactivation domain (1-67), the proline-rich region (67-94), the nuclear localization signal (NLS)-containing region (303-323), and the C-terminal negative regulatory domain (360-393) are also intrinsically disordered.

The N-terminal transactivation domain[246, 247] activates transcription factors, the proline-rich domain[248-250] is important in the apoptotic activity, the central DNA-binding core domain[251, 252] is the responsible for the binding to DNA, the tetramerization domain[253] maintains the four chains together, which is essential for the activity of the protein in vivo, and the C-terminal domain[254, 255] is involved in down regulation of DNA binding of the central domain.

The structure of the DNA binding domain was solved by NMR[256] and crystallography.[257, 258] This was also the case for the tetramerization domain (NMR[259] and crystallography[260, 261]). However, solving the whole protein at once to see the quaternary structure was extremely difficult because of the globular domains being linked by disordered regions. Researchers were unable to crystallize the complete architecture. Lately some approaches were reached in this sense[245] by applying an strategy based on fitting the high resolution structures of individual domains, solved by x-ray crystallography or NMR, into low resolution SAXS (small angle x-ray structures scattering) and EM (electron microscopy data), guided by TROSY NMR constraints on interfaces. Approximate models showed in Figure 42 and Figure 43 of p53 alone and p53 complex with DNA were obtained.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

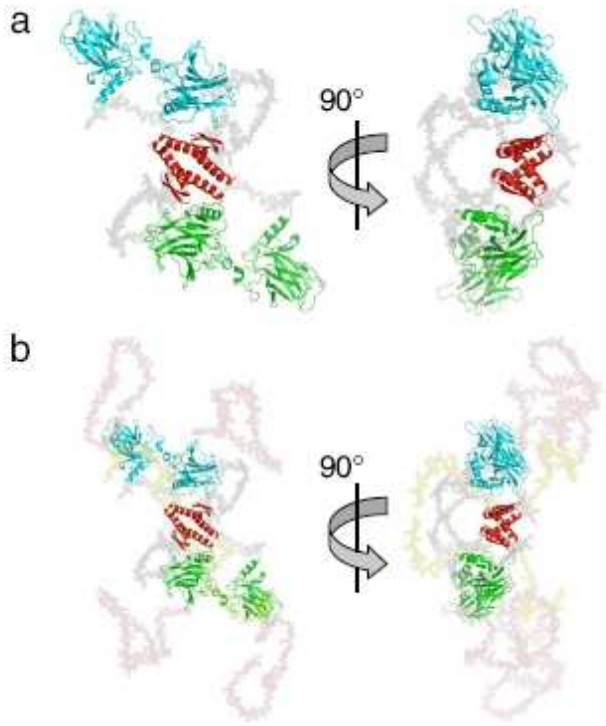


Figure 42. SAXS models of free p53 in solution from rigid body analysis and addition of missing fragments. (a) Only protein core (in blue) and tetramerization domain (in red) are shown. (b) The whole protein is shown. Core domains and tetramerization domain are displayed in cartoon representations, connecting linkers in gray, N-termini in salmon and C-termini in yellow, in semitransparent spacefill mode. Figure reproduced from ref [245].

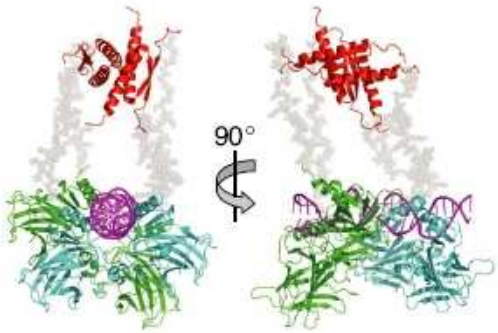


Figure 43. Rigid body model of a p53-DNA complex from SAXS data. Core domains, tetramerization domain and DNA are shown in cartoon representation, connecting linkers in semitransparent spacefill mode. The model is displayed in two orthogonal views. Figure reproduced from ref [245].

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

V.1.3. p53 in human cancer

A decrease of the levels of functional p53 can lead to the development of different kinds of cancers.[212-214] Indeed p53 mutations can be found in 50% of human cancers. Inactivation of the p53 gene is essentially due to small mutations which lead either to the expression of a mutant protein in the 90% of the cases or to a complete absence of the protein.

Li-Fraumeni Syndrome[262] is an example of autosomal dominant genetic cancer caused by mutations in p53 gene. People that suffer Li-Fraumeni Syndrome suffer from different types of tumours: breast cancer, osteosarcoma, soft tissue sarcoma adrenocortical carcinoma, leukaemia, astrocytoma and meningioma are the most common but also others have been associated such as gastric cancer, pharyngeal cancer, choroid plexus carcinoma, pancreatic cancer, melanoma, germ cell tumours, Willm´s tumour, colorectal cancer, Phyllodes tumour, ovarian cancer, thyroid cancer, endometrial cancer, prostate cancer and cervical cancer. This syndrome represents an extreme case since, as commented, multiple kinds of cancers are involved which moreover appear at early ages.

In some cases p53 mutations appear because of exposure to external factors. Skin cancer by radiation[263], lung cancer by exposure to smoking[264, 265] or liver cancer by exposure to Aflatoxin B1[266] are among them. p53 mutations are also important in gastric cancer[267], breast cancer[268], colorectal cancer[269], prostate cancer[270], cervical cancer[271], oral cancer[272], lymphoma[273, 274], leukaemia[275], ovary cancer[276], bladder cancer[277] and pancreatic cancer.[278] However, as well as mutations, there are also other forms of suffering a decrease in p53 function. For instance, there are various DNA viruses that encode proteins that target p53.[279, 280] Some tumours show nuclear exclusion of the protein[281] or an overexpression of MDM2, which is one of the proteins that maintain p53 in an inactive state in normal conditions.[282] Also mutations in the proteins involved in the cascade that activates p53 by phosphorylation can cause a decrease of its functions. [283]

V.1.4. Anti-cancer drugs related with p53

There already exist some anticancer drugs[284] based somehow on the activation or deactivation of the wild-type and the mutant p53s. The induction of the transcription and/or translation towards the wild type protein as well as the inhibition of its degradation, leads to an accumulation of p53.

Some DNA damaging agents such as radiation, DNA intercalation molecules and DNA strand breaking drugs are the most used chemotherapeutics. The latter are those that cause a greater increase of p53 protein levels.[285] DNA damaging agents are supposed to induce p53 by inhibiting MDM2 transcription. In normal conditions p53 induces the transcription of several proteins, MDM2 between them. At the same time, MDM2 is one of the proteins that regulate p53 stabilization. It binds to the N-terminus of the protein causing its ubiquitination and degradation by the proteosoma.[286-290] Thus, in a negative feed-back manner wild type p53 activates its own degradation. If the anticancer drugs low the levels of MDM2, the levels of p53 increase.

Specific inhibitors of the RNA metabolism cause p53- and p21-dependent reversible G1 arrest in the absence of detectable DNA damage and trigger effects that end up in an increase of wt p53 levels.[291-293] Other anti-carcinogenic are inhibitors of the nuclear export of the protein. The degradation of p53 takes place out of the nucleus so that the previous export is necessary.[294] It is carried out by MDM2[295] which has a nuclear export signal and acts as a shuttle. It is possible that ubiquitination of p53 is necessary to go through the membrane of the nucleus. Some drugs inhibit the transport from the nucleus to the cytoplasm.[294-296]

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

On the other part Geldanamycin (GA) causes a decrease in the levels of mutant p53 by restoring its ubiquitination and degradation by the proteasome[297, 298], and also avoids its translocation to the nucleus of the cell.[299] Although GA does not repair p53 normal functions, a decrease in the protein levels is already very important to prevent gain-of functions.[300]

Reintroduction of wild type p53 into tumour cells suppresses tumour growth and potentiates cytotoxicity of DNA damaging drugs.[301-305] The main difficulty in-vivo is for p53 to get the tumour cells. However, the strategy can be used ex-vivo to eliminate tumorigenic cells in bone marrow before transplantations.

Finally, another possibility is restoring the wild-type function from a mutant p53. A synthetic 22-mer peptide able to bind p53 C-terminus has been reported which can restore in certain cases the protein activity as a tumour suppressor.[306, 307] Some other synthetic molecules were found to bind the DNA binding domain and stabilize it allowing re-establishment of p53 activity in mice.[308] It has been recently demonstrated that the reactivation of p53 function really leads to a tumour regression in vivo by apoptosis, cellular senescence or suppression of cell growth.[309]

V.1.5. p53TD

Most of the mutations studied in p53 are mutations on its DNA-binding domain. However some mutations in the tetramerization domain have been found which also unleash cancer.[310] The present work focuses on one of them and possible ways to offset its effects.

The role of tetramerization domain is essential for p53 to carry out its function in vivo.[253] It is necessary that the protein is a tetramer to bind DNA, to interact with other proteins, to be post-translationally modified and even to be degraded.

The tetramerization domain p53TD is composed of four chains each of one containing a ß-strand (residues 326-333) linked to an α-helix (residues 335-355) by a single residue (Gly-334).[245] p53TD has been reported to be a dimer of dimers[311, 312] (Figure 44). Two monomers, A and D, associate in an antiparallel form by establishing eight backbone hydrogen bonds between their ß-strand regions. A hydrophobic core is composed of Phe328, Leu330, Ile332, Phe338, Phe341 and Asn345. Also a salt bridge is reported to be found between Arg337 of one monomer and Asp352 of the other one in crystal structures. Once a primary dimer is formed, it binds another equal primary dimer B-C to form the tetramer.[313]

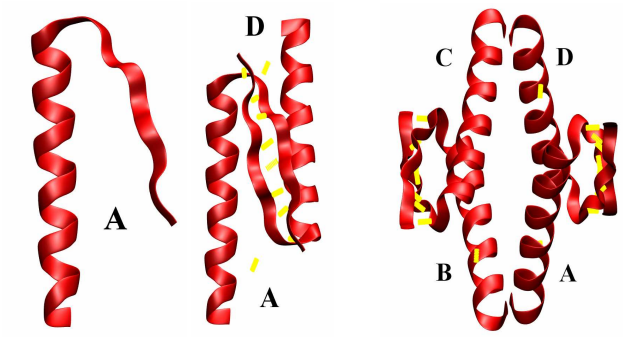


Figure 44. Evolution from a monomer to the formation of p53TD.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

The interaction takes place between the α-helices leading to a structure with D_2 symmetry. The dimer-dimer interface is mainly supported by hydrophobic interactions (Figure 45). Several studies on directed mutagenesis on residues of the hydrophobic core[314] and the observed conservation of them in p53 proteins of different species[315, 316] demonstrate their importance. The tetramer can be reversibly unfolded under physiological conditions.[314] The tetramerization domain has been found to increase the strength of binding to the DNA and to influence the preorganization and the conformation of DNA-p53 complexes.[252] p53TD is also important because there are some proteins that interact directly with it or with other parts of the protein but only when it exists in its tetrameric conformation. This is the case, for instance, of MDM2. Moreover there are several post-transcriptional modifications that depend on the quaternary structure of p53 such as phosphorylation by some kinases[317], very important in the regulation of its function. p53TD also contains a nuclear export signal[318] from residue 340 to residue 351 indispensable for the protein to go through the nucleus membrane.

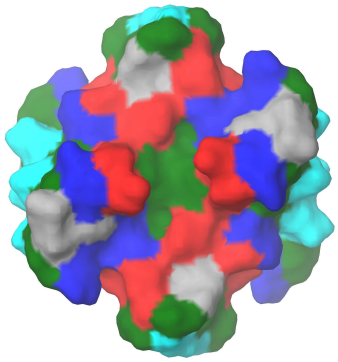


Figure 45. Surface of p53TD with residues coloured by properties. A hydrophobic pocket (in green) in the middle of the four chains supports the structure.

Several mutations in p53TD have shown to cause an increment of the dissociation constant so that, compared to the wild-type protein, a greater amount of monomers is needed to get the same levels of tetramer. Thus, its activity depends on the levels of expression. Some other mutations definitely inhibit the formation of the tetramer so that the phenotypic expression is often more serious in this case.

V.2. Motivation

Because of the inherent symmetry of p53TD, punctual mutations can significantly reduce its stability. Specifically, mutation of arginine in position 337 to histidine (R337H) was linked to pediatric adrenal cortical carcinoma (ACC).[22] It was reported that in a localized population in southern Brazil, 97% of children that developed tumours in the adrenal gland had the R337H mutation in p53. It was striking that the patients only expressed ACC and no other kinds of cancer. Even more surprisingly two independent reports demonstrated that R337H p53TD function was exactly the same as p53TD function.[319, 320] The mystery remained until DiGiammarino et al. found out in 2002 that the loss of stability in the mutant protein was directly related with a basic pH.[22] That was why the cancer appears in the adrenal glands, where the pH has a value around 8.

Firstly, the researchers checked by gel filtration chromatography and glutaraldehyde cross-linking that both p53TD and R337H p53TD were tetramers. Second, circular dichroism (CD) showed that

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

both of them had similar secondary structure. Also 2D ^{1}H-^{15}N TROSY spectra showed that they were structurally very similar. The experiments were done under standard conditions. However, when they studied both proteins by CD at different temperatures and at different pH values the difference became evident. At all pH values the mutant protein unfolded at lower temperatures than the wild type one. Moreover R337H p53TD became significantly less stable as pH increased from 5 toward the physiological range between 7 and 8. At pH 8 at 37ºC 70% of the mutant protein was unfolded whereas 100% of the wild type protein was folded. Therefore intracellular pH might modulate the function of the mutant protein ranging from a function equal to the wild type protein to a total dysfunction in adrenal tumour cells.

DiGiammarino et al. concluded their work, after some more tests, attributing the pH-dependence to the fact that at basic pH, in position 337 a deprotonated His was present instead of a protonated Arg. The lack of the positive charge together with the fact that His sidechain is shorter than Arg´s were supposed to be weakening the four hydrogen bonding interactions with the carboxylate groups of Asp352, considered very important interactions in the stabilization of the tetramer. Moreover the side chain of His is less hydrophobic than the side chain of Arg, so that in the mutant protein also the hydrophobic interactions might be diminished.

Finally, the explanation for the disease being suffered only by children could be explained by the fact that "the adrenal gland is known to undergo extensive cellular remodelling during pre- and postnatal development[321, 322] and this process requires the functional integrity of an apoptotic response. Under certain circumstances, apoptotic cells are known to have increased intracellular pH (~7.9).[323] In the developing adrenal gland, these factors may collaborate to create a cellular environment characterized by an elevated pH that destabilizes the tetramerization domain of R337H p53TD, leading to loss of tumour suppressor function".

V.3. Goals

The final goal of this study consisted in finding, in collaboration with experimentalists, a synthetic ligand which would bind to the mutant protein R337H p53TD and stabilize it in a conformation as similar as possible to that of the tetrameric structure of the wild type.

The first objective, presented in this chapter, was to establish a computational methodology that would permit to find out the main differences in structure and behaviour between the wild type protein and the mutant one.

Once this achieved, the same methodology was used in Chapter VI to study the effect of some ligands bound on both protein surfaces. First of all, oligoguanidinium ligands were tested because they had been already proved to bind p53TD by NMR techniques.[324] The theoretical study pursued a better understanding of the protein-ligand interaction in the wild type protein, as well as a prediction on the possible stabilization effect of the ligands on the mutant protein. Further on, the same goals were posed with two different calixarene ligands.

V.4. Methods

Molecular Dynamics simulations were carried out on p53TD and R337H p53TD at different conditions. The starting structure of the wild-type protein p53 was taken from the X-Ray structure deposited in the Protein Data Bank (PDB) with ID 1aie[325], which constitutes the tetramerization

domain of p53, p53TD. The initial structure for the R337H p53TD mutant was generated by a single residue replacement in the wild-type structure, thus arginine in position 337 was replaced by histidine. The protonation state of every amino acid residue was the corresponding to a medium of pH 7. The histidines in the mutated protein were deprotonated. Amber94[81] force field parameters were used to describe the bonding, van der Waals and electrostatic interactions. All Molecular Dynamics simulations were carried out using the GROMACS simulation package.[90] Each structure was embedded into a rectangular box of water molecules having a volume of 6.2 nm^3 in the case of p53TD (7206 water molecules) and of the same volume in the case of R337H p53TD (7212 water molecules). This particular size of the box was the used in order to let a distance of 9Å from every wall to the nearest part of the protein-ligand complex. Water molecules were represented applying the SPC[326, 327] model. Na+ counter-ions defined for Amber94 were added to ensure the overall electric neutrality of the system. Periodic boundary conditions were used. The geometry of the hydrated systems was initially optimised using a steepest-descent algorithm, and then the system was subjected to a three-stage equilibration protocol: 100ps constant temperature; 100ps constant pressure, allowing volume changes; 100ps constant temperature. After this, a 10 ns long simulation was submitted, collecting data every 2ps. The same protocol was used to carry out molecular dynamics simulations at 300 and 400K. Long-range electrostatic interactions were taken into account by the Particle-mesh Ewald (PME) method. VdW interactions were truncated at a cut-off distance of 10Å. Newton equations were integrated using a time step of 2fs. The simulations were performed in the NVT ensemble using Berendsen coupling method[95] to keep the temperature constant, with coupling times of 0.1ps and 0.5ps, respectively. LINCS algorithm was chosen to satisfy the constraints to the equilibrium for all bond lengths. GROMACS analysis tools and VMD program[107] were used to analyze the trajectories.

V.5. Results

V.5.1. RMSD of the protein backbone

By analyzing the root mean square deviation of the backbone of the protein it was observed (Figure 46) that p53TD at 300K experimented insignificant changes respect to the crystal structure and remained stable during the 10 nanoseconds (red trace in Figure 46). The small fluctuations in the structure and the huge stability of the protein were in good agreement with experimental evidences.

R337H p53TD at 300K experienced slight structural changes compared to p53TD (green trace in Figure 46). The difference with respect to p53TD profile was only around 0.5-1.0Å. Nevertheless a minor tendency of the RMSD to progressively increase from the beginning to the end of the trajectory has to be underlined. The gradually increment pointed that the structure slowly lost its original conformation but 10 ns was not time enough to appreciate significant differences respect to the wild type protein. For this reason, a high temperature of 400K was used to stress structural changes during the simulation. Otherwise considerable changes would occur on a time scale unreachable with currently available computational resources.

At 400K the tetramerization domain of the wild type protein (blue trace in Figure 46) kept on being very similar to the original structure although it experienced larger fluctuations during the whole trajectory and especially from 3 to 4 and from 8.5 to 9 ns. This means that the structure was to some extent disturbed by the denaturing conditions but not as much as to lose its original conformation. The biggest changes were undone in a leeway of time of 1 ns maximum. Instead, R337H p53TD (violet trace in Figure 46) behaved very differently from p53TD. In this case the RMSD increased substantially at the beginning of the MD and kept rising until the end. These results nicely agreed

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

with experiments of Circular Dichroism and other experimental evidences mentioned above and presented by Enrico L. DiGiammarino et al.[22] Indeed the dynamics corroborated that the mutant tetramerization domain adopts a native-like fold but is less stable than the wild type domain in certain conditions.

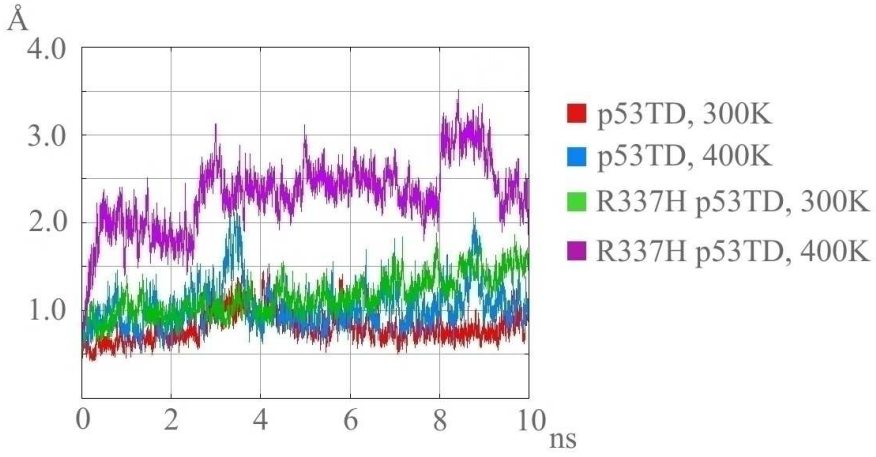


Figure 46. RMSD of the backbone of the protein versus time, taking X-ray structure as reference. Comparison of p53TD and R337H p53TD obtained from the MD simulations at 300 and 400K. Time represented in nanoseconds and distance in angstroms.

From a visual analysis of the trajectories it was obvious that there were important disruptions involving the secondary and the tertiary structures of the mutant protein at 400K. For their better identification and for the understanding on the evolution of the disruption process, further analysis was needed.

V.5.2. Time-evolution of the secondary structure

The secondary structure of each residue in the protein was analyzed at every step for each trajectory with the Timeline plugin in VMD program[107], the results plotted in Figure 47, Figure 48, Figure 49 and Figure 50. (For more details on the analysis see Chapter II).

The four chains composing the tetramer showed the same profile along the whole simulation. The sequences from glutamate 326 to glycine 334 adopted a ß-sheet conformation due to the existence of hydrogen bonds between residues from different chains belonging to the same primary dimer. From arginine 335 to glycine 356 α-helix conformations were adopted except at some times for the last two residues. This occurred because the tails of the helices were completely free so the merely fluctuations with the solvent diverted somewhat untwisted them. These was a little artefact created by the simulations as the chains in the real protein are longer so that these two residues may always have an ideal α-helix conformation. The heads of the helices were never perturbed, as can be seen from the plot.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

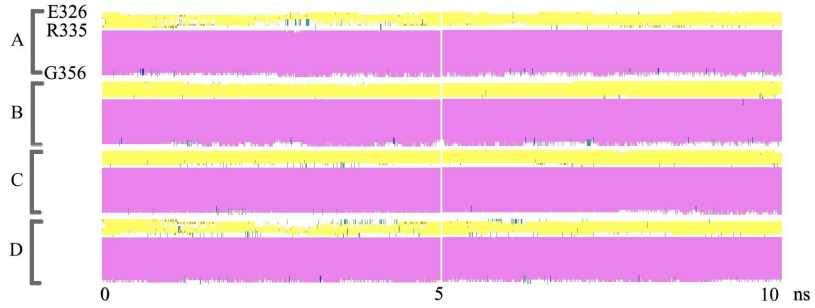


Figure 47. Time-evolution of the secondary structure for p53TD at 300K. All the residues from chains A to D, represented at the y-axis; time at the x-axis.

At 400K the plot (Figure 48) was very similar except for a slight alteration of the ß-sheets between chains B-C. The visual study of several structures from 3 ns on showed that all the hydrogen bonds were always present, but the most ideal conformation of a ß-sheet was not reached most of the time. The heads of the helices were never disrupted except for very punctual moments in chain A and B, as marked in Figure 48.

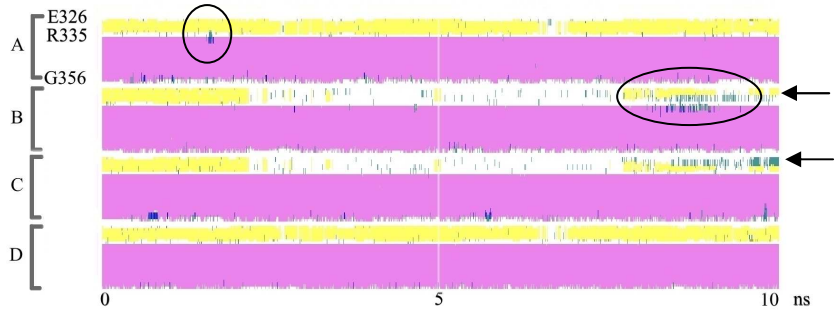


Figure 48. Time-evolution of the secondary structure for p53TD at 400K.

At 300K the mutant protein showed a strong invariable secondary structure along the time with perfect α-helices and ß-sheets (Figure 49), as the wild type protein.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

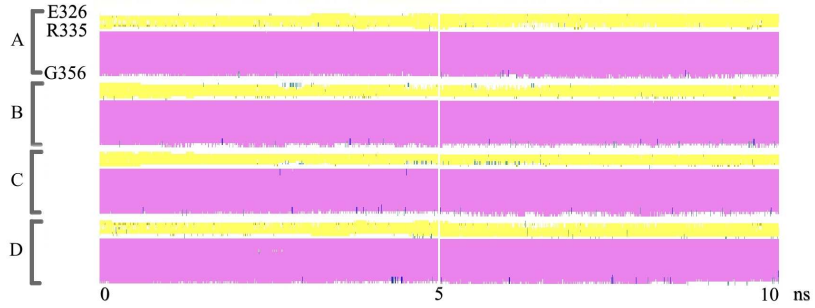


Figure 49. Time-evolution of the secondary structure for R337H p53TD at 300K.

However, at 400K an important disruption of the structure in general showed evident (Figure 50). Not only ß-sheets lost their canonical conformation in both pairs of primary dimers but also the heads of the helices were completely distorted.

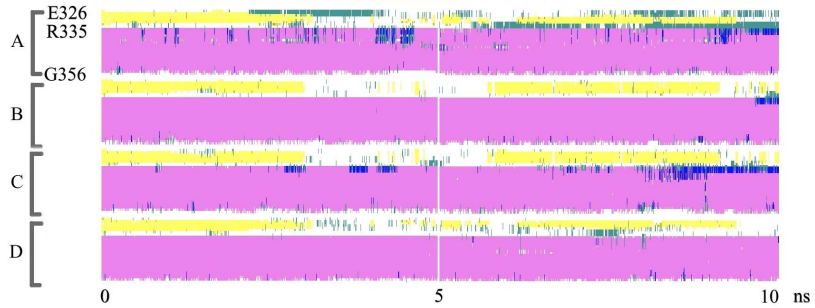


Figure 50. Time-evolution of the secondary structure for R337H p53TD at 400K.

The disruption began always with a great punctual alteration of the residues Met340 and Phe341 that spread up to the first residues composing the helices.
Thus this analysis permitted to zoom in the site where the mutated protein in denaturing conditions experiment the greatest changes.

Picking some structures from the trajectory it was observed that the monomers bended by this moiety so that a first fragment of the helix, from Arg335 to Arg342, was displaced off the axis of the helix, forming and angle different from 180º with the second fragment (see Figure 51). This disruption was the first in coming up and gave rise to subsequent alterations in the secondary structure of the first residues in the α-helices.

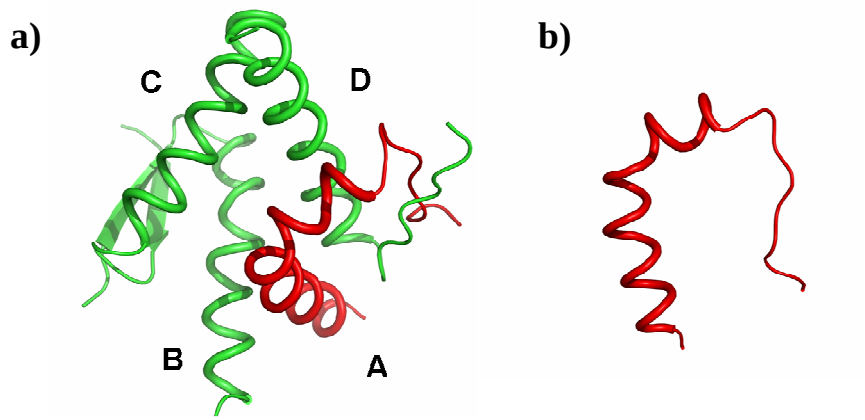


Figure 51. a) Disrupted structure picked from R337H p53TD simulation at 400K. b) The most disturbed monomer presented at the right with three residues drawn as reference.

V.5.3. Hydrogen bond distances

Aimed at understanding the process by which R337H p53TD mutant loses its stability, many interactions were analyzed in detail. Firstly, the distances between pairs of residues forming hydrogen bonds between two chains composing a primary dimer were analyzed. Thus, following a nomenclature like in the Figure 44 and Figure 51 for the chains composing the tetramer, the interactions $Gln331_A$-$Asn345_D$; $Arg/His337_A$-$Asp352_D$; $Arg333_A$-$Asp352_D$; $Arg/His337_A$-$Asn345_D$; $Arg/His337_A$-$Glu349_D$; $Arg333_A$-$Glu349_D$ were taken into account as well as their homologous interactions between chains D-A, B-C and C-B. The most important features are summarized as follows, and are graphically plotted in Figure 52.

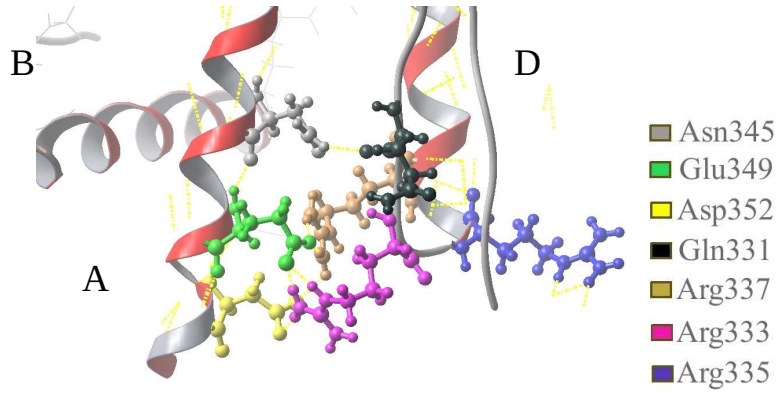


Figure 52. Hydrogen bonding interactions essential in maintaining the tetrameric structure.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Even in p53TD at 300K all these interactions showed very variable along the time. The salt bridge interactions between Arg337 and Asp352, the most evident in the X-ray structure, turn out to be not such static interactions in the dynamics. They always fluctuated between 2 and around 5 Å staying fairly fixed for some periods and being lost during others. The interaction $Arg337_A$-$Asp352_D$ (red trace in Figure 53), for instance, showed an ideal fixed value of 2Å from the beginning to around 3ns trajectory. Afterwards it increased to values fluctuating between 2-3 Å and 5Å.

