

USE OF CALIX[4]ARENES TO  
RECOVER THE SELF-ASSEMBLY  
ABILITY OF MUTATED p53  
TETRAMERIZATION DOMAINS

Susana Gordo Villoslada

2008

Memòria presentada per

Susana Gordo Villoslada

per optar al grau de doctor per la Universitat de Barcelona

Revisada per:

Prof. Ernest Giralt i Lledó

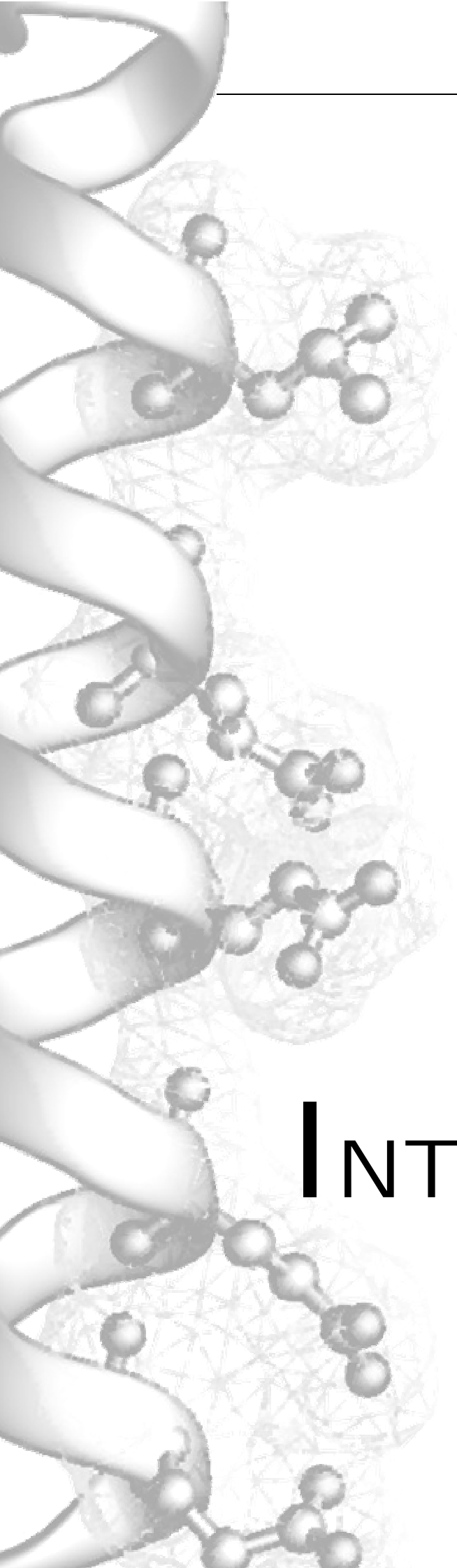
Universitat de Barcelona

Director

Programa de Química Orgànica

Bienni 2003-2005

Barcelona, abril de 2008



# INTRODUCTION



# In the net or

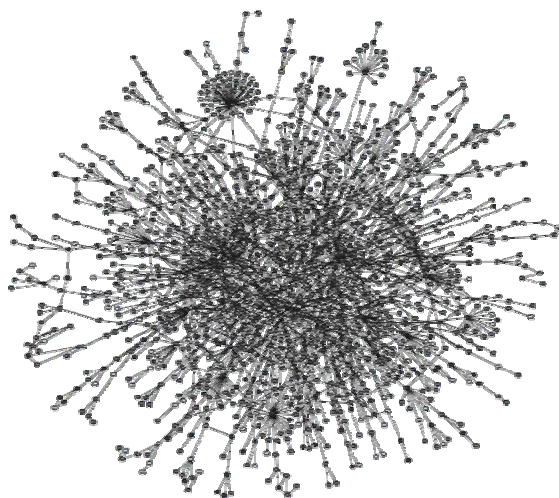
Proteins constitute the working machinery and structural support of all organisms. In performing a given function, they must adopt highly specific structures that can change with their level of activity, often through the direct or indirect action of other proteins. Indeed, proteins typically function within an ensemble, rather than individually. Hence, they must be sufficiently flexible to interact with each other and execute diverse tasks. The discovery that errors within these groups can ultimately cause disease has led to a paradigm shift in drug discovery, from an emphasis on single protein targets to a holistic approach whereby entire ensembles are targeted.



## 1.1 Protein network or

Proteins are the main players within cells. They can carry out myriad functions, including structural support, cell cycle control, enzymatic activity, cell signaling, and immune response among many others. Whoever they are and whatever they do, the key characteristic for the functional diversity in proteins is their ability to recognize and bind other molecules, whether they are other proteins, other biomolecules (*e.g.* sugars, fatty acids, nucleic acids), small ligands or simple ions.

Within the cell, protein-protein interactions are organized into an exquisite highly complex *network*, known as *the interactome*,<sup>1</sup> whereby proteins can be depicted as *nodes* and interactions as *edges* (**Figure 1**). Essential proteins are more connected (*hubs*) than non-essential ones, hence they constitute the structural basis of the network, phenomenon recently named as the *centrality-lethality rule*, since deletion of a hub protein is likely to be lethal for the organism.<sup>2</sup> Nonetheless, deletion of non-hub proteins can also prove fatal. There are only a few essential proteins with tens, hundreds or thousands of links; These high levels of hub connectivity are somehow reflected in the protein structure, given that protein *promiscuity* and *diversity* can be only achieved by certain intrinsic *structural flexibility*.<sup>3-8</sup>



**Figure 1.** Yeast protein interaction network map.<sup>9</sup>

The scale-free connectivity network can be dissected into functional sub-networks which represent singular signaling *pathways*, that is, ensembles of distinct proteins that act in concert to transduce extrinsic or intrinsic information. Many interactions within the network are subjected to temporal and spatial dynamic regulation, which depends on signal-induced or context-dependent post-translational modifications. Because of the transient nature of those interactions leading to post-translational changes, they are described as *soft wires* and they are difficult to detect. In contrast, physical interactions that are not so labile and can be readily identified are accounted as *hard wires*.<sup>10,11</sup> However, the hardness of a wire is not related to the importance of the interaction. Each task has its optimal affinity and evolution not always leads to tighter interactions.

Homo-oligomeric proteins are a particularly relevant case in the protein network. In fact, it is estimated that over 35% of proteins self-assemble to carry out their function.<sup>12</sup> The most are dimers and tetramers;<sup>13</sup> higher order oligomers are less common (but in the case of some structural proteins) and odd-number stoichiometries are also rare.<sup>12,14</sup> These defined oligomeric structures characterize for their high degree of symmetry.<sup>12</sup> Many functional advantages have favored the oligomeric state through evolution;<sup>15</sup> self-assembly confers an additional level of control and regulatory flexibility as well as more stability and resistance to degradation, denaturation and mutations, since the chance of transcription errors is lower for short sequences. Nevertheless, oligomerization may also be disadvantageous and it is not always the evolutionary direction.

## 1.2 Protein-protein interaction

*Protein-protein interactions* have garnered ever increasing attention in the past few years. This is due to their central role in the emerging field of *systems biology*, which studies biological systems as the association of complex interaction networks, thereby introducing a new perspective in understanding life processes: “integration instead of reduction”.

### o protein interact

Regardless of the biological importance of an interaction, proteins interact with each other with high specificity under a wide range of affinities, from the high millimolar to the low femtomolar.<sup>16</sup>

Protein-protein complexes can be described according to three parameters:<sup>17</sup>

- (i) the nature of the units: *homo-oligomeric* (same) or *hetero-oligomeric* (different)
- (ii) the life-time of the complex: *permanent* (very stable, only exists the complex form) or *transient* (associates and dissociates *in vivo*, whether in dynamic equilibrium or under influence of a molecule which triggers an equilibrium shift)
- (iii) the existence of the units: *obligate* (partners do not exist on their own and dissociation leads to denaturation; therefore, they are permanent) or *non-obligate* (partners exist independently).

Many interactions do not fall into any distinct type and there is a continuous between obligate-non-obligate and permanent-transient.<sup>18</sup>

The stability of a complex very much depends on the environment.<sup>17</sup> Indeed, all interactions are ultimately driven by the concentration of the components and by the free energy change in the formation of the complex. Thus, protein-protein interactions can be controlled by altering the *local concentration* (gene expression, protein degradation, temporary storage or co-localization in time and space) and by influencing the *binding affinity* (local change in physiological conditions or physicochemical modifications of the interface).



t the inter ace

*Specificity* derives from the *complementarity* of physical shape and chemistry of the binding interfaces, although structural hindrances of the protein as a whole also condition the interaction.

One of the challenges in the study of protein-protein interfaces is their *diversity*, which is a direct consequence of the variety of biological roles resulting from protein-protein interaction.<sup>14</sup> Every protein complex has adopted different solutions to fulfill the binding requirements for specificity and affinity. Consequently, no universal statements can be made about the composition and architecture of protein–protein binding sites.

Protein-protein interface areas range from 600Å<sup>2</sup> to extensions as large as >4000Å<sup>2</sup>, being the average *ca.* 1500Å<sup>2</sup> (*i.e.* *ca.* 170 atoms)<sup>19</sup> although, it clearly depends on the protein size and the nature of the complex. Interfaces are generally circular-like and flat, and interfacial residues tend to protrude.<sup>20</sup> Electrostatic and geometric complementarity enables specificity of the interaction, but the size and the shape of the interface do not directly correlate with the binding energy.<sup>17</sup> Interfaces are not homogenous and their stability usually depends on only a few crucial residues.<sup>21</sup> Those are the so-called *hotspots* and, whether packed together or dispersed through the surface, they contribute the most to the binding energy.<sup>22</sup> The surrounding residues protect the hotspots from the bulk solvent; this is necessary to enable tight interactions.<sup>15,19,23</sup> Paradoxically, water molecules also form part of many protein-protein interfaces, fitting into small cavities of the planar surface and establishing hydrogen bonds between the two sides.<sup>24-27</sup> Nevertheless, not all protein-protein interactions need to be tight and hence, not all interfaces present hotspots.

In general, interfaces are more hydrophobic than protein exteriors but more polar than protein interiors.<sup>23</sup> The residue composition of interfaces does not greatly differ from that of protein surfaces, although aromatic residues and arginine appear somewhat more frequently. Actually, the most frequent hotspots are tryptophane, arginine and tyrosine –all of which can establish multiple types of favorable interactions–<sup>19</sup> while others as valine, lysine or serine are scarce.<sup>23</sup> The hydrophobic effect is most of the times the driving force of the interaction but, differing from a protein core, hydrogen bonds, salt bridges and water molecules also do their bit.<sup>21,28,29</sup> Carbonyl groups from the interface backbone, which comprise nearly 10% of the interface area, have also a role to play in stabilizing interfaces.<sup>19,30</sup>

It is worth noting that the nature of the complex conditions the nature of the interface.<sup>16</sup> Thus, for instance, permanent and obligate protein complexes present interfaces of high hydrophobicity whose interactions result into a tighter binding, whereas proteins interacting only transiently have interfaces of substantially lower hydrophobicity. This is totally coherent, since in transient complexes, proteins exist as free entities, and therefore, their interfaces are most of the time solvent exposed. However, due to protein floppiness, there are exceptions to this trend.<sup>31</sup>

Structural rearrangements of the unbound proteins are part of the binding process. They range from minor movements of the side chains to major reorganizations in the backbone, such as loops switches or domain conformational changes.<sup>23</sup> Rearrangements on the structure can be induced by the binding event itself or promoted by post-translational modifications or environmental changes.

Whatever the trigger, protein conformation can act as a selection mechanism for protein-protein interaction, hence, for activity.

## 1 o e not hen o ethin oe ron

The perfect balance of the protein network is required for the organism to operate normally.<sup>32,33</sup> This in turn demands that the conformation, concentration, localization and timing of interacting proteins must be orchestrated flawlessly. Should any of these conditions not be satisfied, the system would fail and disease would eventually set in.<sup>34-36</sup>

Many factors –at the molecular level– can alter the harmony of the protein network. Disturbance may imply the *loss-of-function* of a protein or a pathway (a *hole* in the net), the *gain* of an undesired one (a *knot* in the net), as well as both phenomena simultaneously. Clearly, the more central the perturbed hub, the worse the consequences.<sup>34,37</sup>

The protein network can suffer from *exogenous* or *endogenous* perturbations. In the former, a pathogenic external factor (e.g. a bacterium or a virus) directly interferes in the protein-protein connections, cutting established wires and establishing new ones; whereas in the latter, the problem lies in malfunctions of the organism itself, usually at the genomic level (*i.e.* mutations).

A point change in the nucleotide sequence of a gene can have catastrophic consequences for the organism; indeed, many diseases are associated to dysfunctional mutated proteins. Substitutions, deletions or truncations in the sequence of a protein can abolish or diminish (loss-of-function) or even enhance (gain-of-function) pre-existing protein-protein interactions. The mutation can directly alter the binding interface or the structure of the binding domain, but it can also affect other distant residues which, despite not interacting physically with the partner proteins, play a key role in controlling modifications required for the binding, expression or localization of the protein.<sup>38</sup> Besides affecting the former connections, mutations can also promote new undesired interactions with other, new partners (*i.e.* gain-of-function) which typically have negative implications for the organism.

There are other endogenous mechanisms which can impair network traffic. This is the case of the well-known, and regrettably frequent, *misfolding disorders*.<sup>39</sup>

### i o din di order

Proteins synthesis and assembly occurs in the endoplasmatic reticulum (ER), which provides the optimal environment for folding and post-translational modifications. These processes are thoroughly supervised by a complex quality control system, that is mainly ruled by endogenous molecular chaperones, which assist folding and prevent accumulation of defective proteins.<sup>40,41</sup> A point mutation in the protein sequence or the perturbation of the folding process by any

environmental trigger (e.g. protein overexpression, oxidative stress, pH changes, osmotic shock, defects in the chaperones pathways) can lead to a misfolded protein which is then polyubiquitylated by the ER quality control system and destroyed by the proteasome.<sup>42,43</sup> Consequently, the protein and its function are lost. Examples here include cystic fibrosis, familial hypercholesterolemia, renitis pygmentosa, nephrogenic diabetes insipidus, Fabry's and Gaucher's disease among others.<sup>44</sup>

Furthermore, the clearance mechanism of unfolded species can fail causing the species to accumulate within the cell up to finally aggregate into cytotoxic macro-structures, whether highly-organized (e.g. amyloid fibrils) or amorphous. This chain of events causes pathologies such as  $\alpha$ -antitrypsin deficiency or early onset cataracts.<sup>44</sup>

Neurodegenerative disorders like Alzheimer's, Parkinson's, Huntington's or prion disease are also the result of the accumulation and aggregation of misfolded proteins, although in these maladies the native protein is converted into structured amyloid fibrillar aggregates, which can occur intra- or extracellularly.<sup>45-48</sup> Rather than a problem in the protein folding process, this accumulation is the result of unbalanced kinetics of the protein misfolding and aggregation and the clearance of aberrant species.<sup>49-52</sup>

Protein misfolding pathologies can also lead to connective tissue diseases, which often result from the incorporation of toxic or imperfect conformations into the functional protein structure. Besides environmental factors, genetic mutations may also affect collagen and keratin resulting in defective constructions.<sup>53</sup> Inherited connective tissue diseases include Marfan syndrome (tissue abnormalities in the heart, aorta, lungs, eyes, and skeleton), Ehlers-Danlos syndrome (loose, fragile skin or loose joints), osteogenesis imperfecta and pediatric rheuma.



synthetic properties)<sup>83-86</sup> or the use of *proteomimetics* (designed small synthetic constructions)<sup>86-91</sup> to mimic structural and functional features of the target protein.<sup>92</sup> The rational design on the basis of key structural features of the target surface structure employing new multivalent scaffolds also gathers a good number of reported successes (section 1.6).

#### Stabilization of protein-protein interaction

Stabilization of protein-protein interactions holds potential as a therapeutic strategy, yet it has garnered little attention.<sup>59</sup> Block and co-workers recently claimed that stabilization could prove even more effective than inhibition.<sup>59</sup> They found that a number of therapeutically interesting targets form a “druggable” cavity at the boundaries of the proteins complex, in which fitting of a small molecule should be thermodynamically favorable.

Some drugs currently on the market act by stabilizing pre-existing protein-protein assemblies (e.g. rapamycin, tacrolimus, brefeldin A, taxol, forskolin), although in their origins they were discovered as “inhibitors” (from high throughput screenings of natural products), since they are able to inhibit the target protein by stabilizing other protein-protein interactions.

From the design perspective, almost all the work has been focused on cell surface receptors which need to oligomerize for signal transduction.<sup>93-95</sup> Several “dimerizers”, composed by two anchoring groups linked by an appropriate spacer,<sup>58</sup> have been reported to date, but they are only used (successfully) in laboratory practices.<sup>96-99</sup>

#### Alternative approaches to deal with misfolding and aggregation disorders

Alternative approaches are followed to deal with misfolding and aggregation disorders,<sup>100</sup> although the recognition of the protein surface is still a requisite for most of them.

