



UNIVERSITÀ DEGLI STUDI DI SALERNO



UNIVERSITÀ DEGLI STUDI DI SALERNO  
Dipartimento di Farmacia

Dottorato di ricerca  
in Scienze Farmaceutiche  
Ciclo XIII — Anno di discussione 2015

Coordinatore: Chiar.mo Prof. *Gianluca Sbardella*

## ***NMR STUDY OF PROTEIN-LIGAND INTERACTION***

settore scientifico disciplinare di afferenza: CHIM/08

**Dottorando**

**Dott. *Manuela Grimaldi***

**Tutore**

**Chiar.ma Prof. *Anna Maria D'Ursi***



## **Preface**

My PhD in Pharmaceutical Sciences at the Department of Pharmacy of Salerno University was started in January 2012 under the supervision of Prof. Anna Maria D'Ursi.

My research activity was mainly focused onto study of protein-ligand interactions by NMR spectroscopy techniques.

These approaches were successfully applied to two protein-ligand complexes. One to characterize a molecular interaction mechanism at the basis of the antiviral activity of C8 peptide. The other to verify the interaction between N6-isopentenyladenosine and the enzyme Farnesyl Diphosphate Synthase (FPPS). The entire work was carried out under the direct supervision of Prof. Anna Maria D'Ursi.

Furthermore, to improve my knowledge on NMR methodologies, I spent a period of research activity at Structural and Computational Biology Unit, EMBL (European Molecular Biology Laboratory) Heidelberg, under the supervision of the Dr. Teresa Carlomagno. During this period I contributed to a methodological work focused on a development of a new NMR-guided rescoring protocol of molecular docking ligand poses.



**List of publications related to the scientific activity performed during the three years PhD course in Pharmaceutical Sciences:**

1) Sublimi Saponetti M, **Grimaldi M**, Scrima M, Albonetti C, Nori SL, Cucolo A, Bobba F, D'Ursi AM. Aggregation of A $\beta$ (25-35) on DOPC and DOPC/DHA Bilayers: An Atomic Force Microscopy Study. *PLoS One*. **2014** Dec 31;9(12).

2) Testa C, Scrima M, **Grimaldi M**, D'Ursi AM, Dirain NM, Lubin-Germain N, Singh A, Haskell-Luevano C, Chorev M, Rovero P, Papini AM. 1,4-Disubstituted-[1,2,3]triazolyl-containing analogs of MT-II - design, synthesis, conformational analysis, and biological activity. *J Med Chem*. **2014** Oct 27.

3) Scrima M, Lauro G, **Grimaldi M**, Di Marino S, Tosco A, Picardi P, Gazzero P, Riccio R, Novellino E, Bifulco M, Bifulco G, D'Ursi AM. Structural Evidence of N6-Isopentenyladenosine As a New Ligand of Farnesyl Pyrophosphate Synthase. *J. Med. Chem*. **2014** Sep. 15.

4) Scrima M, Di Marino S, **Grimaldi M**, Campana F, Vitiello G, Piotta SP, D'Errico G, D'Ursi AM. Structural features of the C8 antiviral peptide in a membrane-mimicking environment. *Biochim Biophys Acta*. **2014**; 1838(3): 1010-8.

5) Skjaerven L, Codutti L, Angelini A, **Grimaldi M**, Latek D, Monecke P, Creyer MK, Carlomagno T. Accounting for Conformational Variability in Protein-Ligand Docking with NMR-Guided Rescoring. *J Am Chem Soc*. **2013**; 135(15): 5819-27.

6) Scrima M, **Grimaldi M**, Di Marino S, Testa C, Rovero P, Papini AM, Chorev M, D'Ursi AM. Solvent independent conformational propensities of

[1,2,3]triazolyl-bridged parathyroid hormone-related peptide-derived cyclonapeptide analogues. *Biopolymers*. **2012**; 98(6):535-45.

7) Di Marino S, Scrima M, **Grimaldi M**, D'Errico G, Vitiello G, Sanguinetti M, De Rosa M, Soriente A, Novellino E, D'Ursi AM. Antifungal peptides at membrane interaction. *Eur J Med Chem*. **2012**; 51:154-62.

8) Vitiello G, **Grimaldi M**, D'Ursi AM, D'Errico G. The iA $\beta$ 5p  $\beta$ -breaker peptide regulates the A $\beta$  (25-35) interaction with lipid bilayers through a cholesterol-mediated mechanism. *Biochem Biophys Res Commun*. **2012** Jan 6;417(1):88-92.

9) Merlino A, Vitiello G, **Grimaldi M**, Sica F, Busi E, Basosi R, D'Ursi AM, Fragneto G, Paduano L, D'Errico G. Destabilization of lipid membranes by a peptide derived from glycoprotein gp36 of feline immunodeficiency virus: a combined molecular dynamics/experimental study. *J Phys Chem B*. **2012** Jan 12;116(1):401-12.

## Table of Contents

Abstract.....	I-III
<b>Introduction.....</b>	<b>1</b>
<b>1.1 Protein-Observed Experiments.....</b>	<b>2</b>
1.1.1 Chemical Shift Mapping.....	3
1.1.2. Techniques Utilizing Isotopically Labeled Protein.....	4
<b>1.2 Ligand-Observed Experiments.....</b>	<b>7</b>
1.2.1 Saturation Transfer Difference (STD).....	9
1.2.2 Water-ligand observed via gradient spectroscopy (WaterLOGSY).....	11
1.2.3 INPHARMA.....	12
<b>References.....</b>	<b>16</b>
<b>1. NMR study of C8 antiviral peptide with Gp36-MPER.....</b>	<b>20</b>
<b>1.1 Introduction.....</b>	<b>20</b>
<b>1.2 Results and Discussion.....</b>	<b>24</b>
1.2.1 Isotope labeled Gp36-MPER gene expression.....	24
1.2.2 NMR spectroscopy.....	26
1.2.2 NMR interaction study: Chemical Shift Mapping.....	33
<b>1.3 Conclusions.....</b>	<b>35</b>
<b>Materials and Methods.....</b>	<b>36</b>
Expression of isotope labeled Gp36-MPER protein.....	36
NMR experiments: sample preparation and analysis.....	36
NMR structure calculation.....	37
<b>Acknowledgment.....</b>	<b>38</b>
<b>References.....</b>	<b>39</b>
<b>2. New promising ligands for FPPS antitumoral target.....</b>	<b>43</b>
<b>2.1 Introduction.....</b>	<b>43</b>
<b>2.2 Results and Discussion.....</b>	<b>48</b>
2.2.1 FPPS gene expression.....	48
2.2.2 NMR Experiments: i6A-FPPS interaction.....	50
2.2.3 NMR Experiments: i6A analogs-FPPS interaction.....	56
2.2.4 Enzymatic Assay of FPPS by NMR.....	64

<b>2.3 Conclusions</b> .....	67
<b>Materials and Methods</b> .....	68
<i>Expression of FPPS enzyme</i> .....	68
<i>NMR Experiments</i> .....	68
<b>References</b> .....	71
<b>3. Development of INPHARMA-STRING filtering protocol</b> .....	73
<b>3.1 Introduction</b> .....	73
<b>3.2 Results and Discussion</b> .....	75
<i>3.2.1 Structural Characterization and Docking Evaluation</i> .....	75
<i>3.2.2 NMR experiments and rescoring</i> .....	77
<i>3.2.3 INPHARMA-STRING</i> .....	85
<i>3.2.4 Docking Scenario</i> .....	87
<b>3.3 Conclusions</b> .....	89
<b>Materials and Methods</b> .....	90
<i>INPHARMA Experiments and Calculations</i> .....	90
<i>Structural Analysis, MD Simulations and Ensemble Docking</i> .....	92
<b>References</b> .....	94

## Abstract

My PhD project was focused on the study of protein-ligands interactions using different NMR techniques. NMR has a long history in drug discovery and hit-to-lead optimization. Compared to many other biophysical techniques, NMR has the advantage of combining structural and functional parameters to characterize protein inhibitor interactions. NMR experiments for protein-ligands interactions can be classified into two main categories: protein-observed and ligand-observed experiments.

Using protein-observed NMR experiments, such as chemical shift mapping, I studied the Gp36-MPER/C8 interaction. In the context of Feline Immunodeficiency Virus (FIV), we previously demonstrated that several short synthetic peptides that mimic the MPER of gp36 reduce the infectivity of FIV. In particular, an octapeptide, dubbed C8, elicited antiviral activity as a result of blocking cell entry, as observed for HIV fusion inhibitors. In the hypothesis that C8, similarly to T20 peptide, behaves as a fusion inhibitor peptide, interacting with NHR portion of gp36 of FIV, blocking the formation of six-helix bundle, we studied C8 for its ability to interact with the MPER region of gp36 (<sup>738</sup>M-K<sup>785</sup>) named Gp36-MPER. The study was performed according to a protein-observed NMR approach, consistently I determined the 3D NMR structure of the full Gp36-MPER by acquiring 3D heteronuclear NMR spectra on a <sup>13</sup>C and <sup>15</sup>N double labeled protein sample. NMR structure of Gp36-MPER was calculated on the basis of NOE data evidencing the presence of multiple segments of  $\alpha$ -helical and  $\beta$ -turn conformations. Interaction of Gp36-MPER with C8 peptide was demonstrated on the basis of the observation of chemical shift perturbation in <sup>15</sup>N-NHQC spectra.

Using ligand-observed NMR experiments, I studied two protein-ligands complexes: one on Farnesyl Diphosphate Synthase (FPPS) protein and the other on Protein Kinase A (PKA) protein.

FPPS is a key enzyme in the mevalonate, isoprenoid biosynthesis pathway. It is known to be the target of bisphosphonates, but it is also involved in microbial infections, cancer, and hypertension. N6-Isopentenyladenosine (i6A) is a modified nucleoside exhibiting anti-tumor effects on human and murine cells. During my PhD, I demonstrated the structural interaction of i6A with FPPS by recording saturation transfer difference (STD) and WaterLOGSY NMR experiments. i6A was demonstrated to occupy FPPS enzymatic pocket with a calculated  $K_D$  of ~1mM. Based on the i6A-FPPS interaction data, new ligands, analogs of i6A, were screened. STD NMR data demonstrate that the introduction in the 6-position of adenosine ring of a benzyl moiety induce a significant increasing in the interaction with FPPS target.

PKA is an enzyme involved in several functions in the cell, including regulation of glycogen, sugar and lipid metabolism. At the EMBL laboratory of Heidelberg, Germany, under the supervision of Prof. Teresa Carlomagno, I analyzed a set of PKA binding ligands provided by Sanofi Aventis, using INPHARMA method. The INPHARMA method, developed by Prof. Carlomagno, is based on the observation of interligand, spin-diffusion mediated, transferred-NOEs (nuclear Overhauser effects), between two ligands L1 and L2 binding competitively and weakly to a receptor T. In particular, my contribution to the INPHARMA project was the development of a new protocol in which the INPHARMA data are used to select the correct relative orientation of ligand pairs in a pool of complex and numerous structures generated by molecular docking calculation. This new approach, called INPHARMA-STRING, allows to improve the degree of selection of molecular

docking data, leading INPHARMA to be a solid tool applicable to an even large amount of targets and ligands.



## ***Introduction***

The study of the binding of small molecules (ligands) to protein targets is determinant to elucidate the molecular mechanism controlling important biological functions and to design novel bioactive molecules endowed with therapeutic action.

Protein-ligand interactions can be studied with a large variety of biophysical techniques: surface plasmon resonance (SPR),<sup>1</sup> isothermal titration calorimetry (ITC),<sup>2</sup> mass spectrometry (MS),<sup>3</sup> microscale thermophoresis (MST)<sup>4</sup> and nuclear magnetic resonance (NMR).<sup>5,6</sup>

NMR has the advantage of combining structural and functional parameters to characterize protein-ligand interactions. Accordingly NMR analysis of protein-ligand interactions may lead to the identification of binding sites, both on protein and on ligand; the characterization of structural changes induced by ligand binding, may provide the basis for the determination of binding affinity and specificity. Furthermore, as NMR reveals protein-ligand interaction with a large range of affinities ( $10^{-9}$ -  $10^{-3}$  M), it is a suitable technique to screen compound library and to identify lead compounds.<sup>7</sup>

NMR methods can be applied in the context of SBDD (Structure-Based Drug Design). SBDD relies on the knowledge of the three dimensional structure of the biological target in complex with ligands of different chemical structure, obtained through methods such as X-ray crystallography or NMR spectroscopy. The principle of SBDD lies in the combination of different chemical moieties with the intend to identify a molecule that, while possessing the pharmacological properties necessary for a drug, is complementary in

shape to the receptor binding pocket. This process requires knowledge of the exact structure of the protein-ligand complexes.

X-ray crystallography is a very important tool to study the three<sup>8-11</sup> dimensional structure of the proteins but often it is limited by high cost and by the insufficiency of good crystallographic data. On the other hand, in recent years NMR has been applied with considerable success to study protein-ligand interaction, becoming a valid instrument in the SBDD. However, the determination of a high resolution NMR structure of ligand-protein complex, may be limited by the chemical-physical properties of the target.

NMR experiments for protein-ligand interactions take advantage of the difference in NMR parameters characterizing the bound and free state of the ligand. In particular, the ligand assumes all the NMR parameters of protein when it is in the bound state. NMR experiments to analyze protein-ligand interaction can be classified into two main categories: protein-observed and ligand-observed NMR experiments.<sup>12-14</sup>

### **1.1 Protein-Observed Experiments**

Protein-observed methods provide specific and powerful information to immediately distinguish specific from non-specific ligand binding site on the protein.

In protein-observed experiments the typical followed approach is the “chemical shift mapping” consisting in the observation of protein chemical shift perturbations in response to the addition of ligands.<sup>15, 16</sup> The major drawback of the protein-observed NMR experiments is the necessity of having chemical shift assignment and structure calculation relative the protein target.

Protein NMR analysis is a very time consuming process including: i) production of isotopically labeled protein samples ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^2\text{H}$ ); ii) acquisition of homo- and hetero- correlated 3D NMR spectra; iii) chemical shift assignment; iv) NMR structure calculation.

### 1.1.1 Chemical Shift Mapping

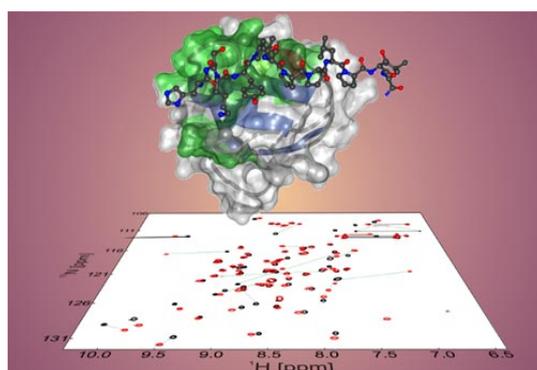
The chemical shift mapping allows for the identification of the ligand that interacts with protein target and the exact location of the binding site on the protein target.

Usually a series of experiments are recorded in which increasing amounts of a given ligand are added to the protein of interest, while the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts are monitored. Typically,  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra are used since  $^1\text{H}$ - $^{15}\text{N}$  HSQC is a highly sensitive experiment in which the peaks are generally well resolved.

The chemical shifts of the bound and free states depend on the binding affinity and on the binding kinetic. The success of the chemical shift mapping indeed, strongly depends on the ligand-protein binding kinetic.

In the *fast exchange systems* the exchange between the bound and free form is faster than the difference in chemical shifts; on the contrary in the *slow exchange systems* the exchange between the bound and free form is slower than the difference in chemical shifts, therefore the bound and free states give separate signals. In the *intermediate system* the rate of exchange between the bound and free states is comparable to the difference in the chemical shifts between the two states.<sup>18</sup> Chemical shift mapping allows for the discrimination between short and long distance effects. Short distance effects are shifts resulting from the interaction of specific residues of the protein with the

ligand, revealing the binding site of the ligand. On the contrary, long distance effects are shifts resulting from structural rearrangement of the protein under ligand binding. These induce misleading information on the binding site zone.



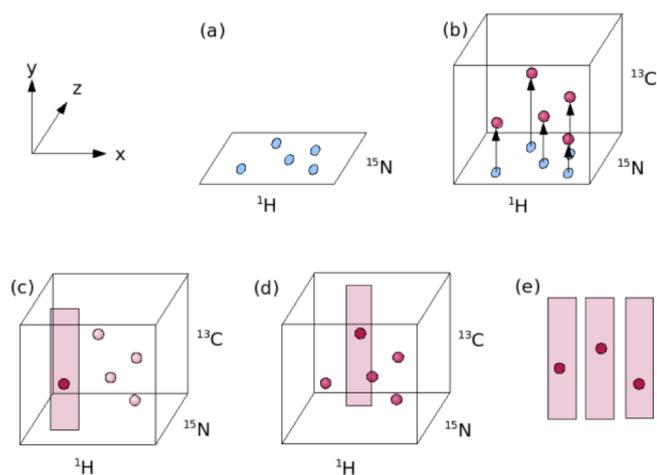
**Figure 1:** Chemical Shift Mapping.

### 1.1.2 Techniques Utilizing Isotopically Labeled Protein

Protein-observed NMR methods rely on protein production using isotopic labelling.<sup>19, 20</sup> Labeled proteins are expressed in *E. coli* using expensive labeled nutrients such as  $^{15}\text{N}$  ammonium chloride or  $^{13}\text{C}$  glucose in the culture media. This is a very common form of labeling for the assignment of backbone and side-chain  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  atoms using triple-resonance spectra. A high proportion of these assignments are required to accurately calculate the structure of protein.

3D experiments are based on 2D HSQC experiments, thus the X and Y axes are  $^1\text{H}$  and  $^{15}\text{N}$  respectively, and in the extended third dimension  $^{13}\text{C}$  signals are observable. 3D spectra are observable from different angles and planes, in form of 2D spectra (*strips*). Starting from HSQC ( $^1\text{H}$  is in the X-dimension

and  $^{15}\text{N}$  in the Y-dimension) the *strips* corresponding to the single HSQC peaks are picked out to be systematically analyzed.



**Figure 2:** Visualizing 3D spectra (Strips (e)).

3D NMR experiments can be classified in intra/inter-residue (HNCA, HNCACB and HN(CA)CO) and inter-residue (HN(CO)CA, HN(CO)CACB, HN(CO)HAHB and HNCO) experiments. The nomenclature for these triple resonance experiments reflects the magnetization transfer pathway of the experiments.<sup>21-24</sup>



The calculation of protein NMR structure critically depends on the assignment of the proton NMR resonances to identify NOE effects. This is possible by analyzing the 3D  $^{15}\text{N}$ -NOESY HSQC and  $^{13}\text{C}$ -NOESY HSQC.

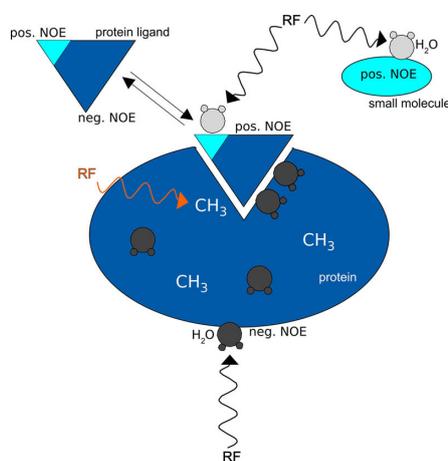
The quantitative evaluation of NOE data allows for the calculation of the interprotonic distances that are imposed as restraints into molecular dynamics calculation to build the 3D protein structure. The structure determination is based on molecular dynamics calculation using a simulated annealing process.<sup>33</sup> In the simulated annealing the protein atoms are virtually heated and cooled successively, while the potential functions are turned on to form the structure. The repeated heating and cooling process is meant to help energetically unfavorable structures to overcome energy barriers and end up in energetically more favorable structures, which may resist the subsequent heating process. Typically, one hundred structures are calculated, and those structures which comply best to the NMR input data and are energetically most favorable, are selected to define NMR structure bundle. The reproducibility of the structure calculation is judged on the basis of the root mean square deviation, RMSD, of every single atom in the structure divided by the number of atoms.<sup>34</sup>

## 1.2 Ligand-Observed Experiments

Ligand-observed experiments can be used to measure the protein-ligand affinity. All ligand-observed experiments are based on the difference in NMR parameters between the bound and free states of ligand. Usually, ligands with low molecular weight exhibit short correlation times ( $\tau_c$ ) and positive NOEs (nuclear Overhauser effects), while proteins with high molecular weight show large  $\tau_c$  and negative NOEs.<sup>13, 35</sup> Therefore, when ligand binds to target, its

NMR properties change and it assumes the NMR properties of high molecular weight molecule.

Ligand-observed experiments have the advantages of requiring minimal amounts of protein even of large size. Their acquisition is based on very fast mono-dimensional NMR experiments, while the complicate process of NMR structure solution is bypassed.



**Figure 4:** Ligand-Observed experiments principles.

Some of the more extensively used ligand-observed NMR experiments are the following: transferred-NOE spectroscopy (NOESY),<sup>35</sup> saturation transfer difference (STD)<sup>36</sup> spectroscopy, water-ligand observed via gradient spectroscopy (WaterLOGSY),<sup>37-39</sup> inter-ligands overhauser effects (ILOEs)<sup>40</sup> and inter-ligands NOEs for pharmacophore mapping (INPHARMA)<sup>41</sup> experiments.

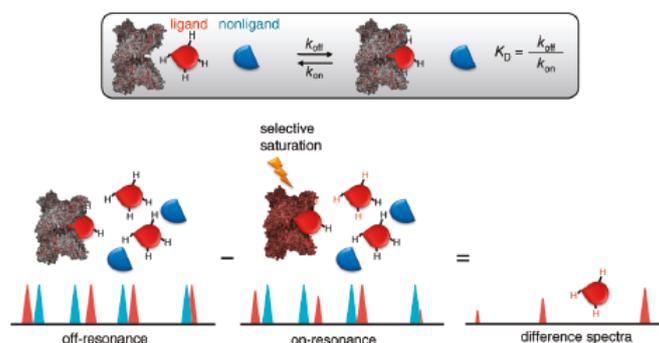
### 1.2.1 Saturation Transfer Difference (STD)

Saturation transfer difference (STD) NMR is one of the most popular ligand-based NMR techniques used for the screening of small molecule libraries.<sup>36, 42</sup> It's based on the nuclear Overhauser effect (NOE) and is focused on the signals of the ligand, without any need of acquiring and processing NMR data for the protein.

STD experiment allows to observe only the signals of the ligands that binds the protein. The STD experiment is based on the subtraction of two mono-dimensional NMR experiments: i) a spectrum is acquired saturating selectively the protein, by irradiating at a region of the spectrum that contains only resonances of the protein (*on-resonance* spectrum, with signals intensities  $I_{SAT}$ ), ii) a spectrum is recorded without protein saturation (*off-resonance* spectrum, with signals intensities  $I_0$ ). In the difference spectrum ( $I_{STD}=I_0-I_{SAT}$ ) only the signals of the ligands that received the magnetization transfer from the protein, via spin diffusion and NOE effect, are observed.<sup>43</sup>

As exchange between the bound and the free ligand state occurs, STD-NMR technique can be applied to protein-ligands systems having dissociation constants  $K_D$ , between  $10^{-8}$  and  $10^{-3}$  M.

Generally, for initial screening, STD experimental conditions consisting of protein-ligand ratio 1:100 and saturation time parameters 1-2s are used. For a quantitative analysis of STD experiments, experiments with different protein-ligand ratio ranging from 1:10 to 1:100 and with different saturation times and relaxation delay parameters are advisable.



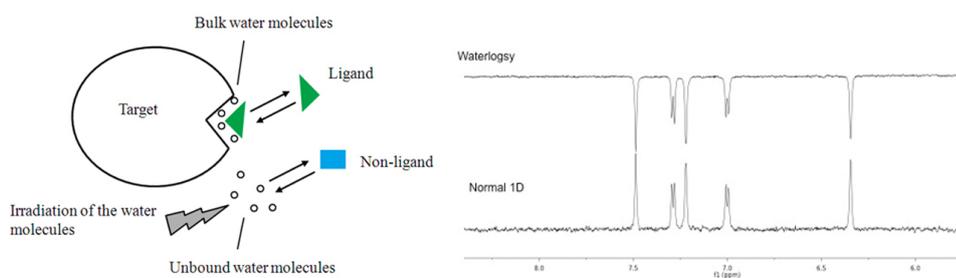
**Figure 5:** Saturation Transfer Difference (STD).<sup>43</sup>

In fact, STD intensities can be quantitatively analyzed to determine the dissociation constant  $K_D$  of protein-ligand complexes. A classic method proposed by Mayer and Meyer<sup>36</sup> provided the calculation of STD amplification factor ( $STD_{AF}$ ), which is the intensity of an STD signal corrected by the molar ratio protein-ligand ( $STD_{AF} = I_{STD}/I_0 \times L/P \text{ molar ratio}$ ). This method may induce an under-estimation of  $K_D$ , therefore two new methods have been developed. One of these is based on the  $K_D$  measurement in the presence of an inhibitor, and the second is based on the direct evaluation of  $K_D$ , determining the initial slopes of the build-up curves of  $STD_{AF}$  values vs the saturation time.<sup>44</sup>

The quantitative interpretation of STD experiments using epitope mapping (GEM)<sup>45</sup> and CORCEMA-STD<sup>46, 47</sup> enables the quantitative estimation of protein-ligand binding, revealing at the same time details on the ligand orientation. Using these two methods it is possible to extract structural information from STD experiments, connecting the relative magnitude of the STD signals to the proximity of the ligand protons to the protein.