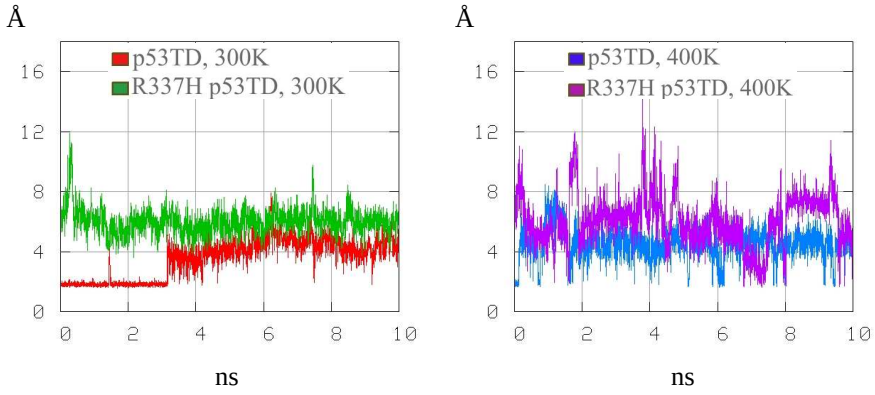


Figure 53. Hydrogen bond distance between Arg/$His337_A$ and $Asp352_D$.

When arginines at position 337 replaced by deprotonated histidines (His337-Asp352) the average distances along the trajectory increased to 4-8 Å and never reached the ideal value of 2 Å. Some interaction distances even reached the 12 Å. Green trace in Figure 53 shows the evolution of the $His337_A$-$Asp352_D$ as an example.

At 400K the interaction distances in p53TD mostly oscillated between 4 and 6 Å, only punctually reaching 2Å (the blue trace in Figure 53 shows the example of the interaction between A and D chains). Thus, this interaction was already strongly weakened in the wild type protein at thermal denaturing conditions. In the mutant protein (violet trace in Figure 53) the distances lengthened to 6-8 Å, sometimes even reaching 12 Å.

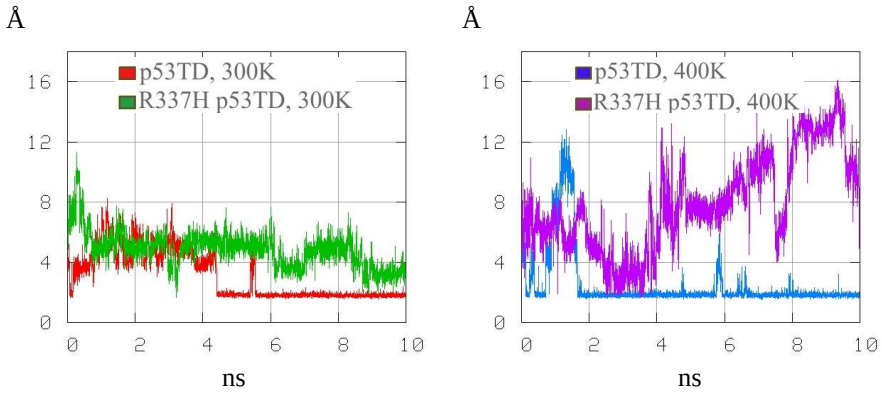


Figure 54. Hydrogen bond distance between Arg/$His337_A$ and $Glu349_D$.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

In p53TD at 300K a swap was evident between 337-352 and 337-349 interactions. Hence, for the periods of time in which a certain 337-352 interaction remained fixed, the interaction 337-349 involving the same monomers was lost, and the other way round. The red traces in Figure 53 and Figure 54 clearly show this effect: the interaction $Arg337_A$-$Asp352_D$ (Figure 53) remained fixed during the 0-3ns interval whereas the interaction $Arg337_A$-$Glu349_D$ (Figure 54) was fixed from 4.5ns on. Concerning the interactions involving other chains the effect was also very evident: $Arg337_D$-$Asp352_A$ was fixed in the interval 6-10ns whereas $Arg337_D$-$Glu349_A$ was fixed during the interval 1-6ns; $Arg337_B$-$Asp352_C$ was ideal from 0 to 6 ns while $Arg337_B$-$Glu349_C$ remained at 2Å from 7 ns on; in the same way $Arg337_C$-$Asp352_B$ was fixed for 0-2ns whereas $Arg337_C$-349_B was fixed from 2ns on. The interactions Arg333-Asp352 in p53TD at 300K were also lost and recovered remaining fairly fixed for long time intervals and being maintained at distances between 4 and 10Å for the rest of the time. $Arg333_A$-$Asp352_D$ (see red trace in Figure 56) remained fixed from 4.5 to 10ns, $Arg333_D$-$Asp352_A$ from 1.5 to 6ns, $Arg333_B$-$Asp352_C$ during the whole dynamics and $Arg333_C$-$Asp352_B$ from 0 to 1.5ns and from 7.5 to 9 ns. For the Arg333-Glu349 interactions (see red trace in Figure 55), the residues were maintained at fairly fixed distances of interaction almost along the whole trajectory but were also lost and recovered several times. All this gave evidence of the interplay between all these possible pairings composing the network. Concerning the two types of hydrogen bonds established with Asn345, the interactions Gln331-Asn345 constantly fluctuated between 2 and 4.5Å. The interactions Arg337-Asn345 fluctuated mostly between 4 and 8Å only in some cases reaching the 2Å.

It is worth to say that, simultaneously to our studies, Lwin[328] et al. also carried out Molecular Dynamics with both the wild type and the mutant tetramerization domain protein. They performed the simulations at several pH and salt conditions although always at the same temperature of 300K. Like us, they simulated neutral to mildly alkaline conditions apart from further simulations in slightly acidic (His337 protonated) and acidic (His337, Glu and Asp protonated) conditions. Their studies completely agreed with ours in pointing out the importance of, specially, the interactions Arg337-Glu349, Arg333-Glu349 and Arg333-Asp352 besides the interaction Arg337-Asp352 which appears evident in the crystal structure of the protein.

In p53TD at 400K all the pairs Arg333-Glu349 and Arg337-Glu349 remained fixed around ideal interaction distances from the beginning to the end of the dynamics. The blue traces in Figure 55 and Figure 54 show the particular cases $Arg333_A$-$Glu349_D$ and $Arg337_A$-$Glu349_D$. The distances between Arg333 and Asp352 experienced larger fluctuations (reaching sometimes values until 8-10Å) but were recovered constantly during the whole trajectory (see the blue trace in Figure 56). Gln331-Asn345 conserved values similar to those at 300K. Arg337-Asn345 were curiously favoured by the thermal denaturing conditions, fluctuating mostly between 2 and 3Å. Indeed Arg337-Asp352, as commented previously, was the most weakened at thermal denaturing conditions. This stated once more the importance of the interactions additional to Arg337-Asp352, observed in the crystal structures, to maintain the tetrameric structure of the protein.

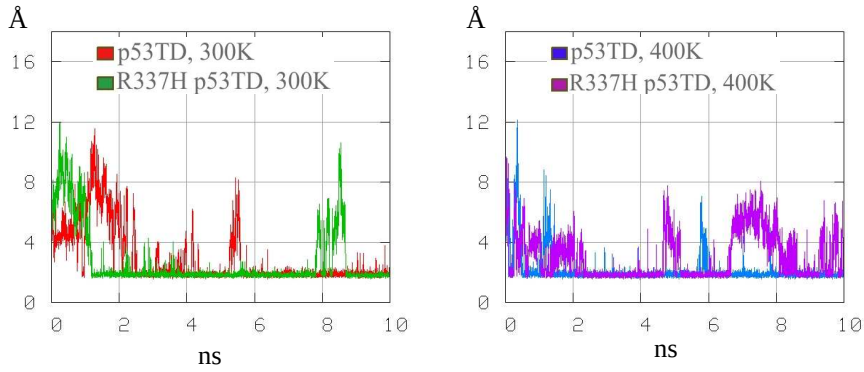


Figure 55. Hydrogen bond distance between $Arg333_A$ and $Glu349_D$.

In R337H p53TD at 300K and at 400K the interactions His337-Asp352 and His337-Glu349 (see green and violet traces in Figure 53 and Figure 54 respectively) were completely lost reaching values between 4 and 12 Å during the whole trajectories. The distances between the pairs Arg333-Asp352 were more favoured, being around the ideal value during some small intervals at 300K (green trace in Figure 56) but during the whole trajectory at 400K (violet trace in Figure 56). The interactions Gln331-Asn345 were disfavoured, experiencing larger fluctuations than in the wild type and reaching ideal values more punctually. The pairs His337-Asn345 never reached values of interaction. The interactions that remained the most fixed for long times were those of the Arg333-Glu349 type (green and violet traces in Figure 55). Thus, Arg333 appeared to play a very important role in the hydrogen bond network existing around the residue 337 and involving monomers from the same primary dimer.

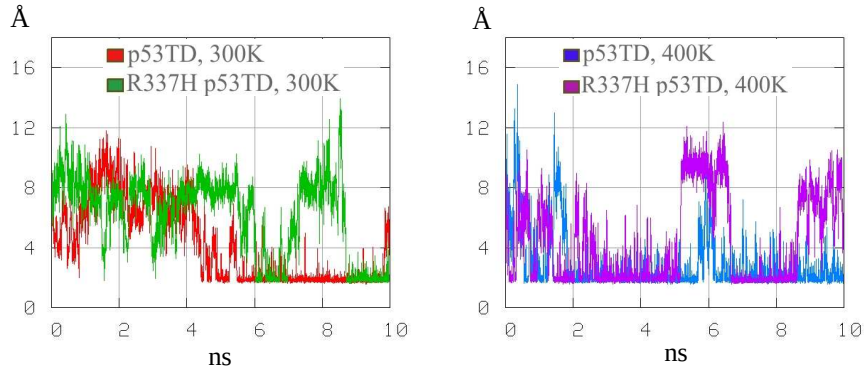


Figure 56. Hydrogen bond distance between $Arg333_A$ and $Asp352_D$.

Lwin[328] reported that two ionic interactions in the dimer-dimer interface were also present in the dynamics of the wild type protein at standard conditions. Based on these results the two interactions were also analyzed in the present work. One of them involved the pair $Arg/His337_A$-$Glu343_C$ (and analogous $Arg/His337_C$-$Glu343_A$, $Arg/His337_B$-$Glu343_D$ and $Arg/His337_D$-$Glu343_B$), which was placed within the region affected by the mutation. The other one was $Lys351_A$-$Glu343_B$ (and analogous $Lys351_B$-$Glu343_A$, $Lys351_C$-$Glu343_D$ and $Lys351_D$-$Glu343_C$), which was far away from the zone destabilized in the mutant protein.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Concerning the Arg/His337-Glu343 interactions, they showed very unfavoured during all the dynamics, including the wild type protein and the mutant protein at 300K and at 400K. The distance values always oscillated between 4 and 10Å except for the case of R337H p53TD at 400K, where the fluctuations were larger, reaching values of even 14Å. Indeed, only very punctual interactions were observed at certain moments in the wild type protein at both standard and thermal denaturing conditions.

In contrast, the interactions Lys351-Glu343 showed much more stable. In most of the cases they fluctuated between 2 and 8 Å being lost and recovered constantly along the whole dynamics. The behaviour was the same for the wild type protein and the mutant protein at both 300K and 400K, with slightly more fluctuations in the latter case.

As summary, in the wild type p53TD at 300K the interactions Arg337-Asp352 showed not as fixed as it could be thought from the observation of the crystal structure. In fact, we observed a whole hydrogen bonding network composed of residues Gln331, Arg333, Arg337, Asn345, Glu349 and Asp352 that constantly formed and broke interacting pairs within all the possibilities. In the wild type protein at 400K all were maintained, being precisely the Arg337-Asp352 the most affected. In the mutant protein at 300K and at 400K those interactions involving the residue 337 were lost and those involving the residue Asn345 were very disturbed. The interactions that best resisted the mutation were those in which Arg333 intervened.

V.5.4. Hydrophobic interaction distances

Hydrophobic interactions between residues were also analyzed. Based on a previous report[314], the distances between the following pairs were studied: $Ile332_A$-$Arg337_A$; $Ile332_A$-$Phe341_A$; $Arg/His337_A$-$Met340_A$; $Arg/His337_A$-$Phe341_A$; $Met340_A$-$Phe341_A$; $Met340_A$-$Met340_C$; $Ile332_A$-$Leu330_D$; $Phe341_A$-$Phe341_D$; $Leu330_D$-$Phe341_D$ (Figure 57). The analogous interactions involving other chains in the tetramer were also analyzed for the wild type protein at 300K and 400K and for the mutant protein at 300K. The disruption originated in the mutant protein at 400K made the protein be asymmetric so that only the interactions previously listed were analyzed, as they were directly related to the most distorted chain A.

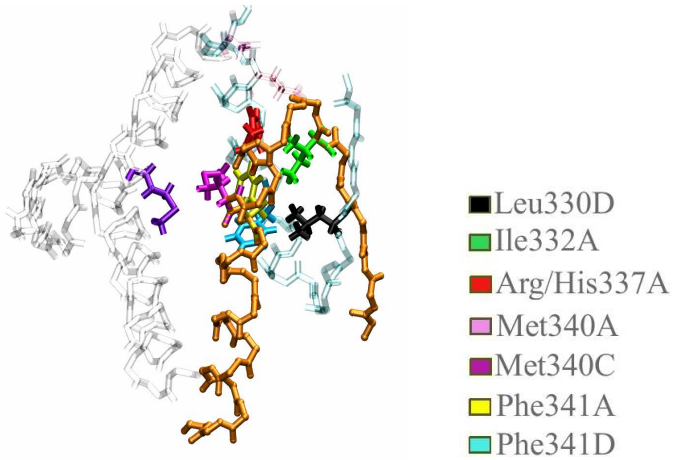


Figure 57. Hydrophobic interactions that maintain the tetrameric structure of p53TD.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

In p53TD at 300K and 400K all the hydrophobic interactions analyzed remained very stable between 2-3 or 2-4 Å. The distances $Leu330_D$-$Phe341_D$ and $Ile332_A$-$Phe341_A$ fluctuated slightly more than the rest.

In R337H p53TD at 300K no important changes were observed with respect to the results in p53TD at the same temperature. However, when working at thermal denaturing conditions most of the interactions showed a different behaviour.

The most weakened interactions from the beginning were $His337_A$-$Phe341_A$, $Ile332_A$-$Phe341_A$ and $Met340_A$-$Met340_C$. The simulation more advanced, at approximately 4ns, $Ile332_A$-$His337_A$ began to fail increasing distances. The interaction $Leu330_D$-$Phe341_D$ was affected in the opposite way, turning into stronger and more stable. The interactions $Phe341_A$-$Met340_A$, $Phe341_D$-$Phe341_A$ and $Leu330_D$-$Ile332_A$ were not affected by the mutation.

Figure 58 shows one of the hydrophobic interactions affected from the beginning in the mutant protein at 400K: $Arg/His337_A$-$Phe341_A$. As commented previously the interaction fluctuated between 2 and 3 Å in the wild type protein and in the mutant protein at 300K (red and green traces in Figure 58). The fluctuations were only slightly larger for the wild type protein at 400K (blue trace in Figure 58). However, the evolution of this interaction was clearly different for the mutant protein (violet trace in Figure 58). The interaction distance reached a value of 5-6 Å from the beginning of the dynamics and was maintained around this value until the end. From the beginning of the trajectory at 400K in the mutant protein the residues went further away from 2 to 4-6 Å never recovering its ideal original distance until the end of the 10ns.

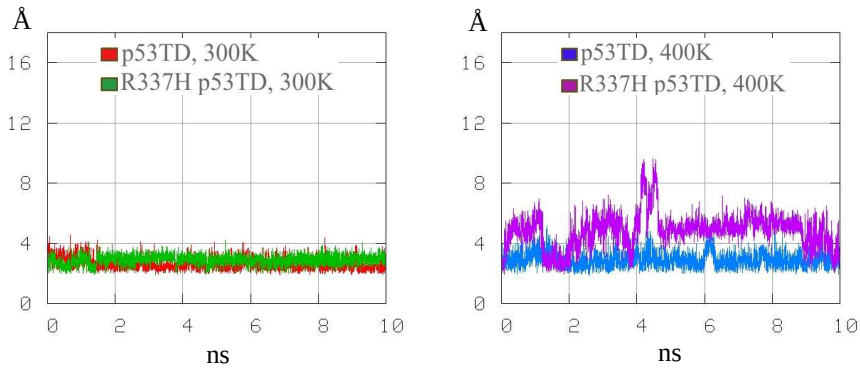


Figure 58. Hydrophobic interaction distance between $Arg/His337_A$ and $Phe341_A$.

Figure 59 shows another of the hydrophobic interaction distances most affected by the mutation, $Ile332_A$-$Phe341_A$. The red and blue traces show the stability of this hydrophobic interaction in the wild type protein. In this case, as exception, the alteration caused by the mutation showed evident even at 300K in chain A (green trace in Figure 59). The residues appeared, from the beginning to the end, approximately 2 Å further apart in comparison with their ideal relative position ($Ile332_B$-$Phe341_B$, $Ile332_C$-$Phe341_C$ and $Ile332_D$-$Phe341_D$ remained at distances of 2-3 Å). At 400K the displacement became much bigger reaching values of 8Å and experiencing much larger fluctuations from the beginning to the end of the trajectory.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

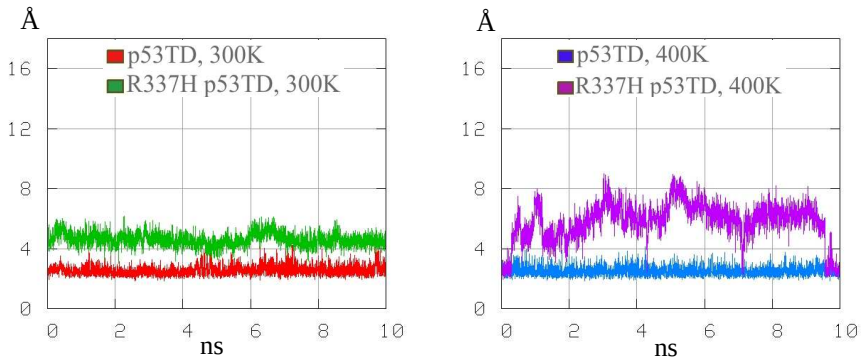


Figure 59. Hydrophobic interaction distance between Ile332$_A$ and Phe341$_A$.

Figure 60 shows the behaviour of the interaction Met340$_A$-Met340$_C$ in the different simulations. In this case the ideal distance was lost and recovered once and again during the whole simulation of the mutant protein at 400K (violet trace). Sometimes it remained lost for 500ps or even 1ns.

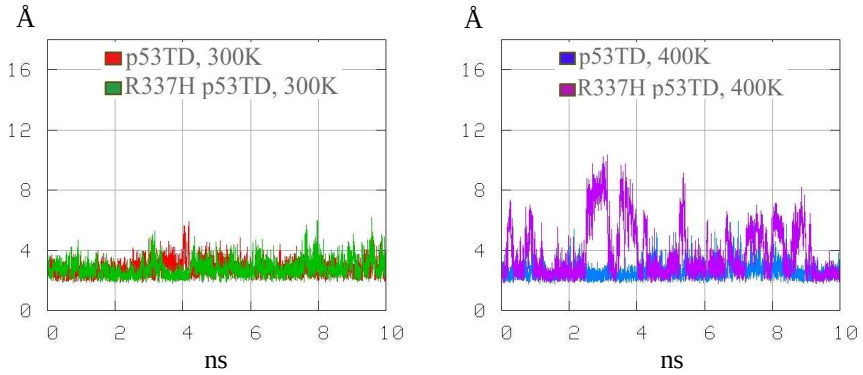


Figure 60. Hydrophobic interaction distance between Met340$_A$ and Met340$_C$.

Figure 61 shows a structure of the R337H p53TD at 400K with the residues involved in the hydrophobic core coloured. The disruption looks evident, existing large distances between them.

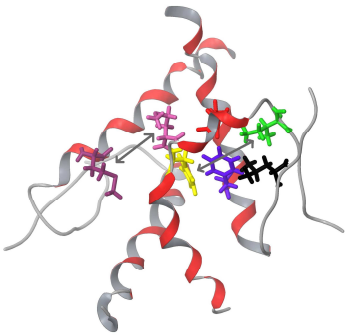


Figure 61. Tetramer in which the increment of the distances among the residues composing the hydrophobic core is evident.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Comparing the graphics from the analysis of the hydrogen bonds and those from the analysis of hydrophobic interactions, a correlation can be found between the largest fluctuations in the different plots. For example, in R337H p53TD at 400K the highly unstable $His337_A$-$Glu349_D$ and $His337_A$-$Asn345_D$ hydrogen bonds were lost definitely one after the other. More or less at the time when the second one misplaced, the hydrophobic interaction $Ile332_A$-$His337_A$ was also lost.

V.5.5. Cluster structures and contact maps

The most representative structures were obtained for all the trajectories by clustering analysis, using an RMSD cut-off 0.75Å. Only one representative structure was found for p53TD at 300K and 400K, and for R337H p53TD mutant at 300K (see Figure 62a). However, for R337H p53TD mutant at 400K, there were 27 significantly different structures (see one example in Figure 62b). Distances between some residues were measured in the averaged structure of the mutant protein at 300K (Figure 62a) and taken as reference: $Asp352_A$-$Asp352_B$ was 8.3; $Asn345_A$-$Asn345_B$ was between 7.4 and 7.8; $Phe341_A$-$Phe341_C$, 9.9; $Phe341_B$-$Phe341_C$, 14.1; $Phe341_C$-$Phe341_D$, 16.8. In general, all these distances as well as the angles between the couples of residues changed in R337H p53TD representative structures (see Figure 62b as an example).

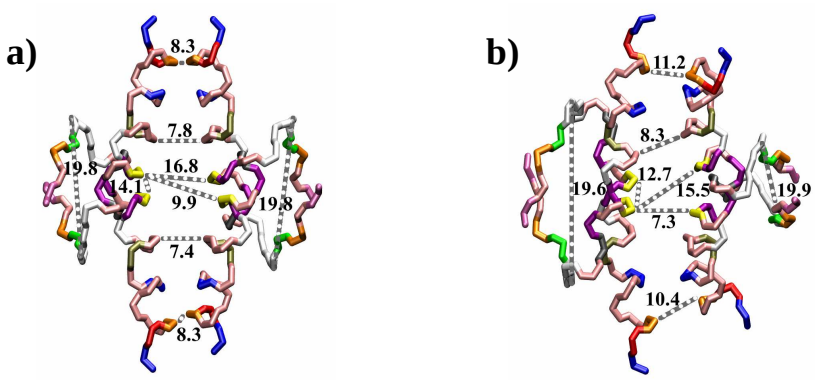


Figure 62. a) Representative structure obtained by clustering from the dynamics of the mutant protein at 300K; b) One of the 27 representative structures obtained at 400K.

This confirmed once again the high instability of the mutant protein in denaturing conditions. Moreover it demonstrated that the suffered changes were not only concerning the secondary structure but also causing distortion of the relative disposition of every chain respect to the rest.
Contact maps (Figure 63) were plotted for the structures showed above pointing that in general the distances between residues in the wild type p53TD protein at 300K remained unaltered at 400K. The map was also very similar for the mutant at 300K. However, R337H P53TD mutant at 400K showed a remarkable decrease of the head-head and centre-centre distances of the helices in relative positions A-C.

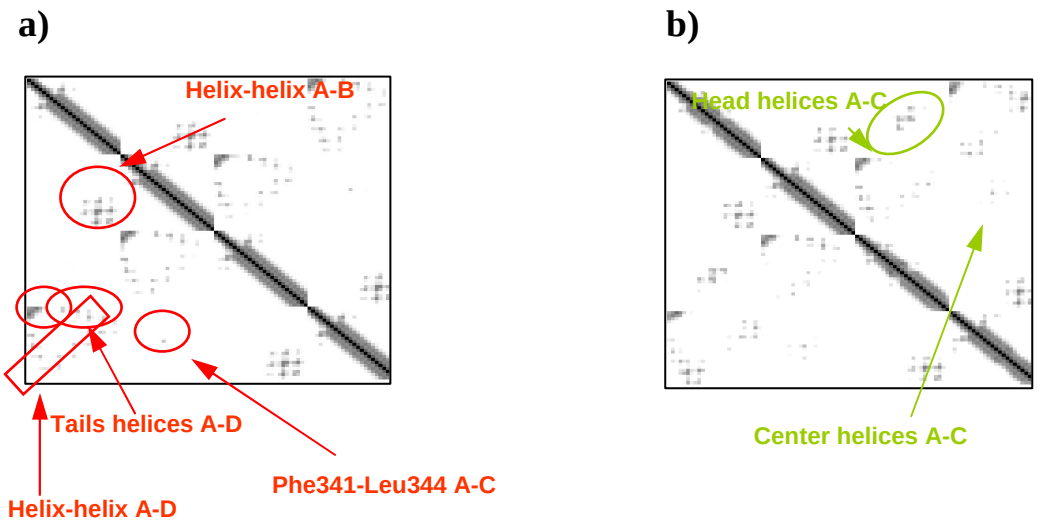
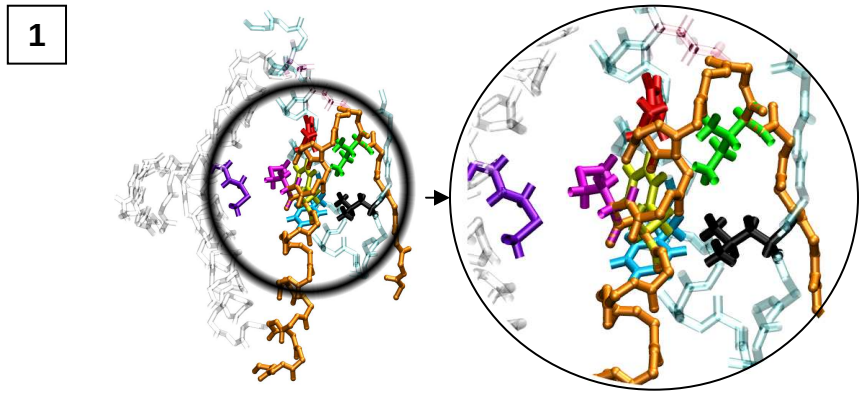


Figure 63. Contact maps for a) R337H p53 TD at 300K and b) 337H p53TD at 400K. A graph square is coloured black at 0 Å and in a linear gray scale from 0 to 10 Å. White squares represent a distance equal to or greater than 10 Å.

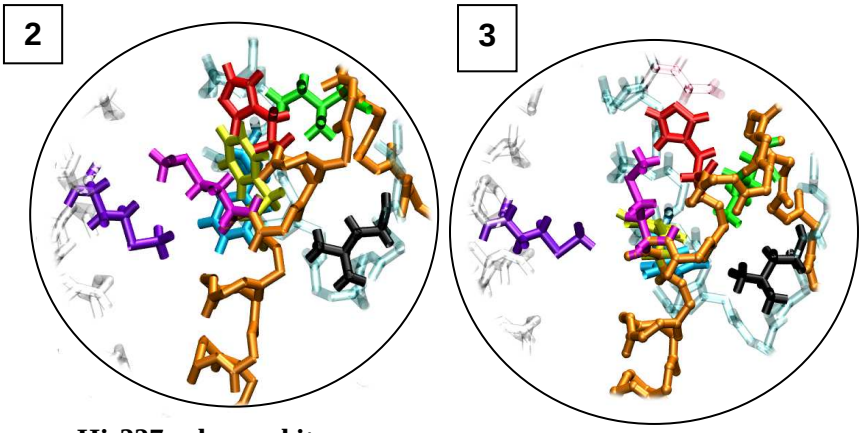
V.5.6. Further analysis

Further analysis of the structural data collected during the simulations, and the study of the correlation between changes in hydrogen bonds and hydrophobic interactions mentioned before, showed that the process of disruption in the mutant was the following: $His337_A$ was not anchored to chain D by hydrogen bonds (as $Arg337_A$ was) so its orientation changed and its side chain pointed out of the nucleus of the protein. The hydrophobic interaction with $Phe341_A$ was then lost and this triggered an alteration in the orientation of $Phe341_A$ causing the lost of its hydrophobic interaction with $Ile332_A$. The residues went 7-8 Å far apart from each other. The most notable change came few picoseconds afterwards, when $Phe341_D$ occupied the space in between $Phe341_A$ and $Ile332_A$ and established new hydrophobic interactions with both of them. From here to the end of the MD simulation the R337H P53TD mutant protein was unable to totally recover its wild secondary and tertiary structures. These consecutive changes altered the binding of the primary dimer (chains A-D) and the structural disruption was afterwards spread to the interactions between pairs of residues involved in keeping the primary dimers together. Some distances became longer and less stable: $Met340_A$- $Met340_C$, $Met340_A$-$Leu348_B$, $Leu348_A$-$Leu348_B$. Some others, more in the centre of the nucleus became shorter, as for instance $Leu344_A$-$Leu344_C$.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

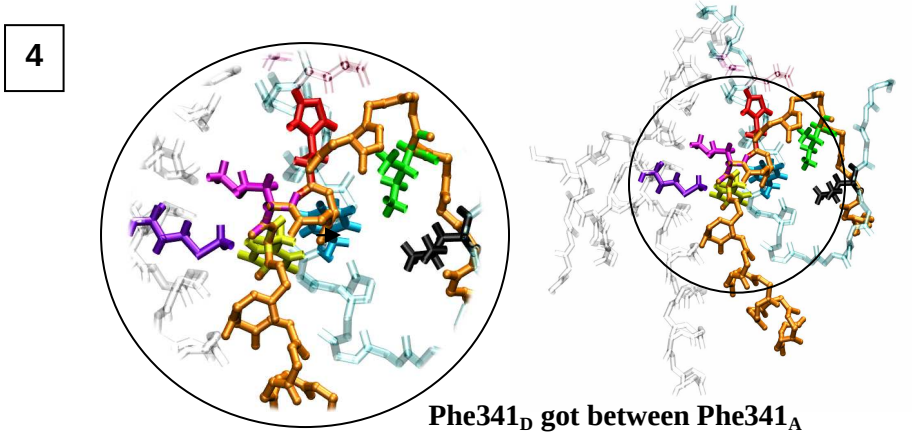


The mutation weakened hydrogen bonding interactions with $His337_A$



$His337_A$ changed its orientation.