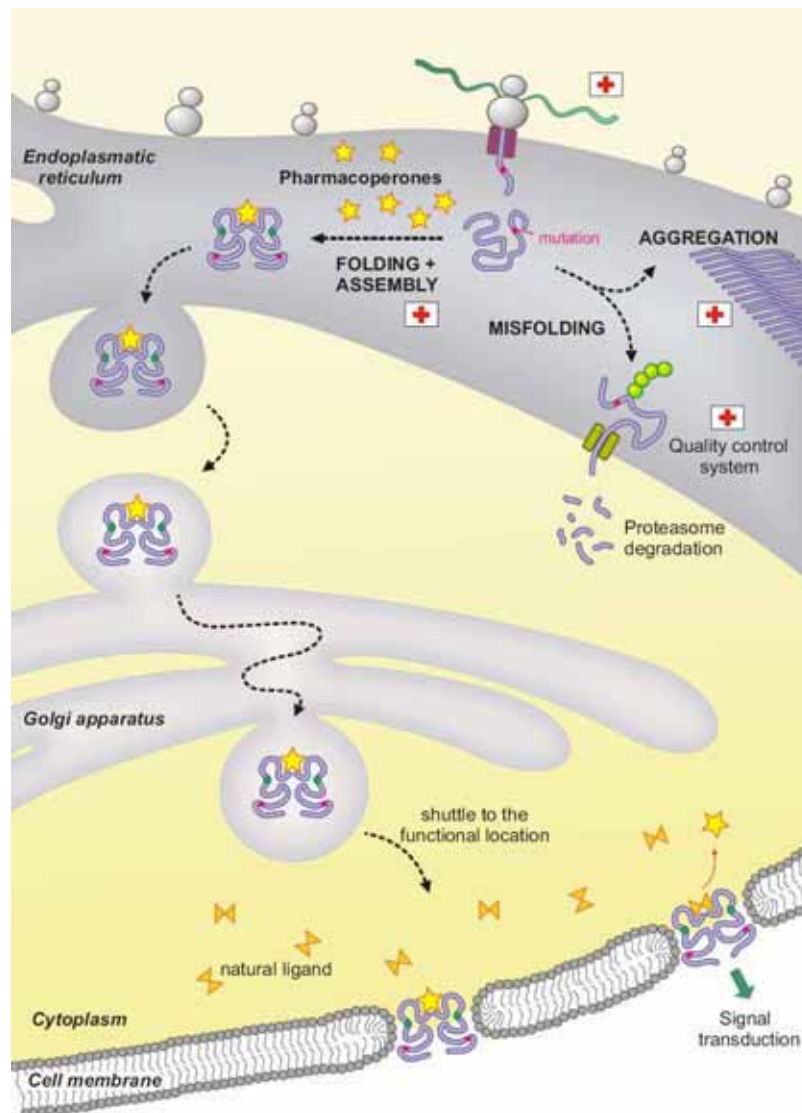
One possibility is to work at the genetic level. When a mutated misfolded protein is prematurely degraded, then the wild type gene can be introduced into the cell by transfection (*i.e.* genetic therapy). Conversely, when the troubles come from the deficient removal of misfolded species, improvements can be made by up-regulating endogenous molecular chaperones or aggregate clearance mechanisms.<sup>101-103</sup>

At the protein level, aggregation can also be inhibited by molecules which bind to the growing aberrant amyloid structures to ultimately inhibit the self-assembly progress.<sup>100</sup> Another option is the use of *kinetic native-state stabilizers* –introduced by Kelly and co-worker– which are drug-like molecules designed to specifically and efficiently interact with the native state of the protein to stabilize it and prevent its misfolding thereby precluding toxic aggregation.<sup>48,104</sup>

Misfolded nascent mutated polypeptides can also be saved through the use of exogenous chaperones, that is, molecules which (whether specifically or not) aid the protein to fold correctly, hence, evading the quality control in the endoplasmic reticulum and shuttling the protein to its functional location.<sup>44,105,106</sup> Artificial chaperones can be classified according to their specificity.

*Chemical chaperones* are low molecular weight chemicals (e.g. glycerol, deuterated water, DMSO, 4-phenylbutiric acid, trimethylamine N-oxide) that act as osmolites; they drive the folding of the defective protein towards the native state by increasing the hydrostatic pressure (and through other not yet well understood mechanisms).<sup>107-110</sup> Despite their proved efficacy, the clinical application of such unspecific molecules is rather unlikely.

More promising are the so-called *pharmacological chaperones* or *pharmacoperones*<sup>101,106,111-113</sup> (**Figure 2**). They are small template molecules that specifically bind the protein in its native conformation, thereby shifting the folding equilibrium towards the correct structure. Most of the reported examples correspond to the rescue of misfolded mutated enzymes by –ironically– inhibitors. This approach has become very promising for genetic diseases, especially for those cases in which structural instability of a protein is caused by a mutation.<sup>114-122</sup>



**Figure 2.** Pharmacological chaperone in action. Rescue of a traffic-defective mutated membrane protein (adapted from Pelmutter<sup>105</sup>).

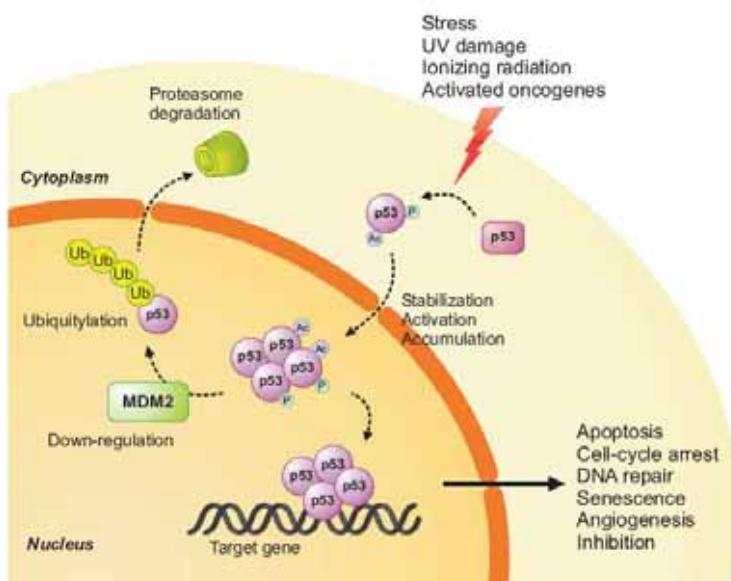
Of course, not all genetic mutations are “pharmacologically treatable”, only some of those that affect the protein trafficking, not those that compromise the activity or the protein integrity.<sup>106</sup> For those instances, the only salvation is genetic therapy (e.g. connective tissue diseases), although the current limitation of suitable methods for cell transfection limit the clinical applicability.

## 1 The e a p e protein p

The tumor suppressor protein p53 perfectly illustrates the relevance of the protein network, the drastic consequences which result from its perturbation and the strategies which can be followed to restore it.

Talking about p53 is talking about cancer. Nearly 60% of human tumors present alterations in the p53 pathway. This proportion, which far exceeds that for other proteins, is testament to the central role of p53 and its pathways (a huge hub in the network) in the regulation and growth of the organism. Hence, p53 is a clear example of the centrality-lethality rule.

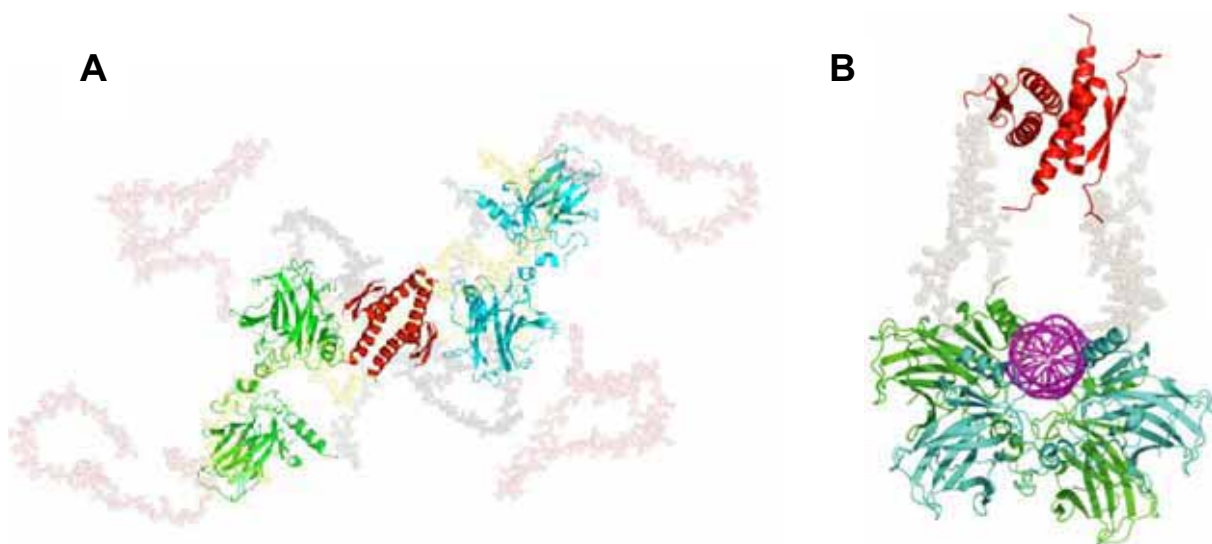
p53 is a transcription factor of genes which control the cell cycle and preserve the genomic integrity of the organism,<sup>123,124</sup> reason why it has been so-called the “genome guardian”.<sup>125</sup> Protein p53 can be activated by numerous factors, primarily, DNA damage and other situations of cellular stress, and can respond in a variety of ways, including cell-cycle arrest and DNA repair or apoptosis (**Figure 3**).<sup>126,127</sup> Because it plays a key role in Life, p53 is one of the most –if not the most– extensively studied proteins and it is the target protein *par excellence* in cancer therapy.



**Figure 3.** Activation of p53 and cellular responses.

From a structural point of view, the active form of p53 is a homotetramer constituted by the self-assembly of four monomers of 393 residues each (**Figure 4**).<sup>128</sup> Five main domains<sup>129</sup> can be distinguished in each monomer chain: the N-terminal acidic transactivation domain (1-42), the Pro-rich domain (61-94), the DNA binding domain (102-292), the tetramerization domain (326-357) and the C-terminal regulatory domain (360-393). By far, the most important is the large, central DNA binding domain. Nevertheless, the rest of domains have not to be overlooked since they are all essential for the appropriate function of the protein.

In line with its central position within the cellular network, p53 presents several regions of intrinsic disorder (**Figure 4**).<sup>128,130</sup> These are mainly the regulatory N- and C-terminal domains although the folded DNA binding domain is also somewhat structurally unstable,<sup>131</sup> which confers the promiscuity required for binding to multitude of partners, from DNA to proteins, most yet to be identified.<sup>132</sup> The activity of the protein is regulated by numerous intricate mechanisms which control expression levels, degradation rate, localization, structure, interaction with biomolecules, post-translational modifications<sup>133,134</sup> ...

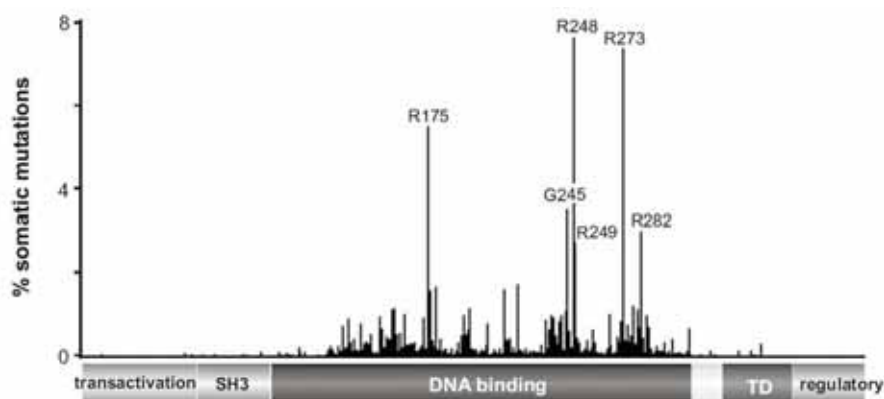


**Figure 4.** SAXS models of **(A)** free p53 and **(B)** p53-DNA complex. The structured core DNA binding domain (green and blue) and tetramerization domain (red) are displayed in cartoon representation, and unstructured connector linkers (grey), N-termini (salmon) and C-termini (yellow) in semitransparent space-fill mode.<sup>128</sup>

The fine balance of the p53 pathway can be altered by many factors. The most common impairment is the direct mutation of *p53* gene,<sup>135-139</sup> which in more of the 95% of the cases affects the prominent DNA binding region (**Figure 5**). Single missense mutations can compromise the structure of the protein or modify key residues either directly involved in the interaction with other biomolecules or required for post-translational modifications. Besides loss-of-function, mutations can also result in a protein with undesired oncogenic activity by promoting the transcription of oncogenes (gain-of-function).<sup>140,141</sup>



In a significant 10% of p53-associated tumor cases, the pathway fails due to alterations in direct or indirect partners of p53, the most relevant of which is the protein MDM2.<sup>138</sup> Interactions with viral proteins (e.g. human papilloma virus E6 protein, HIV1 Tat protein) have also been reported to neutralize p53 activity and establish conditions that favor cancer.<sup>142-144</sup>



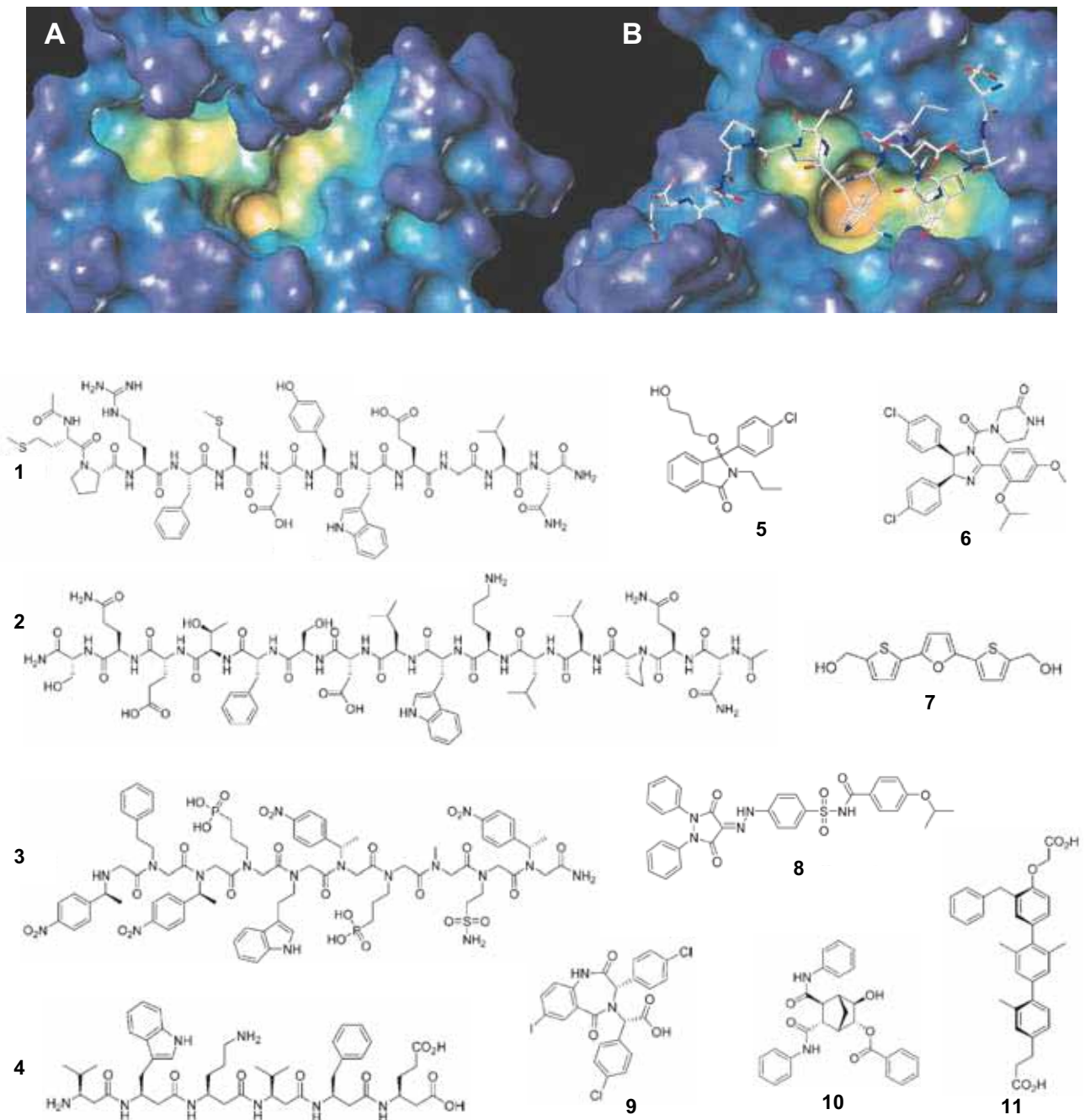
**Figure 5.** Frequency and distribution of p53 mutations along the sequence. From the IARC TP53 Mutation Database, release R12, November 2007.<sup>145-148</sup>

Reconstruction of the p53 tumor suppressor pathway has become one of the most exciting novel concepts in cancer therapy and a growing number of p53-targeting strategies have been proposed in the past few years.<sup>145-148</sup>

For cases in which the wild-type p53 gene is present (ca. 10%), enhancement of the protein activity is usually sought. The main target on this aim is the p53-MDM2 interaction,<sup>149-151</sup> although it is not the only one, since p53 interacts with a wide spectrum of proteins.<sup>152</sup> MDM2 down-regulates p53 by direct binding; inhibition of their interaction results in the accumulation of active p53 protein. Actually, the inhibition of p53-MDM2 interaction has been proposed as the *model of study* for the design of inhibitory strategies for protein-protein interactions and many reviews have been written on the matter.<sup>57,153-156</sup> Given that in the unbound state the p53 binding sequence (at the N-terminus) is not structured, most of the efforts have been focused on the MDM2 surface. Antibodies, oligonucleotides,  $\alpha$ -peptides,  $\beta$ -peptides, retro-enantio-peptides, peptidomimetics of any other kind, peptoids, miniproteins, small molecules from natural product libraries or from chemical libraries, *in silico* screening... successfully or not, almost every current strategy has been employed in the search for inhibitors of the p53-binding site on MDM2 (**Figure 6**).<sup>156</sup>

Besides direct inhibition of MDM2-p53 interface, other explored alternatives deal with the down-regulation of MDM2 protein itself via targeting other MDM2-protein interactions.

For cases in which the p53 gene is mutated (ca. 90%), efforts have been recently focused on the pharmacological rescue of the non-functional protein by small binding molecules. Several molecular chaperones have been reported to recover the folded structure and the activity of some inactive unfolded mutants of the DNA-binding domain.<sup>149-151</sup>



**Figure 6.** Inhibitors for MDM2-p53 interaction.<sup>156</sup> (A) MDM2 unbound surface and (B) MDM2 complexed with the p53 N-terminal peptide. 1 Peptide derived from p53 binding sequence ( $IC_{50}$ : 313nM). 2 D-retroinverso antagonist peptide ( $IC_{50}$ : 15 $\mu$ M). N-methyl peptoid ( $IC_{50}$ : 25 $\mu$ M). Helical  $\beta$ -peptide ( $IC_{50}$ : 80 $\mu$ M). First inhibitor identified by computational screening ( $IC_{50}$ : 25 $\mu$ M). Nutlin (Roche), the first potent inhibitor *in vitro* and *in vivo* ( $IC_{50}$ : 90nM). RITA, molecule believed to bind p53 rather than MDM2. Sulfonamide inhibitor from *in silico* screening ( $IC_{50}$ : 32 $\mu$ M). Inhibitor from combinatorial chemistry Ugi 4-components condensation, later optimized from cell-based activity ( $IC_{50}$ : 80nM). 1 First non-peptide small molecule inhibitor *de novo* designed ( $IC_{50}$ : 10 $\mu$ M). 11 Terphenyl mimic of p53  $\alpha$ -helix ( $IC_{50}$ : 05 $\mu$ M).