### 1.2.2 Water-ligand observed via gradient spectroscopy (WaterLOGSY)

Similar to STD, WaterLOGSY method<sup>37-39</sup> relies on the excitation of the protein-ligand complex. WaterLOGSY, applies indirect saturation of the protein using H<sub>2</sub>O molecules in the binding pocket as an intermediate magnetization pool. In the WaterLOGSY experiment, the source magnetization originates from bulk solvent (H<sub>2</sub>O) protons instead of target protons (STD experiments).



**Figure 6:** WaterLOGSY principles.

The inverted water magnetization can be transferred via three different pathways: i) direct cross-relaxation between the bound ligand and water molecules immobilized in the protein binding site; ii) direct cross-relaxation between exchangeable protein protons within the binding site (amide, hydroxyl, amino, etc.) and the bound ligand; iii) indirect cross-relaxation, transfer via spin diffusion, from the water molecules found in the protein surface to the exchangeable protein protons within the binding site. Accordingly *binders* ligands are characterized by negative NOE with water, while *non-binders* ligands are characterized by positive NOE with water. Similarly to the STD experiment, WaterLOGSY spectra are recorded with the

protein concentration in the low  $\mu\text{M}$  range and protein-ligand ratio of 1:100. Mixing time is critical parameter to optimize WaterLOGSY experiment; in general, the range reported in the literature is from 1 to 3 s.

In drug discovery, WaterLOGSY experiments are used for the screening of weak ligands ( $K_D$   $10^{-6}$  -  $10^{-3}$  M range), to validate the hits identification, to identify the ligand bindings site using competitor and to calculate the dissociation constants with competition and titration experiments.

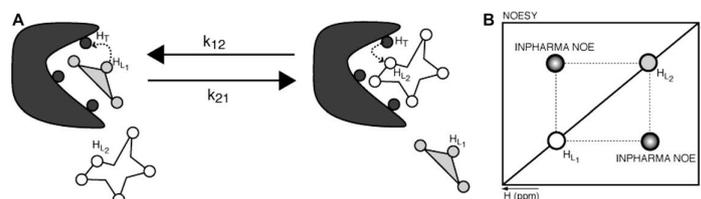
### 1.2.3 INPHARMA

INPHARMA (Interligand NOEs for PHARmacophore MApping)<sup>41, 48-50</sup> is a ligand-observed method to determine the relative binding modes of two low-affinity ligands that bind competitively to a common receptor site.

Molecular docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets, in order to predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational drug design.<sup>51, 52</sup> In order to perform a molecular docking screening, the first requirement is the structure of the protein-target of interest. The data relative to the protein structure and a database of potential ligands serve as inputs for the molecular docking calculation. In this procedure the validity of the binding affinity prediction is judged on the basis of *scoring function*.<sup>53</sup> This parameterizes the energy of the pose on the basis of physics-based molecular mechanics *force field*; a low energy indicates a stable system and thus a likely binding interaction. *Scoring functions* may be consistent with correct binding prediction or sometimes with incorrect binding prediction, potentially yielding a large number of false positive hits.

In order to reduce the number of false positives, one can try to develop new more accurate computational *scoring functions*. An efficient route is to combine computational with experimental data, to select the correct binding modes. INPHARMA is a suitable NMR based technique to produce reliable experimental data. It can be used as an instrument to rank and select binding modes of ligands obtained by docking, to overcome the intrinsic limitations of the molecular docking *scoring functions*<sup>54-56</sup> INPHARMA was successfully applied to various systems.<sup>49, 57</sup> However, in order to improve the degree of selection of molecular docking outputs, INPHARMA-STRING<sup>58</sup> has been implemented. The new approach has been developed as a new, solid tool applicable to a large amount of targets and ligands and having a determinant impact in SBDD.

The INPHARMA method is based on the observation of inter-ligand, spin-diffusion mediated, transferred-NOEs, between two ligands  $L_1$  and  $L_2$  binding competitively and weakly to a receptor T. During the mixing time of the NOESY experiment,  $L_1$  binds to the receptor and its protons  $HL_1$  transfer their magnetization to the receptor protons (HTs). During the same mixing time of the NOESY experiment  $L_1$  dissociates from the receptor and  $L_2$  binds. The magnetization that was transferred from  $HL_1$  to HT can now be transferred from HT to  $HL_2$ . This process generates NOEs peaks between  $HL_1$  and  $HL_2$ , called INPHARMA NOEs. In this way, the binding pocket of the target macromolecule can be mapped on the resonances of the two ligands.<sup>41</sup>



**Figure 7:** Schematic representation of the principle of the INPHARMA NOEs.

For decades, the NOEs have been used for the three dimensional structure determination of bio-molecules in solution using NOESY experiments, converting the intensity of the NOE peaks in distances restraints. However, the analysis of the NOEs, often, had neglected the spin diffusion effect, which is one of the major causes for inaccuracy in deriving distances restraints. Several approaches have been developed in order to correct NOEs restraints for spin diffusion; one of these is the use of the full-relaxation-matrix formalism.<sup>41</sup>

Intensities of NOE cross-peaks of a receptor/ligand complex are described by the following Solomon equation:

$$\frac{dM(t)}{dt} = -(R + K) \cdot (M(t) - M_0) \quad (1)$$

with a solution of the form:

$$M(\tau_m) = \exp(-(R + K)\tau_m) \cdot (M(0) - M_0) + M_0 \quad (2)$$

where  $\mathbf{R}$  is the relaxation matrix,  $\mathbf{K}$  is the exchange matrix describing the exchange of the ligand between the free, and the protein bound forms,  $\mathbf{M}_0$  is the equilibrium magnetization,  $\mathbf{M}(0)$  is the starting magnetization, and  $\mathbf{t}_m$  is

the mixing time. The relaxation matrix is a diagonal block matrix where each sub-matrix describes the proton-proton relaxation pathway of one species; individual elements of the relaxation matrix are expressed as follows:

$$R_{i,i} = \rho_i = \sum_{\substack{H_j \in A \\ j \neq i}} \frac{b^2}{d_{ij}^6} \cdot (J(0) + 3J(\omega) + 6J(2\omega)) \quad (3)$$

$$R_{i,j} = \sigma_i = \frac{b^2}{d_{ij}^6} \cdot (6J(2\omega) - J(0)) \quad (4)$$

where  $\rho_i$  is the longitudinal relaxation rate of proton  $H_i$ ,  $\sigma_{ij}$  is the cross-relaxation rate of protons  $H_i$  and  $H_j$  and  $d_{ij}$  is the distance between two protons  $H_i$  and  $H_j$  in the same chemical species.<sup>41</sup>

On these theoretical bases, in the INPHARMA approach, the interligand NOEs are interpreted exclusively with the help of theoretical calculations by using the full-relaxation matrix approach, to a rapid determination of the binding epitope of different ligands to a common target.

At the current state of art INPHARMA approach has already been successfully applied to protein kinase A (PKA)<sup>49</sup> and Epo A-tubulin<sup>57</sup> complex. Nowadays NMR derived inter-ligand INPHARMA-NOEs are used like a new rescoring function of computationally generated ligand binding modes. Recently INPHARMA-STRING has been developed as an extended protocol which rigorously determines the correct protein-ligand binding modes.<sup>58</sup>

## References

1. Patching, S. G. Surface plasmon resonance spectroscopy for characterisation of membrane protein-ligand interactions and its potential for drug discovery. *Biochimica Et Biophysica Acta-Biomembranes* **2014**, 1838, 43-55.
2. Jecklin, M. C.; Schauer, S.; Dumelin, C. E.; Zenobi, R. Label-free determination of protein-ligand binding constants using mass spectrometry and validation using surface plasmon resonance and isothermal titration calorimetry. *Journal of Molecular Recognition* **2009**, 22, 319-329.
3. Jonker, N.; Kool, J.; Irth, H.; Niessen, W. M. A. Recent developments in protein-ligand affinity mass spectrometry. *Analytical and Bioanalytical Chemistry* **2011**, 399, 2669-2681.
4. Seidel, S. A. I.; Wienken, C. J.; Geissler, S.; Jerabek-Willemsen, M.; Duhr, S.; Reiter, A.; Trauner, D.; Braun, D.; Baaske, P. Label-Free Microscale Thermophoresis Discriminates Sites and Affinity of Protein-Ligand Binding. *Angewandte Chemie-International Edition* **2012**, 51, 10656-10659.
5. Cala, O.; Guilliere, F.; Krimm, I. NMR-based analysis of protein-ligand interactions. *Analytical and Bioanalytical Chemistry* **2014**, 406, 943-956.
6. Pellecchia, M.; Meininger, D.; Dong, Q.; Chang, E.; Jack, R.; Sem, D. S. NMR-based structural characterization of large protein-ligand interactions. *Journal of Biomolecular Nmr* **2002**, 22, 165-173.
7. Śledź, P.; Abell, C.; Ciulli, A. Ligand-Observed NMR in Fragment-Based Approaches. *NMR of Biomolecules: Towards Mechanistic Systems Biology* **2012**, 264-280.
8. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, 274, 1531-1534.
9. Mercier, K. A.; Shortridge, M. D.; Powers, R. A Multi-Step NMR Screen for the Identification and Evaluation of Chemical Leads for Drug Discovery. *Combinatorial Chemistry & High Throughput Screening* **2009**, 12, 285-295.
10. Ross, A.; Salzmann, M.; Senn, H. Fast-HMQC using Ernst angle pulses: An efficient tool for screening of ligand binding to target proteins. *Journal of Biomolecular Nmr* **1997**, 10, 389-396.
11. Feliz, M.; Garcia, J.; Aragon, E.; Pons, M. Fast 2D NMR ligand screening using Hadamard spectroscopy. *Journal of the American Chemical Society* **2006**, 128, 7146-7147.
12. Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T. L.; Homans, S. W.; Kessler, H.; Luchinat, C.; Meyer, B.; Oschkinat, H.; Peng, J.; Schwalbe, H.; Siegal, G. Perspectives on NMR in drug discovery: a technique comes of age. *Nature Reviews Drug Discovery* **2008**, 7, 738-745.
13. Peng, J. W.; Moore, J.; Abdul-Manan, N. NMR experiments for lead generation in drug discovery. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2004**, 44, 225-256.
14. Goldflam, M.; Tarrago, T.; Gairi, M.; Giralt, E. NMR Studies of Protein-Ligand Interactions. *Protein Nmr Techniques, Third Edition* **2012**, 831, 233-259.
15. Medek, A.; Hajduk, P. J.; Mack, J.; Fesik, S. W. The use of differential chemical shifts for determining the binding site location and orientation of protein-bound ligands. *Journal of the American Chemical Society* **2000**, 122, 1241-1242.

16. Krishnamoorthy, J.; Yu, V. C. K.; Mok, Y.-K. Auto-FACE: An NMR Based Binding Site Mapping Program for Fast Chemical Exchange Protein-Ligand Systems. *Plos One* **2010**, *5*.
17. Bax, A.; Grzesiek, S. Methodological advances in protein nmr. *Accounts of Chemical Research* **1993**, *26*, 131-138.
18. Reibarkh, M.; Malia, T. J.; Wagner, G. NMR distinction of single- and multiple-mode binding of small-molecule protein ligands. *Journal of the American Chemical Society* **2006**, *128*, 2160-2161.
19. Roberts, G. C. K. Applications of NMR in drug discovery. *Drug Discovery Today* **2000**, *5*, 230-240.
20. Strauss, A.; Bitsch, F.; Cutting, B.; Fendrich, G.; Graff, P.; Liebetanz, J.; Zurini, M.; Jahnke, W. Amino-acid-type selective isotope labeling of proteins expressed in Baculovirus-infected insect cells useful for NMR studies. *Journal of Biomolecular Nmr* **2003**, *26*, 367-372.
21. Sattler, M.; Schleucher, J.; Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1999**, *34*, 93-158.
22. Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **2011**, *213*, 423-41.
23. Grzesiek, S.; Bax, A. Improved 3d triple-resonance nmr techniques applied to a 31-kda protein. *Journal of Magnetic Resonance* **1992**, *96*, 432-440.
24. Dalvit, C. Semi-selective 2-dimensional homonuclear and heteronuclear nmr experiments recorded with pulsed-field gradients. *Magnetic Resonance in Chemistry* **1995**, *33*, 570-576.
25. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, P.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins-Structure Function and Bioinformatics* **2005**, *59*, 687-696.
26. Johnson, B. A.; Blevins, R. A. NMR view - a computer-program for the visualization and analysis of nmr data. *Journal of Biomolecular Nmr* **1994**, *4*, 603-614.
27. Kneller, D. G.; Kuntz, I. D. UCSF SPARKY - an nmr display, annotation and assignment tool. *Journal of Cellular Biochemistry* **1993**, 254-254.
28. Eccles, C.; Guntert, P.; Billeter, M.; Wuthrich, K. Efficient analysis of protein 2D NMR spectra using the software package EASY. *Journal of biomolecular NMR* **1991**, *1*, 111-30.
29. Linge, J. P.; Habeck, M.; Rieping, W.; Nilges, M. ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* **2003**, *19*, 315-316.
30. Guntert, P.; Mumenthaler, C.; Wuthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *Journal of Molecular Biology* **1997**, *273*, 283-298.
31. Guntert, P. Automated NMR structure calculation with CYANA. *Methods in molecular biology (Clifton, N.J.)* **2004**, *278*, 353-78.
32. Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance* **2003**, *160*, 65-73.
33. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, A package of computer-programs for

applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995**, 91, 1-41.

34. Gottstein, D.; Kirchner, D. K.; Guentert, P. Simultaneous single-structure and bundle representation of protein NMR structures in torsion angle space. *Journal of Biomolecular Nmr* **2012**, 52, 351-364.

35. Meyer, B.; Peters, T. NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie-International Edition* **2003**, 42, 864-890.

36. Mayer, M.; Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie-International Edition* **1999**, 38, 1784-1788.

37. Dalvit, C.; Cottens, S.; Ramage, P.; Hommel, U. Half-filter experiments for assignment, structure determination and hydration analysis of unlabelled ligands bound to C-13/N-15 labelled proteins. *Journal of Biomolecular Nmr* **1999**, 13, 43-50.

38. Dalvit, C.; Pevarello, P.; Tato, M.; Veronesi, M.; Vulpetti, A.; Sundstrom, M. Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *Journal of Biomolecular Nmr* **2000**, 18, 65-68.

39. Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *Journal of Biomolecular Nmr* **2001**, 21, 349-359.

40. Li, D. W.; DeRose, E. F.; London, R. E. The inter-ligand Overhauser effect: A powerful new NMR approach for mapping structural relationships of macromolecular ligands. *Journal of Biomolecular Nmr* **1999**, 15, 71-76.

41. Orts, J.; Griesinger, C.; Carlomagno, T. The INPHARMA technique for pharmacophore mapping: A theoretical guide to the method. *Journal of Magnetic Resonance* **2009**, 200, 64-73.

42. Ji, Z.; Yao, Z.; Liu, M. Saturation transfer difference nuclear magnetic resonance study on the specific binding of ligand to protein. *Analytical Biochemistry* **2009**, 385, 380-382.

43. Viegas, A.; Manso, J.; Nobrega, F. L.; Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *Journal of Chemical Education* **2011**, 88, 990-994.

44. Angulo, J.; Enriquez-Navas, P. M.; Nieto, P. M. Ligand-Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR Spectroscopy: The Binding Isotherm of STD Initial Growth Rates. *Chemistry-a European Journal* **2010**, 16, 7803-7812.

45. Kemper, S.; Patel, M. K.; Errey, J. C.; Davis, B. G.; Jones, J. A.; Claridge, T. D. W. Group epitope mapping considering relaxation of the ligand (GEM-CRL): Including longitudinal relaxation rates in the analysis of saturation transfer difference (STD) experiments. *Journal of Magnetic Resonance* **2010**, 203, 1-10.

46. Jayalakshmi, V.; Krishna, N. R. Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *Journal of Magnetic Resonance* **2002**, 155, 106-118.

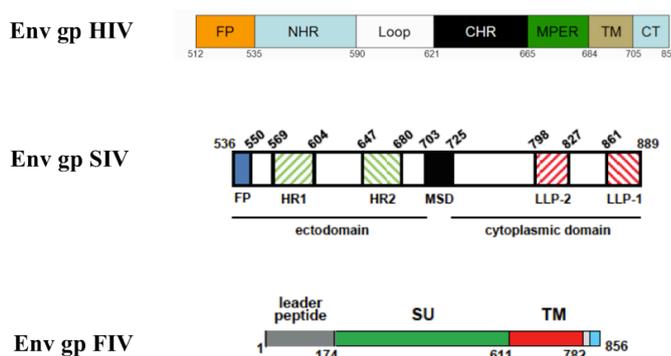
47. Jayalakshmi, V.; Krishna, N. R. CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities. *Journal of Magnetic Resonance* **2004**, 168, 36-45.

48. Orts, J.; Tuma, J.; Reese, M.; Grimm, S. K.; Monecke, P.; Bartoschek, S.; Schiffer, A.; Wendt, K. U.; Griesinger, C.; Carlomagno, T. Crystallography-independent determination of ligand binding modes. *Angewandte Chemie-International Edition* **2008**, *47*, 7736-7740.
49. Orts, J.; Bartoschek, S.; Griesinger, C.; Monecke, P.; Carlomagno, T. An NMR-based scoring function improves the accuracy of binding pose predictions by docking by two orders of magnitude. *Journal of Biomolecular Nmr* **2012**, *52*, 23-30.
50. Bartoschek, S.; Klabunde, T.; Defossa, E.; Dietrich, V.; Stengelin, S.; Griesinger, C.; Carlomagno, T.; Focken, I.; Wendt, K. U. Drug Design for G-Protein-Coupled Receptors by a Ligand-Based NMR Method. *Angewandte Chemie-International Edition* **2010**, *49*, 1426-1429.
51. Carlomagno, T. Ligand-target interactions: What can we learn from NMR? *Annual Review of Biophysics and Biomolecular Structure* **2005**, *34*, 245-266.
52. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery* **2004**, *3*, 935-949.
53. Jain, A. N. Scoring functions for protein-ligand docking. *Current Protein & Peptide Science* **2006**, *7*, 407-420.
54. Plewczynski, D.; Spieser, S. A. H.; Koch, U. Assessing different classification methods for virtual screening. *Journal of Chemical Information and Modeling* **2006**, *46*, 1098-1106.
55. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *Journal of Medicinal Chemistry* **2006**, *49*, 5912-5931.
56. Lexa, K. W.; Carlson, H. A. Protein flexibility in docking and surface mapping. *Quarterly Reviews of Biophysics* **2012**, *45*, 301-343.
57. Sanchez-Pedregal, V. M.; Reese, M.; Meiler, J.; Blommers, M. J. J.; Griesinger, C.; Carlomagno, T. The INPHARMA method: Protein-mediated interligand NOEs for pharmacophore mapping. *Angewandte Chemie-International Edition* **2005**, *44*, 4172-4175.
58. Skjaerven, L.; Codutti, L.; Angelini, A.; Grimaldi, M.; Latek, D.; Monecke, P.; Dreyer, M. K.; Carlomagno, T. Accounting for Conformational Variability in Protein-Ligand Docking with NMR-Guided Rescoring. *Journal of the American Chemical Society* **2013**, *135*, 5819-5827.

# ***1** NMR study of C8 antiviral peptide with Gp36-MPER*

## **1.1 Introduction**

The lentiviral envelope glycoproteins (Env) mediate virus entry by interacting with specific receptors present at the cell surface, thereby determining viral tropism and pathogenesis. The study of simian and feline immunodeficiency viruses (SIV and FIV, respectively), useful models of human immunodeficiency virus (HIV), provides a valuable tool for developing anti-HIV therapies and vaccines. (Figure 1)



**Figure 1:** (top) HIV-1 gp41 pre-fusion structure; (middle) the SIV-gp41 structure; (down) schematic diagram of the feline immunodeficiency virus (FIV) Env glycoprotein.

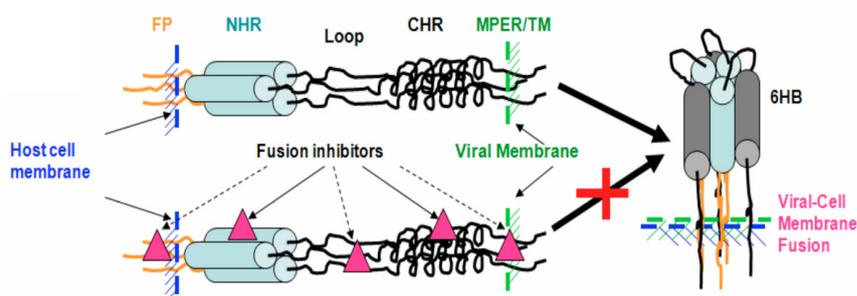
FIV is a lentivirus that resembles the human HIV.<sup>1-4</sup> Figure 1 shows schematic representations of FIV (gp36) and HIV (gp41) envelope glycoproteins. Increasing evidence suggest a common structural framework

for these glycoproteins, corresponding to similar roles in virus cell fusion.<sup>5-12</sup>

The gp41 ectodomain contains several characteristic functional domains, including the fusion peptide (FP), N-terminal heptad repeat (NHR), C-terminal heptad repeat (CHR) and membrane proximal extracellular region (MPER).

During the virus entry, NHR and CHR, automatically, fold back to form a low energy stable six-helical bundle (6HB) with NHR trimer as the inner core and anti-parallel binding of three CHRs.<sup>13</sup> (Figure 2)

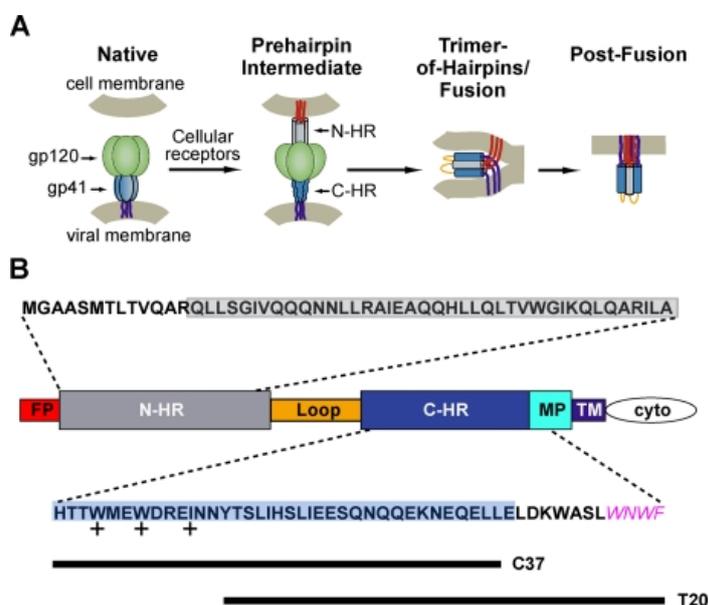
Peptides derived from gp41 NHR or CHR sequences, inhibit HIV-1 infection by interaction with their counterparts in gp41, to prevent 6HB formation and terminate the HIV-1-cell fusion process.



**Figure 2:** Fusion inhibitors that target gp41's functional domains.

In 2003, a 36-amino acid peptide derived from the CHR of the HIV-1 gp41 ectodomain (enfuvirtide T20) was approved by the US FDA for anti-HIV treatment. T20 is the only currently approved drug targeting gp41; however, the high cost and inconvenience of the twice daily injection of this peptide drug prevent it from being used as a regular anti-HIV drug.<sup>14</sup> The development of new fusion inhibitors that overcome the limitations of T20 is of great

importance. The NHR and CHR are still the most intensively investigated targets in gp41.<sup>15</sup> Other less exploited functional domains, such as the FP and the MPER,<sup>16, 17</sup> are receiving increasing attention as potential targets for fusion inhibitors. (Figure 3)<sup>18-26</sup>

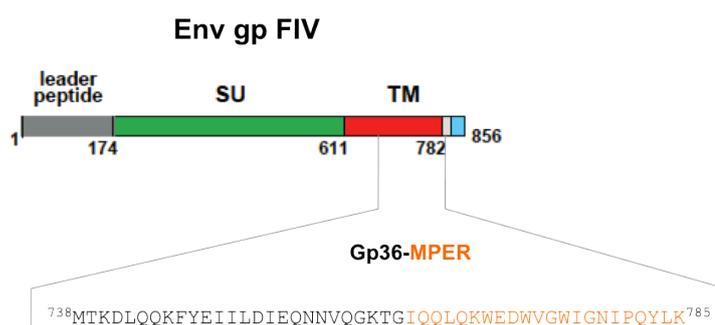


**Figure 3:** A, a model of HIV-1 entry, B, a diagram of HIV-1 gp41. The N-HR and C-HR segments found in the original 5-Helix are boxed in gray and blue, respectively, whereas the sequences of C37 and T20 are denoted by line.<sup>27</sup>

T20 (Fuzeon®, generic name: enfuvirtide), was approved by the U.S. FDA as the first fusion inhibitor.<sup>28-31</sup> Fuzeon® is the only currently approved drug targeting gp41; but the peptide drug is expensive and has the inconvenience of twice daily injection.<sup>32</sup> To overcome the limitations of Fuzeon®, NHR, CHR and others functional domains, such as FP,<sup>17</sup> MPER,<sup>16, 19, 20, 33</sup> are still the most intensively investigated to develop new drugs targeting gp41.<sup>34</sup>

In this context,<sup>5, 6, 10, 35</sup> we previously demonstrated that several short synthetic peptides that mimic the MPER of gp36 reduce the infectivity of FIV.<sup>36</sup> In particular, the fragment <sup>770</sup>WEDWVGWI<sup>776</sup>, dubbed C8, elicited antiviral activity as a result of blocking cell entry, as observed for HIV fusion inhibitors.<sup>23, 24</sup> A structure activity relationship (SAR) study and preliminary NMR conformational analysis demonstrated that C8's antiviral activity depends on the presence of regularly spaced Trp residues and that the orientation of the Trp indolyl rings is critical on a turn-shaped backbone conformations.<sup>37-41</sup>

In the hypothesis that C8, similarly to T20 peptide, behaves as a fusion inhibitor peptide, interacting with NHR coiled coil<sup>27</sup> portion of gp36 and blocking the formation of six-helix bundle (Figure 3), we studied C8 for its ability to interact with the MPER region of gp36 (<sup>738</sup>M-K<sup>785</sup>) named Gp36-MPER. The study was performed according to a protein-observed NMR approach, that included the NMR structure determination of Gp36-MPER and subsequently analysis of C8 binding by following a chemical shift mapping approach.



