DRUG DISCOVERY IN CANCER RESEARCH: HIT IDENTIFICATION USING A FRAGMENT-BASED BY NMR APPROACH

Direttore della Scuola: Ch.mo Prof. Paolo Bernardi

Coordinatore d'indirizzo: Ch.ma Prof.ssa Fiorella Lo Schiavo

Supervisore: Ch.mo Prof. Stefano Mammi

Dottorando: Carlo Baggio

*The only source of knowledge
is experience*

Albert Einstein

to my Family

Contents

ii

ABSTRACT

Bcl-X_L (B-cell-lymphoma-extra large) and the cell surface receptor CD44 are two proteins involved in cancer. Specifically, the anti-apoptotic protein Bcl-X_L, a member of the Bcl-2 family, plays a key role in the maintenance of normal cellular homeostasis. However, its overexpression can lead to oncogenic transformation and it is responsible for drug resistance in certain types of cancer. Structurally, the Bcl-2 family is characterized by the presence of BH domains, which are involved in pivotal protein-protein interactions. The anti-apoptotic protein Bcl-X_L interacts through its hydrophobic pocket with the BH3 domain of the pro-apoptotic members of the Bcl-2 family, such as Bak and Bax. The inhibition of this interaction eventually promotes cell death. The cell surface receptor CD44, a type I transmembrane glycoprotein, is a hyaluronan-binding protein expressed on the surface of many cell types, where it is involved in leukocyte migration to inflamed sites, T-cell activation, and tumor metastasis. Critical for these processes are the interactions between CD44 and its natural ligand, Hyaluronic Acid (HA). The involvement of CD44 in cell migration and its overexpression on a wide spectrum of tumor cells, make this receptor a good target for drug design of new inhibitors and for delivery of chemotherapeutics to cancer cells.

In the first part of thesis, with the aim of identifying an inhibitor of Bcl-X_L, we tested a small library of fragments against Bcl-X_L using solution Nuclear Magnetic Resonance (NMR) spectroscopy. Among the compounds tested, one was found to be active with a low millimolar K_d. Starting from this initial hit, we performed structure-activity relationship (SAR) studies of several analogs using both protein-based NMR experiments (mainly 2D [^{1}H,^{15}N]-HSQC) and ligand-based NMR experiments (Saturation Transfer Difference (STD) and waterLOGSY). Our studies led to a hit

fragment with a K_d of 811 μM, calculated using the chemical shift perturbations (CSP) from the 2D [^{1}H,^{15}N]-HSQC spectra. Molecular docking studies performed using the experimentally obtained CSP information allowed us to obtain a low energy conformation of the final fragment docked into the hydrophobic binding pocket of Bcl-X$_L$.

The second part of this thesis regards CD44, a protein that from structural data is not a particularly druggable target and only a few putative ligand agents are known to bind it in cellular assays. We first attempted to validate these putative ligand agents with biophysical methods, but surprisingly we found that, with the exception of HA and a commercially available antibody, none of these bound recombinant hCD44(21-178) appreciably. In the pursuit of possible novel CD44 antagonists, we performed a fragment screening campaign on ^{15}N-hCD44(21-178) using 1D 1H-aliphatic and 2D [^{1}H,^{15}N]-sofastHMQC NMR experiments. We found two initial structurally related hits and starting from these, further studies were performed using commercially available analogues. These experiments resulted in a fragment hit with a dissociation constant calculated using the chemical shift perturbation from the 2D [^{1}H,^{15}N]-sofastHMQC spectra of 7.43 mM.

In conclusion Bcl-X$_L$ and CD44 are both overexpressed in many types of cancer cells. Hence, finding ligand agents able to inhibit these two proteins is very a appealing task for the development of new chemotherapeutics. Using a fragment-based by NMR approach, we found two hits for each target, which could be at the basis for further fragment evolution studies for the development of more potent lead compounds.

RIASSUNTO

Bcl-X$_L$ (B-cell-lymphoma-extra large) e il recettore transmembrana CD44, sono due proteine coinvolte nel cancro. In particolare, la proteina anti-apoptotica Bcl-X$_L$, membro della famiglia di Bcl-2, gioca un ruolo chiave nel mantenimento della normale omeostasi cellulare. Tuttavia, la sua sovraespressione può portare alla trasformazione oncogenica ed è responsabile per la resistenza ai farmaci in alcuni tipi di cancro. Strutturalmente, questa famiglia è caratterizzata dalla presenza di domini BH, che sono coinvolti in fondamentali interazioni proteina-proteina. La proteina anti-apoptotica Bcl-X$_L$ interagisce attraverso la sua tasca idrofobica con il dominio BH3 dei membri pro-apoptotici della famiglia Bcl-2, come BAK e BAX. L'inibizione di questa interazione favorisce infine la morte cellulare. Il recettore di superficie cellulare CD44, è una glicoproteina di transmembrana di tipo I, ed è una proteina che lega l'acido ialuronico espresso sulla superficie di molti tipi di cellule, in cui è coinvolta nella migrazione dei leucociti ai siti di infiammazione, nell'attivazione delle cellule T e nelle metastasi tumorali. Critico per questi processi sono le interazioni tra CD44 e il suo ligando naturale, l'acido ialuronico (HA). Il coinvolgimento di CD44 nella migrazione cellulare e la sua sovraespressione su un ampio spettro di cellule tumorali, rendono questo recettore un buon bersaglio per la progettazione di nuovi farmaci e per il *drug delivery* di chemioterapici in cellule tumorali.

Nella prima parte della tesi, con l'obiettivo di individuare un inibitore di Bcl-X$_L$, abbiamo testato una piccola libreria di frammenti contro Bcl-X$_L$ utilizzando come tecnica la spettroscopia di Risonanza Magnetica Nucleare (NMR) in soluzione. Tra questi uno è risultato essere attivo con una Kd nel basso millimolare. Partendo da questo hit iniziale, abbiamo effettuato degli studi di relazione struttura-attività (SAR) di diversi analoghi, utilizzando sia esperimenti NMR che guardano alla proteina

(soprattutto 2D [1H, 15N] -HSQC) ed esperimenti NMR concentrati sul ligando (Saturation Transfer Difference (STD) e waterLOGSY). I nostri studi hanno portato ad un frammento con un K_d di 811 μM calcolati utilizzando le perturbazioni di chemical shift (CSP) dagli spettri 2D [1H, 15N]-HSQC. Dagli studi di docking molecolare condotti utilizzando le informazioni di CSP sperimentalmente ottenute, ci hanno permesso di ottenere una conformazione a bassa energia del frammento finale nella tasca di legame idrofobica di Bcl-X_L.

La seconda parte di questa tesi riguarda CD44, una proteina che dai dati strutturali non è un target particolarmente '*druggable*' e solo pochi agenti leganti sono noti per legarla in saggi cellulari. Per prima cosa abbiamo cercato di convalidare tali agenti leganti noti con metodi biofisici, ma sorprendentemente abbiamo trovato che, con l'eccezione di HA e un anticorpo disponibile in commercio, nessuno di questi lega apprezzabilmente la proteina ricombinante hCD44(21-178). Nel perseguire dei possibili nuovi antagonisti di CD44 abbiamo effettuato una campagna di screening di frammenti sulla proteina 15N-hCD44(21- 178) utilizzando esperimenti NMR 1D 1H-alifatici e 2D [1H, 15N]-sofastHMQC come metodi di rilevamento. Abbiamo trovato due hit iniziali strutturalmente affini e partendo da questi abbiamo effettuato ulteriori studi utilizzando analoghi disponibili in commercio. Questi esperimenti hanno portato in un frammento con una costante di dissociazione calcolata con la perturbazione di chemical shift dagli esperimenti 2D [1H, 15N]-sofastHMQC di 7,43 mM.

In conclusione Bcl-X_L e CD44 sono entrambe sovraespresse in molti tipi di cellule tumorali. Per questo motivo trovare agenti leganti capaci di inibire queste due proteine è un compito molto attraente per lo sviluppo di nuovi chemioterapici. Utilizzando un approccio NMR basato sui frammenti, abbiamo trovato per entrambi i

target due hit che potrebbero essere la base per ulteriori studi di evoluzione del frammento per lo sviluppo di composti *lead* più potenti.

Chapter 1

Introduction

B-cell-lymphoma-extra large (Bcl-X$_L$) is a transmembrane protein present on the mitochondria. In humans is encoded by the BCL2L1 gene that encodes a 233 amino acid protein (**Figure 1**). It is one of the many anti-apoptotic proteins belonging to the Bcl-2 family, and it is implicated on the survival of tumor cells. The homology with the other members of the Bcl-2 family is guaranteed by the presence of four Bcl-2 Homologue (BH) domains that are fundamental for protein-protein interactions between Bcl-2 family members.

```
         10         20         30         40         50
MSQSNRELVV DFLSYKLSQK GYSWSQFSDV EENRTEAPEG TESEMETPSA
         60         70         80         90        100
INGNPSWHLA DSPAVNGATG HSSSLDAREV IPMAAVKQAL REAGDEFELR
        110        120        130        140        150
YRRAFSDLTS QLHITPGTAY QSFEQVVNEL FRDGVNWGRI VAFFSFGGAL
        160        170        180        190        200
CVESVDKEMQ VLVSRIAAWM ATYLNDHLEP WIQENGGWDT FVELYGNNAA
        210        220        230
AESRKGQERF NRWFLTGMTV AGVVLLGSLF SRK
```

Figure 1 Amino acid sequence of human Bcl-X$_L$

Bcl-2 family

The peculiarity of this family resides in the high structural homology guaranteed by the presence of particular domains called BH domains[1]. Four BH domains have been identified, commonly called BH1, BH2, BH3, and BH4. A common classification of the Bcl-2 family members could be done based on the ability to induce or prevent apoptosis, and using this criterion, the proteins are divided into pro-apoptotic and anti-apoptotic proteins, respectively, as reported in **Figure 2**[2].
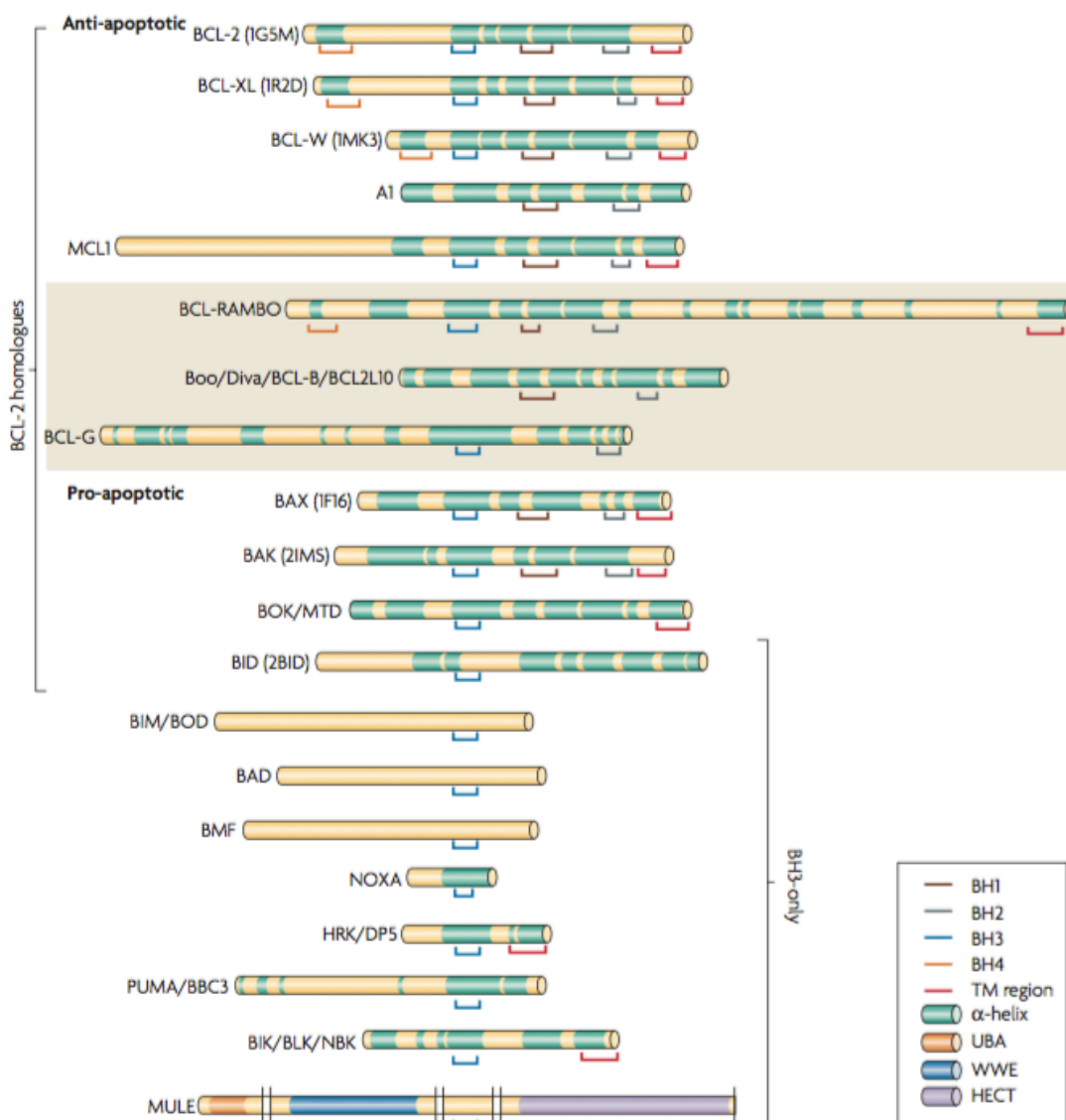

Figure 2 Classification in pro-apoptotic and anti-apoptotic subfamilies of the most important members of the Bcl-2 protein family. Green bars indicate the segments with α-helix structure. The red lines label regions of transmembrane domain (TM). The homology regions are also shown: BH1 (brown lines), BH2 (gray lines), BH3 (blue lines), and BH4 (orange lines). The BH1, BH2, and BH3 domains are folded to form a hydrophobic pocket that can bind BH3-only

peptides. The first five proteins are normally anti-apoptotic. The three proteins in the highlighted area have not been studied much and cannot yet be categorized. The last 12 proteins are considered pro-apoptotic [2].

The anti-apoptotic proteins are multi-BH, and usually possess all four BH domains, as in the case of Bcl-X$_L$. The pro-apoptotic proteins could be in turn subdivided into multi-BH, and BH3-only proteins, depending on the presence of all the BH domains or only BH3. Pro-apoptotic family members that contain BH1 and BH2 regions, such as Bax, can promote apoptosis through their interactions with mitochondrial membranes. This activity is independent of their ability to interact with anti-apoptotic proteins. The BH3 region is responsible for mediating the interactions with anti-apoptotic proteins and for the ability of the proteins to promote programmed cell death[3].

Since its discovery, the role of Bcl-2 was associated with apoptosis. Actually, it is known that the Bcl-2 family proteins regulate all major types of cell death, including apoptosis, necrosis, and autophagy [4].

The mechanisms that regulate tissue homeostasis are governed mainly, but not exclusively, by the proteins of the Bcl-2 family. The daily programmed cell death, in the majority of tissues, involves the mitochondria, organelles that, in addition to producing energy, play a critical role in regulating the balance between life and cell death[5]. The pro-apoptotic proteins of the Bcl-2 family, such as Bax and Bak, induce mitochondrial outer membrane permeabilization, causing the release of caspase-activating proteins and other mediators of cell death. The anti-apoptotic proteins, such as Bcl-2, Mcl-1 and Bcl-X$_L$, act as 'guardians' of the outer membrane and preserve its integrity by binding to Bax and Bak and preventing them from carrying out their function[6]. Many theories have been reported on the molecular mechanism

that leads to apoptosis, but the exact molecular process is not fully clear yet. The key point is, however, the interaction between the different members of the family to promote or inhibit apoptosis, highlighting one of the most distinctive features of this family: the ability to form specific protein-protein interactions.

Structure of Bcl-X$_L$

Bcl-X$_L$ is a 233 residues protein and, as shown in **Figure 3,** consists of two central α-helices (α5 and α6), which are about 30 Å long and are flanked on one side by helices α3 and α4 and on the other side by α1, α2 and α7. The two central helices contain predominantly hydrophobic residues and are arranged in an antiparallel way. The presence of a proline in position 180 inside the α6 helix causes a change in the direction of this helix. The C-terminal helix (α7) is connected to α6 through an irregular turn composed of two glycines (G186 and G187) that are highly conserved[7]. The N-terminal helix forms numerous hydrophobic interactions with helices α2, α5 and α6 that seem to be important for structural stability [7]. Some proteins of the Bcl-2 family share homologous BH1 and BH2 domains and mutations of Bcl-2 or Bcl-X$_L$ in these regions abolish the anti-apoptotic activity and block the hetero-dimerization with other members of the Bcl-2 family (Bax or Bak) that promote programmed cell death. A third region characterized by structural homology between the proteins of the family is the BH3 domain, recognized to be essential for the activity of proteins that promote cell death. These regions with an important function (BH1, BH2 and BH3) are in close spatial proximity, and form an elongated hydrophobic pocket in Bcl-X$_L$, which is the site of interaction with proteins promoting cell death. Helix α1 is connected to α2 through a flexible loop of approximately 60 residues not essential for the anti-apoptotic activity[7]. Like many other proteins of

the Bcl-2 family, Bcl-X$_L$ contains a C-terminal transmembrane domain, the function of which is to anchor Bcl-X$_L$ to various intracellular membranes, in particular the outer mitochondrial membrane [7].
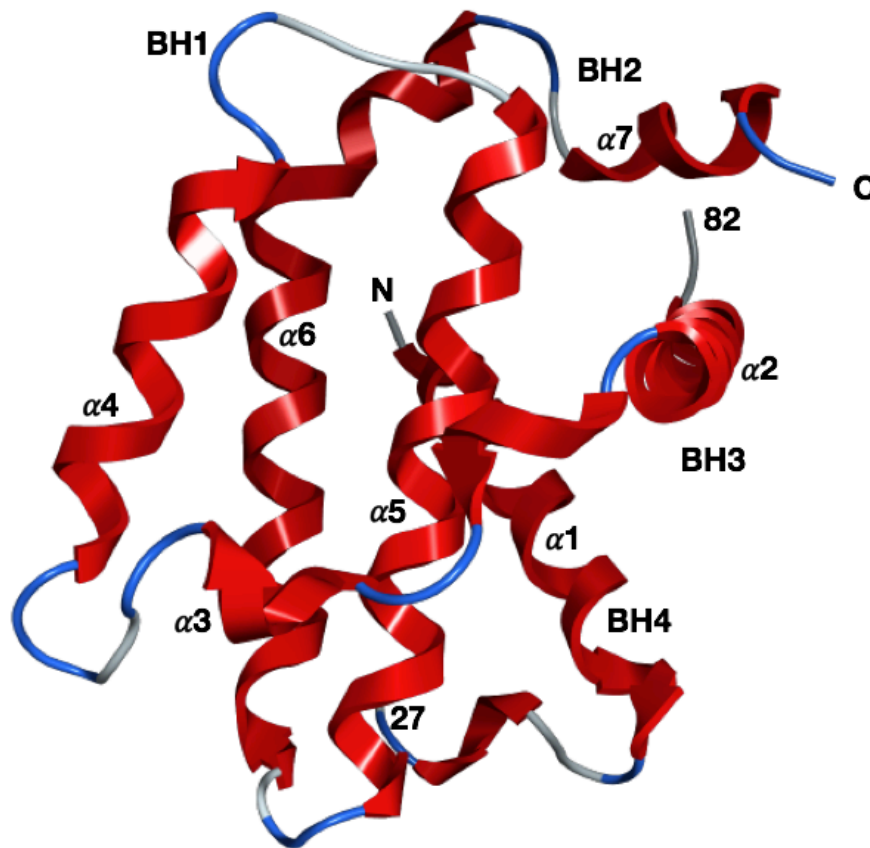

Figure 3 Crystallographic structure of Bcl-X$_L$ (PDB ID 2BZW), shown with the secondary structure (α-helices in red, and β-turn in blue)

Protein-protein interactions

The hydrophobic pocket of Bcl-X$_L$ is able to bind the BH3 domain consisting of a α-helix. The binding interface of BH3 peptides is based on the conservation of some hydrophobic residues (specifically, a Phe and a Leu) that are placed into two cavities in the hydrophobic site of Bcl-X$_L$. In addition, an Asp residue, conserved in all BH3 domains, establishes a polar bond with an Arg present in α5 (**Figure 4**).
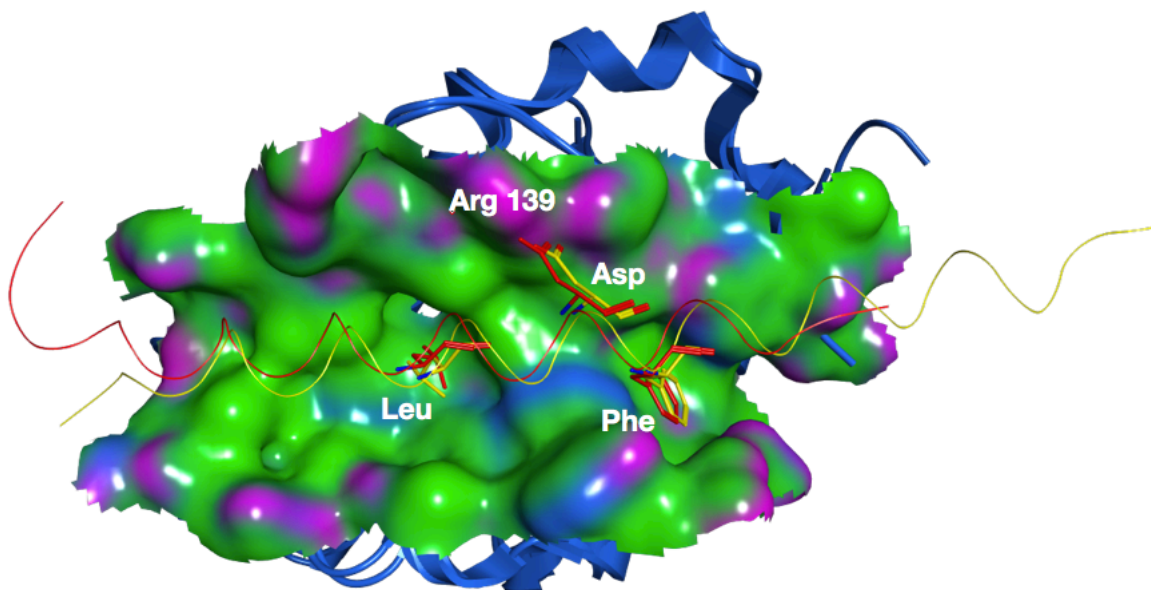
Figure 4 Superposition of two different BH3 domains arising from BIM and BAD, co-crystallized with Bcl-X$_L$ (PDB ID 1PQ1 and 2BZW of BIM and BAD, respectively). The surface of the hydrophobic pocket of Bcl-X$_L$ is color-coded according to the characteristics of lipophilicity (Green: hydrophobic; Blue: mildly polar; Purple: H-bonding area). The backbone of the BH3 peptides and the residues important for interaction are colored in yellow and red (BIM and BAD, respectively).

Physiological role of Bcl-X$_L$

The most important role of Bcl-X$_L$ is related to its ability to promote survival in various cell lines involved in the apoptotic process. The exact molecular mechanism by which Bcl-X$_L$ promotes cell survival is not fully understood, but it is believed to involve the suppression of the release of cytochrome c from mitochondria, possibly through heterodimerization with the neutralization of pro-apoptotic proteins of the Bcl-2 family. The main model to explain the regulation of apoptosis through Bcl-X$_L$ is the model of movement (or indirect activation): the main feature is that the multi-BH pro-apoptotic proteins Bax and Bak are constitutively active and must be continuously bound and inhibited by Bcl-X$_L$ for cell survival[8]. When the apoptotic signal is received, the BH3-only proteins (Bim, tBID, Bik, PUMA, and NOXA) competitively bind to the hydrophobic site of Bcl-X$_L$ displacing Bax and Bak[9]. The

released monomers of Bax and Bak may then interact with each other through oligomerization and form pores that allow the permeabilization of the outer mitochondrial membrane. At this point, cytochrome c is released from mitochondria and activates a cascade of caspases allowing the cleavage of specific cellular proteins, eventually leading to cell death[10]. In this model, it is clear that Bcl-X$_L$ plays an anti-apoptotic role by sequestering the pro-apoptotic proteins in the outer mitochondrial membrane, thus preventing the oligomerization of Bak/Bax (**Figure 5**).
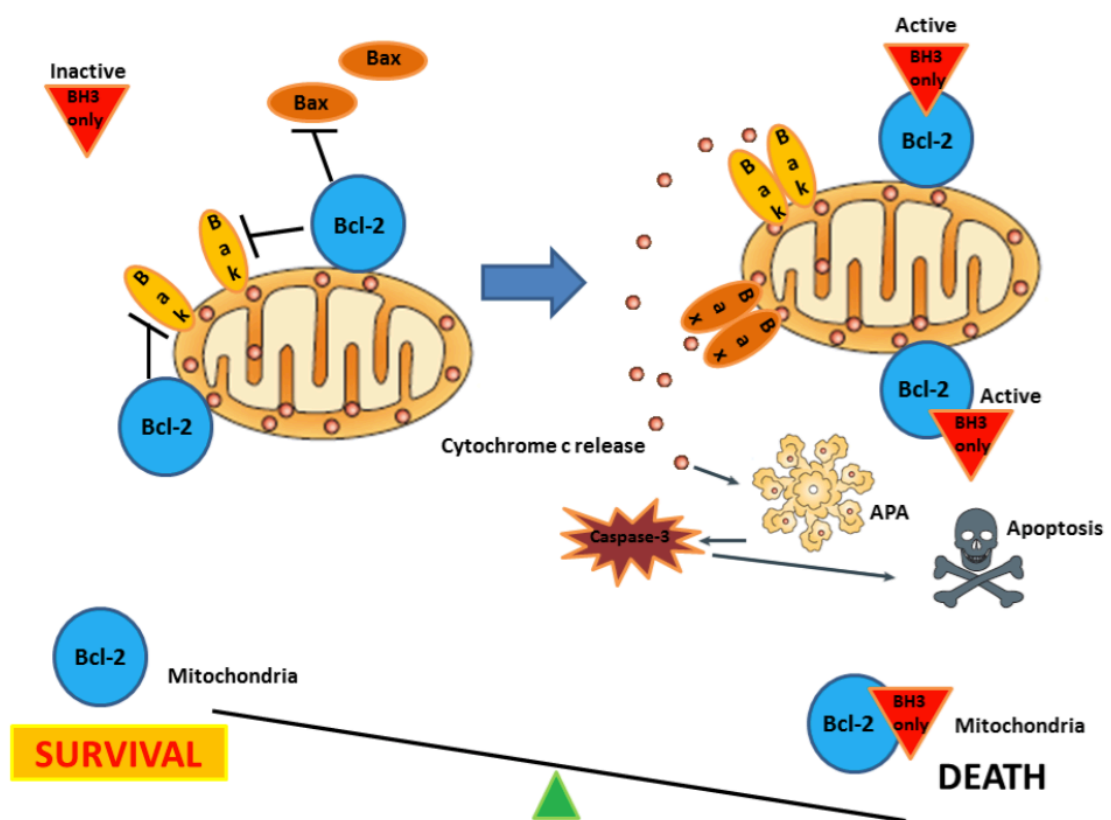

Figure 5 General apoptosis-regulating mechanism mediated by the Bcl-2 family members to which also Bcl-X$_L$ belongs. Upon interactions between proteins of the family, pro-apoptotic factors such as cytochrome c are released that contribute to the formation of the apoptosome and the subsequent activation of caspases.

Diseases related to Bcl-X$_L$

In physiological conditions, apoptosis ensures the homeostasis of tissues by maintaining the number of cells constant. On the other hand, dysregulation of apoptosis produces wide-ranging effects in many patho-physiologies. An excess of apoptosis is implicated in several neurodegenerative disorders such as Alzheimer's disease and multiple sclerosis while the lack of apoptosis is a central step in oncogenesis and many inflammatory conditions[11]. Overexpression of Bcl-X$_L$ is related to several diseases, including particularly aggressive cancers.

Fragment-Based Drug Design of Bcl-X$_L$ inhibitors

The Fragment-Based Drug Design (FBDD) approach has been developed mainly in the last 10 years and is now recognized as a particularly effective methodology in lead discovery. The method was proposed by Abbot in 1996 and has gained increasing interest over the years to become a real alternative to high-throughput screening. Strategies based on FBDD are now routinely used in many pharmaceutical companies and have already yielded promising results. The method consists in the screening of a small set of compounds of low molecular weight (<300Da), called fragments. Because of their low complexity, the fragments normally show low affinity for the target protein, but optimization of the fragments identified, normally through the addition of new chemical moieties or alternatively by connecting two fragments that bind to adjacent sites, can lead to very interesting compounds, as demonstrated in the literature[12] (**Figure 6**).