$Phe341_A$ changed its orientation.



$Phe341_D$ got between $Phe341_A$ and $Leu332_A$

Figure 64. Consecutive stages towards disruption.

V.6. Conclusions

Wild type p53TD is an extremely stable tetrameric protein. It was not distorted even at denaturing conditions. In the dynamics, its secondary and tertiary structures remained very similar to those in the crystal structure. At 300K the same happened with R337H p53TD proving previous experimental results in which any effect of the mutation was detected at standard conditions.[319, 320] The difference between the wild type and the mutant proteins was only detected by means of simulating under denaturing conditions.

Even in the wild type protein a very variable hydrogen bonding network exists, between chains in relative positions A and D, which sustains the tetrameric structure. The residues composing it (Gln331, Arg333, Arg337, Asn345, Glu349 and Asp352) constantly gain and lose interactions and switch couples or interact with more than one residue at a time. Nevertheless the network is very consistent in whole. The mutation R337H affects all the hydrogen bonds, those involving Arg333 being the strongest.

Hydrophobic interactions are more fixed than independent hydrogen bonding interactions in the wild type protein. The mutation alters them as well as hydrogen bonds, pointing out that not only the primary dimer is affected but the whole tetramer is.

The disruption process starts with the lost of the hydrogen bonds and hydrophobic interactions established with the residue mutated R337H. This fact causes a distortion of the hydrogen bonding network by triggering successive changes in the disposition of the residues involved. The cascade of individual changes ends up in the disruption of the secondary structure of the α-helices because of the bending of the monomers with the subsequent formation of new interactions (between chains composing a primary dimer) that were not present in the original structure. The disruption in the primary dimer spreads to the core of the protein and causes the lost of some of the most important interactions in maintaining the tetramer.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Chapter VI

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

VI. Synthetic ligands on the surface of p53TD

VI.1. Tetraguanidinium ligands on the surface of p53TD

VI.1.1. Introduction

An increasing interest in oligoguanidiniums came up when it was discovered in nature that some peptides able to penetrate the cell membranes were rich in arginines, which have a guanidinium group in their side chain. An exceptional example of this was the nuclear transcription activator protein (Tat) encoded by HIV-I.[329],[330] In 2002 guanidinium-rich oligomers based on the previous observations were reported[331] to test their capability to cross membranes, and so their potential application as drug carriers.

In 1999 Giralt, Mendoza and co-workers[332] reported the properties of a family of tetraguanidinium compounds consisting of a highly preorganized chiral bicyclic guanidinium subunits linked together through short thioeter spacers schematically represented in Figure 65.

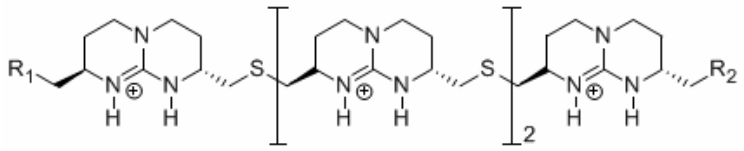


Figure 65. Tetraguanidinium compounds synthesized by Mendoza and co-workers.

The symmetric tetraguanidiniums, where R_1 and R_2 were equal, were expected to bind negatively charged peptides and form stable complexes[332] by means of interactions as shown in Figure 66. Furthermore, these compounds were as amphipatic as proline rich cell penetrating peptides.[333] Thus the internalization and toxicity of these compounds were studied by means of culture of HeLa[334] cells by Confocal Laser Scanning Microscopy , Microplate Fluorimetry Assay and Flow Cytometry.

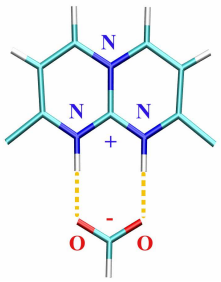


Figure 66. Interaction between a guanidinium and an acid residue through charged hydrogen bonds.

The compounds showed very efficient translocation through HeLa membranes at very low concentrations and, interestingly, accumulated in mitochondria. Mendoza's research group proposed them as a possibility to transport therapeutic agents into cells. In particular they could be very useful

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

to deliver antioxidants in mitochondria for cancer therapies. Other kinds of oligoguanidiniums were synthesized and tested as molecular transporters by other research groups.[335]

VI.1.2. Goals

In 2004, Salvatella et al. reported on the capability of tetraguanidiniums[324] belonging to the family mentioned in the work above to bind p53TD. Each guanidinium, positively charged, was able to form two hydrogen bonds with a negatively charged carboxylate group from an acid residue on the surface of the protein.

The proposal of testing this particular protein-ligand interaction was motivated by the existence of two overlapping helical tetraanionic sequences on each monomer of p53TD that seemed to fit perfectly well with the tetraguanidinium. One of the sequences (called Site1 from now on) is formed by residues Glu336, Glu339, Glu343 and Glu346; the other one (Site2) is formed by Glu343, Glu346, Glu349 and Asp352.
The relative position between a certain residue, *i*, and the next one is *i+3* in all cases except for Glu339-Glu343 in which case it is *i+4*.

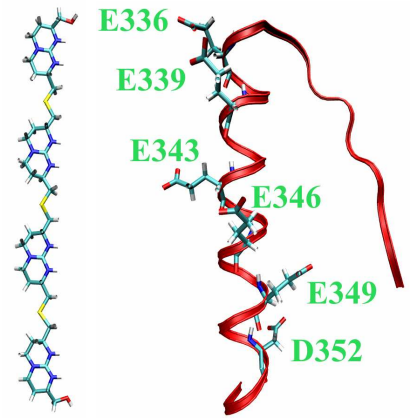


Figure 67. A tetraguanidinium and a monomer of p53

This disposition lets the sidechain of the residues not exactly on the same face of the helix but slightly displaced one another. The spacer provides good adaptation to the distance and angle between the carboxylates (Figure 67).

By comparing Site1 and Site2 in their environment within the whole p53TD structure, it was expected that the ligand would more easily bind to Site1 as Asp352 in Site2 appears to be esterically hindered (Figure 68).

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

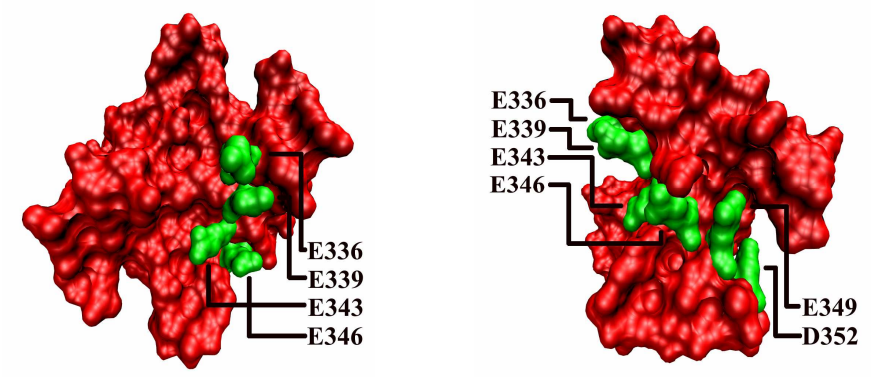


Figure 68. Two perspectives of the p53TD surface. Residues in Site1 and Site2 coloured green.

Mendoza, Giralt and co-workers studied the interaction by NMR chemical shift perturbation (CSP) and saturation transfer difference (STD) spectroscopy. The chemical shift changes were small for all the residues. Those whose chemical shift changed most were Arg337, Met340, Leu344, Ala347 and Leu350 in decreasing order. At first sight these particular residues are closer in space to those composing the Site1 so it was concluded that the anionic patch that interacts noncovalently with the tetraguanidinium ligand is most likely that formed by the four Glu side chains at positions 336, 339, 343 and 346.

In a previous work with tetraanionic peptides the authors had demonstrated that the patch where the tetraguanidinium binds preferably follows the pattern i, i+3, i+6, i+9 regarding the relative positions of the aminoacids. However they presumed that in the case of p53TD there was a change in preference towards Site1 that could be attributed to the fact that the last interactions in Site2 were notably less accessible. Moreover they considered that the existence of an aspartate in position 352 instead of a glutamate, with one more carbon in the sidechain, might cause a lost of plasticity in Site2 in terms of its secondary structure.

The goals posed from a theoretical perspective were to check whether p53TD forms a stable structure with four tetraguanidiniums bounded on its surface, to find other plausible structures for the system, and to see if Site1 is really more favourable than Site2 or others.

VI.1.3. Methods

Molecular Dynamics simulations were carried out with both p53TD and R337H p53TD. The initial structures had one protein with four tetraguanidinium ligands, bounded each of them to one monomer chain. The resulting trajectories were compared with those previously obtained in absence of ligands (see Chapter V).

The structures used to simulate the proteins were the same as in the previous Chapter V, as well as the simulation package used, the force field, the parameters for the water molecules and counter-ions, the conditions of simulation and the preceding procedure of minimization/equilibration.

In both the wild type protein and the mutated protein, a tetraguanidinium molecule was docked by hand to a monomer and replicated applying the D2 symmetry to obtain the entire starting architecture. Each guanidinium group faced its hydrogen atoms towards the oxygens in the

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

carboxylate group of a glutamate or an aspartate, the angle N-H···O being approximately 180º. The database files in GROMACS[90] package were modified to include suitable parameters for the tetraguanidinium. For the bonding and van der Waals interactions, Amber94[81] parameters were applied. The atomic charges were calculated by means of the Restrained Electrostatic Potential fit[82] (RESP) from a quantum mechanics optimization with Gaussian03[140] at the HF/6-31G(d) level.

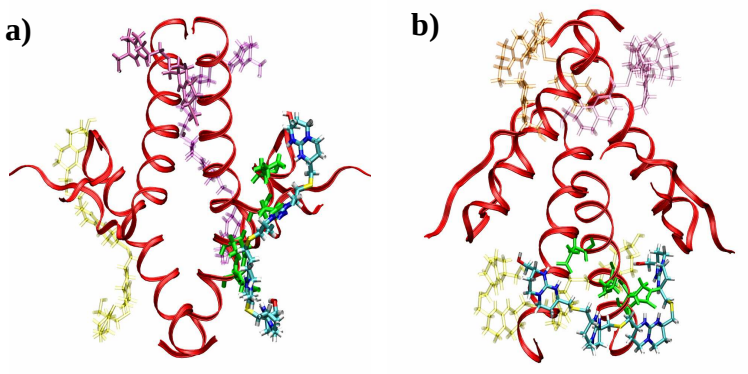


Figure 69. p53TD-4tetraguanidiniums a) in Site1 and b) in Site 2

Each protein-ligand structure was embedded into a rectangular box of around 7600 water molecules. The structures involving p53TD had a volume such as to let a distance of 9Å from every wall to the nearest part of the complex. Simulations were carried out again at 300 and 400K.

In a first step, several analyses were performed in order to establish which site was the most favourable. All of them were done over the trajectories with the wild type protein in standard conditions.

In a second step, further analysis was carried out over the trajectories of the mutant protein as well as the wild type protein, at both temperatures, which tetraguanidinium bound to Site2.

VI.1.4. Results on the comparison of Site1 and Site2

To have a general view about the effect of the ligands in each of the two sites, the evolution of the RMSD of the backbone of the protein was analyzed. The graphs showed that when the tetraguanidiniums were bound to the Site1 on the surface of p53TD, at 300K the protein was slightly disturbed (see green trace in Figure 70). In absence of ligands the average deviation from the X-ray structure was less than 1 Å along the whole trajectory (red trace in Figure 70). The presence of the ligands in Site1 caused a change of 1 Å with respect to the structure in absence of ligands. A clustering analysis of the trajectory with an RMSD cut-off of 0.75Å showed only one cluster. This meant that the change caused by the ligands was not significant. However, a difference between the two sites was noted.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

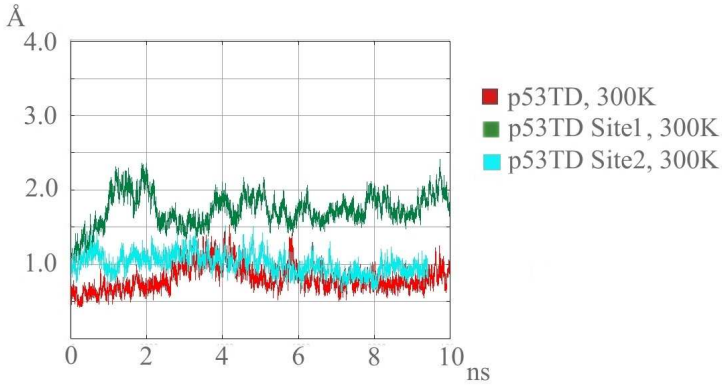


Figure 70. RMSD of the protein backbone versus time, in absence of ligands and in presence of ligands at the two sites under study.

When the tetraguanidinium were bound to the wild-type p53TD at Site2 (blue trace in Figure 70), the trajectories showed that the structure of the protein, highly stable by itself, remained perfectly undisturbed by the presence of the ligands. It could be concluded from this that the tetraguanidinium perturbed to some extent more the protein structure when it was in Site1 than when it was in Site2.

In the next step of the comparative study, attention was pointed to the guanidinium-carboxylate interactions independently. From the visual scrutiny of the trajectories it was known that, in both sites, the sixteen of the interactions established between ligand and protein were lost and recovered once and again along the simulations. When the guanidinium groups definitely lost every contact with the carboxylates, they pulled along neighbour guanidinium groups far away from the anchorage points. In this way, some of the ligands were lost completely.

Regarding this aspect both sites behaved in a similar way, progressively losing interactions. At 5ns, less than half of the original interactions remained conserved in both Site1 and Site2 and some punctual new interactions were established. In both sites at 5ns a tetraguanidinium was found bound to two different chains by means of 2-3 guanidinium-carboxylate interactions. At the end of the 10ns two ligands remained bound to the protein in Site1 and three remained in Site2. However, most of them only conserved one ideal interaction.

The particular interaction with Asp352 present in Site2 behaved as the others. It was lost at some times, the guanidinium group going far away from the asparte and recovered again especially at the end of the simulation. From the visualization of this particular region in the trajectory, it could be seen that the guanidinium went completely out the zone esterically hindered and went in again after some picoseconds. This pointed out that although the anchorage point was in that case slightly more hidden than the others, there was enough space for the guanidinium group to access it.

Losing and recovering guanidinium-carboxylate interactions in both Site1 and Site2 were quantified along the 10ns simulation. The interactions considered for each chain in Site1 were: Glu336 with the first guanidinium group; Glu339 with the second guanidinium group; Glu343 with the third; Glu346 with the fourth. In an analogous way the interactions considered in Site2 were: Glu343 with the first guanidinium group, Glu346 with the second guanidinium group, Glu349 with the third, and Asp352 with the fourth.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

For each interaction, two sets of atoms were defined: one composed by the two oxygen atoms in the carboxylate group of the glutamate or aspartate, and the other composed by the two hydrogen atoms of the N-H groups in the guanidinium. The minimum distance between the two sets was measured along the time and the resulting values were averaged. Figure 71 shows the results in bars, coloured dark red for the interactions in Site1 and coloured yellow for the interactions in Site2. In each case, 4 bars correspond to every residue involved in the binding, each of them representing the average value in a different chain of the protein. Figure 72 represents the corresponding standard deviation values.

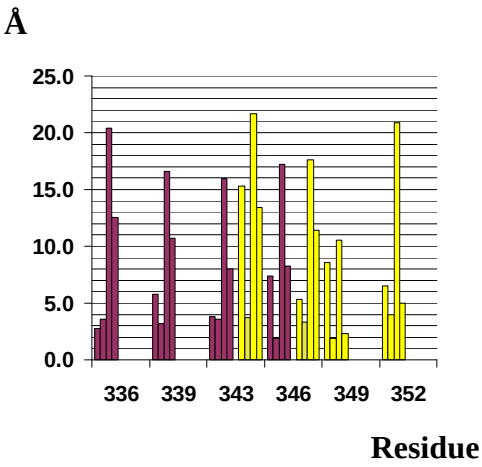


Figure 71. Interaction distances averaged over time for each residue. The chains A, B, C and D composing the tetrameric protein represented in four consecutive bars in each case. Site1 represented in dark red and Site2, in yellow.

The highest values in Figure 71 clearly indicate which interactions were lost definitely already in early stages of the trajectories. Their corresponding standard deviation values are the highest as well. This indicates that the distances varied from the beginning to the end of the trajectories, going from 2Å at the beginning, to values in which the recovery was impossible further on. For instance, if attention is paid to the third interactions in each of the residues composing Site1 (third red columns), it can be observed that all of them have high average distance values and high standard deviation values. This indicates that at some moment in the trajectory all of them were lost with the consequent release of a tetraguanidinium ligand.

A value of 4Å in Figure 71 can be taken as reference to identify those guanidinium groups which remained near their initial anchorage points losing and recovering interactions. It can be observed that only 6 interaction distances in Site1 and 5 in Site2 remained bellow this value. The values remaining around 2Å with corresponding low standard deviation values indicated the guanidinium-carboxylate interactions were kept during the whole trajectory. This was the case for one interaction only in Site1 and two interactions in Site2.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

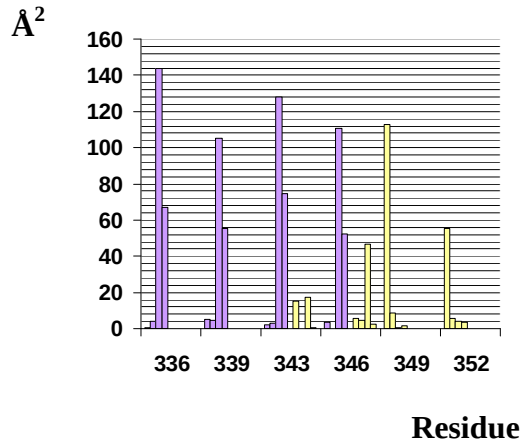


Figure 72. Values of standard deviation corresponding to the average values in Figure 71. Site1 represented in blue and Site2, in light yellow.

Thus, the analysis carried out confirmed that there were not significant differences between the two sites when considering the geometries of independent guanidinium-carboxylate interactions. Moreover there was no difference deserving to be emphasized in the behaviour of Asp352 interactions. Their evolution along the time was very similar to the others. To assess the strength of the protein-ligand interaction, X-CSCORE[104](see Chapter II for details) was evaluated for some structures along the dynamics using the program X-Score. The binding score was obtained for every specific tetraguanidinium with the whole tetrameric protein, removing the rest of the three ligands. Thus every 50 picoseconds 4 values of X-CSCORE were obtained and averaged for both sites. The result was that the binding in Site2 (with a value of 4.2) showed slightly better score than the binding in Site1 (with a value of 3.8).

Four particular cases of one tetramer-one ligand structures in each simulation were visualized trying to identify a correlation between their X-CSCORE values and the number of guanidinium-carboxylate interactions present in the complex. Table 12 shows these data. The protein-ligand interactions are identified by the number of the residue involved.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Table 12. X-CSCORE values and guanidinium-carboxylate interactions present in 8 structures (from struct1 to struct8), composed of one tetramer and one ligand.

	X-CSCORE	Interactions present
Site1		
struct1	4.66	336, 343
struct2	4.17	336, 339, 343, 346
struct3	4.59	336, 339, 343, 346
struct4	4.18	336
Site2		
struct5	4.80	343, 346, 349, 352
struct6	4.24	346, 349, 352
struct7	4.49	343, 346, 352
struct8	4.1	346, 352

Some tetraguanidiniums to which a similar score was attributed had quite different number of anchorage points. The other way around existed too. There were some structures with the same number of interactions but different X-CSCORE values. This showed that the strength of the interaction was not only dependent on the number of guanidinium-carboxylate interactions but also on the disposition of the rest of the ligand respect to the protein.

To compare the results of the chemical shifts previously obtained by Salvatella et al[324] with our calculations, values of RMSD per residue were obtained. The measure was also carried out for the simulation of the protein alone and for both Site1 and Site2 simulations.

For each trajectory, the RMSD per residue were averaged for homolog residues in the different monomers of a same protein. Afterwards, the averaged values obtained for each residue in the absence of ligands were subtracted from the averaged values obtained for the homolog residues in the Site1 and the Site2 simulations.
For instance, the RMSD for the residue Arg335 in the Site1 was calculated as shown in Eq. 1. The RMSD subscript indicated the number of the residue and the chain of the protein (A, B, C or D). The superscript indicated the simulation analized.

$$\left(RMSD^{p53}_{335\text{-}A}+RMSD^{p53}_{335\text{-}B}+RMSD^{p53}_{335\text{-}C}+RMSD^{p53}_{335\text{-}D}\right)/4=RMSD^{p53}_{335\text{-}ABCD}$$

$$\left(RMSD^{Site1}_{335\text{-}A}+RMSD^{Site1}_{335\text{-}B}+RMSD^{Site1}_{335\text{-}C}+RMSD^{Site1}_{335\text{-}D}\right)/4=RMSD^{Site1}_{335\text{-}ABCD}$$

$$RMSD^{Site1}_{335}=RMSD^{Site1}_{335\text{-}ABCD}-RMSD^{p53}_{335\text{-}ABCD}$$

Eq. 19

The results are shown in Table 13. A mean value is also shown for both Site1 and Site2. The residues whose RMSD was larger than the corresponding mean value are underlined in colours.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Table 13. RMSD per residue in presence of ligands minus RMSD per residue in absence of ligands for both Site1 (left column) and Site2 (right column).

	1	2
Arg335	0.1	-0.9
Glu336	0.7	0.0
Arg337	2.2	1.2
Phe338	1.5	0.8
Glu339	0.6	-0.2
Met340	1.0	0.3
Phe341	0.9	0.2
Arg342	-0.1	-1.0
Glu343	1.5	0.5
Leu344	1.1	0.4
Asn345	0.6	0.0
Glu346	1.0	0.2
Ala347	0.9	0.4
Leu348	1.6	0.6
Glu349	1.9	0.1
Leu350	1.1	0.5
Lys351	0.7	0.0
Asp352	0.7	0.0
Ala353	0.9	0.3
	Mean	
	1.0	**0.2**

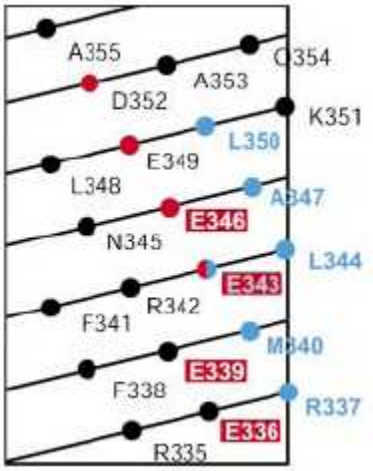


Figure 73. Projection of the surface defined by the helix of one of the monomers in p53TD. Residues that form part of Site1 are labelled with red boxes. Those that form part of Site2 are shown as red spheres. Those which experienced the greatest changes in presence of tetraguanidinium are shown in blue. Reproduced from ref [324].

Figure 73 shows the location of the Site1 and the Site2 in one p53TD chain, together with the position of the residues observed to change the most in presence of tetraguanidinium by NMR techniques. From this representation it was possible to observe that the residues which underwent the most important changes, arginine 337 and methionine 340, agreed with those mentioned in the previous study.[324] However, strikingly, both of the residues changed notably not only in the Site1 simulation but also in the Site2.

The consequent conclusion was that the fact that Arg337 and Met340 experimented great changes did not necessarily mean that the tetraguanidinium was on Site1. The ligand on Site2, which was supposed to be far away from them, in some way was also causing changes on them. By observing in detail the simulations, an explanation could be given to this fact. The ligand bound to one chain A of the tetramer was far away from the Arg337 forming part of this particular chain. However, the ligand bound to chain D was certainly close to this Arg337 in chain A (see Figure 74).

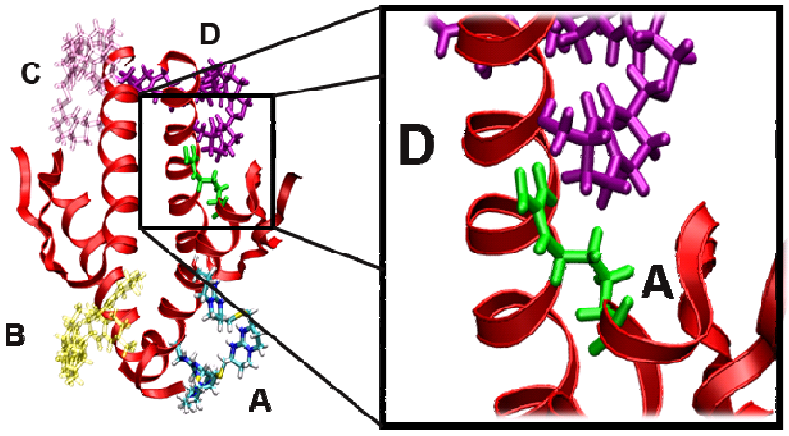


Figure 74. Ligand (in violet) bound to chain D of the tetramer is close to Arg337 (in green) in chain A.

For the case of Met340, the ligand bound to its same chain was not as far away as it was expected. The fact that the secondary structure of the chain was a helix let Met340 close to Glu343, where the ligand was anchored (see Figure 75).

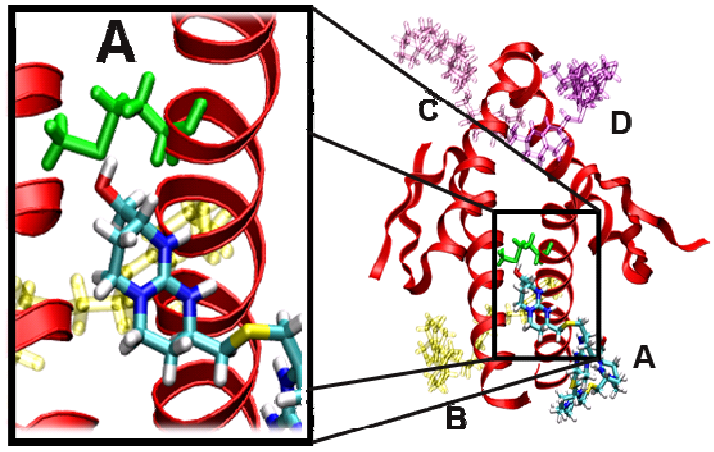


Figure 75. Ligand bound to chain A of the tetramer is close to Met340 in chain A (in green).

VI.1.5. General conclusions on Site1 versus Site2

The simulations confirmed that the tetraguanidinium was able to bind p53TD, and remained for some time on its surface. As demonstrated, the experimental observation of a change in Arg337 and Met340 when tetraguanidinium was present was not a sufficient argument to assure that it bounded Site1, as the change was also present in Site2. Moreover, it was proved that Asp352 is not as

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

esterically hindered as to avoid the interaction with a guanidinium group. Thus both sites could be considered as competitors. Furthermore, taking into account the existence of multiple carboxylate groups in the protein it could be considered a great number of dispositions for the ligands as interactions were lost and recovered continuously along the trajectory the ligand adopting several conformations and even some times occupying sites not only implying one single monomer. All of these possibilities could be as favourable as the others so that it that a mixture of all of them could be expected.

VI.1.6. Further results on Site2

Molecular Dynamics taking the proteins with the tetraguanidiniums on the Site2 as starting structures were carried out at 400K. At these thermal denaturing conditions all the ligands had lost their original guanidinium-carboxylate interactions at 2-3 Å. From then on, multiple structures with the tetraguanidiniums bound to different places were originated along the trajectory, showing very dynamic behaviour of the ligands.

The analysis performed previously for p53TD and R337H p53TD at 300 and 400K were repeated in order to compare the new Site2 trajectories with those where the ligand was not present: RMSD of the backbone of the protein, time-evolution of the secondary structure, essential dynamics, distances of interaction between relevant residues forming hydrogen bonds or hydrophobic interactions, clusters, contact maps.

RMSD of the protein backbone

The RMSD trace almost superposed to the RMSD of p53TD in absence of tetraguanidiniums, at standard and thermal denaturing conditions (Figure 76). This meant that the structure of the protein did not experienced important changes because of the ligands.

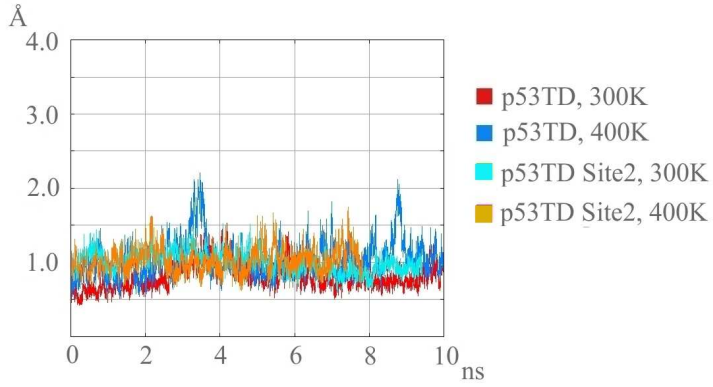


Figure 76. Comparison of the RMSD versus time for the wild type protein in presence and in absence of the ligands in Site2.

Moreover, it was observed that the fluctuations were larger when the oligoguanidiniums were not present, as can be observed comparing the dark blue and orange traces in Figure 76. This meant that

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

the ligand did not disturb the wild type protein at room temperature and even stabilized it to some extent at 400K.

The same analysis carried out on the mutant protein showed that the evolution of the protein structure was the same in presence and in absence of ligands at 300K (green and brown traces in Figure 77). Furthermore, the tetrameric architecture was slightly stabilized by the tetraguanidiniums at 400K (violet and pink traces in Figure 77). The RMSD decreased from a value of 2.5-3 angstroms to a value of 1.5-2. However the trace of R337H p53TD with ligands was very irregular showing that the structure had still not reached stability.