**Figure 4:** aminoacidic sequence of protein Gp36-MPER.

## 1.2 Results and Discussion

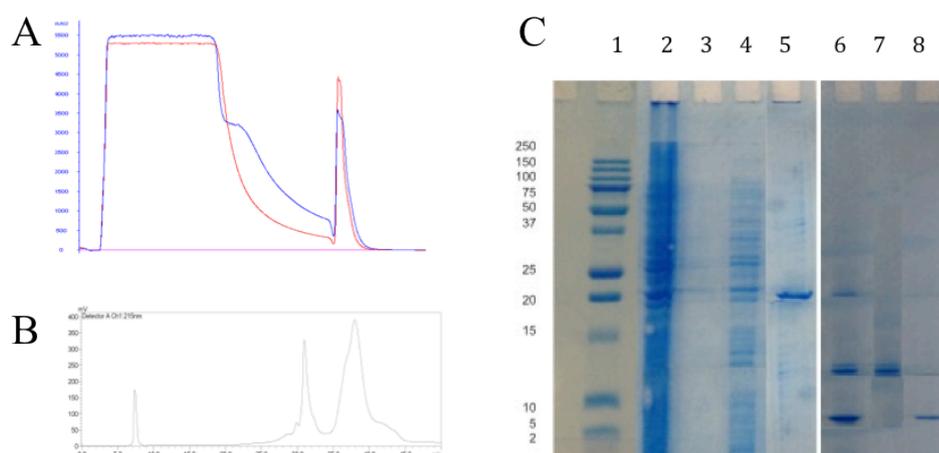
Gp36-MPER is a 48aa protein characterized in its sequence, by many aliphatic and long side chains aminoacids. Therefore the calculation of NMR structure using exclusively homonuclear NMR experiments is prohibitive. The production of  $^{15}\text{N}/^{13}\text{C}$  enriched Gp36-MPER protein is necessary to have resolved NMR spectra using multidimensional heteronuclear ( $^{15}\text{N}/^{13}\text{C}$ ) experiments.

### 1.2.1 Isotope labeled Gp36-MPER gene expression

The expression vector pET-31b(+) transformed into BLR21(DE3)-pLysS cells, was used for the expression in *E. coli*. For the expression in *E. coli* of uniformly  $^{15}\text{N}$  and  $^{13}\text{C}$ -isotope labeled Gp36-MPER protein, I prepared and autoclaved M9 minimal media<sup>42</sup> (dissolve 6g  $\text{Na}_2\text{HPO}_4$ , 3g of  $\text{KH}_2\text{PO}_4$ , 0,5g of NaCl in distilled water 900mL) containing 50  $\mu\text{g}/\text{ml}$  ampicillin and 1g of  $^{15}\text{N}$  ammonium sulfate. Then I added 20%  $^{13}\text{C}$  glucose solution, prepared with 2g of  $^{13}\text{C}$  glucose, 10mL of vitamin cocktails and trace metal mixtures (1mL of thiamine,  $\text{MgSO}_4$ ,  $\text{CaCl}_2$ ; 1M stock solutions sterilized by filtration), finally, we adjusted the final volume to 1L. The overnight culture (20 ml of LB (Luria Bertani) medium was used to inoculate 1 L of M9 medium. When bacterial clones were performed to an  $\text{OD}_{600}$  of 0.5 at 37 °C, the 1mM IPTG was added.

After 18h, cells were pelleted by centrifugation and re-suspended in lysis buffer (0,5M NaCl and 20mM Tris-HCl), sonicated and re-centrifuged. The pellet obtained, was suspended in a binding buffer (10mM Guanidine, 0,5M NaCl, 20mM Tris-HCl and 15mM Imidazole) and stirred for one night at 4°C. Then, the solution was purified with His-Trap<sup>TM</sup> HP column at 1 ml/min using

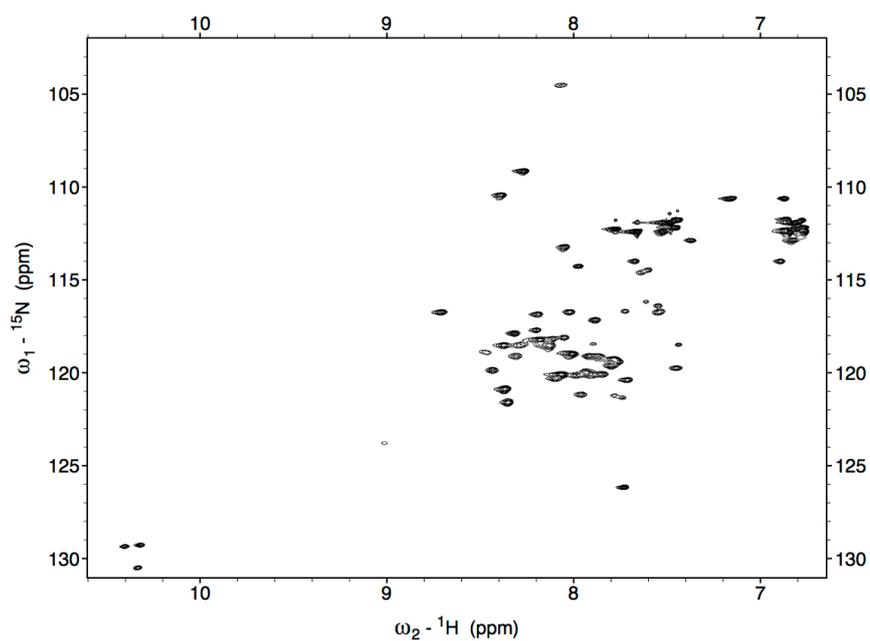
an AKTA purifier system, the protein was eluted from the column with elution buffer (10mM Guanidine, 0,5M NaCl, 20mM Tris-HCl and 50mM Imidazole). The eluted fraction was dialyzed and lyophilized. To remove the KSI tag<sup>43, 44</sup> lyophilized protein was dissolved in a minimum quantity of 70 % (v/v) formic acid and was treated with 0.5 M cyanogen bromide; the reaction stirred for 3h in the dark. After 3h, the solution was dialyzed and lyophilized, finally the protein was purified by HPLC.



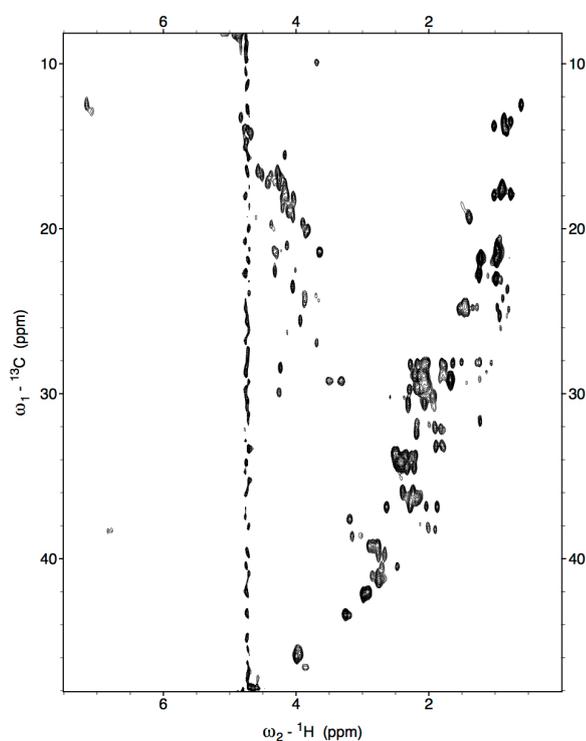
**Figura 5:** A) FPLC chromatogram ; B) HPLC chromatogram, obtained using gradient: 10-40 B in 20 min poi 40-70 B in 30 min, Buffer A: water + TFA 0.1%, Buffer B: Acetonitril + TFA 0.1%; C) SDS-PAGE of FPLC fractions (4-5), after cleavage (6) and HPLC fractions (7 KSI - 8 isotope labeled Gp36-MPER).

### 1.2.2 NMR spectroscopy

The preliminary inspection of 2D  $^{13}\text{C}$ -HSQC/ $^{15}\text{N}$ -HSQC of isotope labeled Gp36-MPER, showed a well dispersed HSQC spectra. (Figures 6-7)

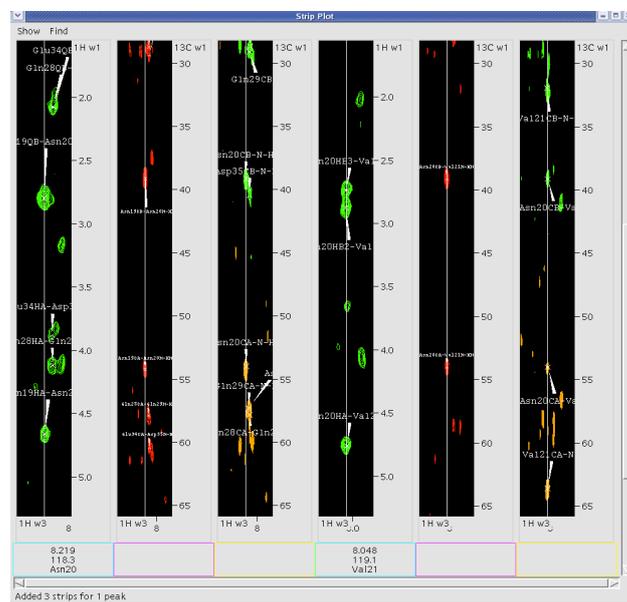
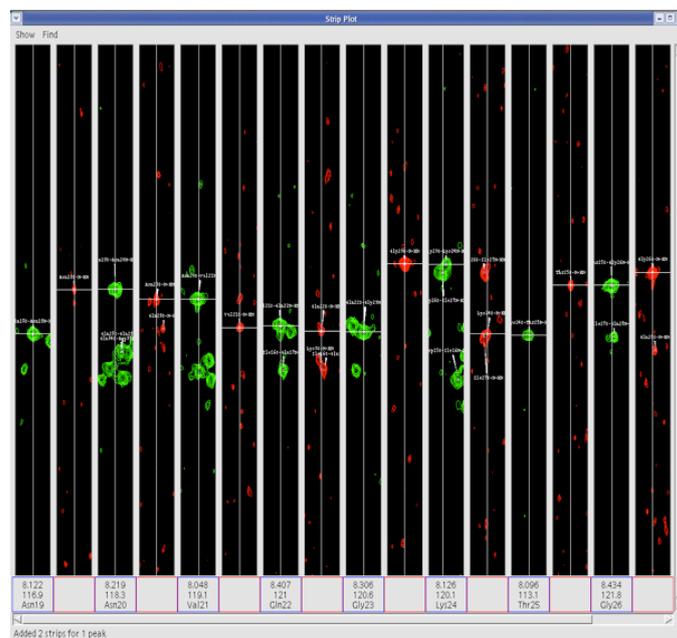


**Figure 6:** Full  $^{15}\text{N}$ -HSQC of isotope labeled Gp36-MPER.



**Figure 7:** Full folded  $^{13}\text{C}$ -HSQC of isotope labeled Gp36-MPER.

The 2D  $^{13}\text{C}$ -HSQC/ $^{15}\text{N}$ -HSQC spectra were used to generate the strip plots of acquired 3D experiments. To achieve chemical shift assignment, the strips of 3D experiments were generated. Figure 8 shows an example of backbone sequential assignment starting from  $^{13}\text{C}$ -HSQC/ $^{15}\text{N}$ -HSQC and iteratively analyzing 3D-HNCO and HN(CA)CO spectra. Similar procedure was used for the sequential assignment of protein side chains, by inspecting HBHA(CO)HN, CBCA(CO)HN and HNCACB heteronuclear spectra.



**Figure 8:** Sequential assignment of isotope labeled Gp36-MPER 3D spectra.

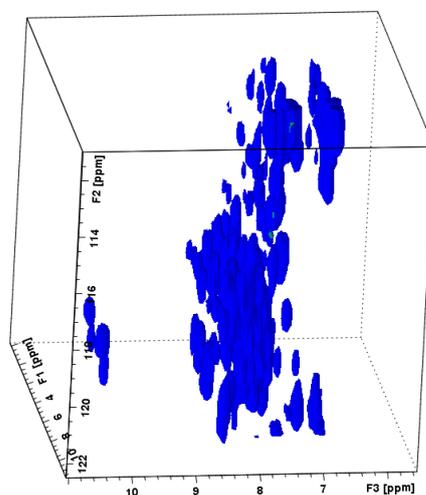
Table I reports the chemical shift of all  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  assigned atoms.

**Table I:** chemical shifts of Gp36-MPER in DPC/SDS mixed micelles.

Number Residue	Residue	HN	$^{15}\text{N}$	$^{13}\text{C}\alpha$	H $\alpha$	$^{13}\text{C}\beta$	H $\beta$	H $\gamma$	Others
738	M			54.721	4.755	30.415	2944/2.402		
739	T	8.691	117.037	65.428	4.317	68.285	4.240		
740	K	8.367	120.996	58.027	4.224	32.059	1.853		
741	D	7.798	119.368	56.500	4.619	40.919	2.829		
742	L	7.742	120.039	58.351	4.020	41.093	1.677		QQ $\delta$ 0.924
743	Q	8.337	117.844	60.200	3.840	28.657	2.277		
744	Q	7.932	117.198	58.724	4.144	28.082	2.252		
745	K	7.930	118.544	58.332	4.174		1.983	1.559	Q $\delta$ 1804
746	F	8.368	111.160	61.388	4.203	38.609	3.135		
747	Y	8.203	117.703	61.011	4.122	37.898	3.193		Q $\epsilon$ 7115
748	E	8.087	117.916	58.059	4.010	29.593	2.081	2.337	
749	I	7.918	119.996	59.189	4.064		1.896	1.524	Q $\delta$ 1 0.793
750	I	7.831	118.343	60.388	3.825		1.919	1.266/1.011	Q $\delta$ 1 0.830
751	L	7.938	120.068	62.522	3.999		1.701		QQ $\delta$ 0.807
752	D	7.733	119.459	56.732	4.516	41.157	2.704		
753	I	8.004	120.001	64.420	3.881	38.321	1.906	1.080/0.894	
754	E	8.392	121.035	58.875	4.073	29.445	2.129		
755	Q	8.222	117.382	58.145	4.131	28.752	2.159		
756	N	8.129	116.894	54.239	4.663	39.124	2.808		
757	N	8.243	118.270	53.917	4.747	39.112	2.903/2.728		
758	V	8.047	119.047	63.520	4.048	32.061	2.199	0.955	
759	Q	8.404	121.035	56.740	4.258	28.847	2.095/2.013		
760	G	8.214	108.979	45.836	4.160				
761	K	8.140	120.072	57.423	4.132	32.980	1.828	1.335	
762	T	8.104	113.124	59.757	4.362	69.866	4.650		
763	G	8.438	110.298	45.904	3.963				
764	I	8.166	120.087	59.331	4.029	38.117	1.913	1.420/0.894	
765	Q	8.420	121.463	58.899	4.095	28.727	2.350/2.062		
766	Q	8.152	118.457	57.054	4.241	28.681	2.065		
767	L	7.875	119.914	62.603	4.322	38.284	1.738		QQ $\delta$ 0.807
768	Q	7.802	119.419	55.532	4.437		1.918/1.813	2.191	

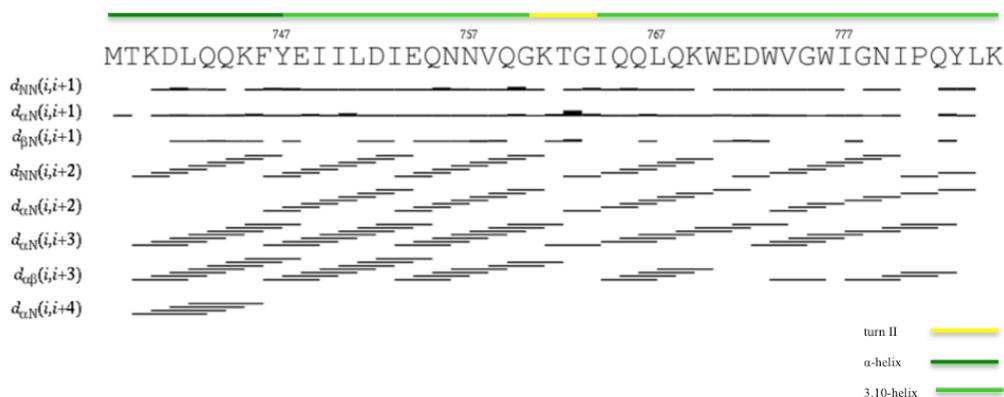
769	K	7.619	116.117	57.223	4.181	33.084	1.839	
770	W	8.025	115.524	57.216	4.612	29.990	3.399	He110.499
771	E	8.409	119.575	59.643	3.864	29.319	2.085	
772	D	8.149	118.465	56.752	4.345	40.326	2.673/2.241	
773	W	7.931	120.964	61.352	4.360	29.189	3.537/3.306	He110.408
774	V	8.280	118.496	61.411	3.626	26.617	2.131	1.052/0.938
775	G	8.097	108.790		4.231			
776	W	7.865	120.025	61.455	4.297		3.389	He110.516
777	I	8.487	118.534	61.471	3.655		2.050	1.484/0.98
778	G	8.072	104.570	46.598	3.859			
779	N	7.585	116.704	53.215	4.826	39.693	2.768/2.689	
780	I	7.794	121.075	60.065	3.783		2.101	1.663
781	P							Qδ1 0.904
782	Q	7.947	119.651	65.547	3.983		2.012	Qδ 3.848
783	Y	7.938	119.881	57.352	4.251		3.154	Qε 7.149
784	L	7.742	116.451	57.971	4.134		1.862	QQδ 0.872
785	K							

Once the backbone chemical shift assignment was achieved, identification of NOE effects was performed by analyzing the heteronuclear 3D  $^{13}\text{C}$ ,  $^{15}\text{N}$ -NOESY spectra. (Figure 9)



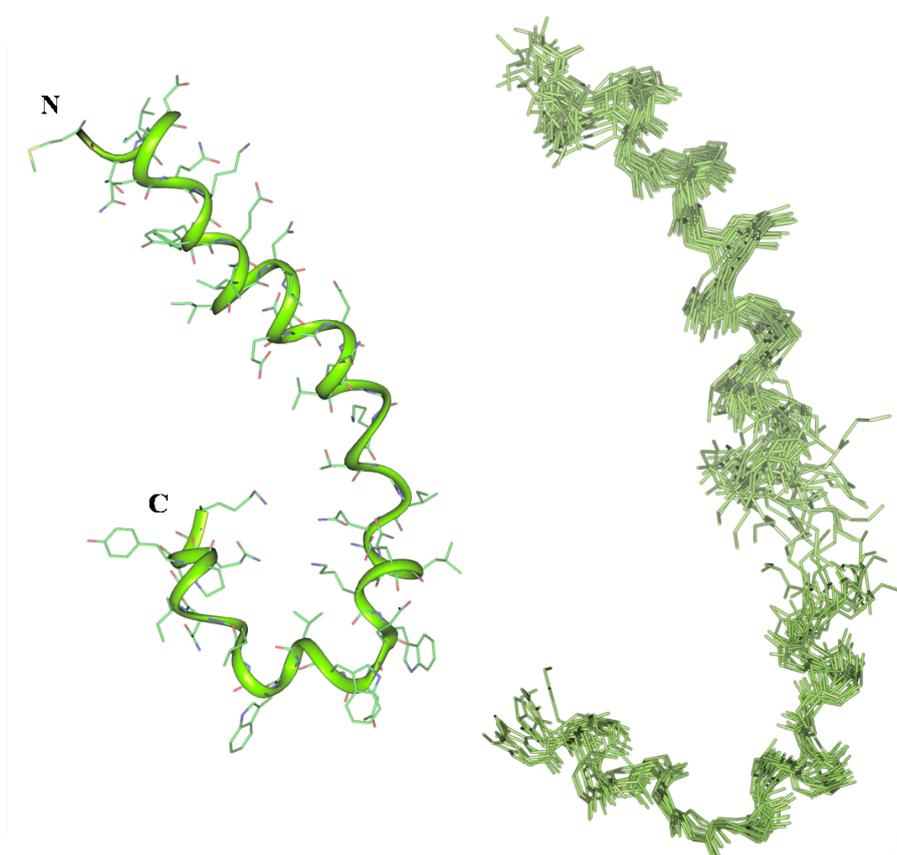
**Figure 9:** 3D-NMR  $^{15}\text{N}$ -NOESY of Gp36-MPER.

The quantitative evaluation of NMR data led to the calculation of Gp36-MPER NMR structure. NOE effects were translated in interprotonic distances using the CALIBA routine of CYANA 2.1 software.<sup>45</sup> Figure 10 shows the NOE connectivities of the protein as derived from 3D-NOESY spectra. Regular sequential medium range - N,N(i, i+2),  $\alpha$ ,N(i, i+2),  $\alpha$ ,N(i, i+3) and  $\alpha$ , $\beta$ (i, i+3) - NOE effects, are observable.



**Figure 10:** NOE connectivities of isotope labeled Gp36-MPER 3D-NOESY spectra in DPC/SDS mixed micelles.

Figure 11 shows the superposition of 20 Gp36-MPER NMR structures as derived from CYANA calculation. The structures show a good level of structural definition at level of the residues  $^{739}\text{Thr}$ - $^{760}\text{Gly}$  and  $^{761}\text{Lys}$ - $^{784}\text{Leu}$  where the superimposing on the backbone heavy atoms reveals 1.2Å and 1.6Å RMSD values respectively. The quantitative estimation of the backbone dihedral angles according to PROMOTIF procedure<sup>46</sup> using Kabsh and Sanders parameters points to the presence of the following secondary structure segments: i)  $\alpha$ -helix on  $^{739}\text{Thr}$ - $^{747}\text{Tyr}$  residues, ii) 3,10-helix on  $^{748}\text{Glu}$ - $^{760}\text{Gly}$  and  $^{765}\text{Gln}$ - $^{784}\text{Leu}$  residues, iii) type II  $\beta$ -turn on  $^{761}\text{Lys}$ - $^{764}\text{Ile}$  residues.



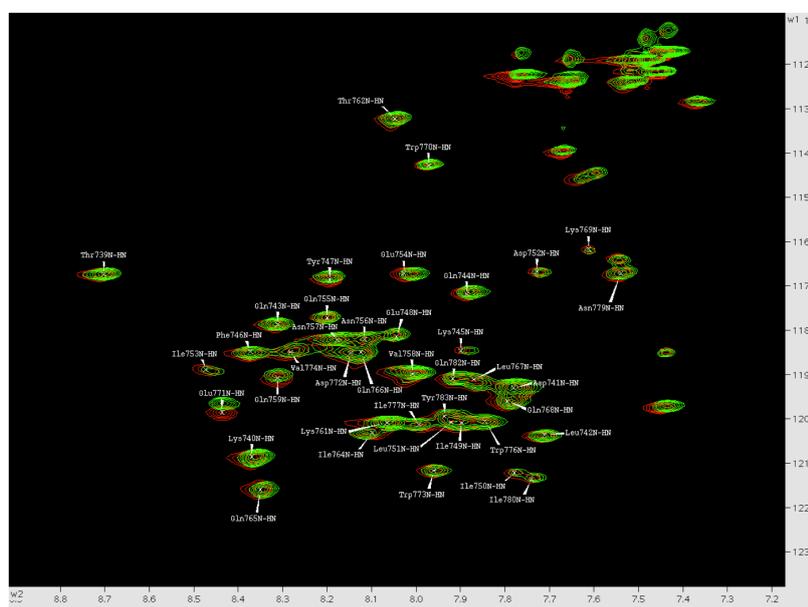
**Figure 11:** Gp36-MPER NMR structure.