This method of 'lead-conception' relies on the premise that the binding site and the way of binding of the fragments should be retained in case of fragment modification,

and when two fragments are linked together, their orientation in their respective binding sites should remain the same[12].
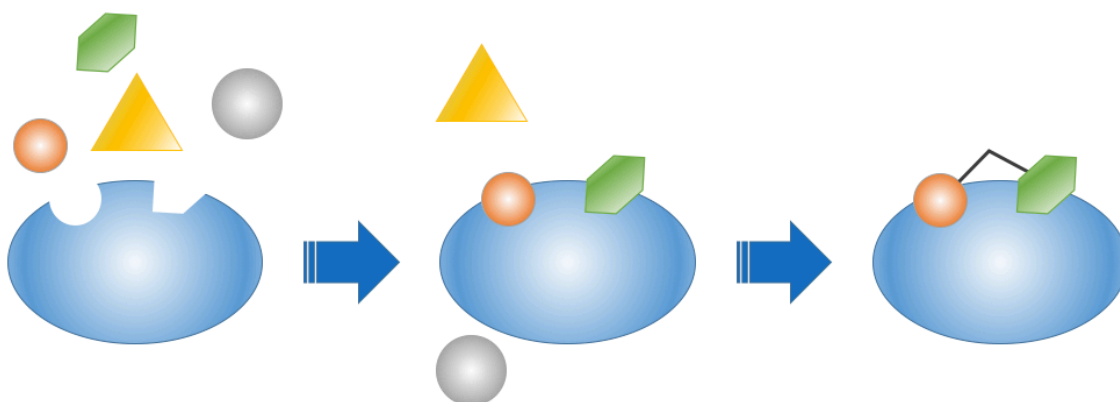

Figure 6 In the Fragment-Based Drug Design, a screening of a library of fragments is done (eg. through NMR) to find those that bind into sub-pocket of the binding surface of the target protein. The fragments are then linked together to make a larger molecule that binds better to the binding pocket of the target protein.

The screening techniques most commonly used in the literature are the solution Nuclear Magnetic Resonance (NMR), the X-ray crystallography, the Surface Plasmon Resonance (SPR), and some biochemical assays. The NMR is the technique of choice in screening of fragments because it is a technique particularly sensitive to intermolecular interactions of low affinity, typical for small molecules (affinity constants in the order of low millimolar). Furthermore, compared with other techniques, thorough the NMR can be derived structural data, and is also very robust, returning a low rate of false-positives[13] [14].

ABT-737 (**Figure 7A**) is a small molecule developed as a binder of some anti-apoptotic proteins of the Bcl-2 family, including Bcl-X_L (**Figure 7B**), which, by upon binding to it, promotes the oligomerization of Bax and Bak causing programmed cell death of malignant cells. ABT-737 was developed by Abbott Laboratories using a screening of fragments by Nuclear Magnetic Resonance, and is the result of the

binding of two different small molecules of weak affinity to give a single molecule of high affinity[15]. However ABT-737 is not suitable as a drug for human treatment because of its toxicity in humans, but it has opened a process field for the development of new molecules derived from it, still in experimentation.
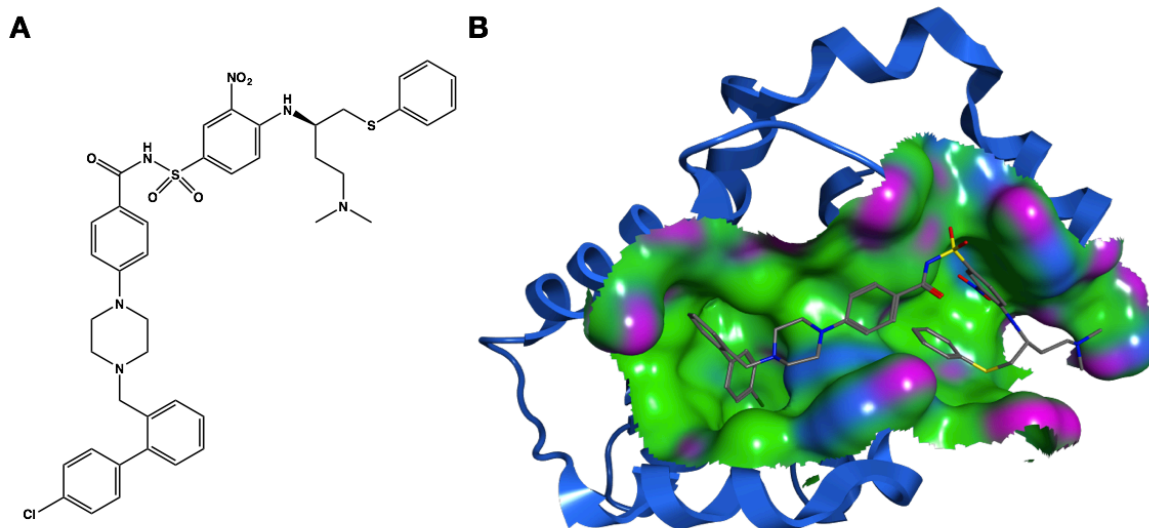

Figure 7 (A) Chemical structure of ABT-737; (B) Bcl-X$_L$ co-crystallized with ABT-737 (PDB ID 2YXJ). The surface of Bcl-X$_L$ is color-coded according to the characteristics of lipophilicity (Green: hydrophobic; Blue: mildly polar; Purple: H-bonding area).

Methods and Materials

Protein expression and purification

The use of NMR spectroscopy requires a substantial amount of very pure protein. For this reason, the protein was produced through biotechnological techniques using an *E. coli* expression system. The cDNA of human Bcl-X$_L$ ΔTM was cloned into an expression plasmid pET15b and introduced into *E. coli* BL21 (DE3) strain. The expression product contained an N-terminal His-tag, a sequence of cleavage for thrombin, and the sequence coding for Bcl-X$_L$, except for the last 24 amino acids, which constitute the transmembrane domain. The overexpression of the protein was obtained in *E. coli* BL21 (DE3), by growing the bacteria in LB medium at 37 °C in the presence of the antibiotic ampicillin until an OD_{600} of 0.6 was reached, followed by induction with 0.6 mM isopropyl β-D-thiogalactopyranoside (IPTG). After 4 hours, the cells were harvested by centrifugation, resuspended in 25 mM Tris-HCl buffer pH = 8, and sonicated. After a further centrifugation step, the soluble fraction, containing Bcl-X$_L$, was purified through a nickel affinity column (His-Trap) using an FPLC system. To obtain the monomer fraction with greater purity, a further step of purification was carried out by means of size exclusion chromatography (gel filtration) using a HiLoad 16/60 Superdex 75 column in 20 mM Tris-HCl buffer at pH 7.4, 1 mM DTT, and 150 mM NaCl. For the expression of the ^{15}N-labeled protein used in the 2D-NMR experiments, the procedure was the same, but the bacteria were grown in M9 minimal medium supplemented with 1 g/L ^{15}NH$_4$Cl for labeling purposes. The pure protein was stored at a concentration of 40 μM at -80 °C. The concentration of the

protein was estimated through UV spectroscopy measuring the absorbance at 280 nm, using the molar extinction coefficient $\varepsilon = 47440$ $M^{-1}cm^{-1}$.

Nuclear Magnetic Resonance (NMR)

Nuclear magnetic resonance (NMR) is one of the most powerful analytical techniques present in chemistry, and together with X-ray crystallography, is the technique of choice for protein structure determination at high resolution. The advantages of NMR are the possibility to study macromolecules in solution (conditions closer to physiological ones) and the possibility to study the dynamic behavior of the protein. Over 9000 NMR structures are deposited in the Protein Data Bank (PDB). Despite these advantages, the NMR can be applied only to proteins of a certain number of amino acids; proteins above 40 kDa are difficult to study through classic strategies, because of their relaxation properties.

One-dimensional ^{1}H -NMR experiment

The nucleus most widely investigated by NMR is ^{1}H, as a result of several factors: ^{1}H is abundant in nature, the hydrogen forms a wide range of compounds with a large number of elements, it is a fermion (non zero spin), and has a high gyromagnetic ratio, which is related to the intensity of the signal. Experiments based on proton allow the study of small molecules and small peptides. In the study of proteins, homonuclear experiments based on hydrogen are useful as fingerprints to identify a protein or to check for chemical shift perturbation in the more shielded area of the spectrum.

Two-dimensional spectroscopy

Peptides are characterized by the presence of many protons in their structure and it is easy to understand that even short sequences generate a certain number of peaks that may lie in the same region of the spectrum. As the size of the peptide increases, more and more resonances become degenerate, and it is impossible to assign the signals using one-dimensional techniques. In a two-dimensional NMR experiment, the signal is registered as a function of two time variables, t_1 and t_2, and the resulting data are subjected to two Fourier transforms to obtain the spectrum as a function of two frequency variables. In **Figure 8**, the general scheme for two-dimensional NMR spectroscopy is shown.
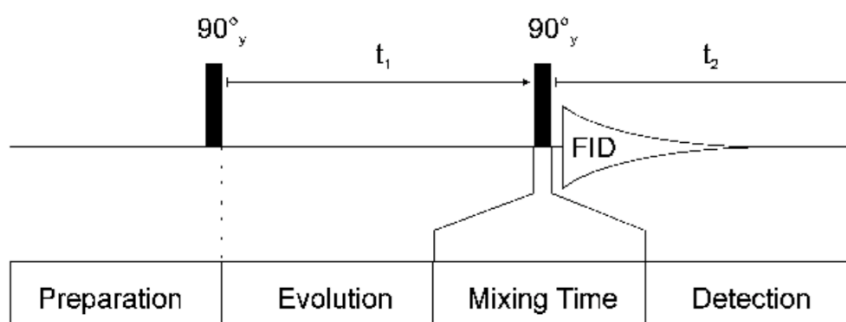
Figure 8 General pulse sequence of a 2D NMR experiment

In the first part, called preparation time, the sample is excited by one or more pulses. The net magnetization is left to evolve during the evolution period, t_1. Then, the mixing time follows, consisting in a further series of pulses, the purpose of which is to transfer the magnetization between different nuclei. After the mixing time, the signal is recorded as a function of the second time variable (t_2). This sequence of events is called pulse sequence and the exact nature of the preparation and mixing periods determines the information obtainable in the NMR spectrum[16].

Heteronuclear experiments

As it is possible to detect correlations through J-coupled proton spins, it is also possible to observe correlations between protons and different nuclei with non-zero spin, such as nitrogen and carbon, to which the protons themselves are coupled (heteronuclear J-correlated spectroscopy). In the case of NMR of proteins, the isotopes ^{13}C and ^{15}N are particularly important, as they have spin 1/2. In a heteronuclear NMR spectrum, along one axis are the proton chemical shifts, and along the other axis the chemical shifts of the heteronucleus. Each peak identifies the scalar correlation between a particular proton-heteronucleus pair. The wide dispersion of heteronuclear chemical shifts allows the overlaps between different peaks to be reduced, and this is particularly important in the study of proteins that have a large number of signals. Because the proteins are mainly composed of C, N, H, and O, the importance of techniques able to give information between nuclei different from proton is evident. In **Table 1**, it can be seen that only isotopes ^{13}C and ^{15}N have spin 1/2, and so they are ideal for study through NMR. However, these isotopes have low natural abundance and need to be artificially introduced into the macromolecule to be studied by NMR.

Table 1 NMR properties of some nuclei

nucleus	I	γ $(10^7$ rad s^{-1} T$^{-1})$	Q $(10^{-28}$ m$^2)$	Obs. freq. at 9.4 T (MHz)	Nat. abund. %	Abs. sensitivity rel. to ^{1}H
^{1}H	1/2	26.7522	-	400.00	99.98	1.00
^{2}H	1	4.1066	2.87×10^{-3}	61.402	0.015	1.45×10^{-6}
^{6}Li	1	3.9371	-6.4×10^{-4}	58.862	7.42	6.31×10^{-4}
^{7}Li	3/2	10.3976	-0.037	155.454	92.58	0.27
^{11}B	3/2	8.5847	0.041	128.355	80.42	0.13
^{13}C	1/2	6.7283	-	100.577	1.108	1.76×10^{-4}
^{14}N	1	1.9338	0.0167	28.894	99.63	1.01×10^{-3}
^{15}N	1/2	-2.7126	-	40.531	0.37	3.85×10^{-6}
^{17}O	5/2	-3.6280	-0.026	54.227	0.037	1.08×10^{-5}
^{19}F	1/2	25.1815	-	376.308	100	0.83
^{23}Na	3/2	7.0704	0.10	105.805	100	0.0925
^{27}Al	5/2	6.9762	0.14	104.229	100	0.21
^{29}Si	1/2	-5.3190	-	79.460	4.7	3.69×10^{-4}
^{31}P	1/2	10.8394	-	161.923	100	0.063
^{33}S	3/2	2.0557	-0.064	30.678	0.76	1.72×10^{-5}
^{51}V	7/2	7.0492	-0.052	105.152	99.76	0.38
^{59}Co	7/2	6.3015	0.42	94.457	100	0.28
^{77}Se	1/2	5.1214	-	76.270	7.58	5.25×10^{-4}
^{197}Au	3/2	0.4692	0.55	6.850	100	2.51×10^{-5}

Nowadays, the expression of isotopically labeled proteins is a common strategy carried out in many laboratories. The NMR samples require a large amount of protein to reach the ideal concentration for this technique. The easiest and cheapest strategy to obtain them is the expression and subsequent purification using bacterial vectors, such as *E. coli*. With this approach, it is possible in the first place to produce large amounts of protein, and also to express proteins enriched with ^{13}C and ^{15}N using specific media containing ^{15}NH$_4$Cl as sole nitrogen source, and ^{13}C-glucose as carbon source. One of the main problems of heteronuclear NMR is the relative sensitivity of ^{13}C and ^{15}N nuclei, such that the direct detection of these nuclei can be problematic. The population difference (at Boltzmann equilibrium) is proportional to the gyromagnetic ratio γ. The relative sensitivity of nucleus X with respect to the proton is defined as:

$$\rho_i = \frac{S_x}{S_{^1H}} = \frac{N_x \gamma_x^3 (I + 1)}{N_{^1H} \gamma_{^1H}^3 \left(\frac{1}{2} + 1\right)}$$

Where S_X and $S_{^1H}$ are the signal to noise ratio of the heteronucleus X, and ^1H, respectively, and N is the natural abundance of the nuclei. Neglecting the term due to the natural abundance (*e.g.*, proteins uniformly enriched with ^{13}C and ^{15}N), for nuclei with spin 1/2 the sensitivity relative to ^1H is given by:

$$\rho_i = \left(\frac{\gamma_x}{\gamma_{^1H}}\right)^3$$

For ^{13}C and ^{15}N, the relative sensitivity are 1.6×10^{-2}, and 1.0×10^{-3} respectively. Nowadays, the strategy used to resolve the problem of sensitivity in heteronuclear experiments is to excite and observe always the magnetization of the most sensitive nucleus (^1H), and to transfer the same magnetization to the heteronucleus and back during the pulse sequence to detect its resonance frequency indirectly.

Insensitive Nuclei Enhanced by Polarization Transfer: INEPT

The INEPT (Insensitive Nuclei Enhanced by Polarization Transfer)[17] sequence is the basic building block of many heteronuclear NMR experiments correlating protons and heteronuclei that are coupled together. Basically, the INEPT consist to transfer the magnetization from a sensitive nucleus with a high gyromagnetic ratio (such as proton) to less sensitive nuclei with a low gyromagnetic ratio (such as nitrogen and carbon), through the scalar coupling interaction.

In **Figure 9[18]** the vector representation of the magnetization during the INEPT pulse sequence is shown. The first 90° ^1H pulse rotates the ^1H magnetization onto the y-axis at point a. After the first $1/(4J_{CH})$ delay (here carbon is taken as an example, but generally is $1/(4J_{XH})$, where X is the less sensitive heteronucleus), each of the

doublets caused by the J_{CH} coupling is 45° away from the y-axis. The simultaneous 180° pulses on both ^{1}H and ^{13}C refocus the chemical shift but allow the coupled ^{1}H vectors to continue to diverge. This can be understood by considering the effect of the ^{1}H and ^{13}C 180° pulses individually. After the ^{1}H 180° pulse flips the doublets about the x-axis, the ^{13}C 180° inverts the population of ^{13}C states of the coupled doublets but has no effect on uncoupled chemical shift. Consequently, the magnetization in the ^{13}C α state becomes that in the β state, and vice versa. In the vector diagram at point c, it is shown that the slower and faster vectors exchange places, resulting in the vectors rotating in the opposite direction (c). After the second $1/(4J_{CH})$ delay, the two vectors are 180° out of phase and aligned on the x-axis (d). The following ^{1}H 90° pulse rotates the ^{1}H magnetization components back to the z and –z axes (e), and the ^{13}C 90° pulse brings the coupled magnetization components back to the xy plane, which results in detectable anti-phase magnetization (f)[18].
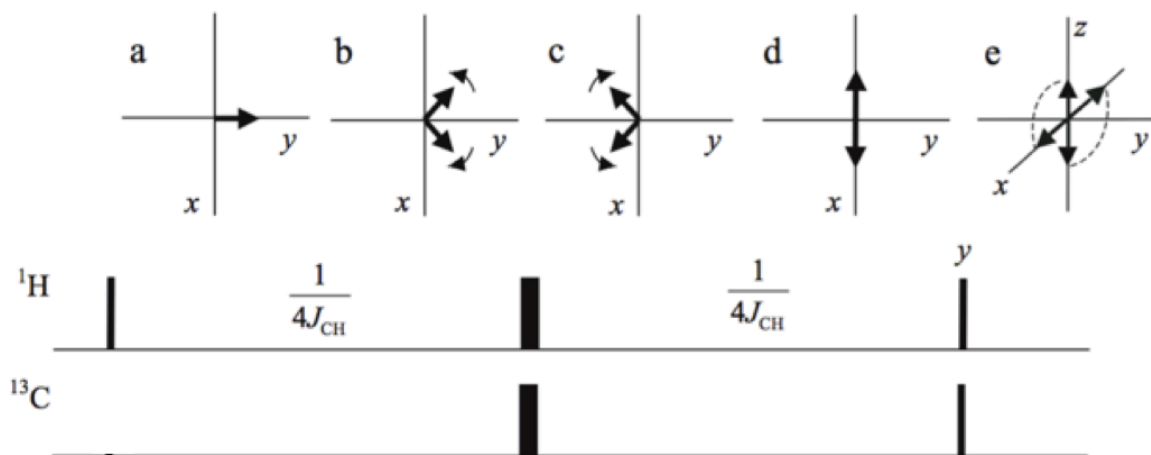
Figure 9 Vector representation of an INEPT experiment. The narrow and wide bars are 90° and 180° pulses, respectively. All pulses are x phase except the last ^{1}H 90° y.

HSQC experiment (Heteronuclear Single-Quantum Coherence)

The HSQC experiment[19] [20] [21] is the most used experiment in protein-NMR. It is frequently used to characterize protein-protein or ligand-protein interactions.

To overcome the problem of sensitivity, the HSQC experiment usually includes two INEPT blocks; the first one transfers the magnetization excited by the first 90° pulse from the proton to the heteronucleus while the second block follows the opposite path to transfer the magnetization back to the proton. In **Figure 10**, the pulse sequence of a basic HSQC experiment, that consists of five main blocks, is shown. At the end of the 2τ interval of the first INEPT, the proton antiphase magnetization is generated. In the middle of the 2τ period, a 180° pulse is applied on both spins to refocus the chemical shift evolution. In the second step, coherence is transferred from the proton to the bound heteronucleus (spin S) through two simultaneous 90° pulses. Then, the S-spin coherence is frequency-labeled during the period, t1. A 180° pulse on the proton in the middle of t1 refocuses the evolution of heteronuclear J_{HS} coupling. In the fourth step, a 90° pulse on both spins transfers the magnetization back to the proton as antiphase ^{1}H-spin magnetization. The final spin-echo period converts this antiphase term into in-phase proton magnetization, which is acquired during the period, t2[22].
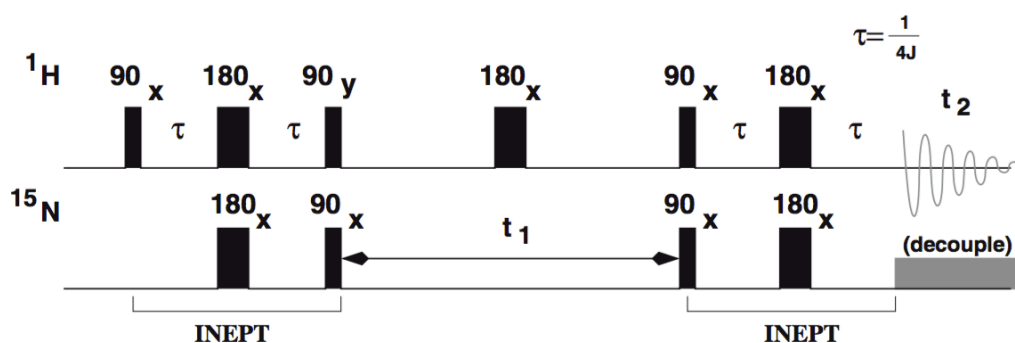
Figure 10 Pulse sequence of a general HSQC experiment

The [^{1}H,^{15}N]-HSQC experiment is the most standard experiment and shows all H-N correlations. These are mainly the backbone amide groups, but Trp side-chain Nε-Hε groups and Asn/Gln side-chain Nδ-Hδ2/Nε-Hε2 groups are also visible. The Arg Nε-Hε peaks are in principle also visible, but because the Nε chemical shift is outside the region usually recorded, the peaks are folded/aliased (this essentially means that they appear as negative peaks and the Nε chemical shift has to be specifically calculated). If working at low pH, the Arg Nη-Hη and Lys Nζ-Hζ groups can also be visible, but are also folded/aliased (**Figure 11**)[23]. The spectrum is rather like a fingerprint and is usually the first heteronuclear experiment performed on proteins.
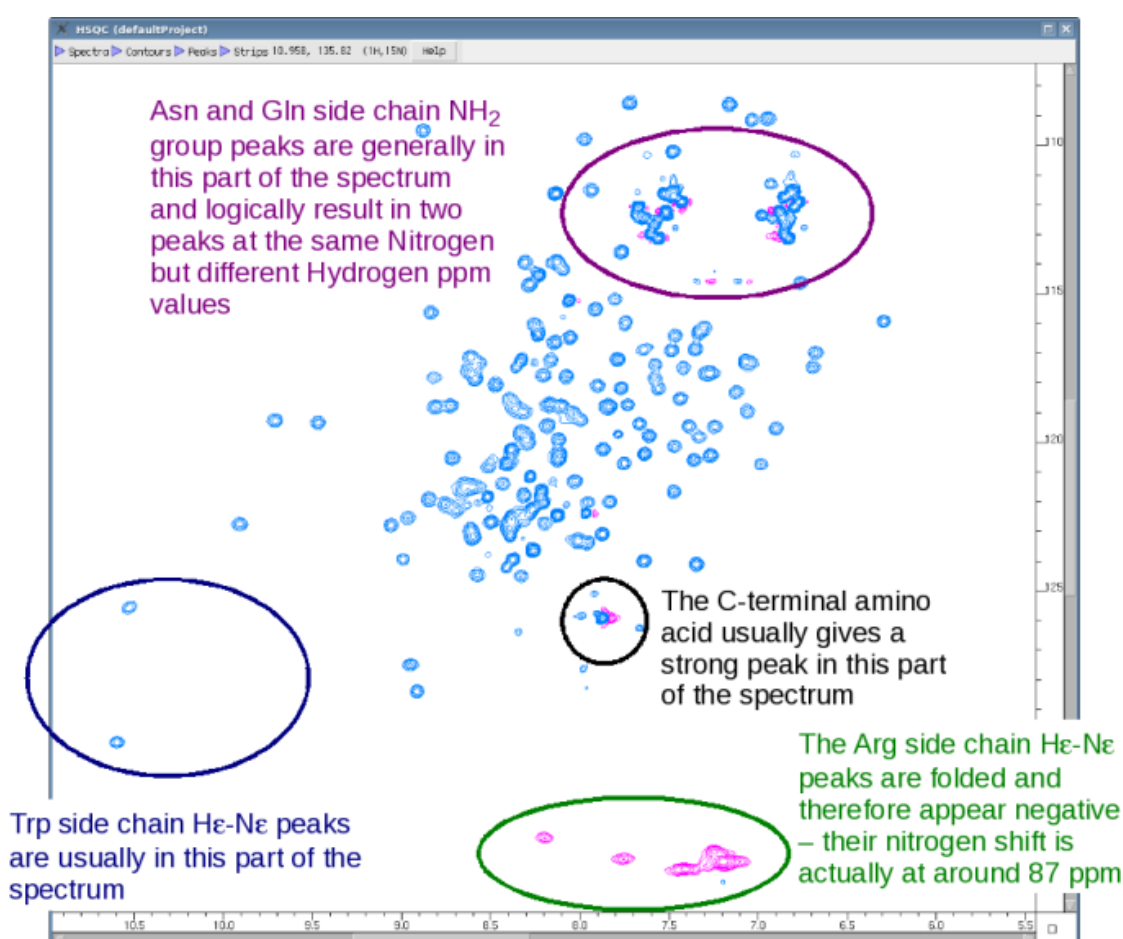
Figure 11 Example of a [^{1}H,^{15}N]-HSQC spectrum of a protein; the characteristic zones of the spectrum are indicated[23].

Nuclear Overhauser Effect (NOE)

While scalar coupling is mediated by electron density and provides information on the proximity between the atoms in the chemical structure, there is another type of interaction between nuclear spins, which depends on the local magnetic fields produced by the magnetic moments of the spins themselves. The effect of the dipolar interaction can be measured as a change in the populations of the energy levels of the coupled spin. This effect is known as NOE (Nuclear Overhauser Enhancement): given two dipolar coupled spins, I and S, if the equilibrium of S is perturbed, there is a change in the magnetization of I proportional to a cross-relaxation constant (σ_{IS}), and to the variation of S from its equilibrium:

$$\frac{d(I_Z - I_Z^0)}{dt} = -R_I(I_Z - I_Z^0) - \sigma_{IS}(S_Z - S_Z^0)$$

where I_Z and S_Z are the components of the magnetization along the external magnetic field and R_I is a self-relaxation constant. The efficacy of this polarization transfer between spins close in space (below 5-5.5Å) is proportional to $1/d^6$ (where d is the internuclear distance)[18] [22]. Referring to **Figure 12**, if a resonance A is irradiated, an increase (positive NOE) or decrease (negative NOE) of the signal intensity of other resonances is observed; for example, resonance C, when spin C is close in space to spin A[24].
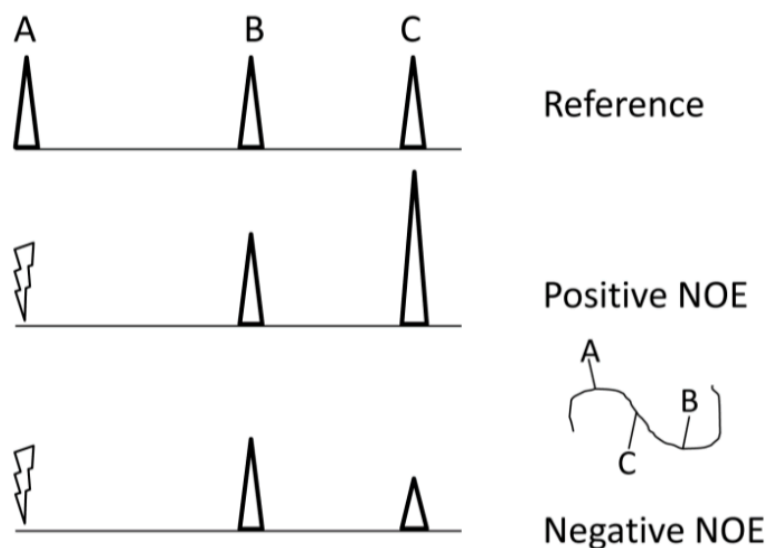
Figure 12 Irradiation of resonance A leads to a increase of peak intensity of the neighboring spin C (positive NOE) or to a decrease of peak intensity (negative NOE)[24].

STD experiment (Saturation-Transfer Difference)

The STD NMR experiment is one of the most popular ligand-based techniques for the study of ligand-protein interactions by NMR. The success of this technique is a consequence of the fact that it is focused on the signal of the ligand, and can be used even with very large proteins and even if the assignment of the resonances of the protein is not available. It also allows the use of small amounts of non-labeled protein. The STD NMR experiment is based on the NOE effect and on the observation of the resonance signals of ligands. Basically, an STD experiment consists in the subtraction of a spectrum in which the protein is selectively saturated (on-resonance spectrum, obtained by irradiating a region of the spectrum that contains only resonances of the receptor/protein and not of the ligand, normally between 0 ppm and -1 ppm) with signal strength I_{SAT}, from one recorded without saturation of the protein (off-resonance spectrum), with signal intensity I_0. In the difference spectrum, $I_{STD} = I_0 - I_{SAT}$, only the signals of the ligand/s which receives the saturation transfer from the

protein (through spin diffusion, NOE effect) remain. This is because, for a ligand with weak affinity (dissociation constant, K_D, in the range of 10^{-8} M - 10^{-3} M), there is an exchange between the bound and the free state (**Figure 13**)[24].
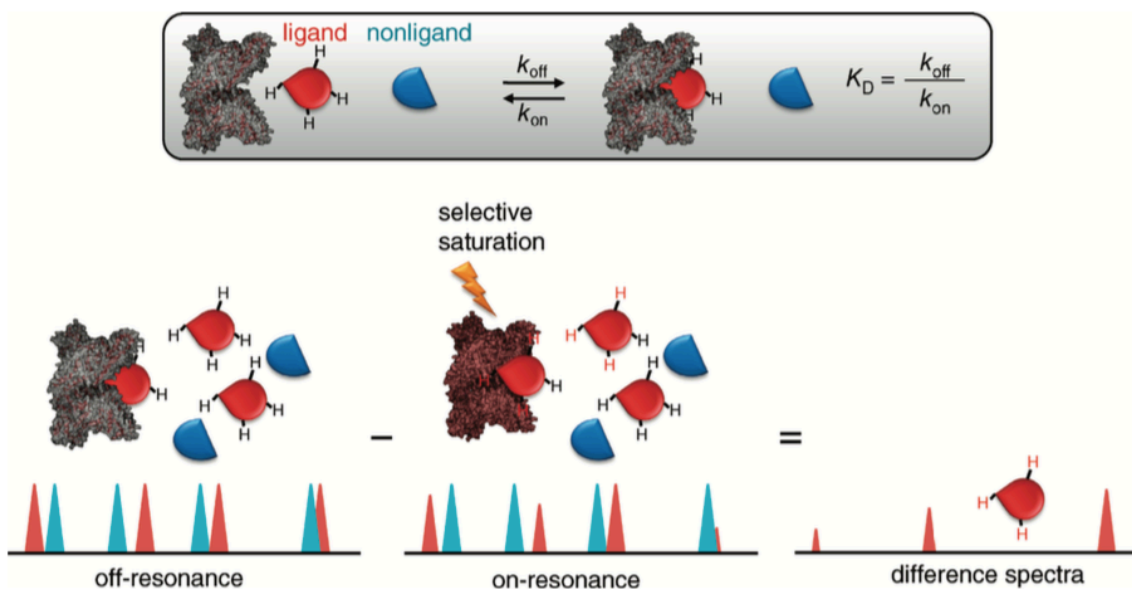
Figure 13 Scheme of the STD-NMR experiment. The exchange between free and bound ligand allows intermolecular transfer of magnetization from the receptor to the bound small molecule[24].