Given that mutations mainly affect the DNA-binding domain, little attention has been paid to other regions, despite the fact that mutations in other domains have also been found in human cancer. For instance, regarding the tetramerization domain, successful attempts of improving the stability of p53 tetramerization have been reported, although they always entail synthetic or genetic modification of the native protein sequence.<sup>157-164</sup>

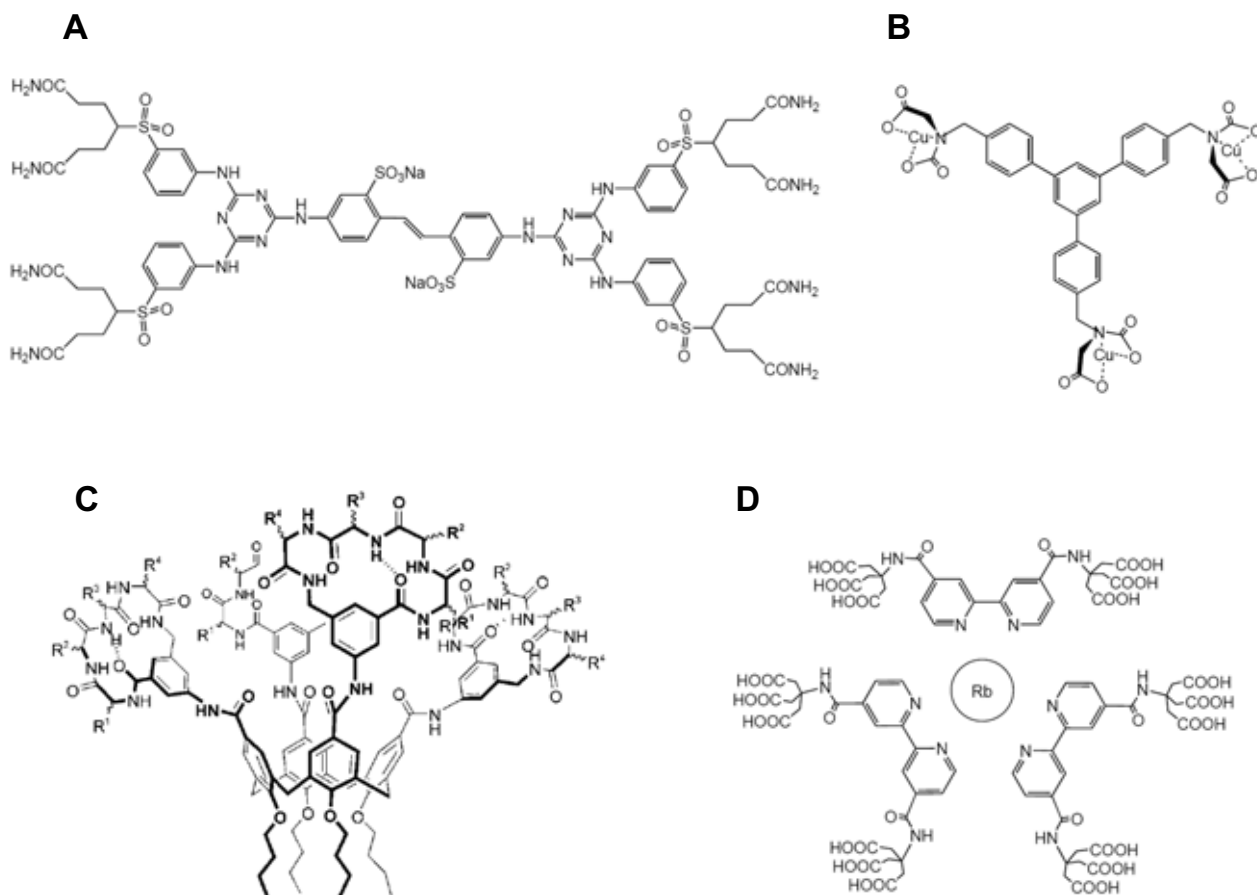
Of course, in the case of mutations affecting key residues for the direct function, or hotspots implied in protein-biomolecule interactions, regardless of whether they enable correct folding, functionality cannot be rescued through pharmacological chaperones. For these cases and many others p53 activity restoration has been approached through genetic therapy, using suitable viral vectors to introduce intact or improved *p53* genes into the tumor tissue.<sup>165-168</sup> Success has been reported for some clinical trials; however, this strategy is still too limited for immediate practical application.

## 1 u ti a enc

The design of drug-like molecules of low molecular weight which can selectively and efficiently recognize a large featureless protein surface is sometimes unattainable. These limitations have inspired novel rational strategies in which larger scaffolds are employed to attach multiple anchoring groups that are specially selected to interact with unique features of the protein surface. The area covered by these potentials ligands is larger and, even if the protein-protein interface does not present any relevant anchoring point, other residues out of the boundary can also be used. In addition to the anchoring groups, the scaffold itself can be made up with functional attributes which further contribute to a tighter binding.

The laboratory copies once again tricks from wise Nature, applying the concept of *multivalency*<sup>169</sup> in the design of these ligand molecules. That is, a tighter interaction can be obtained from the simultaneous contribution of multiple low to moderate affinity interactions. In addition to increase the kinetic and thermodynamic stability of the complex,<sup>170</sup> multiple contributions also lead to a higher specificity and to easier induction of conformational changes.<sup>169,171</sup>

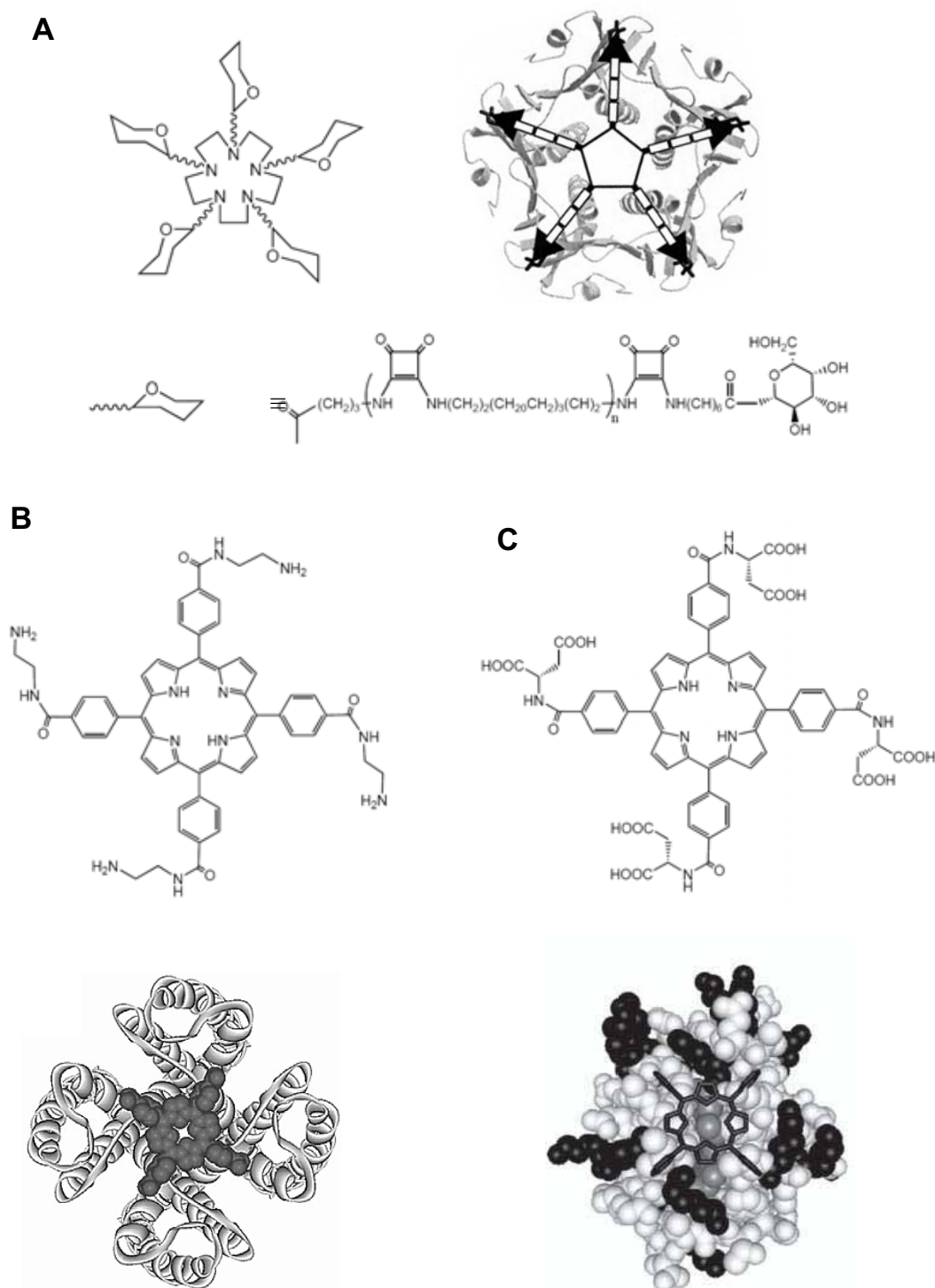
A multivalent ligand consists of a main core, the *scaffold*, bearing several covalent connections, *linkers* or *spacers*, to the peripheral *ligating* (binding) units.<sup>172</sup> These kind of constructions have been widely applied in drug discovery, especially in the fight against infectious diseases; viruses and bacteria adhere to the cell surface by multiple lectin-sugar recognition sites, thus multivalent inhibitors have been designed to “beat Nature at her own game” (**Figure 8A**).<sup>171</sup> The same strategy has been applied in the design of effectors that promote responses via signal transduction by clustering surface receptors.<sup>93-95</sup> In both cases, the ligand is designed to exploit the sugar-binding pocket on the target multivalent receptor.



**Figure 7.** (A) Bistriazine compound that inhibits the tetrameric assembly of RS virus fusion protein.<sup>173</sup> (B)  $\text{Cu}^{2+}$ -IDA-containing trivalent ligand that targets the imidazole side-chain of three histidines at the surface of carbonic anhydrase.<sup>174</sup> (C) Calixarene-based ligand functionalized with multiple recognition elements designed for disrupting the interaction of the homodimer PDGF with its receptor.<sup>175</sup> (D)  $[\text{Rb}(\text{bpy})_3]$  complex for the recognition of the positively charged cytochrome c surface.<sup>176</sup>

In the surface recognition challenge, the multivalent approximation has also been successfully applied, usually by taking advantage of the symmetry of the target molecule (**Figure 7A**, **Figure 8A** and **B**). Branching chelators are selected on the basis of the chemical characteristics of the target residues on the surface. The scaffold, whether natural<sup>95,171</sup> or synthetic, can also actively participate in the binding or simply remain as a rigid support whose entropic penalty is low.

A small selection of reported multivalent ligands for surface recognition is shown in **Figure 7** and **Figure 8**; many other scaffold, both smaller and larger, have been also described. Despite their huge size –far from that of the classic drug-like molecule– these ligands have proven successful in their inhibitory tasks *in vitro* and *in vivo*, thus opening new doors in the “biomimetic approach” in rational drug design.

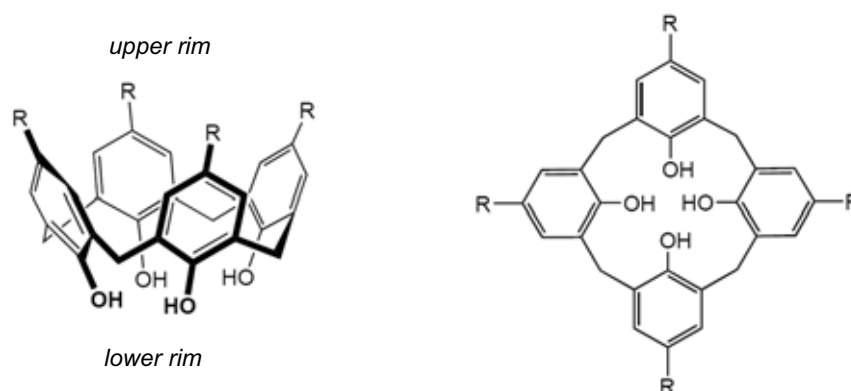


**Figure 8.** (A) Pentavalent saccharide-inhibitor for the receptor binding site of cholera toxin B-pentamer; at the right, a schematic representation of the ligand binding mode.<sup>177</sup> (B) Porphyrine-based tetraivalent ligand for blocking human  $K_v1.3$  potassium channel; at the bottom the complex protein-ligand.<sup>178</sup> (C) Porphyrine-based tetraivalent ligand TPP derivatives for the recognition of cytochrome c exterior surface; at the bottom the complex protein-ligand.<sup>179</sup>

## 1 Calixarene

Calixarenes have become one of the most popular scaffolds currently employed in the design of multivalent ligands, not only for protein recognition but also for other many scopes.<sup>180,181</sup>

*Calixarenes* –or more accurately said, *calix[n]arenes*–<sup>182</sup> are synthetic macrocycles derived from the condensation of phenol and formaldehyde. Its chemistry is well established; depending on the experimental conditions, they can be formed by four to eight aromatic units arranged cyclically (**Figure 9**). The phenol group defines the *narrow* or *lower rim* of the molecule, and is an excellent point for diversity. The *wide* or *upper rim* of the molecule presents the reactive *para*-position of the aromatic rings where many other functional units can be also introduced. The stereochemical orientation of the ligating arms can be properly tuned by shaping the calixarene macrocycle in any of its possible conformations.<sup>183</sup> For instance, pinched and winged conformations can be avoided by conveniently bridging the positions at the lower or the upper rims.<sup>184,185</sup>

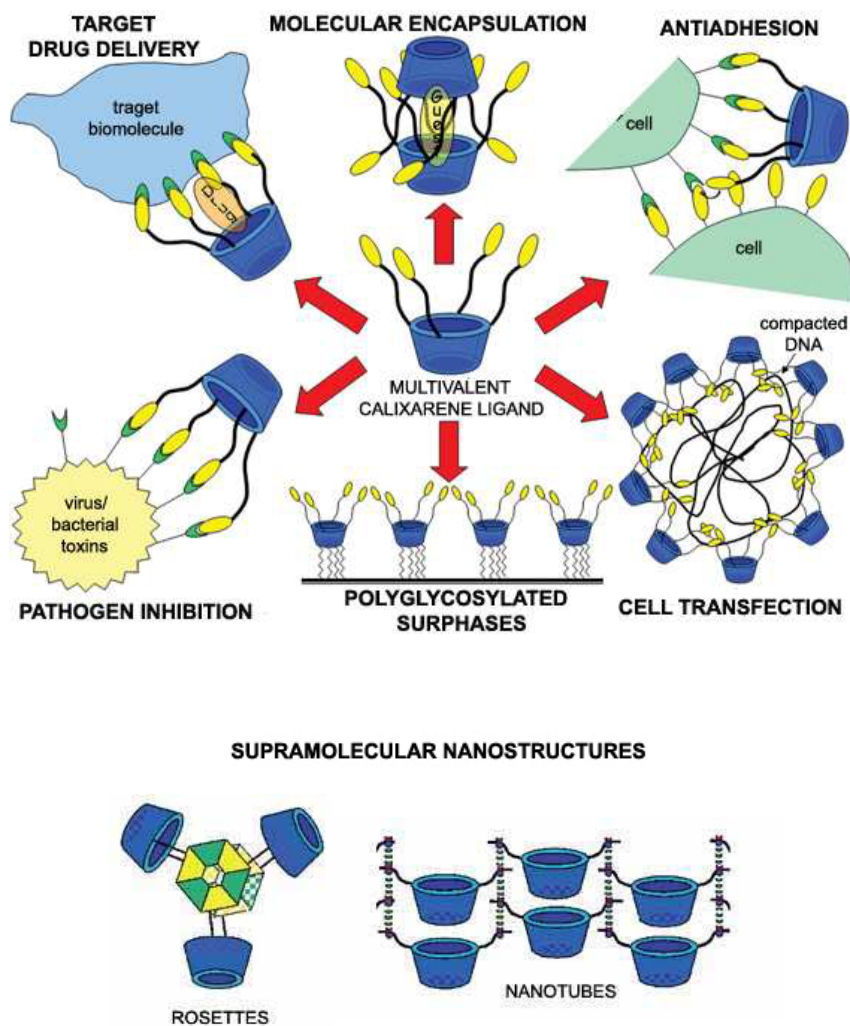


**Figure 9.** Calix[4]arene. At the left, as a cone, and at the right, in an upper view.

Traditionally, calix[n]arenes, like cyclodextrins and crown ethers, have displayed a central role in supramolecular chemistry, mainly as hosts –“model receptors”– for ions, neutral molecules and other organic guests. Far from a basic model of study, those properties have been exploited into the real world for the construction of electrodes and membranes for transport, for the production of selective sensors for analytical applications and medical diagnosis, or for the decontamination of waste water among others.<sup>186,187</sup>

More recently, calixarenes have been used as platforms for the synthesis of multivalent ligands. They are very attractive synthetic molecules since possess several variable reactive positions that can be functionalized selectively in diversity-oriented syntheses; in addition, the flexibility or rigidity of the calixarene scaffold can be modulated at will. Hence, ornamented calixarene have proven success as, for instance, biomimetic receptors for the encapsulation of guest species of biological interest (e.g. insoluble drugs, steroids...),<sup>188,189</sup> building blocks for the non-covalent synthesis of nanostructures,<sup>180,181</sup> multivalent ligands able to modulate biological processes (antiviral,<sup>190</sup>

antibacterial, channel blockers,<sup>191</sup> protein inhibitors,<sup>192</sup> protein surface binders<sup>193</sup>...), artificial enzymatic activity,<sup>194</sup> detection and immobilization of proteins in microarrays,<sup>195</sup> DNA-transfection properties<sup>196</sup>... (**Figure 10**) and the list is unceasingly growing.<sup>172,180,188</sup> Hence, calixarenes are very promising tools in the field of the *bionanotechnology*.