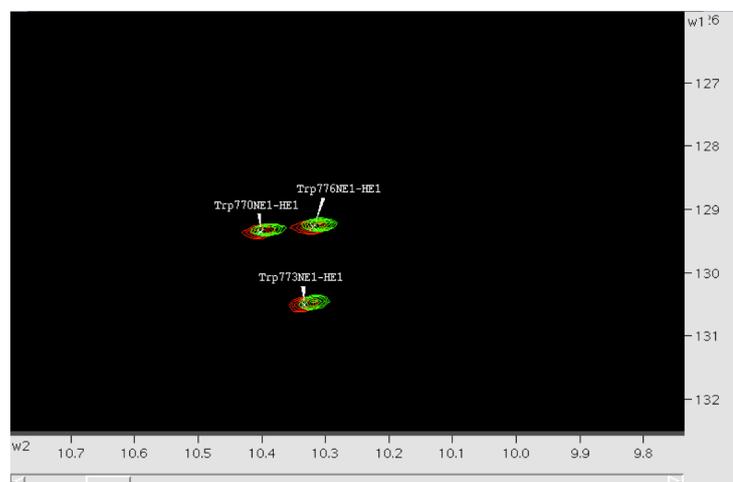
### 1.2.2 NMR interaction study: Chemical Shift Mapping

Perturbations in the chemical shift<sup>47,48</sup> can be used as a very sensitive probe for the identification of interaction surfaces in protein complexes. Usually a series of experiments is recorded in which increasing amounts of a ligand are added to the protein of interest, while chemical shifts changes are monitored. The chemical shifts subjected to perturbation belong to amino acids involved in interaction with the ligand under scrutiny. Typically, a H-<sup>15</sup>N HSQC

spectrum is used, since this is a highly sensitive experiment in which the peaks are generally well resolved.

As previously reported, peptide C8, belonging to the Gp36 of FIV,<sup>10, 19, 33</sup> elicited antiviral activity as a result of blocking cell entry, as observed for HIV fusion inhibitors.<sup>20, 49</sup> In the hypothesis that C8 antiviral activity may be related to its ability to interact with Gp36-MPER in analogy to the molecular mechanism of anti-HIV fusion inhibitors, we investigated the Gp36-MPER in interaction with C8 by following chemical shift mapping approach. We recorded H-<sup>15</sup>N HSQC spectra of Gp36-<sup>15</sup>N-MPER (1mM) containing increasing amounts of C8 (0.2-0.6 mM). Figure 12 shows the superposition of H-<sup>15</sup>N HSQC spectra relative to Gp36-MPER in absence and in presence of C8 (0.4 mM).





**Figure 12:** Gp36-MPER H-<sup>15</sup>N HSQC spectra in absence (red) and in presence (green) of C8 peptide.

The comparison of the spectra points to modifications in the HN backbone chemical shifts of <sup>744</sup>Gln, <sup>745</sup>Lys, <sup>753</sup>Ile, <sup>754</sup>Glu, <sup>759</sup>Gln, <sup>771</sup>Glu and HN indole of <sup>770</sup>Trp, <sup>773</sup>Trp and <sup>776</sup>Trp. These results indicate that C8 binds Gp36-MPER, confirming the hypothesis that its antiviral action may be related to the ability of interacting with MPER region of gp36 modifying the properties of virus cell entry.

### 1.3 Conclusions

C8 peptide is an octapeptide belonging to the gp36 of FIV. It elicited antiviral activity by blocking the penetration of the virus into the host cell. In the hypothesis that this activity was based on the ability of C8 to block the structural arrangement of gp36, essential for the virus cell entry, I studied Gp36-MPER/C8 complex by following a protein-observed NMR approach, specifically chemical shift mapping. Our results show that C8 binds Gp36-

MPER through the involvement of several residues. We demonstrated that the antiviral activity of C8 is possibly related to its ability of modifying the structural properties of Gp36-MPER.

## **Materials and Methods**

### *Expression of isotope labeled Gp36-MPER protein*

The expression vector pET-31b(+) was designed for cloning and high-level expression of Gp36-MPER sequences fused with the 125aa ketosteroid isomerase protein,<sup>44</sup> by Novagen. The vector contained the T7/Lac promoter and ampicillin resistance, was transformed into BLR21(DE3)-pLysS cells. Isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG) was purchased from Sigma-Aldrich (St. Louis, MO). For the expression in *E. coli* of uniformly <sup>15</sup>N and <sup>13</sup>C-isotope labeled Gp36-MPER protein, I prepared and autoclaved M9 minimal media containing <sup>15</sup>N ammonium sulfate and <sup>13</sup>C glucose from CIL (Cambridge Isotopes Laboratory, Andover, MA) and enriched with supplements such as trace metal mixtures and vitamin cocktails commercially available (Sigma-Aldrich (St. Louis, MO)). Cell growth was monitored spectrophotometrically by measuring OD<sub>600</sub> nm periodically.

### *NMR experiments: sample preparation and analysis*

3D-NMR experiments were acquired at Bijvoet Center for Biomolecular Research, (Utrecht University), thanks to a grant funded from Instruct-UB, within Bio-NMR project financed by the European Commission's Framework Programme 7. 3D-NMR experiments of Gp36-MPER were acquired at Bruker 900 MHz in dodecylphosphocoline/sodium dodecyl sulfate (DPC/SDS) mixed micelles. 0.5 mg of isotope labeled Gp36-MPER was dissolved in DPC/SDS

micelles solution (DPC concentration used was 27mM (27 times higher than DPC critical micellar concentration (c.m.c.)) and SDS concentration was 80mM (10 times higher than SDS c.m.c.)).<sup>50</sup> DPC/SDS molar ratio was of 90/10 (27mM/3mM) to produce the partial (2-3%) negative charge present in the typical membrane of eukaryotic cells.<sup>51</sup> The final pH was 7.4. For NMR experiments, <sup>d25</sup>SDS and <sup>d38</sup>DPC were used from Avanti® Polar Lipids, Inc.

NMR experiments were recorded at 300K. Standard backbone and side chain assignment experiments (CBCACONH, HNCACB, HNCO, HNCACO, HBHACONH and HCCH-TOCSY)<sup>52</sup> were recorded at a 600MHz Bruker Avance III machine equipped with TXI probe, running under Topspin 2.1. At this machine also DOSY, CLEANEX and T1, T2 and relaxation experiments were recorded. For T1 relaxation curves were fit from signal intensities in HSQCs recorded as pseud-3D experiment using relaxation delays of 10, 20, 40, 80, 160, 400, 1000, 2000, 4000, 6000 and 7000 ms, T2 times were obtained from fitting HSQC signal intensities after recording pseudo-3D experiment with 31.7 (2 times), 63.4, 95.1, 126.8 (2 times), 158.6, 190.3, 222.0, 253.7 and 285.4 ms. 2D NOESY, 3D NOESY-<sup>13</sup>C-HSQC and NOESY-<sup>15</sup>N-HSQC were recorded with 100 ms mixing times using a 900 MHz Bruker Avance III NMR system, equipped with a TCI cryoprobe and running under Topspin 3.0. Qualitative and quantitative analysis of 3D-NMR spectra were achieved using SPARKY software.<sup>53</sup>

#### *NMR structure calculation*

Peak volumes were translated into upper distance bounds with the CALIBA routine from the CYANA 2.1 software package.<sup>45</sup> After discarding redundant and duplicated constraints, the final list of constraints was used to generate an

ensemble of 100 structures by the standard CYANA protocol of simulated annealing in torsion angle space implemented (using 6000 steps). No dihedral angle restraints and no hydrogen bond restraints were applied.

The best 20 structures, which had low target function values (0.83-1.19) and small residual violations (maximum violation=0.38Å) were analyzed using the Insight II 98.0 program (Molecular Simulations, San Diego, CA, USA).

### **Acknowledgment**

Financial support in the form of Access to the Bio-NMR Research Infrastructure co-funded under the 7th Framework Programme of the EC (FP7/2007-2013, grant agreement 261863 for conducting the research at SONNMRLSF NMR facility in Utrecht (The Netherlands) is gratefully acknowledged. Financial support by the European Union (Bio-NMR, Project 261863) for access to the SONNMR Large-Scale Facility in Utrecht (The Netherlands) is gratefully acknowledged. We are very grateful for the assistance of Dr. Hans Wienk and Dr. Mario Scrima.

## References

1. Pedersen, N. C.; Ho, E. W.; Brown, M. L.; Yamamoto, J. K. Isolation of a t-lymphotropic virus from domestic cats with an immunodeficiency-like syndrome. *Science* **1987**, *235*, 790-793.
2. Elder, J. H.; Phillips, T. R. Feline immunodeficiency virus as a model for development of molecular approaches to intervention strategies against lentivirus infections. *Advances in Virus Research*, Vol 45 **1995**, *45*, 225-247.
3. Lee, T.; Laco, G. S.; Torbett, B. E.; Fox, H. S.; Lerner, D. L.; Elder, J. H.; Wong, C. H. Analysis of the S3 and S3' subsite specificities of feline immunodeficiency virus (FIV) protease: Development of a broad-based protease inhibitor efficacious against FIV, SIV and HIV in vitro and ex vivo. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 939-944.
4. Willett, B. J.; Flynn, J. N.; Hosie, M. J. FIV infection of the domestic cat: An animal model for AIDS. *Immunology Today* **1997**, *18*, 182-189.
5. Pancino, G.; Camoin, L.; Sonigo, P. Structural-analysis of the principal immunodominant domain of the feline immunodeficiency virus transmembrane glycoprotein. *Journal of Virology* **1995**, *69*, 2110-2118.
6. Serres, P. F. Molecular mimicry between the trimeric ectodomain of the transmembrane protein of immunosuppressive lentiviruses (HIV-SIV-FIV) and interleukin 2. *Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences* **2000**, *323*, 1019-1029.
7. Giannechini, S.; Bonci, F.; Pistello, M.; Matteucci, D.; Sichi, O.; Rovero, P.; Bendinelli, M. The membrane-proximal tryptophan-rich region in the transmembrane glycoprotein ectodomain of feline immunodeficiency virus is important for cell entry. *Virology* **2004**, *320*, 156-166.
8. Pierson, T. C.; Doms, R. W. HIV-1 entry and its inhibition. *Cellular Factors Involved in Early Steps of Retroviral Replication* **2003**, *281*, 1-27.
9. de Parseval, A.; Grant, C. K.; Sastry, K. J.; Elder, J. H. Sequential CD134-CXCR4 interactions in feline immunodeficiency virus (FIV): Soluble CD134 activates FIV Env for CXCR4-dependent entry and reveals a cryptic neutralization epitope. *Journal of Virology* **2006**, *80*, 3088-3091.
10. Frey, S. C. S.; Hoover, E. A.; Mullins, J. I. Feline immunodeficiency virus cell entry. *Journal of Virology* **2001**, *75*, 5433-5440.
11. Wyatt, R.; Sodroski, J. The HIV-1 envelope glycoproteins: Fusogens, antigens, and immunogens. *Science* **1998**, *280*, 1884-1888.
12. Harrison, S. C. Viral membrane fusion. *Nature Structural & Molecular Biology* **2008**, *15*, 690-698.
13. Lorizate, M.; Huarte, N.; Saez-Cirion, A.; Nieva, J. L. Interfacial pre-transmembrane domains in viral proteins promoting membrane fusion and fission. *Biochimica Et Biophysica Acta-Biomembranes* **2008**, *1778*, 1624-1639.
14. Dwyer, J. J.; Wilson, K. L.; Davison, D. K.; Freel, S. A.; Seedorff, J. E.; Wring, S. A.; Tvermoes, N. A.; Matthews, T. J.; Greenberg, M. L.; Delmedico, M. K. Design of helical, oligomeric HIV-1 fusion inhibitor peptides with potent activity against enfuvirtide-resistant

virus. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, 104, 12772-12777.

15. Cai, L.; Gochin, M.; Liu, K. Biochemistry and Biophysics of HIV-1 gp41-Membrane Interactions and Implications for HIV-1 Envelope Protein Mediated Viral-Cell Fusion and Fusion Inhibitor Design. *Current Topics in Medicinal Chemistry* **2011**, 11, 2959-2984.

16. Liu, J.; Deng, Y.; Li, Q.; Dey, A. K.; Moore, J. P.; Lu, M. Role of a Putative gp41 Dimerization Domain in Human Immunodeficiency Virus Type 1 Membrane Fusion. *Journal of Virology* **2010**, 84, 201-209.

17. Muench, J.; Staendker, L.; Adermann, K.; Schuz, A.; Schindler, M.; Chinnadurai, R.; Poehlmann, S.; Chaipan, C.; Biet, T.; Peters, T.; Meyer, B.; Wilhelm, D.; Lu, H.; Jing, W.; Jiang, S.; Forssmann, W.-G.; Kirchhoff, F. Discovery and optimization of a natural HIV-1 entry inhibitor targeting the gp41 fusion peptide. *Cell* **2007**, 129, 263-275.

18. Suarez, T.; Nir, S.; Goni, F. M.; Saez-Cirion, A.; Nieva, J. L. The pre-transmembrane region of the human immunodeficiency virus type-1 glycoprotein: a novel fusogenic sequence. *FEBS Letters* **2000**, 477, 145-149.

19. Suarez, T.; Gallaher, W. R.; Agirre, A.; Goni, F. M.; Nieva, J. L. Membrane interface-interacting sequences within the ectodomain of the human immunodeficiency virus type 1 envelope glycoprotein: Putative role during viral fusion. *Journal of Virology* **2000**, 74, 8038-8047.

20. Salzwedel, K.; West, J. T.; Hunter, E. A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity. *Journal of Virology* **1999**, 73, 2469-2480.

21. Schibli, D. J.; Montelaro, R. C.; Vogel, H. J. The membrane-proximal tryptophan-rich region of the HIV glycoprotein, gp41, forms a well-defined helix in dodecylphosphocholine micelles. *Biochemistry* **2001**, 40, 9570-9578.

22. Ji, H.; Shu, W.; Burling, F. T.; Jiang, S. B.; Lu, M. Inhibition of human immunodeficiency virus type 1 infectivity by the gp41 core: Role of a conserved hydrophobic cavity in membrane fusion. *Journal of Virology* **1999**, 73, 8578-8586.

23. Judice, J. K.; Tom, J. Y. K.; Huang, W.; Wrin, T.; Vennari, J.; Petropoulos, C. J.; McDowell, R. S. Inhibition of HIV type 1 infectivity by constrained alpha-helical peptides: Implications for the viral fusion mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, 94, 13426-13430.

24. Wild, C. T.; Shugars, D. C.; Greenwell, T. K.; McDanal, C. B.; Matthews, T. J. Peptides corresponding to a predictive alpha-helical domain of human-immunodeficiency-virus type-1 gp41 are potent inhibitors of virus-infection. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, 91, 9770-9774.

25. Neurath, A. R.; Strick, N.; Lin, K.; Jiang, S. Multifaceted consequences of anti-gp41 monoclonal-antibody 2f5 binding to hiv type-1 virions. *Aids Research and Human Retroviruses* **1995**, 11, 687-696.

26. Kilby, J. M.; Lalezari, J. P.; Eron, J. J.; Carlson, M.; Cohen, C.; Arduino, R. C.; Goodgame, J. C.; Gallant, J. E.; Volberding, P.; Murphy, R. L.; Valentine, F.; Saag, M. S.; Nelson, E. L.; Sista, P. R.; Dusek, A. The safety, plasma pharmacokinetics, and antiviral activity of subcutaneous enfuvirtide (T-20), a peptide inhibitor of gp41-mediated virus fusion, in HIV-infected adults. *Aids Research and Human Retroviruses* **2002**, 18, 685-693.

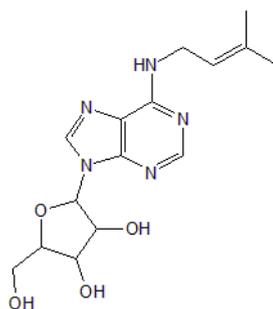
27. Champagne, K.; Shishido, A.; Root, M. J. Interactions of HIV-1 Inhibitory Peptide T20 with the gp41 N-HR Coiled Coil. *Journal of Biological Chemistry* **2009**, *284*, 3619-3627.
28. Walmsley, S.; Henry, K.; Katlama, C.; Nelson, M.; Castagna, A.; Reynes, J.; Clotet, B.; Hui, J.; Salgo, M.; DeMasi, R. Enfuvirtide (T-20) cross-reactive glycoprotein 41 antibody does not impair the efficacy or safety of enfuvirtide. *Journal of Infectious Diseases* **2003**, *188*, 1827-1833.
29. Hardy, H.; Skolnik, P. R. Enfuvirtide, a new fusion inhibitor for therapy of human immunodeficiency virus infection. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* **2004**, *24*, 198-211.
30. Kilby, J. M.; Hopkins, S.; Venetta, T. M.; DiMassimo, B.; Cloud, G. A.; Lee, J. Y.; Alldredge, L.; Hunter, E.; Lambert, D.; Bolognesi, D. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nature medicine* **1998**, *4*, 1302-1307.
31. Wexler-Cohen, Y.; Shai, Y. Demonstrating the C-terminal boundary of the HIV 1 fusion conformation in a dynamic ongoing fusion process and implication for fusion inhibition. *The FASEB Journal* **2007**, *21*, 3677-3684.
32. Lazzarin, A.; Clotet, B.; Cooper, D.; Reynes, J.; Arastéh, K.; Nelson, M.; Katlama, C.; Stellbrink, H.-J.; Delfraissy, J.-F.; Lange, J. Efficacy of enfuvirtide in patients infected with drug-resistant HIV-1 in Europe and Australia. *New England Journal of Medicine* **2003**, *348*, 2186-2195.
33. Barbato, G.; Bianchi, E.; Ingallinella, P.; Hurni, W. H.; Miller, M. D.; Ciliberto, G.; Cortese, R.; Bazzo, R.; Shiver, J. W.; Pessi, A. Structural analysis of the epitope of the Anti-HIV antibody 2F5 sheds light into its mechanism of neutralization and HIV fusion. *Journal of Molecular Biology* **2003**, *330*, 1101-1115.
34. Steffen, I.; Poehlmann, S. Peptide-Based Inhibitors of the HIV Envelope Protein and Other Class I Viral Fusion Proteins. *Current Pharmaceutical Design* **2010**, *16*, 1143-1158.
35. Scrima, M.; Grimaldi, M.; Di Marino, S.; Testa, C.; Rovero, P.; Papini, A. M.; Chorev, M.; D'Ursi, A. M. Solvent independent conformational propensities of 1,2,3 triazolyl-bridged parathyroid hormone-related peptide-derived cyclo-nonapeptide analogues. *Biopolymers* **2012**, *98*, 535-545.
36. Lombardi, S.; Massi, C.; Indino, E.; LaRosa, C.; Mazzetti, P.; Falcone, M. L.; Rovero, P.; Fissi, A.; Pieroni, O.; Bandecchi, P.; Esposito, F.; Tozzini, F.; Bendinelli, M.; Garzelli, C. Inhibition of feline immunodeficiency virus infection in vitro by envelope glycoprotein synthetic peptides. *Virology* **1996**, *220*, 274-284.
37. Williams, A. B. New horizons: antiretroviral therapy in 1997. *Journal of the Association of Nurses in AIDS Care* **1997**, *8*, 26-38.
38. Yeni, P. Update on HAART in HIV. *Journal of Hepatology* **2006**, *44*, S100-S103.
39. Aghokeng, A. F.; Ayouba, A.; Mpoudi-Ngole, E.; Loul, S.; Liegeois, F.; Delaporte, E.; Peeters, M. Extensive survey on the prevalence and genetic diversity of SIVs in primate bushmeat provides insights into risks for potential new cross-species transmissions. *Infection, Genetics and Evolution* **2010**, *10*, 386-396.
40. Esposito, C.; D'Errico, G.; Armenante, M. R.; Giannecchini, S.; Bendinelli, M.; Rovero, P.; D'Ursi, A. M. Physicochemical characterization of a peptide deriving from the glycoprotein gp36 of the feline immunodeficiency virus and its lipoylated analogue in micellar systems. *Biochimica et biophysica acta* **2006**, *1758*, 1653-61.

41. Giannecchini, S.; D'Ursi, A. M.; Esposito, C.; Scrima, M.; Zabogli, E.; Freer, G.; Rovero, P.; Bendinelli, M. Antibodies generated in cats by a lipopeptide reproducing the membrane-proximal external region of the feline immunodeficiency virus transmembrane enhance virus infectivity. *Clinical and vaccine immunology : CVI* **2007**, *14*, 944-51.
42. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular cloning*. Cold spring harbor laboratory press New York: 1989; Vol. 2.
43. Kuliopulos, A.; Walsh, C. T. PProduction, purification, and cleavage of tandem repeats of recombinant peptides. *Journal of the American Chemical Society* **1994**, *116*, 4599-4607.
44. Rodriguez, J. C.; Wong, L.; Jennings, P. A. The solvent in CNBr cleavage reactions determines the fragmentation efficiency of ketosteroid isomerase fusion proteins used in the production of recombinant peptides. *Protein expression and purification* **2003**, *28*, 224-31.
45. Guntert, P. Automated NMR structure calculation with CYANA. *Methods in molecular biology (Clifton, N.J.)* **2004**, *278*, 353-78.
46. Hutchinson, E. G.; Thornton, J. M. PROMOTIF - A program to identify and analyze structural motifs in proteins. *Protein Science* **1996**, *5*, 212-220.
47. Medek, A.; Hajduk, P. J.; Mack, J.; Fesik, S. W. The use of differential chemical shifts for determining the binding site location and orientation of protein-bound ligands. *Journal of the American Chemical Society* **2000**, *122*, 1241-1242.
48. Krishnamoorthy, J.; Yu, V. C. K.; Mok, Y.-K. Auto-FACE: An NMR Based Binding Site Mapping Program for Fast Chemical Exchange Protein-Ligand Systems. *Plos One* **2010**, *5*.
49. Bagby, S.; Tong, K. I.; Liu, D. J.; Alattia, J. R.; Ikura, M. The button test: A small scale method using microdialysis cells for assessing protein solubility at concentrations suitable for NMR. *Journal of Biomolecular Nmr* **1997**, *10*, 279-282.
50. Pellegrini, M.; Mierke, D. F. Structural characterization of peptide hormone/receptor interactions by NMR spectroscopy. *Biopolymers* **1999**, *51*, 208-220.
51. Weller, K.; Lauber, S.; Lerch, M.; Renaud, A.; Merkle, H. P.; Zerbe, O. Biophysical and biological studies of end-group-modified derivatives of Pep-1. *Biochemistry* **2005**, *44*, 15799-15811.
52. Cavanagh, J.; Fairbrother, W. J.; Palmer Iii, A. G.; Skelton, N. J. *Protein NMR spectroscopy: principles and practice*. Academic Press: 1995.
53. Kneller, D. G.; Kuntz, I. D. UCSF SPARKY - an nmr display, annotation and assignment tool. *Journal of Cellular Biochemistry* **1993**, 254-254.

## 2 *New promising ligands for FPPS antitumoral target*

### 2.1 Introduction

N6-Isopentenyladenosine (i6A) is a modified nucleoside (Figure 1), formed by an adenosine harboring an isopentenyl chain derived from dimethylallyl pyrophosphate in the N6 position. It belongs to the cytokinin family, involved in control of many processes in plants. Interestingly, it represents the unique cytokinins found also in mammals, bound to tRNA (tRNA) or as free nucleoside.<sup>1-3</sup>

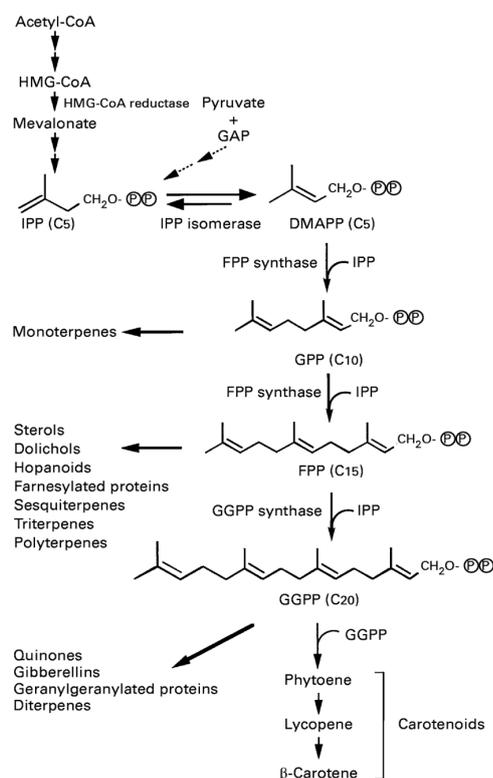


**Figure 1:** N6-isopentenyladenosine (i6A) structure.

In humans, many biological actions, both *in vitro* and *in vivo*, including antitumoral effects, can be attributed to i6A. Although its precise mechanism of action has not been fully clarified, recently i6A was proved able to modulate the activity of farnesyl pyrophosphate synthase (FPPS), a key enzyme involved in the mevalonate pathway and in downstream proteins prenylation that appears deregulated in many tumors.<sup>4</sup> Indeed, i6A exerts anti-proliferative effects in thyroid K-ras (KiMol) transformed cells and

untransformed, FRTL-5 wild-type cells. These effects were due to the inhibition of FPPS, both in the expression and in the activity. FPPS inhibition was also correlated to the inhibition of tumor cell proliferation and protein prenylation. The anti-proliferative action of i6A was corroborated also in an in vivo system, because the growth of murine xenograft (where tumoral KiMol cells were implanted subcutaneously) that resulted was inhibited by its treatment.<sup>5</sup> More recently, another evidence on the ability of i6A to modulate FPPS expression and activity has been reported.<sup>6</sup> On natural killer cells, i6A at lowest concentrations (sub 1 $\mu$ M) was able to directly stimulate the proliferation and the cytotoxic activity *vs* tumor cells by the induction of the expression and the activity of FPPS.