The difference in intensity due to the saturation transfer can be quantified and is an indication of the affinity. For a molecule that binds to the receptor, only the signals of the protons that are in close contact with the protein (≤ 5 Å) and receive transfer of magnetization appear in the STD spectrum.

Water-LOGSY experiment

The water-LOGSY is another 1D experiment based on ligand-observation for detection of protein-ligand interactions[25]. As the STD, water-LOGSY is based on the transfer of magnetization via intermolecular NOE and spin diffusion. In this experiment, water molecules are involved in the transfer pathway. Upon irradiation

of the water signal, the molecules that interact with water via water-ligand-protein or protein-ligand complexes exhibit a negative NOE with water while the molecules that do not bind the protein will have weak and positive intermolecular NOEs with water (**Figure 14A**). Therefore, binders and non-binders can be discriminated in a water-LOGSY spectrum, as they display opposite signs for their corresponding peaks (**Figure 14B**).
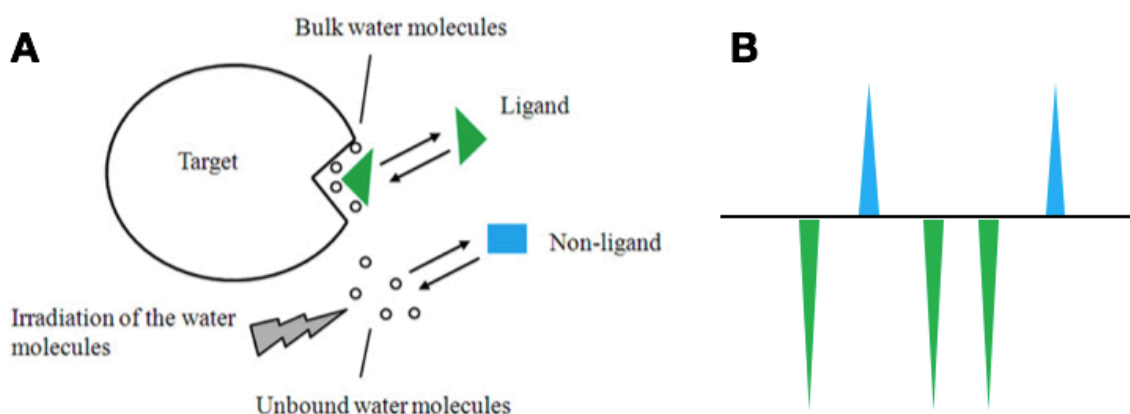
Figure 14 (A) water-LOGSY principle and (B) representation of a water-LOGSY spectrum.

Computational Chemistry

Molecular Docking

The aim of molecular docking protocols is the prediction of the three dimensional structure of the macromolecular complex between the protein and the ligand. The problem is generally attributed to the optimization of a mathematical function (scoring function) through an appropriate algorithm. The scoring functions associate a value that estimates the extent of interaction in terms of free energy of the complex to a vector with the atomic coordinates of protein and ligand. The number of degrees of freedom of the system makes it unthinkable to be able to treat it without introducing simplifications. First, a binding box is typically specified, inside which the

ligand should remain (this is equivalent to impose limitations on the coordinates of the atoms). In the second place, bond lengths and bond angles are fixed, reducing the conformational analysis to the rotation around single bonds. The further degree of simplification distinguishes between the following approaches:

- Rigid molecular docking: in addition to the simplifications set out above, also rotations around single bonds are not permitted, both for the protein and for the ligand. Both molecules are then treated as rigid bodies. The number of degrees of freedom of the system are then reduced to 6 (3 translational and 3 rotational of the ligand within the binding pocket). In this approach, the choice of the initial conformation of the ligand is crucial because it is not changed during the docking protocol.

- Semi-flexible molecular docking: the treatment of the protein is the same as in the rigid molecular docking. The ligand however, is considered flexible. Currently, almost all algorithms implement a conformational search that maintains bond distances and bond angles fixed and allow the variation of the dihedral angles. The number of degrees of freedom then rises to $6 + N_{lrb}$, where N_{lrb} is the number of rotatable bonds of the ligand.

- Flexible molecular docking: it is the most comprehensive approach and the most computationally demanding. A conformational search is performed on both the protein and the ligand. Obviously, these types of algorithm are the last ones in chronological order to be exploited and have probably not yet achieved their full potential. The flexibility of the protein is generally limited to the side chains of the amino acids of the binding site and is treated in the same way as the ligand or through the use of libraries of rotamers.

The free energy of binding is given by the Gibbs-Helmholtz equation:

$$\Delta G = \Delta H - T\Delta S$$

where ΔG is the free energy of binding, ΔH the entropy, T is the temperature, and ΔS the entropy. ΔG is associated with the binding constant K_i by the equation:

$$\Delta G = -RTlnK_i$$

where R is the gas constant. The scoring function is a key element of the docking, to estimate the affinity of the compound for the protein. We can divide the scoring functions in three types:

- *Empirical scoring functions*: these are based on the physical-chemical properties of the system; the total free energy is composed of several energy terms corresponding to hydrogen bonds, hydrophobic and electrostatic interactions, entropic effects and in some cases to interactions with metal ions. A multilinear regression is then used to optimize the coefficients of the equation using a training set of protein-ligand complexes for which both the binding affinity and the three-dimensional structure are known.

- *Scoring functions based on force field*: these are based on terms of molecular mechanics of force field.

- *Knowledge-based scoring functions*: these are based on the principle that situations that are more frequently seen in the three-dimensional structures are energetically more favored. The free energy of the protein-ligand complex is then calculated as the sum of the free energy of interatomic contacts, derived from the frequency of the interatomic distances in the database of experimentally determined three-dimensional structures.

As a scoring function may be simplified, the complexity of its optimization combinatorially grows with the number of degrees of freedom, and this in practice often precludes the use of deterministic algorithms (*i.e.,* able to ensure the optimal

solution). Usually, the research algorithm is heuristic; therefore, it does not guarantee that the solution found is indeed the best possible but, on the other hand, it is able to determine it in a reasonable time. In our case, the molecular docking software PLANTS[26] [27] was used, which employs a search algorithm called ACO, acronym of Ant Colony Optimization, and is a class of optimization algorithms initially introduced by Marco Dorigo[28] and based on the observation of the behavior of an ant colony. The molecular docking software AutoDock is based on another heuristic algorithm, a genetic algorithm that mimics the process of natural selection[29].

Tanimoto similarity

Chemical similarity is one of the most important concepts in chemoinformatics[30]. It plays an important role in the modern approach to the prediction of the properties of chemical compounds, in the production of chemicals with a predetermined set of properties and, specifically, on the development of studies on the design of drugs considering large archives containing structures of available chemicals (or potentially available). These studies are based on the principle of the similar properties of Johnson and Maggiora, which states: similar compounds have similar properties[31]. To determine if two compounds are similar to each other or not is a major problem in chemistry. A simple count of shared functions (substructures-fragments in common) can be a measure of the chemical distance when a similarity coefficient is used. Dictionaries of predefined structural fragments, such as keys MACCS of the MDL Information System, are used to identify the functions contained into a molecule[32]. The structural fragments or functional groups present in the molecule are set on (bit set as 1) and those that are absent are kept off (bit set as 0). In this way, every molecule is defined by a string containing a sequence of 1 and 0 (bit string). Once the

molecules have been represented by such bit strings, the Tanimoto coefficient can be used for evaluating the similarity. If we are comparing two molecules, A and B, and N_A is the number of functions (in bits) in A, N_B is the number of features (in bits) in B, and N_{AB} is the number of characteristics (on bits) common to both A to B, then the coefficient of Tanimoto is simply:

$$T = \frac{N_{AB}}{N_A + N_B - N_{AB}}$$

Note that the "off" bits do not determine the similarity. Compounds that have Tanimoto coefficient values > 0.85 are generally considered similar to each other. A value of 1 does not necessarily mean that the molecules are identical.

Results and Discussion

Expression and characterization of Bcl-X$_L$ ΔTM

For all the experiments, we used a soluble construct of human Bcl-X$_L$ without the last 24 amino acids corresponding to the transmembrane domain. The protein was expressed uniformly ^{15}N labeled growing *E. coli* bacteria in M9 medium supplemented with [15]NH$_4$Cl for the labeling (see Methods and Materials). After the purification steps, to test the correct folding of the protein, the 2D [1H,15N]-HSQC spectrum of Bcl-X$_L$ was recorded (**Figure 15**). The signals of the protein are well spread and closely resemble the deposited [1H,15N]-resonance assignments (adapted from BMRB 18250) confirming the correct folding of our construct.
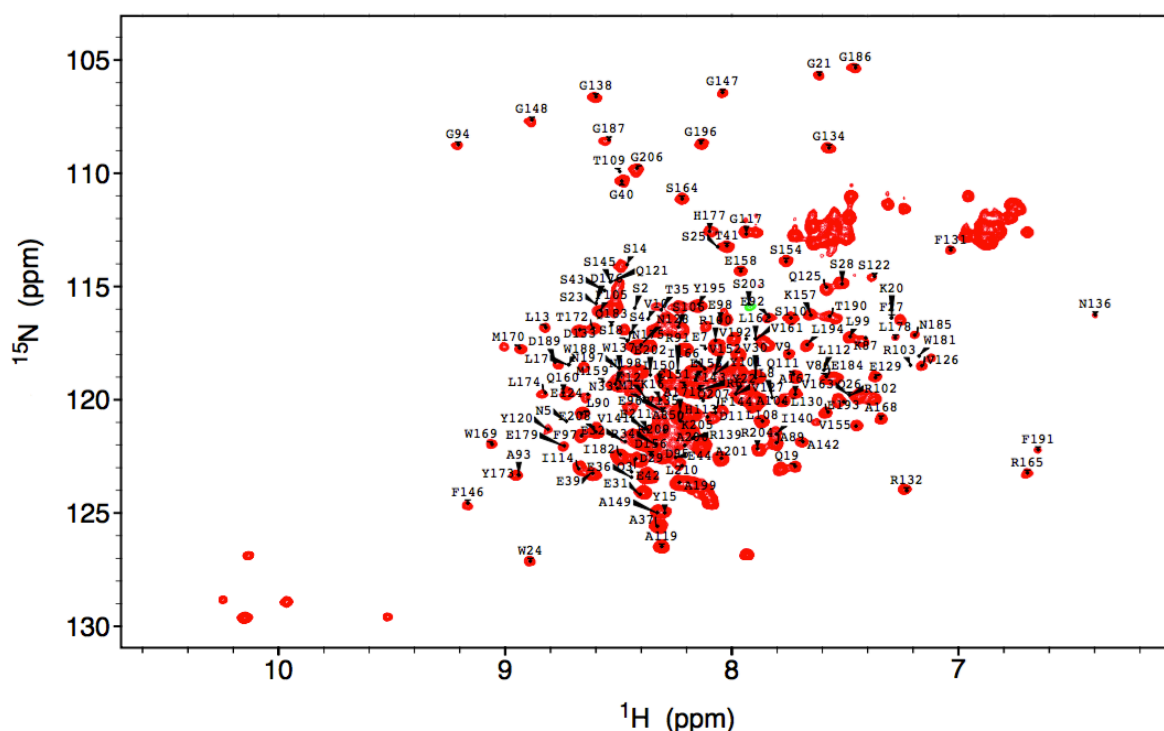
Figure 15 2D [1H,15N]-HSQC spectrum of 150 μM Bcl-X$_L$ ΔTM with superimposed the [1H,15N] resonance assignments adapted from the deposited BMRB 18250.

In silico and NMR screening of a fragments library

After having proved the correct folding of Bcl-X$_L$, and with the aim to find possible novel agents that bind to the target protein, we decided to select a small library of fragments in the pursue of possible small molecules with low affinity toward Bcl-X$_L$ which can then be modified to obtain ligands with higher affinity[12]. For this purpose, we used a library of 653 compounds synthesized in the laboratory of Professor Zagotto (Dipartimento di Scienze del Farmaco, University of Padova, Italy). To select only a few fragments to be tested experimentally, we first selected some compounds using chemoinformatic methods, such as Tanimoto similarity, and molecular docking. This library was first compared with a second library (from ChEMBL) of 625 compounds known from the literature to bind Bcl-X$_L$, with the aim of selecting the fragments of our library that have some common features known to be important for the interaction with the target protein. Before comparing the two libraries, our compounds were filtered using the rule of three[33], a variant of the Lipinski's rule of five[34]. The Lipinski's rule of five is a simple algorithm devised by Christopher Lipinski in 1997 that turns out to be very important for the design and development of drugs. It is a rule derived empirically using information accumulated after decades of research in the pharmacological field. The rule of three was adapted for the construction of fragment libraries for lead generation. This rule suggests that a fragment must have a molecular weight ≤ 300 Da, a number of hydrogen bond donors ≤ 3, a number of hydrogen bond acceptors ≤ 3 and a partition coefficient (ClogP) ≤ 3. This simple rule has reduced the field of investigation decreasing the number of molecules in which to look for possible hits. The fragments obtained from the rule of three were then compared with the library of the Bcl-X$_L$ known binding agents and submitted to molecular docking.

For each compound of both libraries, the functional groups were identified by assigning to each molecule a string based on the MACCS coding, and a fingerprint of each fragment that identified its substructures. With these strings that identify each molecule, the Tanimoto similarity coefficient (T) of our compounds with any molecule known to interact with Bcl-X_L was calculated. The comparison was made using the fingerprints by writing a bash script based on OpenBabel. Usually, compounds are considered similar if the value of the Tanimoto coefficient is greater than 0.85. However, we were selecting fragments not thought to interact with Bcl-X_L and also we were not looking for compounds too similar to those already known (because the ones already known are not used as drugs for different reasons); instead, we were looking for fragments that have common features, but are substantially different from the known Bcl-X_L binding agents. Indeed, as expected, the Tanimoto coefficient of our molecules exceeded at most the value of 0.5, and were selected precisely the fragments with a T value greater than 0.5.

The same library of compounds was subjected to a semi-flexible molecular docking, where the protein is maintained rigid while the ligands are free to rotate around the single bonds (see Methods and Materials). For the molecular docking, the crystallographic structure of Bcl-X_L corresponding to the PDB ID 2YXJ was used. Using the software MOE, the protein structure was minimized energetically, the protons was added, and the partial charges were attributed according to the force field AMBER99. Also the compound library was subjected to a structure optimization protocol, which consists in the energy minimization of the geometry using the force field MMFF94, and the assignment of partial charges. PLANTS (Protein-Ligand ANT System) was used as software for molecular docking. It is based on a class of stochastic optimization algorithms called ant colony optimization (ACO), and

CHEMSCORE as scoring function of the geometries of the complex formation. Before

starting the molecular docking, both compounds with high affinity toward Bcl-X$_L$

were added to our library, and other molecules known not to bind the target protein,

as internal controls of the goodness of the docking. For each ligand, 10 conformers

with the best scoring were returned as output. The results can be considered good as

the first compounds in the list are precisely those known to bind Bcl-X$_L$ (**Figure 16**)

while those with the lowest scores are those without affinity for the protein.
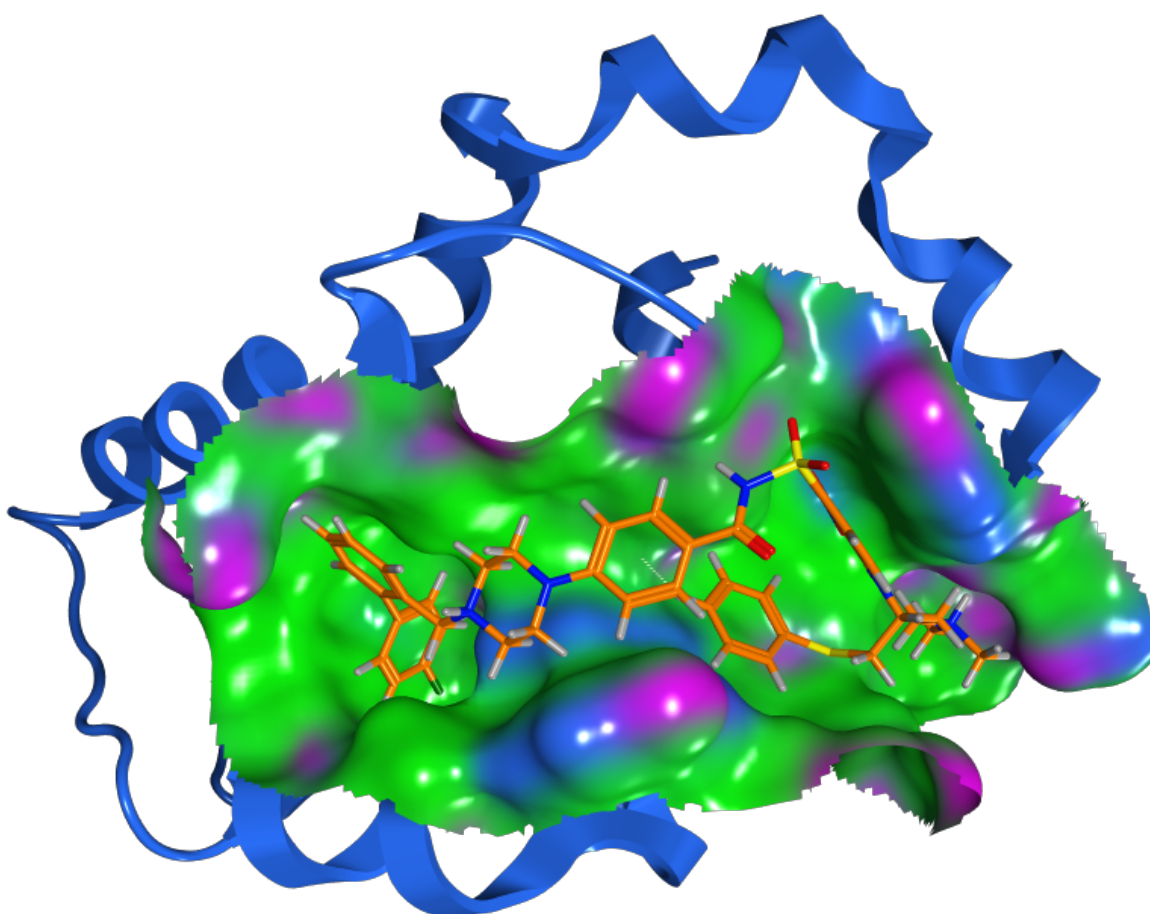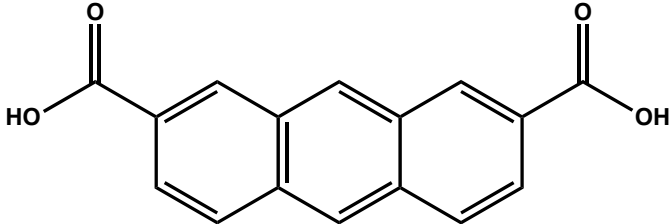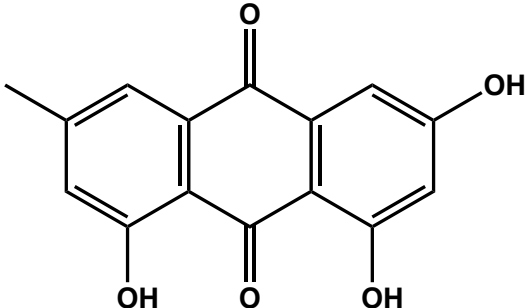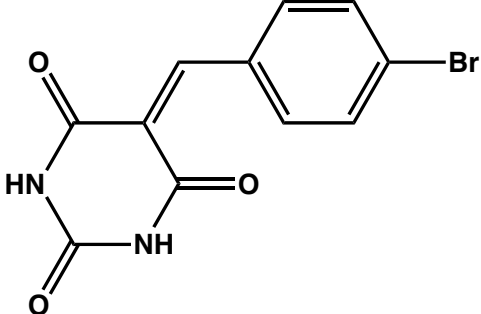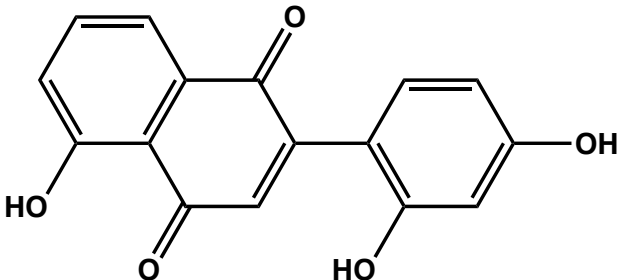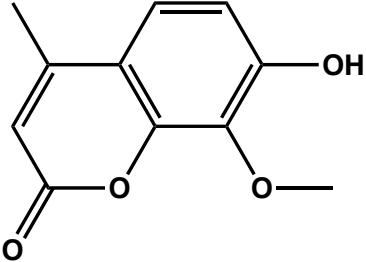
Figura 16 Docking result with the highest scores, which is ABT-737, as expected, a compound known to bind Bcl-X$_L$ with nanomolar affinity. The compound was added as an internal control to check the goodness of the molecular docking. The pose of the compound into the binding pocket closely resembles the pose that ABT-737 has in the crystallographic structure with PDB ID 2YXJ. The binding surface is color-coded: hydrophobic area in green, H-bonding area in purple, and mild polar area in blue.

Using the Tanimoto coefficient and the molecular docking, the 5 fragments reported in **Table 2** were selected. In **Figure 17**, the lowest energy pose of fragment Z314 is shown, in which it is docked into the first hydrophobic pocket of Bcl-X$_L$.

Table 2. Fragments selected using the molecular docking and the Tanimoto coefficient for experimental analysis.

ID	STRUCTURE	MW (Da)
Z65		266
Z119		270
Z166		295
Z176		282

Z314		206
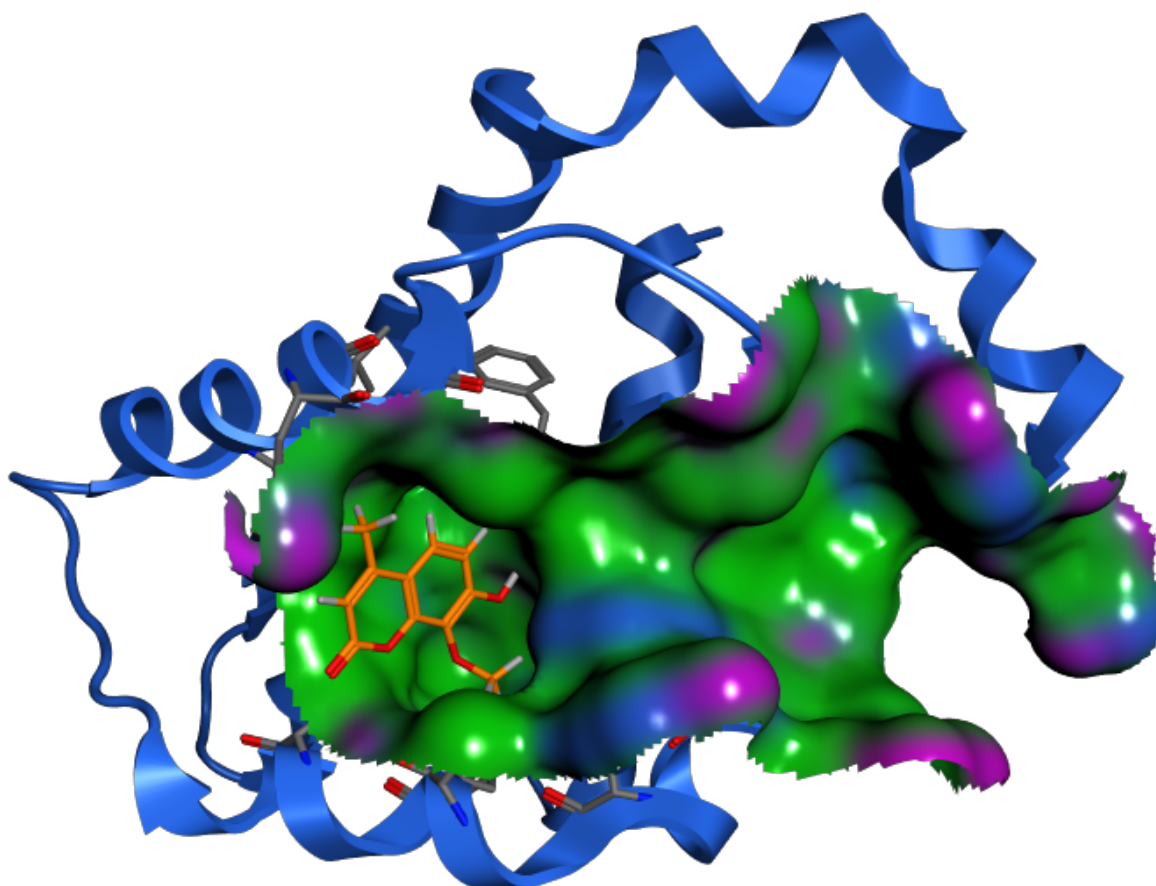
Figure 17 Docking pose of fragment Z314 into the binding pocket of Bcl-X$_L$. The fragment is well docked into the first hydrophobic pocket. The binding surface is color-coded: hydrophobic area in green, H-bonding area in purple, and mild polar area in blue.

These 5 fragments were tested using the Saturation Transfer Difference (STD) and the water-LOGSY experiments, which are ligand based NMR experiments. In the first experiment, the aliphatic region of the spectra below 0 ppm where the fragments have no resonances was irradiated. In this way, the resonances of the protein were saturated and in case of binding, the saturation was transferred to the ligand

decreasing the intensity of its signals. The STD spectrum is the difference between the spectrum recorded without saturating the protein and the saturated one. An STD spectrum presents signals of the ligand only in the case of binding with the protein. The water-LOGSY experiment is similar to the STD, but instead of saturating the target protein, the protons of the bulk solvent (H_2O) are saturated. The saturation is transferred to the binding ligand by the water molecules on the protein surface. The spectrum results in positive signals for non-binding molecules and negative signals for binding fragments (see Methods and Materials). Both the STD and the water-LOGSY NMR experiment were recorded in the presence of 30 μM Bcl-X$_L$ and 900 μM of each fragment in 50 mM phosphate buffer at pH 7.4 with 150 mM NaCl. Fragment Z314 was found to bind to Bcl-X$_L$ with both experiments. In **Figure 18**, the aromatic region of the spectra is shown; the STD spectrum recorded with a saturation time of 2 s is indicated in green.
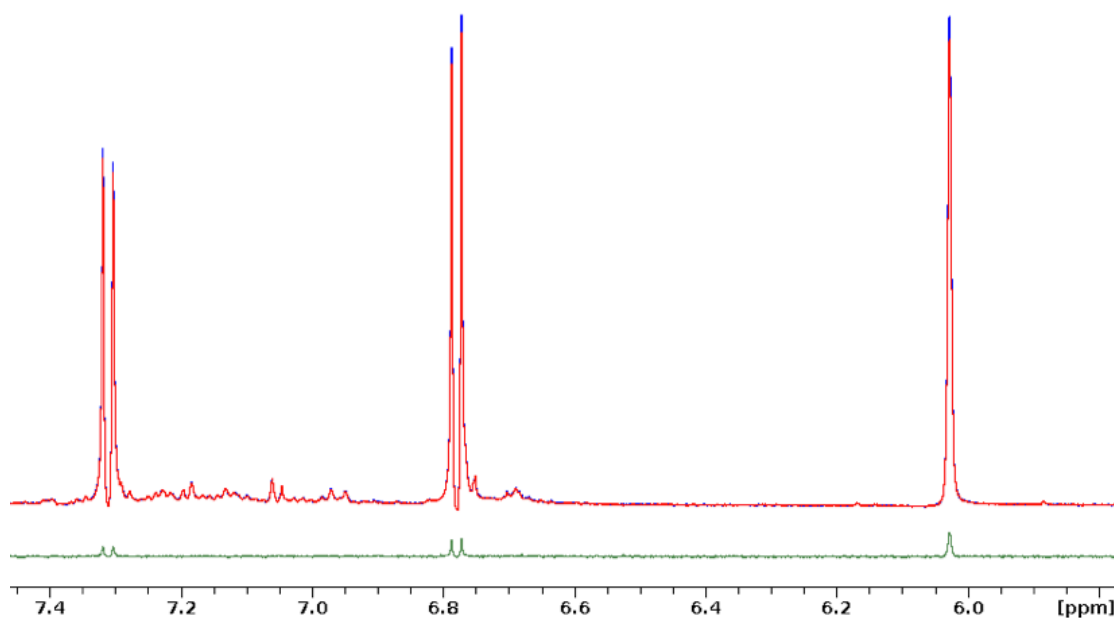
Figure 18 Aromatic region of the Saturation Transfer Difference (STD) experiment of 30 μM of Bcl-X$_L$ in the presence of 900 μM of Z314. the off-resonance spectrum is shown in blue, the on-resonance in red, and the STD in green.

