Tesi Doctoral

NMR IN DRUG DISCOVERY. FROM SCREENING TO STRUCTURE-BASED DESIGN OF ANTITUMORAL AGENTS

Ricard A. Rodríguez Mias

Departament de Química Orgànica

Facultat de Química

Universitat de Barcelona

Barcelona, Juliol 2006

1 NMR METHODS IN DRUG DISCOVERY: A HANDS ON EXPERIENCE

Work presented within this first chapter was carried out in Dr Pellecchia's Lab. at the Burnham Institute -San Diego (CA)- during an internship in 2002. Dr Pellecchia's laboratory focus falls in the use and development of NMR methodologies in Drug Discovery; particularly on the design of low MW inhibitors of protein-protein interactions. At the time of my visit the most relevant targets in the laboratory were a series of proteins involved in the regulation, or rather miss-regulation, of apoptosis in tumoral cells. Inhibition of these proteins is both an interesting and challenging strategy for cancer treatment, and is currently been pursued by a number of research groups. From a practical perspective, such targets will also fulfill an educational goal as they will allow us to get first hand experience on some of the techniques outlined in the introduction.

1.1 SELECTIVE LABELING SCHEMES: BIR3-XIAP DOMAIN A CASE STUDY

1.1.1 BIR3-XIAP DOMAIN AND ITS BIOLOGICAL RELEVANCE

Apoptosis, or otherwise programmed cell death is a delicately controlled phenomenon in pluricellular organisms, with utter importance to development and tissue homeostasis. A variety of pro and anti-apoptotic proteins control this process through a fine network of signaling events; and very often, such opposing signals coexist in a dynamic equilibrium.[1, 2] This balance is altered through stimuli such as cell damage or other types of stress, and ultimately determines cellular fate. Missregulation however, has been pointed out as the cause leading to several conditions, both when balance is tilted towards excessive cell death (degenerative diseases), or disproportionate proliferation (cancer).

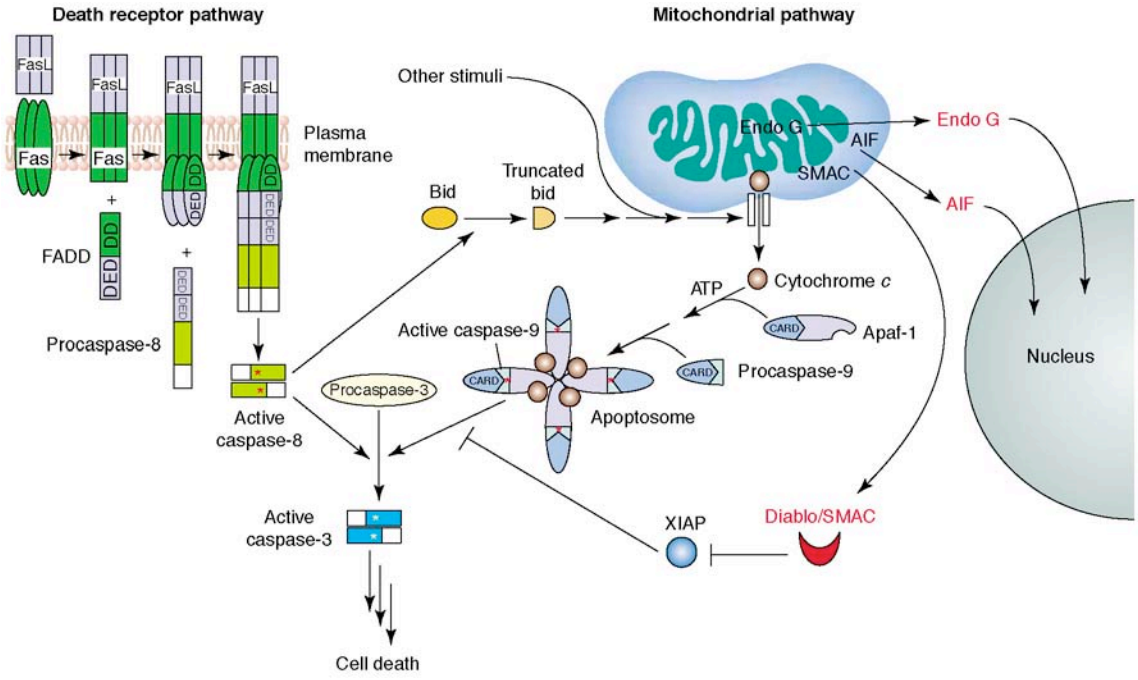


Figure 3 Multiple pathways leading to caspase activation.(see text for more details) Extracted from Kaufmann et al.[2]

Caspases are an important family of distinct cysteine proteases that share the ability to cleave substrates after aspartate residues (hence its name Cys-ASP-ases). They have a central role in the apoptotic process and are largely responsible for its signal transduction, amplification, and execution. In fact one could regard the entire network of protein-protein interactions controlling cellular suicide as a sophisticated trigger for a revolver, the bullet: caspases.