The enzyme FPPS is one of the key enzymes involved in the mevalonate, isoprenoid biosynthesis pathway, it catalyzes the two step synthesis of the C15 isoprenoid farnesyl pyrophosphate (FPP): isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are coupled to produce geranyl pyrophosphate (GPP), which is then condensed with an additional IPP to produce farnesyl pyrophosphate (FPP). (Figure 2)



**Figure 2:** Mevalonate pathway.

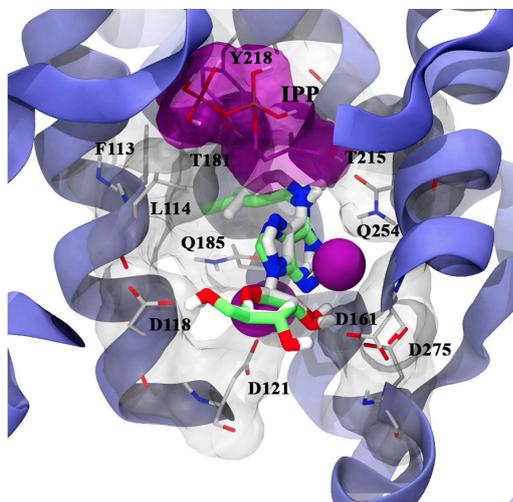
FPP is a crucial precursor in the synthesis of several classes of essential metabolites, such as sterols, ubiquinones and carotenoids. It is also involved in the protein prenylation, visual pigments, constituents of membranes and components of signal transduction.<sup>7, 8</sup>

FPPS is identified as the main biochemical target of bisphosphonates for the treatment of bone-related disorders. The major groups of FPPS inhibitors are nitrogen-containing bisphosphonates (N-BP) used in clinical treatment of osteoporosis diseases, Paget's disease, and more recently, metastatic bone-related tumors and cancer. In view of the implication of FPPS in cancer related

pathways, it is considered an interesting target, prompting the search of new specific anticancer compounds. On the other hand, the employment of bisphosphonates FPPS inhibitor in different tumor or infective diseases is limited by their adverse pharmacokinetic properties.

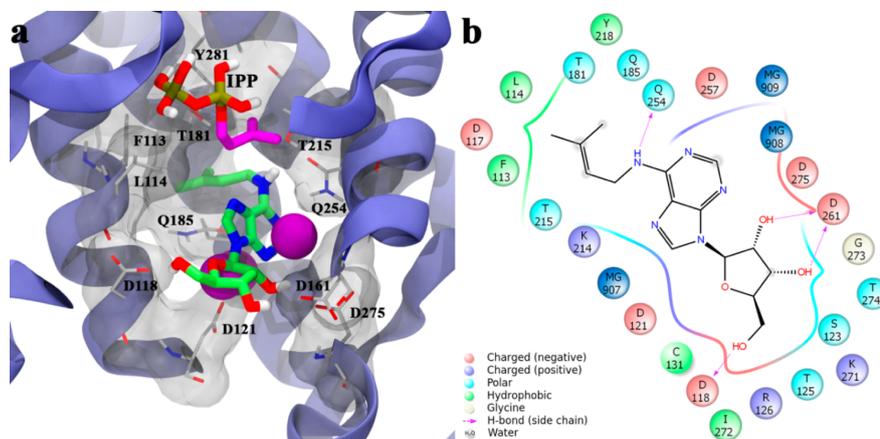
In the hypothesis that i6A may be able to modulate the activity of FPPS, in cancer related pathways, through a direct structural interaction, in silico inverse virtual screening was carried out using i6A on a panel of 296 3D protein structures involved in cancer processes.<sup>9-11</sup> Notably, in this screening, FPPS protein was identified in the first two positions among the 296 proteins investigated. Therefore two models of this target were built, starting from the same crystallized structure (PDB code: 1ZW5)<sup>12</sup> but differing in the absence or the presence of co-crystallized ligand isopentenyl pyrophosphate (IPP).

Molecular docking demonstrated that i6A binds FPPS even in the presence of IPP. In presence and in absence of IPP the binding modes characteristics are conserved, (Figure 3) indicating that the orientation of i6A is not influenced by the absence/presence of co-crystallized FPPS ligand IPP. Overall, these data strengthened the hypothesis of i6A interaction in the canonical binding site of FPPS.<sup>13</sup>



**Figure 3:** Superposition between i6A docking pose in the FPPS binding site in presence of IPP and in absence of IPP.

Analysis of i6A binding poses in the binding pocket of FPPS, revealed a favorable accommodation of i6A through a large set of both hydrophobic and polar interactions (Figure 4 a, b). Sugar moiety of i6A establishes a network of H-bonds with <sup>118</sup>Asp and <sup>261</sup>Asp and polar interactions with <sup>121</sup>Asp, <sup>123</sup>Ser, <sup>125</sup>Thr, <sup>126</sup>Arg, <sup>274</sup>Thr, and <sup>271</sup>Lys. Adenine core of the molecule is located between the three Mg<sup>++</sup> ions co-crystallized with FPPS, undertaking polar contacts with <sup>214</sup>Lys, <sup>257</sup>Asp, and H-bonding with <sup>254</sup>Gln. Isopentenyl part of the molecule is placed in the deeper region of the binding site, and hydrophobic interactions are observable with <sup>113</sup>Phe, <sup>114</sup>Leu, and <sup>218</sup>Tyr.



**Figure 4:** (a) Docking model of i6A in the FPPS binding site in the presence of IPP, (b) 2D interactions panel showing interactions between i6A and residues in the FPPS binding site.

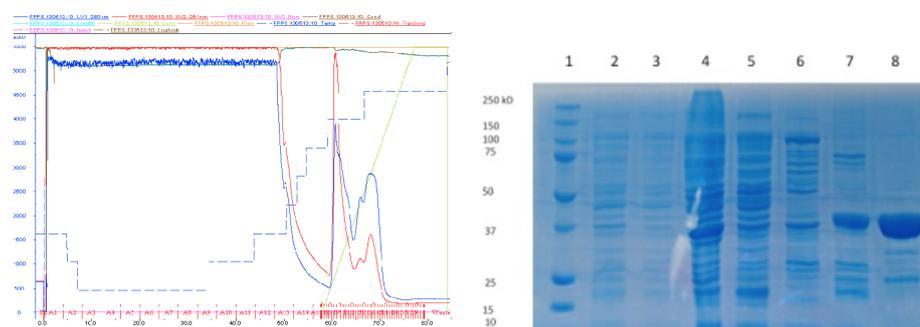
To provide experimental data in support of the inverse virtual screening procedure, I performed NMR investigation of i6A-FPPS protein interaction, by recording saturation transfer difference (STD) and WaterLOGSY experiments. Subsequently, based on the i6A-FPPS interaction data, I analyzed the structural interaction of new analogs of i6A with FPPS by recording STD-NMR experiments. Finally, I evaluated the enzymatic activity of FPPS using a known colorimetric<sup>14</sup> assay, and a NMR based enzymatic assay, newly developed in our laboratory.

## 2.2 Results and Discussion

### 2.2.1 FPPS gene expression

The plasmid p11, transformed into BL21(DE3)-pLysS cells, were used for the expression in *E. coli* as described in Materials and Methods. After 6h of cell growth, cells were pelleted by centrifugation and re-suspended in lysis buffer (50 mL of 5% glycerol, 5 mM imidazole, 500 mM NaCl, 50 mM PBS

(pH 7.5)) and sonicated. Protein was purified with His-Trap HP column at 1 mL/min using an AKTA purifier system, the soluble extract was applied to a nickel-chelated agarose affinity column that had been equilibrated with the same buffer. The protein was eluted from the column with elution buffer (5% glycerol, 250 mM imidazole, 500 mM NaCl, 50 mM PBS (pH 7.5)). Affinity chromatography on a nickel chelated agarose column permitted a simple one-step protein purification. Enzyme purity was judged by using SDS-polyacrylamide gel electrophoresis with Coomassie Blue staining (Figure 5). The eluted fraction (Figure 5, fraction 8, the molecular mass of FPPS-His-tag is 43kDa) was transferred into Vivaspin 20 concentrator, cutoff 3 kDa, to exchange the buffer for NMR studies.



**Figure 5:** left) FPPS chromatogram obtained by FPLC. right) SDS-PAGE: 1 Marker (Dual Color Biorad), 2-3 Pellet after lysis, 4 Soluble extract after lysis, 5 eluted fraction that doesn't bind column, 6 first peak eluted at 60min, 7 peaks eluted at 65min and 8 peak eluted at 70min.

2.2.2 NMR Experiments: *i6A-FPPS interaction*

NMR analysis of FPPS-*i6A* interaction was based on saturation transfer difference (STD) and water-ligand observed via gradient spectroscopy (WaterLOGSY) NMR experiments.<sup>16, 17</sup> STD and WaterLOGSY are powerful NMR techniques that enable the identification of protein-ligand binding sites and the determinations of protein-ligand dissociation constants ( $K_D$ ). NMR sample containing 8  $\mu$ M of FPPS was titrated with *i6A* to have STD build-up at protein-ligand molar ratios: 1:10, 1:20, 1:30, 1:50, 1:70, and 1:100. For each titration point, STD experiments were carried out using different saturation times (0.50, 1.00, 1.50, 2.00, 3.00, 4.00, and 5.00 s).<sup>18</sup> Standard  $^1\text{H}$  mono-dimensional and 2D COSY NMR experiments allowed the  $^1\text{H}$  chemical shift assignment of *i6A* proton signals. (Figure 6)

Figures 7-8 show STD and WaterLOGSY NMR spectra, respectively, recorded at 1:100 FPPS-*i6A* molar ratio (2s saturation time).

<b><i>i6A</i> Proton</b>	<b><math>^1\text{H}</math> chemical shift</b>
<b>H2</b>	8.22
<b>H8</b>	8.17
<b>H16</b>	5.99
<b>H12</b>	5.35
<b>H17</b>	4.33
<b>H20-21</b>	4.20
<b>H11</b>	4.07
<b>H18</b>	3.83
<b>H19</b>	3.74
<b>H14-15</b>	1.68

**Figure 6:** *i6A*  $^1\text{H}$  chemical shift.

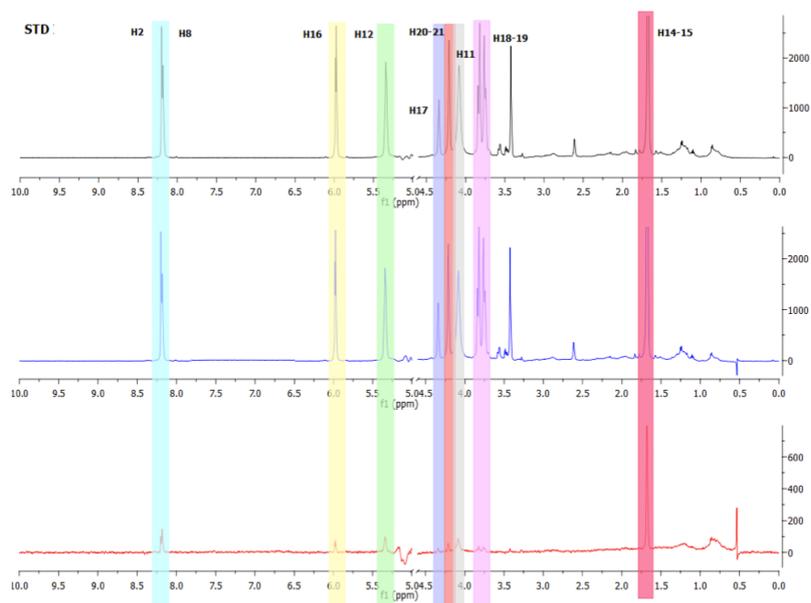


Figure 7: STD at 1:100 FPPS-i6A molar ratio.

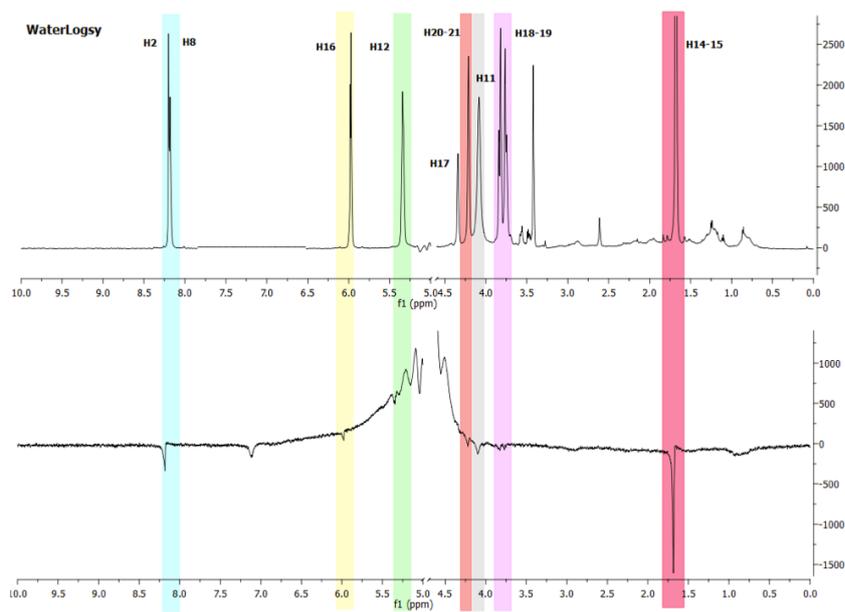


Figure 8: WaterLOGSY at 1:100 FPPS-i6A molar ratio.

Both STD and WaterLOGSY NMR spectra, shown in Figures 7-8, evidence significant variations in intensity for the H11, H12, H14, and H15 signals belonging to the i6A isopentenyl moiety. H2 and H8 protons of the adenine ring are moderately perturbed, and weak STD effects are also observable on the protons belonging to the ribose sugar. The data extracted from the STD experiments indicate that the isopentenyl moiety and the purine ring of i6A are directly involved in the interaction with FPPS binding pocket, while the ribose portion may undertake interactions that are less evident in the current experimental conditions.

The quantitative estimation of STD effects in experiments collected at different protein-ligand ratios and saturation time conditions allowed for the calculation of i6A-FPPS binding constant  $K_D$ , according to the methodology recently developed by Angulo et al.<sup>19</sup> Using this procedure, the calculation of protein-ligand affinities is independent from contingent experimental factors, such as STD saturation time, ligand residence time in the complex, and the intensity of the signal.<sup>19</sup> Table I reports  $K_D$  for the single i6A protons involved in the binding with FPPS. In agreement with the previous qualitative evaluation, protons H11, H12, and H14-H15 of the isopentenyl moiety and the proton H2 of purine ring show the lowest values of  $K_D$ . The mean  $K_D$  value calculated for i6A-FPPS interaction is in the mM range.

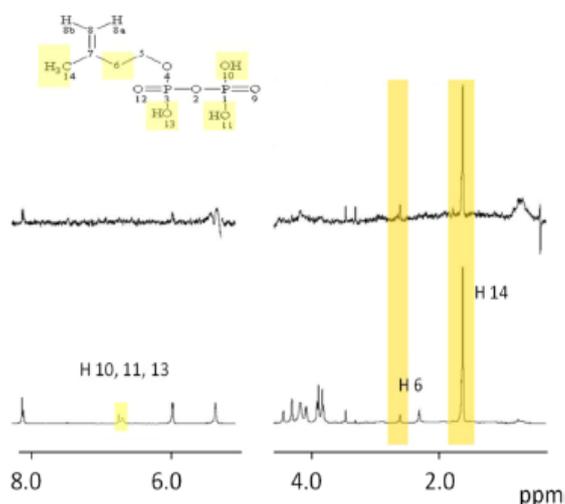
**Table I:** Dissociation Constants of the i6A-FPPS complex.

Protein-ligand sistem	$K_D$ [mM]
proton H2/H8	$2.02 \pm 0.31$
proton H12	$2.21 \pm 0.32$
proton H11	$0.55 \pm 0.06$
proton H14-H15	$1.21 \pm 0.12$

Moreover to verify the molecular docking prediction, suggesting that the binding mode of i6A is not influenced by the absence/presence of co-crystallized FPPS ligand IPP, we recorded STD NMR experiments by titrating 1:100 FPPS-i6A sample with increasing amounts of IPP to have FPPS-IPP molar ratios: 1:10, 1:20, 1:30, 1:50, 1:70, 1:100. Figure 9 shows  $^1\text{H}$  NMR spectrum of FPPS in the presence of both i6A and IPP (FPPS-i6A and FPPS-IPP 1:100 molar ratio, respectively).<sup>20</sup>

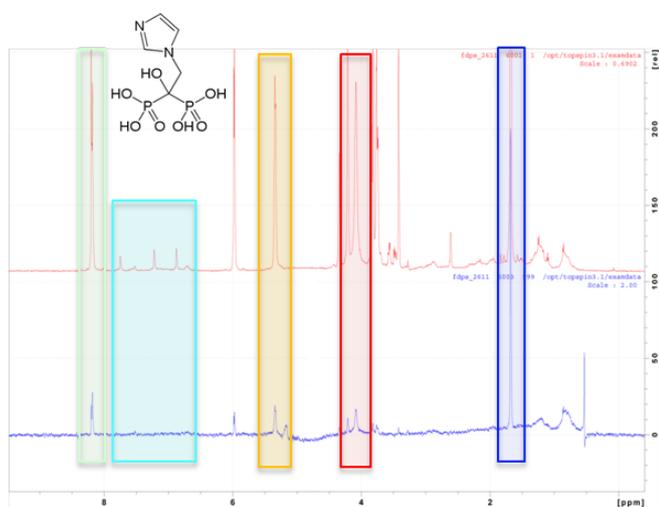
Unfortunately, H5 and H8 IPP protons were not observable in the reported experimental conditions due to the water suppression. The spectrum shown in Figure 9 confirms the participation of the previously mentioned (H11, H12, H14, and H15) protons of i6A to the binding with FPPS, moreover, it evidences STD effects for H14 and H6 protons of IPP.

The qualitative evaluation of these STD effects at different FPPS-IPP ratios indicates a different modulation of STD effects relative to IPP with respect to i6A. These data prove that both IPP and i6A bind FPPS; their binding is characterized by different kinetic and no evidence of competition is observable between the two ligands.

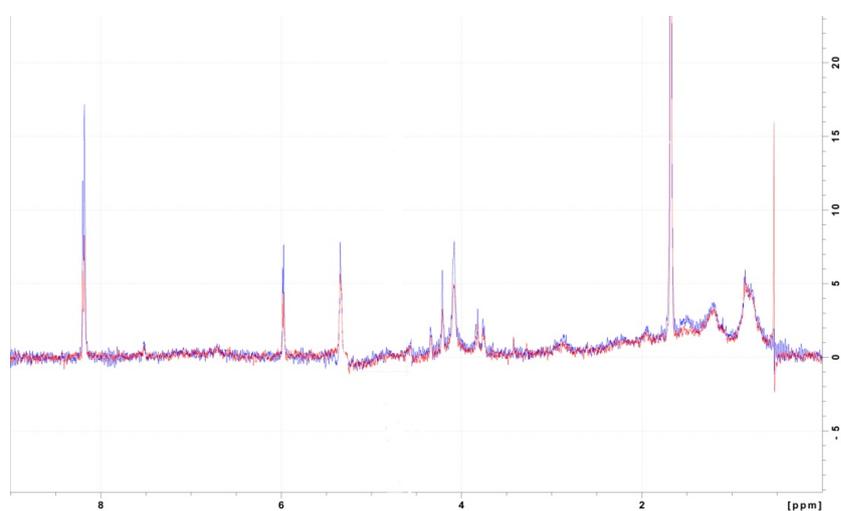


**Figure 9:** STD NMR spectra of FPPS containing both i6A and IPP (1:100 FPPS-i6A; 1:100 FPPS-IPP). On top, the STD spectrum; on bottom, the off-resonance spectrum.

To discriminate the binding of detected interaction of i6A with FPPS, STD-NMR competition experiments were recorded against zoledronic acid used as FPPS modulator. Figure 10 shows STD-NMR spectra of FPPS-i6A, 1:100 molar ratio, at saturation time of 2s, to which zoledronic acid (80 $\mu$ M) was added. Inspection of the Figure indicates the absence of STD effects for zoledronic acid coherently with its  $K_D$  value 3-10 nM, that is out of the range allowed for the observation of STD effect. Figure 10 shows that the addition of minimal quantities of zoledronic acid results in the decrease of the STD effect on i6A protons. In particular STD effects, relative to H2, H4, H7 and H9 protons, show respectively 3% decrement with respect to the STD effects observable in the spectrum recorded without zoledronic acid. The decrease in STD effect is a strong evidence of the competition of i6A and zoledronic acid, for the same binding site. (Figure 11)



**Figure 10:** NMR spectra of FPPS-i6A, 1:100 molar ratio, at saturation time of 2s, to which was added zoledronic acid. In red the off-resonance spectrum, in blue the STD spectrum.

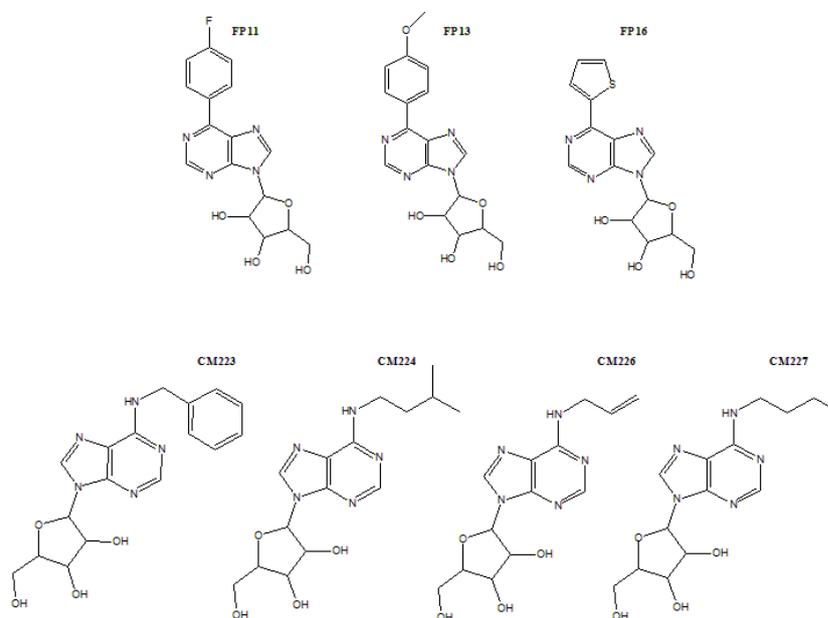


**Figura 11:** overlapping of STD spectrum of FPPS-i6A, 1:100 molar ratio, at saturation time of 2s without zoledronic acid in blue, and spectrum of FPPS-i6A, 1:100 molar ratio, at saturation time of 2s with zoledronic acid.

In conclusion all the reported evidence show that NMR data fit with the prediction of the inverse virtual screening, indicating a direct binding of i6A with FPPS; interestingly, the intensity of the STD effects are coherent with the binding pose hypothesized in the docking calculation. Indeed, the protons showing the lowest values of  $K_D$  (Table I) occupy the deepest region of the binding pocket, whereas the protons of the sugar moiety evidencing weak STD effects are those more exposed to the solvent and possibly, more blandly involved in the interaction with the binding site.<sup>20</sup>

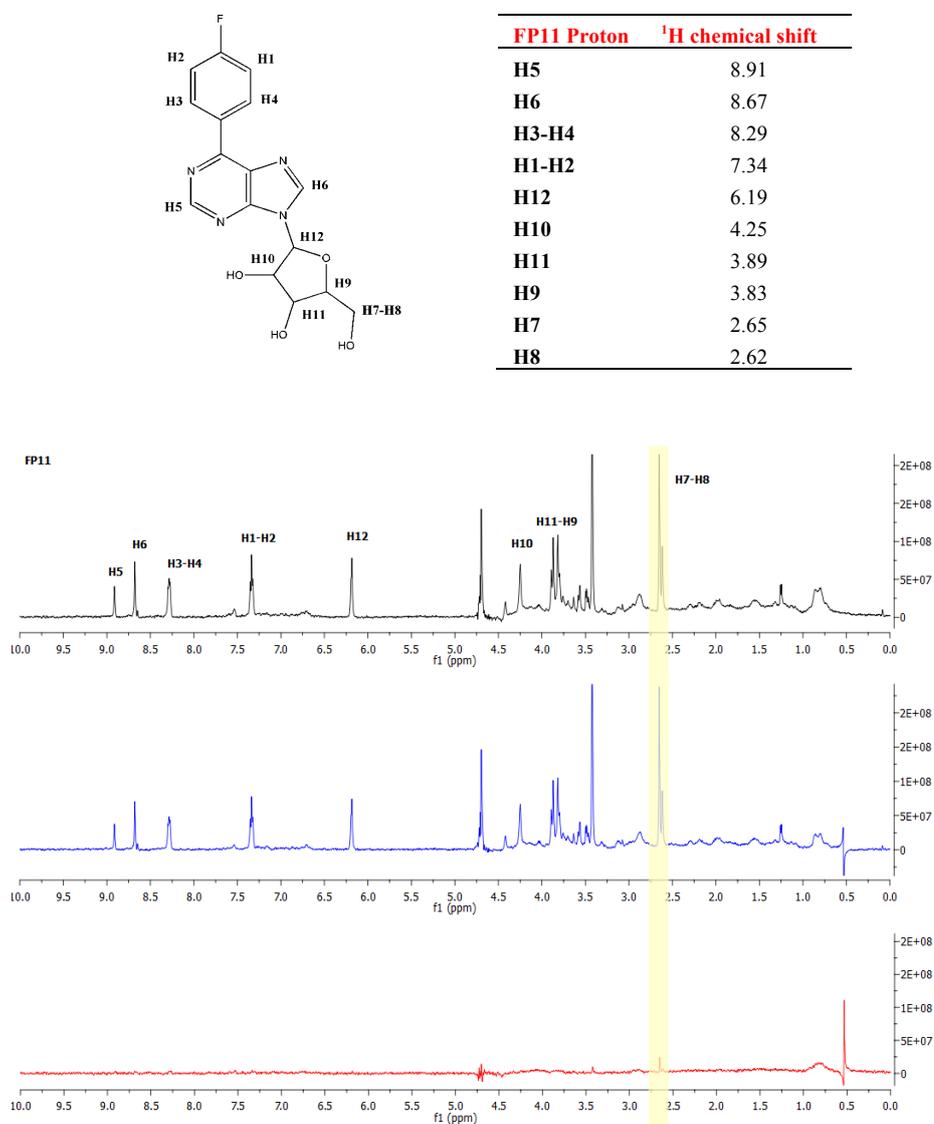
### 2.2.3 NMR Experiments: i6A analogs-FPPS interaction

Computational and NMR experimental evidence show that i6A binds FPPS with an important participation of the isopentenyl moiety. Based on these data, i6A analogs including modification of the isopentenyl moiety were synthesized (Figure 12): double bond elimination (CM224), insertion of a phenyl ring characterized by increasing steric hindrance at 6 position of adenosine (FP11, FP13, FP16), insertion, in the same position, of a benzyl ring (CM223).