To verify the binding and to avoid the risk of incurring in a false positive, the same experiment was carried out varying the saturation time. In this way, only the signals of the ligand agent increase in intensity increasing the saturation time, which does not happen for false positives. The signal of the DMSO can be used as internal control since, not being a ligand agent, its signal remains unchanged increasing the saturation time (**Figure 19).**
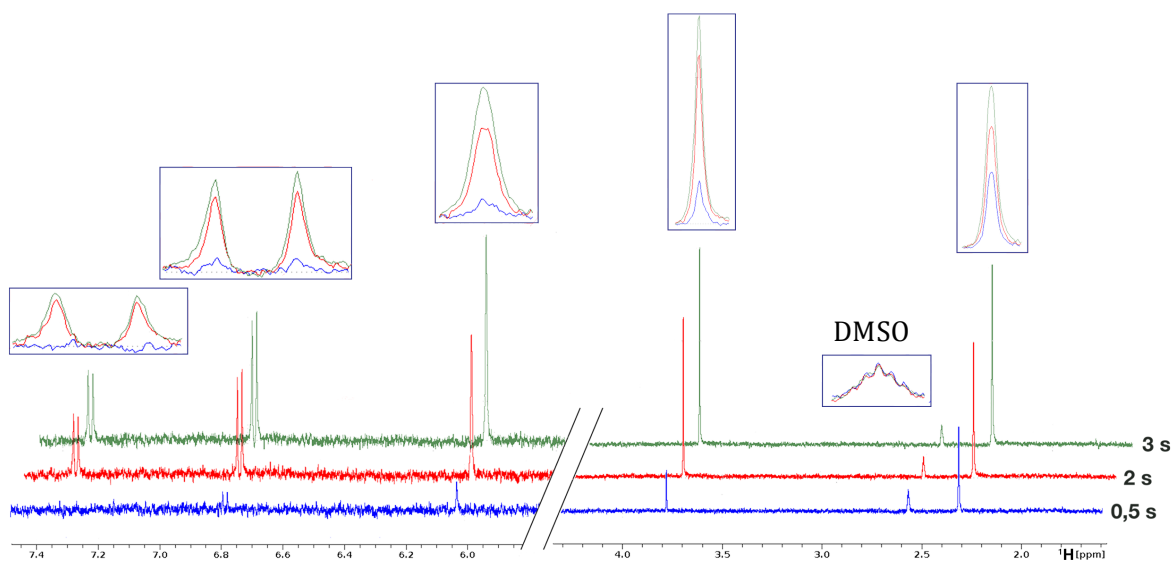
Figure 19 Saturation Transfer Difference spectra (STD) of 30 µM Bcl-X$_L$ in the presence of 900 µM of the fragment Z314 recorded at increasing saturation times. In blue, red and, green the spectra recorded with a saturation time of 0.5 s, 2 s, and 3 s, respectively, are shown. The signal of the DMSO does not change increasing the saturation time indicating that fragment Z314 is not a false positive.

In **Figure 20**, the water-LOGSY spectrum of Bcl-X$_L$ in the presence of fragment Z314 is reported; the signals of our compound are all negative indicating the presence of binding. Also in this case, the positive signal of DMSO can be used as an internal control.
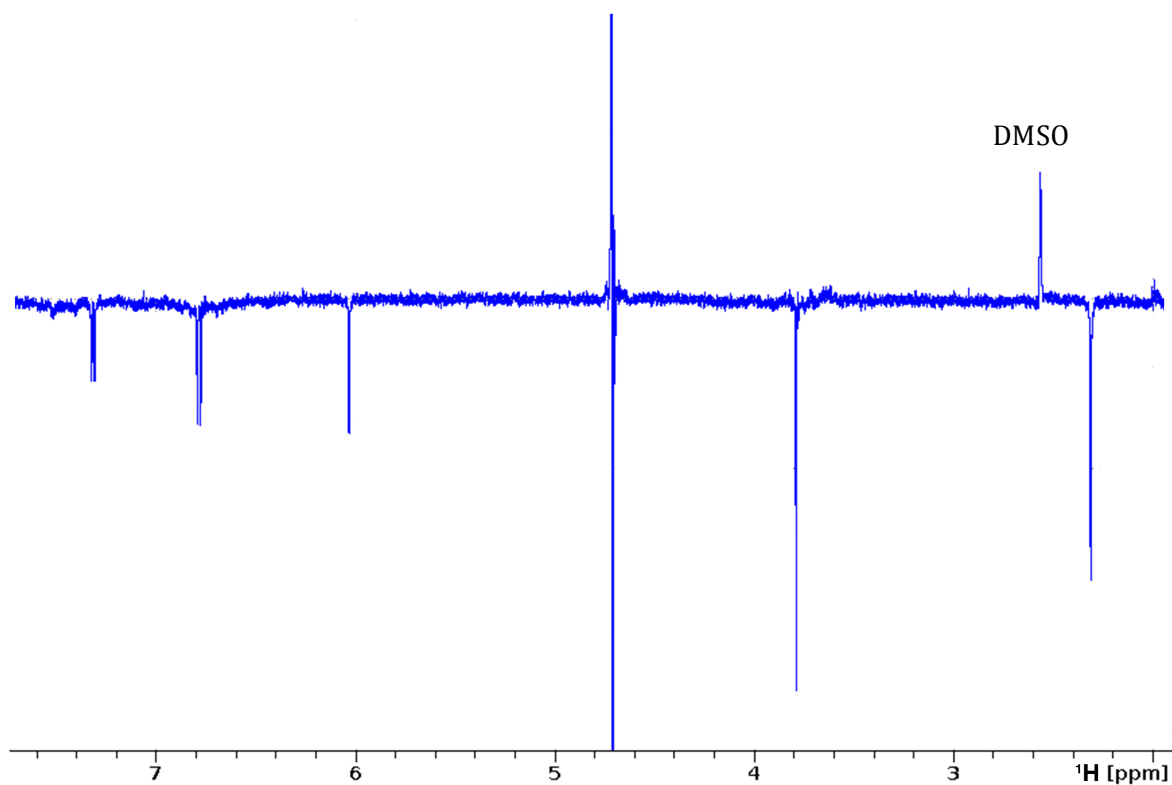
Figure 20 Water-LOGSY spectrum of 30 μM Bcl-X$_L$ in the presence of 900 μM of fragment Z314. All the signals of the compound are negative indicating the presence of binding.

To further confirm the binding, a more robust protein based NMR experiment was also recorded, the 2D [^{1}H,^{15}N]-HSQC with 110 μM Bcl-X$_L$ in the presence of 2 mM of our compound. From this experiment (**Figure 21**), a chemical shift perturbation of the resonances of some residues situated in the hydrophobic binding pocket of Bcl-X$_L$ can be appreciated.
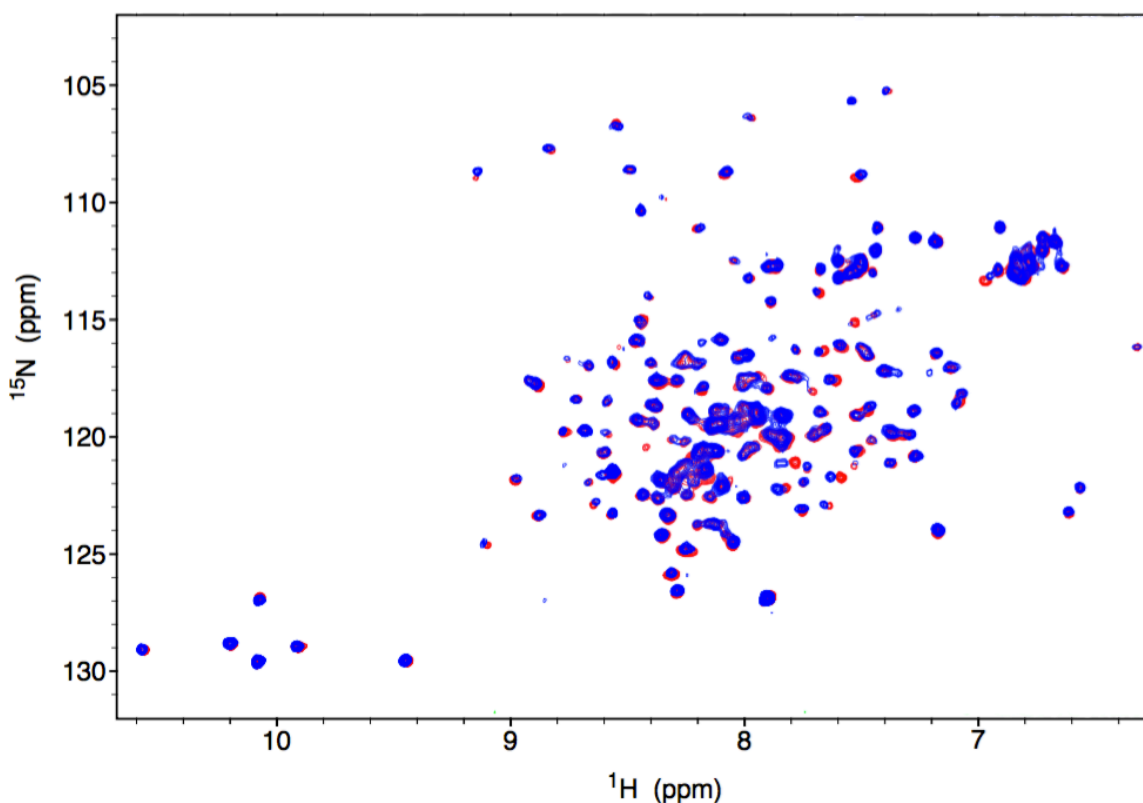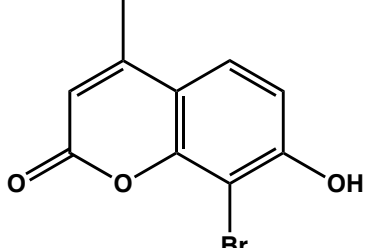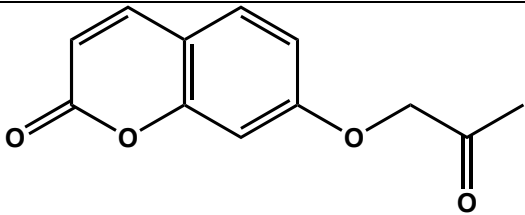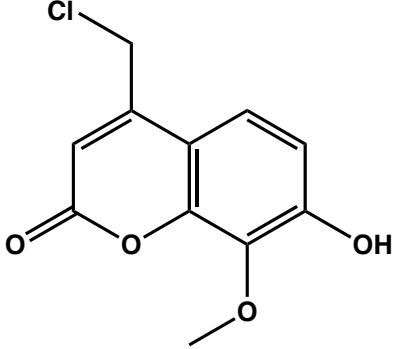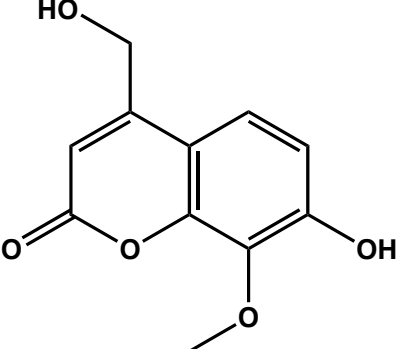
Figure 21 2D [^{1}H,^{15}N]-HSQC spectra of 110 μM ^{15}N-Bcl-X$_L$ recorded in the absence (red) and in the presence (blue) of 2 mM of fragment Z314.
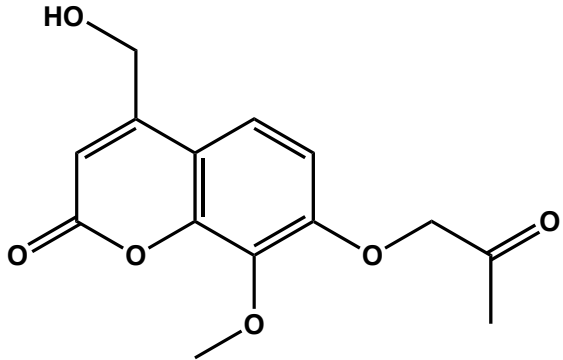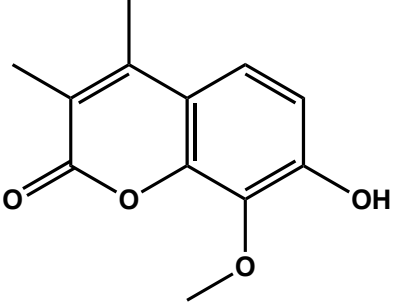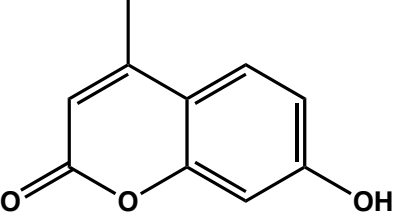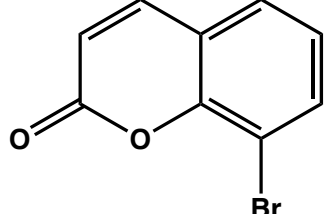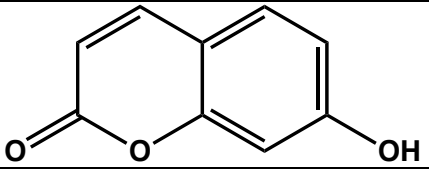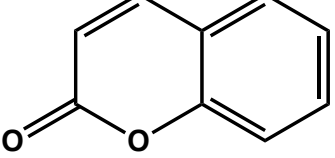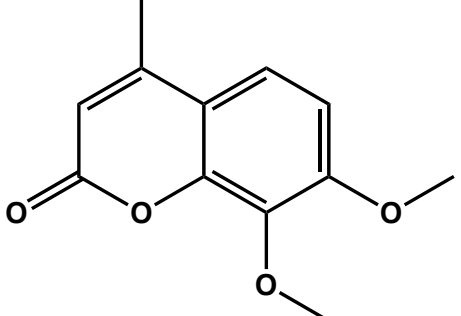
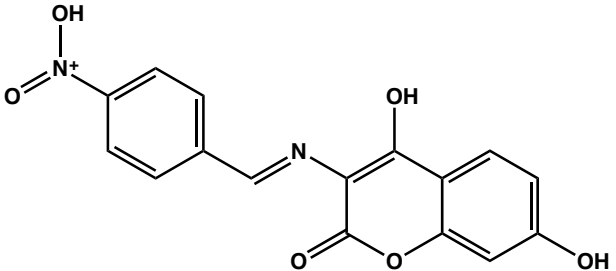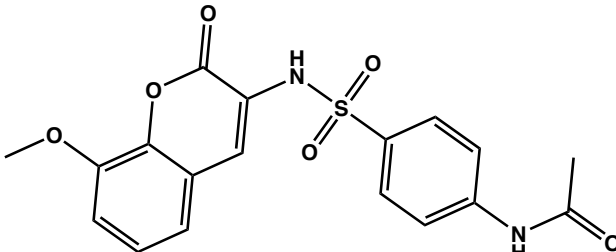## Hit optimization and molecular docking

After this initial screening, and after confirming the binding between fragment Z314 and Bcl-X$_L$, initial fragment evolution studies were performed using compounds analogous to our initial hit. The scaffold of our compound is a coumarin; for this reason, we chose 14 different molecules with coumarin as main scaffold (**Table 3**), with different substituents in different positions, with the aim of identifying a more potent hit. All the 14 compounds in **Table 3** were synthesized in the laboratory of Professor Zagotto (Dipartimento di Scienze del Farmaco, University of Padova, Italy). For each of these compounds, the STD, the water-LOGSY, and the 2D [^{1}H,^{15}N]-HSQC experiments were recorded. Both the STD and the water-LOGSY NMR experiment were recorded in the presence of 30 μM of Bcl-X$_L$ and 900 μM of each fragment in 50

mM phosphate buffer at pH 7.4 with 150 mM NaCl. The 2D [^{1}H,^{15}N]-HSQC experiments were recorded using 150 μM Bcl-X$_L$ in the presence of 3 mM of each molecules. The results for each experiment are summarized in **Table 3**.

Table 3 Compounds selected for their homology to the initial hit Z314, and tested against Bcl-X$_L$. For all the experiments, the sign "-" means that under our experimental conditions there is no binding. For the STD and water-LOGSY experiments, the sign "+" stands for presence of binding. In the 2D [^{1}H,^{15}N]-HSQC experiment, the symbols +, ++, +++, and ++++ stand for an average chemical shift perturbation CSP < 0.02 ppm, 0.02 ppm < CSP < 0.03 ppm, 0.03 ppm < CSP < 0.06 ppm, and CSP > 0.06 ppm, respectively.

ID	STRUCTURE	STD	Water-LOGSY	[^{1}H,^{15}N]-HSQC
Z69		+	-	+++
Z262		-	-	-
Z406		+	-	++
Z413		+	-	++

45

Z415		-	-	-
Z426		+	-	++
Z585		+	-	++
Z622		+	+	+++
Z1474		+	-	+
Z1486		+	-	+
Z1537		-	-	-

Z1707		-	-	-
Z1777		+	+	++++

In **Figure 22** and **Figure 23**, the experiments for compound Z622 and Z1777, respectively, are reported. The saturation of the signals in the STD spectra was very weak for both ligands, but with the confirmation of the interaction from the water-LOGSY experiment and from the more robust 2D [^{1}H,^{15}N]-HSQC experiment, we can say that, under our experimental conditions, the two compounds bind to Bcl-X$_L$.
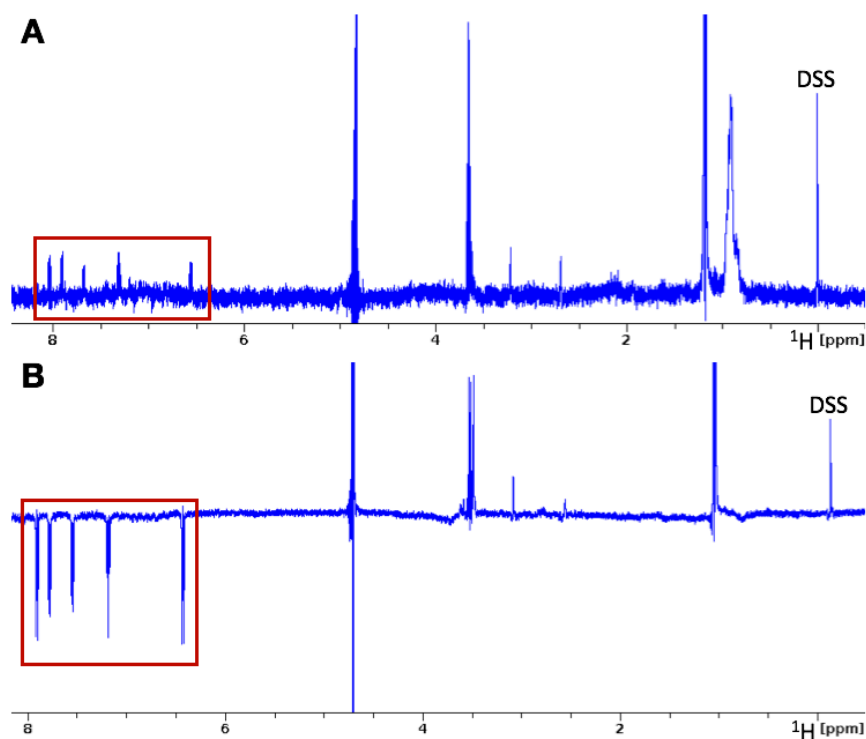
Figure 22 STD spectrum (A) and water-LOGSY spectrum (B) of 30 μM of Bcl-X$_L$ in the presence of 900 μM of Z622. For both spectra, the signals of compound Z622 are boxed in red. Both experiments suggest the presence of binding under our experimental conditions.
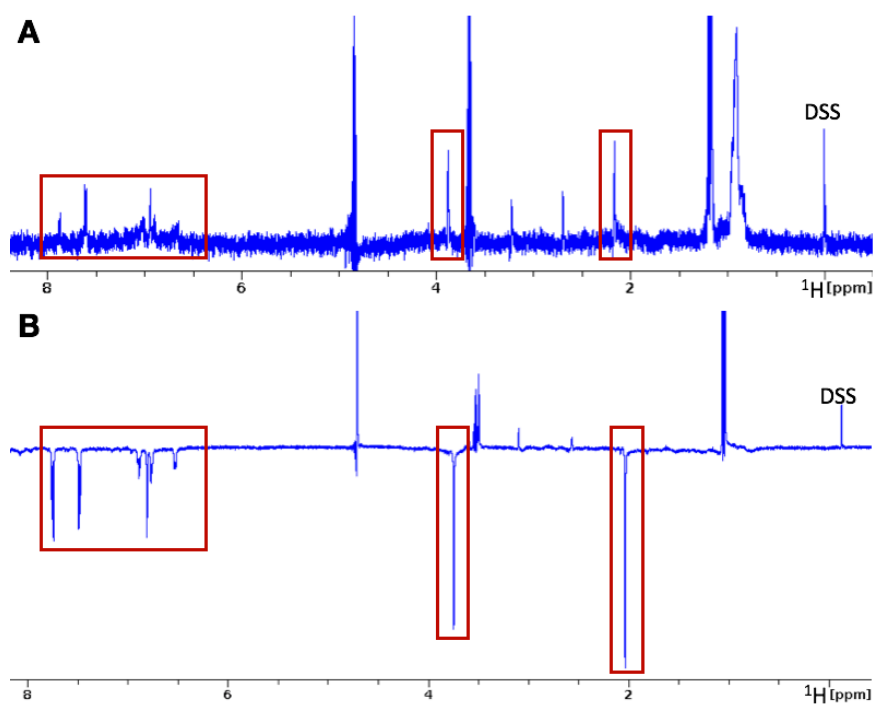
Figure 23 STD spectrum (A) and water-LOGSY spectrum (B) of 30 μM of Bcl-X$_L$ in the presence of 900 μM of Z1777. For both spectra, the signals of compound Z1777 are boxed in red. Both experiments suggest the presence of binding under our experimental conditions.

In addition, a 2D [^{1}H,^{15}N]-HSQC titration was performed for compounds Z622 and Z177 that were found to be active in all the experiments. In **Figure 24**, the superposition of the 2D [^{1}H,^{15}N]-HSQC spectra of 150 μM Bcl-X$_L$ alone and in the presence of 0.5 mM, 1 mM, 2 mM, and 3 mM of compound Z1777, respectively, is reported.

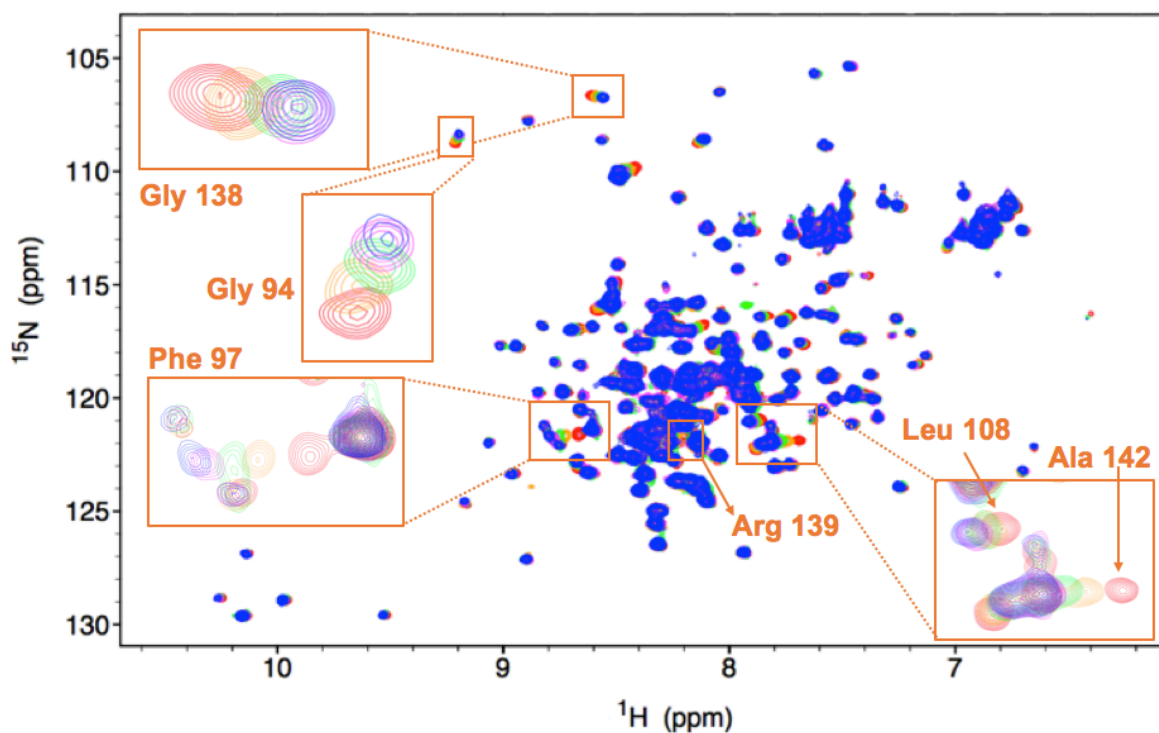**Figure 24** NMR titration of Z1777 against 150 μM ^{15}N-Bcl-X$_L$ using 2D [^{1}H,^{15}N]-HSQC spectra. The spectrum of the apo protein is depicted in red while the spectra in orange, green, purple, and blue were collected in the presence of Z1777 in increasing concentrations: 0.5 mM, 1 mM, 2 mM, and 3 mM, respectively. The resonances corresponding to the most perturbed residues are boxed in orange.

From the Chemical Shift Perturbation (CSP) data, the average dissociation constant (K$_d$) were calculated. A K$_d$ value of 1.45 mM was found for compound Z622 (**Figure 25B**), and a K$_d$ value of 950 μM was found for compound Z1777 (**Figure 25A**). The chemical shift deviations for compound Z1777 are plotted for each residue to summarize the magnitude of the shift in the spectrum (**Figure 26**). The residues presenting largest shifts are concentrated in the proximity of the hydrophobic pocket of Bcl-X$_L$, suggesting an interaction of our compound with the target protein that resembles the interaction with the already known binding agents of Bcl-X$_L$ as indicated by the chemical shift perturbation map (**Figure 27**).
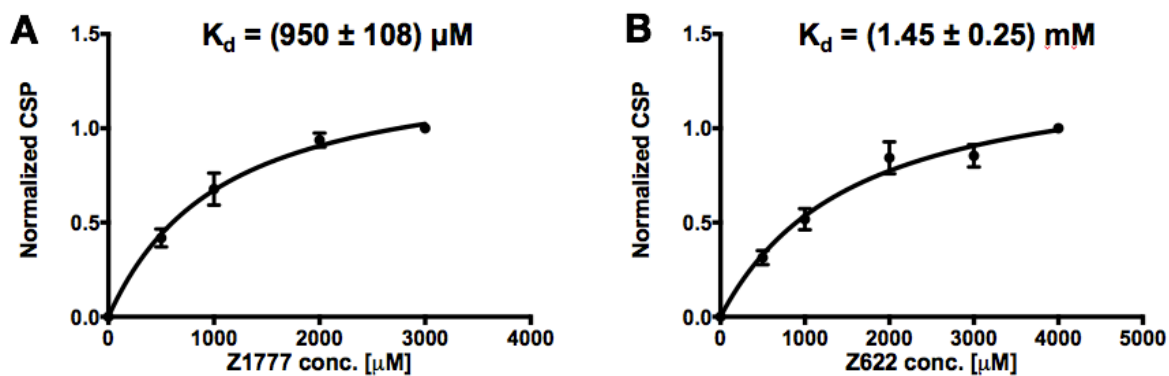
Figure 25 Determination of K_d of Z1777 (panel A), and Z622 (panel B) using the chemical shift perturbation from the 2D [^{1}H,^{15}N]-HSQC experiments. K_d = (950 ± 108) μM for Z1777, and K_d = (1.45 ± 0.25) mM for Z622.
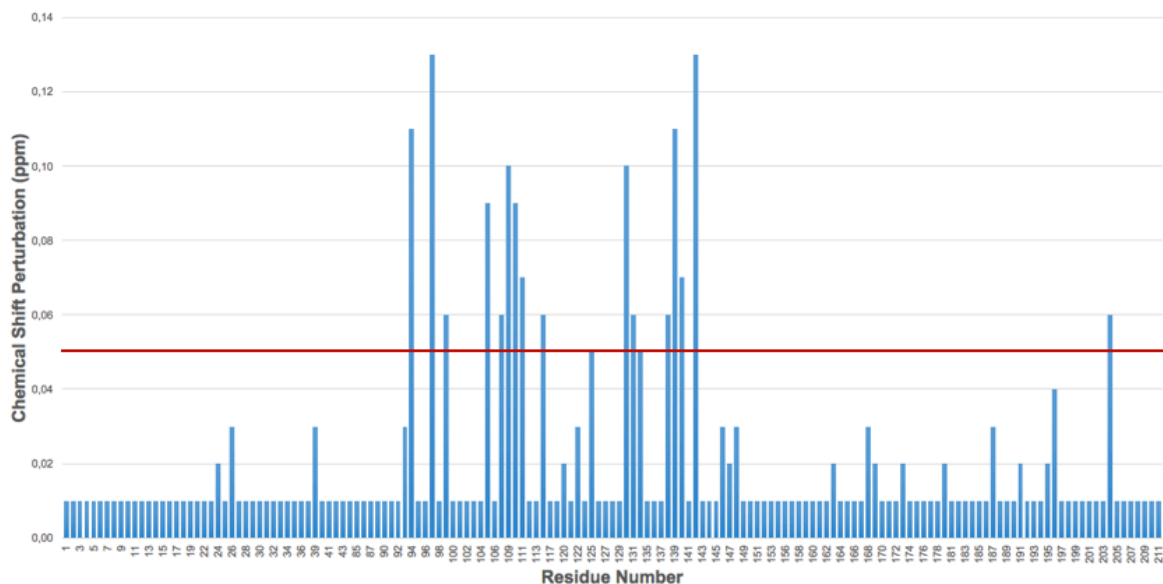

Figure 26 Chemical Shift Perturbation (CSP) plot of the residues of Bcl-X$_L$ from the 2D [^{1}H,^{15}N]-HSQC titration with compound Z1777. The red line corresponds to 0.05 ppm, the value above which the CSP was considered significant.
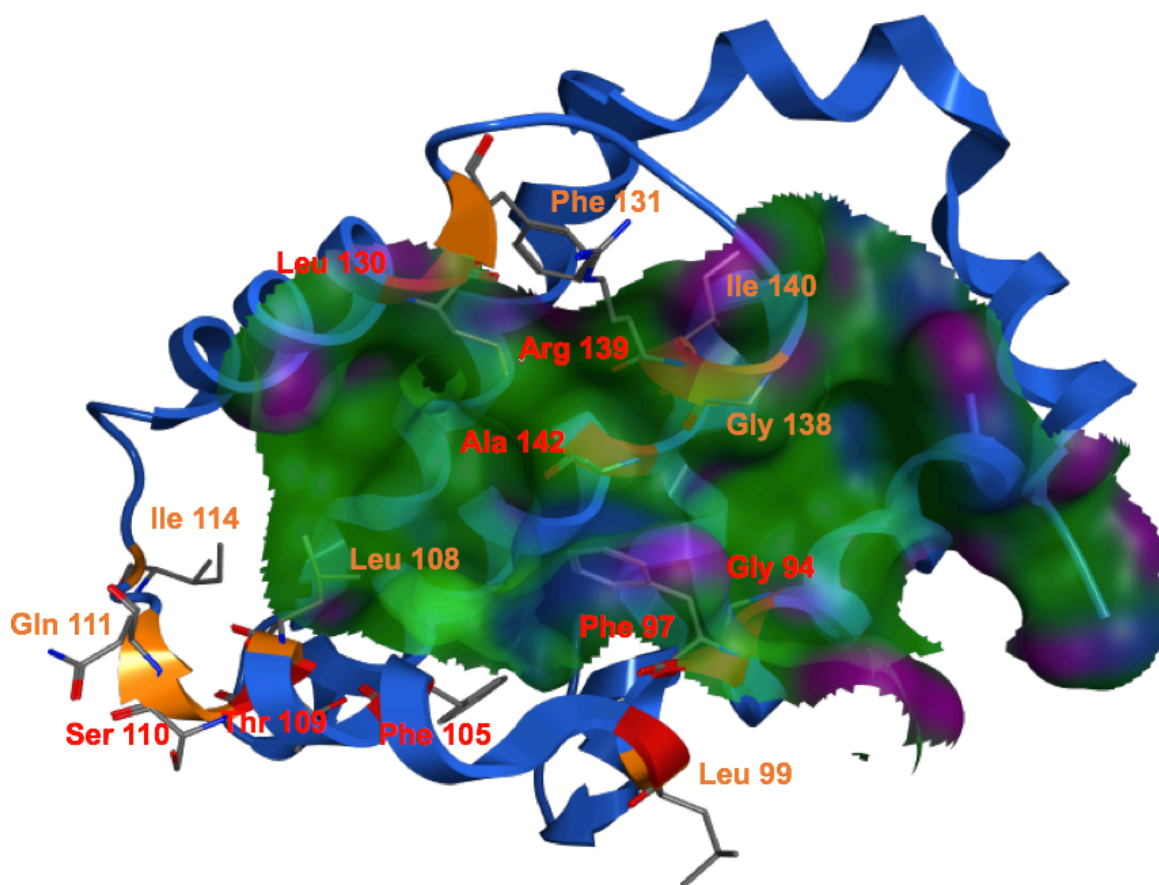
50

Figure 27 Chemical shift perturbation map consequent to the binding between Bcl-X_L and Z1777. The secondary structure of Bcl-X_L is depicted in red for the residues with CSP > 0.08 ppm, and in orange for the residues with 0.05 < CSP < 0.08 ppm.