Typically, two canonical caspase activation pathways have been described (Figure 3). One begins with ligand binding to the outer cell-surface receptors known as death receptors; which include tumor necrosis factor-α and Fas/CD95. As an example, Fas Ligand (FasL) recognition initiates recruitment of FADD protein to the membrane, mediated by the interaction of its multiple death domains (DD) with Fas; formation of this complex known as DISC (death inducing signaling complex) is followed by recruitment of pro-caspase 8,

which becomes activated upon binding and autocatalytic pro-peptide cleavage. In certain cells, (Type I cells) caspase-8 activation is strong enough to be followed by procaspase-3 processing, activation, death signal amplification and execution. For the rest of the cells -type II- however, an alternative set of proteins modulates caspase-8 signal: this is known as the mitochondrial pathway. In the latter the cytoplasmatic protein Bid is processed by caspase-8 spawning a fragment that relocates in the mitochondrial outer membrane (MOM), there hetero-oligomerizes with Bax permeabilizing the membrane and inducing extrusion of various signals from the mitochondrion. Bax oligomerization at the mitochondrial membrane is another important cross point, and is controlled by various pro and anti-apoptotic factors; we will cover this event in more detail later in this chapter (pg. 64).

Among the polypeptides produced by mitochondria during apoptosis we find AIF (apoptosis inducing factor) and endonuclease G, both implicated in DNA fragmentation and degradation. But most importantly cytochrome C, whose release and cytoplasm accumulation readily follows MOM permeabilization.[3] Later cytochrome C is assembled in presence of ATP, pro-caspase-9 and Apaf-1 (apoptotic protease activator factor 1) scaffolding protein into the apoptosome, a large complex where caspase-9 becomes active and subsequently induces pro-caspase-3 through proteolysis.

The last fail-safe mechanism in the mitochondrial pathway are IAP (Inhibitor of Apoptosis Proteins), another important family of anti-apoptotic proteins.[4] Their mission consists in blocking caspase-8 and 9 initiated events; and for this some family members are able to inhibit proteolytic activity of various cysteine proteases, among others: caspases 3, 7 and 9.

XIAP (X-linked Inhibitor of apoptosis) is a remarkable paradigm of IAP's caspase inhibitory properties. It is composed of a series of three homologous BIR domains: in particular, the second BIR domain (BIR2) is believed to be largely responsible for caspase 3 and 7 inhibition. Whereas its inhibitory role towards caspase 9 has been characterized by Fesik and col.[5] and demonstrated to stem from BIR3 domain; which binds to the protease and hinders substrate access to the catalytic site. XIAP BIR-2 and BIR-3 domains are structurally very similar; they both contain three-stranded antiparallel $\beta-$sheets and three α-helices with similar orientations on either domain. Both are held together by a Zn atom, chelated to 3 histidines and a cysteine; the main difference between the two consists on the extra C-terminus helix present in BIR-3. But despite their similarities, BIR2 and BIR3 inhibition mechanisms towards Caspase-3 and 9 are slightly different. [6] BIR3 exerts its effect on caspase-9 by binding to its N-terminus AVPI consensus motif (Figure 4) through a remarkably localized set of interactions, occluding access to the catalytic cysteine. In fact, as we will see later on, so localized is this interaction that the sole presence of this N-terminus motif is able to disrupt Caspase9/BIR3 interaction. BIR2 on the other hand, operates somewhat similarly by modulating binding site accessibility, and although it shows certain affinity towards the former tetrapeptidic motif, its caspase-3 inhibitory activity involves a different binding surface.

In this "rope pulling" game there is still another player: Smac/DIABLO; yet another mitochondrial protein released during cellular suicide with pro-apoptotic activity. Its effects, however, are opposed to XIAP; in fact its job consists in capturing the former and liberating caspase-9 active sites.

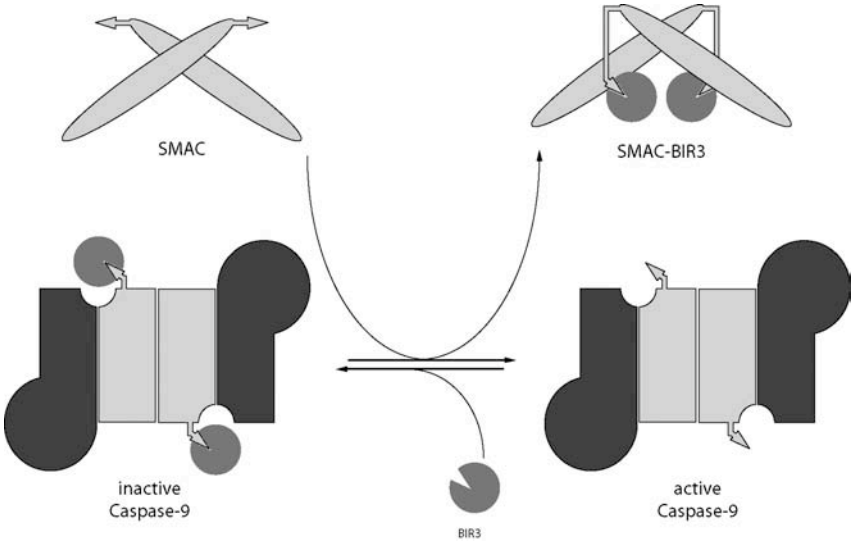


Figure 4 Caspase-9 inhibition by XIAP-BIR3 domain and subsequent release by SMAC/Diablo protein.