**Figure 10.** Multivalent calixarene ligands in Bionanotechnology.<sup>172</sup>



## i o r a p h

1. Cusick, M. E., Klitgord, N., Vidal, M. & Hill, D. E. Interactome: gateway into systems biology. *Hum. Mol. Genet.* **14 Spec No. 2**, R171-R181 (2005).
2. He, X. & Zhang, J. Why do hubs tend to be essential in protein networks? *PLoS. Genet.* **2**, e88 (2006).
3. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197-208 (2005).
4. Keskin, O. & Nussinov, R. Similar binding sites and different partners: implications to shared proteins in cellular pathways. *Structure.* **15**, 341-354 (2007).
5. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321-331 (1999).
6. Dyson, H. J. & Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54-60 (2002).
7. Uversky, V. N. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **11**, 739-756 (2002).
8. Plaxco, K. W. & Gross, M. Cell biology. The importance of being unfolded. *Nature* **386**, 657, 659 (1997).
9. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **272**, 5129-5148 (2005).
10. Phizicky, E. M. & Fields, S. Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59**, 94-123 (1995).
11. Sali, A., Glaeser, R., Earnest, T. & Baumeister, W. From words to literature in structural proteomics. *Nature* **422**, 216-225 (2003).
12. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105-153 (2000).
13. Goodsell, D. S. Inside a living cell. *Trends Biochem. Sci.* **16**, 203-206 (1991).
14. Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A* **93**, 13-20 (1996).
15. Ali, M. H. & Imperiali, B. Protein oligomerization: how and why. *Bioorg. Med. Chem.* **13**, 5013-5020 (2005).
16. Keskin, O., Gursoy, A., Ma, B. & Nussinov, R. Principles of Protein-Protein Interactions: What are the Preferred Ways For Proteins To Interact? *Chem. Rev.* **108**, 1225-1244 (2008).
17. Nooren, I. M. & Thornton, J. M. Diversity of protein-protein interactions. *EMBO J.* **22**, 3486-3492 (2003).
18. Nooren, I. M. & Thornton, J. M. Structural characterisation and functional significance of transient protein-protein interactions. *J. Mol. Biol.* **325**, 991-1018 (2003).
19. Lo Conte L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177-2198 (1999).
20. Jones, S. & Thornton, J. M. Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.* **272**, 121-132 (1997).
21. Keskin, O., Ma, B. & Nussinov, R. Hot regions in protein-protein interactions: the organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.* **345**, 1281-1294 (2005).
22. Clackson, T. & Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383-386 (1995).
23. Stites, W. E. Protein-Protein Interactions: Interface Structure, Binding Thermodynamics, and Mutational Analysis. *Chem. Rev.* **97**, 1233-1250 (1997).
24. Reichmann, D. *et al.* The modular architecture of protein-protein binding interfaces. *Proc. Natl. Acad. Sci. U. S. A* **102**, 57-62 (2005).
25. Chakrabarti, P. & Janin, J. Dissecting protein-protein recognition sites. *Proteins* **47**, 334-343 (2002).
26. Raschke, T. M. Water structure and interactions with protein surfaces. *Curr. Opin. Struct. Biol.* **16**, 152-159 (2006).
27. Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H. & Schreiber, G. The molecular architecture of protein-protein binding sites. *Curr. Opin. Struct. Biol.* **17**, 67-76 (2007).
28. Hubbard, S. J. & Argos, P. Cavities and packing at protein interfaces. *Protein Sci.* **3**, 2194-2206 (1994).
29. Hu, Z., Ma, B., Wolfson, H. & Nussinov, R. Conservation of polar residues as hot spots at protein interfaces. *Proteins* **39**, 331-342 (2000).
30. Cho, K. I., Lee, K., Lee, K. H., Kim, D. & Lee, D. Specificity of molecular interactions in transient protein-protein interaction interfaces. *Proteins* **65**, 593-606 (2006).
31. Gunasekaran, K. & Nussinov, R. How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *J. Mol. Biol.* **365**, 257-273 (2007).
32. Stelzl, U. *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957-968 (2005).
33. Rual, J. F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173-1178 (2005).
34. Ruffner, H., Bauer, A. & Bouwmeester, T. Human protein-protein interaction networks and the value for drug discovery. *Drug Discov. Today* **12**, 709-716 (2007).
35. Hanash, S. Disease proteomics. *Nature* **422**, 226-232 (2003).



36. Hsu, W. T., Pang, C. N., Sheetal, J. & Wilkins, M. R. Protein-protein interactions and disease: use of *S. cerevisiae* as a model system. *Biochim. Biophys. Acta* **1774**, 838-847 (2007).
37. Goh, K. I. *et al.* The human disease network. *Proc. Natl. Acad. Sci. U. S. A* **104**, 8685-8690 (2007).
38. Fishman, M. C. & Porter, J. A. Pharmaceuticals: a new grammar for drug discovery. *Nature* **437**, 491-493 (2005).
39. Ellis, R. J. & Pinheiro, T. J. Danger - misfolding proteins. *Nature* **416**, 483-484 (2002).
40. Brooks, D. A. Introduction: molecular chaperones of the ER: their role in protein folding and genetic disease. *Semin. Cell Dev. Biol.* **10**, 441-442 (1999).
41. Trombetta, E. S. & Parodi, A. J. Quality control and protein folding in the secretory pathway. *Annu. Rev. Cell Dev. Biol.* **19**, 649-676 (2003).
42. Werner, E. D., Brodsky, J. L. & McCracken, A. A. Proteasome-dependent endoplasmic reticulum-associated protein degradation: an unconventional route to a familiar fate. *Proc. Natl. Acad. Sci. U. S. A* **93**, 13797-13801 (1996).
43. Schubert, U. *et al.* Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**, 770-774 (2000).
44. Chaudhuri, T. K. & Paul, S. Protein-misfolding diseases and chaperone-based therapeutic approaches. *FEBS J.* **273**, 1331-1349 (2006).
45. Stefani, M. & Dobson, C. M. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81**, 678-699 (2003).
46. Dobson, C. M. Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329-332 (1999).
47. Martin, J. B. Molecular basis of the neurodegenerative disorders. *N. Engl. J. Med.* **340**, 1970-1980 (1999).
48. Cohen, F. E. & Kelly, J. W. Therapeutic approaches to protein-misfolding diseases. *Nature* **426**, 905-909 (2003).
49. Hubmacher, D., Tiedemann, K. & Reinhardt, D. P. Fibrillins: from biogenesis of microfibrils to signaling functions. *Curr. Top. Dev. Biol.* **75**, 93-123 (2006).
50. Cole, W. G. Collagen genes: mutations affecting collagen structure and expression. *Prog. Nucleic Acid Res. Mol. Biol.* **47**, 29-80 (1994).
51. Vandenberg, P. Molecular basis of heritable connective tissue disease. *Biochem. Med. Metab Biol.* **49**, 1-12 (1993).
52. Prockop, D. J. What holds us together? Why do some of us fall apart? What can we do about it? *Matrix Biol.* **16**, 519-528 (1998).
53. Strong, M. & Eisenberg, D. The protein network as a tool for finding novel drug targets. *Prog. Drug Res.* **64**, 191, 193-191, 215 (2007).
54. Zutshi, R., Brickner, M. & Chmielewski, J. Inhibiting the assembly of protein-protein interfaces. *Curr. Opin. Chem. Biol.* **2**, 62-66 (1998).
55. Cochran, A. G. Protein-protein interfaces: mimics and inhibitors. *Curr. Opin. Chem. Biol.* **5**, 654-659 (2001).
56. Kishan, K. V. Structural biology, protein conformations and drug designing. *Curr. Protein Pept. Sci.* **8**, 376-380 (2007).
57. Fry, D. C. & Vassilev, L. T. Targeting protein-protein interactions for cancer therapy. *J. Mol. Med.* **83**, 955-963 (2005).
58. Veselovsky, A. V. *et al.* Protein-protein interactions: mechanisms and modification by drugs. *J. Mol. Recognit.* **15**, 405-422 (2002).
59. Block, P., Weskamp, N., Wolf, A. & Klebe, G. Strategies to search and design stabilizers of protein-protein interactions: a feasibility study. *Proteins* **68**, 170-186 (2007).
60. Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.* **3**, 301-317 (2004).
61. Cochran, A. G. Antagonists of protein-protein interactions. *Chem. Biol.* **7**, R85-R94 (2000).
62. Pagliaro, L. *et al.* Emerging classes of protein-protein interaction inhibitors and new tools for their development. *Curr. Opin. Chem. Biol.* **8**, 442-449 (2004).
63. Chene, P. Drugs targeting protein-protein interactions. *ChemMedChem.* **1**, 400-411 (2006).
64. Huang, Z. The chemical biology of apoptosis. Exploring protein-protein interactions and the life and death of cells with small molecules. *Chem. Biol.* **9**, 1059-1072 (2002).
65. Huang, Z. Structural chemistry and therapeutic intervention of protein-protein interactions in immune response, human immunodeficiency virus entry, and apoptosis. *Pharmacol. Ther.* **86**, 201-215 (2000).
66. Perez-Montfort, R., de Gomez-Puyou, M. T. & Gomez-Puyou, A. The interfaces of oligomeric proteins as targets for drug design against enzymes from parasites. *Curr. Top. Med. Chem.* **2**, 457-470 (2002).
67. Dall'Acqua, W. & Carter, P. Antibody engineering. *Curr. Opin. Struct. Biol.* **8**, 443-450 (1998).
68. Vaughan, T. J., Osbourn, J. K. & Tempest, P. R. Human antibodies by design. *Nat. Biotechnol.* **16**, 535-539 (1998).
69. Junghans, R. P. *et al.* Anti-Tac-H, a humanized antibody to the interleukin 2 receptor with new features for immunotherapy in malignant and immune disorders. *Cancer Res.* **50**, 1495-1502 (1990).
70. Braisted, A. C. & Wells, J. A. Minimizing a binding domain from protein A. *Proc. Natl. Acad. Sci. U. S. A* **93**, 5688-5692 (1996).
71. Domingues, H., Cregut, D., Sebald, W., Oschkinat, H. & Serrano, L. Rational design of a GCN4-derived mimetic of interleukin-4. *Nat. Struct. Biol.* **6**, 652-656 (1999).
72. Nygren, P. A. & Uhlen, M. Scaffolds for engineering novel binding sites in proteins. *Curr. Opin. Struct. Biol.* **7**, 463-469 (1997).
73. Cooper, W. J. & Waters, M. L. Molecular recognition with designed peptides and proteins. *Curr. Opin. Chem. Biol.* **9**, 627-631 (2005).

74. Osborne, S. E. & Ellington, A. D. Nucleic Acid Selection and the Challenge of Combinatorial Chemistry. *Chem. Rev.* **97**, 349-370 (1997).
75. Jing, N. & Hogan, M. E. Structure-activity of tetrad-forming oligonucleotides as a potent anti-HIV therapeutic drug. *J. Biol. Chem.* **273**, 34992-34999 (1998).
76. Kay, B. K., Kurakin, A. V. & Hyde-DeRuyscher, R. From peptides to drugs via phage display. *Drug Discov. Today* **3**, 370-378 (1998).
77. Sidhu, S. S., Fairbrother, W. J. & Deshayes, K. Exploring protein-protein interactions with phage display. *Chembiochem.* **4**, 14-25 (2003).
78. Fletcher, S. & Hamilton, A. D. Protein-protein interaction inhibitors: small molecules from screening techniques. *Curr. Top. Med. Chem.* **7**, 922-927 (2007).
79. Che, Y., Brooks, B. R. & Marshall, G. R. Development of small molecules designed to modulate protein-protein interactions. *J. Comput. Aided Mol. Des* **20**, 109-130 (2006).
80. Gonzalez-Ruiz, D. & Gohlke, H. Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr. Med. Chem.* **13**, 2607-2625 (2006).
81. Zhong, S., Macias, A. T. & MacKerell, A. D., Jr. Computational identification of inhibitors of protein-protein interactions. *Curr. Top. Med. Chem.* **7**, 63-82 (2007).
82. Ivanov, A. S. *et al.* Protein-protein interactions as new targets for drug design: virtual and experimental approaches. *J. Bioinform. Comput. Biol.* **5**, 579-592 (2007).
83. McDowell, R. S. *et al.* From Peptide to Non-Peptide. 1. The Elucidation of a Bioactive Conformation of the Arginine-Glycine-Aspartic Acid Recognition Sequence. *J. Am. Chem. Soc.* **116**, 5069-5076 (1994).
84. McDowell, R. S. *et al.* From Peptide to Non-Peptide. 2. The de Novo Design of Potent, Non-peptidal Inhibitors of Platelet Aggregation Based on a Benzodiazepinedione Scaffold. *J. Am. Chem. Soc.* **116**, 5077-5083 (1994).
85. Sillerud, L. O. & Larson, R. S. Design and structure of peptide and peptidomimetic antagonists of protein-protein interaction. *Curr. Protein Pept. Sci.* **6**, 151-169 (2005).
86. Hershberger, S. J., Lee, S. G. & Chmielewski, J. Scaffolds for blocking protein-protein interactions. *Curr. Top. Med. Chem.* **7**, 928-942 (2007).
87. Fletcher, S. & Hamilton, A. D. Protein surface recognition and proteomimetics: mimics of protein surface structure and function. *Curr. Opin. Chem. Biol.* **9**, 632-638 (2005).
88. Oldfield, C. J. *et al.* Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* **44**, 12454-12470 (2005).
89. Sillerud, L. O. & Larson, R. S. Design and structure of peptide and peptidomimetic antagonists of protein-protein interaction. *Curr. Protein Pept. Sci.* **6**, 151-169 (2005).
90. Davis, J. M., Tsou, L. K. & Hamilton, A. D. Synthetic non-peptide mimetics of alpha-helices. *Chem. Soc. Rev.* **36**, 326-334 (2007).
91. de Vega, M. J., Martin-Martinez, M. & Gonzalez-Muniz, R. Modulation of protein-protein interactions by stabilizing/mimicking protein secondary structure elements. *Curr. Top. Med. Chem.* **7**, 33-62 (2007).
92. Neduva, V. & Russell, R. B. Peptides mediating interaction networks: new leads at last. *Curr. Opin. Biotechnol.* **17**, 465-471 (2006).
93. Germain, R. N. T-cell signaling: the importance of receptor clustering. *Curr. Biol.* **7**, R640-R644 (1997).
94. Klemm, J. D., Schreiber, S. L. & Crabtree, G. R. Dimerization as a regulatory mechanism in signal transduction. *Annu. Rev. Immunol.* **16**, 569-592 (1998).
95. Gestwicki, J. E., Cairo, C. W., Strong, L. E., Oetjen, K. A. & Kiessling, L. L. Influencing receptor-ligand binding mechanisms with multivalent ligand architecture. *J. Am. Chem. Soc.* **124**, 14922-14933 (2002).
96. Amara, J. F. *et al.* A versatile synthetic dimerizer for the regulation of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A* **94**, 10618-10623 (1997).
97. Blau, C. A., Peterson, K. R., Drachman, J. G. & Spencer, D. M. A proliferation switch for genetically modified cells. *Proc. Natl. Acad. Sci. U. S. A* **94**, 3076-3081 (1997).
98. Keenan, T. *et al.* Synthesis and activity of bivalent FKBP12 ligands for the regulated dimerization of proteins. *Bioorg. Med. Chem.* **6**, 1309-1335 (1998).
99. MacCorkle, R. A., Freeman, K. W. & Spencer, D. M. Synthetic activation of caspases: artificial death switches. *Proc. Natl. Acad. Sci. U. S. A* **95**, 3655-3660 (1998).
100. Sciarretta, K. L., Gordon, D. J. & Meredith, S. C. Peptide-based inhibitors of amyloid assembly. *Methods Enzymol.* **413**, 273-312 (2006).
101. Smith, D. F., Whitesell, L. & Katsanis, E. Molecular chaperones: biology and prospects for pharmacological intervention. *Pharmacol. Rev.* **50**, 493-514 (1998).
102. Barral, J. M., Broadley, S. A., Schaffar, G. & Hartl, F. U. Roles of molecular chaperones in protein misfolding diseases. *Semin. Cell Dev. Biol.* **15**, 17-29 (2004).
103. Rochet, J. C. Novel therapeutic strategies for the treatment of protein-misfolding diseases. *Expert. Rev. Mol. Med.* **9**, 1-34 (2007).
104. Miroy, G. J. *et al.* Inhibiting transthyretin amyloid fibril formation via protein stabilization. *Proc. Natl. Acad. Sci. U. S. A* **93**, 15051-15056 (1996).
105. Perlmutter, D. H. Chemical chaperones: a pharmacological strategy for disorders of protein folding and trafficking. *Pediatr. Res.* **52**, 832-836 (2002).
106. Ulloa-Aguirre, A., Janovick, J. A., Brothers, S. P. & Conn, P. M. Pharmacologic rescue of conformationally-defective proteins: implications for the treatment of human disease. *Traffic.* **5**, 821-837 (2004).
107. Singh, L. R., Chen, X., Kozich, V. & Kruger, W. D. Chemical chaperone rescue of mutant human cystathionine beta-synthase. *Mol. Genet. Metab* **91**, 335-342 (2007).