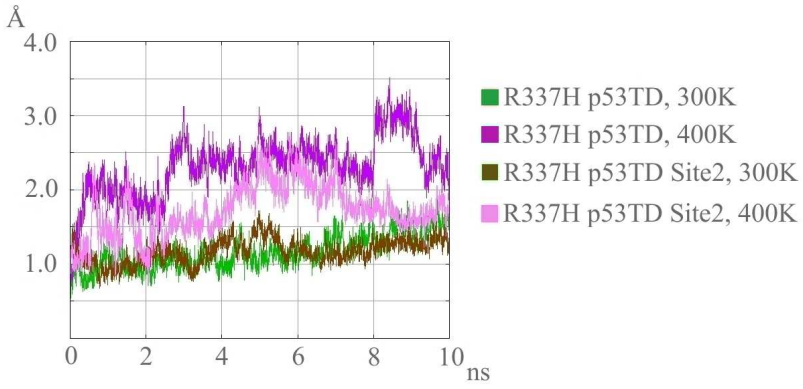


Figure 77. Comparison of the RMSD versus time for the mutant protein in presence and in absence of the ligands.

Time-evolution of the secondary structure

The plot of the time evolution of the secondary structure in p53TD with tetraguanidiniums at room temperature confirmed again that the ligands did not disturb the wild type protein nor at 300 neither at 400K.

Concerning R337H p53TD, at 300K the secondary structure remained intact more or less until 4ns time. After that a deformation at the first residues in the α-helix of chain D was appreciable. When the deformation appeared the ligands had not completely gone but they had changed their disposition forming hydrogen bonds at different sites. This fact disturbed the secondary structure of the two first residues in the helix but had nothing to do with the distortion experienced in the mutant protein in absence of tetraguanidiniums.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

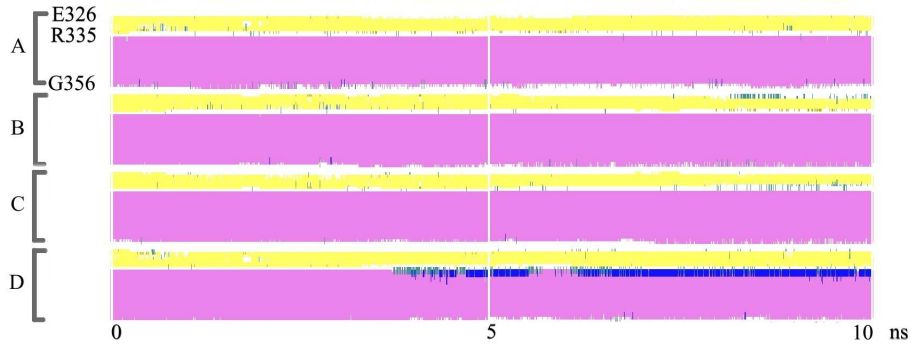


Figure 78. Time-evolution of the secondary structure for R337H p53TD with tetraguanidiniums at 300K.

At 400K, there were little distortions at the end of the helices, especially in chains B and C. At the beginning of the helices there were only punctual distortions always recovered. Thus, the distortion was different from the suffered in the mutant protein.

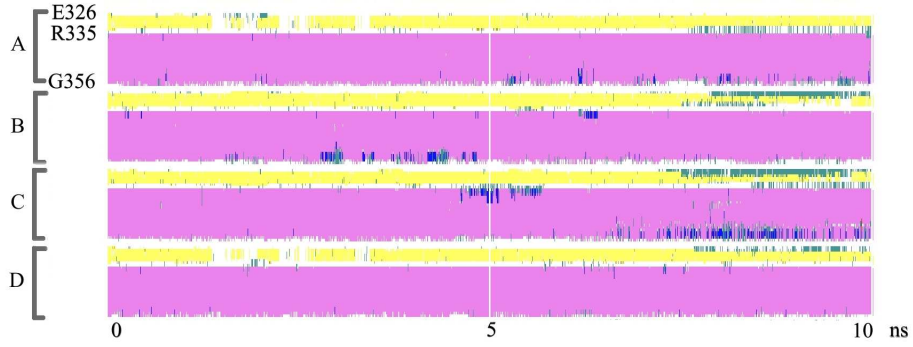


Figure 79. Time-evolution of the secondary structure for R337H p53TD with tetraguanidiniums at 400K.

The general conclusion was that tetraguanidiniums, during the time they were bound to the mutant protein, avoided the deformation of the α-helix from Arg335 to Arg342 experienced when the ligands were not present. However, from the previous graphs representing the RMSD of the backbone of the protein we saw that the mutant protein with the ligands had not a RMSD as high as in their absence but it had not an RMSD as low as the wild type protein. So that it could be expected that there was a change in the tertiary structure. In other words the distribution of the four chains in the space might have changed.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Hydrogen bond distances

The hydrogen bond distances crucial in maintaining the tetrameric structure were analyzed in R337H p53TD with tetraguanidiniums at 400K aimed at comparing them with those existing in p53TD and R337H p53TD at 400K. I would like to remind the reader the analysis in the previous chapter, on interactions $Gln331_A$-$Asn345_D$; $Arg/His337_A$-$Asp352_D$; $Arg333_A$-$Asp352_D$; $Arg/His337_A$-$Asn345_D$; $Arg/His337_A$-$Glu349_D$; $Arg333_A$-$Glu349_D$ and analogous. The same analysis was carried out with the present simulations.

It could be seen that the only interaction favoured by the presence of tetraguanidinium was the one between Gln331 and Asn345 in chains A-D. Figure 80 shows the profile of the interaction distance for the wild type protein in blue, fluctuating always between 2 and 4 Å. The trace in violet corresponds to that of the mutant protein and shows values between 4 and 8 Å during the most of the simulation. The pink trace, which corresponds to that of the mutant protein in presence of ligand, superposes to the blue trace with even lower fluctuations compared with it. The other hydrogen bonds behaved the same as when the ligands were not present.

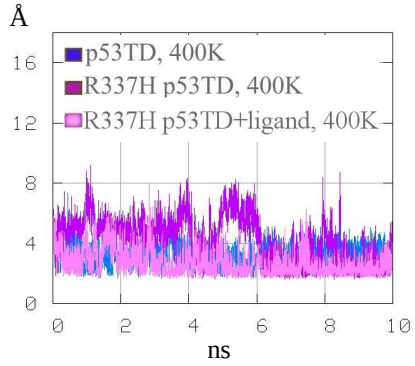


Figure 80. Comparison of hydrogen bonding distances in p53TD, R337H p53TD and R337H p53TD-ligands at 400K. $Gln331_A$-$Asn345_D$ interaction.

Hydrophobic interaction distances

The same hydrophobic interactions analyzed when comparing p53TD and R337H p53TD at 400K, were analyzed in the case of R337H p53TD at 400K with tetraguanidiniums bound: $Ile332_A$-$Arg337_A$; $Ile332_A$-$Phe341_A$; $Arg/His337_A$-$Met340_A$; $Arg/His337_A$-$Phe341_A$; $Met340_A$-$Phe341_A$; $Met340_A$-$Met340_C$; $Ile332_A$-$Leu330_D$; $Phe341_A$-$Phe341_D$; $Leu330_D$-$Phe341_D$. Analogous interactions between residues of the other chains were also analyzed.

Most of the hydrophobic distances were found to be favoured by the presence of the ligands. Figure 81 comparing $Ile332_A$-$Phe341_A$ interaction is shown as an example. In the graph, the pink trace which represents the distances in the mutant protein in presence of ligand was almost superposed to the blue trace of the wild type protein.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

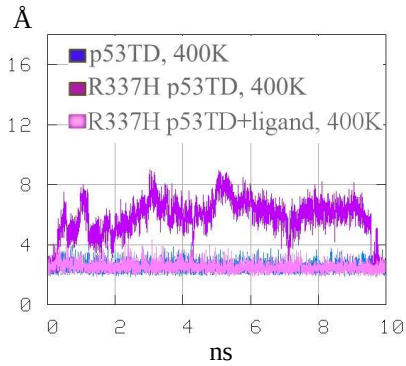


Figure 81. Comparison of the hydrophobic interaction distance 332A-341A in p53TD, R337H p53TD and R337H p53TD-ligands at 400K.

There were only two of the hydrophobic interactions analyzed that were not favoured by the tetraguanidinium: $His337_B$-$Phe341_B$ and $His337_C$-$Phe341_C$. However, this was not sufficient to preserve the original structure of the protein.

Cluster structures and Contact Maps

When the structures obtained in the trajectory of the mutant protein with tetraguanidinium bound to Site2 at 400K were grouped following the same method as previously done with p53TD and R337H p53TD, 12 representative structures were found. It must be borne in mind that for R337H p53TD structure at 400K 27 different structures had been obtained. Thus, the tetraguanidinium ligands caused a slight stabilization while they were present.

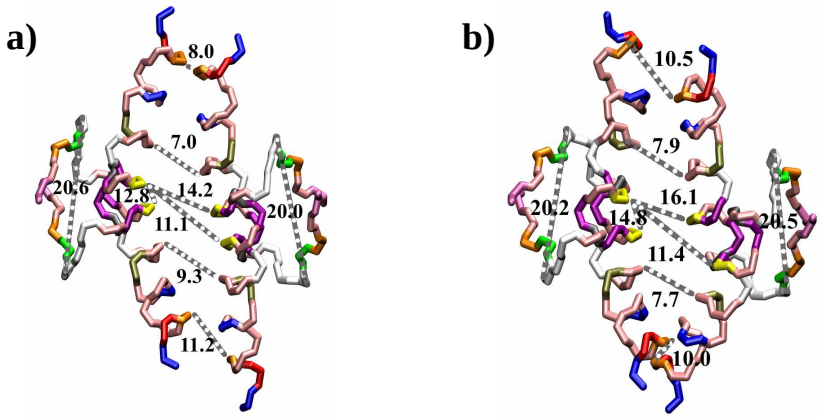


Figure 82. Representative structures obtained by clustering from the dynamics of the mutant protein with tetraguanidinium at 400K

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Figure 82 shows two examples of representative structures. In both of them a lost of symmetry was evident, which lead to a considerable distortion in the distances and orientation between couples of homolog residues. Figure 83 shows their corresponding contact maps.

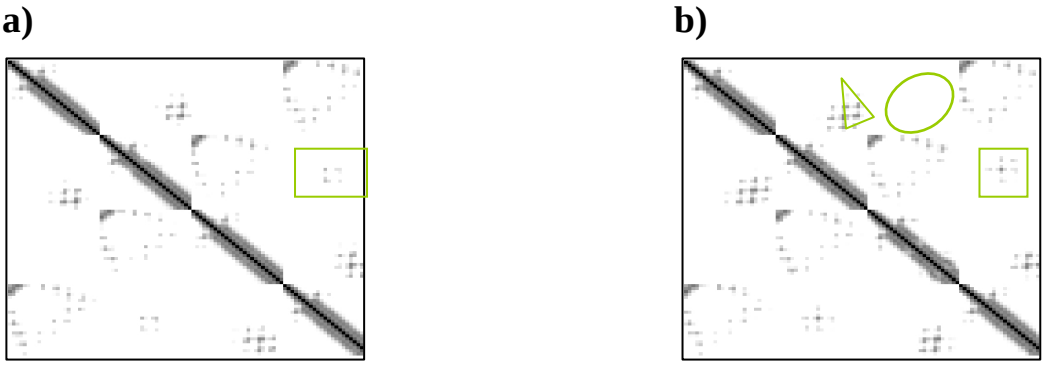


Figure 83. Contact maps corresponding to two of the 12 representative structures R337H p53TD Site2 at 400K.

VI.1.7. Conclusions on the effect of tetraguanidiniums on Site2

When positioned on the surface of the mutant protein at thermal denaturing conditions, the tetraguanidinium ligands do not remain bound at its original site. Indeed around the 2-3ns almost all the initial guanidinium-carboxylate interactions had been lost. From 3ns to the end of the simulation many protein-ligand complexes originated, showing the ligands bound to many different binding sites. The interaction between protein and tetraguanidinium is not strong enough to overcome the swings of the solvent around. It has to be taken into account that the backstroke of the tetraguanidinium is slightly hydrophobic and it is in direct contact with water.

Although the protein-ligand complex never reaches a stable structure in 10ns time trajectory, the binding avoids the typical disruption caused by the mutation R337H. The secondary structure of the helices appears less disturbed than in the absence of the ligands. However, they cause other kind of disorder, more generalized and especially related to the tertiary structure.

The presence of the ligands does not favour at all hydrogen bonds but does enforce most of the hydrophobic interactions within the region of the residue 337. Nevertheless, this latter effect is not sufficient to maintain the tetrameric structure as in the wild type but it suffers a more generalized disruption than the mutant protein with no ligands.

VI.2. Hexaguanidinium ligands on the surface of p53TD

VI.2.1. Motivation and methods

Based on the previous results on the tetraguanidinium, we proposed that a hexaguanidinium could bind the six anionic residues simultaneously (Glu336, Glu339, Glu343, Glu346, Glu349, Asp352). To test this hypothesis a simulation of the mutant protein with four hexaguanidiniums was directly

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

carried out at thermal denaturing conditions with exactly the same methodology as in previous calculations (see Chapter V). The box of 7010 water molecules in which the system was embedded had 6.2nm^3 volume. The dynamics was brought to 6ns. The results were compared to those obtained previously at the same temperature with the mutant protein in the presence of tetraguanidinums.

VI.2.2. Results

The system behaved very similar as in the case of having tetraguanidiniums. Also in this case all the interactions were lost after 3-4 nanoseconds. Indeed, only few of them remained after 2 ns trajectory.

It is worth mentioning that the guanidinium-carboxylate interactions were not gone following a consecutive order, in a zip fashion, from the ends of the ligands to the inner guanidinium groups but in a more disordered way. In the case of tetraguanidiniums, two ligands remained bound to the protein until the end of the trajectory by establishing punctual interactions only with the Asp352. In the case of hexaguanidiniums, two of them were completely lost but two others remained bound by one or two interactions with different chains.

The mean value for the X-CSCORE computed following the same methodology as in previous calculations was 3.8, similar to that for the simulation of tetraguanidiniums on Site1 and slightly lower than that for the simulation of the tetraguanidiniums on Site2. A plot of the RMSD of the backbone showed that, somehow, the hexaguanidiniums slightly stabilized the tetramer, delaying the disruption which typically appeared in their absence. However a more generalized disruption of the structure occurred.

The evolution of the secondary structure showed a disruption of the first part of the helix D since around 2 ns trajectory. The effect on hydrogen bond distances and hydrophobic interactions was similar to the experimented with the tetraguanidiniums.

VI.2.3. Conclusions

Hexaguanidiniums showed behaviour similar to that of the tetraguanidiniums on the protein surface. They proved to poorly stabilize the mutant structure due to their highly dynamic behaviour which leaded to structures in which the hexaguanidiniums were bound to multiple alternative sites, implying carboxylate groups of the same chain or of different chains. This caused a more generalized disruption of the structure.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

VI.3. Calixarene ligands on the surface of p53TD

VI.3.1. Introduction

Calixarenes are very attractive building blocks in supramolecular chemistry. They are easily synthesized, have a very particular architecture with an inner cavity that can be modified to obtain the desirable size. Concerning their applications, they are very versatile molecules.

Calixarenes[336] are [1_n]metacyclophanes that acquired their name because of the resemblance of the shape of one of the conformers of the smallest member of their family to a type of Greek vase called a calix crater. Calixarenes were classified as such since in 1978 Gutsche and coworkers[337] discovered them while working on products previously reported by Zinke.[338]

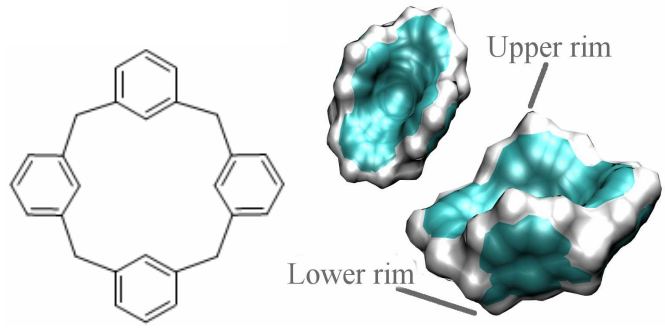


Figure 84. Scheme and different perspectives of the surface view of a [1_4]metacyclophane.

Calixarenes can exist in several conformations: depending on the position (up or down) of each of the aryls.[339] Several computational studies were carried out on the subject.[340] Thus in the case of calix[4]arenes[341] such as those considered in the present study, there are four possibilities as schematized in Figure 85. That preferred depends on the existing substituents in the upper and in the lower rim as well as on the physical state. In the solid state calix[4]arenes containing four endo-OH groups, as well as the O-substituted compounds, exist in the cone conformation.[342] However, in solution they are conformationally flexible at room temperature so that mixed conformations can be found. When containing four endo-OH groups the cone conformation is favoured against the others. Changing these groups for H-, MetO- or even EtO favours partial cone and alternate conformations. The interconversion between conformers[343] involves a “lower rim through” the annulus pathway and an EtO group is not still bulky enough to avoid it. Larger substituents do block this movement conformationally immobilizing the calixarenes. Another possibility to prevent complete conformational interconversion is the immobilization by establishing bridges[344] between the groups in the lower rim.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

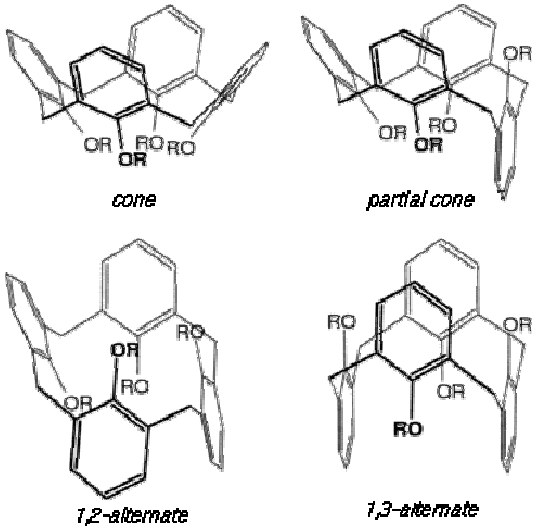


Figure 85. Conformations of calix[4]arenes

Both the lower and the upper rim of calixarenes can be modified by organic synthesis in infinite manners.[344] Many works have been reported on esterificated, etherificated and bridged calixarenes in their lower rim. Also intermolecular bridging has been reached. Most of the selective functionalization of calix[4]arenes at the lower rim has the aim of blocking them in a rigid cone conformation.[345] Other works deal with the modification of the upper rim: halogenation, sulfonation, alkylation, acylation, aroylation, arylation, aminomethylation, etc. Bridges have also been constructed in the upper rims of a single molecule or between molecules. Further modifications include methylene group oxidations, aromatic ring oxidation to quinones, aromatic ring oxidation to spirodienones, and the formation of chiral calixarenes by attaching a chiral group, selective functionalization, polymerization, etc.

The field in which the calixarenes are more useful is analytical chemistry[346, 347] because of their chemical sensing properties. Depending on their functionalization they can bind ions or even neutral molecules. They are also important in liquid-liquid extraction, membrane transport and ion selective electrodes. Specific calixarenes have been found to bind C_{60} and C_{70} fullerenes, forming 1:1 supramolecular complexes.[348]

VI.3.2. Goals

In collaboration with Javier de Mendoza' s Group at ICIQ and Ernest Giralt's Group at PCB, the binding affinity of two different tetraguanidilated calix[4]arene ligands (see Figure 86) with p53TD was tested expecting them to stabilize the mutant protein. Both calixarenes have four chains of one carbon and a positively charged guanidinium group in their upper rim. They differ in their lower rim. "CalixBridge" has ether chains as bridges between one and other fenolate. "CalixProp" has four propyl chains.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

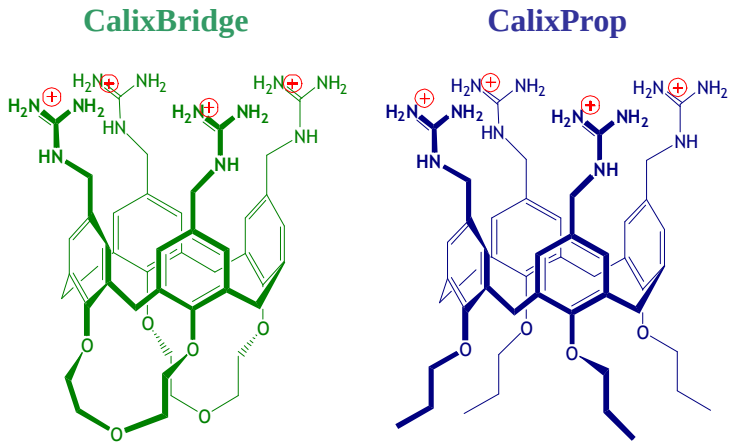


Figure 86. Calixarenes tested. At the left, the called CalixBridge from now on in this work. At right, the called CalixProp.

VI.3.3. Methods

Molecular Dynamics simulations were carried out on p53TD and R337H p53TD plus the ligands, at standard and at denaturing conditions. The database files of GROMACS[90] were modified accordingly to define suitable parameters for the calixarenes.

The protein-ligand complexes were composed of one protein and two equal calixarenes, placed in the specific way as to have their lower rims interacting with the hydrophobic pockets of the protein between the four chains forming the tetramer (Figure 87). The guanidinium groups were placed with the appropriate orientation to interact with glutamate 336 and glutamate 339 of two different chains situated in an "A-C" position (See the left part of Figure 87). Thus the calixarenes were surrounded by parts of the protein that have notable affinity for corresponding parts of the ligand.

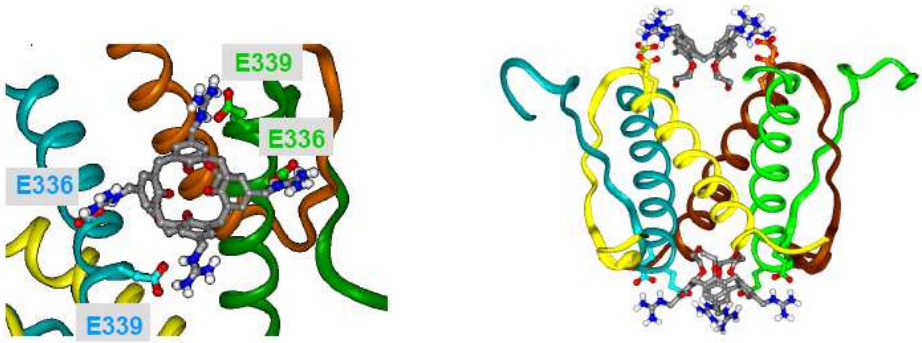


Figure 87. Initial structure, which had calixarenes bound to p53TD by establishing hydrophobic interactions with their lower rim and hydrogen bonds with their upper rim.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Each protein-ligand structure was embedded into a rectangular box. The structures involving p53TD had a volume such as to let a distance of 9Å from every wall to the nearest part of the complex.

The analyses were carried out over the trajectories for the new systems.

VI.3.4. Results

RMSD of the protein backbone

From the analysis of the trajectories obtained during the molecular dynamics simulations on the wild type protein, we observed that in the presence of CalixBridge the RMSD was slightly higher than in its absence at 300K (light blue trace compared to red trace in Figure 88) as well as at 400K (orange trace compared to dark blue trace in Figure 88). The largest difference was of around 0.5 Å. Moreover the fluctuations experienced by the protein at 400K (dark blue trace in Figure 88) were minimized in the presence of the ligand.

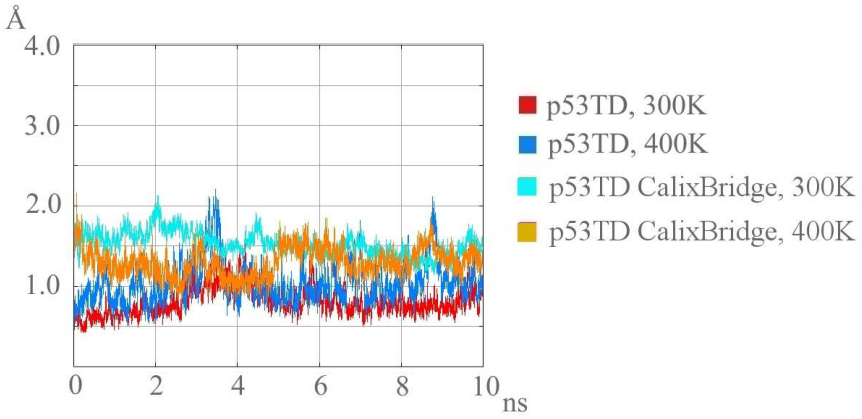


Figure 88. Comparison of the RMSD versus time for the wild type protein in presence and in absence of CalixBridge.

In the presence of CalixProp the behaviour at 400K was very similar to that in the case of CalixBridge (orange trace in Figure 89). At 300K (light blue trace) the trace evolution almost superposed with those from p53TD alone. Again sudden fluctuations like the suffered in p53TD at 400K were present neither at standard conditions nor at denaturing conditions.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

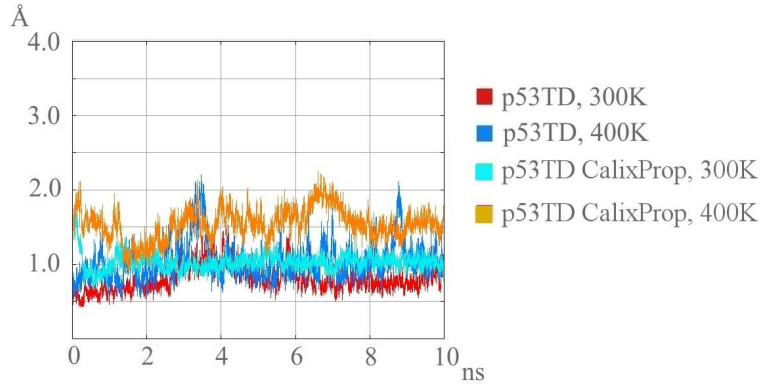


Figure 89. Comparison of the RMSD versus time for the mutant protein in presence and in absence of CalixProp.

Concerning the mutant protein, the RMSD revealed that at 300K the structures visited during the simulations when CalixBridge and CalixProp were present were almost identical to those obtained with R337H p53TD with no ligands (brown traces in Figure 90 and Figure 91).

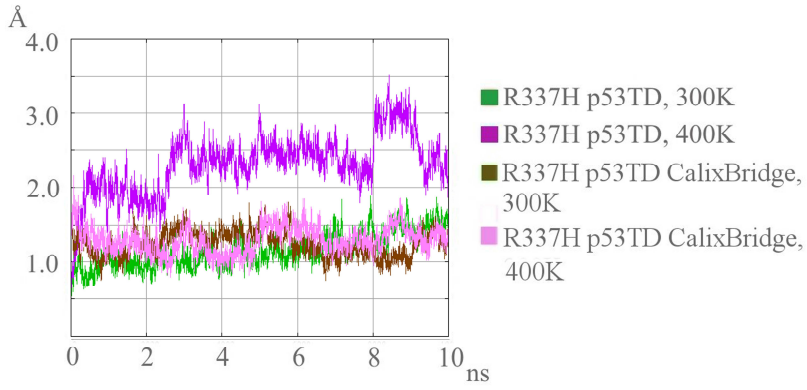


Figure 90. Comparison of the RMSD versus time for the mutant protein in presence and in absence of CalixBridge.

At 400K, a highly remarkable stabilization due to the calixarenes was revealed. Whereas the trace of the mutant protein (violet traces in Figure 90 and Figure 91) reached large values from the beginning and kept on raising until the end, the traces of the mutant protein with the calixarenes remained among the values characteristic for the protein at 300K (pink traces in Figure 90 and Figure 91).

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

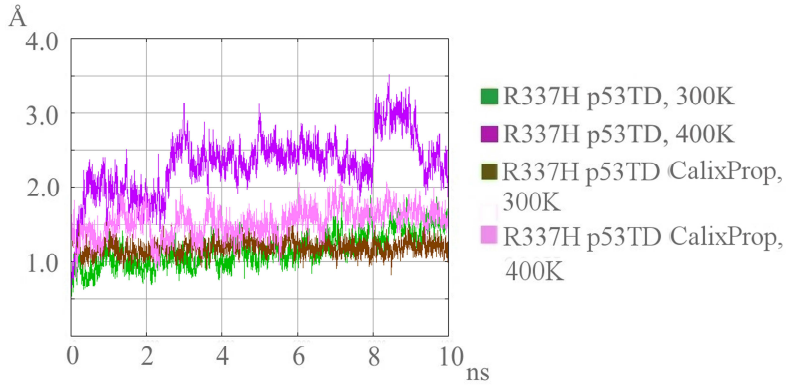


Figure 91. Comparison of the RMSD versus time for the mutant protein in presence and in absence of CalixProp.

In other words, we found that the calixarenes did not disturb the dynamics of p53TD, but also strongly stabilized R337H p53TD.

Time-evolution of the secondary structure

Figure 92 shows the time-evolution of the secondary structure during the simulation of the mutant protein with CalixBridge bound. The plot revealed a behaviour similar to that of p53TD at 400K. Also in this case, the ß-sheets between residues from Glu326 to Gly334 suffered slight alterations between chains B and C in the second half of the trajectory. The secondary structure of the four α-helices remained unchanged during the whole trajectory except for some punctual distortions at the tail of helix C around 8.5ns and some disruptions of the first residues of the helices at certain steps. However, these alterations were very small and always easily recovered avoiding spread of the distortion.

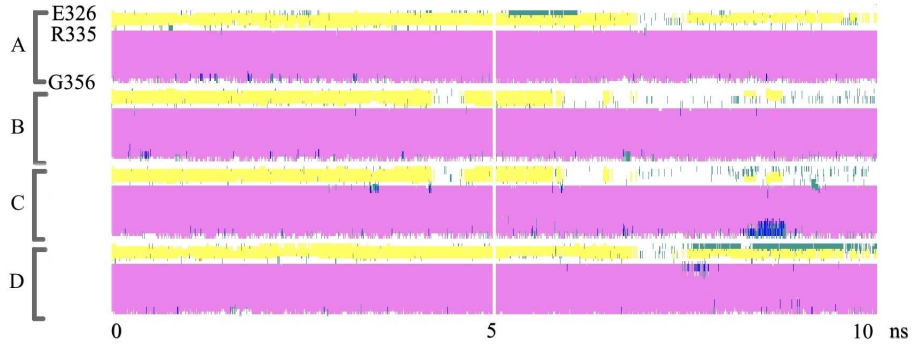


Figure 92. Time-evolution of the secondary structure of R337H p53TD at 400K when CalixBridge was bound.