Structurally, Smac/DIABLO is a dimer of elongated helix bundles and its N-termini peptides share the same tetrapeptidic motif involved in caspase9–BIR3 interaction. As a result, they are able to supplant XIAP in its interaction with the cysteine protease (Figure 4) and the sole presence of Smac liberates caspase-9 to perform its grim duty. That is process procaspase-3, transmit and amplify the death signal. [7]

A recurrent observation in a variety of tumors is the presence of large amounts of naturally occurring anti-apoptotic factors; this is a common strategy used by malignant cells to silence suicide signals. In this sense XIAP has been found overproduced in different cancer phenotypes.[8] Naturally, the most obvious alternative to achieve sensitization to apoptosis for such cells is to increase the presence pro-apoptotic counterparts: i.e. Smac or Smac heptapeptidic surrogate. Indeed, several authors have demonstrated the apoptotic effects of such peptides and validated XIAP as an interesting therapeutic target,[9] and for this reason a number of academic and non-academic groups have attempted to develop Smac non-peptidic surrogates.

Given the biological background presented above, it is undeniable that XIAP's BIR3 domain is very attractive from a therapeutic perspective. From a methodological standpoint, this system is particularly enticing, since instead of classic enzyme target this is a protein-protein interaction; which so far has proven rather elusive to traditional drug discovery programs. Still, Smac/XIAP binding interface is remarkably small –a tetra peptide– and may be regarded nearly as a catalytic site. Predictably, it will be much more amenable to small MW inhibitors, and seems as an ideal transition between enzyme-type and protein-protein interaction targets.

In the following section, we will be using XIAP-BIR3 domain as a test-bench in the implementation of different selective labeling strategies with interesting features for eventual HTS (High Through-put Screening) programs.

1.1.2 XIAP-BIR3 PRODUCTION AND SYSTEM VALIDATION

The first milestone we have to accomplish if we are to perform any NMR study on XIAP-BIR3 domain is being able to produce and purify it in sufficient amounts. For this purpose several constructs were tested and finally $XIAP_{252-340}$ sequence plus a His-tag chosen for its yield and ease of purification.

Later, we attempted a standard ^{15}N uniform labeling: protein over expression was performed on BL21 host cells, followed by sonication, cell lysis and Ni affinity purification. NMR samples were obtained after protein concentration and dialysis into sample buffer.