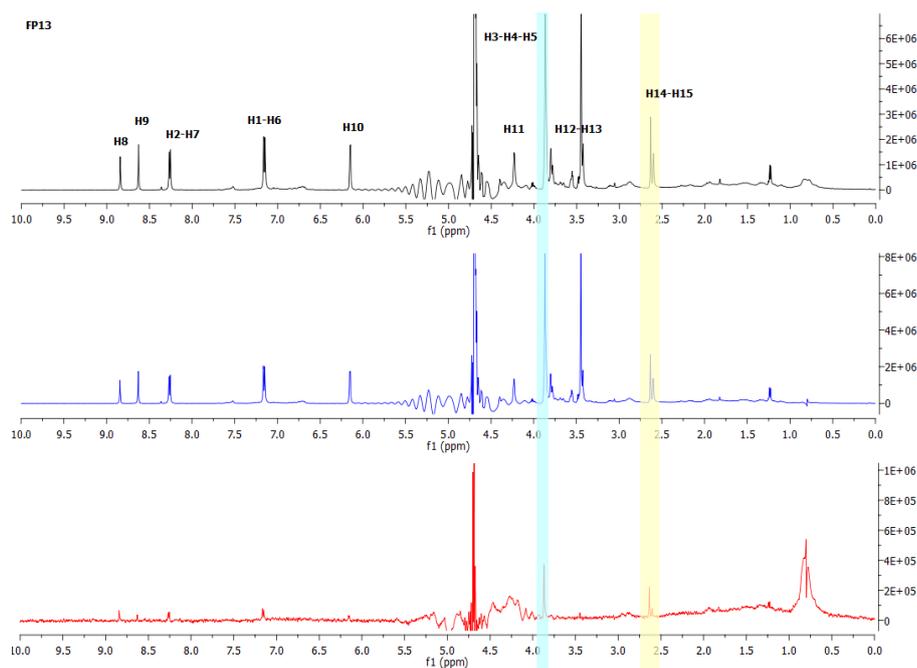
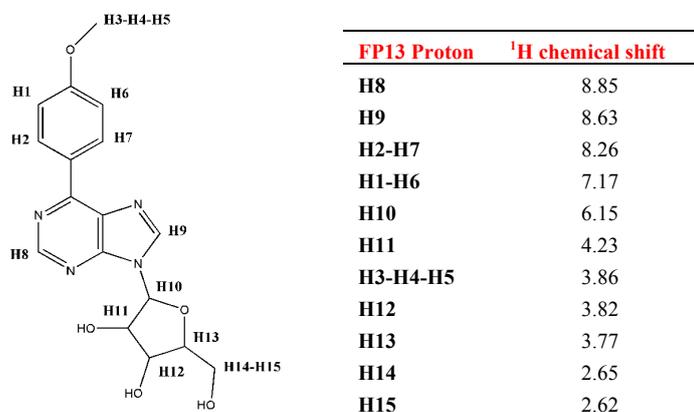


**Figure 12:** i6A analogs.

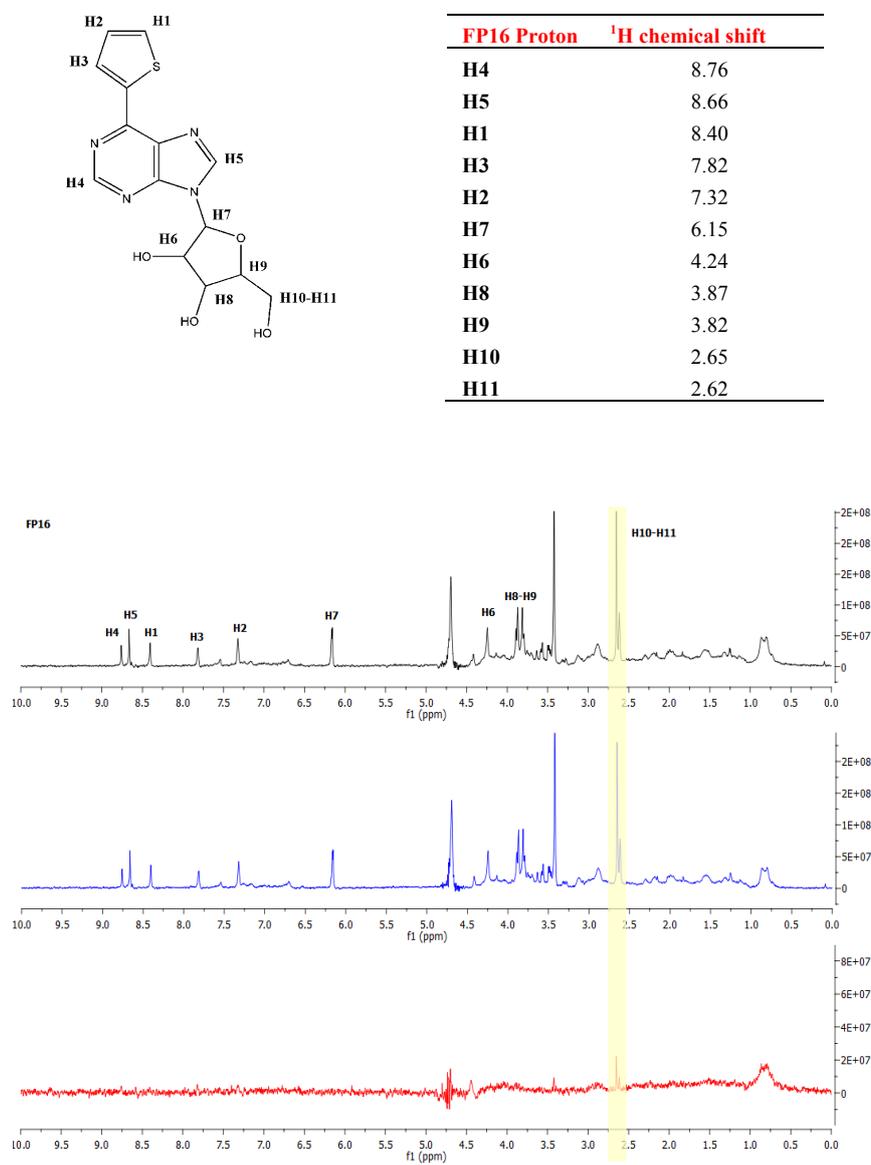
Structural interaction of i6A analogs with FPPS protein was evaluated by recording STD-NMR experiments (Figures 13). STD experiments show that, for the compounds of FP series, containing the aromatic ring directly linked to the 6-position of adenosine, the most consistent STD effect (10%) involve the protons on the C5 of ribose sugar (FP11: H7-H8; FP13: H14-H15; FP16: H10-H11).



**Figure 13 a:** STD-NMR spectra of FPPS-FP11, in black the off-resonance spectra, in blue the on-resonance spectra and in red the STD spectra.

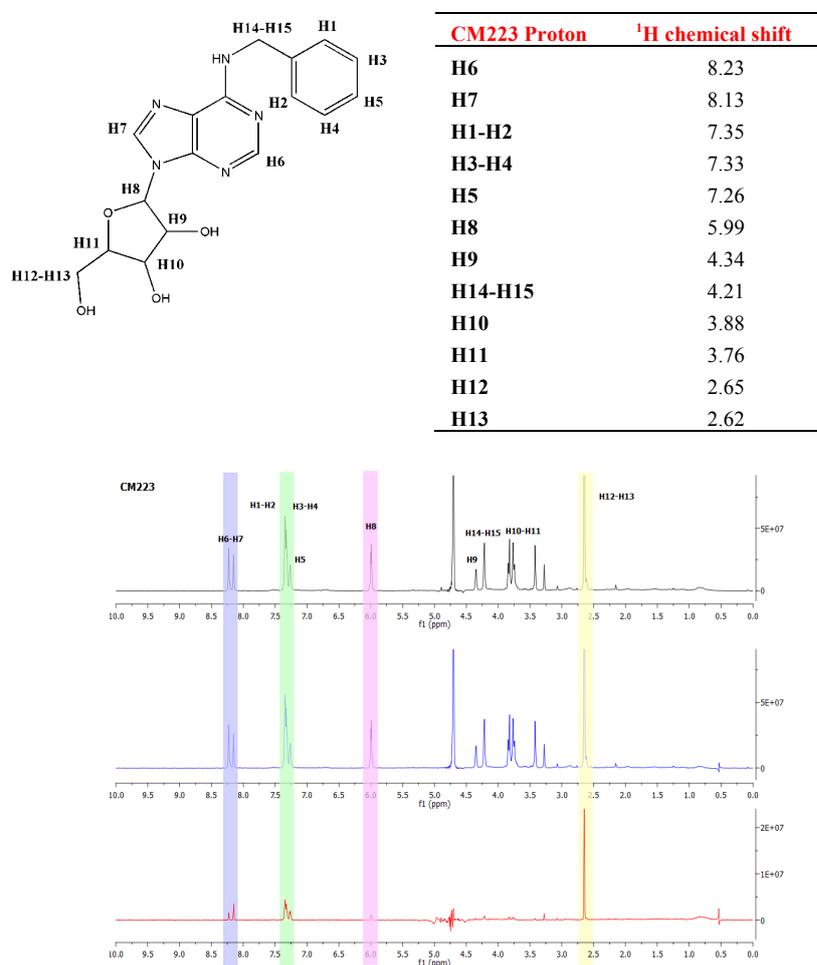


**Figure 13 b:** STD-NMR spectra of FPPS-FP13, in black the off-resonance spectra, in blue the on-resonance spectra and in red the STD spectra.

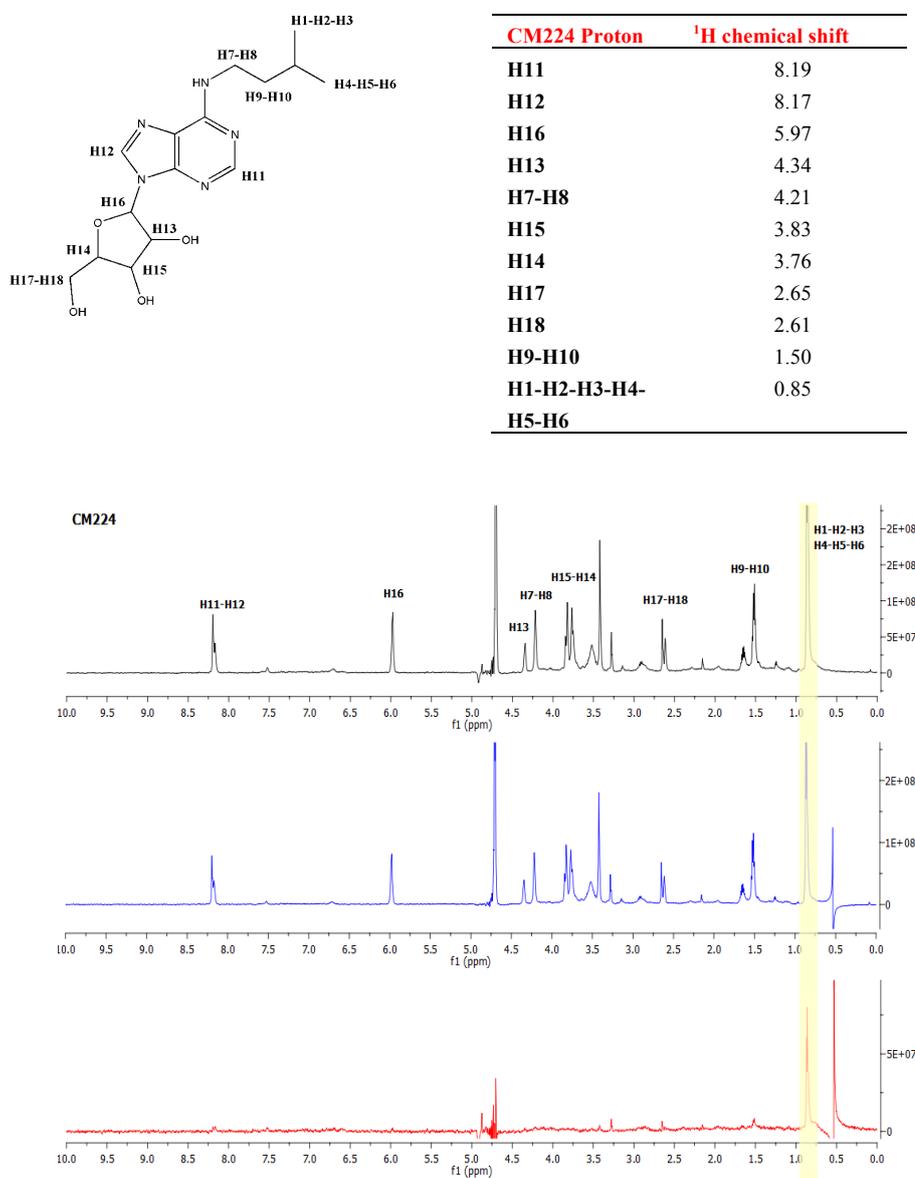


**Figure 13 c:** STD-NMR spectra of FPPS-FP16, in black the off-resonance spectra, in blue the on-resonance spectra and in red the STD spectra.

In the compounds of CM series, characterized by the insertion of the phenyl ring (CM223) and isopentyl moiety (CM224), the following STD effects were observed: CM223: H1-H2-H3-H4-H5; CM224: H1-H2-H3-H4-H5-H6. These effects reached the significant value of 30% (CM223) and 20% (CM224) respectively. (Figure 14)



**Figure 14 a:** STD-NMR spectra of FPPS-CM223, in black the off-resonance spectra, in blue the on-resonance spectra and in red the STD spectra.



**Figure 14 b:** STD-NMR spectra of FPPS-CM224, in black the off-resonance spectra, in blue the on-resonance spectra and in red the STD spectra.

The data extracted from the STD experiments on the i6A-analogs allow us to hypothesize that, for the compounds of FP series, the portion most involved in the interaction with FPPS binding pocket is the ribose sugar; conversely, for the compounds of CM series, the protons in the N6 adenosine position are those most significantly involved in the interaction with FPPS binding site.

As CM223 showed the most significant STD effect (30%), we calculated the  $K_D$  for CM223-FPPS interaction. The quantitative estimation of STD effects in experiments collected at different protein-ligand ratios and saturation time conditions allowed for the calculation of i6A-FPPS binding constant  $K_D$ , according to the methodology recently developed by Angulo et al.<sup>18,19</sup> Using this procedure, the calculation of protein-ligand affinities is not affected by contingent experimental factors, such as STD saturation time, ligand residence time in the complex, and intensity of the signal.

NMR sample containing 8  $\mu$ M of FPPS was titrated with CM223 to have STD build-up at protein-ligand molar ratios: 1:10, 1:20, 1:30, 1:50, 1:70, and 1:100. For each titration point, STD experiments were carried out using different saturation times (0.50, 1.00, 1.50, 2.00, 3.00, 4.00, and 5.00 s). The mean  $K_D$ , value calculated for the single protons of CM223 (Table II) is 0.25mM.

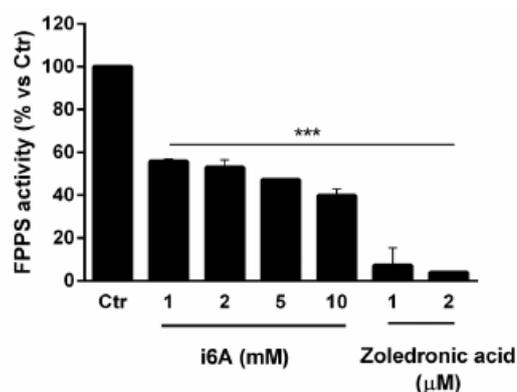
**Table II:** Dissociation constants of the FPPS-CM223 complex.

Protein-ligand sistem	$K_D$ [mM]
proton H6	$0.20 \pm 0.01$
proton H7	$0.20 \pm 0.01$
proton H1-H2-H3-H4	$0.20 \pm 0.01$
proton H5	$0.16 \pm 0.11$
proton H8	$0.19 \pm 0.02$
proton H12-H13	$0.78 \pm 0.08$

This value, that is 4 fold lower as i6A-FPPS  $K_D$  value, indicates an improvement in the binding with FPPS target for CM223 and opens the perspective that a modification of the newly introduced benzyl portion of the molecule may lead to more active FPPS inhibitors.

#### 2.2.4 Enzymatic Assay of FPPS by NMR

Once we established the i6A-FPPS structural interaction, we measured i6A enzymatic activity following the colorimetric method developed by Gao J et al.,<sup>14</sup> and based on the recognition of pyrophosphate. Figure 15 shows FPPS enzymatic activity in the presence of i6A (1 to 10 mM concentration range). FPPS activity in the presence of high affinity FPPS inhibitor, zoledronic acid (1-2 $\mu$ M), 30min 37°C), is considered as control. Inspection of the curves indicates the ability of i6A (1-10mM) to inhibit FPPS activity. Therefore, the interaction of i6A with FPPS is effective to moderately inhibit the FPPS enzymatic activity.



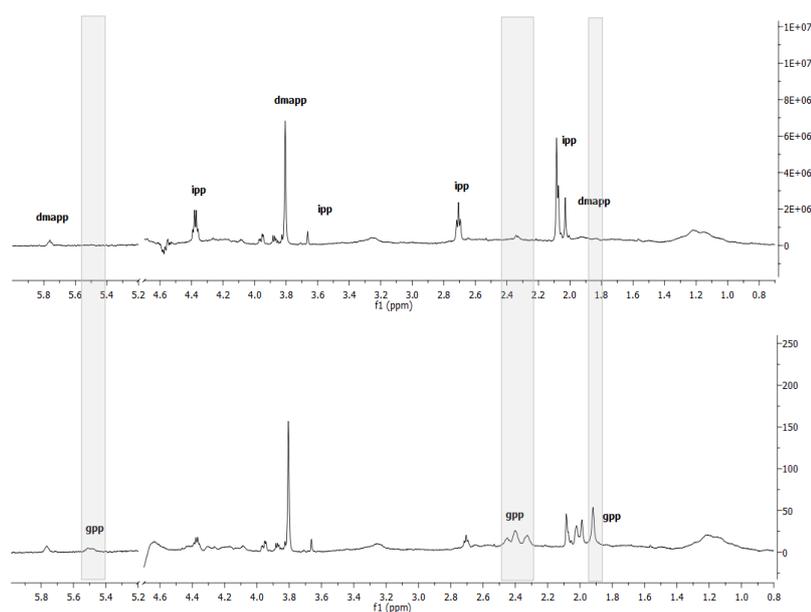
**Figure 15:** Colorimetric enzymatic assay of FPPS.

FPPS colorimetric enzymatic assay revealed to be low sensitive test, with possibility of obtaining false positive results. In order to overcome the low sensitivity of this test and the problematic of the enzymatic assay related to the use of radio-labeled ligand, we started the experimental process to develop an NMR based FPPS enzymatic assay. As previously reported the FPPS enzymatic activity consists in the synthesis of GPP (geranyl pyrophosphate) from IPP (isopentenyl pyrophosphate) and DMAPP (dimethylallyl pyrophosphate) substrates. Both the substrates and the products are small molecules observable with NMR techniques, therefore we recorded a set of mono-dimensional  $^1\text{H}$  NMR experiments as follows:

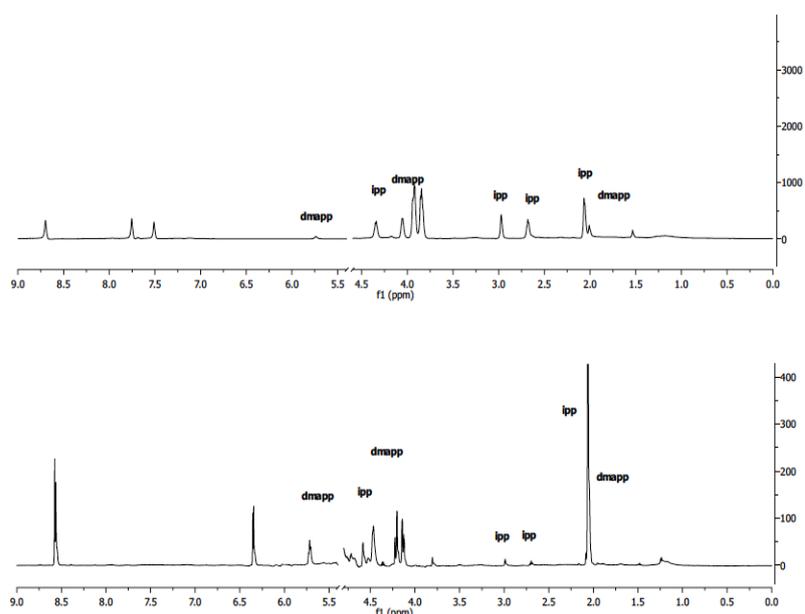
- $^1\text{H}$  NMR spectra of the single IPP and DMAPP substrates to have reference spectra.
- $^1\text{H}$  NMR spectra of 2:1 molar ratio IPP (200 $\mu\text{M}$ ) and DMAPP (100 $\mu\text{M}$ ) mixture in absence of FPPS enzyme, to exclude the presence of reactivity in absence of enzyme.
- $^1\text{H}$  NMR spectra of 2:1 molar ratio IPP (200 $\mu\text{M}$ ) and DMAPP (100 $\mu\text{M}$ ) mixture in presence of FPPS enzyme. (Figure 16)
- $^1\text{H}$  NMR spectra of 2:1 molar ratio IPP (200 $\mu\text{M}$ ) and DMAPP (100 $\mu\text{M}$ ) mixture in presence of FPPS enzyme and the known high affinity inhibitor zoledronic acid.
- $^1\text{H}$  NMR spectra of 2:1 molar ratio IPP (200 $\mu\text{M}$ ) and DMAPP (100 $\mu\text{M}$ ) mixture in presence of FPPS enzyme and i6A.

We recorded experiments at 37°C at the time of sample preparation ( $t=0$ ) and after 1h of incubation ( $t=60'$ ). Figure 5 shows the comparison of the spectra recorded at  $t=0$  and  $t=60'$ , as evident, signals relative to the side chain of GPP appear. These spectra prove the possibility that, by monitoring the intensity of

substrate and products H signals, it is possible to quantitatively measure the FPPS enzymatic activity and consequently, to measure ligand inhibitory activity. Accordingly, NMR spectra recorded in experimental conditions analogous to those of the spectra shown in Figure 16 and including zoledronic acid or i6A as inhibitor, show an appreciable decrease in the intensity of  $^1\text{H}$  GPP signals with respect to the spectra recorded in absence of zoledronic acid or i6A respectively (Figure 17). The quantitative evaluation of these data is currently in progress, but these results provide preliminary represent the initial steps on the route of the optimization an NMR based enzymatic assay useful to appreciate the inhibitory activity of newly identified FPPS ligands.



**Figure 16:** NMR enzymatic activity of FPPS.



**Figure 17:** NMR enzymatic activity of FPPS: up) in presence of zoledronic acid, down) in presence of i6A.

### 2.3 Conclusions

In the reported work, I proved that FPPS is a molecular target for i6A, in agreement with the biological evidences and in silico inverse virtual screening evidence. STD NMR experiments allowed to confirm computational data indicating that i6A is able to occupy the FPPS active site, with the isopentenyl moiety and adenosyl ring possibly oriented toward the deep region of the binding site. Based on the i6A-FPPS binding data, i6A analogues were studied for their interaction with FPPS. STD experiments show that the insertion of benzyl ring on N6-adenosine position induces a significant improvement in the interaction with FPPS enzyme, in particular CM223 was identified as a new lead compound in the search of FPPS inhibitors. From the methodological point of view, my experimental work demonstrates that NMR techniques are

powerful and fast tools to provide valuable information on the protein-ligand interaction and to quantitatively measure the enzymatic activity.