Figure 93 shows the time-evolution of the secondary structure during the simulation of the mutant protein with CalixProp bound. In this case very small distortions in the ß-sheets were observed

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

between chains A and D as green traces indicate. The α-helices also remained undisturbed from the beginning to the end except for punctual recovered disruptions. Note that the alterations in Met340 and Phe341 previously seen in the simulation of R337H p53TD at 400K did not appear in any case when the ligands, CalixBridge or CalixProp, were bound to the mutant protein.

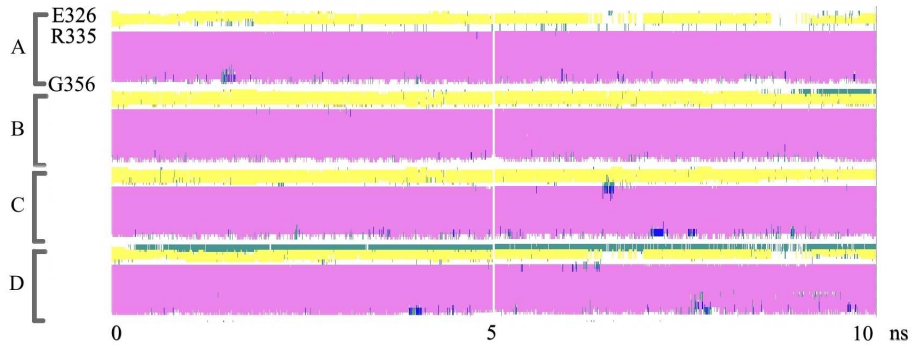


Figure 93. Time-evolution of the secondary structure of R337H p53TD at 400K when CalixProp was bound.

Hydrogen bond distances

The same analysis involving hydrogen bonds, carried out in the study of R337H p53TD versus p53TD at thermal denaturing conditions, was performed for R337H p53TD bound to calixarenes. Thus the interactions $Gln331_A$-$Asn345_D$, $Arg/His337_A$-$Asp352_D$, $Arg333_A$-$Asp352_D$, $Arg/His337_A$-$Asn345_D$, $Arg/His337_A$-$Glu349_D$ and $Arg333_A$-$Glu349_D$ were measured as well as the analogous interactions between chains D-A, B-C and C-B. The two hydrogen bonding interactions in the dimer-dimer interface were also measured: $Arg/His337_A$-$Glu343_C$ (and analogous $Arg/His337_C$-$Glu343_A$, $Arg/His337_B$- $Glu343_D$ and $Arg/His337_D$- $Glu343_B$) and $Lys351_A$-$Glu343_B$ (and analogous $Lys351_B$- $Glu343_A$, $Lys351_C$- $Glu343_D$ and $Lys351_D$- $Glu343_C$).

The analysis of the interactions between residues belonging to the same primary dimer revealed that the hydrogen bonding interactions remained completely disturbed in presence of the calixarenes. However in the cases in which the interactions were completely lost, the distances between the residues from each pair were shorter than in the mutant protein alone. This indicated that although the hydrogen bonding interactions were not recovered, the relative disposition of the residues was to some extent closer to their relative disposition in the wild type protein. For the sake of brevity only two plots are shown in Figure 94a and Figure 94b. The blue traces in the figures represent the wild type protein, the violet traces represent the mutant protein and the pink traces, the mutant protein with calixarenes bound on its surface. In this case the interactions correspond to $Arg333_A$-$Glu349_D$ and $Arg333_A$-$Asp352_D$. It can be observed that the pink traces representing the mutant protein with ligands did not superpose to the blue traces but had behaviour more similar to the violet.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

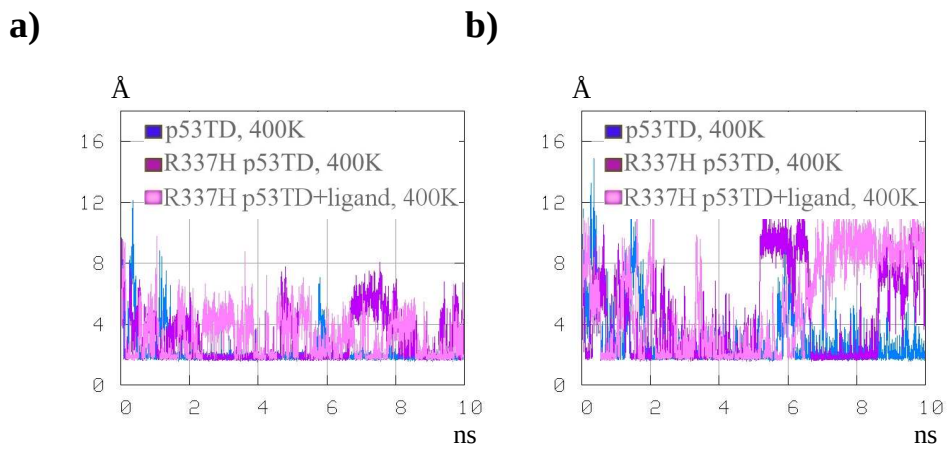


Figure 94. Hydrogen bonding distances a) $Arg333_A$-$Glu349_D$ and b) $Arg333_A$-$Asp352_D$ in p53TD, R337H p53TD and R337H p53TD-CalixBridge at 400K.

In the case of CalixBridge the distances were sometimes shorter than in the case of CalixProp but the difference was not significant. Concerning the interactions in the dimer-dimer interface, both Arg337-Glu343 and Lys351-Glu343 behaved exactly equal as in the mutant protein alone. The interactions Arg337-Glu343 were never at ideal distances but at distances even reaching values above 10Å. The interactions Lys351-Glu343 remained very stable during the whole dynamics fluctuating between 2 and 6-8 Å.

Therefore although the calixarenes stabilized the protein, the hydrogen bonds in general were not reinforced. Further analysis on the hydrophobic interactions would reveal the secret for the ligands to maintain the architecture near by the conformation of that characterized by X-ray diffraction.

Hydrophobic interaction distances

The same analysis performed on the hydrophobic interaction distances in the comparison of the wild type protein and the mutant protein was carried out here for the mutant protein in the presence of calixarenes at 400K. Thus the interactions taken into account were again $Ile332_A$-$Arg337_A$, $Ile332_A$-$Phe341_A$, $Arg/His337_A$-$Met340_A$, $Arg/His337_A$-$Phe341_A$, $Met340_A$-$Phe341_A$, $Met340_A$-$Met340_C$, $Ile332_A$-$Leu330_D$, $Phe341_A$-$Phe341_D$ and $Leu330_D$-$Phe341_D$ (and analogous interactions involving different chains).

Plots like in the previous section were obtained here. The wild type protein was represented in blue, the mutant protein in violet and the mutant protein with ligands, in pink. The three were compared at 400K. All the possible hydrophobic interactions existing in the mutant protein showed recovered, the pink traces always fluctuating between 2 and 4Å, perfectly superposed to the blue traces. Figure 95 shows two examples: $Arg/His337_A$-$Phe341_A$ and $Ile332_A$-$Phe341_A$ in the case of CalixBridge. These were two of the interactions that experienced the larger changes in the mutant protein at 400K. In these graphics, the difference due to the binding with calixarenes was evident.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

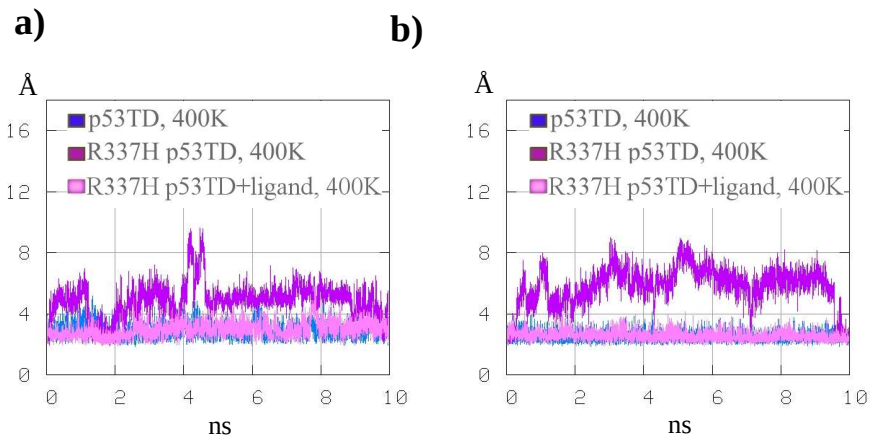


Figure 95. Comparison of hydrophobic interaction distances in p53TD, R337Hp53TD and R337H p53TD-CalixBridge at 400K; a) Arg/His337$_A$-Phe341$_A$;b) Ile332$_A$-Phe341$_A$.

Hence the hydrophobic distances maintained like in the wild type protein provided the preservation of the original functional structure even at denaturing conditions.

Cluster structures and Contact Maps

For both calixarenes, and as a proof of their stabilizing power on the protein, only one representative organization was found when clusters (Figure 96) in the ordinary conditions were searched for the mutant protein at 400K. The found structures were very similar to the averaged structure of the wild type trajectory. The three-dimensional disposition of the residues in the presence of CalixBridge appeared to be slightly more ideal. However the structure reached with CalixProp was as stable as the one reached with CalixBridge or even slightly more.

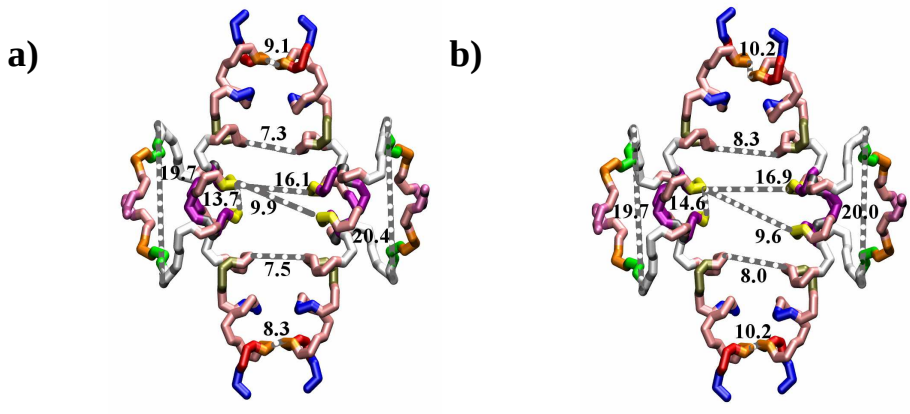


Figure 96. Representative structures obtained by clustering from the dynamics of the mutant protein bound to a) CalixBridge and of the mutant protein bound to b) CalixProp at 400K.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

VI.3.5. Conclusions

Both CalixBridge and CalixProp were able to bind p53TD and R337H p53TD for long time trajectories of 10ns. Their lower rims interacted with several sidechains of the residues involved in the hydrophobic core, which belonged to different monomers. At the same time the four guanidinium moieties in their upper rim interacted with the carboxylate groups on the surface of the protein also belonging to different chains. This did not disturb the wild type protein in a noticeable way, and strongly stabilized the mutant protein in a conformation very similar to that of the wild type protein.

The hydrogen bonding interactions essential in maintaining the tetrameric structure and that were highly disturbed in the mutant protein were not maintained in the presence of calixarenes either. However, the most important hydrophobic interactions in supporting the tetramer and which were also largely perturbed in the mutant protein were maintained at their ideal distances when the calixarenes were present.

The two calixarenes presented similar affinity for the R337H p53TD protein and stabilized it in a very similar way. However, CalixProp seemed to change its structure slightly more.

VI.3.6. Experimental results

Both CalixBridge and CalixProp were synthesised by V. Martos and J. de Mendoza at ICIQ and the protein-ligand binding was studied by S. Gordo and E. Giralt at the PCB. Indeed, CalixProp is still under study.

Experimental results on CalixBridge validated the rational design of J. de Mendoza. NMR experiments permitted to prove that the ligand bound the protein by inserting its lower rim in the hydrophobic core while the upper rim remained outside. Microcalorimetry experiments showed a recovering of the thermal stability of R337H p53TD mutant (see Figure 97), in presence of CalixBridge, to nearly that of the wild type protein. The computational results presented above fully agree with the experimental evidences. Further information provided by NMR provided insight on the mechanism of the protein-ligand recognition events. While for the stable and well packed wild type the two binding sites showed equivalent and independent, for the mutant protein, more flexible and less stable, the binding of the two ligands seemed to be sequential and implying some structural rearrangements to the plasticity of the protein.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

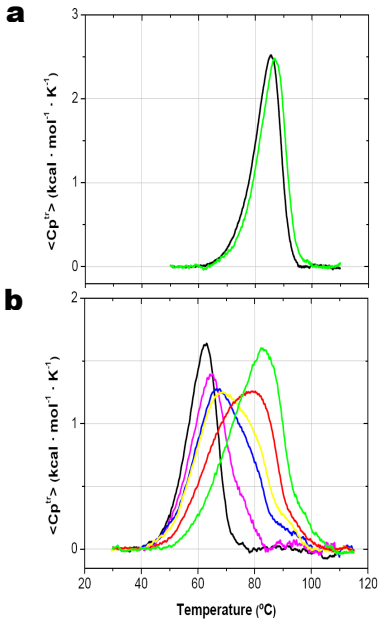


Figure 97. Thermal stabilization induced by CalixBridge. The presence of the ligand shifted the unfolding endotherms for p53TD (a) and p53TD-R337H (b) towards higher temperatures. While the stable wild type p53TD experienced a small uniform shift (from 85.5ºC to 86.9ºC), mutant p53TD-R337H melting temperature was asymmetrically shifted up to 20ºC. Black traces represent the free proteins and colour ones correspond to the presence of a ligand excess of 0.4 (magenta), 2 (blue), 4 (yellow), 8 (red) and 16-fold (green).

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Chapter VII

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

VII. Undecaguanidinium ligands on the surface of a B-DNA molecule

VII.1. Introduction

The Deoxyribonucleic acid (DNA) is the macromolecule which encodes the genetic material. It is a double-stranded molecule twisted into a helix in which the two strands intertwine with each other (see Figure 98a). Each strand is comprised of a sugar (deoxyribose)-phosphate backbone and attached bases through which it binds to the other complementary strand by hydrogen bonding (see Figure 98b and Figure 98c).

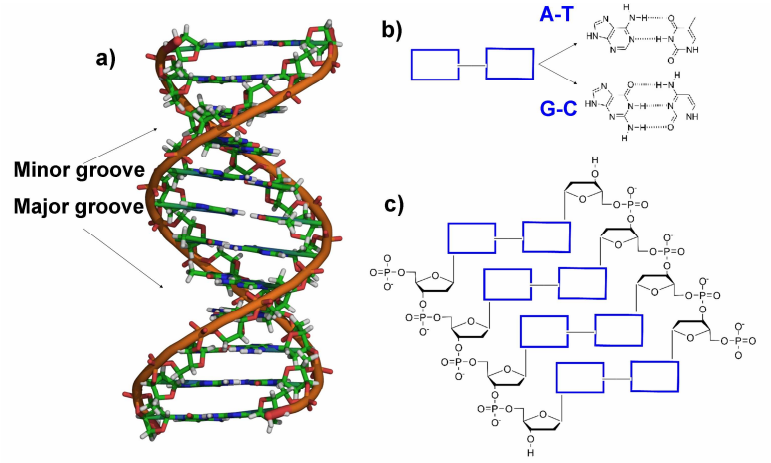


Figure 98. DNA structure.

As a result of the patterns of hydrogen bonding between complementary bases, the glycosidic bonds corresponding to each base pair are not diametrically opposite each other. Thus, taking as reference the glycosidic bonds, each base pair has a larger side and a smaller side which define the so-called major and minor grooves along the DNA molecule. Figure 99 shows the major groove between the two glycosidic bonds in a base pair marked in violet. The minor groove appears marked in green. Figure 98c shows both grooves in the double helix structure.

DNA can be found in several conformations.[349] Its most common form in living organisms is the B-DNA[350], in which the double helix is twisted as a right-handed spiral and the major and minor grooves are respectively 22Å and 12Å wide. As the major groove is more dependent on base composition, is there where proteins interact with specific sequences of DNA.[351]

There are four types of bases: adenine (A), thymine (T), guanine (G) and cytosine (C). Adenine exclusively binds to Thymine by means of two hydrogen bonds and Guanine to Cytosine by means of three hydrogen bonds. The sequential disposition of these bases along the molecule is what encodes the necessary information to construct proteins, RNA and other molecules in the cell.

Besides hydrogen bonds between base pairs, the double helix is also stabilized by a hydrophobic effect and by π-stacking interactions between the bases, which do not depend on the sequence of DNA.

In the DNA molecule the reversibility of hydrogen bonding is crucial as it permits the separation of the two complementary strands and their binding again, which is essential for the molecule to carry out its biological functions. The processes of replication[23], transcription[24, 25] and translation[26] are clear examples. DNA can establish two types of interactions with proteins: specific and non-specific.

Concerning specific interactions with the double helix, the base faces oriented towards their complementary bases to establish hydrogen bonds are not accessible from outside the double helix. However, some base features can still be recognised from the minor groove and especially from the major groove.[352] Figure 99 shows these features in red, being A hydrogen acceptors and D hydrogen donors.

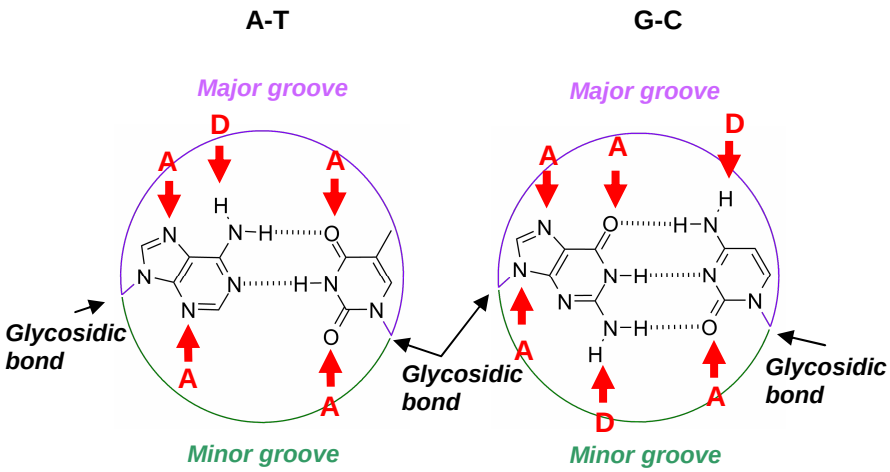


Figure 99. Base pair patterns recognized from the Major and Minor grooves.

When a protein is in contact with the major groove, the two base pairs and their two orientations can be distinguished[353], so the protein can choose one of the four possibilities A-T, T-A, G-C or C-G. The possible chemical contacts in the major groove for A-T are (reading form left to right in Figure 99): hydrogen acceptor, hydrogen donor, hydrogen acceptor and methyl group, or A-D-A-Me. This sequence can be distinguished from Me-A-D-A, which is the possible sequence of contacts in T-A. In the same way the sequences for G-C and C-G are also specific, being A-A-D-(blank) and (blank)-D-A-A respectively. In contrast, contacts in the minor groove appear nearly identical between A-T and T-A (with sequence A-A) and also between G-C and C-G (with sequence A-D-A).
Specific interactions with DNA are established by Transcription Factors, proteins that regulate transcription by facilitating the binding and activity of the RNA polymerase. Transcription Factors recognise determined sequences of bases by interacting directly with them in the major groove.

Non-specific interactions show lower affinity, as they are mostly weak interactions between positively charged residues on the surface of the protein and the negatively charged phosphates in the major and minor grooves of the DNA backbone. Histones[354] are examples of proteins that bind DNA in this way. They are basic proteins that are rich in positively charged lysine and arginine.

These residues bind to the negatively charged phosphates of DNA. The DNA wraps around the proteins thus leading to a more compact and organized form of the genetic material, called chromatin.

In medicinal chemistry, the rational design, inspired on nature, of specific and non-specific DNA-binding ligands has growing importance and could give rise to potent therapeutic drugs in the future.

Concerning specific interactions, the Triplex-forming oligonucleotides[355] (TFOs) and the Peptide Nucleic Acids (PNAs)[356] are two examples of tools widely used. Both kinds of ligands bind to the DNA by invading the double helix and forming triplex, being complementary to one of the DNA strands. TFOs and PNAs can inhibit transcription by competing with transcription factors or, on the contrary, can increase gene expression if they are bound to a transactivator protein domain. Their capability to bind DNA can also be used to increase rates of site-specific mutagenesis and site-specific recombination, by linking them to mutagens or DNA fragments homologous to the targeted site.

Other designed molecules have been synthesized combining different moieties to small molecules known to bind DNA in a specific manner, such as netropsin, distamycin A and lexitropsins.[357] Following this strategy Juan B. Blanco et al. designed in 2006 a synthetic short peptide which binds to DNA in a highly (sequence) specific manner.[358]

Intercalation is another useful type of DNA-ligand interaction, which can be specific or non-specific. It consists of a ligand of the appropriate size and chemical nature that inserts itself in between base pairs of DNA by establishing π-π stacking interactions, and often hydrogen bonds, with the planar aromatic bases. In a lot of cases the intercalators are polar or even charged so electrostatic interactions seem to have also large importance. These ligands are generally polycyclic, aromatic and planar and are used as antibiotics and in chemotherapy to impede the replication of tumour cells.

Other non-specific interactions are used to deliver DNA into the cell nucleus.[359] There are two types of gene carriers: virus[360] and cationic molecular carriers[361] combination of nucleic acids with cationic lipids(lipoplexes) and/or cationic polymers(polyplexes). Cationic lipids and cationic polymers bind DNA molecules by electrostatic interaction with the phosphate backbone. Due to the positive charge the complexes interact with cell surfaces and are absorbed by endocytosis. Polycations used for gene delivery are very promising due to the infinite structural possibilities with different physicochemical properties.[362] Polycations used are normally polyamines, cationic at physiologic conditions. The carriers must show low toxicity, must evade the immune system, minimize interactions with proteins in the plasma and non-targeted cell surfaces and must not aggregate. Physicochemical stability of simple DNA-polycation complexes is often limited by the fact that in physiological saline the charge is masked by the presence of salt and the complexes aggregate. In order to overcome this problem polymers such as polyethylene glycol (PEG)[363], dextran[364] and poly(N-[2-hydroxypropyl]methacrylamide) (PHPMA)[365] have been used to provide steric stabilization.

When a DNA molecule is complexed with a vector the molecular size and surface charge of the complex affect its distribution in the body and the efficiency of transfection.[366-369] Lipoplexes and polyplexes combined with small molecules like folate or galactose or proteins like transferrin and antibodies, which interact with surface receptors, should allow the targeting of specific cell types. For instance, transferrin[370] has been used to target tumour cells and galactose-containing ligands[371], to target hepatocytes.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

VII.2. Goals

In collaboration with the Javier De Mendoza' s group the main objective was the prior-to-experiment in-silico design of several oligoguanidinium ligands to explore whether they can potentially bind DNA in a non-specific manner. The oligoguanidiniums tested were formed by eleven bicyclic charged guanidinium groups linked to each other by means of different kinds of spacers (Figure 100).

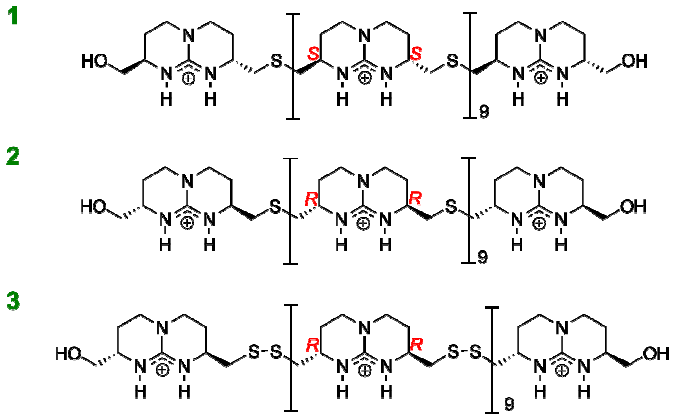


Figure 100. Ligands tested.

A tetraguanidinum compound composed of four bicyclic guanidinium subunits linked through short thioeter spacers (-CH_2-S-CH_2-) had been previously studied bound to two different tetraanionic sequences on p53TD surface (see Chapter VI). Here, seven more guanidinium subunits were added to the tetraguanidinium structure leading to an undecaguanidium ligand. This was tested on its two enantiomeric forms SS (ligand **1** in Figure 100) and RR (ligand **2** in Figure 100). By changing the thioeter spacers to disulfide spacers (-CH_2-S-S-CH_2-) ligand **3** in Figure 100 was also constructed and tested.

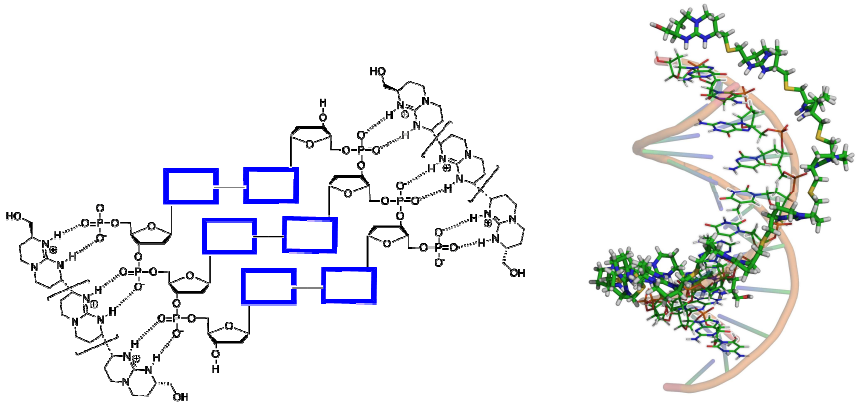


Figure 101. Undecaguanidiniums on the phosphate backbones of a dodecanucleotide (only one ligand shown for clarity).

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

The motivation to test these particular ligands came from several experimental reasons. Ligands **1** and **2** had been previously synthesized and proved to be active as cell penetrating vectors by the experimental group.[372] Ligand **3** was suggested as its synthesis could be easily carried out by means of the polymerisation of bicyclic guanidinium moieties with sulphydril groups in oxidative conditions. The length of the oligoguanidinium could be controlled by applying the principles of Dynamical Combinatorial Chemistry[373], which allows the synthesis to occur through reversible covalent reactions while in the presence of a template. Thus in presence of a dodecanucleotide the guanidinium-SH moieties would bind the phosphates establishing at the same time disulfide bonds with other neighbour guanidinium-SH moieties until reaching the length desired of the undecaguanidinum. This would be applicable to DNA fragments of different lengths.

Once demonstrated high affinity of the oligoguanidiniums proposed for DNA molecules, their applicability would be widely versatile. The ligands confer lipophilicity to the DNA structure, as phosphates are prevented from establishing strong hydrogen bonds with water molecules, these being displaced by guanidinium groups. The gain in lipophilicity would permit a separation of the fragments from water to organic phases. Having a mix of DNA fragments of different lengths, their separation by means of binding to oligoguanidiniums would be specific and dependent of the ligand length.

Further applications taking advantage of the lipophilic properties transferred to the DNA fragments could be related to transfection. Indeed oligoguanidinium-based cell delivery systems have gained interest in drug delivery since one decade ago due to the previous identification of guanidinium-containing proteins, such as Antennapedia homeodomain (Antp) and Tat peptide, able to transport non-permeant molecules across the cell membrane.[374-376] Non-peptidic structures with similar functionality are pursued since peptides are easily hydrolyzed by peptidases in vivo and have complex pharmacokinetic properties. Tetraguanidinium compounds analogous to **1** and **2** in the present chapter have been proved to efficiently penetrate cells.[372] These compounds have shown more efficient translocation through HeLa membranes than Antp or Tat at very low concentrations. Furthermore the latter do not cross the mitochondrial membrane[377] whereas tetraguanidinium specifically targets mitochondria.

One step forward could be the addition of new moieties to the structure of the ligands under study in the present work, in order to reach new anchorage points with the DNA base pairs. These new moieties could be, for instance, intercalant groups which would bind to the nitrogenated bases by π-stacking. Some intercalant molecules have proved to have more affinity for some particular bases so if they were included in the design of the ligand, a sequence-specific binding could be reached.

VII.3. Structures under study

The DNA structure used was the so-called Dickerson Dodecamer (5'-CGCGATATCGCG-3'). The structure was constructed with the module Biomolecules in the InsightII software (Biosym Technologies Inc. San Diego, CA).

The undecaguanidiniums were docked by hand to the dodecanucleotide strands by orienting the –NH donor groups in the guanidiniums towards the –O atoms in the phosphate backbone of the DNA molecule, therefore establishing eleven charged hydrogen bonds with each strand. The hydrogen bonding distances were lead to ~2Å. The angle between the atoms –N-H-O intervening in the interaction was fixed at approximately the ideal value of 180º, thus the hydrocarbon crest of the ligand pointing to the outside of the complex which would be surrounded by solvent. The program used to construct the complexes was Maestro (Schrodinger Inc., New York, NY, USA).

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

It should be noted that the $–CH_2-S-CH_2-$ spacer, which nicely fits the distance between phosphates, is much shorter than $–CH_2-S-S-CH_2-$. C-S bonds are 1.8Å in both cases but in the second case S-S bonds lengthen the spacer in 2Å. However, the flexibility of the –C-S-S-C- dihedral angle which ideal value is 90º, allowed also in this case a correct orientation of the guanidinium groups facing the phosphates.

The complexes tested were: DNA plus two ligands in the three cases studied and DNA plus one ligand for the cases of ligands **2** and **3**.

VII.4. Methods

Each of the DNA-ligand structure was set into a box of water molecules, with a volume such as to let a distance of 9Å from every wall to the nearest part of the complex. SPC[326, 327] model was used to simulate water molecules. Counter-ions were added when needed to ensure the overall electric neutrality of the system. Periodic boundary conditions were used. Amber94[81] was the force field chosen to describe the bonding, van der Waals and electrostatic interactions. The simulations were carried out using GROMACS.[90] The database files of GROMACS were modified to include suitable parameters for the ligands.