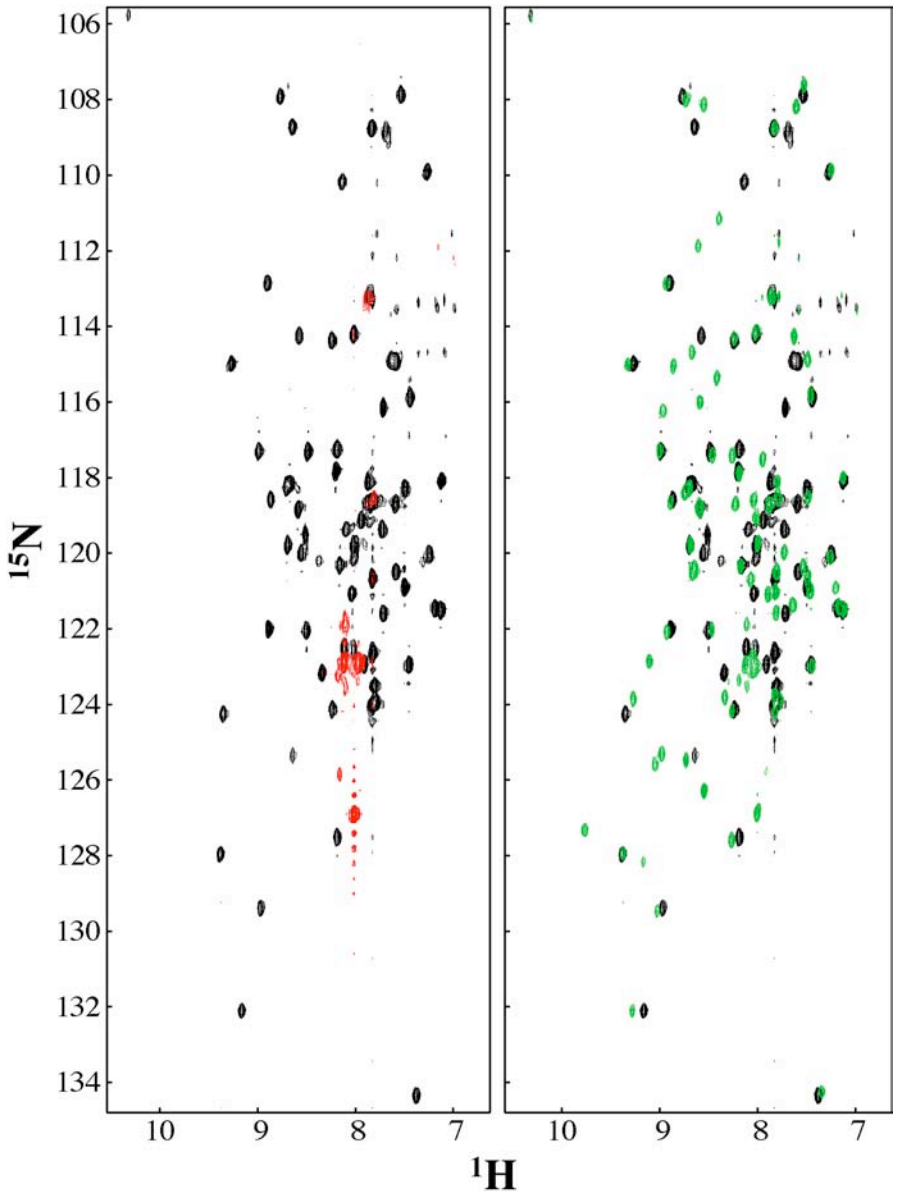


Figure 5 ^{15}N-^{1}H HSQC experiments for XIAP-BIR3 domain. Left (Black) spectrum corresponding to free protein, (red) spectrum after addition of 1:1 Smac full-length protein. On the right we can see (green) the effect of N-terminus Smac heptapeptide addition in excess to a sample containing full length SMAC/Diablo and BIR3 (1:1) compared to the free (black) spectrum.

Despite $XIAP_{252-348}$ construct's solubility and overall exceptional macroscopic behavior, which allowed us to reach up to mM concentrations, we decided to acquire heteronuclear 2D NMR experiments in order to assess fold integrity. As we can see in Figure 5 ^{15}N-^{1}H HSQC spectrum quality is excellent (black), and the observed signal dispersion seems to indicate that the domain is well folded; in fact, the recorded spectroscopic data closely resembles Fesik's published results.[10]

Next, we decided to examine the recognition properties described in the bibliography for this construct; and for this reason titrations were performed with two different versions of its natural ligand: Smac. Increasing amounts of full length Smac/Diablo protein were added to our ^{15}N uniformly labeled domain until a 1:1 ratio was reached (Figure 5, left-red); at this point, most protein signals broaden beyond detection. Given Smac's size and shape –it is a highly anisotropic 40 kD homodimer– together with XIAP-BIR3 molecular weight (13 kD), the observations could well be a consequence of an enhancement to BIR3's correlation time linked to the complexation event. Such apparent size increase has important relaxation implications that only allow the observation of loop and termini residues.

Interestingly, when Smac's N-terminus heptapeptide is titrated into the sample an increasing number of correlations reappear; also, when compared to the free protein spectrum the latter differs considerably (Figure 5, right-black) and nearly almost every correlation has shifted in the presence of heptapeptide. These results can be interpreted as Smac heptapeptide displacing full-length Smac protein to form a 14 kD complex, much more amenable to NMR observation and less prone to the previous relaxation issues. The extent of changes observed for the heptapeptide-complexed-BIR3 is on the other hand remarkable, and is certainly a consequence of a large rearrangement on the protein surface induced upon peptide binding.