At this point, it would be necessary to obtain more detailed structural information about the protein-ligand interaction with the aim of modifying our hit to obtain a compound (lead) with higher affinity towards Bcl-X_L. Usually, both X-ray crystallography and NMR are used to determine a high-resolution co-structure. This approach is very time-consuming and, with small molecules with low affinity, often it does not yield satisfactory results. For this reason, we decided to opt for a bioinformatics approach. The molecular docking could be a rapid alternative to provide protein-ligand structures, even though this method suffers from the problem that the correct docked structure has to be predicted. To increase the reliability of the molecular docking results, we decided to combine the advantages given by the

molecular modeling with the experimental information given by the NMR experiments. The chemical shift perturbations information obtained from the 2D [^{1}H,^{15}N]-HSQC spectra were used to guide and filter the molecular docking calculation using the AutoDockFilter program[35].

First, we ran the molecular docking using AutoDock as docking software[29], using the Autodock/Vina plug-in for pymol. The coordinates of the binding pocket of PDB 2YXJ of Bcl-X$_L$ were set and the output were the 30 lowest energy conformations of Z1777 (**Figure 28B**). At this point, using the experimental CSP information (**Figure 28A**) we used the AutoDockFilter program to filter the low energy conformation that is in accordance with the information derived from the NMR experiment (**Figure 28C**).

The docking pose obtained with AutoDockFilter is shown in **Figure 29**; the interaction is with the residues that present bigger chemical shift perturbations. Specifically, an H-bond interaction is present that is fundamental for binding with Bcl-X$_L$ and also a hydrophobic interaction through the coumarin scaffold.
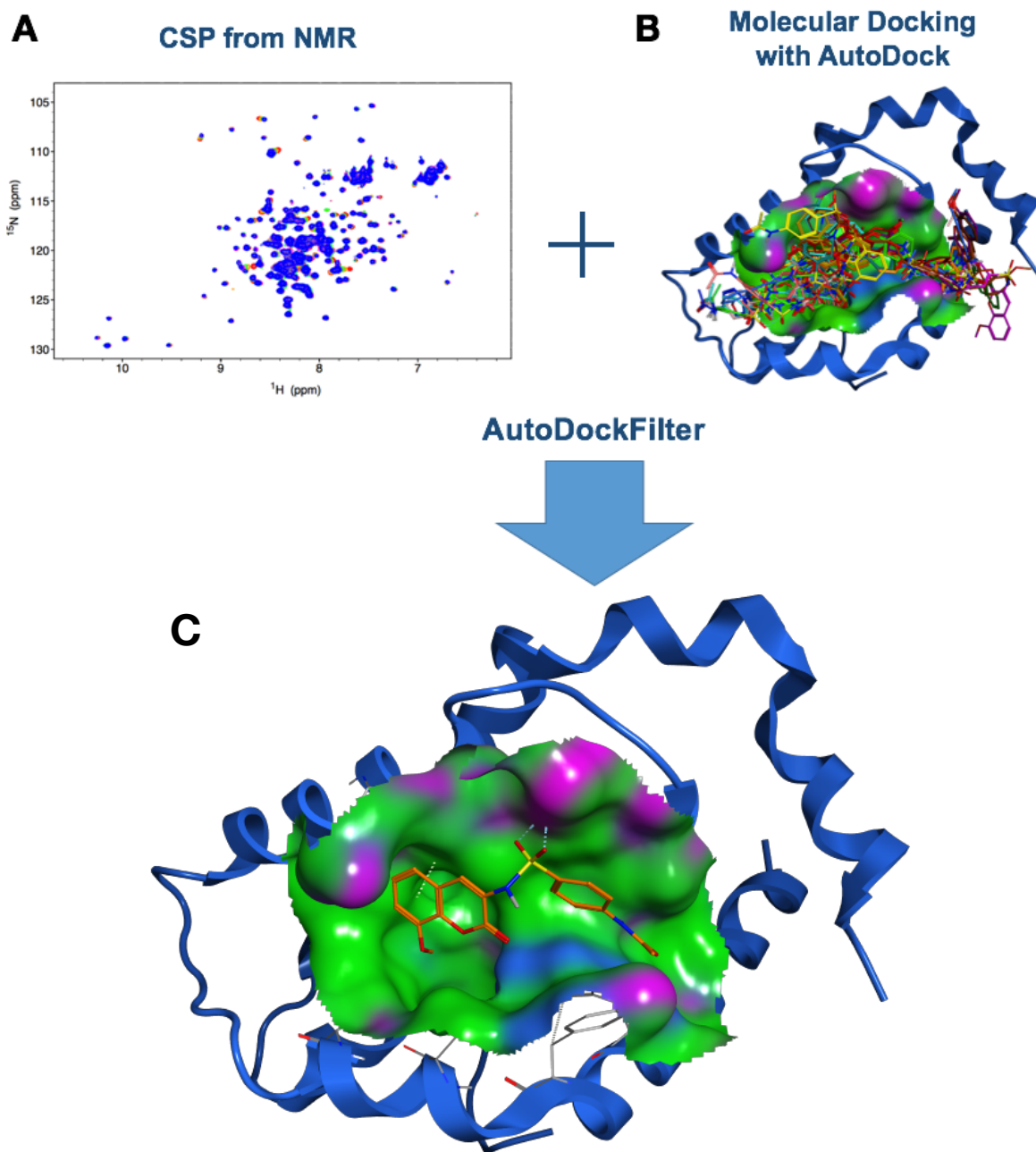
Figure 28 (A) 2D [^{1}H,^{15}N]-HSQC spectra of ^{15}N-Bcl-X$_L$ with increasing concentrations of Z1777. The CSP information from these experiments was used to guide the molecular docking. (B) All the 30 poses obtained from the docking with AutoDock. (C) The pose of Z1777 docked into the binding pocket of Bcl-X$_L$ obtained filtering the 30 solutions with AutoDockFilter.
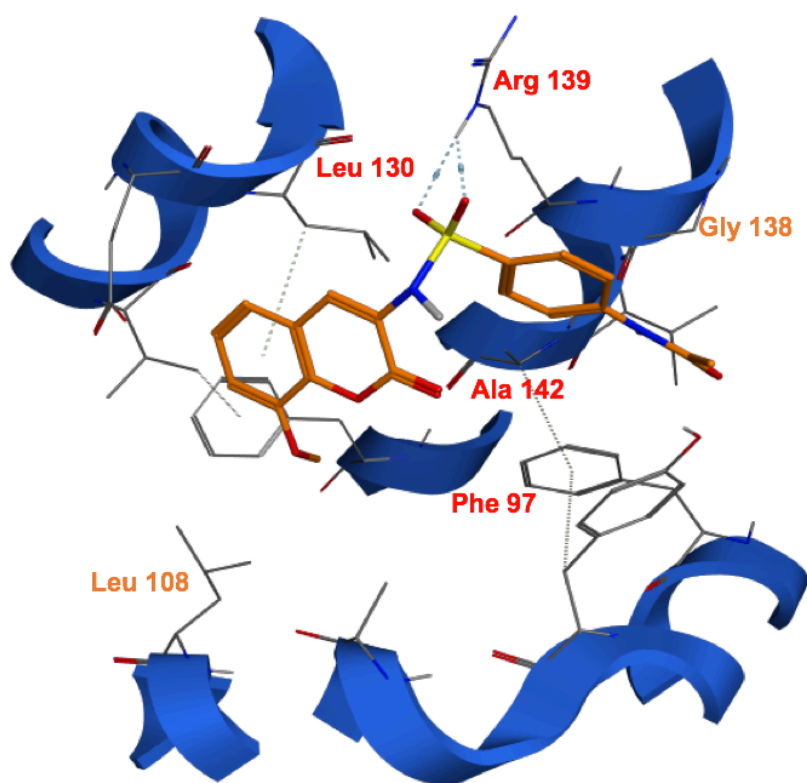
Figure 29 Details of the interaction between Z1777 and Bcl-X$_L$ from the co-structure obtained with AutoDockFilter.

Conclusions

The cellular homeostasis is maintained when the consistency of mitosis (cell proliferation) in a tissue is balanced by death (apoptosis) of an equivalent number of cells[11]. In physiological conditions, apoptosis guarantees the homeostasis of tissues, maintaining the number of cells constant, thanks to the balance between pro- and anti-apoptotic proteins[36]. A dysregulation in the apoptotic process, in particular a lack of apoptosis, is a central point in the development of cancer. The importance of studying this process is evident in order to identify anticancer therapeutics able to interfere with the proteins responsible for this balance. The Bcl-2 protein family regulates the apoptotic process, and in particular the anti-apoptotic proteins belonging to it are responsible for the resistance of cancer cells. In this study, we investigated Bcl-X$_L$, one of the anti-apoptotic proteins belonging to this family, and one of the most involved in the process of programmed cell death, which makes it one of the most attractive proteins for the development of anticancer agents.

In the pursue of binding agents of Bcl-X$_L$, we conducted a fragment-based campaign. Starting from a library of more than 600 compounds, 5 fragments were selected using computational techniques and tested using Nuclear Magnetic Resonance spectroscopy (NMR). One of such fragments was found to weakly bind to the target protein. The scaffold of this compound was used to develop 14 analogues that were tested using ligand-based NMR experiments, such as Saturation Transfer Difference (STD) and water-LOGSY, and the protein-based NMR experiment 2D [^{1}H,^{15}N]-HSQC. The compound Z1777 was found to bind to Bcl-X$_L$ with a dissociation constant of 950 μM. In addition, molecular docking was performed using AutoDock, and the results was guided and filtered through AutoDockFilter using the Chemical Shift Perturbation (CSP) information obtained from the NMR experiment. In this way, we

obtained a convincing structure of the complex Bcl-X$_L$-Z1777, which will allow us to make structural considerations for the hit to lead maturation.

Chapter 2

The cell surface receptor CD44: NMR based characterization of putative ligands.

Dr. Carlo Baggio[1,2], Dr. Elisa Barile[1,2], Dr. Gianluigi Di Sorbo[2], Prof. Dr. Thomas J. Kipps[3], and Prof. Dr. Maurizio Pellecchia[1,2,*]

1. Division of Biomedical Sciences, School of Medicine, University of California Riverside, 900 University Avenue, Riverside, CA 92521, USA.

2. Sanford Burnham Prebys Medical Discovery Institute, 10901 North Torrey Pines Road, La Jolla, CA, 92037

3. Division of Hematology/Oncology, Department of Medicine, University of California San Diego, La Jolla, CA

* Corresponding author: Maurizio Pellecchia, phone number: **(951) 827-7829**; email address: maurizio.pellecchia@ucr.edu

Submittet for publication to ChemMedChem

Introduction

The cell surface receptor CD44, a type I transmembrane glycoprotein, was first described in 1983[37]. This hyaluronan-binding protein is expressed on the surface of many cell types[38] [39], where it is involved in leukocyte migration to inflamed sites, T cell activation, and tumor metastasis[40] [41] [42]. Several studies suggest that the interactions between the adhesion of CD44 with its main natural ligand, the hyaluronic acid (HA) are critical for its function [38] [43] [44]. HA is a linear copolymer of repeating N-acetyl-D-glucosamine and D-glucuronic acid units, and represents one of the most abundant component of the extra-cellular matrix[45]. There are several isoforms of CD44[46], and this variability is enhanced by the presence of post-translational modifications, most notables is its glycosylation (both N- and O-linked carbohydrate side chains)[47]. Each of the different isoforms, however, is characterized by a fixed N-terminal domain that functions as docking site for HA[48]. In particular, the recognition of HA by CD44 is permitted by the presence in the amino-terminal region of about 90-amino acids containing the highly conserved Link domain (residues 32-124) [39].

It has been observed that CD44 expression enhances metastatic behavior of cancer cells[49] enabling matrix metalloproteases (MMPs) activation, thus enabling many biological processes[50] [48] that result into the cleavage of CD44, and concomitant cancer cells detachment and migration[39]. Moreover, in Chronic Lymphocytic Leukemia (CLL), CD44 has been reported to signal B-cell survival by activating both the PI3K/AKT as well as the MAPK/ERK pro-survival pathways, and by inducing the expression of the anti-apoptotic protein Mcl-1[51]. Accordingly, Mcl-1

inhibitors or an anti-CD44 antibody can revert these effects and induce apoptosis in primary CLL cells [52] [53].

Hence, in recent years several laboratories have attempted to identify CD44 binding agents for the development of novel therapeutics. The involvement of CD44 in cells migration and its overexpression on a wide spectrum of tumor cells, make this receptor also a suitable target for delivery of chemotherapeutics in cancer cells. Most approaches targeting CD44 have either used antibodies or small peptides binding to CD44 epitopes. Other targeting strategies used the natural CD44 ligand, hyaluronic acid (HA), as a backbone or scaffold for drug attachment, or as both a targeting moiety and delivery agent for small molecule therapeutics[54].

Of these putative CD44 interacting agents, the peptide A6, an 8-amino acid peptide (acetyl-KPSSPPEE-amino), has been shown to have anti-invasive, anti-migratory, and anti-angiogenic activities in a variety of *in vitro* and *in vivo* model systems[55] [56] [57]. Recent studies[58] with human CLL B-cell lymphocytes have shown that A6 down modulates the expression of CD44 and ZAP-70 (a marker for an aggressive form of CLL), and inhibits B-cell receptor (BCR) signaling, resulting in a direct, dose-dependent, cytotoxicity *in vitro*. Currently the putative CD44 binding peptide A6 is in phase 2 clinical trial in patients with CLL and SLL (Small Lymphocytic Lymphoma)[49] (ClinicalTrials.gov Identifier: NCT02046928).

From structural data, however, one might conclude that CD44 is not particularly 'druggable' as a target [59]. The HABD has no obvious well-formed or deep pockets that would serve as attractive binding sites for small molecule inhibitors and is known to undergo small but important conformational changes upon HA binding [60]. Similarly, the stem region is unlikely to possess a well-defined three-dimensioned structure to accommodate a ligand. Nonetheless, the biological relevance of CD44 in

cancer metastasis provides further impetus for the identification of suitable and novel CD44 binding agents, either peptides or small molecules. To this end we sought to first characterize the binding of reported putative CD44 binding compounds. Moreover, we used an NMR guided fragment-based screening campaign to identify possibly novel agents.

Results

Expression and characterization of hCD44(21-178) and HA$_8$ binding studies

For all the experiments we used a soluble construct of human CD44 from residue Q21 to residue V178, corresponding to the HA Binding Domain (HABD). Uniformly ^{15}N labeled hCD44(21-178) was expressed in *E. coli* and purified as reported by *Banerji et al.*[61]. The yield of the protein was about 14 mg per liter of M9 medium in the monomeric form. The 2D [^{1}H,^{15}N]-sofast-HMQC[62] spectrum of hCD44(21-178) closely resembles the previously reported spectrum[63] and the deposited [^{1}H,^{15}N] resonance assignments (BMRB ID 6093). Furthermore, we tested the ability of our construct to bind its natural ligand, Hyaluronic Acid (HA)[64] [43] [44]. Structural studies have been reported for the complex between HA$_8$ (**Figure 30**) and murine CD44 HABD by X-ray crystallography[60] (PDB 2JCR). In addition, the NMR structure of hCD44 HABD in complex with HA$_8$ was also reported[63], although only the coordinates of the protein were deposited (PDB 1POZ). Hence, we prepared a model of the structure of hCD44 HABD in complex with HA$_8$ using SWISS-MODEL[65] [66] [67] [68] and the PDB-ID 2JCR as template (**Figure 30**). To test the ability of our hCD44(21-178) to bind HA$_8$, a 20 µM solution of ^{15}N-hCD44(21-178) was titrated with increasing amounts of HA$_8$ (20, 40, 60, 80, 100, 120, 140, 160, 180, 200, and 220 µM, respectively) collecting both 1D ^{1}H and 2D [^{1}H,^{15}N]-sofast-HMQC NMR spectra at each HA$_8$ concentration (**Figure 31 and 32**). Binding of HA$_8$ to hCD44(21-178) could be appreciated by chemical shift perturbations induced in the [1]H-aliphatic region of the spectrum of hCD44(21-178) (**Figure 31**).
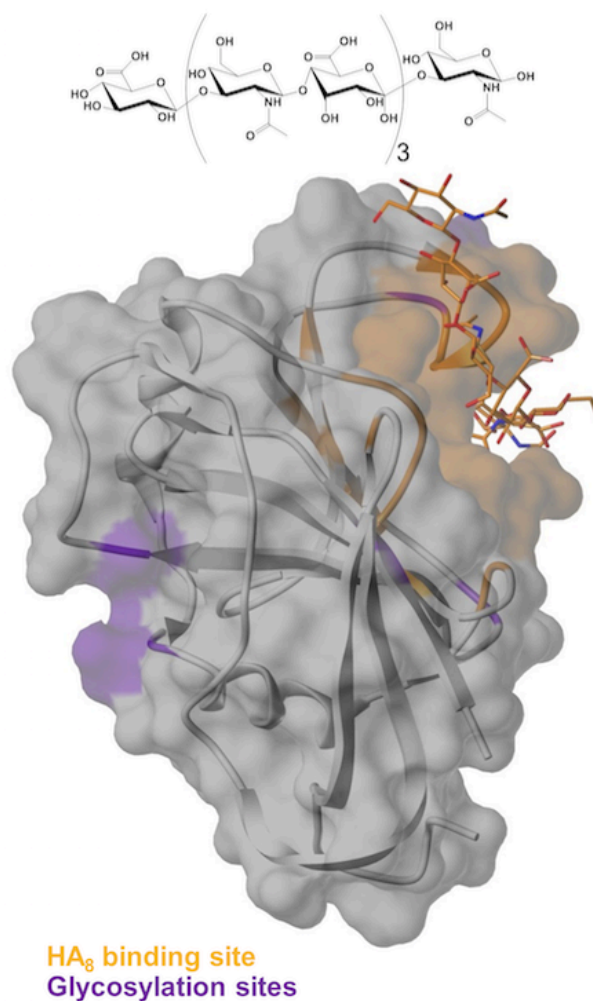
Figure 30 Semitransparent molecular surface representation of the homology model of hCD44 built with SWISS-MODEL [65] [66] [67] [68] using the structure of mCD44 in complex with HA$_8$ (PDB 2JCR) as template. HA$_8$ (chemical structure is reported in the upper panel) and the HA binding pocket are colored in orange. The glycosylation sites are highlighted in purple.
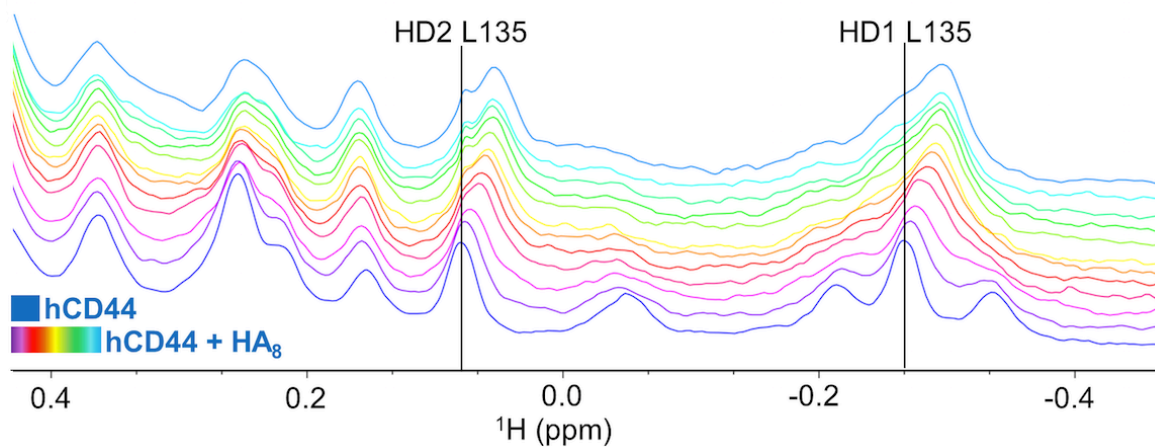

Figura 31 1D ^{1}H-aliphatic spectra of 20 μM hCD44(21-178). The apo spectrum is blue while the spectra recorded in presence of increasing concentrations of HA$_8$ (starting from 20 μM to 220 μM) are in purple to light-blue.

In particular, the resonances corresponding to δ-protons of residue L135 seemed most affected by the presence of HA$_8$. Interestingly, L135 is not in direct contact with HA$_8$, hence the observed perturbations are likely a results of conformational changes in hCD44(21-178) induced by ligand binding, in agreement with previously reported structural studies[69].
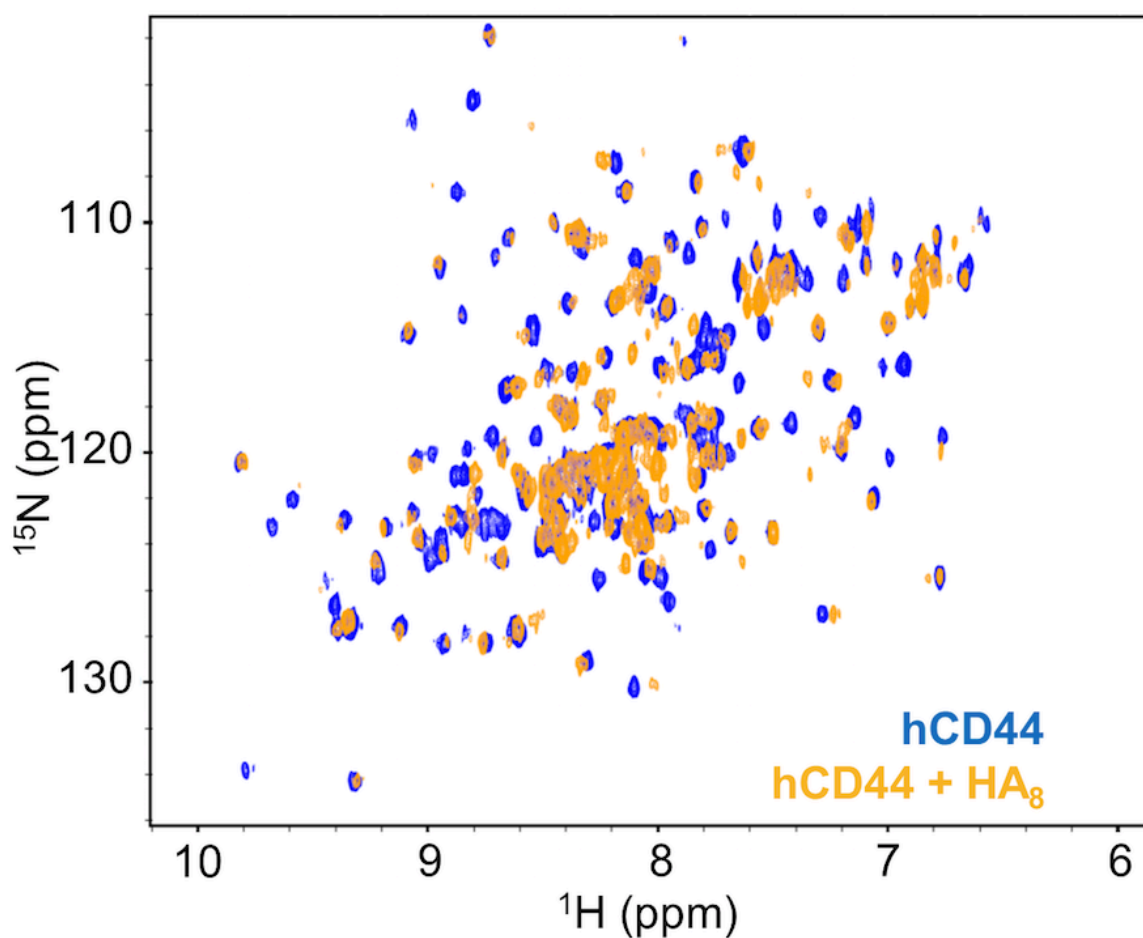

Figure 32 2D [^{1}H,^{15}N]-sofast-HMQC of 20 μM hCD44(21-178) in the apo form (blue) and with a 11-fold molar excess of HA$_8$ (orange).

Similarly, widespread perturbations were observed in the 2D [^{1}H,^{15}N]-sofast-HMQC spectrum of hCD44(21-178) (20 μM) in presence of HA$_8$ (220 μM) (**Figure 32**), again suggestive of binding and conformational rearrangements upon complex formation. Finally, we used Isothermal Titration Calorimetry (ITC) to quantify the dissociation constant between HA$_8$ and our construct (**Figure 33**).
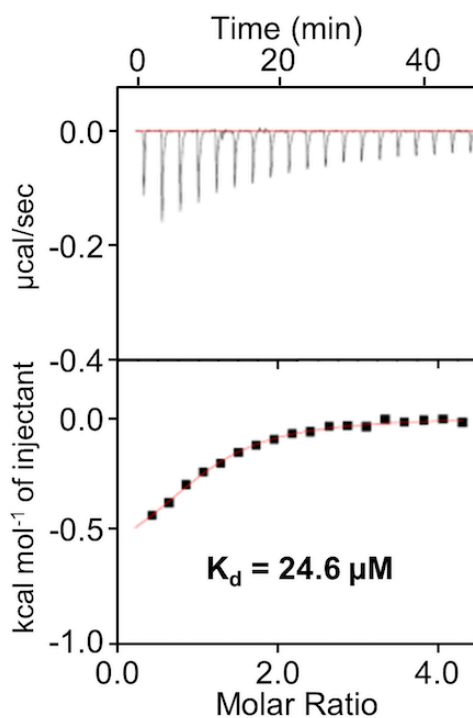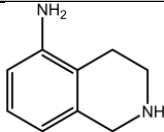
Figura 33 Isothermal Titration Calorimetry (ITC) for hCD44(21-178) titrated with HA_8. The K_d obtained is 24.6 μM. The relatively low enthalpy of binding (ΔH= -0.7 Kcal/mol) is justified by the weak binding.

We obtained a K_d value of 24.6 μM, which is in agreement with previously reported values in similar studies by ITC[60], and by SPR[59]. These relatively low binding affinities are expected for the *E. coli* expressed non-glycosylated form of hCD44(21-178), and the relatively low molecular weight version of HA used in the binding experiments. Nonetheless, our data collectively suggested that our construct was properly folded and retained the ability to bind to small HA derived oligosaccharides. As such, our CD44 construct was suitable for further studies to evaluate the binding properties of several putative CD44 binding agents.