The atomic charges in the ligands were assigned by following a systematic method which avoided ab-initio calculations on the whole ligands, what was very time-spending. The charges for the molecules in Figure 102 were calculated by applying the Restrained Electrostatic Potential fit[82] (RESP) from a quantum mechanics optimization with Gaussian03[140] at the HF/6-31G(d) level. From a triguanidinium like the represented in Figure 102a, atomic charges for any terminal bicyclic guanidinium moiety and also for any central bicyclic guanidinium moiety were obtained (marked with a red rectangle). The charges for analogous atoms were slightly different between the two fragments. From calculations on molecules as the represented in Figure 102b, the charges for the atoms in the spacers were assigned. Then, all the values were fit in order to conserve overall neutrality of the ligands (the final charge assignment is detailed in the Appendix). Thus, if new ligands based on bicyclic guanidinium moieties were constructed, the charges would have to be calculated only for the spacers, in case they were different from the used in the present work.

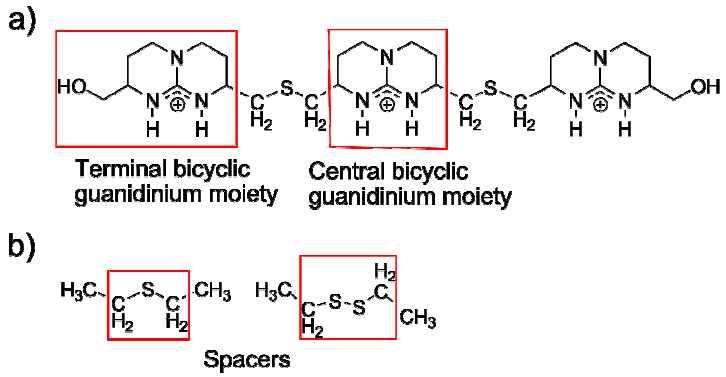


Figure 102. Scheme showing the molecules on which RESP calculations were applied to obtain the atomic charges for the ligands under study.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

The equilibration protocol used was three-stage: 100ps constant temperature; 100ps constant pressure; 100ps constant temperature. The total MD time was of 5ns except for the case of one simulation in which the total time was of 10ns. Structures were picked every 2 ps. Long-range electrostatic interactions were taken into account by the Particle-mesh Ewald (PME) method.[378] VdW interactions were truncated at a cut-off distance of 10Å. Newton equations were integrated using a time step of 2fs. The simulations were performed at constant temperature in the NVT ensemble using Berendsen coupling method[95], with a coupling time of 0.1ps. The constraints to the equilibrium for the bond lengths were managed by means of the LInear Constraint Solver LINCS algorithm.[379]

The molecular dynamics carried out were the following:

- 5ns DNA plus two ligands **1** at 300K with fixed DNA.
- 10 ns DNA plus two ligands **2** at 300K with fixed DNA.
- 5ns DNA plus two ligands **3** at 300K with fixed DNA.
- 5ns DNA plus one ligand **3** at 300K with flexible DNA.
- 5ns DNA plus one ligand **2** at 400K with flexible DNA.
- 5ns DNA in absence of ligands at 300K
- 5ns DNA in absence of ligands at 400K

In simulations in which two ligands were present no counter-ions were added to the system as the 11x2 positive charges of the guanidinium groups were neutralized by the 11x2 negative charges present in the DNA dodecamer. Indeed oligoguanidiniums with exactly 11 bicyclic groups were constructed on this purpose, in order to avoid competition with Na+ ions in these particular dynamics. In the simulations of one only ligand bound to the DNA molecule 11 Na+ ions were added in order to neutralize the system. In absence of ligands, 22 Na+ ions were added for the same reason.

GROMACS analysis tools were used to analyze the trajectories. Time-averaged interaction distances were computed for each guanidinium phosphate interaction in a simulation. The distances were considered between the atoms directly involved in the interactions. Root Mean Square Deviation of the DNA molecule was analyzed in different cases in order to see to what extent the ligands disturbed or not the structure of the dodecanucleotide. Radial Distribution Functions were computed to observe the changes in the solvent and ion distributions around the double helix due to the presence of the ligands.

VII.5. Results

The simulation of a fixed oligonucleotide with two ligands **1** bounded on the surface at standard conditions was compared to the simulation of the DNA molecule with two ligands **2** bounded at the same conditions. The ligands, only differing in their chiral centers behaved differently along the dynamics. In the case of ligands **1**, both of them changed their initial conformation in such a way that the central guanidinium bicycles lost interactions and made loops falling into the DNA groove (see Figure 103a). This was caused by the tendency of the ligand with RR centers to turn left forming a left-handed helix whereas the DNA molecule chosen in the present study was right-handed. The fallen of a part of the ligand in the groove was evident before 5ns. On the contrary, the global right-handed orientation of ligand **2**, with SS centers, permitted the molecule to remain perfectly aligned with the phosphate path with the –N-H donor groups pointing to the phosphates and the opposite site of the bicycles oriented towards the solvent (see Figure 103b). Ligand **1** would probably fit nicely a Z-DNA molecule, which is left-handed.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

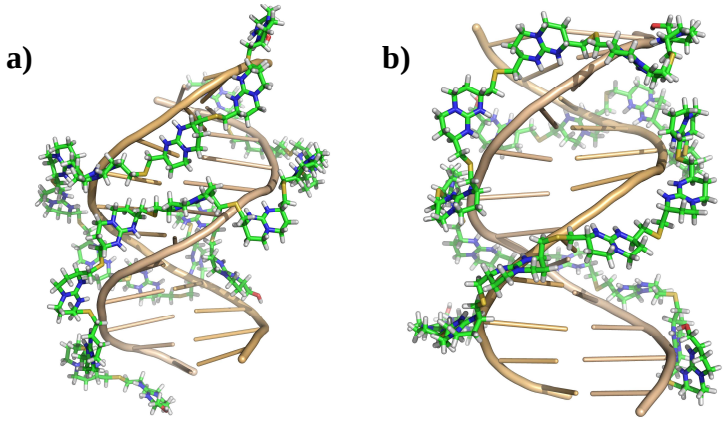


Figure 103. Side view of the structures picked at 4.6ns time from trajectories with the DNA atoms fixed and two ligands 1 and 2.

Ligands **2** remained on the surface of the fixed DNA molecule for 10 ns total time trajectory. The average distance during this time was calculated for each of the 11 hydrogen bonding interactions established between the DNA and the ligand (see Figure 104 b). For the sake of clarity two equal ligands bound to the same DNA molecule are called herein ligand A and ligand B. The DNA-ligand interactions are referred to with one number from 1 to 11, being interaction 1 the one established with one guanidinium group at the end of the ligand. The interaction established with the following guanidinium is named number 2 and those consecutive have increasing numbers until number 11, corresponding to the interaction at the opposite end of the undecaguanidinium. The average interaction distances were measured between the hydrogen atoms of the –NH donor groups of the guanidiniums and the acceptor atoms –O of the phosphate groups.

Figure 104 b shows that most of the interactions remained fixed at around 2Å, which is the ideal interaction distance. At the end of 10ns one of the two ligands, ligand A, had lost only interactions number 1 and 11 (values larger than 2-3 Å). Ligand B had lost number 1, 9, 10 and 11. 1 and 11 were terminal interactions and were the first gone. Further on, 10 and after, 9 were also lost. The 5 central interactions remained especially stable along the whole trajectory.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

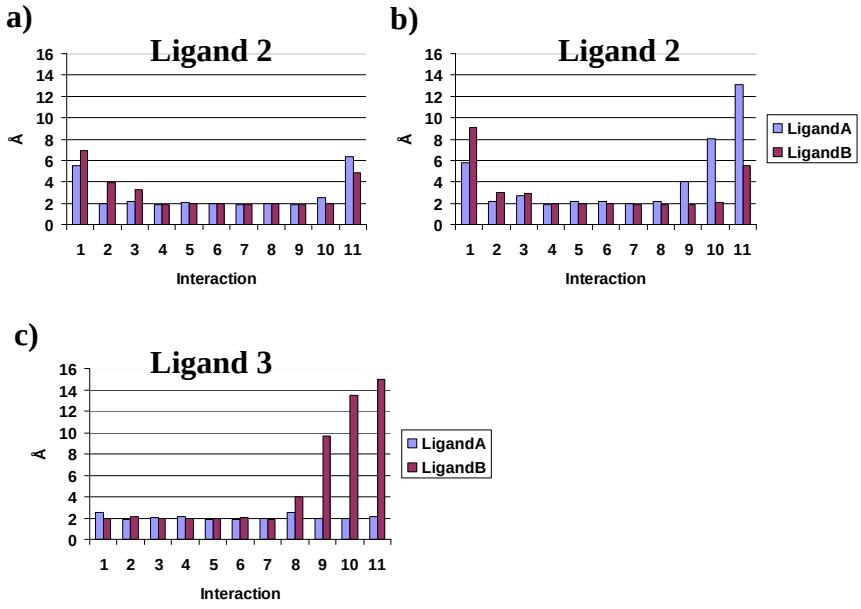


Figure 104. Time-averaged interaction distances for the consecutive DNA-ligand contacts (1 - 11) of each of the two oligoguanidiniums A and B bound to the double helix; a) , and c) refer to the first 5ns trajectory with ligands 2 and 3 on fixed DNA at standard conditions; b) refers to the first 10ns with ligands 2.

In a 5ns analogous simulation, ligand **3** showed similar behaviour to that of **2**, as it had also RR chirality. Figure 104c shows the time-averaged distances in this case. Although the simulation was extended to 10ns in the case of ligand **2**, the distances were also averaged for the first 5 ns and shown in Figure 104 a, for a fair comparison with the other ligand.

At 5ns only one/two interactions had been lost in the case of ligand **2**. In the **3** case (Figure 104 c) all the interactions in ligand A remained stable whereas 3-4 of those terminal in ligand B were lost. The evolution of the binding of ligands **3** was similar to that of ligands **2**. In both cases the visualization of the dynamics showed a higher instability of the interactions at both ends of the undecaguanidinium in comparison with the rest of interactions. The terminal interactions were first lost, sweeping out the subsequent units, which consecutively dragged along further guanidine units in the same way. The central units remained relatively undisturbed and attached to their original binding sites.

Figure 105 shows a top view of the structure picked at 4.6ns from the trajectory of ligand 2. Only one ligand molecule has been drawn for clarity. It can be observed that the hydrogen bonds (shown in yellow) are well oriented in most of the guanidinium-phosphate pairs and that four of the terminal interactions are lost.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

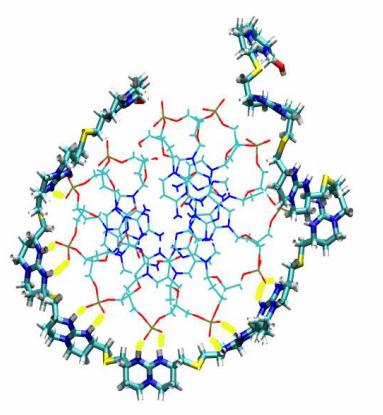


Figure 105. Top view of the DNA-ligand 2 structure at 4600ps at 300K.

It is remarkable that the simulations permitted to prove a similar behaviour for oligoguanidiniums with spacers as different in length as $–CH_2-S-CH_2-$ and $–CH_2-S-S-CH_2-$. Indeed the disposition of ligands **2** and **3** around DNA strands was to some extent different, ligand **3** adopting more winding conformations to maintain the 11 hydrogen bonds. This adaptation was possible due to the flexibility of the –C-S-S-C- dihedral angle.

5ns Molecular Dynamics with ligands **2** and **3** with no restraints for the DNA molecule showed very similar results. Also comparable results were obtained from simulation of an only ligand **2** bound to one of the strands of a flexible DNA double helix. This stated that the dynamics of a ligand onto a DNA strand was independent of the existence of a second ligand onto the complementary strand.

The strength of the DNA-ligand interactions was evident from a 4ns trajectory obtained at 400K in which a single ligand **2** was interacting with a flexible DNA molecule. The time-averaged interaction distances were around 2Å except for 3 of the terminal, which reached values of around 5 Å.

Figure 106 and Figure 107 show the RMSD values obtained for the entire DNA molecules in simulations carried out applying no restraints to the dodecanucleotide at 300K and 400K respectively. The structure of reference was the initial DNA structure.

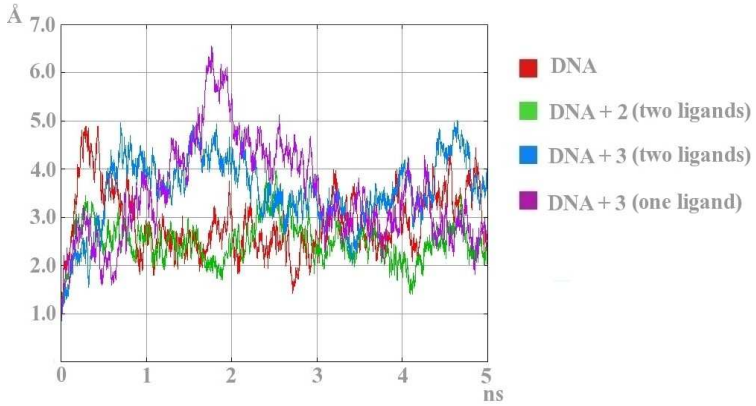


Figure 106. DNA RMSD in trajectories carried out with flexible DNA at 300K.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

At 300K the DNA molecule in absence of ligands fluctuated between 2 and 4 Å (red trace in Figure 106). The presence of two ligands **3** caused an increase in the fluctuations to values of 2-5 Å (blue trace in Figure 106). The presence of one unique ligand **3** showed structures with higher RMSD values of 5-6.5 Å in the 1.5-2 ns interval and values of 2-5 Å further on (violet trace in Figure 106). However it is remarkable that in any case the ligands caused an irrecoverable change in the dodecanucleotide as the highest values in RMSD were not caused by changes in the binding of the couples of nitrogenated bases but by bending of the whole DNA double helix.

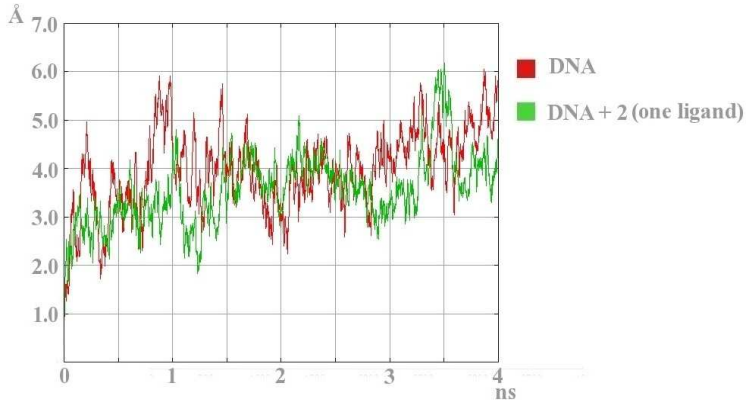


Figure 107. DNA RMSD in trajectories carried out with flexible DNA at 400K.

Ligand **2** (green trace in Figure 107) did not disturb at all the double helix, showing values very similar to those obtained when no ligand was present, even reducing to some extent the fluctuations. This profile was reproduced at 400K in the presence of an only ligand **2** (Figure 107).

It is remarkable that in the simulations the undecaguanidiniums had to compete with water molecules and with Na+ ions. Figure 108a shows the representation of Radial distribution functions (RDF) between a 6th phosphate group and the water molecules in presence and in absence of ligand **2**. At short distances some obvious picks appeared in the RDFs, which indicated that the water molecules packed around the phosphate groups in shells. The higher peak appeared at around 1.7Å showing a higher organization at this distance due to the influence of the phosphate groups with which the water molecules are able to establish hydrogen bonds. The following peaks were probably due to the existence of the particular organization of this first layer. From around 6.5 Å on the RDF tend to a value of 1, as the function described the average density at long range. Both in presence and in absence of ligand the peaks appeared at the same distances. However in the latter case their intensities were lower due to the ligand occupying space in one direction at these distances.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

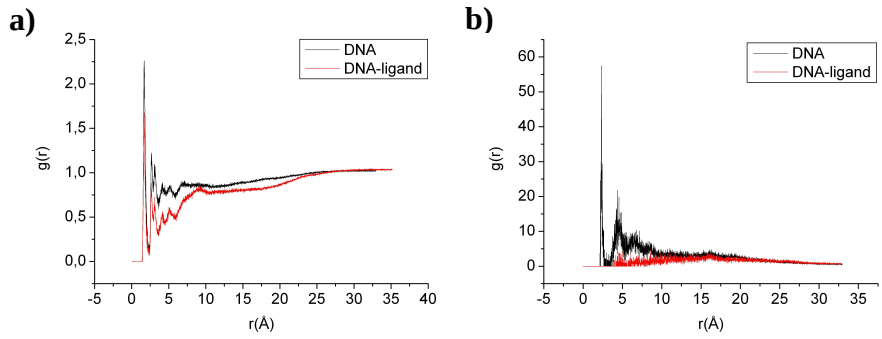


Figure 108. Radial distributions between 6th phosphate group and water molecules (a) and Na+ (b).

Figure 108b shows the representation of the RDFs between a 6th phosphate group and the Na+ ions. In this case the functions showed more drastic differences in presence and in absence of ligand. In absence of ligand (black trace) a highest peak appeared at around 2.3Å showing a much higher probability of finding Na+ ions within this distance. This was in agreement with a study reported in 2006 on the condensation of Na+ ions around the DNA.[380] The following peaks were much lower and broader indicating higher motion of the ions. In the presence of ligand no peaks appeared, the function profile tending to 1 from the shortest differences. This meant that the oligoguanidinium drastically displaced the Na+ ions from the proximity of the phosphate groups.

Thus, even in a highly competitive environment for the phosphate groups, the guanidinium-DNA interactions were maintained along the trajectory. Moreover, some of the terminal interactions which were lost were some nanoseconds further on recovered thanks to the anchorage provided by successive guanidinium groups.

VII.6. Conclusions

The ligands tested proved to fit the phosphate backbone of the DNA binding to it with high affinity by means of 11 hydrogen bonds and remaining on the surface during long time simulation. This took place in simulations were no restrictions were applied to the DNA molecule as well as in simulations were all the DNA atoms were fixed.

It was necessary that the chiral centres of the ligands had an RR configuration to adopt a right-handed orientation able to align with the helix of the dodecanucleotide simulated (B-DNA).

The length of the oligoguanidinium was important in order to reach a cooperative action of the hydrogen bonds strengthen by the existence of nearby anchorage points (consecutive hydrogen bond interactions).

The most terminal DNA-ligand hydrogen bonds showed to be the weakest, being the first in being lost and pulling along consecutive interactions. Ligands lost one to three terminal interactions sometimes recovered some nanoseconds afterwards thanks to the existence of neighbouring anchorage points.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Ligands **2** and **3** behaved very similarly in spite of having very different spacers. Although in principle the distance between the guanidinum moieties in ligand **2** is more similar to that between the phosphates, the flexibility of the $–CH_2$-S-S-CH_2- dihedral angle in ligand **3** permitted also an appropriate fit in this case.

When dealing with a flexible DNA, none of the ligands caused irrecoverable changes in the dodecanucleotide structure, although ligand **3** seemed to cause a slight bending of the molecule.

Even at thermal denaturing conditions the ligand remained on its initial binding site for all the time computed (4ns) without losing interactions and without disturbing the DNA structure.

The presence of one ligand on a particular strand of the DNA molecule did not influence the behaviour of a second ligand on the complementary strand.

High affinity of the oligoguanidiniums was proved to successfully compete with water molecules and Na+ ions with high tendency to interact with phosphate groups.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Chapter VIII

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

VIII. General conclusions of the thesis

The following conclusions can be drawn from this thesis:

1- The methods BP86/TZP and B3LYP/6-311G** were shown to be suitable for the study of the tautomeric equilibrium involving the 2-ureidopyrimidinone (UPy). The Donor-Donor-Acceptor-Acceptor (DDAA) dimer was demonstrated to be stronger than the Acceptor-Donor-Acceptor-Donor (ADAD), due to the more favourable secondary interactions.

2- The presence of a CF_3 group in position R_1 of the UPy was shown to lower the dimerization energy for DDAA and raise it for ADAD and also induced a thermodynamic destabilization of the DDAA dimer with respect to the most stable monomer involved in the equilibrium. This was in full agreement with experimental data. When the solvent effects of $CHCl_3$ were included, the equilibrium between the monomers was shown to be displaced towards the DDAA monomer.

3- The method BP86/TZP was shown to be suitable to study the interaction between $(CTV\text{-}3UPy)_2$ nanocapsules and fullerenes. However, an empirical correction term was required in order to correctly describe the π-stacking interactions between the host and the guest. RbRb was the $(CTV\text{-}3UPy)_2$ nanocapsule that presented the best balance between the energy released for forming the host and the host-guest interaction. The sequence for the fullerene with highest affinity for the capsule to that with lowest affinity is the following: $C_{84}>C_{90}>C_{78}>C_{76}>C_{70}>C_{60}$. Experimentally, it is known that $C_{84} > C_{70} > C_{60}$. Therefore our results are in agreement with experiment and are reliable in a predictive sense.

4- Molecular dynamics at thermal denaturing conditions was an appropriate method to study the difference in stability between the protein p53TD and its R337H mutant. As recently reported, the hydrogen bonding network supporting the tetramer was observed to be disturbed by the mutation. However, in this study, the hydrophobic interactions in the R337H p53TD were demonstrated to also be disturbed at 400K while they were unaffected in the p53TD under the same conditions. The disruption process started with the lost of hydrogen bonds and hydrophobic interactions with the residue 337, which caused subsequent changes in nearby residues ending up in a distortion of the secondary structure of the α-helices. The disruption was shown to later spread to the core of the protein.

5- The interactions of certain tetraguanidine ligands with the surface of p53TD and its R337H mutant were investigated. The interaction between the wild type protein and this kind of ligands had been previously tested by NMR, showing protein-ligand affinity. A certain site herein called Site1 had been thought to be the most favourable. However, the simulations carried out in the present work showed that the ligands did not form a stable complex when bound to this site but they had a highly dynamic behaviour, leading to structures in which the tetraguanidines bound to multiple alternative sites. Thus the ligands proved to poorly stabilise the mutant protein structure due to this high dynamic behaviour on the protein surface. This suggested the design of a longer hexaguanidine in order to have two additional anchorage points. However, the behaviour of these ligands showed very similar to that of the tetraguanidines. They did not remain bound to a specific site during reasonable time either.

6- The calix[4]arene ligands herein called CalixBridge and CalixProp did reveal to be suitable for the stabilisation of the mutant protein. Their lower rim interacted with several sidechains of the residues involved in the hydrophobic core, which belonged to different

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

monomers. At the same time the four guanidinium moieties in their upper rim interacted with the carboxylate groups on the surface of the protein, also belonging to different chains. This did not disturb the wild type protein in a noticeable way and strongly stabilized the mutant protein in a conformation similar to that of the wild type protein. Experimental results on CalixBridge showed a recovering of the thermal stability of the R337H p53TD mutant, in the presence of the ligand, to nearly that of the wild type protein. Therefore, the results obtained from the molecular dynamics were in agreement with experiment.

7- The interactions of undecaguanidines with the phosphate backbone of a B-DNA molecule were investigated aimed at predicting whether the ligands would have affinity for the DNA, binding in a non-specific manner. The undecaguanidines were composed of eleven bicyclic charged guanidinium groups bound by means of thioether or disulfide spacers. Those ligands with an RR configuration of the chiral carbon atoms showed to have considerable affinity for the dodecanucleotide, competing with Na^+ ions and remaining bound to the DNA surface for long time molecular dynamics.

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

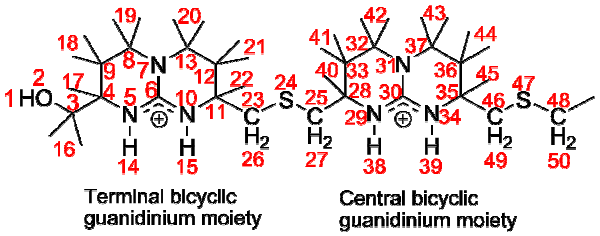
Appendix

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

Appendix

Amber atomtypes and charges assigned to the ligands in Chapters VI and VII



Tetraguanidinium (Chapter VI)

Atom	Atomtype	Charge
1	amber94_25	0.459245
2	amber94_43	-0.624955
3	amber94_11	0.034918
4	amber94_19	0.066985
5	amber94_72	-0.509523
6	amber94_73	0.437666
7	amber94_68	-0.053325
8	amber94_69	-0.028916
9	amber94_70	-0.033551
10	amber94_72	-0.509523
11	amber94_19	0.066985
12	amber94_70	-0.033551
13	amber94_69	-0.028916
14	amber94_76	0.352456
15	amber94_76	0.352456
16	amber94_19	0.066985
17	amber94_19	0.125850
18	amber94_18	0.075735
19	amber94_19	0.088600
20	amber94_19	0.088600
21	amber94_18	0.075735
22	amber94_19	0.125850
23	amber94_11	-0.262651
24	amber94_47	-0.232367
25	amber94_11	-0.262651
26	amber94_19	0.162944
27	amber94_19	0.162944
28	amber94_71	0.111760
29	amber94_72	-0.509523
30	amber94_73	0.437666
31	amber94_68	-0.053325
32	amber94_69	-0.028916
33	amber94_70	-0.033551
34	amber94_72	-0.509523
35	amber94_71	0.111760
36	amber94_70	-0.033551
37	amber94_69	-0.028916

ISBN:978-84-691-8850-7/DL:T-1276-2008

38	amber94_76	0.352456
39	amber94_76	0.352456
40	amber94_19	0.125850
41	amber94_18	0.075735
42	amber94_19	0.088600
43	amber94_19	0.088600
44	amber94_18	0.075735
45	amber94_19	0.125850
46	amber94_11	-0.262651
47	amber94_47	-0.232367
48	amber94_11	-0.262651
49	amber94_19	0.162944
50	amber94_76	0.352456

Hexaguanidinium (Chapter VI)

Atom	Atomtype	Charge
1	amber94_25	0.477149
2	amber94_43	-0.651953
3	amber94_11	0.054137
4	amber94_71	0.092340
5	amber94_72	-0.483468
6	amber94_73	0.382835
7	amber94_68	-0.040140
8	amber94_69	-0.027761
9	amber94_70	-0.009741
10	amber94_72	-0.483468
11	amber94_71	0.092340
12	amber94_70	-0.009741
13	amber94_69	-0.027761
14	amber94_76	0.347905
15	amber94_76	0.347905
16	amber94_19	0.063251
17	amber94_19	0.129726
18	amber94_19	0.069937
19	amber94_19	0.086395
20	amber94_19	0.086395
21	amber94_19	0.069937
22	amber94_19	0.129726
23	amber94_11	-0.231497
24	amber94_47	-0.228611
25	amber94_11	-0.231497
26	amber94_76	0.347905
27	amber94_76	0.347905
28	amber94_71	0.092340
29	amber94_72	-0.483468
30	amber94_73	0.382835
31	amber94_68	-0.040140
32	amber94_69	-0.027761
33	amber94_70	-0.009741
34	amber94_72	-0.483468
35	amber94_71	0.092340
36	amber94_70	-0.009741
37	amber94_69	-0.027761

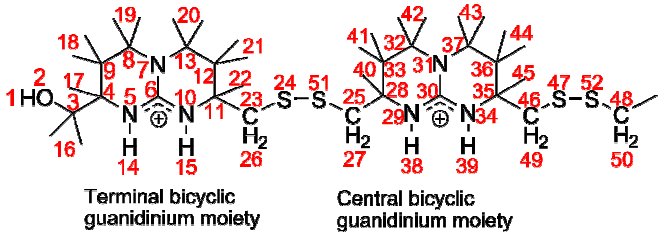
38	amber94_76	0.347905
39	amber94_76	0.347905
40	amber94_19	0.129726
41	amber94_19	0.069937
42	amber94_19	0.086395
43	amber94_19	0.086395
44	amber94_19	0.069937
45	amber94_19	0.129726
46	amber94_11	-0.231497
47	amber94_47	-0.228611
48	amber94_11	-0.231497
49	amber94_19	0.152510
50	amber94_19	0.152510

Undecaguanidinium (1 and 2 in Chapter VII)

Atom	**Atomtype**	**Charge**
1	amber94_25	0.474400
2	amber94_43	-0.654700
3	amber94_11	0.051400
4	amber94_71	0.089600
5	amber94_72	-0.486200
6	amber94_73	0.380100
7	amber94_68	-0.042800
8	amber94_69	-0.030500
9	amber94_70	-0.012400
10	amber94_72	-0.486200
11	amber94_71	0.089600
12	amber94_70	-0.012400
13	amber94_69	-0.030500
14	amber94_76	0.345300
15	amber94_76	0.345300
16	amber94_19	0.060700
17	amber94_19	0.127100
18	amber94_19	0.067300
19	amber94_19	0.083800
20	amber94_19	0.083800
21	amber94_19	0.067300
22	amber94_19	0.127100
23	amber94_11	-0.219900
24	amber94_47	-0.217000
25	amber94_11	-0.219900
26	amber94_19	0.164200
27	amber94_19	0.164200
28	amber94_71	0.089300
29	amber94_72	-0.486500
30	amber94_73	0.379900
31	amber94_68	-0.043100
32	amber94_69	-0.030800
33	amber94_70	-0.012600
34	amber94_72	-0.486500

ISBN:978-84-691-8850-7/DL:T-1276-2008

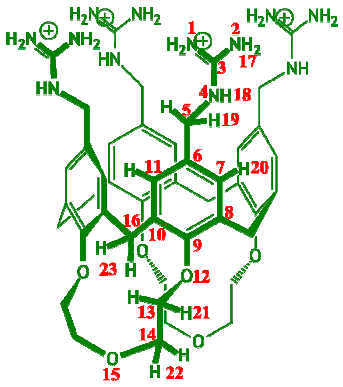
35	amber94_71	0.089300
36	amber94_70	-0.012600
37	amber94_69	-0.030800
38	amber94_76	0.344900
39	amber94_76	0.344900
40	amber94_19	0.126700
41	amber94_19	0.066900
42	amber94_19	0.083400
43	amber94_19	0.083400
44	amber94_19	0.066900
45	amber94_19	0.126700
46	amber94_11	-0.219900
47	amber94_47	-0.217000
48	amber94_11	-0.219900
49	amber94_19	0.164200
50	amber94_19	0.164200