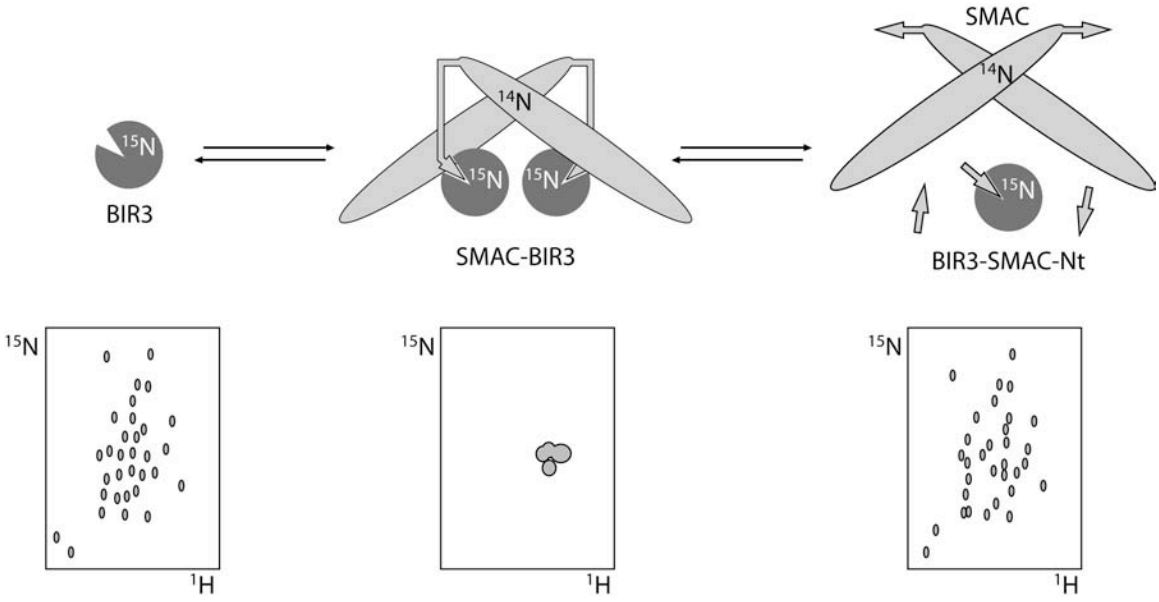


Figure 6 Schematic representations for an NMR assay to detect inhibitors of protein-protein interactions. Left: HSQC spectrum for small free protein (i.e. XIAP-BIR3). Center: protein-protein complex spectrum XIAP-BIR3/SMAC. Right: low MW inhibitor has disrupted Smac/XIAP-BIR3 interaction and spectrum corresponds to BIR3-Smac N-terminus surrogate complex.

All in all these data demonstrates that our construct $XIAP_{252-340}$ is able to recognize both full length and shortened Smac versions, and that as reported in the literature this interaction mainly involves a short Smac N-terminal section. In fact, the experiment depicted in Figure 5 could be envisaged as a general assay to screen for protein-protein interaction inhibitors. This takes advantage of a frequent nuisance in protein complex studies; usually complexes exceed the NMR accessible molecular weight and only individual complex members can be approached. In such assay only one part of the complex is labeled, and this remains unobservable in its complexed form as long as no inhibitor is added to the sample. But whenever a complex disruptor is introduced, the labeled fragment is released from the macromolecular complex and its HSQC spectrum recovers the appearance of a small protein. (Figure 6)

Lately D'Silva et al. [11] have used this very same idea to screen for inhibitors of MDMD2-p53 interaction and suggest this can also be extended to small protein-protein complexes –which are still observable by NMR- by artificially increasing the complex size with protein tags. Rizo and collaborators [12] have proposed a fairly similar approach to screen for protein-protein interactions; this elegant method uses non-labeled proteins and is able to differentiate between their free and complexed forms by relaying solely on ^{1}H spectra. There, the authors anticipate enhanced relaxation properties for protein complexes and are able to detect such behavior on the basis of simple spectra addition and subtraction.

1.1.3 W RELEVANCE AND HOT SPOT COMPOSITION

Although protein-protein binding surfaces seem at first sight somewhat unexceptional, being relatively large (some authors estimate around 600-1200 $Å^2$) and flat, the truth is that most display a large degree of surface and charge complementarity with their interacting partners. Moreover, contrary to what one could think, most often binding energy attained in the protein complex does not come from burial of large hydrophobic surfaces. Instead, as several alanine-scanning studies suggest, [13] a large fraction of this energy is contributed by a small number of residues or otherwise named "hot spots". These key amino acids tend to be situated at the center of binding interfaces, and are usually surrounded by a shell of residues engaged in relatively unimportant contacts. This shell of amino acids performs an important protective task in such flat surfaces: occluding water from the hot spot; which allows for efficient burial of key amino acids, avoids interaction with competing solvent, and improves contacts between complementary hot spots therefore reducing complex dissociation constant.

Certain amino acid residues, as shown in Table 1, are very common in hot spots, particularly tryptophan (21%), arginine (13%), and tyrosine (12%). On the contrary, residues such as leucine, methionine, serine, threonine, and valine, account for less than 3% of all hot-spot residues. Intriguingly, the nature of these favored amino acids is not strictly hydrophobic, and due to the combination of various chemical features they have the possibility to mediate a variety of interactions (hydrogen bonds, hydrophobic, π–π aromatic interactions) etc. and even more so within the lowered dielectric environment found in hot spots.

All in all, tryptophan is by far the amino acid with the highest fold increase in its hot spot occurrence, and this is even more interesting considering it is the least abundant amino acid in proteins. In the same sense, and despite its hydrophobic character, it is as often found in protein cores as it is on surfaces. The same happens for tyrosine, which as tryptophan, has a large hydrophobic surface and is able to contribute with aromatic π-interactions and hydrogen bonds. [13, 14] Certainly, hot spot amino acid frequency is not random and has been shaped throughout evolution to perform a precise task: mediating protein-protein

interaction; and if we are to target such interactions it seems only natural benefit from this privileged residues by using them to monitor binding events.

Residue	Total	(%)	Hot spots	(%)	Fold enrichment in hot spots
Arg	218	9.38	29	13.3	2.47
Asn	99	4.26	5	5.05	0.93
Asp	177	7.61	16	9.04	1.67
Cys	3	0.13	0	0	0
Gln	160	6.88	5	3.13	0.58
Glu	220	9.46	8	3.64	0.68
Gly	28	1.2	1	3.57	0.45
His	50	2.15	4	8	1.49
Ile	104	4.47	10	9.62	1.79
Leu	242	10.41	2	0.83	0.01
Lys	143	6.15	9	6.29	1.17
Met	69	2.97	2	2.9	0.54
Phe	166	7.14	5	3.01	0.56
Pro	89	3.83	6	6.74	1.25
Ser	178	7.66	2	1.12	0.21
Thr	131	5.63	2	1.53	0.28
Trp	19	0.82	4	21.05	3.91
Tyr	122	5.25	15	12.3	2.29
Val	107	4.6	0	0	0

Table 1 Amino acid preferences in protein hot spots according to Bogan et al. [13]. Hot spot amino acids are those contributing to Gibbs binding energy in 2 kcal/mol or more.