## Materials and Methods

### *Expression of FPPS enzyme*

The expression vector p11 was purchased from SGC-Oxford and contained the T7/Lac promoter and ampicillin resistance. The plasmid was transformed into *E. coli* BL21(DE3)-pLysS cells. FPPS was expressed in *Escherichia coli* as a fusion protein (67–419 residues) with a N-terminal poly-histidine tail and a mutation (threonine with serine) on residue 266, molecular weight 43kDa. Isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG) was purchased from Sigma-Aldrich (St. Louis, MO). Cell growth was monitored spectrophotometrically by measuring OD<sub>600</sub> nm periodically. For expression in *E. coli*, bacterial clones were grown in 1L of LB (Luria-Bertani) medium containing 50 $\mu$ g/mL ampicillin. When growth was performed to an OD<sub>600</sub> of 0.7 at 37°C, the 1mM isopropyl-D-thiogalactoside (IPTG) was added.

### *NMR Experiments*

All chemicals were purchased from Microtech Srl. i6A was purchased from Iris Biotech GMBH and zoledronic acid from Sigma-Aldrich (St. Louis, MO). i6A analogs were synthesized at University of Pisa and University of Milano. STD-NMR and WaterLOGSY experiments were recorded at 25°C on Bruker AV600 MHz spectrometer at a <sup>1</sup>H resonance frequency of 600 MHz equipped with a 5mm triple resonance <sup>1</sup>H(<sup>13</sup>C/<sup>15</sup>N), z-axis pulsed-field gradient probe head. For characterization purposes, i6A samples consisted of a 5 mM solution in 25 mM d-Tris, pH 7.4, 0.5 mM MgCl<sub>2</sub>, and 25 mM NaCl with 1% dimethyl

sulfoxide-d<sub>6</sub> as a co-solvent, and the spectra were referenced to residual solvent. <sup>1</sup>H-<sup>1</sup>D spectra were acquired at a resolution of 16k complex points in the time domain with 32 accumulations each (sw=6000 Hz, d1=3s). The FPPS protein at 8μM concentration was titrated with i6A to have protein/ligand molar ratios 1:10, 1:20, 1:30, 1:50, 1:70, and 1:100. For each addition of ligand, STD build-up experiment was carried out using different saturation times (0.50, 1.00, 1.50, 2.00, 3.00, 4.00, and 5.00 s and different relaxation delay 1.50, 2.00, 2.50, 3.00, 4.00, 5.00, and 6.00 s). For each experiment in the frequency list (FQ2LIST), the on-resonance and off-resonance pulse were 320 and 50000 Hz, respectively.

Briefly, two free induction decay (FID) data sets were collected in an interleaved manner to minimize temporal fluctuations with the protein irradiation frequency set on-resonance (-0.5 ppm) and off-resonance (40 ppm), respectively (sw = 6000 Hz, 16 steady state scans, 2048 transients, 4k complex points, d1 = 3s). Protein saturation was obtained using a train of individual 50 ms long, frequency selective Gaussian radio frequency (rf) pulses separated by an interpulse delay of 1 ms. The number of selective pulses was set to 50, leading to a total saturation time (τ<sub>sat</sub>) of 2.5 s. gradient tailored excitation (WATERGATE) scheme was employed to suppress the residual water signal.<sup>15</sup> Suppression of the background signals arising from the protein was not required. The FID acquired with off-resonance irradiation generated the reference spectrum (I<sub>off</sub>), whereas the difference FID (off-resonance on-resonance) yielded the STD spectrum (I<sub>STD</sub> = I<sub>off</sub> - I<sub>on</sub>). Spectra were processed with an exponential apodization function (<sup>1</sup>Hz line-broadening) and zero-filling to 8k complex points before Fourier transformation and baseline correction with a thirdorder Bernstein polynomial

fit. The STD measurements were done in duplicate, and all data were processed and analyzed using TopSpin software (Bruker v 1.3).

One-dimensional  $^1\text{H}$  NMR WaterLOGSY experiments were acquired. A reference experiment was recorded followed by the WaterLOGSY magnetization transfer spectrum. Acquisition parameters for the WaterLOGSY spectra were 256-512 scans with a mixing time of 1.4 s and a 2 s relaxation delay.

## References

1. Bifulco, M.; Malfitano, A. M.; Proto, M. C.; Santoro, A.; Caruso, M. G.; Laezza, C. Biological and pharmacological roles of N-6-Isopentenyladenosine: An emerging anticancer drug. *Anti-Cancer Agents in Medicinal Chemistry* **2008**, *8*, 200-204.
2. Laten, H. M.; Zahareasdoktor, S. Presence and source of free isopentenyladenosine in yeasts. *Proceedings of the National Academy of Sciences of the United States of America* **1985**, *82*, 1113-1115.
3. Kersten, H. On the biological significance of modified nucleosides in tRNA. *Progress in nucleic acid research and molecular biology* **1984**, *31*, 59-114.
4. Thurnher, M.; Nussbaumer, O.; Gruenbacher, G. Novel Aspects of Mevalonate Pathway Inhibitors as Antitumor Agents. *Clinical Cancer Research* **2012**, *18*, 3524-3531.
5. Laezza, C.; Notarnicola, M.; Caruso, M. G.; Messa, C.; Macchia, M.; Bertini, S.; Minutolo, F.; Portella, G.; Fiorentino, L.; Stingo, S.; Bifulco, M. N6-isopentenyladenosine arrests tumor cell proliferation by inhibiting farnesyl diphosphate synthase and protein prenylation. *Faseb Journal* **2006**, *20*, 412-418.
6. Ciaglia, E.; Pisanti, S.; Picardi, P.; Laezza, C.; Malfitano, A. M.; D'Alessandro, A.; Gazzero, P.; Vitale, M.; Carbone, E.; Bifulco, M. N6-isopentenyladenosine, an endogenous isoprenoid end product, directly affects cytotoxic and regulatory functions of human NK cells through FDPS modulation. *Journal of Leukocyte Biology* **2013**, *94*, 1207-1219.
7. Sacchettini, J. C.; Poulter, C. D. Biochemistry - Creating isoprenoid diversity. *Science* **1997**, *277*, 1788-1789.
8. Szkopinska, A.; Plochocka, D. Farnesyl diphosphate synthase; regulation of product specificity. *Acta Biochimica Polonica* **2005**, *52*, 45-55.
9. Lauro, G.; Romano, A.; Riccio, R.; Bifulco, G. Inverse Virtual Screening of Antitumor Targets: Pilot Study on a Small Database of Natural Bioactive Compounds. *Journal of Natural Products* **2011**, *74*, 1401-1407.
10. Wang, J.-C.; Chu, P.-Y.; Chen, C.-M.; Lin, J.-H. idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. *Nucleic Acids Research* **2012**, *40*, W393-W399.
11. Chen, Y. Z.; Zhi, D. G. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins-Structure Function and Genetics* **2001**, *43*, 217-226.
12. Kavanagh, K. L.; Guo, K. D.; Dunford, J. E.; Wu, X. Q.; Knapp, S.; Ebetino, F. H.; Rogers, M. J.; Russell, R. G. G.; Oppermann, U. The molecular mechanism of nitrogen-containing bisphosphonates as anti osteoporosis drugs. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 7829-7834.
13. Scrima, M.; Lauro, G.; Grimaldi, M.; Di Marino, S.; Tosco, A.; Picardi, P.; Gazzero, P.; Riccio, R.; Novellino, E.; Bifulco, M.; Bifulco, G.; D'Ursi, A. M. Structural Evidence of N6-Isopentenyladenosine As a New Ligand of Farnesyl Pyrophosphate Synthase. *Journal of Medicinal Chemistry* **2014**, *57*, 7798-7803.
14. Gao, J.; Chu, X.; Qiu, Y.; Wu, L.; Qiao, Y.; Wu, J.; Li, D. Discovery of potent inhibitor for farnesyl pyrophosphate synthase in the mevalonate pathway. *Chemical Communications* **2010**, *46*, 5340-5342.
15. Piotta, M.; Saudek, V.; Sklenar, V. Gradient-tailored excitation for single-quantum nmr-spectroscopy of aqueous-solutions. *Journal of Biomolecular Nmr* **1992**, *2*, 661-665.

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. *Journal of the American Chemical Society* **2001**, *123*, 6108-6117.
17. Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *Journal of Biomolecular Nmr* **2001**, *21*, 349-359.
18. Angulo, J.; Nieto, P. M. STD-NMR: application to transient interactions between biomolecules-a quantitative approach. *European Biophysics Journal with Biophysics Letters* **2011**, *40*, 1357-1369.
19. Angulo, J.; Enriquez-Navas, P. M.; Nieto, P. M. Ligand-Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR Spectroscopy: The Binding Isotherm of STD Initial Growth Rates. *Chemistry-a European Journal* **2010**, *16*, 7803-7812.
20. Wang, Y. S.; Liu, D. J.; Wyss, D. F. Competition STD NMR for the detection of high-affinity ligands and NMR-based screening. *Magnetic Resonance in Chemistry* **2004**, *42*, 485-489.

## 3 *Development of INPHARMA-STRING filtering protocol*

### 3.1 Introduction

INPHARMA (Interligand NOEs for PHARmacophore MApping) is a NMR method to determine the relative binding modes of two low-affinity ligands that bind competitively a common receptor site.

INPHARMA method is based on the observation of interligand, spin-diffusion mediated, transferred-NOEs between two ligands, which bind competitively and weakly to a receptor. During the mixing time of the NOESY experiment, inter-ligand NOE cross-peaks result from magnetization transfer mediated by protein protons. Inter-ligand INPHARMA NOEs are estimated for pairs of protein-ligand complexes and subsequently fitted to the theoretical inter-ligand NOEs to discriminate the correct binding modes.<sup>1,2</sup>

Molecular docking is a rapid and inexpensive computational technique. It offers, for a given ligand, a large number of plausible binding poses. At the early stage of the computational technique development, molecular docking methods were based on traditional “key-and-lock” principle. However they revealed limitations in the correct calculation of protein flexibility and entropic effects, often leading to the identification of false positive hits.<sup>3-5</sup>

Therefore, molecular docking methods have been improved over the years, taking in account the local flexibility of the complex, in a strategy called *soft-docking*. Subsequently *ensemble docking* procedures were implemented where multiple receptor structures were used to describe protein flexibility.<sup>6-9</sup> *Induced-fit docking* is nowadays, the most used molecular docking technique

that combines the advantages of each of the above-mentioned approaches together.

This is Schrödinger's innovative method for fast and accurate prediction of ligand induced conformational changes in receptor active sites.<sup>10</sup>

In each of the above mentioned docking approaches, the results are ranked according to a consensus *scoring function* and to molecular dynamics force field interaction energies.

Although computational techniques offer a significant contribution to the calculation and the selection of the ligand binding poses, the most efficient strategy of ranking to select the correct binding modes, takes advantage of the combination of calculated and experimental data.

Reliable experimental data can be obtained by ligand-observed NMR techniques including experiments such as transfer NOESY,<sup>11, 12</sup> saturation transfer difference (STD),<sup>13</sup> inter-ligand NOEs (ILOES),<sup>14</sup> inter-ligand NOEs for PHARmacophore Mapping (INPHARMA).<sup>15</sup>

INPHARMA methodology is currently used as an experimental tool to select the correct binding modes within a pool of pairs of complex structures generated by molecular docking.<sup>1, 2</sup> However, this method might lead to false positive hits or ambiguous answers when the starting structure for the docking experiments is inaccurate.<sup>16</sup>

In this context, INPHARMA-STRING<sup>17</sup> has been developed as a new scoring protocol employing *induced-fit docking* data and NMR-derived inter-ligand NOEs (INPHARMA NOEs).

The participation to the development of INPHARMA-STRING protocol represented my PhD experience in the group of Prof. Teresa Carlomagno at Structural and Computational Biology Unit, EMBL (European Molecular

Biology Laboratory), Heidelberg, Germany. INPHARMA-STRING was developed using as model protein kinase A (PKA)<sup>16</sup> that had been previously, successfully studied together with Epo A-tubulin<sup>15</sup> complex for the development of INPHARMA protocol.<sup>2</sup>

INPHARMA-STRING protocol was tested on five diverse ligands and employing data from a series of realistic docking scenarios, ranging from the fully open to the inhibitor bound closed form of PKA. In particular, my contribution in the development of INPHARMA string protocol consisted in the acquisition and analysis of INPHARMA spectra of a set of selected ligands pairs.

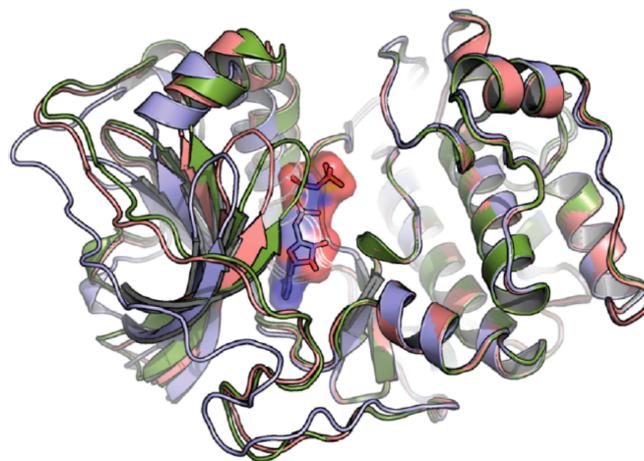
## 3.2 Results and Discussion

### 3.2.1. Structural Characterization and Docking Evaluation

Protein Kinase A (PKA) is an extensively characterized pharmacological target, exerting important actions in the cell, such as regulation of glycogen, sugar and lipid metabolism.

PKA is an enzyme that consists of two regulatory subunits and two catalytic subunits. When the concentration of cyclic-AMP increases, after activation of Gs-protein coupled receptor, two molecules of cyclic-AMP bind the two regulatory subunits of PKA, that move out and to free catalytic subunits and interact with protein to phosphorylate serine and threonine residues.

Three different sets of structural coordinate are present for PKA in RCSB protein databank, corresponding to three distinctive PKA conformational states: (1) fully open, (2) intermediate, and (3) fully closed, (heavy-atom RMSD at binding site 1.9/2.1 Å) (Figure 1).



**Figure 1:** Visualization of three representative PKA structures superimposed on the large lobe: fully open apo (blue; PDB id 1CMK), intermediate (red; PDB id 3DND), and fully closed (green; PDB id 1CDK). Ligand present in PDB id 1CDK is shown in stick and surface representation.

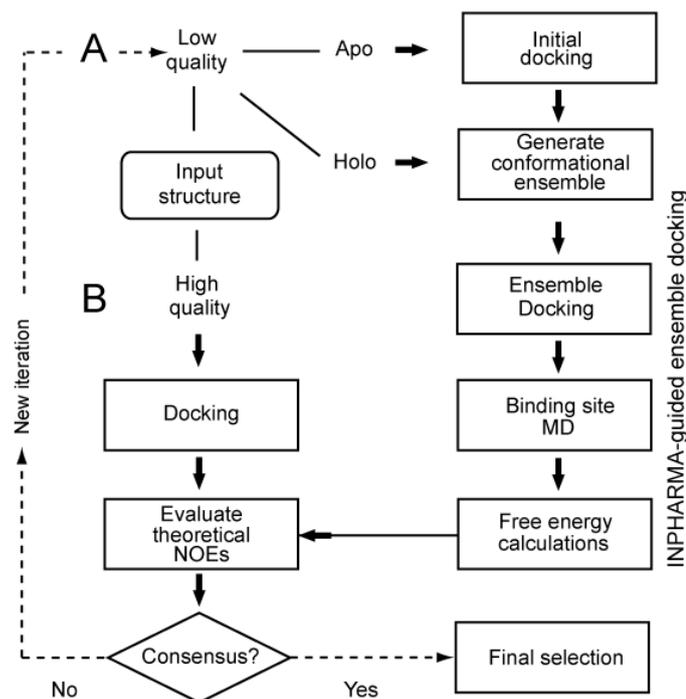
Standard molecular docking experiments were carried out for all the PKA conformations (fully open, fully close and intermediate). Additionally *ensemble docking* calculation was carried out for a pool of X-ray structures, and including a low quality homology model of human PKA built from PKB (PDB id: 2JDO).

Molecular docking to the high-resolution X-ray structures with PDB ids 3DND, 3AGM, and 1Q8W yielded a success rate of 60-80%. Molecular docking to the fully closed structure (PDB id 1CDK) resulted in a lower 40% success rate. Finally, the PKA homology model and the fully open apo enzyme (PDB id 1CMK) yielded docking success rates as low as 20% and 0%, respectively. This demonstrates the general challenge of obtaining reliable results with current docking protocols, even when high-resolution structures are available.

Docking data evidenced the necessity of new rescoring protocol based on experimentally derived NMR data: INPHARMA protocol meets these requirements.

### 3.2.2. NMR experiments and rescoring

Conventional or ensemble based docking experiments were the starting point to design new extended protocol for rescoring protein-ligand binding modes. Figure 2 shows a flow chart describing a protocol relative to two paths determined either by the a priori knowledge of the query system, or by INPHARMA-guided evaluation of the docking modes.



**Figure 2:** Overview of the INPHARMA-guided ensemble docking protocol.

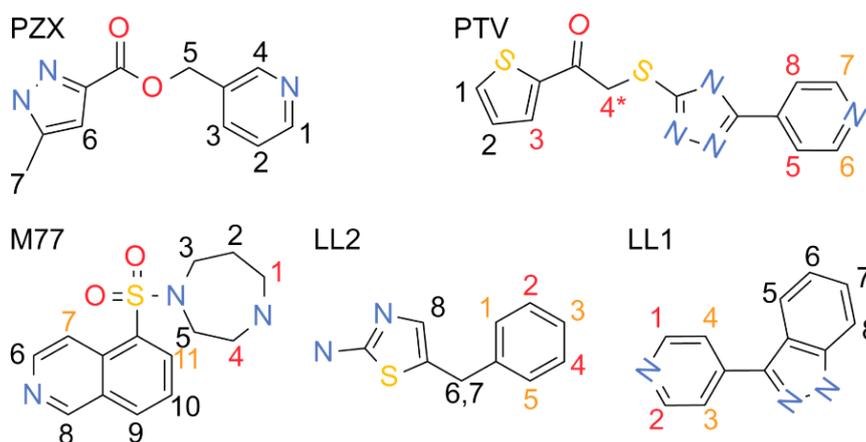
In the case of a high quality “rigid” receptor structure, filtering by INPHARMA-STRING is directly performed on the docking modes to the single static receptor. In this scenario, flexibility is only incorporated through minimization of the docking modes. In the presence of several X-ray structures of the target protein, an analogous approach can be applied for *ensemble docking* to the pool of X-ray structures. (Figure 2A) Conversely, with a low-quality structure or in case of a flexible target, the protocol includes a more extensive procedure, utilizing MD-generated *ensemble docking* (Figure 2B). Therefore, prior to the calculation of the theoretical INPHARMA-NOEs two additional computational steps were included: i) a short binding site MD of each docking mode, which allows for mutual protein-ligand adaptation (*induced-fit*), and ii) subsequent filtering of high energy conformations.

NMR derived inter-ligand INPHARMA-NOEs represent the key element in the new rescoring function of computationally generated ligand binding modes. Based on different docking scenarios, INPHARMA-STRING protocol, was developed, which rigorously determines the correct protein-ligand binding modes.

In the case of PKA, we illustrate by combining molecular modeling and INPHARMA-guided docking, that selected receptor conformers result in a relatively small structural deviation from the target X-ray structures. This demonstrates the remarkable feature of the INPHARMA data, which contain information not only on the ligand binding pose, but also, indirectly, on the conformation of the binding pocket mediating the magnetization transfer.

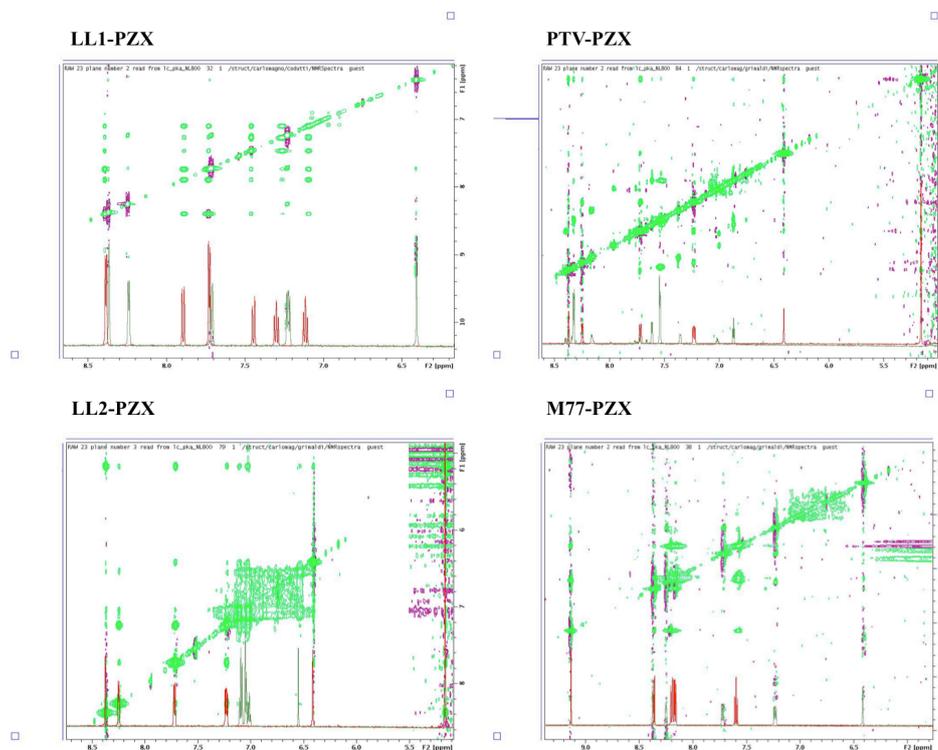
To develop the new INPHARMA-STRING protocol, INPHARMA experiments were carried out using the five ligands LL1, LL2, M77, PTV and

PZX. (Figure 3)



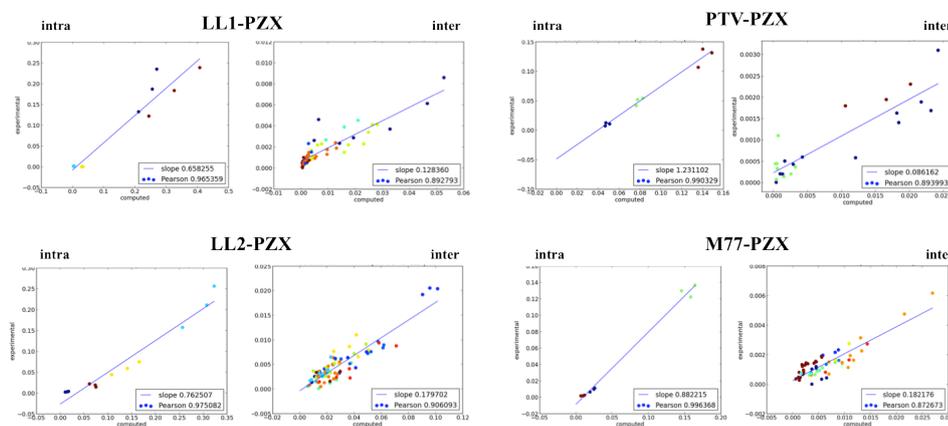
**Figure 3:** Structures of PKA ligands. Positions for NMR-detectable hydrogen moieties are marked numerically. Red and orange fonts denote overlapping chemical shifts between two or more protons, whereas protons whose resonance overlaps with the solvent are marked by an asterisk.

Experiments with a total of 10 ligand pair combinations, in the presence of PKA protein, were acquired: 30 INPHARMA experiments, corresponding to three NOESY mixing times for each ligand pair combination. Within the described procedure, my contribution consisted in the analysis of four pairs of ligand PZX over ten possible combinations. (Figure 4)



**Figure 4:** Zoom of INPHARMA spectra of PZX pairs.

In order to verify the reliability of the experimental data relative to the crystal structure of PKA complex, we compared the observed intra-ligands NOEs and INPHARMA NOEs (inter-ligands) with computed-NOEs calculated from the corresponding crystallographic structure. The kinetic INPHARMA parameters, for the analysis of the NOEs, were set as in Materials and Method. (Figure 5)



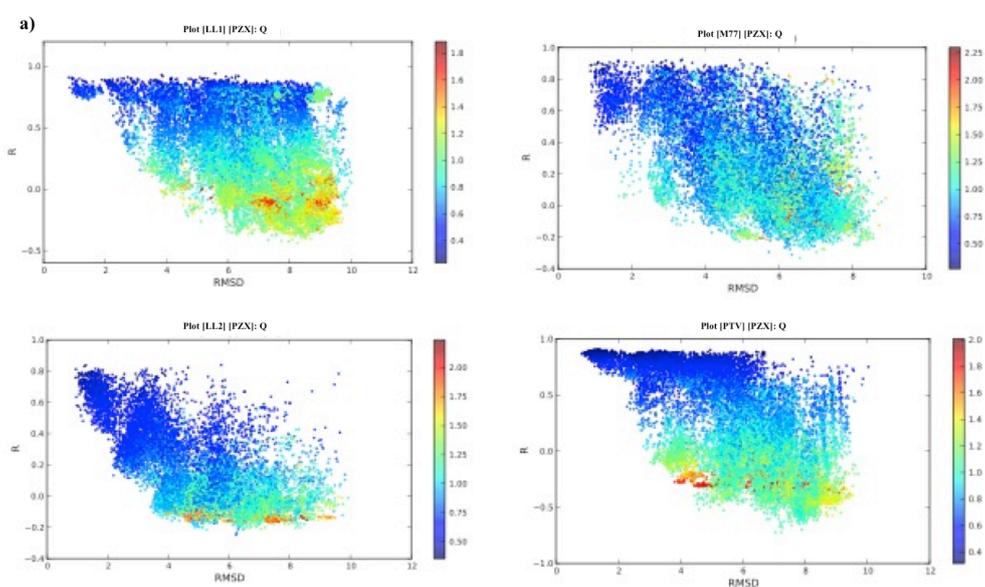
**Figure 5:** Fitting of the experimental NOEs (y-axis) and computed NOEs (x-axis) for all the pairs of the ligand PZX.