Undecaguanidinium (3 in Chapter VII)

Atom	Atomtype	Charge
1	amber94_25	0.474400
2	amber94_43	-0.654700
3	amber94_11	0.051400
4	amber94_71	0.089600
5	amber94_72	-0.486200
6	amber94_73	0.380100
7	amber94_68	-0.042800
8	amber94_69	-0.030500
9	amber94_70	-0.012400
10	amber94_72	-0.486200
11	amber94_71	0.089600
12	amber94_70	-0.012400
13	amber94_69	-0.030500
14	amber94_76	0.345300
15	amber94_76	0.345300
16	amber94_19	0.060700
17	amber94_19	0.127100
18	amber94_19	0.067300
19	amber94_19	0.083800
20	amber94_19	0.083800
21	amber94_19	0.067300
22	amber94_19	0.127100
23	amber94_11	-0.006000
24	amber94_47	-0.180600

25	amber94_11	-0.006000
26	amber94_19	0.093300
27	amber94_19	0.093300
28	amber94_71	0.089300
29	amber94_72	-0.486500
30	amber94_73	0.379900
31	amber94_68	-0.043100
32	amber94_69	-0.030800
33	amber94_70	-0.012600
34	amber94_72	-0.486500
35	amber94_71	0.089300
36	amber94_70	-0.012600
37	amber94_68	-0.043100
38	amber94_76	0.344900
39	amber94_76	0.344900
40	amber94_19	0.126700
41	amber94_19	0.066900
42	amber94_19	0.083400
43	amber94_19	0.083400
44	amber94_19	0.066900
45	amber94_19	0.126700
46	amber94_11	-0.006000
47	amber94_47	-0.180600
48	amber94_11	-0.006000
49	amber94_19	0.093300
50	amber94_19	0.093300
51	amber94_47	-0.180600
52	amber94_47	-0.180600

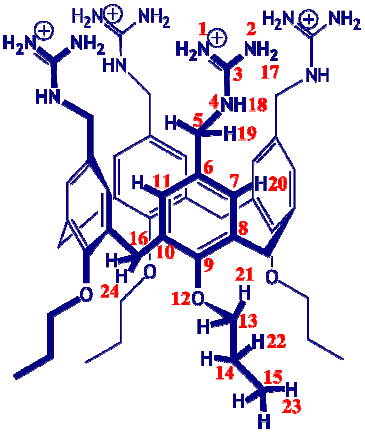


CalixBridge

Atom	Atomtype	Charge
1	amber94_72	-0.876075
2	amber94_72	-0.876075
3	amber94_73	0.787906
4	amber94_68	-0.447133
5	amber94_11	-0.022129
6	amber94_3	-0.005553

ISBN:978-84-691-8850-7/DL:T-1276-2008

7	amber94_3	-0.236344
8	amber94_3	0.027615
9	amber94_2	0.138216
10	amber94_3	0.003065
11	amber94_3	-0.217282
12	amber94_44	-0.225913
13	amber94_11	-0.012167
14	amber94_11	0.038280
15	amber94_44	-0.354283
16	amber94_11	-0.038354
17	amber94_76	0.454753
18	amber94_17	0.327986
19	amber94_19	0.100876
20	amber94_22	0.162149
21	amber94_19	0.100940
22	amber94_19	0.085174
23	amber94_18	0.049430



CalixProp

Atom	Atomtype	Charge
1	amber94_72	-0.876075
2	amber94_72	-0.876075
3	amber94_73	0.787906
4	amber94_68	-0.447133
5	amber94_11	-0.022129
6	amber94_3	-0.005553
7	amber94_3	-0.236344
8	amber94_3	0.027615
9	amber94_2	0.138216
10	amber94_3	0.003065
11	amber94_3	-0.217282
12	amber94_44	-0.225913
13	amber94_11	-0.012167
14	amber94_11	-0.043000
15	amber94_11	-0.066000
16	amber94_11	-0.038354

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

17	amber94_76	0.454753
18	amber94_17	0.327986
19	amber94_19	0.100876
20	amber94_22	0.162149
21	amber94_19	0.100940
22	amber94_18	0.023600
23	amber94_18	0.018600
24	amber94_18	0.049430

UNIVERSITAT ROVIRA I VIRGILI
COMPUTATIONAL STUDIES ON SUPRAMOLECULAR HYDROGEN-BONDED STRUCTURES:
FROM NANOCAPSULES TO PROTEINS
Eva Santos Garcia
ISBN:978-84-691-8850-7/DL:T-1276-2008

References

[1] L. Pauling, *The Nature of the Chemical Bond*, New York, **1939**, p.
[2] M. L. Huggins, *Chem. Rev.* **1943**, *32*, 195-218.
[3] J. D. Watson and F. H. C. Crick, *Nature* **1953**, *171*, 737-738.
[4] I. Alkorta, I. Rozas and J. Elguero, *Chem. Soc. Rev.* **1998**, *27*, 163-170.
[5] T. Steiner, *J Chem Soc Perk T 2* **1995**, 1315-1319.
[6] T. Steiner, *J Phys Chem A* **1998**, *102*, 7041-7052.
[7] T. Steiner, *Acta Crystallographica B* **1998**, *54*, 464-470.
[8] G. A. Jeffrey, *An introduction to hydrogen bonding*, **1997**, p.
[9] R. M. Badger and S. H. Bauer, *J Chem Phys* **1937**, *5*, 839-851.
[10] A. C. Legon and D. J. Millen, *Chem. Rev.* **1986**, *86*, 635-657.
[11] Desiraju, *Crystal engineering. The design of organic solids*, Amsterdam, **1989**, p.
[12] J. P. Glusker, M. Lewis and M. Rossi, *Crystal structure analysis for chemists and biologists*, New York, **1989**, p.
[13] F. H. Allen, J. E. Davies, J. J. Galloy, O. Johnson, O. Kennard, C. F. Macrae, E. M. Mitchell, G. F. Mitchell, J. M. Smith and D. G. Watson, *J Chem Inf Comp Sci* **1991**, *31*, 187-204.
[14] E. E. Abola, J. L. Sussman, J. Prilusky and N. O. Manning, *Method Enzymol* **1997**, *277*, 556-571.
[15] H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S. H. Hsieh, A. R. Srinivasan and B. Schneider, *Biophys J* **1992**, *63*, 751-759.
[16] G. Bergerhoff, R. Hundt, R. Sievers and I. D. Brown, *J Chem Inf Comp Sci* **1983**, *23*, 66-69.
[17] M. D. Joesten and L. J. Schaad, *Hydrogen Bonding*, New York, **1974**, p.
[18] U. Koch and P. L. A. Popelier, *J Phys Chem-Us* **1995**, *99*, 9747-9754.
[19] K. Morokuma, *J Chem Phys* **1971**, *55*, 1236-&.
[20] Y. F. Chen and J. J. Dannenberg, *J Am Chem Soc* **2006**, *128*, 8100-8101.
[21] I. Visiers, H. Weinstein, G. Rudnick and M. M. Stephan, *Biochemistry* **2003**, *42*, 6784-6793.
[22] E. L. DiGiammarino, A. S. Lee, C. Cadwell, W. X. Zhang, B. Bothner, R. C. Ribeiro, G. Zambetti and R. W. Kriwacki, *Nat Struct Biol* **2002**, *9*, 12-16.
[23] M. D. Sutton and G. C. Walker, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8342-8349.
[24] A. L. Gnatt, P. Cramer, J. H. Fu, D. A. Bushnell and R. D. Kornberg, *Science* **2001**, *292*, 1876-1882.
[25] H. Boeger, D. A. Bushnell, R. Davis, J. Griesenbeck, Y. Lorch, J. S. Strattan, K. D. Westover and R. D. Kornberg, *Febs Lett* **2005**, *579*, 899-903.
[26] M. V. Rodnina, M. Beringer and W. Wintermeyer, *Trends Biochem Sci* **2007**, *32*, 20-26.
[27] H. Luecke and F. A. Quiocho, *Nature* **1990**, *347*, 402-406.
[28] M. A. Massiah, C. Viragh, P. M. Reddy, I. M. Kovach, J. Johnson, T. L. Rosenberry and A. S. Mildvan, *Biochemistry* **2001**, *40*, 5682-5690.
[29] D. S. Lawrence, T. Jiang and M. Levett, *Chem. Rev.* **1995**, *95*, 2229-2260.
[30] J. Yang, J. L. Marendaz, S. J. Geib and A. D. Hamilton, *Tetrahedron Lett.* **1994**, *35*, 3665-3668.
[31] J. A. Zerkowski, J. C. Macdonald, C. T. Seto, D. A. Wierda and G. M. Whitesides, *J Am Chem Soc* **1994**, *116*, 2382-2391.
[32] J. A. Zerkowski, C. T. Seto and G. M. Whitesides, *J Am Chem Soc* **1992**, *114*, 5473-5475.
[33] I. Tabushi, Y. Kuroda and K. Yokota, *Tetrahedron Lett.* **1982**, *23*, 4601-4604.
[34] T. Gulikkrzywicki, C. Fouquey and J. M. Lehn, *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 163-167.
[35] K. D. Shimizu and J. Rebek, *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 12403-12407.
[36] I. N. Levine, *Quantum Chemistry*, **2000**, p. 739.
[37] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry. Introduction to Advanced Electronic Structure Theory*, New York, **1996**, p.
[38] W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Weinheim, **2001**, p.
[39] P. Hohenberg and W. Kohn, *Phys Rev B* **1964**, *136*, B864-&.
[40] W. Kohn and L. J. Sham, *Phys Rev* **1965**, *140*, 1133-&.
[41] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, New York, **1989**, p.
[42] J. R. Smith, *Phys Rev Lett* **1970**, *25*, 1023-&.
[43] D. M. Ceperley and B. J. Alder, *Phys Rev Lett* **1980**, *45*, 566-569.
[44] S. H. Vosko, L. Wilk and M. Nusair, *Can J Phys* **1980**, *58*, 1200-1211.
[45] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098-3100.
[46] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822-8824.

[47] J. P. Perdew, *Phys. Rev. B* **1986**, *34*, 7406-7406.
[48] A. D. Becke, *J Chem Phys* **1993**, *98*, 1372-1377.
[49] C. Moller and M. S. Plesset, *Phys. Rev.* **1934**, *46*, 0618-0622.
[50] J. A. Pople, M. Headgordon and K. Raghavachari, *J Chem Phys* **1987**, *87*, 5968-5975.
[51] R. J. Bartlett, *J Phys Chem-Us* **1989**, *93*, 1697-1708.
[52] A. D. Becke, *J Chem Phys* **1993**, *98*, 5648-5652.
[53] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J Phys Chem-Us* **1994**, *98*, 11623-11627.
[54] X. Xu and W. A. Goddard, *J. Phys. Chem. A* **2004**, *108*, 2305-2313.
[55] M. P. Waller, A. Robertazzi, J. A. Platts, D. E. Hibbs and P. A. Williams, *J Comput Chem* **2006**, *27*, 491-504.
[56] A. Robertazzi and J. A. Platts, *J. Phys. Chem. A* **2006**, *110*, 3992-4000.
[57] B. J. Lynch, P. L. Fast, M. Harris and D. G. Truhlar, *J. Phys. Chem. A* **2000**, *104*, 4811-4815.
[58] Y. Zhao and D. G. Truhlar, *Acc. Chem. Res.* **2008**, *41*, 157-167.
[59] C. T. Lee, W. T. Yang and R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785-789.
[60] H. J. Song, H. M. Xiao and H. S. Dong, *J Chem Phys* **2006**, *124*, -.
[61] X. Xu and W. A. Goddard, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673-2677.
[62] J. Cerny and P. Hobza, *Phys Chem Chem Phys* **2005**, *7*, 1624-1626.
[63] S. Grimme, *J. Comput. Chem.* **2004**, *25*, 1463-1473.
[64] J. M. Ducere and L. Cavallo, *J Phys Chem B* **2007**, *111*, 13124-13134.
[65] S. Grimme, C. Muck-Lichtenfeld and J. Antony, *J Phys Chem C* **2007**, *111*, 11199-11207.
[66] A. Bondi, *J Phys Chem-Us* **1964**, *68*, 441-&.
[67] K. J. Miller, *J Am Chem Soc* **1990**, *112*, 8543-8551.
[68] J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.
[69] A. D. Rabuck and G. E. Scuseria, *Theor Chem Acc* **2000**, *104*, 439-444.
[70] E. J. Baerends, D. E. Ellis and P. Ros, *Chem. Phys.* **1973**, *2*, 41-51.
[71] E. J. Baerends and P. Ros, *Chem. Phys.* **1973**, *2*, 52-59.
[72] E. J. Baerends, *Theoretical Chemistry*, Vrije Universiteit, Amsterdam, The Netherlands, **2004**, p.
[73] C. F. Guerra, J. G. Snijders, G. te Velde and E. J. Baerends, *Theor Chem Acc* **1998**, *99*, 391-403.
[74] P. Jurecka, J. Sponer, J. Cerny and P. Hobza, *Phys Chem Chem Phys* **2006**, *8*, 1985-1993.
[75] J. Sponer, P. Jurecka, I. Marchan, F. J. Luque, M. Orozco and P. Hobza, *Chem-Eur J* **2006**, *12*, 2854-2865.
[76] F. Jensen, *Introduction to Computational Chemistry*, **1999**, p.
[77] A. R. Leach, *Molecular Modelling: Principles and Applications*, **2001**, p.
[78] C. D. Sherrill, *Introduction to Molecular Mechanics*, **2005**, p.
[79] D. C. Young, *Computational Chemistry. A Practical Guide for Applying Techniques to Real-World Problems*, **2001**, p.
[80] A. D. Mackerell, *J. Comput. Chem.* **2004**, *25*, 1584-1604.
[81] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J Am Chem Soc* **1996**, *118*, 2309-2309.
[82] C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J Phys Chem-Us* **1993**, *97*, 10269-10280.
[83] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, New York, **1989**, p.
[84] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, San Diego, **1996**, p.
[85] T. Darden, D. York and L. Pedersen, *J Chem Phys* **1993**, *98*, 10089-10092.
[86] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans and B. Pullman, *Intermolecular Forces*, Reidel, Dordrecht, **1981**, p.
[87] G. W. Robinson, S. B. Zhu, S. S. and M. W. Evans, *Water in Biology, Chemistry and Physics; Experimental Overviews and Computational Methodologies*, Singapore, **1996**, p.
[88] M. E. Tuckerman and G. J. Martyna, *J Phys Chem B* **2000**, *104*, 159-178.
[89] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen and H. J. C. Berendsen, *Gromacs User Manual. Version 3.3*, **2006**, p.
[90] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J Comput Chem* **2005**, *26*, 1701-1718.
[91] V. Daggett and M. Levitt, *Curr Opin Struc Biol* **1994**, *4*, 291-295.
[92] V. Daggett, *Curr Opin Struc Biol* **2000**, *10*, 160-164.
[93] C. L. Brooks, *Curr Opin Struc Biol* **1998**, *8*, 222-226.
[94] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, *Science* **1983**, *220*, 671-680.

[95] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *J Chem Phys* **1984**, *81*, 3684-3690.
[96] X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, *J Mol Biol* **1998**, *280*, 925-932.
[97] M. Eleftheriou, R. S. Germain, A. K. Royyuru and R. H. Zhou, *J Am Chem Soc* **2006**, *128*, 13388-13395.
[98] J. Klein-Seetharaman, M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson and H. Schwalbe, *Science* **2002**, *295*, 1719-1722.
[99] M. Dumoulin, A. M. Last, A. Desmyter, K. Decanniere, D. Canet, G. Larsson, A. Spencer, D. B. Archer, J. Sasse, S. Muyldermans, L. Wyns, C. Redfield, A. Matagne, C. V. Robinson and C. M. Dobson, *Nature* **2003**, *424*, 783-788.
[100] N. Ferguson and A. R. Fersht, *Curr Opin Struc Biol* **2003**, *13*, 75-81.
[101] Y. Duan and P. A. Kollman, *Science* **1998**, *282*, 740-744.
[102] S. Chowdhury, M. C. Lee, G. M. Xiong and Y. Duan, *J Mol Biol* **2003**, *327*, 711-717.
[103] S. Kmiecik and A. Kolinski, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 12330-12335.
[104] R. X. Wang, L. H. Lai and S. M. Wang, *J Comput Aid Mol Des* **2002**, *16*, 11-26.
[105] P. Ferrara, H. Gohlke, D. J. Price, G. Klebe and C. L. Brooks, *J Med Chem* **2004**, *47*, 3032-3047.
[106] C. Obiol-Pardo and J. Rubio-Martinez, *J. Chem. Inf. Model.* **2007**, *47*, 134-142.
[107] W. Humphrey, A. Dalke and K. Schulten, *J Mol Graphics* **1996**, *14*, 33-&.
[108] A. Muller, M. Losada and S. Leutwyler, *J. Phys. Chem. A* **2004**, *108*, 157-165.
[109] A. E. Shchavlev, A. N. Pankratov, V. B. Borodulin and O. A. Chaplygina, *J. Phys. Chem. A* **2005**, *109*, 10982-10996.
[110] K. Adachi, Y. Sugiyama, K. Yoneda, K. Yamada, K. Nozaki, A. Fuyuhiro and S. Kawata, *Chem-Eur J* **2005**, *11*, 6616-6628.
[111] R. P. Sijbesma and E. W. Meijer, *Chem. Commun.* **2003**, 5-16.
[112] P. S. Corbin and S. C. Zimmerman, *J Am Chem Soc* **1998**, *120*, 9710-9711.
[113] U. Luning, C. Kuhl and A. Uphoff, *Eur J Org Chem* **2002**, 4063-4070.
[114] F. H. Beijer, H. Kooijman, A. L. Spek, R. P. Sijbesma and E. W. Meijer, *Angew Chem Int Edit* **1998**, *37*, 75-78.
[115] G. B. W. L. Ligthart, H. Ohkawa, R. P. Sijbesma and E. W. Meijer, *J Am Chem Soc* **2005**, *127*, 810-811.
[116] O. A. Scherman, G. B. W. L. Ligthart, R. P. Sijbesma and E. W. Meijer, *Angew Chem Int Edit* **2006**, *45*, 2072-2076.
[117] H. Q. Zeng, R. S. Miller, R. A. Flowers and B. Gong, *J Am Chem Soc* **2000**, *122*, 2635-2644.
[118] E. D. Raczynska, W. Kosinska, B. Osmialowski and R. Gawinecki, *Chem. Rev.* **2005**, *105*, 3561-3612.
[119] J. J. Gonzalez, P. Prados and J. de Mendoza, *Angew Chem Int Edit* **1999**, *38*, 525-528.
[120] X. Q. Li, D. J. Feng, X. K. Jiang and Z. T. Li, *Tetrahedron* **2004**, *60*, 8275-8284.
[121] L. Sanchez, N. Martin and D. M. Guldi, *Angew Chem Int Edit* **2005**, *44*, 5374-5382.
[122] H. Sun, J. Steeb and A. E. Kaifer, *J Am Chem Soc* **2006**, *128*, 2820-2821.
[123] F. H. Beijer, R. P. Sijbesma, H. Kooijman, A. L. Spek and E. W. Meijer, *J Am Chem Soc* **1998**, *120*, 6761-6769.
[124] W. L. Jorgensen and J. Pranata, *J Am Chem Soc* **1990**, *112*, 2008-2010.
[125] D. W. Guo, R. P. Sijbesma and H. Zuilhof, *Org Lett* **2004**, *6*, 3667-3670.
[126] B. J. B. Folmer, R. P. Sijbesma, H. Kooijman, A. L. Spek and E. W. Meijer, *J Am Chem Soc* **1999**, *121*, 9001-9007.
[127] A. El-ghayoury, E. Peeters, A. P. H. J. Schenning and E. W. Meijer, *Chem. Commun.* **2000**, 1969-1970.
[128] A. P. H. J. Schenning, P. Jonkheijm, E. Peeters and E. W. Meijer, *J Am Chem Soc* **2001**, *123*, 409-416.
[129] J. H. K. K. Hirschberg, L. Brunsveld, A. Ramzi, J. A. J. M. Vekemans, R. P. Sijbesma and E. W. Meijer, *Nature* **2000**, *407*, 167-170.
[130] S. H. M. Sontjens, R. P. Sijbesma, M. H. P. van Genderen and E. W. Meijer, *Macromolecules* **2001**, *34*, 3815-3818.
[131] H. M. Keizer, R. van Kessel, R. P. Sijbesma and E. W. Meijer, *Polymer* **2003**, *44*, 5505-5511.
[132] N. Robertson and C. A. McGowan, *Chem. Soc. Rev.* **2003**, *32*, 96-103.
[133] G. L. Xu, R. J. Crutchley, M. C. DeRosa, Q. J. Pan, H. X. Zhang, X. P. Wang and T. Ren, *J Am Chem Soc* **2005**, *127*, 13354-13363.
[134] J. J. Gonzalez, S. Gonzalez, E. M. Priego, C. P. Luo, D. M. Guldi, J. de Mendoza and N. Martin, *Chem. Commun.* **2001**, 163-164.
[135] L. Shi, X. W. Wang, C. A. Sandoval, M. X. Li, Q. Y. Qi, Z. T. Li and K. L. Ding, *Angew Chem Int Edit* **2006**, *45*, 4108-4112.
[136] P. M. Boerrigter, G. T. Velde and E. J. Baerends, *Int J Quantum Chem* **1988**, *33*, 87-113.
[137] G. T. Velde and E. J. Baerends, *J Comput Phys* **1992**, *99*, 84-98.
[138] L. Versluis and T. Ziegler, *J Chem Phys* **1988**, *88*, 322-328.

[139] S. F. Boys and F. Bernardi, *Mol. Phys.* **1970**, *19*, 553-&.
[140] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr.;, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukada, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian03, Revision C.02, Gaussian, Inc,Wallingford CT*, **2004**, p.
[141] B. G. Johnson and M. J. Fisch, *J Chem Phys* **1994**, *100*, 7429-7442.
[142] B. G. Johnson and M. J. Frisch, *Chem Phys Lett* **1993**, *216*, 133-140.
[143] R. E. Stratmann, J. C. Burant, G. E. Scuseria and M. J. Frisch, *J Chem Phys* **1997**, *106*, 10175-10183.
[144] C. F. Guerra, *Structure and Bonding of DNA. Development and Application of Parallel and Order-N DFT Methods*, Amsterdam, **2000**, p.
[145] M. Schulz-Dobrick, T. Metzroth, H. W. Spiess, J. Gauss and I. Schnell, *Chemphyschem* **2005**, *6*, 315-327.
[146] A. Schafer, H. Horn and R. Ahlrichs, *J Chem Phys* **1992**, *97*, 2571-2577.
[147] A. Schafer, C. Huber and R. Ahlrichs, *J Chem Phys* **1994**, *100*, 5829-5835.
[148] M. S. Dresselhaus, G. Dresselhaus and P. C. Eklund, *Science of fullerenes and carbon nanoutubes*, San Diego, **1996**, p. 1-985.
[149] A. Hirsch and M. Brettreich, *Fullerenes-Chemistry and Reactions*, Weinheim, **2005**, p.
[150] H. W. Kroto, J. R. Heath, S. C. Obrien, R. F. Curl and R. E. Smalley, *Nature* **1985**, *318*, 162-163.
[151] W. Kratschmer, L. D. Lamb, K. Fostiropoulos and D. R. Huffman, *Nature* **1990**, *347*, 354-358.
[152] J. Aihara, *J Am Chem Soc* **1995**, *117*, 4130-4136.
[153] H. W. Kroto, A. W. Allaf and S. P. Balm, *Chem. Rev.* **1991**, *91*, 1213-1235.
[154] E. Albertazzi, C. Domene, P. W. Fowler, T. Heine, G. Seifert, C. Van Alsenoy and F. Zerbetto, *Phys Chem Chem Phys* **1999**, *1*, 2913-2918.
[155] E. E. B. Campbell, P. W. Fowler, D. Mitchell and F. Zerbetto, *Chem Phys Lett* **1996**, *250*, 544-548.
[156] K. Kobayashi, S. Nagase, M. Yoshida and E. Osawa, *J Am Chem Soc* **1997**, *119*, 12693-12694.
[157] X. Lu and Z. F. Chen, *Chem. Rev.* **2005**, *105*, 3643-3696.
[158] H. W. Kroto, *Nature* **1987**, *329*, 529-531.
[159] P. W. Fowler and D. E. Manolopoulos, *An Atlas of Fullerenes*, Oxford, **1995**, p.
[160] K. M. Kadish and R. S. Ruoff, *Fullerene: Chemistry, Physics and Technology*, New York, **2002**, p.
[161] P. W. Fowler and T. Heine, *J Chem Soc Perk T 2* **2001**, 487-490.
[162] Z. Chen, *Angew Chem Int Edit* **2004**, *43*, 4690-4691.
[163] X. Lu, Z. F. Chen, W. Thiel, P. V. Schleyer, R. B. Huang and L. S. Zheng, *J Am Chem Soc* **2004**, *126*, 14871-14878.
[164] R. C. Haddon, *Accounts Chem Res* **1992**, *25*, 127-133.
[165] R. C. Haddon and L. T. Scott, *Pure Appl. Chem.* **1986**, *58*, 137-142.
[166] D. J. Klein, T. G. Schmalz, G. E. Hite and W. A. Seitz, *J Am Chem Soc* **1986**, *108*, 1301-1302.
[167] Y. Murata, M. Murata and K. Komatsu, *Chem-Eur J* **2003**, *9*, 1600-1609.
[168] K. Komatsu, M. Murata and Y. Murata, *Science* **2005**, *307*, 238-240.
[169] M. M. Roubelakis, G. C. Vougioukalakis and M. Orfanopoulos, *J Org Chem* **2007**, *72*, 6526-6533.
[170] S. H. Friedman, D. L. Decamp, R. P. Sijbesma, G. Srdanov, F. Wudl and G. L. Kenyon, *J Am Chem Soc* **1993**, *115*, 6506-6509.
[171] C. Toniolo, A. Bianco, M. Maggini, G. Scorrano, M. Prato, M. Marastoni, R. Tomatis, S. Spisani, G. Palu and E. D. Blair, *J. Med. Chem.* **1994**, *37*, 4558-4562.
[172] G. L. Marcorin, T. Da Ros, S. Castellano, G. Stefancich, I. Bonin, S. Miertus and M. Prato, *Org Lett* **2000**, *2*, 3955-3958.
[173] A. S. Boutorine, H. Tokuyama, M. Takasugi, H. Isobe, E. Nakamura and C. Helene, *Angew Chem Int Edit* **1995**, *33*, 2462-2465.
[174] H. Tokuyama, S. Yamago, E. Nakamura, T. Shiraki and Y. Sugiura, *J Am Chem Soc* **1993**, *115*, 7918-7919.
[175] Y. Yamakoshi, N. Umezawa, A. Ryu, K. Arakane, N. Miyata, Y. Goda, T. Masumizu and T. Nagano, *J Am Chem Soc* **2003**, *125*, 12803-12809.