As already mentioned, NMR monitorization of protein-protein interactions faces all the difficulties associated with high molecular weight macromolecules, namely relaxation and spectrum complexity issues; some of these can be alleviated by selectively introducing stable isotopes and monitoring a few resonances, rather than hundreds of them. And what better choice do we have than to carefully place these probes on residues

found in up to 46% of the hot spots: Trp, Tyr, and Arg. In the following section we will explore selective introduction of NMR observable isotopes on tryptophan, the most relevant of the three.

XIAP/caspase-9 interaction is an excellent example of tryptophan's ubiquity mediating protein-protein interactions. In this particular case the interaction surface is limited to a relatively small surface, and most of the interaction energy befalls on AVPI tetrapeptidic motif –shared by caspase-9 and SMAC N-terminus- and on two tryptophan residues in XIAP-BIR3. Liu and collaborators [15] structurally dissected this complex and observed a set of crucial interactions: D314 forms a salt bridge with peptide alanine (A1) α-amino, while A1's methyl group establishes a tight hydrophobic interaction with W310 side chain. Additionally alanine's carbonyl forms a hydrogen bond with a second tryptophan, W323. The previous tryptophan in BIR3 is also involved in important hydrophobic contacts with the peptide proline, thanks to tetra peptide extended conformation. Tryptophans 310 and 323 are thus the foundation of BIR3 interaction with its biological partners: Caspase 9 and Smac. Consequently, introduction of probes at these positions would provide a simple and reliable method to monitor such an important pocket. For the same reason XIAP-BIR3 makes an excellent test bench in our attempt to devise a selective labeling strategy for tryptophan.

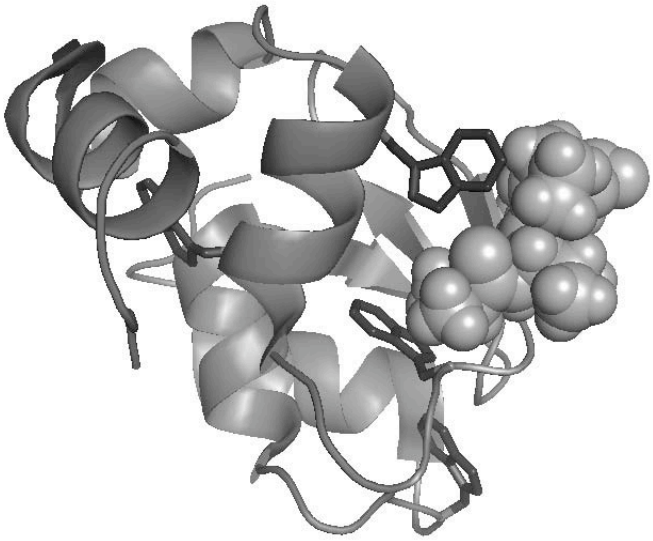


Figure 7 XIAP-BIR3 domain in complex with Smac N-terminus tetra peptide (AVPI). Tryptophan side chains are depicted in sticks.

1.1.4 TRYPTOPHAN SELECTIVE LABELING

At the time of our internship in Dr. Pellechia's Laboratory, the available methodologies to address specific tryptophan labeling were sparse, and to some extent they were rather general to almost every residue. Among these, probably the simplest strategy consists on supplying bacterial media with the labeled amino acid one wishes to introduce the isotope at. However, care must be taken to avoid this being processed by bacteria and scrambling the spin label to metabolically close residues. Isotope dilution can be minimized to some degree by supplying the rest of amino acids together with the labeled residue, since under these conditions the bacterial enzymes responsible for amino acid interconversion appear to be low or absent. Still, success is largely dependant on the particular amino acid and where the isotope is introduced; for

instance Asn, Asp, Gln or Glu, being biosynthetically central products, are typically used as precursors to other amino acids and thus prone to suffer from scrambling. While metabolically peripheral amino acids such as Trp or His can be labeled with higher isotope purity.[16]

Another general strategy takes advantage of amino acid specific auxotrophic bacterial strains. These strains have tailored defects on their metabolic machinery that hamper the fabrication of one or more amino acids, which have to be supplemented in the media for them to survive. There are currently available auxotrophic *E. coli* strains for every amino acid and although protein yields may suffer from the use of such strains, they can be useful when one wishes to introduce non-proteinogenic residues. Duewel and collaborators [17, 18] provide several examples for the latter; in their work they introduce fluorinated methionine analogs into phage lysozyme by using a methionine auxotrophic *E. coli* strain. Similarly, Eichler and collaborators have recently introduced fluorohistidine analogs into PapD chaperone. [19]

As we have seen above, amino acid production blockade can be introduced at a genetic level; however, one can achieve the same result by chemically inhibiting certain biosynthetic pathways. Kim and collaborators reported successful introduction of fluorinated phenylalanine analogs into *E. coli* ATPase through glyphosate-mediated inhibition of aromatic amino acids biosynthesis.[20] There, tryptophan and tyrosine were supplemented together with 4-fluorophenylalanine to obtain remarkable incorporations for the non-natural amino acid into ATPase.