Validation of putative binders of hCD44

Over the past several years, a variety studies reported on numerous putative CD44 binding agents, ranging from monoclonal antibodies, to peptides and to small organic molecules (**Table 4**). Surprisingly, we found that with the notable exception of HA_8 and antibodies, none of these previously reported agents bound appreciably to hCD44(21-178) as described below.

Table 4

Putative CD44 binders		*References and comments*
Natural Ligand		
HA₈ : [GlcNAc-GlcUA]₄	[64] [43] [44] [63]	K_d = 24.6 µM by ITC[a]
Antibodies		
DF1485	[70]	From Santa Cruz Biotechnolgy K_d << 400 nM by NMR[b]
Roche		Patent Number: WO 2011095498 A1[c]
Peptides		
A6 Peptide: Ac-KPSSPPEE-NH₂	[70] [49]	In phase 2 for CLL
A5G27: H-RLVSYNGIIFFLK-NH₂	[71]	Laminin α5 peptide
Ac-SRPQGPFL-NH₂	[72]	From blade I of MMP-9
Fragment		
Compound 3:	[59]	Reported K_d by SPR is 0.9 mM No appreciable binding by NMR or ITC[e]

(Peptides section right column note: No direct bind to recombinant hCD44(21-178) by NMR[d])

[a]Isothermal Titration Calorimetry (ITC) data as reported in Figure 1. The data are in agreement with similar measurements from ref. [60].

[b]No K_d was reported in literature for this antibody. By NMR spectroscopy with recombinant ¹⁵N-hCD44(21-178) we estimated a K_d value << 400 nM.

[c]This antibody is not commercially available. We tested peptides regions cited in the patent as putative CD44 binding elements. However, under the reported experimental condition, none of these peptides bound to hCD44(21-178) significantly.

[d]Using ¹⁵N-hCD44(21-178) at 20 µM we could not detect significant binding for these peptides when tested at a concentration up to 500 µM.

[e]In [¹H,¹⁵N]-sofast-HMQC correlation spectra with ¹⁵N-hCD44(21-178) at 20 µM, compound 3 did not show appreciable binding up to 20mM. In addition, the compound was not able to significantly displace the binding of HA_8 to CD44 at 55 mM by ITC (see text and supporting information).

HCAM (DF1485), from Santa Cruz Biotechnology, is a mouse anti-CD44 monoclonal antibody of about 90-95 kDa. The 2D [^{1}H,^{15}N]-sofast-HMQC of hCD44(21-178) at 10 µM is reported in **Figure 34** in absence (blue) and in presence (magenta) of the DF1485 antibody at 1:1 stoichiometry.
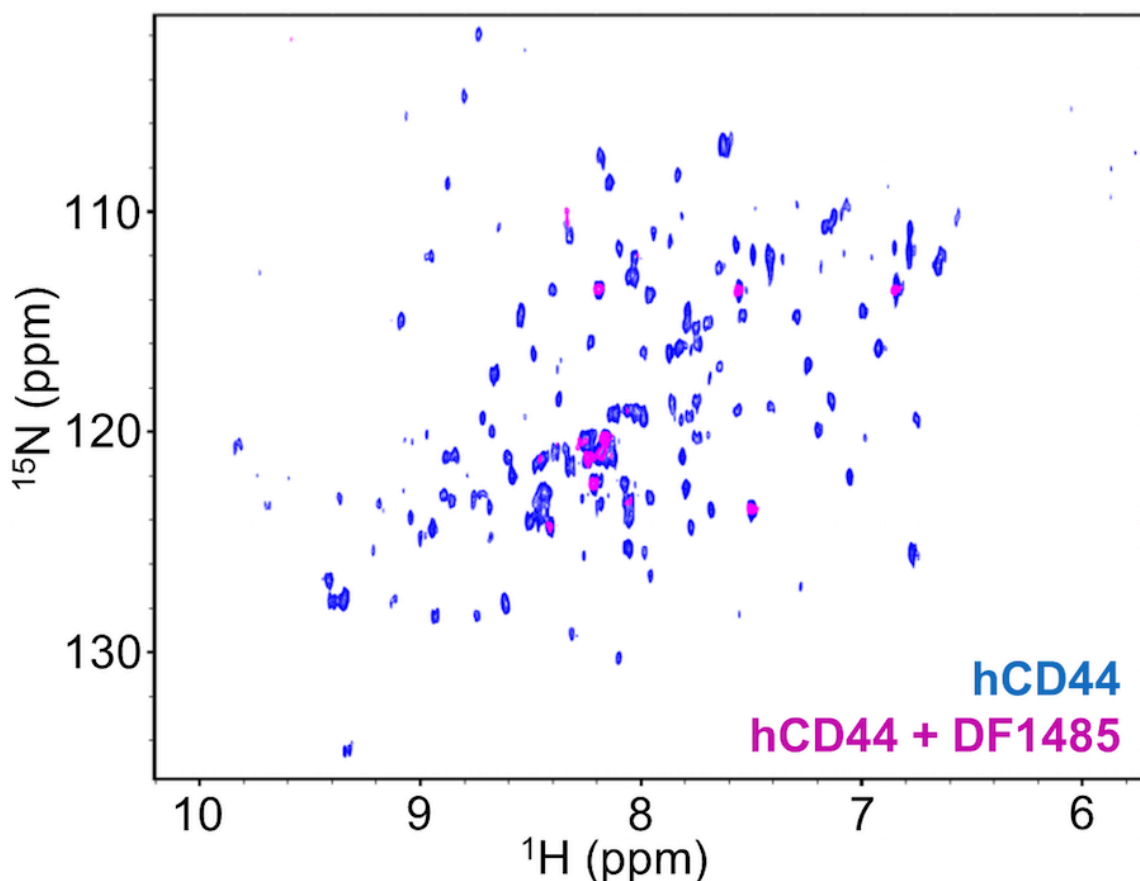

Figure 34 *CD44 antibody DF1485*. 2D [^{1}H,^{15}N]-sofast-HMQC spectra of 10 µM ^{15}N-hCD44(21-178) recorded in absence (blue) and in presence (magenta) of 10 µM DF1485 antibody. Due to the high molecular weight of the antibody-CD44 complex, as a result of the binding, the signals of the protein are broadened beyond detection. After the addition of 500 µM A6 peptide the antibody is not displaced, corroborating the findings of panel B that A6-peptide does not appreciably bind to hCD44(21-178).

Complex formation was evident by the extensive line broadening beyond detection, typical of a complex of large molecular weight (> 100 kDa). Similar data were obtained at 5 µM hCD44(21-178) and 4 µM DF1485 concentration, indicating an upper limit for the K_d of the complex of ~ 400 nM (**Table 4**). A second antibody

was recently reported (patent number US 20130224108 A1), however it is not commercially available, hence could not be tested.

The most advanced putative ligand of hCD44 is the A6-peptide (**Table 4**), that is currently in phase 2 clinical trial in patients with Chronic Lymphocytic Leukemia (CLL) and Small Lymphocytic Lymphoma (SLL)[49](ClinicalTrials.gov Identifier: NCT02046928). Similar to what we reported for HA$_8$ and DF1485, in **Figure 35** are shown the superimposed 2D [^{1}H,^{15}N]-sofast-HMQC spectra of ^{15}N-hCD44(21-178) (20 mM) in absence (blue) and in presence (green) of the A6-peptide (500 μM).
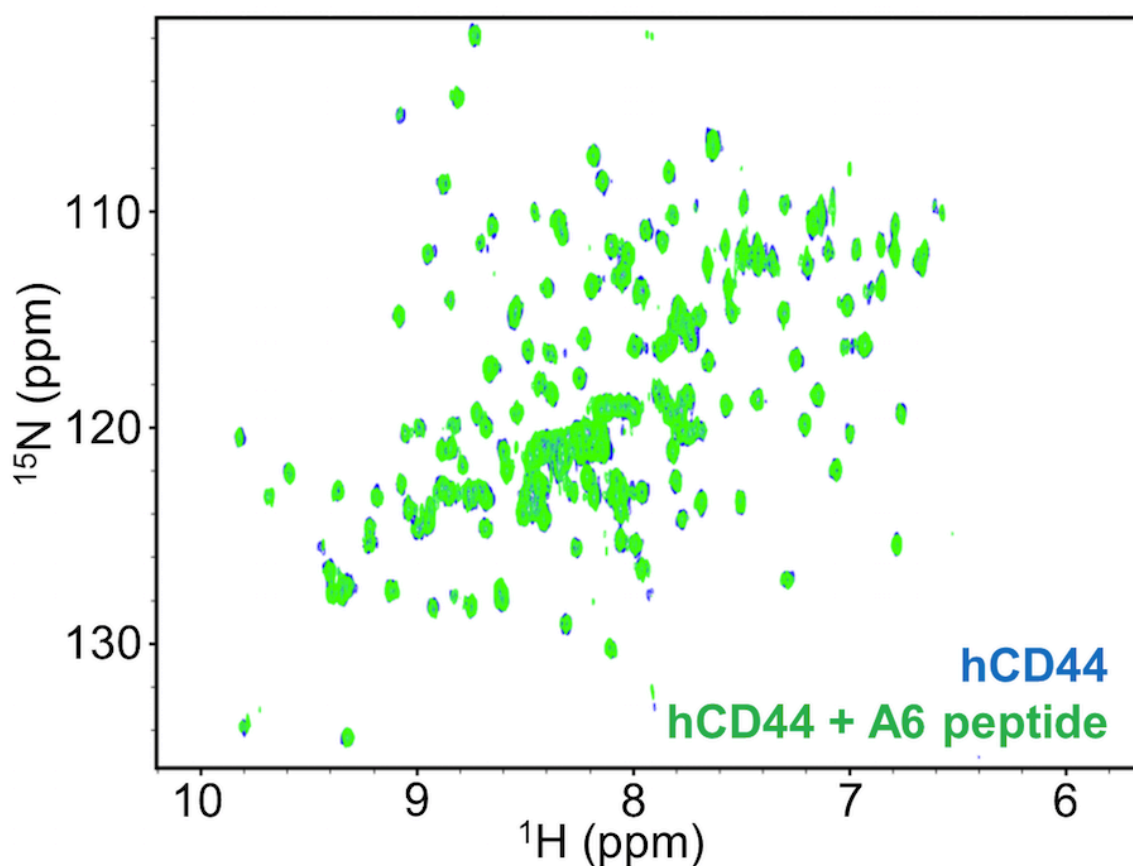
Figure 35 2D [^{1}H,^{15}N]-sofast-HMQC spectra of 20 μM ^{15}N-hCD44(21-178) in absence (blue) and in presence (green) of 500 μM A6 peptide. No appreciable binding is detected under those experimental conditions.

The superposition revealed no significant changes in chemical shifts upon titration of the ligand, suggesting the absence of appreciable binding under our experimental conditions. The absence of direct interactions between A6-peptide and

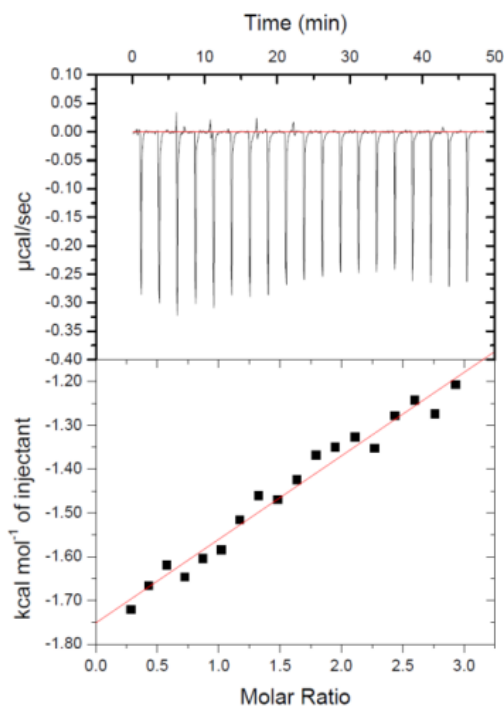recombinant non-glycosylated hCD44(21-178) was also confirmed by ITC (**Figure 36**).

Figure 36 Isothermal Titration Calorimetry (ITC) of hCD44(21-178) titrated with A6. There is no evidence of binding in our experimental condition.

These results are in agreement with what was found very recently by *Liu et al.*[73] using a sensitive Surface Plasmon Resonance (SPR) binding assay. Because the activity of A6 against CD44 was probed in cellular studies using DF1485[70], we also investigated the possibility of a direct interaction between the antibody and A6. For these studies we synthesized a ^{13}C labeled A6 peptide by introducing a ^{13}C methyl in its N-terminal acetyl group (^{13}C-A6 peptide). Subsequently we recorded 2D [^{1}H,^{13}C]-HSQC spectra of 5 µM ^{13}C-A6-peptide in absence and in presence of 4 µM DF1485 antibody (**Figure 37**). Unlike CD44, no signal broadening or chemical shift perturbations could be detected for ^{13}C-A6 in presence of DF1485. Hence, it is possible that A6 bound to CD44 only when it is fully glycosylated and/or that it recognized a CD44 region that is outside the HA binding domain.
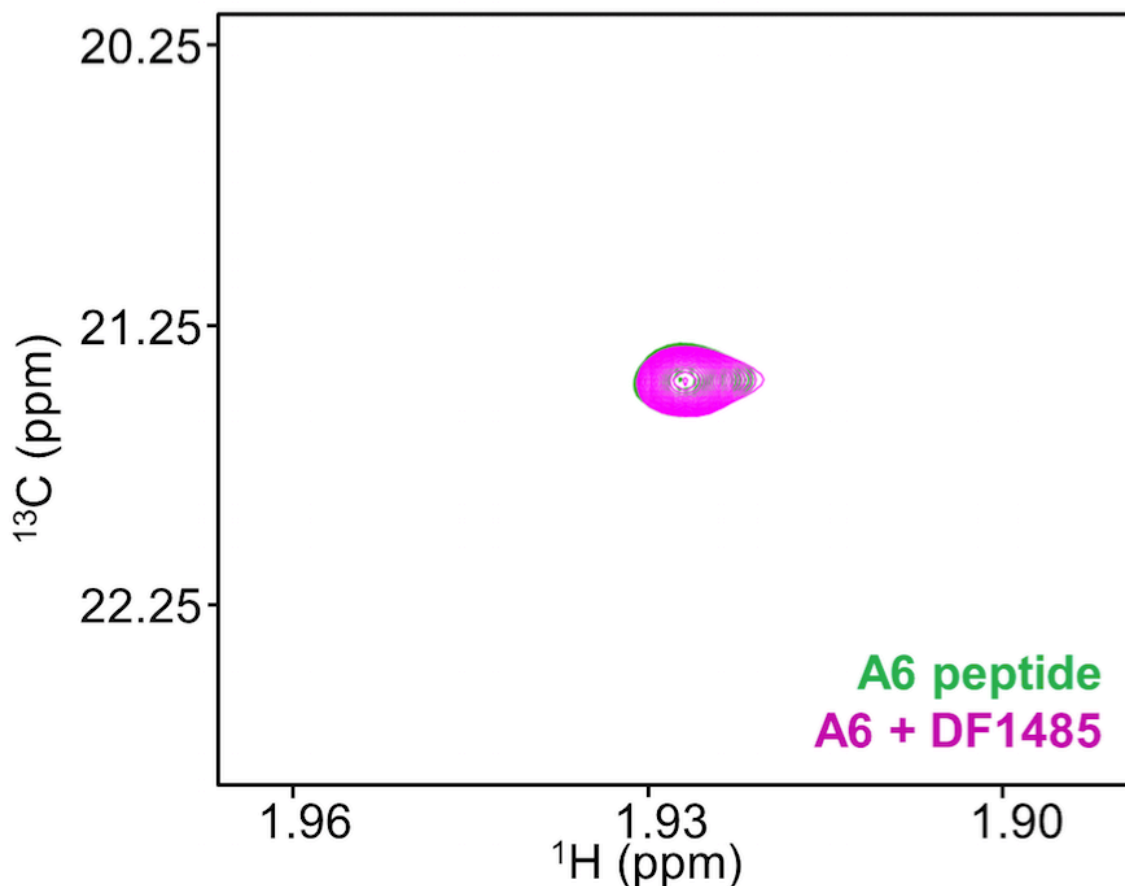
Figure 37 2D [^{1}H,^{13}C]-HSQC spectra of 5 μM ^{13}C-A6 peptide in absence (green) and in presence (magenta) of 4 μM DF1485 antibody. The signal of the ^{13}C-labeled acetyl methyl of the peptide is not perturbed by the presence of the antibody, indicating the absence of significant interactions between the A6-peptide and the antibody.

Because of the possible therapeutic potential of A6, other laboratories pursued studies aimed at identifying other putative CD44 binding peptides. *Hibino et al.*[71] found that the laminin α5 synthetic peptide A5G27 (**Table 4**) inhibits cell migration, invasion and angiogenesis by binding the glycosaminoglycans on CD44. However, the peptide presented limited solubility for binding studies, and again it is an unlikely binder for the non-glycosylated version of hCD44. Accordingly from the 2D [^{1}H,^{15}N]-sofast-HMQC spectra of 20 μM ^{15}N-hCD44(21-178) with 200 μM of A5G27 there was no evidence of binding.

Based on evidences related to the interactions between the Matrix Metalloproteinase-9 (MMP-9) and CD44[74] [75] [76], *Dufour et al.[72]* suggested that MMP-9 could be involved in a heterodimer formation with CD44 and in particular that the interactions could take place between CD44 and an 8-amino acid peptide (SRPQGPFL) constituting the blade I of the PEX domain of MMP-9. This peptide was purchased acetylated at the N-terminus and amidated at the C-terminus (**Table 4**) and was tested at a concentration of 400 μM against 20 μM ^{15}N-hCD44(21-178). However, also in this case, under these experimental conditions, the 2D [^{1}H,^{15}N]-sofast-HMQC spectra of hCD44(21-178) indicated no significant binding of this peptide to our CD44 construct.

Recently *Liu et al.*[59] performed a fragment screening of the Maybridge Ro3 Diversity Library with the goal of identifying initial fragment hits. The authors of this work screened the fragments using immobilized CD44 HABD in a SPR based assay. The most potent fragment reported (compound 3 in **Table 4**) was crystallized (PDB 4MRG). To validate the binding of this fragment, initially we recorded 2D [^{1}H,^{15}N]-sofast-HMQC spectra of 20 μM ^{15}N-hCD44(21-178) (**Figure 38**) in absence (blue) and in presence (red) of 2 mM of compound 3. Because under these conditions there was no evidence of binding, we increased the concentration of the fragment up to 20 mM. At this concentration we detected large perturbations in the chemical shift of the ^{15}N-hCD44(21-178) resonances, but after further investigation we noticed that the high concentration of compound 3 had also increased the pH of the solution from the value of 6.7 to about 8. After bringing the pH to the initial value of 6.7, the signals on the 2D [^{1}H,^{15}N]-sofast-HMQC spectrum of ^{15}N-hCD44(21-178) in presence of 20 mM compound 3 matched the spectrum of the apo form, indicating the absence of binding in these experimental conditions (**Figure 39**).
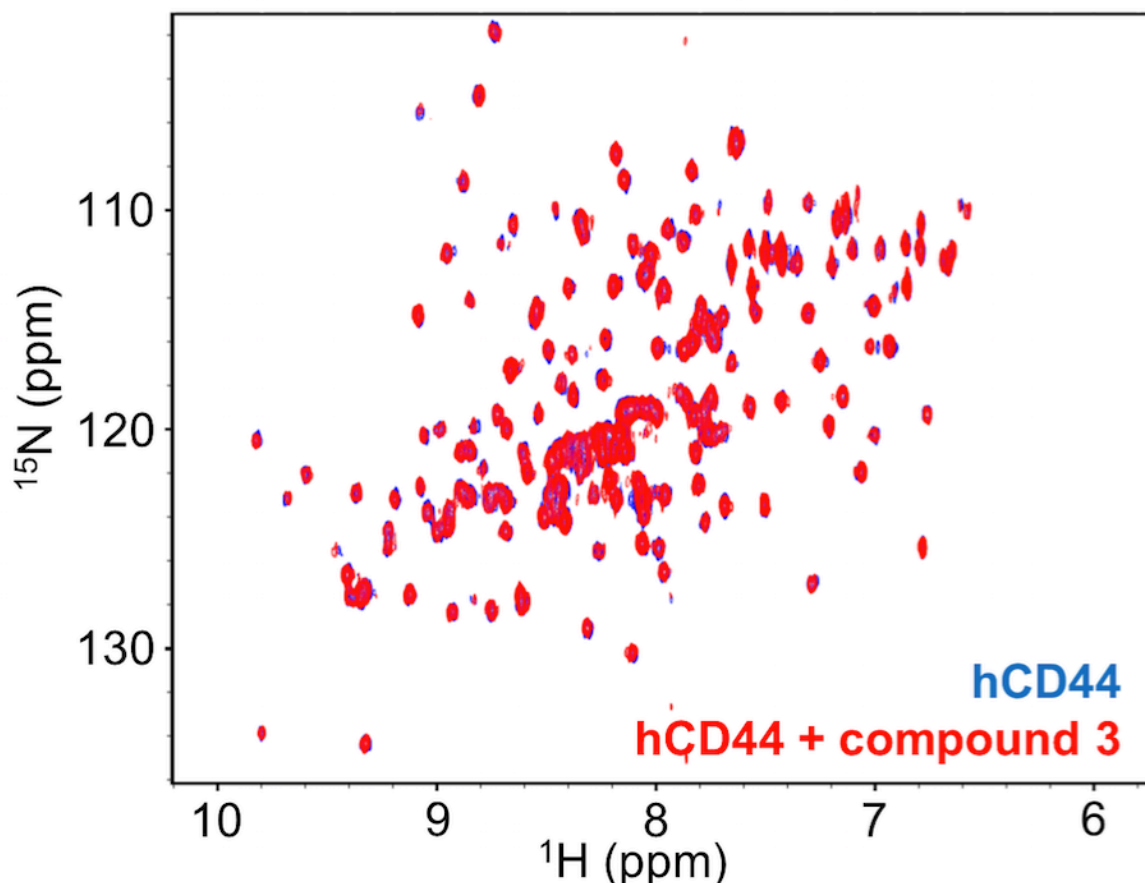
Figure 38 *Compound 3*[59]. 2D [^{1}H,^{15}N]-sofast-HMQC spectra of 20 μM ^{15}N-hCD44(21-178) recorded in absence (blue) and in presence (red) of 2 mM compound 3 [59]. No binding is detected. The absence of significant interactions is confirmed by Isothermal Titration Calorimetry in a competition assay in which hCD44(21-178) is incubated with a 750-fold molar excess of compound 3 and subsequently the binding of HA$_8$ is probed. The K$_d$ detected for HA$_8$ under this condition was 37.5 μM with a ΔH= -0.83 Kcal/mol, hence not significantly different than the binding of HA$_8$ to hCD44(21-178) in absence of compound 3 (**Figure 40**).

In addition, we performed a displacement assay using ITC. The dissociation constant between HA$_8$ and hCD44(21-178) was determined in absence and in presence of a 750-fold molar excess of compound 3. The K$_d$ detected in presence of such large excess of compound 3 was 37.5 μM (**Figure 40**), hence similar to the value obtained in absence of the fragment (**Figure 33**), suggesting that the affinity of HA$_8$ for hCD44(21-178) was not significantly affected by the presence of compound 3.
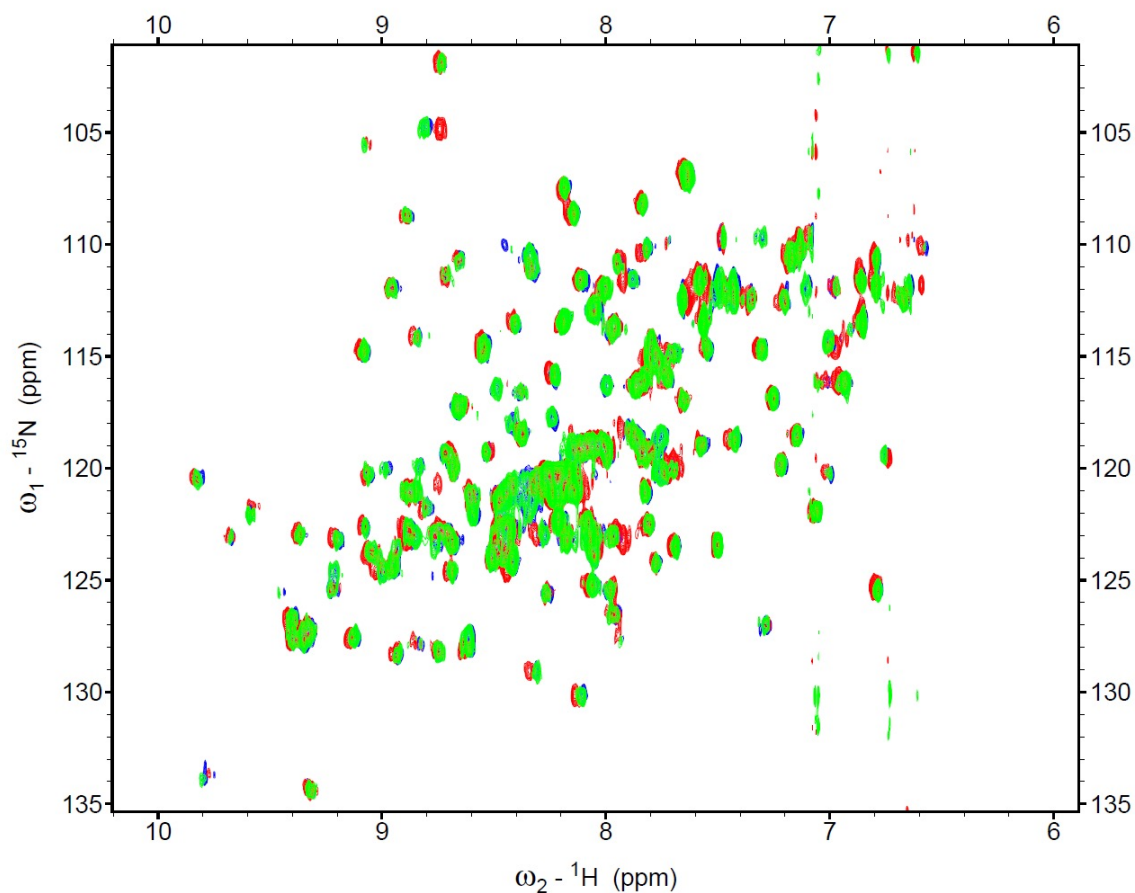
Figure 39 2D [^{1}H,^{15}N]-sofastHMQC NMR spectra of 20 μM ^{15}N-hCD44(21-178) in its apo form (blue), in presence (red) of 20 mM of compound 3[59] with the pH increase at about 8, and in presence (green) of 20 mM of compound 3[59] at the correct pH. At the pH of 6.7 (green spectrum) there are no evidences of binding.
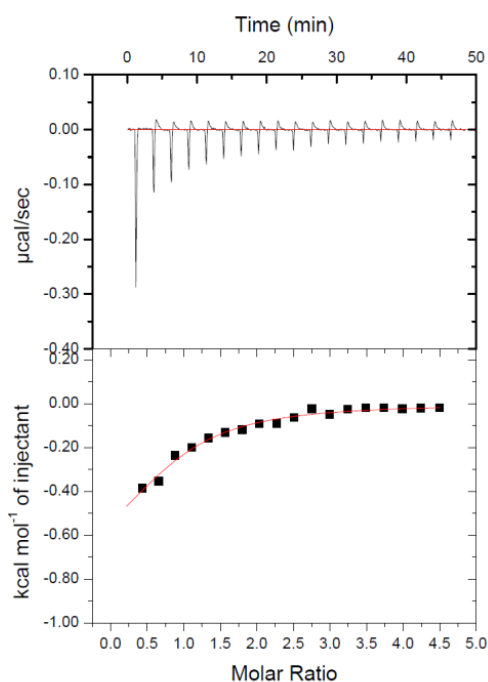

Figure 40 Isothermal Titration Calorimetry (ITC) of hCD44(21-178) in presence of 750-fold molar excess, titrated with HA$_8$. The K$_d$ of 37.5 μM is not significantly different than the K$_d$ in absence of the compound 3, suggesting that this fragment not significantly compete with HA$_8$.

Fragment based ligand discovery (FBLD) by NMR

In the pursue of possible novel agents that bind to CD44 we performed a fragment screening campaign using 1D ^{1}H-aliphatic and 2D [^{1}H,^{15}N]-sofast-HMQC of ^{15}N-hCD44(21-178) as detection methods[77]. Hence we tested a library of 500 fragments from Maybridge (Fisher Scientific). To maximize the sensitivity of the binding assay, we used ^{15}N-hCD44(21-178) at 20 μM while each fragment was tested at 400 μM. To reduce the number of samples, fragments were tested in pools of 10, and mixtures containing hits were further deconvoluted. Hence, of the 50 mixtures tested only 2 mixtures presented significant chemical shift perturbation in both 1D ^{1}H-aliphatic and 2D [^{1}H,^{15}N]-sofast-HMQC spectra. Further deconvolutions of these mixtures were accomplished by testing individual compounds at 400 μM against 20 μM ^{15}N-hCD44(21-178). Of these resulting 20 samples, two structurally related hits produced significant chemical shift perturbations in both 1D and 2D NMR spectra of ^{15}N-hCD44(21-178) (**Figure 41**).
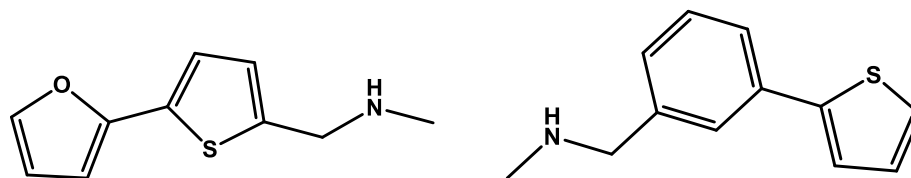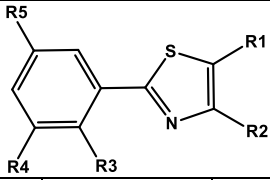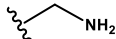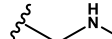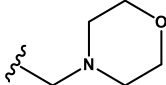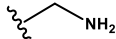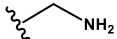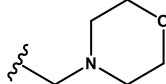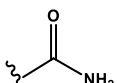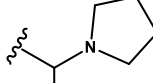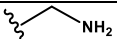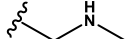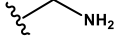

Figure 41 The two initial hits came from the 500 fragment screening (Maybridge).