108. Arakawa, T., Ejima, D., Kita, Y. & Tsumoto, K. Small molecule pharmacological chaperones: From thermodynamic stabilization to pharmaceutical drugs. *Biochim. Biophys. Acta* **1764**, 1677-1687 (2006).
109. Perlmutter, D. H. Chemical chaperones: a pharmacological strategy for disorders of protein folding and trafficking. *Pediatr. Res.* **52**, 832-836 (2002).
110. Papp, E. & Csermely, P. Chemical chaperones: mechanisms of action and potential use. *Handb. Exp. Pharmacol.* 405-416 (2006).
111. Gregersen, N. Protein misfolding disorders: pathogenesis and intervention. *J. Inherit. Metab. Dis.* **29**, 456-470 (2006).
112. Conn, P. M., Leanos-Miranda, A. & Janovick, J. A. Protein origami: therapeutic rescue of misfolded gene products. *Mol. Interv.* **2**, 308-316 (2002).
113. Loo, T. W. & Clarke, D. M. Chemical and pharmacological chaperones as new therapeutic agents. *Expert. Rev. Mol. Med.* **9**, 1-18 (2007).
114. Amaral, M. D. Therapy through chaperones: sense or antisense? Cystic fibrosis as a model disease. *J. Inherit. Metab. Dis.* **29**, 477-487 (2006).
115. Wang, Y., Bartlett, M. C., Loo, T. W. & Clarke, D. M. Specific rescue of cystic fibrosis transmembrane conductance regulator processing mutants using pharmacological chaperones. *Mol. Pharmacol.* **70**, 297-302 (2006).
116. Conn, P. M., Ulloa-Aguirre, A., Ito, J. & Janovick, J. A. G protein-coupled receptor trafficking in health and disease: lessons learned to prepare for therapeutic mutant rescue in vivo. *Pharmacol. Rev.* **59**, 225-250 (2007).
117. Robben, J. H. & Deen, P. M. Pharmacological chaperones in nephrogenic diabetes insipidus: possibilities for clinical application. *BioDrugs.* **21**, 157-166 (2007).
118. Stankunas, K. *et al.* Rescue of degradation-prone mutants of the FK506-rapamycin binding (FRB) protein with chemical ligands. *Chembiochem.* **8**, 1162-1169 (2007).
119. Wildin, R. S., Antush, M. J., Bennett, R. L., Schoof, J. M. & Scott, C. R. Heterogeneous AVPR2 gene mutations in congenital nephrogenic diabetes insipidus. *Am. J. Hum. Genet.* **55**, 266-277 (1994).
120. Janovick, J. A., Maya-Nunez, G. & Conn, P. M. Rescue of hypogonadotropic hypogonadism-causing and manufactured GnRH receptor mutants by a specific protein-folding template: misrouted proteins as a novel disease etiology and therapeutic target. *J. Clin. Endocrinol. Metab.* **87**, 3255-3262 (2002).
121. Petaja-Repo, U. E. *et al.* Ligands act as pharmacological chaperones and increase the efficiency of delta opioid receptor maturation. *EMBO J.* **21**, 1628-1637 (2002).
122. Chen, J. K., Taipale, J., Young, K. E., Maiti, T. & Beachy, P. A. Small molecule modulation of Smoothed activity. *Proc. Natl. Acad. Sci. U. S. A* **99**, 14071-14076 (2002).
123. Levine, A. J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323-331 (1997).
124. Sengupta, S. & Harris, C. C. p53: traffic cop at the crossroads of DNA repair and recombination. *Nat. Rev. Mol. Cell Biol.* **6**, 44-55 (2005).
125. Lane, D. P. Cancer. p53, guardian of the genome. *Nature* **358**, 15-16 (1992).
126. Prives, C. & Hall, P. A. The p53 pathway. *J. Pathol.* **187**, 112-126 (1999).
127. Levine, A. J., Hu, W. & Feng, Z. The P53 pathway: what questions remain to be explored? *Cell Death. Differ.* **13**, 1027-1036 (2006).
128. Tidow, H. *et al.* Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. *Proc. Natl. Acad. Sci. U. S. A* **104**, 12324-12329 (2007).
129. Wang, P. *et al.* p53 domains: structure, oligomerization, and transformation. *Mol. Cell Biol.* **14**, 5182-5191 (1994).
130. Bell, S., Klein, C., Muller, L., Hansen, S. & Buchner, J. p53 contains large unstructured regions in its native state. *J. Mol. Biol.* **322**, 917-927 (2002).
131. Joerger, A. C. & Fersht, A. R. Structure-function-rescue: the diverse nature of common p53 cancer mutants. *Oncogene* **26**, 2226-2242 (2007).
132. Braithwaite, A. W., Del, S. G. & Lu, X. Some p53-binding proteins that can function as arbiters of life and death. *Cell Death. Differ.* **13**, 984-993 (2006).
133. Bode, A. M. & Dong, Z. Post-translational modification of p53 in tumorigenesis. *Nat. Rev. Cancer* **4**, 793-805 (2004).
134. Lavin, M. F. & Gueven, N. The complexity of p53 stabilization and activation. *Cell Death. Differ.* **13**, 941-950 (2006).
135. Levine, A. J., Momand, J. & Finlay, C. A. The p53 tumour suppressor gene. *Nature* **351**, 453-456 (1991).
136. Greenblatt, M. S., Bennett, W. P., Hollstein, M. & Harris, C. C. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res.* **54**, 4855-4878 (1994).
137. Petitjean, A. *et al.* Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum. Mutat.* **28**, 622-629 (2007).
138. Asker, C., Wiman, K. G. & Selivanova, G. p53-induced apoptosis as a safeguard against cancer. *Biochem. Biophys. Res. Commun.* **265**, 1-6 (1999).
139. Sigal, A. & Rotter, V. Oncogenic mutations of the p53 tumor suppressor: the demons of the guardian of the genome. *Cancer Res.* **60**, 6788-6793 (2000).
140. Picksley, S. M., Spicer, J. F., Barnes, D. M. & Lane, D. P. The p53-MDM2 interaction in a cancer-prone family, and the identification of a novel therapeutic target. *Acta Oncol.* **35**, 429-434 (1996).
141. Zheleva, D. I., Lane, D. P. & Fischer, P. M. The p53-Mdm2 pathway: targets for the development of new anticancer therapeutics. *Mini. Rev. Med. Chem.* **3**, 257-270 (2003).
142. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307-310 (2000).
143. Ariumi, Y., Kaida, A., Hatanaka, M. & Shimotohno, K. Functional cross-talk of HIV-1 Tat with p53 through its C-terminal domain. *Biochem. Biophys. Res. Commun.* **287**, 556-561 (2001).

144. Li, C. J., Wang, C., Friedman, D. J. & Pardee, A. B. Reciprocal modulations between p53 and Tat of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A* **92**, 5461-5464 (1995).
145. Bullock, A. N. & Fersht, A. R. Rescuing the function of mutant p53. *Nat. Rev. Cancer* **1**, 68-76 (2001).
146. Selivanova, G. & Wiman, K. G. Reactivation of mutant p53: molecular mechanisms and therapeutic potential. *Oncogene* **26**, 2243-2254 (2007).
147. Hupp, T. R., Lane, D. P. & Ball, K. L. Strategies for manipulating the p53 pathway in the treatment of human cancer. *Biochem. J.* **352 Pt 1**, 1-17 (2000).
148. Lane, D. P. & Lain, S. Therapeutic exploitation of the p53 pathway. *Trends Mol. Med.* **8**, S38-S42 (2002).
149. Foster, B. A., Coffey, H. A., Morin, M. J. & Rastinejad, F. Pharmacological rescue of mutant p53 conformation and function. *Science* **286**, 2507-2510 (1999).
150. Maurici, D. *et al.* Amifostine (WR2721) restores transcriptional activity of specific p53 mutant proteins in a yeast functional assay. *Oncogene* **20**, 3533-3540 (2001).
151. Issaeva, N. *et al.* Rescue of mutants of the tumor suppressor p53 in cancer cells by a designed peptide. *Proc. Natl. Acad. Sci. U. S. A* **100**, 13303-13307 (2003).
152. Vitali, R. *et al.* Activation of p53-dependent responses in tumor cells treated with a PARC-interacting peptide. *Biochem. Biophys. Res. Commun.* **368**, 350-356 (2008).
153. Chene, P. Inhibiting the p53-MDM2 interaction: an important target for cancer therapy. *Nat. Rev. Cancer* **3**, 102-109 (2003).
154. Fry, D. C., Graves, B. & Vassilev, L. T. Development of E3-substrate (MDM2-p53)-binding inhibitors: structural aspects. *Methods Enzymol.* **399**, 622-633 (2005).
155. Vassilev, L. T. *et al.* In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844-848 (2004).
156. Murray, J. K. & Gellman, S. H. Targeting protein-protein interactions: lessons from p53/MDM2. *Biopolymers* **88**, 657-686 (2007).
157. Blankenship, J. W. & Dawson, P. E. Thermodynamics of a designed protein catenane. *J. Mol. Biol.* **327**, 537-548 (2003).
158. Brokx, R. D., Bolewska-Pedyczak, E. & Gariepy, J. A stable human p53 heterotetramer based on constructive charge interactions within the tetramerization domain. *J. Biol. Chem.* **278**, 2327-2332 (2003).
159. Conseiller, E. *et al.* CTS1: a p53-derived chimeric tumor suppressor gene with enhanced in vitro apoptotic properties. *J. Clin. Invest* **101**, 120-127 (1998).
160. Davison, T. S. *et al.* Structure and functionality of a designed p53 dimer. *J. Mol. Biol.* **307**, 605-617 (2001).
161. Poon, G. M., Brokx, R. D., Sung, M. & Gariepy, J. Tandem dimerization of the human p53 tetramerization domain stabilizes a primary dimer intermediate and dramatically enhances its oligomeric stability. *J. Mol. Biol.* **365**, 1217-1231 (2007).
162. Poon, G. M. Enhancement of oligomeric stability by covalent linkage and its application to the human p53tet domain: thermodynamics and biological implications. *Biochem. Soc. Trans.* **35**, 1574-1578 (2007).
163. Waterman, M. J., Waterman, J. L. & Halazonetis, T. D. An engineered four-stranded coiled coil substitutes for the tetramerization domain of wild-type p53 and alleviates transdominant inhibition by tumor-derived p53 mutants. *Cancer Res.* **56**, 158-163 (1996).
164. Yan, L. Z. & Dawson, P. E. Design and Synthesis of a Protein Catenane This work was supported by The Skaggs Institute for Chemical Biology, The Sloan Foundation, and NIH-GM570132 (PED). We thank Dr. Songpong Deechongkit for assistance with the analytical ultracentrifuge and CD measurements. *Angew. Chem. Int. Ed Engl.* **40**, 3625-3627 (2001).
165. Bykov, V. J. & Wiman, K. G. Novel cancer therapy by reactivation of the p53 apoptosis pathway. *Ann. Med.* **35**, 458-465 (2003).
166. Roth, J. A. *et al.* Retrovirus-mediated wild-type p53 gene transfer to tumors of patients with lung cancer. *Nat. Med.* **2**, 985-991 (1996).
167. Roth, J. A. p53 Prognostication: Paradigm or Paradox? *Clin Cancer Res* **5**, 3345 (1999).
168. Roth, J. A. Adenovirus p53 gene therapy. *Expert Opinion on Biological Therapy* **6**, 55-61 (2006).
169. Mammen, M., Choi, S. K. & Whitesides, G. M. Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angewandte Chemie International Edition* **37**, 2754-2794 (1998).
170. Mulder, A., Huskens, J. & Reinhoudt, D. N. Multivalency in supramolecular chemistry and nanofabrication. *Org. Biomol. Chem.* **2**, 3409-3424 (2004).
171. Fan, E. & Merritt, E. A. Combating infectious diseases through multivalent design. *Curr. Drug Targets. Infect. Disord.* **2**, 161-167 (2002).
172. Baldini, L., Casnati, A., Sansone, F. & Ungaro, R. Calixarene-based multivalent ligands. *Chem. Soc. Rev.* **36**, 254-266 (2007).
173. Ding, W. D. *et al.* Novel and specific respiratory syncytial virus inhibitors that target virus fusion. *J. Med. Chem.* **41**, 2671-2675 (1998).
174. Fazal, M. A., Roy, B. C., Sun, S., Mallik, S. & Rodgers, K. R. Surface recognition of a protein using designed transition metal complexes. *J. Am. Chem. Soc.* **123**, 6283-6290 (2001).
175. Sebti, S. M. & Hamilton, A. D. Design of growth factor antagonists with antiangiogenic and antitumor properties. *Oncogene* **19**, 6566-6573 (2000).
176. Takashima, H., Shinkai, S. & Hamachi, I. Ru(bpy)<sub>3</sub>-based artificial receptors toward a protein surface: selective binding and efficient photoreduction of cytochrome c. *Chem. Comm.* 2345-2346 (1999).
177. Fan, E. *et al.* High-Affinity Pentavalent Ligands of Escherichia coli Heat-Labile Enterotoxin by Modular Structure-Based Design. *Journal of the American Chemical Society* **122**, 2663-2664 (2000).

178. Gradl, S. N., Felix, J. P., Isacoff, E. Y., Garcia, M. L. & Trauner, D. Protein surface recognition by rational design: nanomolar ligands for potassium channels. *J. Am. Chem. Soc.* **125**, 12668-12669 (2003).
179. Aya, T. & Hamilton, A. D. Tetrabiphenylporphyrin-based receptors for protein surfaces show sub-nanomolar affinity and enhance unfolding. *Bioorg. Med. Chem. Lett.* **13**, 2651-2654 (2003).
180. Casnati, A., Sansone, F. & Ungaro, R. Peptido- and glycolixarenes: playing with hydrogen bonds around hydrophobic cavities. *Acc. Chem. Res.* **36**, 246-254 (2003).
181. Wei, A. Calixarene-encapsulated nanoparticles: self-assembly into functional nanomaterials. *Chemical Communications* 1581-1591.
182. Gutsche, D. *Calixarenes Revisited*. The Royal Society of Chemistry. Blackhorse Road, Letchworth (1998).
183. Iwamoto, K., Araki, K. & Shinkai, S. Conformations and structures of tetra-O-alkyl-p-tert-butylcalix[4]arenes. How is the conformation of calix[4]arenes immobilized? *J. Org. Chem.* **56**, 4955-4962 (1991).
184. Arduini, A. *et al.* Calix[4]arenes Blocked in a Rigid Cone Conformation by Selective Functionalization at the Lower Rim. *The Journal of Organic Chemistry* **60**, 1454-1457 (1995).
185. Galán, H., de Mendoza, J. & Prados, P. Conformational Control of Calix[6]arenes Through Multiple Bridges. *European Journal of Organic Chemistry* **19**, 4093-4097 (2005).
186. Böhmer, V. Calixarenes, Macrocycles with (Almost) Unlimited Possibilities. *Angew. Chem. Int. Ed Engl.* **34**, 713-745 (1995).
187. Jose, P. & Menon, S. Lower-Rim Substituted Calixarenes and Their Applications. *Bioinorganic Chemistry and Applications* **2007**, 16 pages (2007).
188. Perret, F., Lazar, A. N. & Coleman, A. W. Biochemistry of the para-sulfonato-calix[n]arenes. *Chem. Commun. (Camb.)* 2425-2438 (2006).
189. Agrawal, Y. K. & Bhatt, H. Calixarenes and their biomimetic applications. *Bioinorg. Chem. Appl.* 237-274 (2004).
190. Arosio, D. *et al.* A synthetic divalent cholera toxin glycolix[4]arene ligand having higher affinity than natural GM1 oligosaccharide. *J. Am. Chem. Soc.* **127**, 3660-3661 (2005).
191. Droogmans, G., Prenen, J., Eggermont, J., Voets, T. & Nilius, B. Voltage-dependent block of endothelial volume-regulated anion channels by calix[4]arenes. *Am. J. Physiol* **275**, C646-C652 (1998).
192. Consoli, G. M., Cunsolo, F., Geraci, C. & Sgarlata, V. Synthesis and lectin binding ability of glycosamino acid-calixarenes exposing GlcNAc clusters. *Org. Lett.* **6**, 4163-4166 (2004).
193. Sun, J. *et al.* Blocking angiogenesis and tumorigenesis with GFA-116, a synthetic molecule that inhibits binding of vascular endothelial growth factor to its receptor. *Cancer Res.* **64**, 3586-3592 (2004).
194. Cacciapaglia, R. *et al.* Efficient and selective cleavage of RNA oligonucleotides by calix[4]arene-based synthetic metallonucleases. *J. Am. Chem. Soc.* **129**, 12512-12520 (2007).
195. Oh, S. W. *et al.* Calixarene derivative as a tool for highly sensitive detection and oriented immobilization of proteins in a microarray format through noncovalent molecular interaction. *The FASEB Journal* **19**, 1335-1337 (2005).
196. Sansone, F. *et al.* DNA Condensation and Cell Transfection Properties of Guanidinium Calixarenes: Dependence on Macrocycle Lipophilicity, Size, and Conformation. *Journal of the American Chemical Society* **128**, 14528-14536 (2006).



# Biophysical methods

The physical properties of protein-protein and protein-ligand interactions can be thoroughly studied by any of the numerous *biophysical methods* developed to date. *In vitro* biochemical approaches have also been designed for the detection, characterization and understanding of protein-protein interactions and as a result, it starts to be possible to integrate structural information gathered at multiple levels of the biological hierarchy –from atoms to cells– into a common framework. Therefore, we are closer to understand the intriguing principles that underlie the protein network.

It is not the scope of this chapter to review all the biophysical and biochemical techniques available for characterizing proteins interactions. The following pages present a brief introduction to the methodologies applied in the present work, placing emphasis on their application to protein-ligand systems.





## 2 Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) spectroscopy is the best known method to determine three-dimensional structures of biological macromolecules in solution.<sup>1</sup> Even so, the study of protein-protein and protein-ligand interactions is its widest applications nowadays.<sup>2,3</sup> By NMR, the complex is not only structurally characterized; valuable information about the thermodynamics, the kinetics and the mechanisms of the binding event can be also attained.

NMR is based on the quantum mechanical magnetic properties of the nuclei of atoms. Regardless of the experiment design, NMR spectroscopy ultimately studies the magnetic nuclei by aligning them with an applied magnetic field and then perturbing this alignment using an alternating orthogonal magnetic field.

The magnetic properties of nuclei are extremely sensitive to the environment and therefore, to the interaction between molecules. There are many NMR parameters whose change can be used to detect and study the binding process, such as for instance chemical shifts, coupling constants, relaxation or translational diffusion properties. NMR experiments to characterize protein-ligand complexes can be divided according to which part of the complex is observed: on the one hand, those that observe the macromolecule (*i.e.* the receptor), and on the other hand, those that measure NMR parameters for the small binding partner (*i.e.* the ligand).

### 2 Receptor based experiments

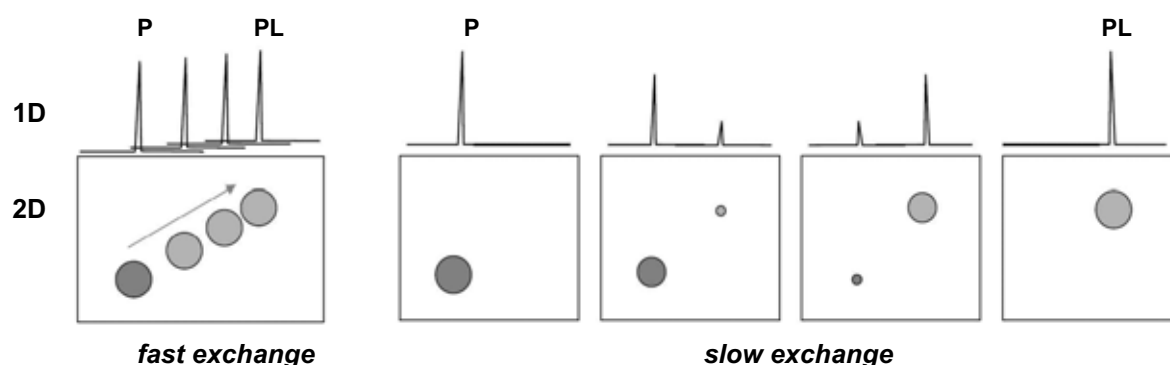
The most popular and probably the simplest experiment among those focused on protein observation to characterize the binding event, is the so-called *Chemical Shift Perturbation (CSP)*. This involves monitoring the changes in the protein resonances upon ligand addition; it is commonly performed in 2D NMR experiments (*e.g.*  $^1\text{H}$ - $^{13}\text{C}$  or  $^1\text{H}$ - $^{15}\text{N}$ ) on an appropriately labeled protein. CSP takes advantage of the high sensitivity of the nuclei chemical shift to their environment; hence, perturbations on the chemical shifts of the protein are used to assess whether the binding event does take place, its dissociations constant and its kinetics (**Figure 11**). Nevertheless, the real power of the technique is unleashed when the protein structure and the assignment of its resonances are available. CSP will primarily occur on resonances nearby the binding site and consequently, by representing the most affected residues on the protein structure, the protein-ligand interface can be outlined.