So far, in all mentioned strategies one needs to supplement isotopically labeled amino acids to the media, which is often an expensive or synthetically difficult choice. A cheaper approach, especially suitable when trying to label several amino acid types, is the use of common biosynthetic precursors. In this case, rather than providing the whole amino acid, a precursor is supplemented; usually the end isotopic incorporation depends on the particular precursor selected and the reactions this undergoes within the host's metabolism. There are many successful examples of such strategies in the literature; particularly, Kay and collaborators have widely used them to incorporate stable isotopes on aliphatic residues, especially on methyl and methylene groups. Regarding labeling specificity, this can be controlled to some extent with the precursor choice; for instance ^{13}C pyruvate will introduce labels in Ala, Ile, Val, and Leu methyl carbons, while ^{13}C α-keto-isovalerate will primarily label Leu, Val residues, and ^{13}C ϵketo-butyrate will only affect isoleucines.[21-23]

Several authors have also used precursor approaches to introduce spin labels on aromatic side chains, these operate similar to methyl and methylene labeling schemes, except that shikimate or chorismate are used instead as common precursors to aromatic residues. In this context, Rajesh and collaborators have validated the former scheme by selectively incorporating protons into aromatic side chains in an otherwise perdeuterated protein background; in this way relevant aromatic nOe restraints were collected for yeast ubiquitin hydrolase 1. [24]

Despite the apparent variety of methods available to incorporate spin labels on proteins, thus far the above methods have something in common: they all require a host to produce the protein; and although this is a well established technology that provides sufficient yields, it often fails to unorthodox labeling schemes. In this respect, several recent breakthroughs promise to flatten the way to a wider universe in protein production. Cell free expression systems [25-27] have been proposed, and once some efficiency issues are worked out, they are expected to be much more versatile to the incorporation of non-natural amino acids

and less inclined to isotope dilution; since the supplemented non-natural amino acid and its potential toxic effects will not compromise the viability of the protein expression system. On the other hand, semi synthetic strategies like Expressed Protein Ligation (EPL) allow for virtually any chemical modification on a given protein section. As a result they have raised considerable interest to produce distinct isotope labeling schemes within segments of the same protein (segmental labeling). [28-30] But probably the "Holy Grail" in the field is currently being explored by Peter Schultz's group [31, 32]; they intend to expand the host cell genetic code in order to incorporate more amino acids than the 20 naturally available.

Keeping in mind our protein production expertise, it seemed wise to concentrate our efforts on selective labeling techniques involving *E. coli* strains as hosts. At the time, the most appealing options to achieve tryptophan selective labeling implicated either the use of biosynthetic precursors or some flavor of labeled amino acid in combination with Trp auxotrophic *E. coli* strain or chemical "knockdown". Whichever the choice, it is necessary to understand how tryptophan is produced within *E. coli*.

1.1.4.1 TRYPTOPHAN BIOSYNTHESIS

In bacteria, fungi and plants the branched pathway leading to tryptophan, tyrosine and phenylalanine [16] is fundamental; not only are these amino acids essential, but also for this is the main route to aromatic ring formation and thus to many secondary metabolites. This proceeds by cyclization of the aliphatic derivative (2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate) and subsequent steps introducing double bonds in a 6-member ring precursor (Figure 8). The first steps produce shikimate, which after three enzymatic reactions is transformed into chorismate, the last common precursor to aromatic amino acids. Next, for Phe and Tyr, biosynthesis proceeds through prephenate as intermediate; while for Trp, chorismate is transformed into anthranilate and subsequently condensed with 5-phospho-D-ribosyl-1-pyrophosphate (PRPP) to form N-(5-Phospho-D-ribosyl)anthranilate. Later, indole-3-glycerol-phosphate synthase will produce the indole ring from anthranilate's amino group and two carbons provided by PRPP. Finally, the very last two steps leading to tryptophan biosynthesis are catalyzed by tryptophan synthase. All in all, it appears that aromatic amino acids are peripheral metabolites, sitting considerably distant from trafficked crossroads and intermediates, (Figure 8) which is an important consideration if they are to be supplemented in the bacterial medium as labeling sources, for they will be less prone to scrambling.