Initial fragment evolution studies were performed using commercially available analogues of the two hits (**Table 5**). These studies resulted in a validated hit, compound 131B6 (**Figure 42**). Interestingly, the perturbations induced by 131B6 in the 1D ^{1}H-aliphatic NMR spectrum of hCD44(21-178) (**Figure 42**) closely resemble the perturbations induced by HA$_8$ as reported in **Figure 31**.

Table 5 Reported fragments tested against hCD44(21-178), grouped according to structural homology in thiazoles, thiophenes and furans. Rank ordering according to the following criteria: +++, when shifts of HD1 L135, and HD2 L135 peaks are larger than 7 Hz; ++, when shifts of HD1 L135, and HD2 L135 peaks are in between 3 Hz and 7 Hz; when shift of HD1 L135, and HD2 L135 peaks smaller than 3 Hz; -, no shift of the peaks HD1 L135, and HD2 L135. All the shifts refer to the chemical shift difference to the apo for spectrum.

THIAZOLES						
ID	**R1**	**R2**	**R3**	**R4**	**R5**	**Ranking by NMR**
131B5	(CH2NH2)	-H	-H	(OCH3)	-H	-
131B6	-H	(CH2CH2NHCH3)	-H	-H	-H	+++
131B7	-H	(morpholine ethyl)	-H	-CH$_3$	-H	-
131B8	(CH2NH2)	-H	(OCH3)	-H	-H	-
131B9	-H	(CH2CH2NH2)	-H	-CH$_3$	-H	-
131B10	-CH$_3$	-CH$_3$	-H	-H	(morpholine ethyl)	-
131B11	-H	(CH2C(O)NH2)	-H	-H	(pyrrolidine)	-
131B12	-H	-H	-H	-H	(CH2CH2NH2)	+
131C1	-H	-H	-H	(CH2CH2NHCH3)	-H	++
131C2	-H	(CH2CH2NH2)	-H	-H	-H	++
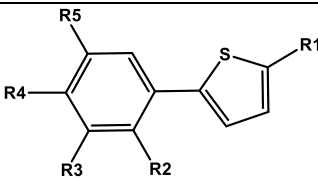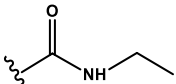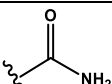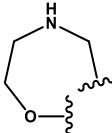
THIOPHENES						
ID	**R1**	**R2**	**R3**	**R4**	**R5**	**Ranking by NMR**
131C3	(C(O)NHEt)		(piperidine HN)	-H	-H	-
131C4	(C(O)NH2)		(piperidine HN)	-H	-H	-
131C5	-CH$_3$	-H	(OCH3)	(oxazepane)		-

74

THIOPHENES

(core structure: R5, R4, R3, R2-substituted phenyl ring connected to a thiophene bearing R1)

ID	R1	R2	R3	R4	R5	Ranking by NMR
131C6	-H	-H	[–O–CH₃ (–CH₂OCH₃)]	[1,4-oxazepane ring, NH, O]		-
131C7	-CH₃	-H	[–CH₂OH]	[1,4-oxazepane ring, NH, O]		+
131C8	-H	-H	[–CH₂OH]	[1,4-oxazepane ring, NH, O]		++
131C9	[–C(=O)N(CH₃)₂]	[piperidine ring, HN]		-H	-H	-
131C10	-Cl	-H	[pyridin-2(1H)-one ring, HN, C=O]		-H	-
131C11	[–SO₂NH₂]	[piperidine ring, HN]		-H	-H	-
131C12	[–C(=O)OH]	-H	[–C(=O)NH–CH₂CH₃]	-H	-H	-

FURANS

(core structure: R3, R2-substituted phenyl ring connected to a furan bearing R1)

ID	R1	R2	R3	Ranking by NMR
131D1	[–C(=O)OH]	-H	[pyrrolidine-N–CH(CH₃)–]	-
131D2	[–C(=O)N(CH₃)₂]	[piperidine ring, HN]		-
131D3	[–C(=O)–N(pyrrolidine)]	[piperidine ring, HN]		-

75

Figure 42 NMR titration of 131B6 (structure on the upper panel) against 20 μM of ^{15}N-hCD44(21-178) using 1D ^{1}H-aliphatic spectra. The spectra of the apo protein are depicted in blue, while the spectra in purple to light-blue were collected in presence of the fragment in increasing concentration starting from 1 mM to 15 mM respectively. The resonances corresponding to the most perturbed residues are labeled.

However, chemical shift perturbations in 2D [^{1}H,^{15}N]-sofast-HMQC spectrum (**Figure 43**), indicated that only a distinct subset of resonances were perturbed by the li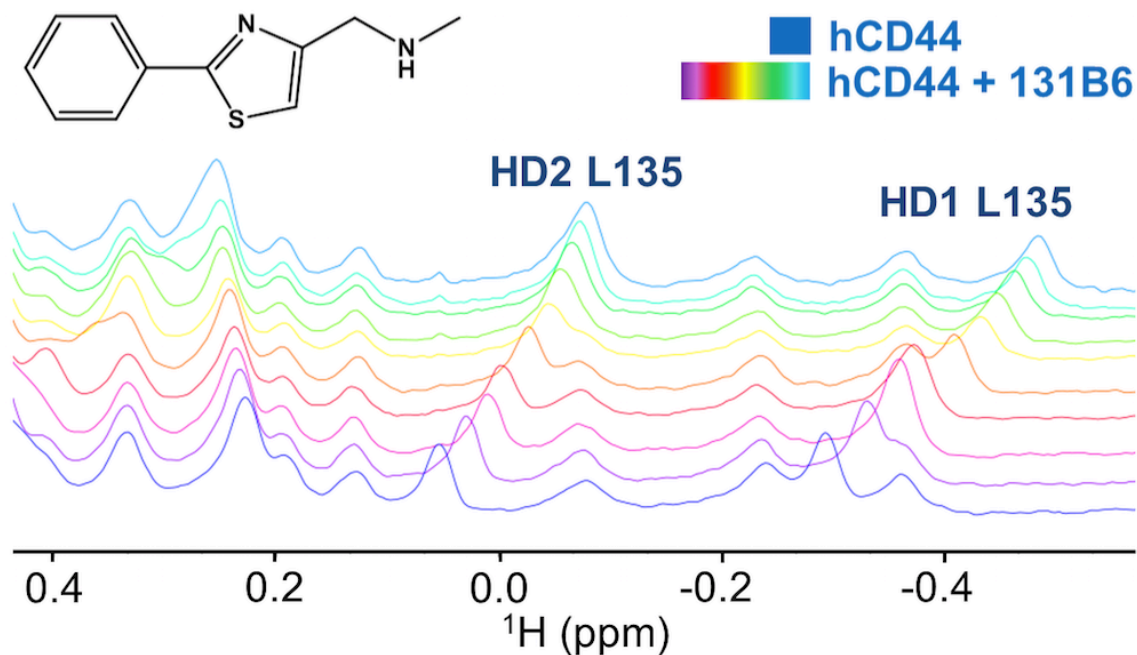gand. Intriguingly, these changes map in a region of the structure of hCD44(21-178) that is opposite to the reported HA$_8$ binding site (**Figure 44**). The dissociation constant calculated using the chemical shift perturbation from the 2D [^{1}H,^{15}N]-sofast-HMQC spectra results in a K_d of 7.43 mM (**Figure 45**), hence in the same affinity range of the fragments previously reported[59]. Further evaluations of the binding of 131D6 to CD44 where enabled by measurements of intermolecular NOEs (**Figure 46**).
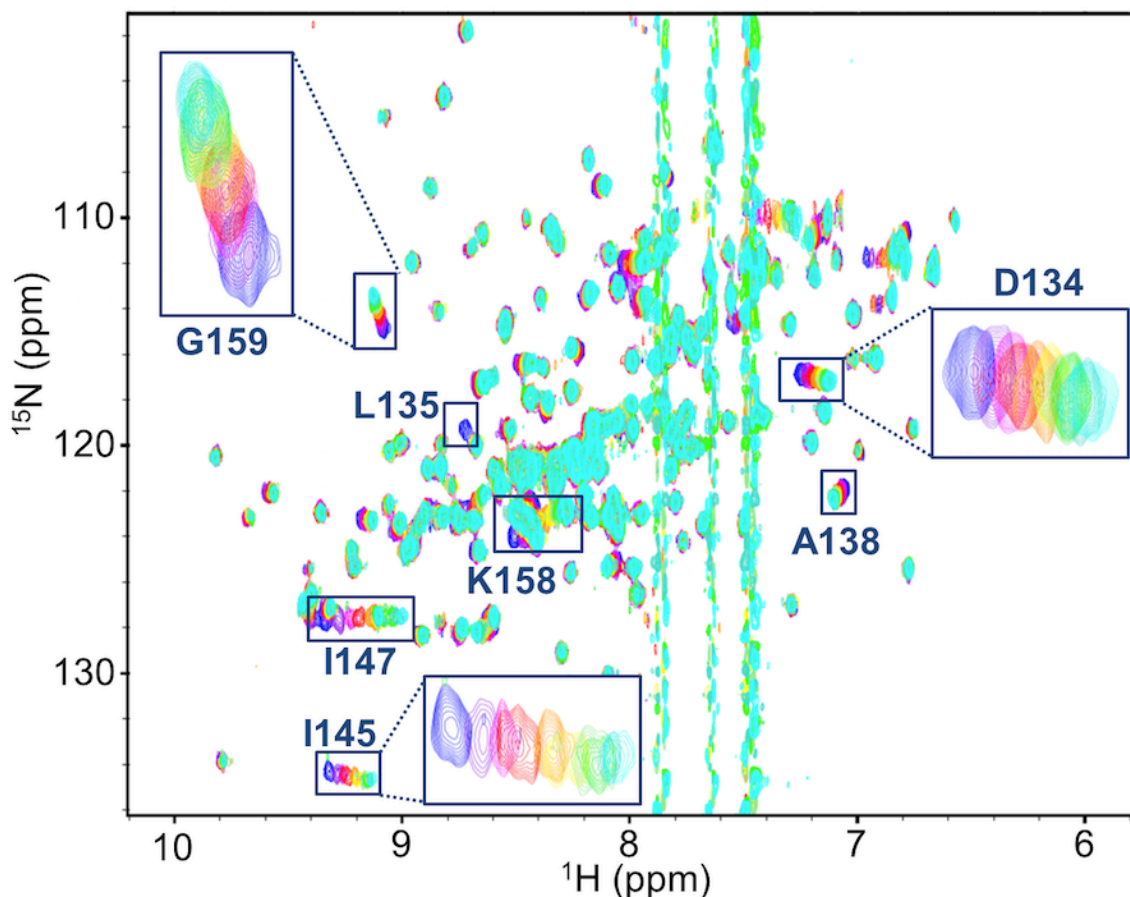
Figure 43 NMR titration of 131B6 against 20 μM of ^{15}N-hCD44(21-178) using 2D ^{15}N-sofasHMQC spectra. The spectra of the apo protein are depicted in blue, while the spectra in purple to light-blue were collected in presence of the fragment in increasing concentration starting from 1 mM to 15 mM respectively. The resonances corresponding to the most perturbed residues are labeled.

Hence, 2D [^{1}H, ^{1}H] NOESY spectra were recorded with 100 μM of hCD44(21-178) in absence (blue) and in presence (red) of 500 μM of 131B6. In the spectrum of the complex, we could clearly observe in the aromatic region of the spectra, the presence of a cross-peak that connected the resonances of the proton of the thiazol (131B6) with the resonances of the δ-protons of the leucine 135 in hCD44(21-178), suggesting a spatial proximity of the compound to this aminoacid (< 5 Å). However, in a subsequent NMR displacement assay the presence of 131B6 did not prevent the binding of HA$_8$, suggesting that under these conditions, these two binding events can exist simultaneously (**Figure 47A and 47B**). Nonetheless, the screening revealed a

possible binding pocket on the surface of CD44 that may be 'druggable' and that may have functional implications in cell.
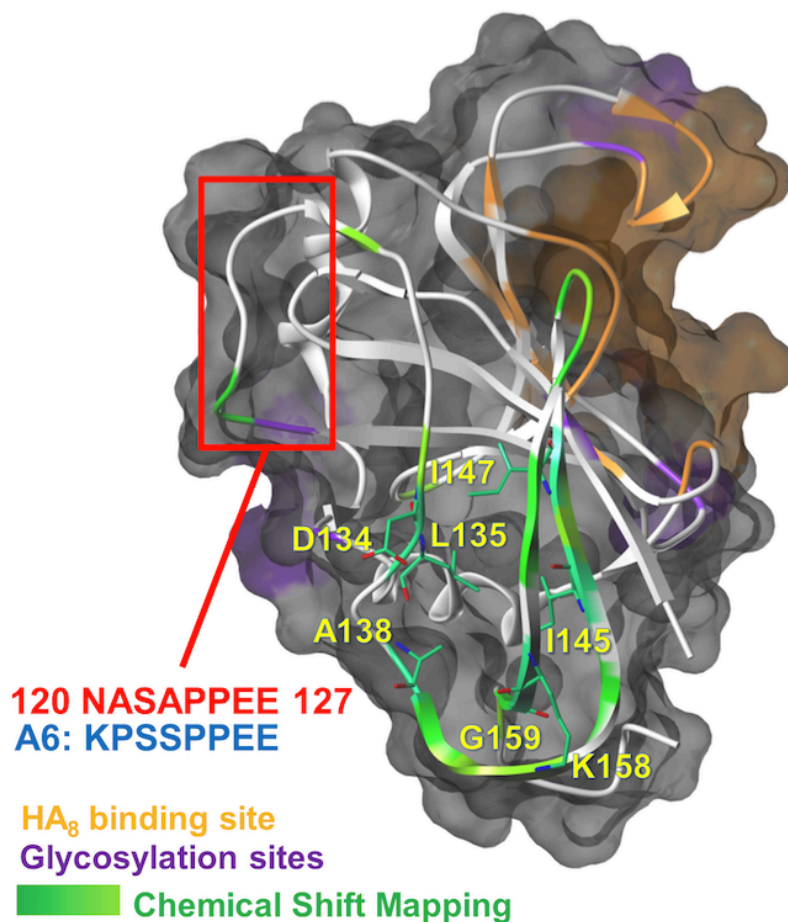
Figure 44 Semitransparent molecular surface representation of the homology model of hCD44 built with SWISS-MODEL [65] [66] [67] [68] using the structure of mCD44 in complex with HA$_8$ (PDB 2JCR) as template. The HA$_8$ binding pocket is depicted in orange while in purple are highlighted the putative glycosylation sites. The Chemical Shift Perturbations induced by 131B6 are mapped on the secondary structure (ribbon): $\Delta\delta$ > 0.14 ppm in green; $\Delta\delta$ < 0.14 ppm in light green. The most perturbed residues are labeled and localized in a putative back pocket opposite from the HA$_8$ binding pocket. The red square indicates the portion of hCD44 with sequence homology with A6 peptide.
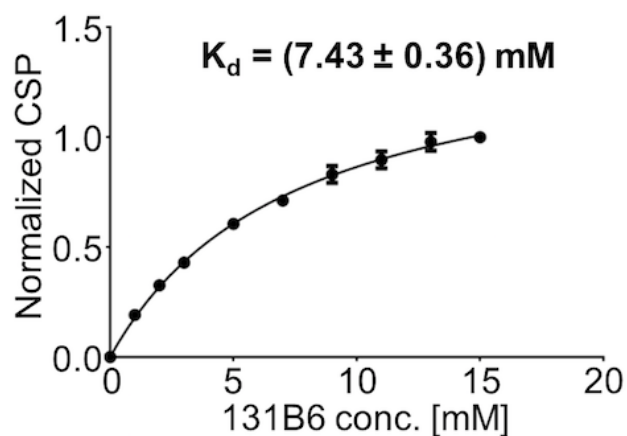
Figure 45 Determination of K_d of 131B6 using the chemical shift perturbation titrations from the 2D [^{1}H,^{15}N]-sofasHMQC experiments; K_d = 7.43 mM)
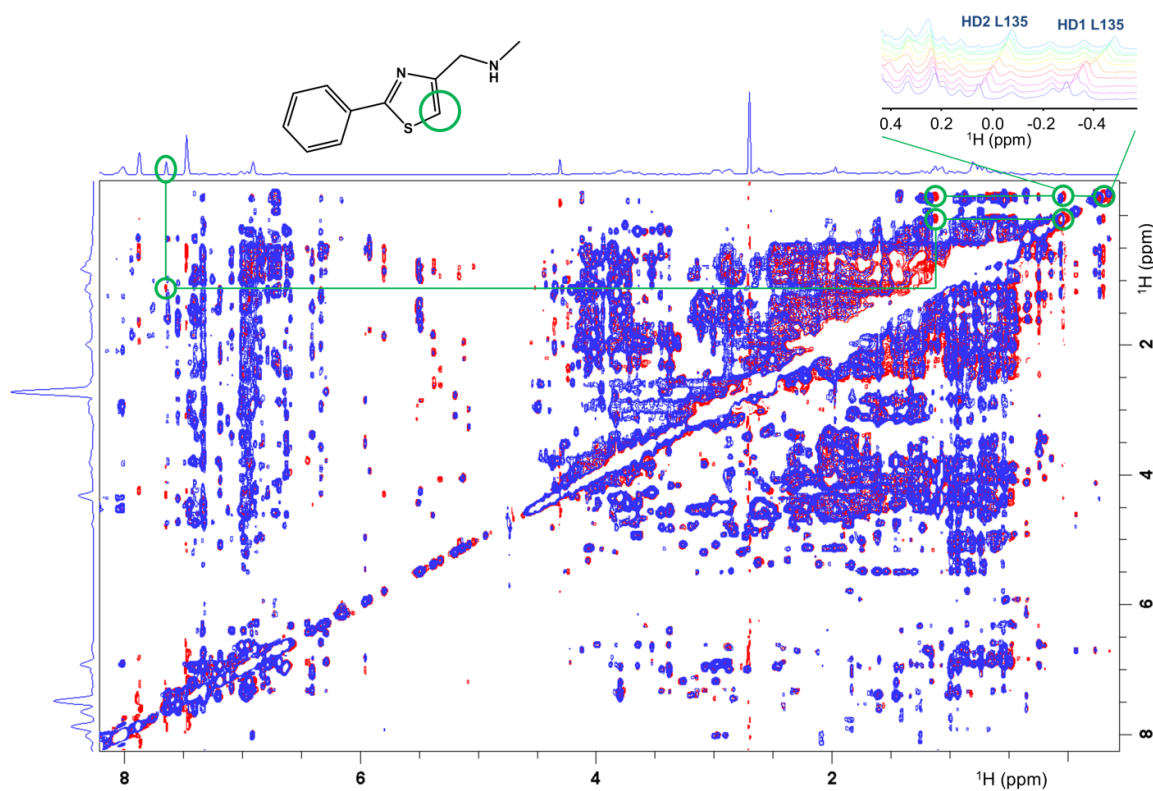
Figure 46 2D [^{1}H,^{1}H]-NOESY NMR spectra of 100 μM hCD44(21-178) in absence (blue) and in presence (red) of 500 μM of compound 131B6. In green is depicted the correlation between the proton of the thiazole (compound 131B6) and probably the β-protons of the leucine 135 of hCD44(21-178).
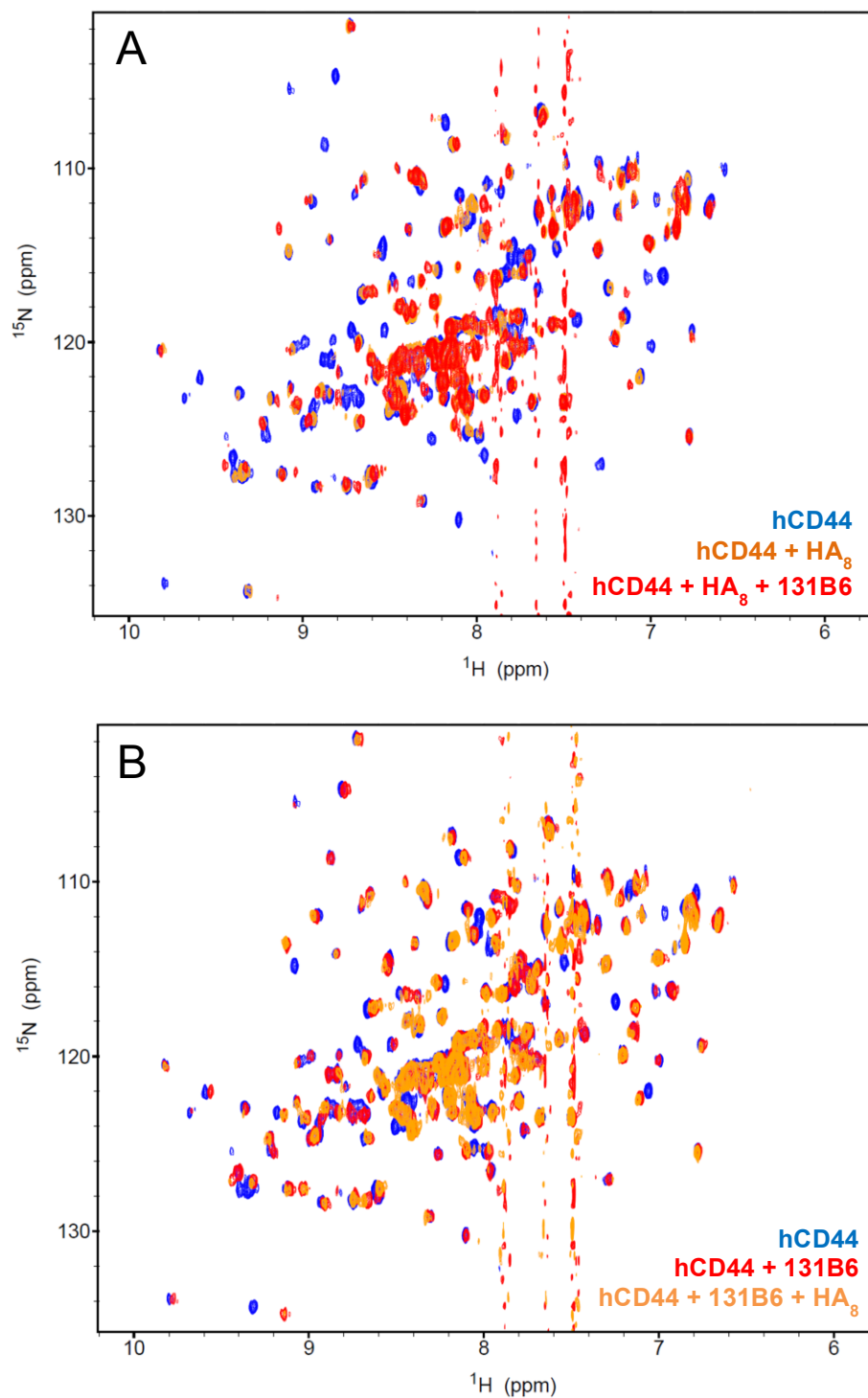
Figure 47 (A) In blue 20 µM ^{15}N-hCD44(21-178) in its apo form. In orange the spectrum of the protein in presence of 80 µM HA$_8$, and in red the spectrum after the addition of 15 mM of the compound 131B6. (B) In blue 20 µM ^{15}N-hCD44(21-178) in its apo form. In red the spectrum of the protein in presence of 15 mM of the compound131B6, and in orange the spectrum after the addition of 8=60 µM HA$_8$.

Methods and Materials

Protein expression and purification

The human CD44 HA Binding Domain (HABD) residues 21-178 was expressed in *Escherichia Coli* and purified from insoluble inclusion body as described previously [61]. The cDNA of hCD44(21-178) was cloned into the vector pET19b and transformed into *E. coli* strain BL21(DE3) gold pLysS (Novagen). The overexpression of the protein was obtained growing the transformed cells in LB medium at 37°C with 100 µg/L of ampicillin until reaching an OD_{600} of 0.6 followed by induction with 0.4 mM IPTG. After OVT incubation at 20°C, cells were harvested by centrifugation. The inclusion body was isolated from the pellet and solubilized in 8 M urea in presence of 1mM dithiothreitol (DTT) at 4°C OVT. The denatured recombinant hCD44(21-178) was subsequently refolded dropping the supernatant solution at 600 µL/h using a syringe pump by 200-fold dilution into buffer containing 250 mM L-Arginine, 100 mM tris-HCl pH = 8, 2 mM reduced glutathione and 1 mM oxidized glutathione. After 48h at 4°C stirring, the solution was ultrafiltrated through Amicon YM10 membranes. The monomeric protein was purified through size exclusion chromatography using a HiLoad 26/60 Superdex 75 prep grade column. For the expression of the ^{15}N-labeled protein used for all heteronuclear NMR experiments, the procedure was the same, but the bacteria were grown in M9 minimal medium supplemented with trace elements, 0.09%v/v glycerol and 0.5 g/ L $^{15}NH_4Cl$ for the labeling.

Reagents

HA$_8$ was purchased from Iduron (UK). The antibody DF1485 was from Santa Cruz Biotechnology (Dallas, TX). Compound 3 [59] was obtained from Matrix Scientific (Columbia, SC, USA). The peptides A6 (Ac-KPSSPPEE-NH$_2$), A5G27 (H-RLVSYNGIIFFLK-NH$_2$), the peptide from the blade I of the PEX domain of MMP-9 (Ac-SRPQGPFL-NH$_2$), the three peptides from the light chain of the Hoffmann-La Roche antibody (Ac-SRYWMS-NH$_2$, Ac-EVNPDSTSINYTPSLKD-NH$_2$, and Ac-PNYYGSRYHYYAMDY-NH$_2$), and the three from the heavy chain (Ac-RASQDINNYL-NH$_2$, Ac-YTSRLHS-NH$_2$, and Ac-QQGSTLPFT-NH$_2$) were purchased from Innopep (San Diego, CA). ^{13}C-A6 peptide was synthesized in our laboratory. 131B6 and all the compounds in **Table 5** were purchased from Hit2Lead ChemBridge Corporation (San Diego, CA) except for compound 131C1 from Maybridge (Fisher Scientific) and compound 131B12 from Acros Organics.

^{13}C-A6 peptide synthesis

^{13}C-A6 peptide was synthesized using a fmoc solid-phase synthesis using the Rink amide resin. For the coupling reaction each amino acid was dissolved into 5 mL of dry DMF containing 6 equivalents of Oxima pure, DIC and DIAE. Each coupling reaction was carried shaking the resin for 3 hours. After the reaction the resin has been washed 3 times with 8 mL DMF followed by 3 washings with DCM and 3 washings with DMF. The deprotection of the terminal amino acid was achieved by adding to the resin-bound peptide a 20% piperidine solution in DMF (1 mL and 4 mL respectively) for 40 minutes twice. After the last coupling step, the N-terminus was acetylated with 5 mL of dry DMF solution containing 0.05 equivalents of DMAP, 3 equivalents of DIPEA and 2 equivalents of ^{13}CH$_3$COCl (Cambridge Isotope

Laboratories, Inc.). The cleavage from the resin and the removal of all the side chains protecting groups was performed by shaking at RT with a cleavage cocktail composed by 94% TFA, 2% phenol, 2% TIPS, and 2% of water for 5.5 hours. TFA was subsequently removed under reduced pressure using a rota-vapor and the peptide was subjected to few cycles of precipitation in diethyl ether and centrifugation before OVT drying in high vacuum. The peptide was purified by a HPLC Breeze system from Waters Co. using preparative reverse phase column to get a purity level > 95% . The final peptide was characterized by NMR and MALDI-TOF spectrometry.

Isothermal Titration Calorimetry (ITC)

The ITC measurements were performed with a ITC200 calorimeter from Microcal (Northampton, MA, USA) at 23°C. For the direct experiments, hCD44 (73 μM) was titrated with a solution containing 1.6 mM HA_8 (**Figure 33**) or 1 mM A6 peptide (**Figure 36**). For the competition assay (**Figure 40**) the protein was incubated with a 55 mM solution of compound 3 and then titrated with a solution of 1.6 mM HA_8. All experiments were conducted in buffer containing 50 mM phosphate, 150 mM NaCl at pH = 6.7 in presence of 5% DMSO. The data were analyzed using Origin software provided by Microcal.

NMR Spectroscopy

All the NMR spectra were acquired on a 600 or a 700 MHz Bruker Avance spectrometers equipped with TCI cryoprobes and z-shielded gradient coils. The data were processed and analyzed using TOPSPIN 2.1 (Bruker Biospin, MA) and Sparky 3.1 (University of California, San Francisco, CA). In general, 1D ^1H experiments were

acquired using 128 scans with 2048 complex data points, 2D [^{1}H,^{15}N]-sofast-HMQC experiments using 48 scans with 2048 and 128 complex data points in the ^{1}H and ^{15}N dimensions, respectively, and the 2D [^{1}H,^{13}C]-HSQC using 64 scans with 4096 and 128 complex points in the ^{1}H and ^{13}C dimensions, respectively. All samples were acquired at 300 K in buffer containing 50 mM phosphate, 150 mM NaCl, and 0.02% NaN$_3$, pH = 6.7.