In spite of the simplicity of the experiment, CSP is not devoid of difficulty. The first intrinsic limitation falls on the requirement of a stable isotope enriched protein which, besides expensive, may not be always accessible. Overexpression in *E. coli* is the regular procedure to achieve labeled samples; however, these systems are no always successful in producing any protein (*e.g.* post-translational modification, chaperone-aided folding or toxic proteins). But even if large production is possible, at the high concentrations required for NMR experiments, the protein might no be stable long enough. In addition, there is a limit in the molecular weigh of the target protein;

the observation of complexes beyond 40kDa is often hampered by severe relaxation and important signal overlapping. New methodologies are being developed for the study of larger proteins,<sup>4</sup> such as for instance deuterium labeling,<sup>5</sup> residue-selective labeling,<sup>6,7</sup> TROSY, *etc.*

Besides these inherent limitations of the technique, observed CSP can not be always distinctive features of the ligand binding site. Proteins floppiness can lead to conformational rearrangements in response to the recognition event; that makes difficult to differentiate between CSP directly caused by the ligand and those induced by distal effects. Fortunately, the interest in NMR for the study of complexes in solution has raised new strategies and experiments to unequivocally determine binding interfaces.<sup>8</sup>

Certain regions of the one-dimensional <sup>1</sup>H spectrum of a protein have also proven informative in molecular recognition processes. This is the case of the <sup>1</sup>H *up-field region*, comprised between 0.4ppm and -1.0ppm. Resonances at these high fields correspond to protons nuclei close to the center of an aromatic ring (*i.e.* a ring current) within the protein. Even if they are not directly perturbed by the ligand, they can be sensitive conformational probes of the structural rearrangement induced on the protein upon ligand interaction.<sup>9-11</sup>



**Figure 11.** How the chemical shift of the protein changes during a titration with ligand is determined by the kinetics of the interaction in the NMR time scale. If the complex dissociation is very fast:  $k_{off}/\Delta\delta > \pi/\sqrt{2}$  (namely,  $k_{off} > 10^4 \text{ s}^{-1}$ ), only a single set of resonances is detected, whose chemical shifts are the fractional weighted average of those of the free and the bound forms. Contrariwise, if the dissociation is very slow:  $k_{off}/\Delta\delta < \pi/\sqrt{2}$ , during the titration with ligand the “free set” disappears and is replaced by the “bound set” of resonances. For intermediate chemical exchange, an intermediate behavior is observed and resonances become poorly defined along the titration. As a rule of thumb, interactions with  $K_D < 10\mu\text{M}$  are slow exchange, and intermediate or fast otherwise.<sup>12</sup> However, it is important to understand that the NMR time scale depends on both the kinetic dissociation constant,  $k_{off}$ , and the distance between the free and bound resonances,  $\Delta\delta$  (in Hz). Large  $\Delta\delta$  can also result in a slow-intermediate exchange behavior and hence, can be misinterpreted as a high affinity association. Likewise,  $\Delta\delta$ , in Hz, also depends on the magnetic field; higher magnetic fields make the exchange faster.

## 2.2.1 Ligand-based experiments

NMR experiments based on the ligand are more numerous and diverse than those for the protein, mainly due to the dramatic changes that small molecules can undergo upon binding to a large receptor. Most of these properties are related to the apparent molecular weight of the ligand, which in the presence of the target increases up to the complex one. Consequently, size-dependent properties such as longitudinal (R1) and transverse (R2) relaxation, diffusion coefficients and inter- or intramolecular magnetization transfer will report of such molecular weight change and hence, prove and characterize the interaction.

Detecting the effects of the binding on the ligand overcomes all the limitations previously described for the receptor-based experiments. First, no labeled protein is required. Second, experiments are less protein demanding, since large ligand-to-protein ratios are used. And third, there is no limit on the molecular weight of the macromolecule; indeed, the larger the receptor the easier to detect the changes on the ligand.

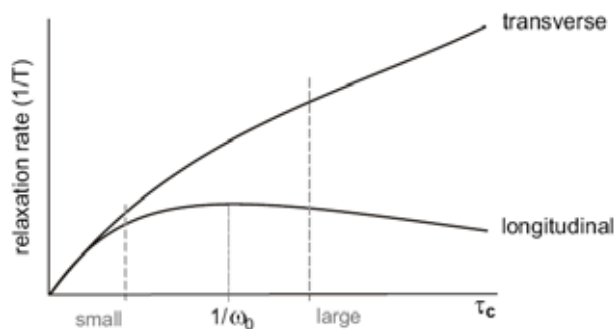
Nevertheless, observing the ligand has its own limitations, too. The major stumbling block which plays down importance to ligand based experiments is the little –or none– information regarding the binding on the macromolecule. Additionally, those experiments are only suitable when the binding proceeds under fast exchange ( $k_{off} > 10^4 \text{s}^{-1}$ ). Tighter binders with slow or intermediate dissociation rates may appear as non-interacting ligands. Even so, it has been proven that this limitation can be overcome by competition experiments, using a weaker binder as a reporter of the binding effects of the tighter one.<sup>2,13</sup>

From the NMR point of view, *relaxation* is likely the most different property between small and large molecules (**Figure 12**). Upon complexation with the receptor, the ligand happens to behave as a large entity and so, the rate at which it relaxes changes dramatically. Given the fast exchange rate between both states, measured magnitudes are the weighted average accounting for the free ( $f_F$ ) and bound ( $f_B$ ) relative populations:

$$A = f_F A_F + f_B A_B.$$

Although the ligand is in large excess over the receptor ( $f_F \gg f_B$ ), relaxation-depending parameters can change several orders of magnitude when the interaction occurs ( $A_F \ll A_B$ ), and consequently, the bound term can dominate the value detected experimentally.

A distinction is made between longitudinal, R1, and transverse, R2, relaxations. The use of R1 relaxation rates to detect binding can be only performed on individual resonances; this is because of the different relaxation mechanisms that resonances can undergo. R2 relaxation is more convenient, informative and successful to use. Spin-spin relaxation times for small molecules rapidly tumbling are usually long, while for resonances of large molecules or protein-bound ligands, T2 is usually much shorter (**Figure 11**). The resonance *linewidth* directly depends on the transversal relaxation rate; hence, upon interaction with the receptor protein, the ligand resonances immediately broaden, and such a direct effect can be used to qualitatively and quantitatively assess binding.

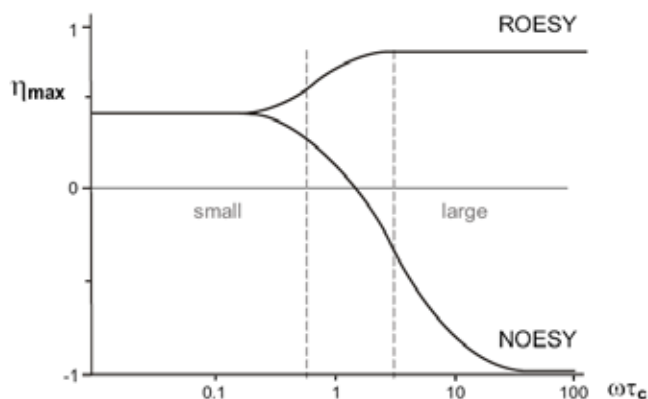


**Figure 12.** Relaxation rates  $R_1$  and  $R_2$  in NMR as a function of the correlation time,  $\tau_c$  (*i.e.* time that the molecule takes to rotate one radian over itself); at a given temperature,  $\tau_c$  mainly depends on the size of the molecule.

In addition to the relaxation rate, another parameter widely employed in ligand-based experiments –due to its high sensitivity to the tumbling of the molecule– is the *nuclear Overhauser effect* (*nOe*).

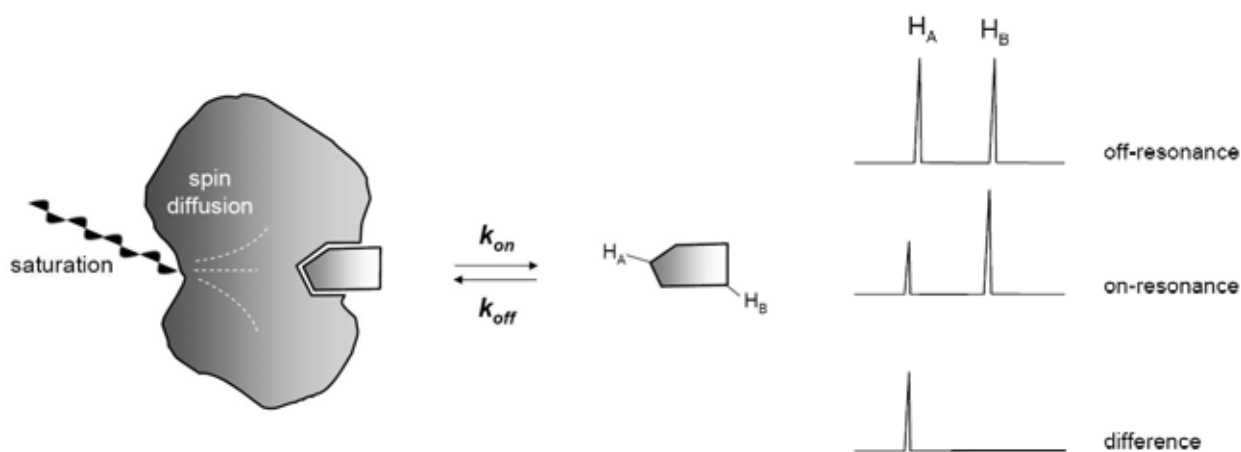
The *nOe* arises from the spatial transfer of polarization from one spin to another.<sup>2,14</sup> Large molecules exhibit strongly negative *nOes*, whereas for small or medium size molecules *nOes* can be positive, null or slightly negative, depending on the molecular weight, shape and magnetic field (**Figure 13**). When the ligand binds to the receptor, it adopts the *nOe* behavior corresponding to the complex and therefore, it shows strong negative *nOes*. This phenomenon is known as *transferred nOe* (*tr-nOe*). Hence, binding can be easily proven by simple observation of the sign and size of the recorded *nOes*.

In addition to the change of the pre-existing *nOes*, new ones can appear if the conformation of the ligand changes upon binding. Likewise, *nOes* between the ligand and the receptor (*i.e.* intermolecular *nOes*) can also be detected under appropriate experimental conditions. Nonetheless, the real power of transferred *nOe* experiments lies in the possibility to determine the actual conformation of the bound ligand.<sup>15</sup>



**Figure 13.** Maximum *nOes* ( $\eta_{\max}$ ) detectable by NOESY or ROESY, plotted as a function of the correlation time ( $\tau_c$ ) and the frequency ( $\omega$ ), since *nOes* not only depend on the molecule tumbling but also on the magnetic field were they are recorded.

Taking advantage of the large negative  $nOes$  in the ligand-receptor complex, several experiments have been designed based on the transfer of magnetization between ligand and protein.<sup>2</sup> This is the case of the *Saturation Transfer Difference (STD)* experiment.<sup>16</sup> In STD, the receptor is magnetically saturated by irradiation at a selective frequency, known as the *on-resonance*. Part of this magnetization can be eventually transferred to the bound ligand by spin diffusion mechanisms. When the ligand dissociates from the target, the magnetization transferred in the bound state still remains in the free ligand; this transfer can be easily determined by subtracting a reference spectrum –acquired with *off-resonance* irradiation– to the spectrum recorded in the on-resonance experiment. The resulting spectrum only contains those ligand signals that have been magnetically perturbed by binding to the target (**Figure 14**). Moreover, the magnitude of the transferred magnetization is proportional to the actual distance of the ligand nuclei to the protein in the bound state; hence, the ligand binding mode can also be determined.



**Figure 14.** STD experiment. The receptor is selectively irradiated at the selective on-resonance frequency. Upon binding, part of the magnetization is transferred to the ligand, the closest regions being the most affected. Subtraction of the on-resonance experiment to an off-resonance reference one, directly leads to the mapping of the ligand binding mode.

Due to the dependence of the STD signal intensity on both ligand excess and ligand concentration, it is usually employed the so-called *STD amplification factor* to better assess the absolute magnitude of the STD effect. The STD amplification factor, in effect, is the relative intensity of the STD signal compared to that of the signal of the protein. It is calculated multiplying the *fractional saturation* of a given proton by the excess of the ligand over the protein (equation [1]). The *fractional saturation* expresses the signal intensity in the STD spectrum,  $I_{sat}$ , as a function of the intensity of an unsaturated reference spectrum,  $I_0$ . The fraction of saturated ligand is continuously reduced when the ligand excess is increased. On the contrary, the STD signal increases as a function of ligand excess.<sup>16</sup> Consequently, the STD amplification factor is used to compare the STD effect for samples of different composition.

$$\text{STD amplification factor} = \frac{I_o - I_{\text{sat}}}{I_o} \times \text{ligand excess} \quad [1]$$

Although the saturation transferred to the ligand depends on the proximity to the receptor, the size of the observed STD signal not only depends on that. Saturation of ligand protons in the bound state is counteracted by their longitudinal relaxation times,  $T_1$ , in the free state;<sup>17</sup> hence, a large STD effect can be misinterpreted for a strong contact if the protein is irradiated at the on-resonance frequency during long saturation times. This effect is especially troublesome for molecules with protons differing markedly in their  $T_1$  values.<sup>17</sup>

Short saturation times are more accurate for epitope mapping experiments, since the  $T_1$  effect has smaller impact.<sup>17</sup> However, under those conditions protein saturation is not efficient and STD signals intensity is not enough for an accurate quantitative analysis. STD misinterpretation can be bypassed tracing the *STD build-up curves* (equation [2]).<sup>16,18</sup> To eliminate the  $T_1$  bias, the probe of the proximity to the receptor is the slope of the STD build-up curve at zero saturation time (equation [3]).

$$\text{STD} = \text{STD}_{\text{max}} \times [1 - \exp(-k_{\text{sat}} \times t_{\text{sat}})] \quad [2]$$

$$v_o = \text{STD}_{\text{max}} \times k_{\text{sat}} \quad [3]$$

## 2.2. Crystallization

X-ray crystallography is the technique *par excellence* for the determination of the three-dimensional structure of proteins and protein complexes at atomic resolution in solid state.<sup>19</sup> In fact, it is by far much more prolific than NMR, and at present more of the 80% of biomolecules structures have been solved by X-ray.<sup>20</sup> NMR cannot compete with X-ray crystallographic methods in accomplishments such as the structural determination of large proteins or supramolecular assemblies. However, not all proteins and protein-complexes may crystallize or, at least, not homogeneously and in a biological relevant conformation.

Crystallization drawbacks are magnified in non-covalent protein-ligand complexes. Even if the crystallization conditions are well established for the target protein, obtaining co-structures with the ligand is too often not successful.<sup>21,22</sup> Two approaches are used in seeking the complex structure: *co-crystallization* and *soaking*. In the soaking experiments, the ligand is incubated with a preformed crystal of the target protein, in the hopes that it will diffuse through the lattice and interact with the protein binding site. In the co-crystallization approach, the ligand is combined with the protein in solution and then the complex is crystallized.

Protein crystals are loosely packed, with 30-80% being solvent molecules.<sup>23</sup> Channels traversing the lattice contain the bulk solvent bathing the crystal and are an access door for small ligands to the crystallized protein.<sup>24</sup> The soaking approach makes faster the crystallization process of the protein-ligand complex;<sup>25</sup> however, the protein crystal used must be compatible with the ligand binding mode. Therefore, channels within the crystal must be large enough as to allow the ligand to diffuse through them, and the binding site of the protein must be present (*i.e.* the protein must be in the right conformation) and accessible. Although both requirements are satisfied, the ligand interaction mode in the crystal structure under the crystallographic conditions may not accurately represent the solution binding mode.<sup>26</sup>

Direct co-crystallization of the complex can circumvent these issues, although that means to deal with the crystallization problem from the beginning. In addition to the difficulties already existent for the crystallization of a protein, introduction of a ligand only carries inconveniences; this is because the sample to be crystallized is not homogeneous anymore. Homogeneity is essential in the preparation of useful crystals; the presence of impurities or any other perturbation factor can make impossible the crystallization, or may incorporate defects into the lattice and lead to a poor quality crystal. For the crystallization of a protein-ligand complex, heterogeneity is unavoidably introduced, since the free and the bound species are present. Whether this heterogeneity negatively affects crystal growth will depend on the ability of each form of the protein to incorporate into the lattice. In this respect, ligand affinity is crucial in order to minimize heterogeneities. Nevertheless, the affinity of the ligand for the protein can be seriously perturbed under the stringent buffer conditions employed in crystallography, and –most of the times– the ligand solubility is reduced so much that it simply precipitates.

But the crystallographic problem does not end when crystals are finally formed.

The loose packing of the protein molecules within the lattice makes the crystal to be extremely fragile, not only to physical manipulation but also to the X-ray radiation during the diffraction. The damage can be delayed by working at very low temperatures (*i.e.* 100K), as *cryocrystallography* does. In order to keep unaltered the crystal when it is frozen –because of the transition of liquid water into the solid state–, it is first required to find an appropriate *cryoprotectant* buffer.

When finally the crystal is diffracted, it may occur that its quality is not enough as to collect high-resolution data ( $<3\text{\AA}$ ; a C-C bond is approximately  $1.5\text{\AA}$ ). *Crystal size* and *mosaicity* are the two key characteristics which determine the quality of the diffraction. Large crystals present a larger number of dispersing elements and hence, its diffraction is more intense. Mosaicity refers to the degree of order –or disorder– within the crystal. Large mosaicities correspond to imperfect and disordered packages, thereby resulting in poor quality diffractions.