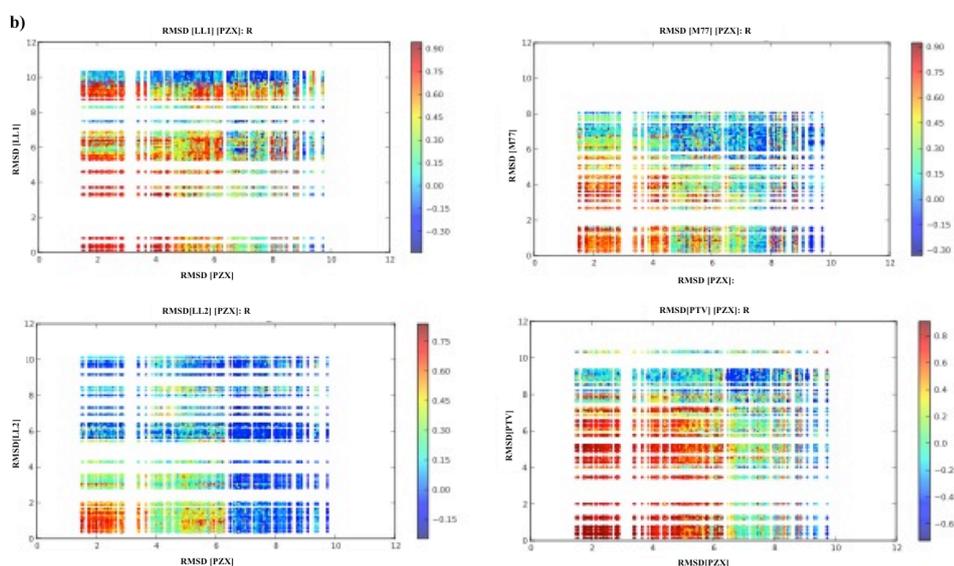
After this preliminary step, to check the validity of our experimental data, we approached INPHARMA protocol. We used the NMR derived experimental data to select for each pair the docking modes that were in the best agreement with the INPHARMA data. An ensemble of 100 docking modes were usually generated for each ligand, all protein-ligand structures were minimized prior to the INPHARMA calculation using Amber<sup>18</sup>. The resulting theoretical NOEs were evaluated with respect to their agreement with the experimental NOEs using the parameters, Pearson correlation coefficient (R), Quality factor (Q), and RMSD parameters as *INPHARMA score* (see Materials and Methods). The value of the *INPHARMA score* for ligand PZX pairs is correlated in Figure 6 to the structural deviation of the ligand from their correct orientations.

Figure 6 a shows plots for each pairs of PZX, two graphical distributions of the docking modes compared with the experimental data using the *R*, *Q* and *RMSD* parameters. We show four plots, where on the x-axis and y-axis, the

values of *RMSD* from the crystal structures (best values close to 0) the values of *R* (best values close to 1) are reported.

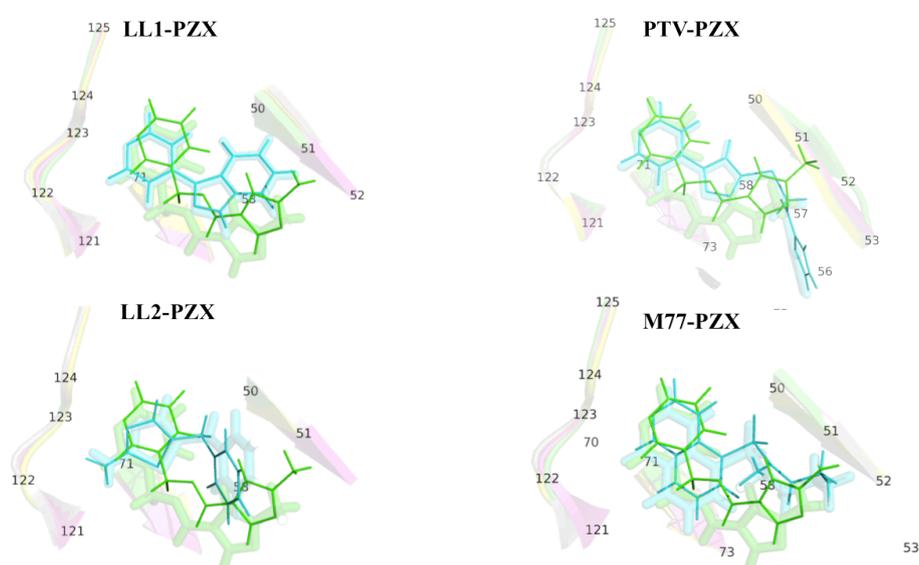
Figure 6 b shows four plots for each pairs of PZX, where on the x/y-axis the values of *RMSD* of two ligands are reported.





**Figure 6:** Evaluation of INPHARMA score. a) in color code the Q value parameter, b) in color code the RMSD value parameters.

By following the described procedure, we selected the 10 best binding poses. Figure 7 shows the comparison between the binding poses of selected ligands in the crystal structures and the binding poses selected thanks to the NMR based experimental data. For the ligands LL1, LL2, M77 and PTV the overlapping with the structures present in database leads to RMSD values close to 0. For the ligand PZX the RMSD value is close to 2 in each pair.



**Figure 7:** Comparison between binding poses of the ligands in the crystal structures and the docking modes selected by experimental data.

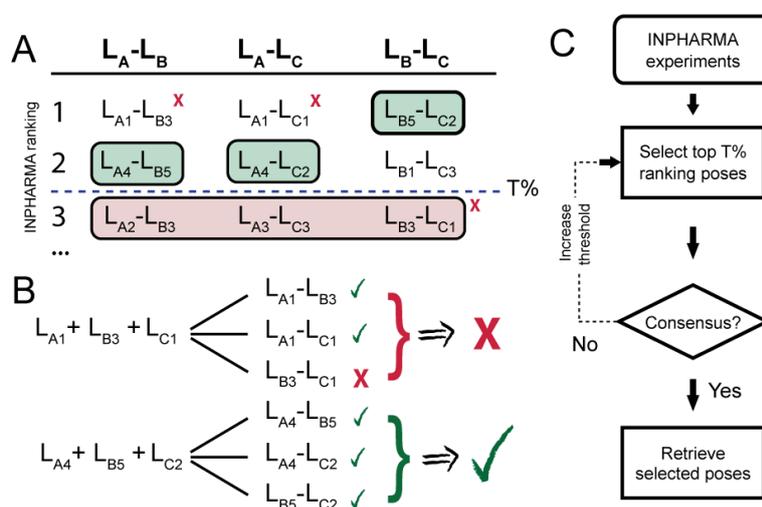
However, the fitting of theoretical INPHARMA-NOEs estimated from a pair of docking modes of two ligands to the experimental NOEs does not always provide a unique solution. As recurrent presence of false positive is observable, better rescoring algorithm are necessary to leverage the ambiguities and deliver a precise and accurate selection of binding modes.

### 3.2.3 INPHARMA-STRING

In view of the presence of false positives in the INPHARMA selection, we sought a method that could simultaneously interpret, in an automated approach, the INPHARMA information from multiple ligand pairs, to exclude false positive hits. Since each individual INPHARMA experiment comprises a pair of ligands (i.e.,  $L_A$ - $L_B$ ), each ligand also occurs in combinations with other ligands (e.g.,  $L_A$  with  $L_C$ ,  $L_D$ , etc.). The property of such a composition of the experiments can be exploited to determine a consensus selection of the docking modes.

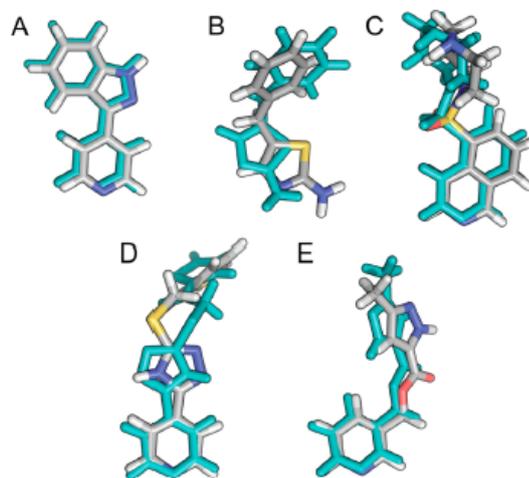
Let ( $L_A$ ,  $L_B$ ,  $L_C$ ) be a set of 3 ligands whose docking modes are characterized by indices 1,2,3 etc., respectively. Ranking by INPHARMA score of docking modes associated with three ligand pairs  $L_A$ - $L_B$ ,  $L_A$ - $L_C$ ,  $L_B$ - $L_C$  is shown in Figure 8A. A consensus criterion is searched for each “string” of docking modes in which all pair-wise combinations should be above the threshold ( $T\%$ ). Here, pairs of docking modes -  $L_{A1}$ - $L_{B3}$  and  $L_{A1}$ - $L_{C1}$ - are discarded because  $L_{B3}$ - $L_{C1}$  is below the threshold  $A$ , while a string of docking modes  $L_{A4}$ ,  $L_{B5}$ ,  $L_{C2}$  is selected if and only if all combinations of binding modes ( $L_{A4}$ - $L_{B5}$ ,  $L_{A4}$ - $L_{C2}$ ,  $L_{B5}$ - $L_{C2}$ ) score high with respect to the INPHARMA experimental data. (Figure 8)

This approach entitled INPHARMA-STRING, is based on the simultaneous analysis of multiple ligands combined pair-wise in the presence of the protein target; in our study we have 5 ligands and the protein PKA.



**Figure 8:** INPHARMA-STRING filtering protocol.

In our system, the consensus final selection of docking modes was applied to PDB id 3AGM of the ligands LL1 (A), LL2 (B), M77 (C), PTV (D), and PZX (E) and a consensus quintuplet of binding modes was found for the five ligands. Figure 9 shows the ligand binding poses in the X-ray structures compared to the docking modes selected by INPHARMA-STRING. Remarkably the RMSD values vs the X-ray structures were low: 0.6, 1.6, 0.8, 1.3, and 1.2 Å, respectively (Figure 9). Similar result were obtained for the rescoring of the ligand docking modes vs the two different receptor structures representing the intermediate conformation (3DND and 1Q8W) (Table I). Analysis of the data shows that the application of the NMR-based rescoring and the consensus criterion selection protocol effectively filters out false positives resulted from the standard docking calculation and from the pairwise INPHARMA selection.



**Figure 9:** Consensus selection of 3AGM docking modes. The final selection of docking modes to PDB id 3AGM of the ligands LL1 (A), LL2 (B), M77 (C), A14 (D), and INH (E). The ligand binding pose in the X-ray structures are shown as grey sticks, while docking modes selected by INPHARMA-STRING are shown in cyan sticks.

### 3.2.4 Docking Scenario

INPHARMA-STRING protocol was applied to i) high-resolution PDB (ids 3DND, 1Q8W and 3AGM) in the intermediate state of receptor, obtained with a 60– 80% success rate ii) the fully closed receptor structure (PDB id 1CDK) and iii) the fully open receptor structure (PDB id 1CMK), where the docking success rate was low from 40% to 20%-0%. In these last cases, the proposed INPHARMA-STRING protocol allowed to recognize incorrect binding modes in the pool: the consensus selection was not found. Consequently, these cases demonstrate that a single static protein structure might not be sufficient for a reliable molecular docking calculation. To overcome this problem, the INPHARMA-STRING protocol was implemented to consider the full protein flexibility by employing *ensemble docking* to MD-generated conformers.

The transferred-NOEs information was discarded and flexible ligands were used in the docking protocol. The molecular modeling of *induced-fit* effects represents one of the major challenges in protein-ligand docking, and the ability to rapidly reach a reliable model of the complex would potentially accelerate SBDD. Our objective was to demonstrate that INPHARMA-NOEs alone are sufficient to discriminate between docking modes and bound ligand conformations, even in the absence of high quality receptor structures.

In this scenario, the approach INPHARMA guided *ensemble docking* to MD-generated conformers was applied to the fully close structure (PDB id 1CDK), to the homology model PKB<sub>MD</sub> (PKA built from PKB, with a sequence identity of 42%) and to fully open enzyme of PKA (PDB id 1CMK). In Table I, we reported the results of proposed protocol based on the combination of *MD simulations*, *ensemble docking* and INPHARMA-STRING rescoring (Figure 2). Analysis of data indicates a remarkable improvement of the ligand binding pose prediction, even in the presence of large structural changes.

**Table I:** Performance of INPHARMA-STRING in docking scenario.

	LL1	LL2	M77	PTV	PZX
<b>3AGM</b>	0.3-0.6	0.6-1.6	0.8	1.3	0.5-1.3
<b>3DND</b>	0.4-1.1	1.5	0.8	2.1	0.2
<b>1Q8W</b>	1.9	1.6-1.9	2.2	1.2-4.3	1.6
<b>Pooled</b>	0.4	1.6	1.0	4.3	0.2
<b>1CDK<sub>MD</sub></b>	1.3	1.3-1.5	0.9	2.3	0.9
<b>PKB<sub>MD</sub></b>	1.3	0.9	1.3	0.6	1.7
<b>1CMK<sub>MD</sub></b>	0.8	1.5	1.6	0.9	2.1

<sup>a</sup>RMSD values (in Å) for the predicted binding modes (by INPHARMA-STRING; to their respective X-ray structures) of the five ligands (column-wise) are provided for each docking scenario (row-wise). Results are presented for conventional and X-ray ensemble docking above the empty row, and MD ensemble docking (noted with MD in subscript) below the empty row. PKB: Homology model of PKA built from PKB (PDB id: 2JDO).

This work demonstrates the potential of combining extensive molecular modeling with NMR-guided *ensemble docking* to decipher receptor-ligand interactions in the case of *induced-fit*, and highlights the importance of NMR in SBDD.<sup>19, 20</sup>

### 3.3 Conclusions

In conclusion, during my research experience in the research group of Prof. Carlomagno, I contributed to the development of new INPHARMA method, named INPHARMA-STRING. The new protocol was tested on ten ligand pairs including 3 different ligands, INPHARMA STRING combined with extensive molecular modeling and simulation techniques, demonstrated to be an efficacy method to accurately and reliably select the docking mode, reducing the possibility of selecting false positives in INPHARMA procedure;

in this respect it showed to be a suitable method for high-throughput campaigns.

### **Materials and Methods**

A set of five PKA inhibitors was collected from our inhouse compound library based on their chemical diversity and low affinity for PKA (Figure 3). The ligands, referred to by their PDB ligand ids (LL1, LL2, M77, PTV, and PZX) were characterized by dissociation constants ( $K_D$ ) ranging from 6 to 30  $\mu\text{M}$ . High resolution structures of PKA bound to ligands LL1, LL2, and M77 were collected from PDB with ids: 3DNE, 3DND, and 1Q8W, respectively. Structures of PKA bound to ligands with ids PTV and PZX were obtained by X-ray crystallography, at 2.6 and 1.9 Å resolution, respectively. Coordinates are available from the RCSB protein data bank with accession codes 4IJ9 (ligand PTV) and 4IE9 (ligand PZX).

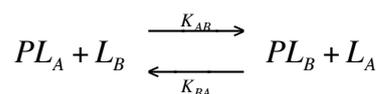
Hamster PKA for NMR experiments was prepared starting from constructs described elsewhere, and stored in per-deuterated PBS buffer (pH 7.4, [NaCl]=150mM). The buffers for both the ligand characterization and the INPHARMA experiments contained an addition of 5% in volume of d-DMSO. NMR samples contained ligands-protein in a 10:1 stoichiometric ratio for a final protein concentration of 25  $\mu\text{M}$ .

### *INPHARMA Experiments and Calculations*

We collected a total of 10 INPHARMA experiment sets, corresponding to all pairwise combinations of the five ligands. Each INPHARMA set was recorded as a fully interleaved scan by scan series of NOESY spectra with mixing times 300, 500, and 700 ms on an 800 MHz spectrometer (Bruker).

Spectral assignment and peak integration was made using Felix 2007. The relative  $K_{DS}$  of the ligands were estimated by means of STD competition experiments.<sup>21</sup>

Kinetic parameters and protein concentration used for the INPHARMA evaluation were set as follows: ligands  $\tau_c$  was set to 0.1 ns, the concentration of the ligands was 250  $\mu\text{M}$ , with the exception of PTV, which degraded over time. Its concentration was estimated for every experiment set and varied between 150  $\mu\text{M}$  and 80  $\mu\text{M}$ . The kinetic parameters were optimized using a direct exchange kinetic model:



where  $k_{AB}$  and  $k_{BA}$  are the kinetic constants for the forward and the backward reaction, respectively (scaled by  $10^{-3}$ ). The concentration of the complex was tuned by fitting experimental intra-ligand NOEs. Theoretical inter-ligand INPHARMA NOEs were calculated for all protein-ligand pairs in each of the docking scenarios using an in-house written program in accordance to the previous theoretical development, using a direct exchange model to describe the protein-ligand kinetics.<sup>2</sup>

The quality of the data fitting and the selection of the correct binding modes by experimental INPHARMA are performed using the Pearson's correlation coefficient (R) (eq (1)), the slope (m) (eq (2)) and Q (Quality Factor) parameters:

$$R \equiv \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (1)$$

$$m \equiv \frac{y_2 - y_1}{x_2 - x_1} \quad (2)$$

$$Q_{(a)} \equiv \sqrt{\frac{\sum_i \left( \frac{y_i - ax_i}{y_i} \right)^2}{N}} \quad (3)$$

The R has been defined in the equations (1), to have a good selection of the docking data, the values of R should be close to 1. The Q parameter is a measure of the quality of the fitting of the data, to have a good fitting the value of Q should be close to 0.

The method is successful, if INPHARMA selects docking modes close to the crystallographic binding pose. Initially, to verify the quality of the experimental data, the experimental INPHARMA NOEs were compared with the theoretical data calculated from the crystallographic structures.

Subsequently, the actual selection of the docking modes was conducted by comparing the experimental NOEs observed in the NMR spectra and the theoretical NOEs calculated for the single binding pose. In order to simulate a real case, in which the x-ray structure is not known, we used PKA as a starting protein structure for docking, after removal of the ligands.

### *Structural Analysis, MD Simulations and Ensemble Docking*

All PKA structures with a sequence identity higher than 98% to the human sequence were collected from the RCSB protein databank.<sup>22</sup> Five representative high resolution structures with PDB id 3DND, 3AGM, 1Q8W,

1CDK, and 1CMK were selected as targets for cross-docking.

Homology model of human PKA was generated with Modeller<sup>23</sup> based on the high resolution structure of protein kinase B (PKB; PDB id 2JDO). All-atomic models of the high resolution X-ray structures with PDB id 1CDK, 1CMK, and the PKA homology model were generated with AmberTools<sup>24</sup> using the corresponding Amber force field.<sup>25</sup>

Molecular Dynamics (MD) simulations were carried out in explicit solvent employing a 1 fs time step under periodic boundary conditions. Ligands, with which the structure was crystallized, were included in the simulations for PDB id 1CDK and the homology model. Ligand parameters were generated using the generalized Amber force field.<sup>26</sup>

*Ensemble docking* was carried out with Surflex<sup>27</sup> and Glide on the PKA X-ray structures with PDB ids 3DND, 3AGM, 1Q8W, 1CDK, and 1CMK. All docking experiments were conducted using flexible ligands and up to 20 docking modes per ligand were collected for each receptor structure.

## References

1. Orts, J.; Tuma, J.; Reese, M.; Grimm, S. K.; Monecke, P.; Bartoschek, S.; Schiffer, A.; Wendt, K. U.; Griesinger, C.; Carlomagno, T. Crystallography-independent determination of ligand binding modes. *Angewandte Chemie-International Edition* **2008**, *47*, 7736-7740.
2. Orts, J.; Griesinger, C.; Carlomagno, T. The INPHARMA technique for pharmacophore mapping: A theoretical guide to the method. *Journal of Magnetic Resonance* **2009**, *200*, 64-73.
3. Davis, I. W.; Raha, K.; Head, M. S.; Baker, D. Blind docking of pharmaceutically relevant compounds using RosettaLigand. *Protein Science* **2009**, *18*, 1998-2002.
4. Verdonk, M. L.; Mortenson, P. N.; Hall, R. J.; Hartshorn, M. J.; Murray, C. W. Protein-Ligand Docking against Non-Native Protein Conformers. *Journal of Chemical Information and Modeling* **2008**, *48*, 2214-2225.
5. Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Current Opinion in Structural Biology* **2008**, *18*, 178-184.
6. Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews* **2006**, *26*, 531-568.
7. Carlson, H. A.; Meagher, K. L. Protein flexibility and drug design: How to hit a moving target. *Abstracts of Papers of the American Chemical Society* **2003**, *225*, U752-U752.
8. Teodoro, M. L.; Kavraki, L. E. Conformational flexibility models for the receptor in structure based drug design. *Current Pharmaceutical Design* **2003**, *9*, 1635-1648.
9. Barril, X.; Fradera, X. Incorporating protein flexibility into docking and structure-based drug design. *Expert Opinion on Drug Discovery* **2006**, *1*, 335-349.
10. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *Journal of Medicinal Chemistry* **2006**, *49*, 534-553.
11. Clore, G. M.; Gronenborn, A. M. Theory and applications of the transferred nuclear Overhauser effect to the study of the conformations of small ligands bound to proteins. *Journal of Magnetic Resonance (1969)* **1982**, *48*, 402-417.
12. Ni, F.; Zhu, Y.; Scheraga, H. A. Thrombin-bound structures of designed analogs of human fibrinopeptide-a determined by quantitative transferred noe spectroscopy - a new structural basis for thrombin specificity. *Journal of Molecular Biology* **1995**, *252*, 656-671.
13. Mayer, M.; Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie-International Edition* **1999**, *38*, 1784-1788.
14. Wyss, D. F.; Eaton, H. L. Fragment-based approaches to lead discovery. *Front Drug Des Discov* **2007**, *3*, 171-202.
15. Sanchez-Pedregal, V. M.; Reese, M.; Meiler, J.; Blommers, M. J. J.; Griesinger, C.; Carlomagno, T. The INPHARMA method: Protein-mediated interligand NOEs for pharmacophore mapping. *Angewandte Chemie-International Edition* **2005**, *44*, 4172-4175.
16. Orts, J.; Bartoschek, S.; Griesinger, C.; Monecke, P.; Carlomagno, T. An NMR-based scoring function improves the accuracy of binding pose predictions by docking by two orders of magnitude. *Journal of Biomolecular Nmr* **2012**, *52*, 23-30.
17. Skjaerven, L.; Codutti, L.; Angelini, A.; Grimaldi, M.; Latek, D.; Monecke, P.; Dreyer, M. K.; Carlomagno, T. Accounting for Conformational Variability in Protein-Ligand

Docking with NMR-Guided Rescoring. *Journal of the American Chemical Society* **2013**, 135, 5819-5827.

18. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995**, 91, 1-41.

19. Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T. L.; Homans, S. W.; Kessler, H.; Luchinat, C.; Meyer, B.; Oschkinat, H.; Peng, J.; Schwalbe, H.; Siegal, G. Perspectives on NMR in drug discovery: a technique comes of age. *Nature Reviews Drug Discovery* **2008**, 7, 738-745.

20. Carlomagno, T. NMR in natural products: understanding conformation, configuration and receptor interactions. *Natural Product Reports* **2012**, 29, 536-554.

21. Wang, Y. S.; Liu, D. J.; Wyss, D. F. Competition STD NMR for the detection of high-affinity ligands and NMR-based screening. *Magnetic Resonance in Chemistry* **2004**, 42, 485-489.

22. Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. The Protein Data Bank. *Acta Crystallographica Section D-Biological Crystallography* **2002**, 58, 899-907.

23. Sali, A.; Overington, J. P. Derivation of rules for comparative protein modeling from a database of protein-structure alignments. *Protein Science* **1994**, 3, 1582-1596.

24. Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2013**, 3, 198-210.

25. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics* **2006**, 65, 712-725.

26. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, 25, 1157-1174.

27. Jain, A. N. Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *Journal of Computer-Aided Molecular Design* **2007**, 21, 281-306.