The 2D [^{1}H,^{1}H]-NOESY NMR spectra were recorded using 72 scans with 2048 and 256 complex data points in the direct and indirect dimensions, respectively, with a NOE mixing time of 0.2 seconds. The Saturation Transfer Difference (STD) NMR spectra were recorded using 128 scans and saturating for 2 seconds at -0.28 ppm or at -10 ppm for the reference spectra.

The dissociation constant (K$_d$) was calculated monitoring the chemical shift perturbations in the 2D [^{1}H,^{15}N]-sofast-HMQC spectra of hCD44 (20 μM) due to the presence of increasing concentrations of compound 131B6 (1, 2, 3, 5, 7, 9, 11, 13, and 15mM). The weight average perturbations induced on backbone amide resonances corresponding to residues D134, I145, I147 and G159 were determined using the following equation: [78]

$$\Delta\delta = \sqrt{(\Delta_d^1 H)^2 + (0.17 * (\Delta_d^{15} N)^2}$$

Subsequently, K$_d$ values were obtained by fitting these averaged chemical shift perturbation data into the following equation: [79] [80]

$$\Delta\delta_{obs} = \Delta\delta_{max} \frac{(K_d + [L]_0 + [P]_0) - \sqrt{(K_d + [L]_0 + [P]_0)^2 - 4[P]_0[L]_0}}{2[P]_0}$$

where $\Delta\delta_{obs}$ is the chemical shift perturbation value observed at each point of the titration, $\Delta\delta_{max}$ is the maximum chemical shift change of the fully complexed protein, and $[L]_0$ and $[\ ]_0$ are the total concentrations of compound and protein.

Screening by NMR

The Maybridge library of 500 fragments was assembled in 50 mixtures containing 10 fragments each in d_6-DMSO. 1D ^{1}H and 2D [^{1}H,^{15}N]-sofast-HMQC experiments were recorded for each mixture at the final concentration of 4 mM (400 μM individual fragment concentration) against 20 μM ^{15}N-hCD44(21-178). The spectra were recorded in 50 mM phosphate buffer pH 6.7, 150 mM NaCl and 0.02% NaN$_3$ with a 2% of d_6-DMSO. The hit mixtures were subsequently deconvoluted testing each fragment individually in the same experimental conditions. For the HTS by NMR of combinatorial libraries approach [81] was tested a four-position combinatorial library (Ac-XXXX-NH$_2$) using 19 amino acids (all the natural amino acid except for cysteine) as building blocks. In this way, 76 (19×4) mixtures each containing 6,859 (19×19×19) individual peptides with one amino acid fixed at one position were synthesized and tested. Hence, we were able to test 130,321 combinations of natural amino acid (19×19×19×19) just by synthesizing and testing 76 mixtures. With the same approach, we tested a tri-peptide library assembled using 46 natural and non-natural amino acid as building blocks (Pepscan Presto BV, Lelystad, The Netherlands) maintaining one amino acid fixed at the first position. In this way we tested 2,116 peptides (1×46×46) just testing 1 mixture.

Discussion

CD44, via its interactions with its ligand, the hyaluronic acid, [38] [43] [44] (**Figure 30**), plays an important role in tumor cell migration and tumor metastasis[42] and as such it is expressed on the surface of several types of tumors [38]. Despite its demonstrated relevance as viable drug target in oncology, further assessments on the biological role of CD44 in the onset and progression of cancer is hampered by the availability of suitable pharmacological tools. CD44 indeed is known to be not an easily 'druggable' target [59], due to the absence of a clear well-defined ligand binding pocket on its surface (**Figure 30**). In addition, there are several post-translational modifications that occur in cell, mostly glycosylation, that are critical for the activity of CD44 [47]. However, the exact glycosylation state of the protein cannot be easily determined and reproduced in a test tube. Nonetheless, the recombinant form of the HA binding domain of CD44, retaining some of its binding affinity for HA, may provide a valid platform onto which to attempt to derive CD44 binding agents. Therefore we expressed the hCD44(21-178) as it was reported[43] [61] by using that such construct which possesses a proper folding for binding HA. Accordingly, we demonstrated the construct's ability to bind HA_8 using both solution nuclear magnetic resonance spectrometry (NMR) (**Figure 31 and 32)** and isothermal titration calorimetry (ITC) techniques (**Figure 33**). However and surprisingly, when subsequently we tried to validate the binding of the other reported putative ligands of CD44 (**Table 4**), these resulted inactive under our experimental conditions, with the sole exception of the antibody DF1485 (**Figure 34**). The results for these activities are summarized in **Table 4**. Several possible causes may be attributable to the lack of binding for these agents to CD44. For example, the peptide A5G27 has a fairly low solubility in our

conditions it is also known to bind the glycosaminoglycans on CD44[71]. We could not test the antibody patented by Hoffman-La Roche and the University of Miami (Patent number US 20130224108 A1) because it is not commercially available, but based on the information disclosed in the patent, we could select and test six peptides of which three belongs at the heavy chain of the antibody, and the other three to the light chain. Unfortunately, also all these peptides tested in isolation didn't seem to bind significantly the protein (see Methods and Materials). Another interesting peptide was reported in literature [72] based on the interactions between CD44 and the PEX domain of MMP-9 (**Table 4**). Nonetheless, as mentioned, also this peptide under our experimental conditions didn't show significant binding to CD44. Next, we focused our studies on the binding properties of the A6 peptide given that it is in phase 2 clinical trials for CLL and SLL patients[49](ClinicalTrials.gov Identifier: NCT02046928). The detailed mechanism of action of A6 has not been defined, but studies on metastatic disease suggested that it functions through a CD44-mediated pathway[70]. While cellular studies indicated a possible direct interaction between CD44 and A6, we could not observe any significant binding of A6 to ^{15}N-labeled sample of hCD44(21-178) in 2D [^{1}H,^{15}N]-sofast-HMQC (**Figure 35**) and ITC (**Figure 36**) experiments. However and intriguingly, the amino acid sequence of A6 (**Table 4**) exhibits marked homology with a linear sequence within CD44 (120-NASAPPEE-127) (**Figure 44**)[70]. This has prompted us to experimentally verify for a possible recognition of the A6 peptide by the CD44 antibody DF1485, given that such antibody was used in cellular assays to study the mechanism of action of A6. To facilitate these binding studies, we synthesized a ^{13}C-labeled version of A6 and collected 2D [^{13}C,^{1}H]-HSQC spectra of such agent in absence and in presence of the antibody DF1485.

However, under these experimental conditions, no appreciable binding was detected (**Figure 37**) between A6 and the antibody.

Subsequently, we also tested a recently discovered small organic molecule that was reported to bind the recombinant CD44 HA binding domain with millimoar affinity[59]. We obtained such compound 3 (Matrix Scientific, Columbia, SC) and tested it at 2 mM (**Figure 38**) and 20 mM (**Figure 39**) against ^{15}N-hCD44(21-178) by 2D [^{1}H,^{15}N]-sofast-HMQC. Several significant chemical shift perturbations in the 2D [^{1}H,^{15}N]-sofast-HMQC spectrum of hCD44(21-178) were observed initially in presence of 20 mM of compound 3. However, the perturbations were suspiciously widespread for such small ligand. Upon further investigations we noticed that such high concentration of compound 3 led to a sizable increase of the pH, likely due to the compound itself (having a primary ammine) or an impurity. Indeed, recording again the spectrum after correcting the pH we observed that there was no significant evidence of binding (**Figure 39**). The 2D [^{1}H,^{15}N]-sofast-HMQC of ^{15}N-hCD44(21-178) at 20 μM measured in absence and presence of 2 mM compound 3 are reported in **Figure 38.** Furthermore we monitored the ability of compound 3 to displace HA$_8$ by ITC. In this experiment, ITC curves for the binding of HA$_8$ to CD44 were collected in absence (**Figure 33**) and in presence of up to 55 mM of compound 3. Also by this displacement assay, the presence of compound 3, even at such high concentration, did not significantly displace the binding between CD44 and HA$_8$ (**Figure 40**). We conclude, unfortunately that the putative binding and HA$_8$ displacement reported by *Liu et al.*[59] using a less sensitive SPR assay, may have been a result of pH changes induced by the compound.

In a follow-up study however *Liu et al.*[73] reported that in resolving the crystal structure of hCD44(20-178), electron density of an unidentified peptide docked into a

novel hydrophobic binding pocket situated on the opposite side of the HA binding site. From the electron density it was possible only to identify a valine residue as the first amino acid of this unknown peptide located into the small binding pocket. Based on this observations we decided to perform a further screening campaign using the HTS by NMR approach[81]. Aimed at identifying a possible CD44 binding peptide, we subsequently decided to screen a four-position combinatorial library composed by 19 natural amino acids (the 20 natural amino acids except cysteine) as building blocks. In this way we tested all the possible combinations of tetra-peptides, including peptides with the valine at the N- and C- terminal. However, no viable hits were detected using 2D [^{1}H,^{15}N]-sofast-HMQC among the tested 76 mixtures (see Methods). At this point we tested with the same approach a tri-peptide combinatorial library assembled using 46, natural and non-natural, amino acids as building blocks. Based on the evidence that only the first amino acid stays inside the new binding pocket, we decided to screen only the peptides with the first position fixed. Again we didn't find any peptide interacting with the target protein using 2D [^{1}H,^{15}N]-sofast-HMQC.

Finally we decided to conduct a fragment screening using NMR as detection technique aimed at finding possibly new hits able to bind hCD44. By screening of a 500 small molecule fragment library (Maybridge) we obtained two structurally related initial hits (**Figure 41**). Starting from these initial hits, we tested other analogs (**Table 5**) that resulted in compound 131B6 (**Figure 42**). Analyses of data resulting from both the 1D ^{1}H aliphatic (**Figure 42**) and 2D [^{1}H,^{15}N]-sofast-HMQC NMR experiments (**Figure 43**), concluded that the residues that were most perturbed were situated in a region of hCD44(21-178) forming a possible back pocket in the opposite site from where HA$_8$ binds and close but different from the pocket observed

by *Liu et al.*[73]. From the chemical shift perturbations we could obtain an average dissociation constant of K_d equal to 7.43 mM (**Figure 45**). For both, 131B6 and compound 3[59], we also collected Saturation Transfer Difference (STD) spectra. First we tested as positive control HA_8 that as expected gave a significant STD effect (**Figure 48A**). However, when we ran the same experiment with 131B6 (**Figure 48B**), or compound 3 (**Figure 48C**), there was no significant STD effect in either compound. These data further confirm that for fragments with fairly low affinity ranging from 1 to 10 mM a protein based experiment (as the 2D [^{1}H,^{15}N]-sofast-HMQC or the 1D ^{1}H-aliphatic NMR experiments) is more sensitive and robust than a ligand based approach (as the STD)[77]. To further confirm the binding mode of this fragment hit and to obtain further structural information about its possible binding pose we recorded 2D [^{1}H,^{1}H]-NOESY NMR spectra (**Figure 46**) for hCD44 (21-178) in absence and in presence of compound 131B6. In the presence of this compound, we observed a cross-peak between the proton of the thiazole of 131B6 and the β-protons of leucine 135 that, from both 1D and 2D NMR experiments, was the most perturbed residue by the compound. These data further confirmed that 131B6 is close in space (< 5 Å) to the leucine 135. Binding of 131B6 to CD44 did not affect the binding of HA_8 as determined by NMR titration with 2D [^{1}H,^{15}N]-sofast-HMQC (**Figure 47A and 47B**).

An important and interesting observation is that the residues most perturbed upon the binding of 131B6 to hCD44(21-178) spatially overlap to a region that encompasses the sequence of CD44 with high homology to the A6 peptide sequence (**Figure 44**). In particular, the chemical shift of several residues within this sequence were perturbed from the 2D [^{1}H,^{15}N]-sofast-HMQC spectrum collected in presence of 131B6. As previously hypothesized, this homology between the A6 peptide and a

linear sequence within CD44 may provide an alternative explanation for the cellular and *in vivo* activity of A6 whereby A6 may function as a decoy for a yet unknown binding partner of CD44[70].
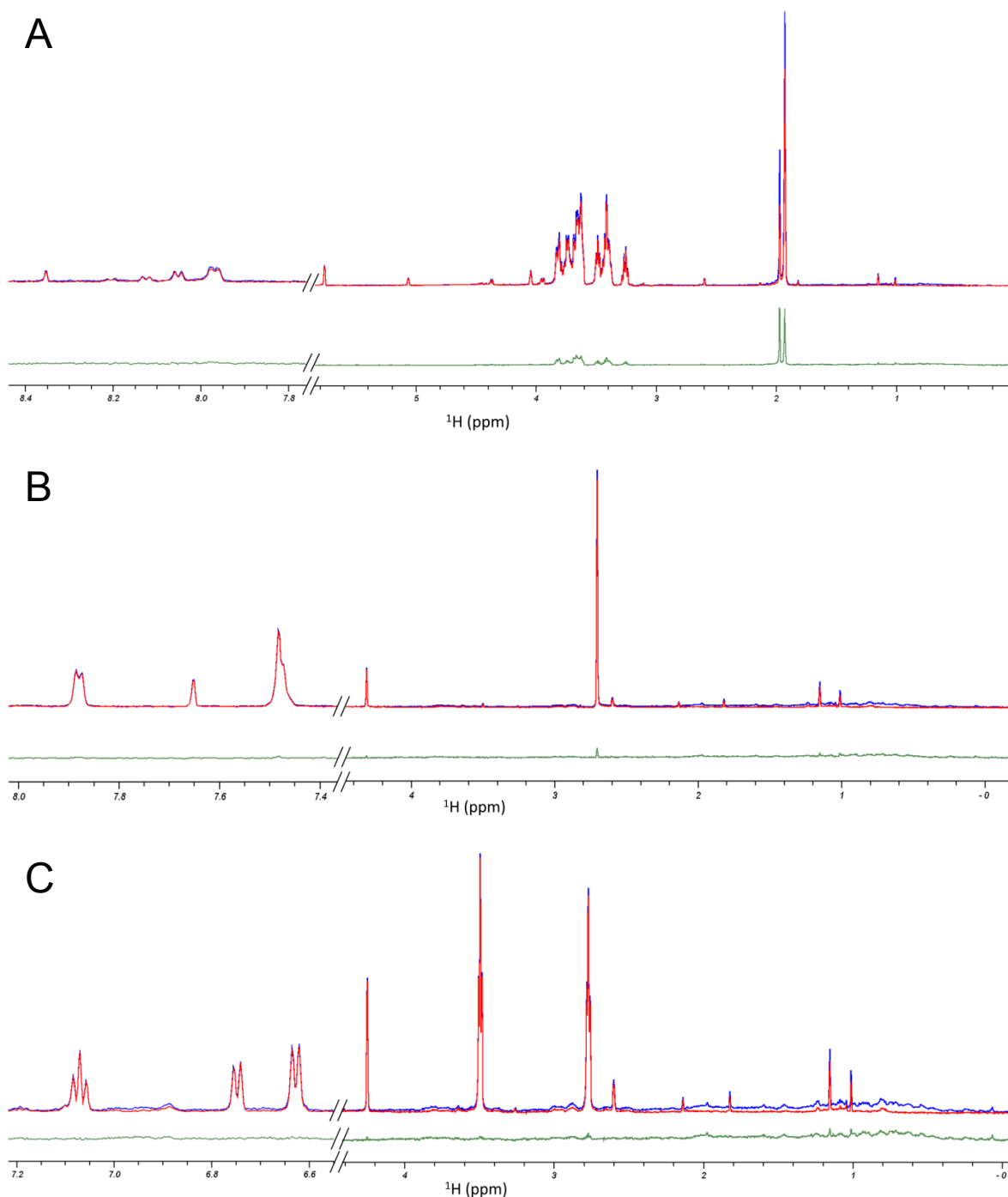

Figure 48 Saturation Transfer Difference (STD) experiments of 5 μM hCD44(21-178) in presence of 400 μM HA$_8$ (A), 400 μM compound 131B6 (B) and, 400 μM compound 3 (C). In all the spectra in blue is represented the off-resonance spectrum, in red the on-resonance and, in green the STD.

If this hypothesis is correct, ligands binding in proximity of this loop of CD44 may exert a biological activity, rendering 131B6 a potentially interesting fragment for further hit maturations and subsequent cellular assays. Hence our studies have potentially significant implications on the possible mechanism of A6, and may provide a new rationale for targeting a possible binding pocket on CD44.

Conclusions

CD44 has revealed to be a potentially very interesting receptor for pharmacological intervention targeting tumor metastasis. However, further research in the validation of CD44 as possible drug target is hampered by the lack of suitable pharmacological agents. In this article we characterized a variety of putative CD44 binding agents. With the exception of an antibody and HA_8, none of these agents, including the clinical peptide A6, showed appreciable binding to a recombinant HA binding domain construct of CD44. However, an NMR guided fragment screening revealed potentially interesting fragment hit 131B6. The compound bound to CD44 in an area that is opposite to the HA binding site and is composed by a loop region of CD44 (loop $\beta6$-$\beta7$) that presents a high homology with the anti-CD44 peptide A6. Based on these data, it is tempting to speculate that A6 may function as a decoy for an unidentified CD44 binding protein that uses loop $\beta6$-$\beta7$ for such interactions[70]. Therefore, further optimizations of 131B6 may result in small molecule agents that similar to A6 may impair the function of CD44.

References

[1]   S. W. Fesik, *Nature reviews. Cancer* **2005**, *5*(11), 876-885.

[2]   R. J. Youle, A. Strasser, *Nature reviews. Molecular cell biology* **2008**, *9*(1), 47-59.

[3]   A. M. Petros, E. T. Olejniczak, S. W. Fesik, *Biochim Biophys Acta* **2004**, *1644*(2-3), 83-94.

[4]   J. C. Reed, *Blood* **2008**, *111*(7), 3322-3330.

[5]   D. R. Green, G. Kroemer, *Science* **2004**, *305*(5684), 626-629.

[6]   X. Wang, *Genes & development* **2001**, *15*(22), 2922-2933.

[7]   S. W. Muchmore, M. Sattler, H. Liang, R. P. Meadows, J. E. Harlan, H. S. Yoon, D. Nettesheim, B. S. Chang, C. B. Thompson, S. L. Wong, S. L. Ng, S. W. Fesik, *Nature* **1996**, *381*(6580), 335-341.

[8]   L. Scorrano, S. J. Korsmeyer, *Biochem Biophys Res Commun* **2003**, *304*(3), 437-444.

[9]   P. Bouillet, A. Strasser, *J Cell Sci* **2002**, *115*(Pt 8), 1567-1574.

[10]   C. Borner, *Mol Immunol* **2003**, *39*(11), 615-647.

[11]   B. Fadeel, S. Orrenius, B. Zhivotovsky, *Biochem Biophys Res Commun* **1999**, *266*(3), 699-717.

[12]   S. Barelier, J. Pons, O. Marcillat, J. M. Lancelin, I. Krimm, *J Med Chem* **2010**, *53*(6), 2577-2588.

[13]   M. Pellecchia, I. Bertini, D. Cowburn, C. Dalvit, E. Giralt, W. Jahnke, T. L. James, S. W. Homans, H. Kessler, C. Luchinat, B. Meyer, H. Oschkinat, J. Peng, H. Schwalbe, G. Siegal, *Nature reviews. Drug discovery* **2008**, *7*(9), 738-745.

[14]   M. J. Harner, A. O. Frank, S. W. Fesik, *Journal of biomolecular NMR* **2013**, *56*(2), 65-75.

[15]   M. H. Kang, C. P. Reynolds, *Clin Cancer Res* **2009**, *15*(4), 1126-1132.

[16]   J. Keller, *Understanding NMR Spectroscopy*, **2002**.

[17]   G. A. Morris, R. Freeman, *Journal of the American Chemical Society* **1979**, *101*(3), 760-762.

[18]   Q. Teng, *Structural Biology: Practical NMR Applications*.

[19]   L. Mueller, *Journal of the American Chemical Society* **1979**, *101*(16), 4481-4484.

[20]   G. Bodenhausen, *Chemical Physics Letters* **1980**, *69*, 185-189.

[21]   A. Bax, R. H. Griffey, B. L. Hawkins, *Journal of Magnetic Resonance (1969)* **1983**, *55*(2), 301-315.

[22]   G. S. Rule, T. K. Hitchens, *Fundamentals of Protein NMR Spectroscopy*.

[23]   http://www.protein-nmr.org.uk, Protein NMR - A Practical Guide.

[24]   A. Viegas, J. Manso, F. L. Nobrega, E. J. Cabrita, *Journal of Chemical Education* **2011**, *88*(7), 990-994.

[25]   C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi, B. Stockman, *Journal of biomolecular NMR* **2001**, *21*(4), 349-359.

[26]   O. Korb, T. Stützle, T. E. Exner, *Swarm Intelligence* **2007**, *1*(2), 115-134.

[27]   http://www.tcd.uni-konstanz.de/research/plants.php, PLANTS - Protein-Ligand ANT System.

[28]   M. Dorigo, V. Maniezzo, A. Colorni, *IEEE Trans Syst Man Cybern B Cybern* **1996**, *26*(1), 29-41.

[29]   G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *Journal of computational chemistry* **2009**, *30*(16), 2785-2791.

[30]   N. Nikolova, J. Jaworska, *QSAR & Combinatorial Science* **2003**, *22*(9-10), 1006-1026.

[31]   A. M. Johnson, G. M. Maggiora, *Concepts and Applications of Molecular Similarity*, **1990**.

[32]   MACCS Keys; MDL Information Systems I, 14600 Catalina Street, San Leandro, CA 94577.

[33]   M. Congreve, R. Carr, C. Murray, H. Jhoti, *Drug Discovery Today* **2003**, *8*(19), 876-877.

[34]   C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Advanced Drug Delivery Reviews* **2001**, *46*(1–3), 3-26.

[35]   J. Stark, R. Powers, *Journal of the American Chemical Society* **2008**, *130*(2), 535-545.

[36]   J. C. Rathmell, C. B. Thompson, *Cell* **2002**, *109*(2, Supplement 1), S97-S107.

[37]   W. M. Gallatin, I. L. Weissman, E. C. Butcher, *Nature* **1983**, *304*(5921), 30-34.

[38]   J. Lesley, R. Hyman, P. W. Kincade, *Advances in immunology* **1993**, *54*, 271-335.

[39]   D. Naor, R. V. Sionov, D. Ish-Shalom, *Advances in cancer research* **1997**, *71*, 241-319.

[40]   H. C. DeGrendele, P. Estess, L. J. Picker, M. H. Siegelman, *The Journal of experimental medicine* **1996**, *183*(3), 1119-1130.

[41]   Y. Shimizu, G. A. Van Seventer, R. Siraganian, L. Wahl, S. Shaw, *Journal of immunology (Baltimore, Md. : 1950)* **1989**, *143*(8), 2457-2463.

[42]   U. Gunthert, M. Hofmann, W. Rudy, S. Reber, M. Zoller, I. Haussmann, S. Matzku, A. Wenzel, H. Ponta, P. Herrlich, *Cell* **1991**, *65*(1), 13-24.

[43]   R. J. Peach, D. Hollenbaugh, I. Stamenkovic, A. Aruffo, *The Journal of cell biology* **1993**, *122*(1), 257-264.

[44]   H. X. Liao, D. M. Lee, M. C. Levesque, B. F. Haynes, *Journal of immunology (Baltimore, Md. : 1950)* **1995**, *155*(8), 3938-3945.

[45]   A. Almond, *Cellular and molecular life sciences : CMLS* **2007**, *64*(13), 1591-1596.

[46]   M. Zoller, *Nature reviews. Cancer* **2011**, *11*(4), 254-267.

[47]   T. A. Brown, T. Bouchard, T. St John, E. Wayner, W. G. Carter, *The Journal of cell biology* **1991**, *113*(1), 207-221.

[48]   H. Ponta, L. Sherman, P. A. Herrlich, *Nature reviews. Molecular cell biology* **2003**, *4*(1), 33-45.

[49]   M. Finlayson, *Frontiers in immunology* **2015**, *6*, 135.

[50]   G. Borland, J. A. Ross, K. Guy, *Immunology* **1998**, *93*(2), 139-148.

[51]   S. Zhang, C. C. Wu, J. F. Fecteau, B. Cui, L. Chen, L. Zhang, R. Wu, L. Rassenti, F. Lao, S. Weigand, T. J. Kipps, *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*(15), 6127-6132.

[52]   Y. Herishanu, F. Gibellini, N. Njuguna, I. Hazan-Halevy, M. Farooqui, S. Bern, K. Keyvanfar, E. Lee, W. Wilson, A. Wiestner, *Leukemia & lymphoma* **2011**, *52*(9), 1758-1769.

[53]   I. M. Pedersen, S. Kitada, L. M. Leoni, J. M. Zapata, J. G. Karras, N. Tsukada, T. J. Kipps, Y. S. Choi, F. Bennett, J. C. Reed, *Blood* **2002**, *100*(5), 1795-1801.

[54]   S. C. Ghosh, S. Neslihan Alpay, J. Klostergaard, *Expert opinion on therapeutic targets* **2012**, *16*(7), 635-650.

[55]   Y. Guo, A. P. Mazar, J. J. Lebrun, S. A. Rabbani, *Cancer research* **2002**, *62*(16), 4678-4684.

[56]   D. D. Boyd, S. J. Kim, H. Wang, T. R. Jones, G. E. Gallick, *The American journal of pathology* **2003**, *162*(2), 619-626.

[57]   K. Mishima, A. P. Mazar, A. Gown, M. Skelly, X. D. Ji, X. D. Wang, T. R. Jones, W. K. Cavenee, H. J. Huang, *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*(15), 8484-8489.

[58]   H. Lai, G. Liu, L. Rassenti, M. Y. Choi, S. B. Howell, M. Finlayson, T. J. Kipps, *Blood* **2013**, *122*(21), 5303-5303.

[59]   L. K. Liu, B. C. Finzel, *J Med Chem* **2014**, *57*(6), 2714-2725.

[60]   S. Banerji, A. J. Wright, M. Noble, D. J. Mahoney, I. D. Campbell, A. J. Day, D. G. Jackson, *Nat Struct Mol Biol* **2007**, *14*(3), 234-239.

[61]   S. Banerji, A. J. Day, J. D. Kahmann, D. G. Jackson, *Protein expression and purification* **1998**, *14*(3), 371-381.

[62]   P. Schanda, E. Kupce, B. Brutscher, *Journal of biomolecular NMR* **2005**, *33*(4), 199-211.

[63]   P. Teriete, S. Banerji, M. Noble, C. D. Blundell, A. J. Wright, A. R. Pickford, E. Lowe, D. J. Mahoney, M. I. Tammi, J. D. Kahmann, I. D. Campbell, A. J. Day, D. G. Jackson, *Molecular cell* **2004**, *13*(4), 483-496.

[64]   A. Aruffo, I. Stamenkovic, M. Melnick, C. B. Underhill, B. Seed, *Cell* **1990**, *61*(7), 1303-1313.

[65]   K. Arnold, L. Bordoli, J. Kopp, T. Schwede, *Bioinformatics (Oxford, England)* **2006**, *22*(2), 195-201.

[66]   F. Kiefer, K. Arnold, M. Kunzli, L. Bordoli, T. Schwede, *Nucleic acids research* **2009**, *37*(Database issue), D387-392.

[67]   N. Guex, M. C. Peitsch, T. Schwede, *Electrophoresis* **2009**, *30 Suppl 1*, S162-173.

[68]   M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli, T. Schwede, *Nucleic acids research* **2014**, *42*(Web Server issue), W252-258.

[69]   M. Takeda, S. Ogino, R. Umemoto, M. Sakakura, M. Kajiwara, K. N. Sugahara, H. Hayasaka, M. Miyasaka, H. Terasawa, I. Shimada, *The Journal of biological chemistry* **2006**, *281*(52), 40089-40095.

[70]   R. S. Piotrowicz, B. B. Damaj, M. Hachicha, F. Incardona, S. B. Howell, M. Finlayson, *Molecular cancer therapeutics* **2011**, *10*(11), 2072-2082.

[71]   S. Hibino, M. Shibuya, M. P. Hoffman, J. A. Engbring, R. Hossain, M. Mochizuki, S. Kudoh, M. Nomizu, H. K. Kleinman, *Cancer research* **2005**, *65*(22), 10494-10501.

[72]   A. Dufour, S. Zucker, N. S. Sampson, C. Kuscu, J. Cao, *The Journal of biological chemistry* **2010**, *285*(46), 35944-35956.

[73]   L. K. Liu, B. Finzel, *Acta Crystallogr F Struct Biol Commun* **2014**, *70*(Pt 9), 1155-1161.

[74]   Q. Yu, I. Stamenkovic, *Genes & development* **1999**, *13*(1), 35-48.

[75]   Q. Yu, I. Stamenkovic, *Genes & development* **2000**, *14*(2), 163-176.

[76]   J. Redondo-Munoz, E. Ugarte-Berzal, J. A. Garcia-Marco, M. H. del Cerro, P. E. Van den Steen, G. Opdenakker, M. J. Terol, A. Garcia-Pardo, *Blood* **2008**, *112*(1), 169-178.

[77]   E. Barile, M. Pellecchia, *Chemical reviews* **2014**, *114*(9), 4749-4763.

[78]   B. T. Farmer, 2nd, K. L. Constantine, V. Goldfarb, M. S. Friedrichs, M. Wittekind, J. Yanchunas, Jr., J. G. Robertson, L. Mueller, *Nature structural biology* **1996**, *3*(12), 995-997.

[79]   M. Pellecchia, *Chem Biol* **2005**, *12*(9), 961-971.

[80]   C. Smet, J. F. Duckert, J. M. Wieruszeski, I. Landrieu, L. Buee, G. Lippens, B. Deprez, *J Med Chem* **2005**, *48*(15), 4815-4823.

[81]   B. Wu, Z. Zhang, R. Noberini, E. Barile, M. Giulianotti, C. Pinilla, R. A. Houghten, E. B. Pasquale, M. Pellecchia, *Chem Biol* **2013**, *20*(1), 19-33.

Ringraziamenti