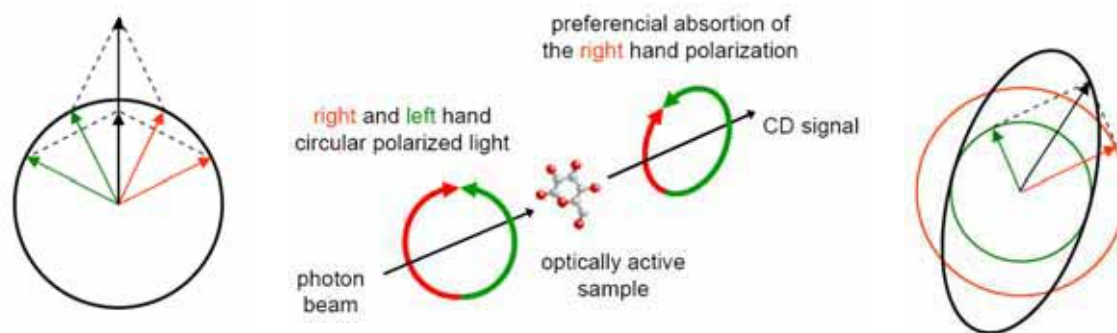
The processing of the collected data starts by the determination of the *unit cell* (*i.e.* the smallest repeating unit that makes up the crystal), its *dimensions* (given as three lengths: a-b-c, and three angles:  $\alpha$ - $\beta$ - $\gamma$ , which determine the spot spacing of the diffraction image) and the *space group* (which determines how the molecules are packed into the crystal lattice and is provided by the symmetry of the diffraction patterns). The diffraction pattern is then indexed and the intensities measured. The *electron density map* is ultimately calculated and the structural model built (*i.e.* each amino acid of the protein sequence is inserted into the electron density map) and refined.

In the case of crystal soaking, data processing is shortened, because the structure of the protein is already known and it is only necessary to determine at which positions the changes in the electron density appear to locate the bound ligand. In order to obtain reliable information, enough ligand has to be incorporated into the crystal (*i.e.*  $>20\%$  occupation). Soaking experiments can even provide binding constants;<sup>27</sup> this is achieved by soaking protein crystals into solutions containing increasing amounts of ligand (like a “titration”), since the degree of occupation depends on the ligand concentration. However, the “thermodynamics” and the structure of the complex obtained in the solid state by X-ray may differ substantially from those of a fluid biological system.



## 2 Circular Dichroism

A beam of light polarized in a plane is composed of two components circularly polarized: the right-handed one and the left-handed one. Both components are in phase and have the same amplitude. When the beam passes through an optically active sample, the right-handed and the left-handed components interact in a different manner with the chiral centers of the molecules and thus, they are differently absorbed. After passing through the sample, each component is still circularly polarized, but their phase and amplitude are not the same anymore. When they are combined, the result is an elliptically distorted polarized light beam. This is the phenomenon known as *circular dichroism* (**Figure 15**).



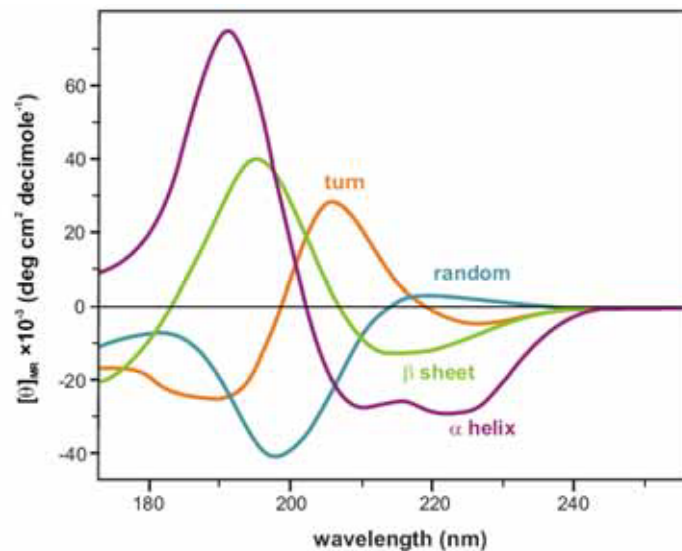
**Figure 15.** Circular dichroism phenomenon. At the left, the former beam of polarized light with the right-handed (in red) and the left-handed (in green) components phased and of equal amplitude. At the right, the resulting beam of light elliptically polarized.

Although all amino acids –except glycine– contain at least one asymmetric carbon atom, most of them only display small CD effects by themselves. It is the conformation of the protein, namely, the spatial asymmetric and periodical arrangements of the residues, that arises the characteristic CD spectra of proteins. The major optically active groups in proteins are the amide bonds of the backbone and the aromatic side chains.

In the *far ultraviolet* region (178-260nm) CD bands mainly arise from the amides of the protein backbone (transitions  $n-\pi^*$  at 220 nm and  $\pi-\pi^*$  at 190 nm). Depending on the orientation of the peptide bonds within the protein structure, the optical transitions of the amide bond can be split into multiple transitions, the wavelengths of the transitions can be increased or decreased, and the intensity of the transitions can be enhanced or decreased. As a consequence, each secondary structure motif has a characteristic CD bands pattern (**Figure 16**). Hence,  $\alpha$ -*helices* display two negative bands at 222nm and 208nm and a positive one at 190nm; for  $\beta$ -*sheets*, the spectrum is less intense, with two negatives bands around 217nm and 180nm and a positive at 195nm, although depending on the relative orientation of the  $\beta$ -strands on the  $\beta$ -sheet the pattern can change substantially;<sup>28</sup> *turns* can adopt several conformations ( $\beta$ -turns type I, II and III) and each

has its own CD profile; and for disordered structures, namely *random coils*, a negative band about 200nm is detected.

Moreover, some proteins also display bands in the *near ultraviolet* (350-260nm) and visible region, which arise from aromatic amino acids (transitions  $\pi$ - $\pi^*$ ). The CD in these regions depends on the environment of the chromophores, hence, on the tertiary structure of the protein.



**Figure 16.** Circular dichroism spectra in the far UV for polypeptides in pure  $\alpha$ -helix,  $\beta$ -strand,  $\beta$ -turn or random coil conformation

As a result, valuable information about the secondary and tertiary structure of the protein can be attained by analysis of its CD spectrum.<sup>29-33</sup> Although CD does not provide information at the atomic detail, it is an excellent tool for *rapid* structural determinations and does not require large protein amounts. However, for an accurate estimation of the secondary structure, data should be collected up to 178nm or lower; unfortunately, at these wavelengths experimental data are fairly uncertain and display poor signal-to-noise ratios.

In addition to structural information, CD changes as a function of temperature, time or presence of ligands can also provide other valuable thermodynamic and kinetic information.<sup>34-37</sup> Regarding the interaction with ligands, since CD is a quantitative technique, the changes in the spectra are directly proportional to the amount of complex formed; hence, they can be used to estimate binding affinities. The changes can arise from structural rearrangements induced upon binding, but also from changes in the environment of the aromatic moieties –from both the protein and the ligand. Furthermore, the thermal stabilization induced on the protein by the interaction with a ligand can also be determined by monitoring the change in the ellipticity when scanning up the temperature; in addition, from the thermal shift, thermodynamic parameters can be calculated.

## 2 i erentia cainn Ca ori etr

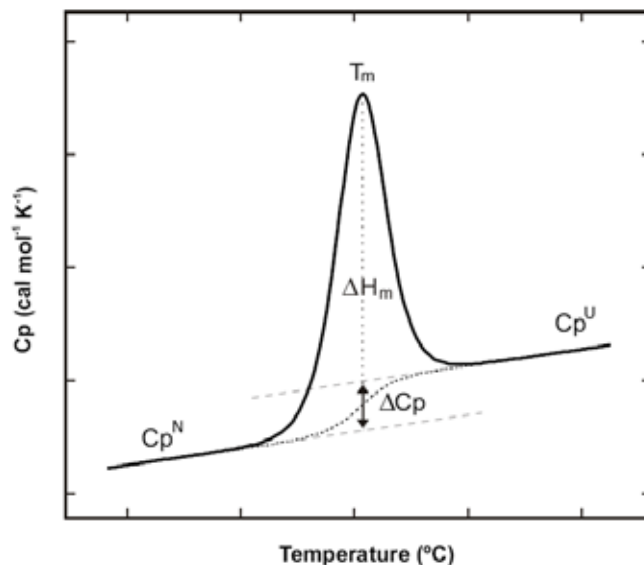
Differential Scanning Calorimetry (DSC) is the most powerful technique for the direct and utter energetic characterization of the thermal stability of a protein. Consequently, DSC provides priceless quantitative information about the thermodynamic effects that result from the interaction with a ligand molecule. Moreover, energetic characterization can further provide qualitative structural description of the protein-ligand system (e.g. hydrogen bonds, electrostatic or hydrophobic dominance, structural rearrangements...). The main pitfall in DSC is the requirement of a large concentration of unrecoverable protein, which might lead to a non-ideal behavior.

Experimentally, DSC records the changes in the apparent heat capacity,  $C_p$ , when scanning up the temperature. During the thermal unfolding process, the hydrophobic core of the folded protein becomes solvent exposed and consequently, the sample experiences an increase in the heat capacity. This means that the heat required to rise the temperature of a solution of unfolded protein is larger than that required for a solution of folded protein, due to the reorganization of the solvent molecules around the non-polar side chains exposed to the solvent during the unfolding.<sup>38</sup>

It is called “differential” because microcalorimeters measure the difference of the heat capacity between two identical cells, one containing the protein sample (sample cell) and the other containing the plain buffer (reference cell). Given that the heat capacity for both cells is different, different thermal power have to be applied to each in order to be scanned at the same rate.<sup>39</sup>

From the raw heat capacity data, it is possible to thermodynamically characterize the unfolding transition of the system. Some parameters can be directly measured from the experimental thermogram (**Figure 17**), such are the *melting temperature*,  $T_m$  (the temperature at which  $C_p$  is at a maximum); the *heat capacity change*,  $\Delta C_p$  (the difference between the initial and final baselines at  $T_m$ ) and the *calorimetric unfolding enthalpy*,  $\Delta H_m$  (the area under the  $C_p$  curve, i.e. the integral of  $C_p$  over the temperature). The *entropy change*,  $\Delta S_m$ , can be determined as the integral of the  $C_p/T$  curve over the temperature. The *free energy change* at the melting temperature,  $\Delta G_m$ , is calculated through the well-known equation  $\Delta G_m = \Delta H_m - T_m \Delta S_m$ . In addition, the *reversibility* of the process can be proven by re-scanning or scanning-back an unfolded sample.

The melting temperature is an excellent probe of the protein *stability*; a larger  $T_m$  means a more stable protein. The melting enthalpy correlates with the content of structure and is the net value from the combination of endothermic contributions –such as the disruption of hydrogen bonds– and exothermic ones –such as the break-up of hydrophobic interactions.<sup>40</sup> In general, the more stable the protein, the larger the energy required to unfold it; that is,  $\Delta H_m$  becomes more positive as  $T_m$  increases.<sup>41</sup> The *sharpness* of the transition peak is an indicator of the *cooperative* nature of the transition from native to unfolded (e.g. two-state or multi-state); if the unfolding occurs within a narrow temperature range, the transition is considered highly cooperative.<sup>40</sup> The unfolding mechanism can be asserted by comparison of the experimental parameters with those calculated by an indirect non-calorimetric approach, the so-called *van't Hoff analysis*.<sup>42</sup>



**Figure 17.** DSC thermogram for a two-state unfolding ( $N \rightleftharpoons U$ ).  $C_p^N$  and  $C_p^U$  correspond to the heat capacity in the native and the unfolded state respectively. The area comprised between the experimental  $C_p$  curve and a hypothetical baseline linking the native and the unfolded state (short-dashed line) corresponds to  $\Delta H_m$ .

The interaction of a protein with a ligand involves changes in the intra- and intermolecular interactions as well as in the dynamics of all the components. These changes are reflected in the energetics of the binding (namely,  $\Delta G_B$ ,  $\Delta H_B$  and  $\Delta S_B$ ) and have immediate consequences on the protein stability.<sup>43</sup>

When a ligand binds to the native state of the protein, the biomolecule becomes thermodynamically stabilized. Therefore, the protein experiences an increase in the melting temperature that is usually accompanied by an increase in the unfolding enthalpy.<sup>44</sup> These changes are proportional to the affinity of the ligand; hence, binding constants –at the melting temperature– can be calculated. However, since the ligand binding is indirectly estimated from the changes in the unfolding equilibrium of the protein, precaution must be taken in the interpretation of the resulting affinities.<sup>45</sup> In addition, the shift in  $T_m$  is not always directly correlated with the affinity. The ligand can also interact with the unfolded protein and shift backwards the transition peak. Alternatively, the ligand might dissociate before getting to  $T_m$ , thus not modifying the protein unfolding profile.

Regardless of these limitations, DSC is the only technique which enables the study of ultra-tight bindings which cannot be managed by other biophysical methods (*i.e.*  $K_D \geq 10^{-20}\text{M}$ ).<sup>42</sup>

## 2 Isothermal Titration Calorimetry

Isothermal Titration Calorimetry (ITC) directly measures the enthalpy change for a bimolecular binding reaction at a constant temperature.<sup>46</sup> In principle, the technique is *universal* and can be applied to any kind of bimolecular process since *heat evolution* is an *inherent* property of all chemical reactions. No reporters are required, no approximations or indirect assumptions have to be taken, and there are no limitations in the size or the nature of the components.

Experimentally, an ITC calorimeter is composed by two identical cells: the sample cell and the reference cell, placed in an adiabatic jacket at constant temperature. The reference cell is filled with the buffer (or water) and the sample cell is filled with a solution of one of the components of the reaction –usually the protein. The second component of the reaction –usually the ligand– is loaded into a syringe, which is then physically introduced into the sample cell, where will act also as a stirring device for sample homogenization. The experiment consists in injecting little aliquots of the titrant into the sample cell, and measuring the heat that is absorbed (*i.e.* endothermic) or released (*i.e.* exothermic) after every injection. In fact, the direct observable measured is the time dependent input of power ( $\text{J}\cdot\text{s}^{-1}$ ) required to maintain equal temperatures in both cells, which appears as a “peak” of power after every injection (**Figure 18**).<sup>40</sup> As the titration progresses, the heats resulting after every injection progressively decrease, due to the saturation of binding sites, up to finally only detected the heat from the dilution of the titrant into the sample solution.

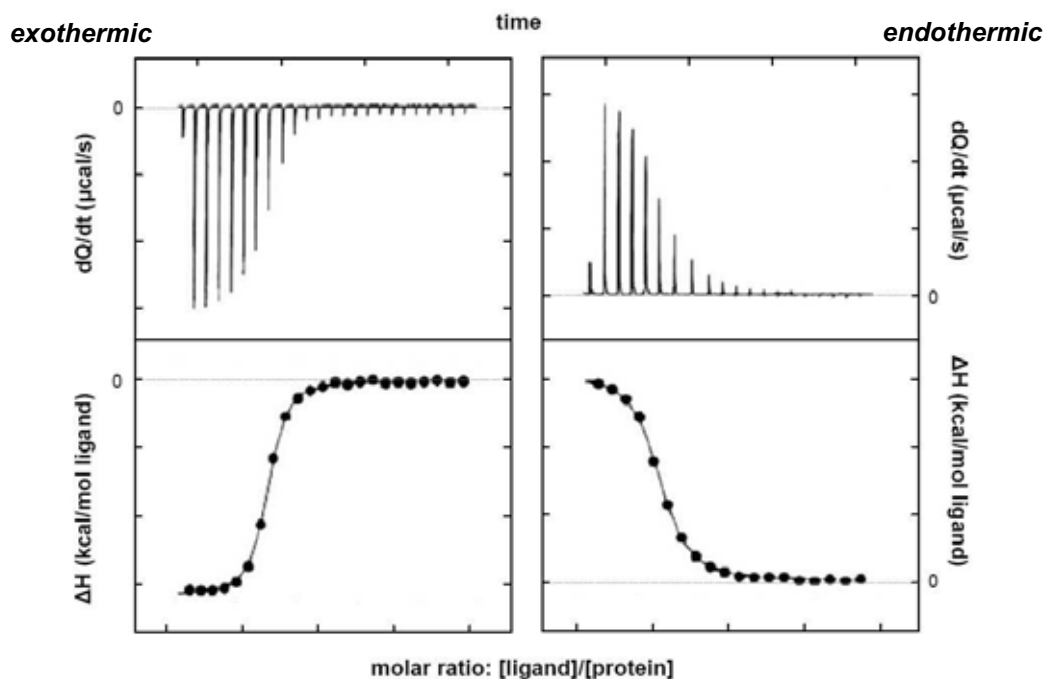
The raw data from the ITC experiment contain the energy of all the processes taking place during the titration; this not only includes the heat of the interaction, but also the dilution of the ligand into the protein solution, the dilution of the protein when the ligand is added, and the heat effects of stirring. These additional factors can be corrected by the corresponding blank controls. The resulting net heat from the interaction protein-ligand is the “probe” of the extent of the binding reaction.

Analysis of the ITC data provides the so-called *binding isotherm* (**Figure 18**). The isotherm gathers all the thermodynamic information regarding the binding reaction: the *stoichiometry*,  $n$ ; the *equilibrium binding constant*,  $K_B$ , and the *binding enthalpy*,  $\Delta H_B$ , at the working temperature. From these parameters, the *free energy change*,  $\Delta G_B$  ( $\Delta G_B = -RT \ln K_B$ ) and the *binding entropy*,  $\Delta S_B$  ( $\Delta G_B = \Delta H_B - T\Delta S_B$ ) are directly calculated.<sup>47</sup>

Further thermodynamic characterization of the system can be achieved by recording several ITC experiments at different temperatures (*e.g.* determination of the binding  $\Delta C_P$ ), at different pH (*e.g.* number of protonation events)<sup>48,49</sup> or in different buffers (*e.g.* ionic strength effects).<sup>50,51</sup>

From a molecular point of view, there are many factors that contribute to the binding energy. The experimental value for  $\Delta H_B$  comprises all the single enthalpic processes taking place in the whole system; this includes intermolecular contacts (*e.g.* van der Waals interactions, hydrogen bonds), conformational changes, solvent rearrangements or ionizations events. They all may not directly result from the pure protein-ligand interaction and thus,  $\Delta H_B$  can be difficult to interpret.  $\Delta S_B$  is even more ambiguous; upon binding, the system loses degrees of freedom, which results in a

negative contribution. However, upon binding, water molecules are realized from the molecule ordered shell into the bulk solvent, which positively contributes to the entropy. As a consequence, interpreting the driving forces of the interaction from the enthalpy and the entropy values is most of the times uncertain.<sup>47</sup>



**Figure 18.** Isothermic calorimetric titration of a protein with a ligand. Upper panels display the recorded experimental data, each peak corresponding to an injection of ligand. In the lower panels, the binding isotherms display the heat per mol of ligand vs. the ligand-to-protein molar ratio. The solid line corresponds to the mathematical adjustment of the data.