[176] C. M. Sayes, J. D. Fortner, W. Guo, D. Lyon, A. M. Boyd, K. D. Ausman, Y. J. Tao, B. Sitharaman, L. J. Wilson, J. B. Hughes, J. L. West and V. L. Colvin, *Nano Lett.* **2004**, *4*, 1881-1887.
[177] P. Schilinsky, C. Waldauf and C. J. Brabec, *Appl Phys Lett* **2002**, *81*, 3885-3887.
[178] F. Padinger, R. S. Rittberger and N. S. Sariciftci, *Adv Funct Mater* **2003**, *13*, 85-88.
[179] M. M. Wienk, J. M. Kroon, W. J. H. Verhees, J. Knol, J. C. Hummelen, P. A. van Hal and R. A. J. Janssen, *Angew Chem Int Edit* **2003**, *42*, 3371-3375.
[180] T. Hasobe, H. Imahori, P. V. Kamat, T. K. Ahn, S. K. Kim, D. Kim, A. Fujimoto, T. Hirakawa and S. Fukuzumi, *J Am Chem Soc* **2005**, *127*, 1216-1228.
[181] S. Kobayashi, S. Mori, S. Iida, H. Ando, T. Takenobu, Y. Taguchi, A. Fujiwara, A. Taninaka, H. Shinohara and Y. Iwasa, *J Am Chem Soc* **2003**, *125*, 8116-8117.
[182] H. Kato, Y. Kanazawa, M. Okumura, A. Taninaka, T. Yokawa and H. Shinohara, *J Am Chem Soc* **2003**, *125*, 4391-4397.
[183] M. Saunders, R. J. Cross, H. A. JimenezVazquez, R. Shimshi and A. Khong, *Science* **1996**, *271*, 1693-1697.
[184] M. Saunders, H. A. Jimenezvazquez, R. J. Cross, S. Mroczkowski, D. I. Freedberg and F. A. L. Anet, *Nature* **1994**, *367*, 256-258.
[185] A. P. H. J. Schenning, J. von Herrikhuyzen, P. Jonkheijm, Z. Chen, F. Wurthner and E. W. Meijer, *J Am Chem Soc* **2002**, *124*, 10252-10253.
[186] D. M. Guldi and N. Martin, *J Mater Chem* **2002**, *12*, 1978-1992.
[187] J. L. Sessler, M. Sathiosatham, C. T. Brown, T. A. Rhodes and G. Wiederrecht, *J Am Chem Soc* **2001**, *123*, 3655-3660.
[188] A. J. Myles and N. R. Branda, *J Am Chem Soc* **2001**, *123*, 177-178.
[189] T. H. Ghaddar, E. W. Castner and S. S. Isied, *J Am Chem Soc* **2000**, *122*, 1233-1234.
[190] M. Segura, L. Sanchez, J. de Mendoza, N. Martin and D. M. Guldi, *J Am Chem Soc* **2003**, *125*, 15093-15100.
[191] K. Chatterjee, D. H. Parker, P. Wurz, K. R. Lykke, D. M. Gruen and L. M. Stock, *J Org Chem* **1992**, *57*, 3253-3254.
[192] K. C. Khemani, M. Prato and F. Wudl, *J Org Chem* **1992**, *57*, 3254-3256.
[193] W. A. Scrivens, P. V. Bedworth and J. M. Tour, *J Am Chem Soc* **1992**, *114*, 7917-7919.
[194] L. Isaacs, A. Wehrsig and F. Diederich, *Helv. Chim. Acta* **1993**, *76*, 1231-1250.
[195] N. Komatsu, T. Ohe and K. Matsushige, *Carbon* **2004**, *42*, 163-167.
[196] T. Andersson, K. Nilsson, M. Sundahl, G. Westman and O. Wennerstrom, *J Chem Soc Chem Comm* **1992**, 604-606.
[197] T. Suzuki, K. Nakashima and S. Shinkai, *Chem. Lett.* **1994**, 699-702.
[198] N. Komatsu, *Org Biomol Chem* **2003**, *1*, 204-209.
[199] T. Haino, C. Fukunaga and Y. Fukazawa, *Org Lett* **2006**, *8*, 3545-3548.
[200] Y. Shoji, K. Tashiro and T. Aida, *J Am Chem Soc* **2004**, *126*, 6570-6571.
[201] J. W. Steed, P. C. Junk, J. L. Atwood, M. J. Barnes, C. L. Raston and R. S. Burkhalter, *J Am Chem Soc* **1994**, *116*, 10346-10347.
[202] H. Matsubara, A. Hasegawa, K. Shiwaku, K. Asano, M. Uno, S. Takahashi and K. Yamamoto, *Chem. Lett.* **1998**, 923-924.
[203] H. Matsubara, S. Oguri, K. Asano and K. Yamamoto, *Chem. Lett.* **1999**, 431-432.
[204] D. Felder, B. Heinrich, D. Guillon, J. F. Nicoud and J. F. Nierengarten, *Chem-Eur J* **2000**, *6*, 3501-3507.
[205] R. P. Sijbesma, F. H. Beijer, L. Brunsveld, B. J. B. Folmer, J. H. K. K. Hirschberg, R. F. M. Lange, J. K. L. Lowe and E. W. Meijer, *Science* **1997**, *278*, 1601-1604.
[206] S. H. M. Sontjens, R. P. Sijbesma, M. H. P. van Genderen and E. W. Meijer, *J Am Chem Soc* **2000**, *122*, 7487-7493.
[207] E. J. Bylaska, W. A. d. Jong, K. Kowalski, T. P. Straatsma, M. Valiev, D. Wang, E. Aprà, T. L. Windus, S. Hirata, M. T. Hackler, Y. Zhao, P.-D. Fan, R. J. Harrison, M. Dupuis, D. M. A. Smith, J. Nieplocha, V. Tipparaju, M. Krishnan, A. A. Auer, M. Nooijen, E. Brown, G. Cisneros, G. I. Fann, H. Frücht, J. Garza, K. Hirao, R. Kendall, J. A. Nichols, K. Tsemekhman, K. Wolinski, J. Anchell, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Deegan, K. Dyall, D. Elwood, E. Glendening, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, T. Nakajima, S. Niu, L. Pollack, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. v. Lenthe, A. Wong and Z. Zhang, *NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.0*, Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA, **2006**, p.
[208] R. A. Kendall, E. Apra, D. E. Bernholdt, E. J. Bylaska, M. Dupuis, G. I. Fann, R. J. Harrison, J. L. Ju, J. A. Nichols, J. Nieplocha, T. P. Straatsma, T. L. Windus and A. T. Wong, *Comput. Phys. Commun.* **2000**, *128*, 260-283.

[209] E. M. Perez, M. Sierra, L. Sanchez, M. R. Torres, R. Viruela, P. M. Viruela, E. Orti and N. Martin, *Angew Chem Int Edit* **2007**, *46*, 1847-1851.
[210] E. Huerta, G. A. Metselaar, A. Fragoso, E. Santos, C. Bo and J. de Mendoza, *Angew Chem Int Edit* **2007**, *46*, 202-205.
[211] E. Huerta, E. Cequier and J. de Mendoza, *Chem. Commun.* **2007**, 5016-5018.
[212] A. J. Levine, J. Momand and C. A. Finlay, *Nature* **1991**, *351*, 453-456.
[213] M. Hollstein, D. Sidransky, B. Vogelstein and C. C. Harris, *Science* **1991**, *253*, 49-53.
[214] P. Hainaut and K. G. Wiman, *25 Years of P53 Research*, Dordrecht, The Netherlands, **2007**, p.
[215] A. B. Deleo, G. Jay, E. Appella, G. C. Dubois, L. W. Law and L. J. Old, *P Natl Acad Sci USA* **1979**, *76*, 2420-2424.
[216] A. J. Levine, C. A. Finlay and P. W. Hinds, *Cell* **2004**, *S116*, S67-S69.
[217] C. A. Finlay, P. W. Hinds and A. J. Levine, *Cell* **1989**, *57*, 1083-1093.
[218] S. J. Baker, E. R. Fearon, J. M. Nigro, S. R. Hamilton, A. C. Preisinger, J. M. Jessup, P. Vantuinen, D. H. Ledbetter, D. F. Barker, Y. Nakamura, R. White and B. Vogelstein, *Science* **1989**, *244*, 217-221.
[219] T. Soussi and K. G. Wiman, *Cancer Cell* **2007**, *12*, 303-312.
[220] C. Miller, T. Mohandas, D. Wolf, M. Prokocimer, V. Rotter and H. P. Koeffler, *Nature* **1986**, *319*, 783-784.
[221] M. F. Lavin and N. Gueven, *Cell Death Diff.* **2006**, *13*, 941-950.
[222] Z. A. Stewart and J. A. Pietenpol, *Chem. Res. Toxicol.* **2001**, *14*, 243-263.
[223] L. E. Giono and J. J. Manfredi, *J Cell Physiol* **2006**, *209*, 13-20.
[224] S. Sengupta and C. C. Harris, *Nat Rev Mol Cell Bio* **2005**, *6*, 44-55.
[225] A. Sancar, L. A. Lindsey-Boltz, K. Unsal-Kacmaz and S. Linn, *Annu Rev Biochem* **2004**, *73*, 39-85.
[226] E. Meulmeester and A. G. Jochemsen, *Curr. Cancer Drug Targets* **2008**, *8*, 87-97.
[227] R. H. Daniels and G. M. Bokoch, *Trends Biochem. Sci.* **1999**, *24*, 350-355.
[228] S. X. Liu, W. R. Bishop and M. Liu, *Drug Resist Update* **2003**, *6*, 183-195.
[229] C. Han, A. J. Demetris, G. K. Michalopoulos, Q. M. Zhan, J. H. Shelhamer and T. Wu, *Hepatology* **2003**, *38*, 167-177.
[230] E. S. Child and D. J. Mann, *Cell Cycle* **2006**, *5*, 1313-1319.
[231] J. G. Jackson and O. M. Pereira-Smith, *Cancer Res.* **2006**, *66*, 8356-8360.
[232] A. Benzinger, N. Muster, H. B. Koch, J. R. Yates and H. Hermeking, *Molecular & Cellular Proteomics* **2005**, *4*, 785-795.
[233] H. L. Yang, R. Y. Zhao and M. H. Lee, *Mol. Cancer Ther.* **2006**, *5*, 253-260.
[234] B. Vogelstein, D. Lane and A. J. Levine, *Nature* **2000**, *408*, 307-310.
[235] M. Malumbres and M. Barbacid, *Trends Biochem. Sci.* **2005**, *30*, 630-641.
[236] R. M. Golsteyn, *Cancer Lett.* **2005**, *217*, 129-138.
[237] G. Draetta, *Trends Biochem. Sci.* **1990**, *15*, 378-383.
[238] X. W. Wu and Y. B. Deng, *Front Biosci* **2002**, *7*, D151-D156.
[239] A. Fortin, S. P. Cregan, J. G. MacLaurin, N. Kushwaha, E. S. Hickman, C. S. Thompson, A. Hakim, P. R. Albert, F. Cecconi, K. Helin, D. S. Park and R. S. Slack, *J. Cell Biol.* **2001**, *155*, 207-216.
[240] O. Ashur-Fabian, K. Adamsky, L. Trakhtenbrot, Y. Cohen, P. Raanani, I. Hardan, A. Nagler, G. Rechavi and N. Amariglio, *Cell Cycle* **2007**, *6*, 589-594.
[241] S. Kumar, *Cell Death Diff.* **2007**, *14*, 32-43.
[242] K. Itahana, G. Dimri and J. Campisi, *Eur. J. Biochem.* **2001**, *268*, 2784-2791.
[243] W. W. Hu, Z. H. Feng, A. K. Teresky and A. J. Levine, *Nature* **2007**, *450*, 721-U728.
[244] C. C. Harris, *J Natl Cancer Inst* **1996**, *88*, 1442-1455.
[245] H. Tidow, R. Melero, E. Mylonas, S. M. V. Freund, J. G. Grossmann, J. M. Carazo, D. I. Svergun, M. Valle and A. R. Fersht, *P Natl Acad Sci USA* **2007**, *104*, 12324-12329.
[246] E. Bochkareva, L. Kaustov, A. Ayed, G. S. Yi, Y. Lu, A. Pineda-Lucena, J. C. C. Liao, A. L. Okorokov, J. Milner, C. H. Arrowsmith and A. Bochkarev, *P Natl Acad Sci USA* **2005**, *102*, 15412-15417.
[247] V. M. Golubovskaya, R. Finch and W. G. Cance, *J Biol Chem* **2005**, *280*, 25008-25021.
[248] N. Baptiste, P. Friedlander, X. B. Chen and C. Prives, *Oncogene* **2002**, *21*, 9-21.
[249] F. Toledo, K. A. Krummel, C. J. Lee, C. W. Liu, L. W. Rodewald, M. J. Tang and G. M. Wahl, *Cancer Cell* **2006**, *9*, 273-285.
[250] F. Toledo, C. J. Lee, K. A. Krummel, L. W. Rodewald, C. W. Liu and G. M. Wahl, *Mol. Cell. Biol.* **2007**, *27*, 1425-1432.
[251] N. P. Pavletich, K. A. Chambers and C. O. Pabo, *Genes Dev.* **1993**, *7*, 2556-2564.
[252] J. L. F. Waterman, J. L. Shenk and T. D. Halazonetis, *Embo J.* **1995**, *14*, 512-519.
[253] P. Chene, *Oncogene* **2001**, *20*, 2611-2617.
[254] W. Gu and R. G. Roeder, *Cell* **1997**, *90*, 595-606.

[255] R. Wolkowicz and V. Rotter, *Pathol. Biol.* **1997**, *45*, 785-796.
[256] J. M. P. Canadillas, H. Tidow, S. M. V. Freund, T. J. Rutherford, H. C. Ang and A. R. Fersht, *P Natl Acad Sci USA* **2006**, *103*, 2109-2114.
[257] Y. J. Cho, S. Gorina, P. D. Jeffrey and N. P. Pavletich, *Science* **1994**, *265*, 346-355.
[258] M. Kitayner, H. Rozenberg, N. Kessler, D. Rabinovich, L. Shaulov, T. E. Haran and Z. Shakked, *Mol Cell* **2006**, *22*, 741-753.
[259] G. M. Clore, J. G. Omichinski, K. Sakaguchi, N. Zambrano, H. Sakamoto, E. Appella and A. M. Gronenborn, *Science* **1995**, *267*, 1515-1516.
[260] P. D. Jeffrey, S. Gorina and N. P. Pavletich, *Science* **1995**, *267*, 1498-1502.
[261] W. Lee, T. Harvey, Y. Yin, P. Yau, D. Litchfield and C. H. Arrowsmith, *Nat Struct Biol* **1994**, *1*, 877-890.
[262] T. Iwakuma, G. Lozano and E. R. Flores, *Cell Cycle* **2005**, *4*, 865-867.
[263] P. W. Fei and W. S. El-Deiry, *Oncogene* **2003**, *22*, 5774-5783.
[264] W. P. Bennett, S. P. Hussain, K. H. Vahakangas, M. A. Khan, P. G. Shields and C. C. Harris, *J Pathol* **1999**, *187*, 8-18.
[265] K. Viktorsson, L. De Petris and R. Lewensohn, *Biochem. Biophys. Res. Commun.* **2005**, *331*, 868-880.
[266] H. M. Shen and C. N. Ong, *Mutat Res-Rev Genet* **1996**, *366*, 23-44.
[267] S. Suzuki, Y. Muroishi, I. Nakanishi and Y. Oda, *J Gastroenterol* **2004**, *39*, 220-230.
[268] B. J. Boersma, T. M. Howe, J. E. Goodman, H. G. Yfantis, D. H. Lee, S. J. Chanock and S. Ambs, *J Natl Cancer Inst* **2006**, *98*, 911-919.
[269] A. J. Munro, S. Lain and D. P. Lane, *Br. J. Cancer* **2005**, *92*, 434-444.
[270] S. R. Downing, P. Jackson and P. J. Russell, *Urol Oncol* **2001**, *6*, 103-110.
[271] J. L. Wang, B. Y. Zheng, X. D. Li, T. Angstrom, M. S. Lindstrom and K. L. Wallin, *Clin Cancer Res* **2004**, *10*, 2407-2414.
[272] D. A. Whyte, C. E. Broton and E. J. Shillitoe, *J Oral Pathol Med* **2002**, *31*, 125-133.
[273] E. Schmitt, A. Steyaert, G. Cimoli and R. Bertrand, *Cell Death Diff.* **1998**, *5*, 506-516.
[274] A. Ichikawa, *Int J Hematol* **2000**, *71*, 211-220.
[275] A. R. Pettitt, E. Matutes and D. Oscier, *Leukemia* **2006**, *20*, 1441-1445.
[276] J. B. Gebhart, P. C. Roche, G. L. Keeney, T. G. Lesnick and K. C. Podratz, *Gynecol Oncol* **2000**, *77*, 232-236.
[277] N. D. Smith, J. N. Rubenstein, S. E. Eggener and J. M. Kozlowski, *J Urology* **2003**, *169*, 1219-1228.
[278] R. Talar-Wojnarowska, A. Gasiorowska, B. Smolarz, H. Romanowicz-Makowska, J. Strzelczyk, A. Juniak and E. Malecka-Panas, *Hepato-Gastroenterol* **2006**, *53*, 608-612.
[279] M. Scheffner, B. A. Werness, J. M. Huibregtse, A. J. Levine and P. M. Howley, *Cell* **1990**, *63*, 1129-1136.
[280] T. Crook, D. Wrede, J. A. Tidy, W. P. Mason, D. J. Evans and K. H. Vousden, *Lancet* **1992**, *339*, 1070-1073.
[281] U. M. Moll, M. Laquaglia, J. Benard and G. Riou, *P Natl Acad Sci USA* **1995**, *92*, 4407-4411.
[282] K. Onel and C. Cordon-Cardo, *Mol Cancer Res* **2004**, *2*, 1-8.
[283] E. Liberzon, S. Avigad, B. Stark, J. Zilberstein, L. Freedman, M. Gorfine, H. Gavriel, I. J. Cohen, Y. Goshen, I. Yaniv and R. Zaizov, *Gene Chromosome Canc* **2004**, *39*, 161-166.
[284] M. V. Blagosklonny, *Int J Cancer* **2002**, *98*, 161-166.
[285] W. G. Nelson and M. B. Kastan, *Mol. Cell. Biol.* **1994**, *14*, 1815-1823.
[286] Y. Haupt, R. Maya, A. Kazaz and M. Oren, *Nature* **1997**, *387*, 296-299.
[287] M. H. G. Kubbutat, S. N. Jones and K. H. Vousden, *Nature* **1997**, *387*, 299-303.
[288] C. A. Midgley and D. P. Lane, *Oncogene* **1997**, *15*, 1179-1189.
[289] C. L. Brooks and W. Gu, *Mol Cell* **2006**, *21*, 307-315.
[290] M. H. Lee and G. Lozano, *Sem. Cancer Biol.* **2006**, *16*, 225-234.
[291] S. P. Linke, K. C. Clarkin, A. DiLeonardo, A. Tsou and G. M. Wahl, *Genes Dev.* **1996**, *10*, 934-947.
[292] D. M. Pritchard, A. J. M. Watson, C. S. Potten, A. L. Jackman and J. A. Hickman, *P Natl Acad Sci USA* **1997**, *94*, 1795-1799.
[293] F. Bunz, P. M. Hwang, C. Torrance, T. Waldman, Y. G. Zhang, L. Dillehay, J. Williams, C. Lengauer, K. W. Kinzler and B. Vogelstein, *J Clin Invest* **1999**, *104*, 263-269.
[294] D. A. Freedman and A. J. Levine, *Mol. Cell. Biol.* **1998**, *18*, 7288-7293.
[295] K. H. Vousden and G. F. Vande Woude, *Nat Cell Biol* **2000**, *2*, E178-E180.
[296] P. Smart, E. B. Lane, D. P. Lane, C. Midgley, B. Vojtesek and S. Lain, *Oncogene* **1999**, *18*, 7378-7386.
[297] M. V. Blagosklonny, J. Toretsky and L. Neckers, *Oncogene* **1995**, *11*, 933-939.
[298] W. G. An, R. C. Schnur, L. Neckers and M. V. Blagosklonny, *Cancer Chemoth Pharm* **1997**, *40*, 60-64.
[299] G. Dasgupta and J. Momand, *Exp Cell Res* **1997**, *237*, 29-37.

[300] M. V. Blagosklonny, *FASEB J.* **2000**, *14*, 1901-1907.
[301] T. Fujiwara, E. A. Grimm, T. Mukhopadhyay, W. W. Zhang, L. B. Owenschaub and J. A. Roth, *Cancer Res.* **1994**, *54*, 2287-2291.
[302] D. Katayose, J. Gudas, H. Nguyen, S. Srivastava, K. H. Cowan and P. Seth, *Clin Cancer Res* **1995**, *1*, 889-897.
[303] M. V. Blagosklonny and W. S. ElDeiry, *Int J Cancer* **1996**, *67*, 386-392.
[304] B. Yang, J. R. Eshleman, N. A. Berger and S. D. Markowitz, *Clin Cancer Res* **1996**, *2*, 1649-1657.
[305] R. D. Meng and W. S. El-Deiry, *Drug Resist Update* **1998**, *1*, 205-210.
[306] G. Selivanova, V. Iotsova, I. Okan, M. Fritsche, M. Strom, B. Groner, R. C. Grafstrom and K. G. Wiman, *Nat Med* **1997**, *3*, 632-638.
[307] G. Selivanova, L. Ryabchenko, E. Jansson, V. Iotsova and K. G. Wiman, *Mol. Cell. Biol.* **1999**, *19*, 3395-3402.
[308] B. A. Foster, H. A. Coffey, M. J. Morin and F. Rastinejad, *Science* **1999**, *286*, 2507-2510.
[309] A. Ventura, D. G. Kirsch, M. E. McLaughlin, D. A. Tuveson, J. Grimm, L. Lintault, J. Newman, E. E. Reczek, R. Weissleder and T. Jacks, *Nature* **2007**, *445*, 661-665.
[310] C. Rollenhagen and P. Chene, *Int J Cancer* **1998**, *78*, 372-376.
[311] M. G. Mateu, M. M. S. Del Pino and A. R. Fersht, *Nat Struct Biol* **1999**, *6*, 191-198.
[312] L. T. Chong, C. D. Snow, Y. M. Rhee and V. S. Pande, *J Mol Biol* **2005**, *345*, 869-878.
[313] G. M. K. Poon, R. D. Brokx, M. Sung and J. Gariepy, *J Mol Biol* **2007**, *365*, 1217-1231.
[314] M. G. Mateu and A. R. Fersht, *Embo J.* **1998**, *17*, 2748-2758.
[315] T. Soussi, C. C. Defromentel and P. May, *Oncogene* **1990**, *5*, 945-952.
[316] T. Soussi and P. May, *J Mol Biol* **1996**, *260*, 623-637.
[317] S. Y. Shieh, J. Ahn, K. Tamai, Y. Taya and C. Prives, *Genes Dev.* **2000**, *14*, 289-300.
[318] J. M. Stommel, N. D. Marchenko, G. S. Jimenez, U. M. Moll, T. J. Hope and G. M. Wahl, *Embo J.* **1999**, *18*, 1660-1672.
[319] R. C. Ribeiro, F. Sandrini, B. Figueiredo, G. P. Zambetti, E. Michalkiewicz, A. R. Lafferty, L. DeLacerda, M. Rabin, C. Cadwell, G. Sampaio, I. Cat, C. A. Stratakis and R. Sandrini, *P Natl Acad Sci USA* **2001**, *98*, 9330-9335.
[320] M. E. Lomax, D. M. Barnes, T. R. Hupp, S. M. Picksley and R. S. Camplejohn, *Oncogene* **1998**, *17*, 643-649.
[321] S. Mesiano and R. B. Jaffe, *Endocr. Rev.* **1997**, *18*, 378-403.
[322] S. J. Spencer, S. Mesiano, J. Y. Lee and R. B. Jaffe, *J Clin Endocr Metab* **1999**, *84*, 1110-1115.
[323] H. Y. Dai, N. Tsao, W. C. Leung and H. Y. Lei, *Radiat Res* **1998**, *150*, 183-189.
[324] X. Salvatella, M. Martinell, M. Gairi, M. G. Mateu, M. Feliz, A. D. Hamilton, J. de Mendoza and E. Giralt, *Angew Chem Int Edit* **2004**, *43*, 196-198.
[325] P. R. E. Mittl, P. Chene and M. G. Grutter, *Acta Crystallographica Section D-Biological Crystallography* **1998**, *54*, 86-89.
[326] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, *Intermolecular Forces*, B. Pullman, Reidel, Dordrecht, **1981**, p. 331.
[327] G. W. Robinson, S. B. Zhu, S. Singh and M. W. Evans, *Water in Biology, Chemistry and Physics; Experimental Overviews and Computational Methodologies*, Singapore, **1996**, p.
[328] T. Z. Lwin, J. J. Durant and D. Bashford, *J Mol Biol* **2007**, *373*, 1334-1347.
[329] K. T. Jeang, H. Xiao and E. K. Rich, *J Biol Chem* **1999**, *274*, 28837-28840.
[330] M. Green, P. M. Loewenstein, R. Pusztai and J. S. Symington, *Cell* **1988**, *53*, 921-926.
[331] J. B. Rothbard, S. Garlington, Q. Lin, T. Kirschberg, E. Kreider, P. L. McGrane, P. A. Wender and P. A. Khavari, *Nat Med* **2000**, *6*, 1253-1257.
[332] T. Haack, M. W. Peczuh, X. Salvatella, J. Sanchez-Quesada, J. de Mendoza, A. D. Hamilton and E. Giralt, *J Am Chem Soc* **1999**, *121*, 11813-11820.
[333] J. Fernandez-Carneado, M. J. Kogan, S. Castel and E. Giralt, *Angew Chem Int Edit* **2004**, *43*, 1811-1814.
[334] J. R. Masters, *Nat Rev Cancer* **2002**, *2*, 315-319.
[335] E. A. Goun, T. H. Pillow, L. R. Jones, J. B. Rothbard and P. A. Wender, *Chembiochem* **2006**, *7*, 1497-1515.
[336] C. D. Gutsche, *Calixarenes Revisited*, Fort Worth, USA, **1998**, p.
[337] C. D. Gutsche and R. Muthukrishnan, *J Org Chem* **1978**, *43*, 4905-4906.
[338] A. Zinke and E. Ziegler, *Ber Dtsch Chem Ges* **1941**, *74*, 1729-1736.
[339] C. D. Gutsche, B. Dhawan, J. A. Levine, K. H. No and L. J. Bauer, *Tetrahedron* **1983**, *39*, 409-426.
[340] P. D. J. Grootenhuis, P. A. Kollman, L. C. Groenen, D. N. Reinhoudt, G. J. Vanhummel, F. Ugozzoli and G. D. Andreetti, *J Am Chem Soc* **1990**, *112*, 4165-4176.
[341] J. D. Vanloon, W. Verboom and D. N. Reinhoudt, *Org. Prep. Proced. Int.* **1992**, *24*, 437-462.

[342] S. Shinkai, T. Nagasaki, K. Iwamoto, A. Ikeda, G. X. He, T. Matsuda and M. Iwamoto, *Bull. Chem. Soc. Jpn.* **1991**, *64*, 381-386.
[343] J. Blixt and C. Detellier, *J Am Chem Soc* **1994**, *116*, 11957-11960.
[344] S. Kanamathareddy and C. D. Gutsche, *J Org Chem* **1995**, *60*, 6070-6075.
[345] A. Arduini, M. Fabbi, M. Mantovani, L. Mirone, A. Pochini, A. Secchi and R. Ungaro, *J Org Chem* **1995**, *60*, 1454-1457.
[346] S. K. Menon and M. Sewani, *Rev Anal Chem* **2006**, *25*, 49-82.
[347] R. Ludwig, *Fresenius Journal of Analytical Chemistry* **2000**, *367*, 103-128.
[348] S. Bhattacharya, S. K. Nayak, S. Chattopadhyay, M. Banerjee and A. K. Mukherjee, *J Phys Chem B* **2003**, *107*, 11830-11834.
[349] A. Ghosh and M. Bansal, *Acta Crystallogr D* **2003**, *59*, 620-626.
[350] R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Nature* **1980**, *287*, 755-758.
[351] C. O. Pabo and R. T. Sauer, *Annu Rev Biochem* **1984**, *53*, 293-321.
[352] P. E. Nielsen, *Chem-Eur J* **1997**, *3*, 505-508.
[353] N. C. Seeman, J. M. Rosenberg and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 804-808.
[354] L. Marino-Ramirez, M. G. Kann, B. A. Shoemaker and D. Landsman, *Exp. Rev. Proteom.* **2005**, *2*, 719-729.
[355] M. P. Knauert and P. M. Glazer, *Hum. Mol. Genet.* **2001**, *10*, 2243-2251.
[356] M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden and P. E. Nielsen, *Nature* **1993**, *365*, 566-568.
[357] U. Pindur, M. Jansen and T. Lemster, *Curr. Med. Chem.* **2005**, *12*, 2805-2847.
[358] J. B. Blanco, V. L. Dodero, M. E. Vazquez, M. Mosquera, L. Castedo and J. L. Mascarenas, *Angew Chem Int Edit* **2006**, *45*, 8210-8214.
[359] A. G. Schatzlein, *Anti-Cancer Drugs* **2001**, *12*, 275-304.
[360] C. Mah, B. J. Byrne and T. R. Flotte, *Clin. Pharmacokinet.* **2002**, *41*, 901-911.
[361] M. E. Davis, *Curr. Opin. Biotechnol.* **2002**, *13*, 128-131.
[362] A. V. Kabanov, *Pharm. Sci. Tech. Today* **1999**, *2*, 365-372.
[363] D. Putnam, A. N. Zelikin, V. A. Izumrudov and R. Langer, *Biomaterials* **2003**, *24*, 4425-4433.
[364] Y. K. Park, Y. H. Park, B. A. Shin, E. S. Choi, Y. R. Park, T. Akaike and C. S. Cho, *J. Controlled Release* **2000**, *69*, 97-108.
[365] D. Oupicky, K. A. Howard, C. Konak, P. R. Dash, K. Ulbrich and L. W. Seymour, *Bioconjugate Chem.* **2000**, *11*, 492-501.
[366] Y. Liu, L. C. Mounkes, H. D. Liggitt, C. S. Brown, I. Solodin, T. D. Heath and R. J. Debs, *Nat. Biotechnol.* **1997**, *15*, 167-173.
[367] D. Goula, C. Benoist, S. Mantero, G. Merlo, G. Levi and B. A. Demeneix, *Gene Ther.* **1998**, *5*, 1291-1295.
[368] H. E. J. Hofland, D. Nagy, J. J. Liu, K. Spratt, Y. L. Lee, O. Danos and S. M. Sullivan, *Pharm. Res.* **1997**, *14*, 742-749.
[369] R. I. Mahato, K. Anwer, F. Tagliaferri, G. Meaney, P. Leonard, M. S. Wadhwa, M. Logan, M. French and A. Rolland, *Hum. Gene Ther.* **1998**, *9*, 2083-2099.
[370] P. R. Dash, M. L. Read, K. D. Fisher, K. A. Howard, M. Wolfert, D. Oupicky, V. Subr, J. Strohalm, K. Ulbrich and L. W. Seymour, *J. Biol. Chem.* **2000**, *275*, 3793-3802.
[371] W. T. Collard, Y. S. Yang, K. Y. Kwok, Y. Park and K. G. Rice, *J. Pharm. Sci.* **2000**, *89*, 499-512.
[372] J. Fernandex-Carneado, M. Van Gool, V. Martos, S. Castel, P. Prados, J. de Mendoza and E. Giralt, *J Am Chem Soc* **2005**, *127*, 869-874.
[373] P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J. L. Wietor, J. K. M. Sanders and S. Otto, *Chem. Rev.* **2006**, *106*, 3652-3711.
[374] D. J. Dunican and P. Doherty, *Biopolymers* **2001**, *60*, 45-60.
[375] M. Silhol, M. Tyagi, M. Giacca, B. Lebleu and E. Vives, *Eur. J. Biochem.* **2002**, *269*, 494-501.
[376] E. Vives, J. P. Richard, C. Rispal and B. Lebleu, *Curr. Protein Pept. Sci.* **2003**, *4*, 125-132.
[377] M. F. Ross, A. Filipovska, R. A. J. Smith, M. J. Gait and M. P. Murphy, *Biochem. J.* **2004**, *383*, 457-468.
[378] T. Darden, D. York and L. Pederson, *J Chem Phys* **1993**, *98*, 10089.
[379] B. Hess, H. Beker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J Comput Chem* **1997**, *18*, 1463-1472.
[380] A. Savelyev and G. A. Papoian, *J. Am. Chem. Soc.* **2006**, *128*, 14506-14518.