The beauty of the ITC is in its simplicity: the entire set of thermodynamic parameters is directly obtained from a single experiment. However, it is not devoid of difficulty, and calorimetric titrations are not always successful. One of the most critical factors in the experimental design is the concentration of the macromolecule and the ligand. For an accurate determination of the binding constant, the concentration of the macromolecules ( $[M]_0$ ) must satisfy the following rule:<sup>46</sup>

$$1 < K_B \times n \times [M]_0 < 1000$$

Ligand must be added until saturation is reached, but depicting a well defined transition, not too smooth neither too sharp. Hence, for low affinity systems, the concentrations required to achieve a well featured binding isotherm are too large and most of the time unattainable. Conversely, for too high affinities, the concentrations required are so low that experimental heats are beyond the sensitivity of the calorimeter. In practice, ITC is limited to  $K_B$  in the range  $10^6$ - $10^{11}$ M.

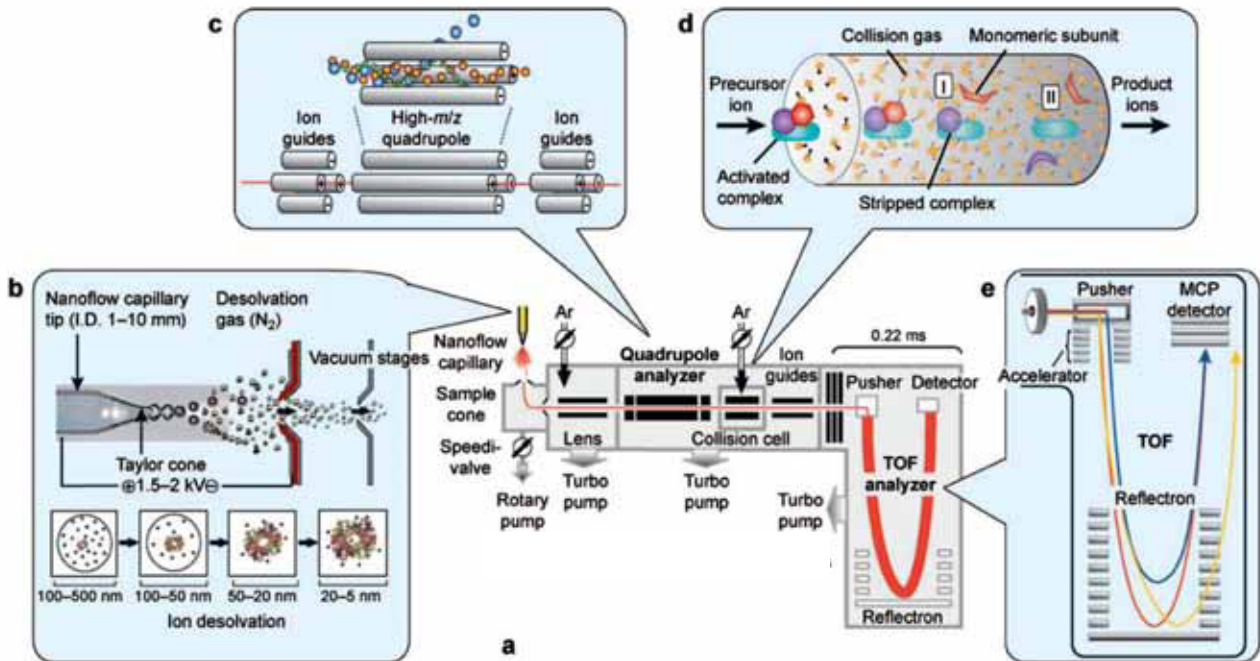
## 2 Mass spectrometry by electrospray ionization

Detection of non-covalent protein-ligand complexes by mass spectrometry (MS) has become very popular in the last decade. It presents many advantages over other classical biophysical techniques, such as the low sample demand and that no label is required –the molecular weight is an inherent property of every molecule. Furthermore, mass spectrometry experiments are *ultra-fast* and do not require long and complicated processing; hence, within seconds, it is possible to know if the ligand binds to the protein and the stoichiometry of their interaction, even if the protein sequence and structure are unknown.

Of course, mass spectrometry also has its limitations, being the worse of the pitfalls the *gas phase* detection of the solution-phase species. For the generation of ions in the gas-phase directly from the liquid solution, it is essential a gentle ionization method that yields no molecular fragmentation. This can be achieved by the well-known *electrospray ionization* (ESI). Nevertheless, the generation of ions is only the first hurdle to overcome. For success, ions must be retained intact in the various differentially pumped regions of the mass spectrometer, focused and guided along their intended flight path, separated according to their mass-to-charge ( $m/z$ ) ratio, and subsequently detected (**Figure 19**).<sup>52</sup>

In MS-ESI, it is assumed that many aspects of the solution-phase protein-ligand structure are maintained in the gas-phase.<sup>53,54</sup> However, it is well known that MS has a bias toward detecting electrostatic interactions, and hydrophobic interfaces are weakened in the gas phase.<sup>55</sup> Hence, the intensity of the detected mass peaks for a protein-ligand complex cannot be directly correlated with the affinity. Moreover, in cases in which the association is mainly hydrophobic, it may be impossible to maintain and detect an intact complex.

Fortunately, the rising interest in the field has prompted the development of new approaches to bypass those limitations and further attain quantitative data about binding from MS experiments. The strategy in question consists in compare the hydrogen-deuterium exchange properties of the ligand-free and the ligand-bound proteins. The existence of the complex in the gas phase is not required anymore since the level of H/D exchange in the protein –that can be determined by MS with great accuracy– reports directly the binding of the ligand and can even provide information regarding the dynamics of the interaction.<sup>55</sup>



**Figure 19.** Schematic representation of a high-mass quadrupole time-of-flight (QTOF) spectrometer.<sup>55</sup> **(a)** Ions are formed by nanoflow electrospray (ES) of the complex-containing solution and are drawn into vacuum through the ion source. The ions are driven through a series of pressure and potential gradients until their ultimate detection (red line). **(b)** Representation of the *nanoflow ES process*. **(c)** The *quadrupole mass filter*. A quadrupole consists of four parallel rods, in which the ions describe a complicated trajectory dependent on their mass-to-charge ( $m/z$ ) value. One of the four rods is removed for clarity. **(d)** Schematic representation of the *collision-induced dissociation process* within the collision cell. The precursor ion selected via the quadrupole undergoes repeated collisions with the collision gas (orange), which increase the internal energy of the complex until it dissociates. **(e)** Different ion trajectories in the *TOF detector*. Ions with different  $m/z$  values have different velocities and will therefore reach the detector at different times.



## i io raph

1. Ferentz, A. E. & Wagner, G. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* **33**, 29-65 (2000).
2. Meyer, B. & Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew. Chem. Int. Ed Engl.* **42**, 864-890 (2003).
3. Lepre, C. A., Moore, J. M. & Peng, J. W. Theory and applications of NMR-based screening in pharmaceutical research. *Chem. Rev.* **104**, 3641-3676 (2004).
4. Fiaux, J., Bertelsen, E. B., Horwich, A. L. & Wuthrich, K. NMR analysis of a 900K GroEL GroES complex. *Nature* **418**, 207-211 (2002).
5. Tugarinov, V. & Kay, L. E. A 2H NMR relaxation experiment for the measurement of the time scale of methyl side-chain dynamics in large proteins. *J. Am. Chem. Soc.* **128**, 12484-12489 (2006).
6. Rodriguez-Mias, R. A. & Pellecchia, M. Use of selective Trp side chain labeling to characterize protein-protein and protein-ligand interactions by NMR spectroscopy. *J. Am. Chem. Soc.* **125**, 2892-2893 (2003).
7. Leone, M., Rodriguez-Mias, R. A. & Pellecchia, M. Selective incorporation of 19F-labeled Trp side chains for NMR-spectroscopy-based ligand-protein interaction studies. *Chembiochem.* **4**, 649-650 (2003).
8. Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. & Shimada, I. A novel NMR method for determining the interfaces of large protein-protein complexes. *Nat. Struct. Biol.* **7**, 220-223 (2000).
9. Hoyt, D. W., Harkins, R. N., Debanne, M. T., O'Connor-McCourt, M. & Sykes, B. D. Interaction of transforming growth factor alpha with the epidermal growth factor receptor: binding kinetics and differential mobility within the bound TGF-alpha. *Biochemistry* **33**, 15283-15292 (1994).
10. Hinshelwood, J. & Perkins, S. J. Conformational changes during the assembly of factor B from its domains by (1)H NMR spectroscopy and molecular modelling: their relevance to the regulation of factor B activity. *J. Mol. Biol.* **301**, 1267-1285 (2000).
11. Perkins, S. J. & Wuthrich, K. Conformational transition from trypsinogen to trypsin: 1H nuclear magnetic resonance at 360 MHz and ring current calculations. *Journal of Molecular Biology* **138**, 43-64 (1980).
12. van Nuland, N. A. *et al.* The NMR determination of the IIA(mtl) binding site on HPr of the Escherichia coli phosphoenol pyruvate-dependent phosphotransferase system. *FEBS Lett.* **315**, 11-15 (1993).
13. Carlomagno, T. Ligand-target interactions: what can we learn from NMR? *Annu. Rev. Biophys. Biomol. Struct.* **34**, 245-266 (2005).
14. Ni, F. Recent developments in transferred NOE methods. *Progress in Nuclear Magnetic Resonance Spectroscopy* **26**, 517-606 (1994).
15. Ni, F. & Scheraga, H. A. Use of the Transferred Nuclear Overhauser Effect To Determine the Conformations of Ligands Bound to Proteins. *Accounts of Chemical Research* **27**, 257-264 (1994).
16. Mayer, M. & Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* **123**, 6108-6117 (2001).
17. Yan, J., Kline, A. D., Mo, H., Shapiro, M. J. & Zartler, E. R. The effect of relaxation on the epitope mapping by saturation transfer difference NMR. *J. Magn Reson.* **163**, 270-276 (2003).
18. Mayer, M. & James, T. L. NMR-based characterization of phenothiazines as a RNA binding scaffold. *J. Am. Chem. Soc.* **126**, 4453-4460 (2004).
19. Cantor, C. R. & Schimmel, P. R. *Biophysical Chemistry Part II: Techniques for the study of biological structure and function.* Bartlett, A. C., Vapnek, P. C. & McCombs, L. W. (eds.), pp. 687-791 (WH Freeman & co., San Francisco, 1980).
20. Westbrook, J. *et al.* The Protein Data Bank: unifying the archive. *Nucleic Acids Res.* **30**, 245-248 (2002).
21. Danley, D. E. Crystallization to obtain protein-ligand complexes for structure-aided drug design. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 569-575 (2006).
22. Hassell, A. M. *et al.* Crystallization of protein-ligand complexes. *Acta Crystallogr. D. Biol. Crystallogr.* **63**, 72-79 (2007).
23. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497 (1968).
24. Geremia, S., Campagnolo, M., Demitri, N. & Johnson, L. N. Simulation of diffusion time of small molecules in protein crystals. *Structure.* **14**, 393-400 (2006).
25. McNae, I. W. *et al.* Studying protein-ligand interactions using protein crystallography. *Crystallography Reviews* **11**, 61-71 (2005).
26. Hiller, N. *et al.* Plasmodium falciparum glutathione S-transferase--structural and mechanistic studies on ligand binding and enzyme inhibition. *Protein Sci.* **15**, 281-289 (2006).
27. Wu, S. S., Dornan, J., Kontopidis, G., Taylor, P. & Walkinshaw, M. D. The First Direct Determination of a Ligand Binding Constant in Protein Crystals. *Angew. Chem. Int. Ed Engl.* **40**, 582-586 (2001).
28. Manning, M. C. & Woody, R. W. Theoretical determination of the CD of proteins containing closely packed antiparallel beta-sheets. *Biopolymers* **26**, 1731-1752 (1987).
29. Brahms, S. & Brahms, J. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J. Mol. Biol.* **138**, 149-178 (1980).
30. Greenfield, N. J. Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal. Biochem.* **235**, 1-10 (1996).
31. Greenfield, N. J. Analysis of circular dichroism data. *Methods Enzymol.* **383**, 282-317 (2004).
32. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876-2890 (2006).

33. Rogers, D. M. & Hirst, J. D. First-principles calculations of protein circular dichroism in the near ultraviolet. *Biochemistry* **43**, 11092-11102 (2004).
34. Greenfield, N. J. Applications of circular dichroism in protein and peptide analysis. *TrAC Trends in Analytical Chemistry* **18**, 236-244 (1999).
35. Greenfield, N. J. Analysis of the kinetics of folding of proteins and peptides using circular dichroism. *Nat. Protoc.* **1**, 2891-2899 (2006).
36. Greenfield, N. J. Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. *Nat. Protoc.* **1**, 2733-2741 (2006).
37. Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **1**, 2527-2535 (2006).
38. Gomez, J., Hilser, V. J., Xie, D. & Freire, E. The heat capacity of proteins. *Proteins* **22**, 404-412 (1995).
39. Jelesarov, I. & Bosshard, H. R. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* **12**, 3-18 (1999).
40. Privalov, P. L. & Khechinashvili, N. N. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *J. Mol. Biol.* **86**, 665-684 (1974).
41. Privalov, P. L. & Dragan, A. I. Microcalorimetry of biological macromolecules. *Biophys. Chem.* **126**, 16-24 (2007).
42. Bruylants, G., Wouters, J. & Michaux, C. Differential scanning calorimetry in life science: thermodynamics, stability, molecular recognition and application in drug design. *Curr. Med. Chem.* **12**, 2011-2020 (2005).
43. Celej, M. S., Dassie, S. A., Gonzalez, M., Bianconi, M. L. & Fidelio, G. D. Differential scanning calorimetry as a tool to estimate binding parameters in multiligand binding proteins. *Anal. Biochem.* **350**, 277-284 (2006).
44. Fukada, H., Sturtevant, J. M. & Quijcho, F. A. Thermodynamics of the binding of L-arabinose and of D-galactose to the L-arabinose-binding protein of *Escherichia coli*. *J. Biol. Chem.* **258**, 13193-13198 (1983).
45. Waldron, T. T. & Murphy, K. P. Stabilization of proteins by ligand binding: application to drug screening and determination of unfolding energetics. *Biochemistry* **42**, 5058-5064 (2003).
46. Leavitt, S. & Freire, E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **11**, 560-566 (2001).
47. Doyle, M. L., Louie, G., Dal Monte, P. R. & Sokoloski, T. D. Tight binding affinities determined from thermodynamic linkage to protons by titration calorimetry. *Methods Enzymol.* **259**, 183-194 (1995).
48. Kresheck, G. C., Vitello, L. B. & Erman, J. E. Calorimetric studies on the interaction of horse ferricytochrome c and yeast cytochrome c peroxidase. *Biochemistry* **34**, 8398-8405 (1995).
49. Baker, B. M. & Murphy, K. P. Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry. *Biophys. J.* **71**, 2049-2055 (1996).
50. Holdgate, G. A. *et al.* The entropic penalty of ordered water accounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: a thermodynamic and crystallographic study. *Biochemistry* **36**, 9663-9673 (1997).
51. Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem. Rev.* **107**, 3544-3567 (2007).
52. Ruotolo, B. T. & Robinson, C. V. Aspects of native proteins are retained in vacuum. *Curr. Opin. Chem. Biol.* **10**, 402-408 (2006).
53. Wyttenbach, T. & Bowers, M. T. Intermolecular interactions in biomolecular systems examined by mass spectrometry. *Annu. Rev. Phys. Chem.* **58**, 511-533 (2007).
54. Robinson, C. V. *et al.* Probing the Nature of Noncovalent Interactions by Mass Spectrometry. A Study of Protein-CoA Ligand Binding and Assembly. *Journal of the American Chemical Society* **118**, 8646-8653 (1996).
55. Sharon, M. & Robinson, C. V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu. Rev. Biochem.* **76**, 167-193 (2007).



# Objectives

Protein-protein interactions are essential in biological processes and thus have become very promising pharmacological targets. Nevertheless, this area of research is still young and modulation of protein complexes remains a challenge. Since most work to date has been focused on the inhibition of protein-protein interactions, there is little precedent on the design of molecules which can induce, stabilize or recover the oligomerization state of proteins. In this context, the system comprised of the tetramerization domain of protein p53 and its oncogenic mutants with defective oligomerization properties is an outstanding case of study for the design and evaluation of molecules which can recover the tetrameric structure.

Previous work in our group showed how linear oligobicycloguanidinium compounds (Salvatella *et al.*, 2004) and linear oligoarginine peptides (Martinell *et al.*, 2006) could recognize an anionic patch of carboxylates on the surface of p53 tetramerization domain. However, these compounds were –presumably– not able to join together the four units of the protein.

Based on the recognition properties of the guanidinium group, and on the structure of the p53 tetramerization domain, Prof. Javier de Mendoza (ICIQ) rationally designed multivalent calix[4]arene ligands that –theoretically– can simultaneously interact with several monomers of the protein and thus stabilize the whole tetrameric assembly.

Hence, the objectives for the present thesis have been the following:

1. **To obtain natural mutants of the tetramerization domain of protein p53** with defective assembly abilities, and to establish the best biophysical methodologies for their **structural characterization**.
2. To test the **molecular recognition abilities of the designed calix[4]arene ligands**, namely, by answering the following questions:
  - Can they interact with the tetramerization domain of p53 as intended?
  - Can they interact with mutated tetramerization domains? If so, what is this interaction like?
  - Can they stabilize the tetrameric assembly?

## Precedents

M. Martinell, X. Salvatella, J. Fernandez-Carneado, S. Gordo, M. Feliz, M. Menendez, and E. Giralt. 2006. Synthetic ligands able to interact with the p53 tetramerization domain. Towards understanding a protein surface recognition event. *Chembiochem*. **7**:1105-1113.

X. Salvatella, M. Martinell, M. Gairi, M. G. Mateu, M. Feliz, A. D. Hamilton, J. de Mendoza, and E. Giralt. 2004. A tetraguanidinium ligand binds to the surface of the tetramerization domain of protein P53. *Angew. Chem. Int. Ed Engl.* **43**:196-198