In bacteria (i.e. *E. coli*), tryptophan synthase is a multi-enzymatic complex formed by two different polypeptidic chains α and β; two copies of each unit are assembled into a $(\alpha,\beta)_2$-hetero-tetramer, or a dimer of hetero-dimers. Each chain presents a distinct activity and whereas α unit catalyzes IGP (3-indole-D-glycerol 3'-phosphate) transformation into indole and G3P (D-glyceraldehyde-3'-phosphate), β subunit is responsible for the condensation between indole and L-serine using a pyridoxal-5'-phosphate (PLP) cofactor. (See Figure 9)

Tryptophan synthase's architecture is noteworthy: individual active sites for α and β subunits are connected through a 25Å tunnel; this is largely hydrophobic and its dimensions allow indole to transit from α to β subunit active sites. The rational behind this structure is efficiency; since a controlled indole transfer within the protein matrix avoids its escape to the cytoplasm and eventually through the bacterial cell wall due to its neutral charge and hydrophobic nature. Coupled to this transport, a fine network of allosteric mechanisms orchestrates a precise choreography where the activity of catalytic sites in α and β subunits occurs in phase. [33]

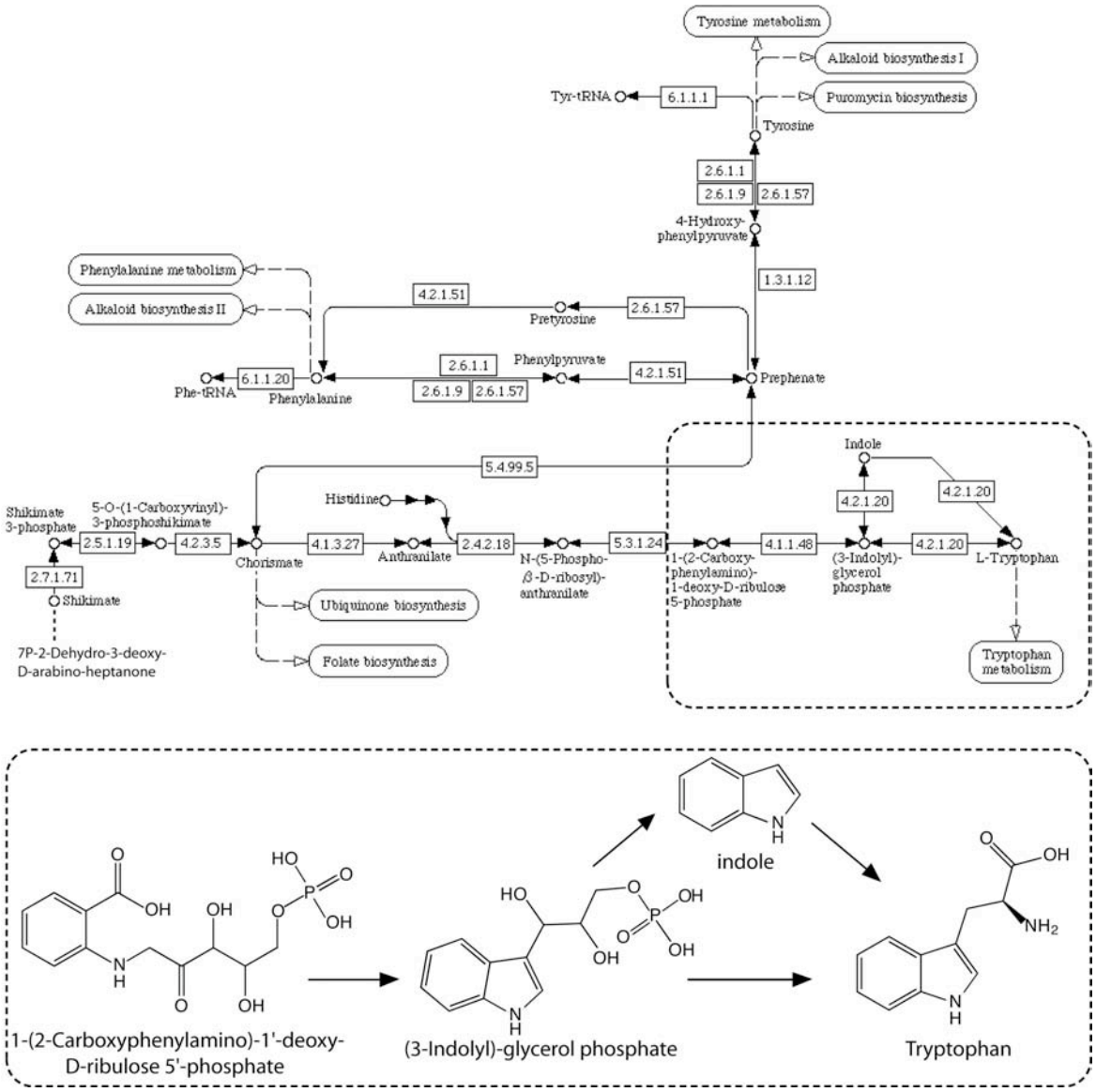


Figure 8 Aromatic amino acids biosynthesis in *E. coli*. Extracted from KEGG resources [16, 34] The last steps for tryptophan assembly, catalyzed by tryptophan synthase, are shown in more detail in the bottom chart.

The fact that indole is an intermediate for the last enzymatic steps towards tryptophan is very appealing one. This simple molecule is synthetically very accessible, in fact, it is commercially available in various labeled flavors; additionaly, indole is nearly as "metabolically isolated" as it is the complete amino acid. Both features are very attractive for labeling purposes as tryptophan side chain precursor. On the other hand, intracellular transport will not likely be an issue, since the very reasons that require the existence of a tunnel in tryptophan synthase will probably play on our side, favoring transit across cell wall.

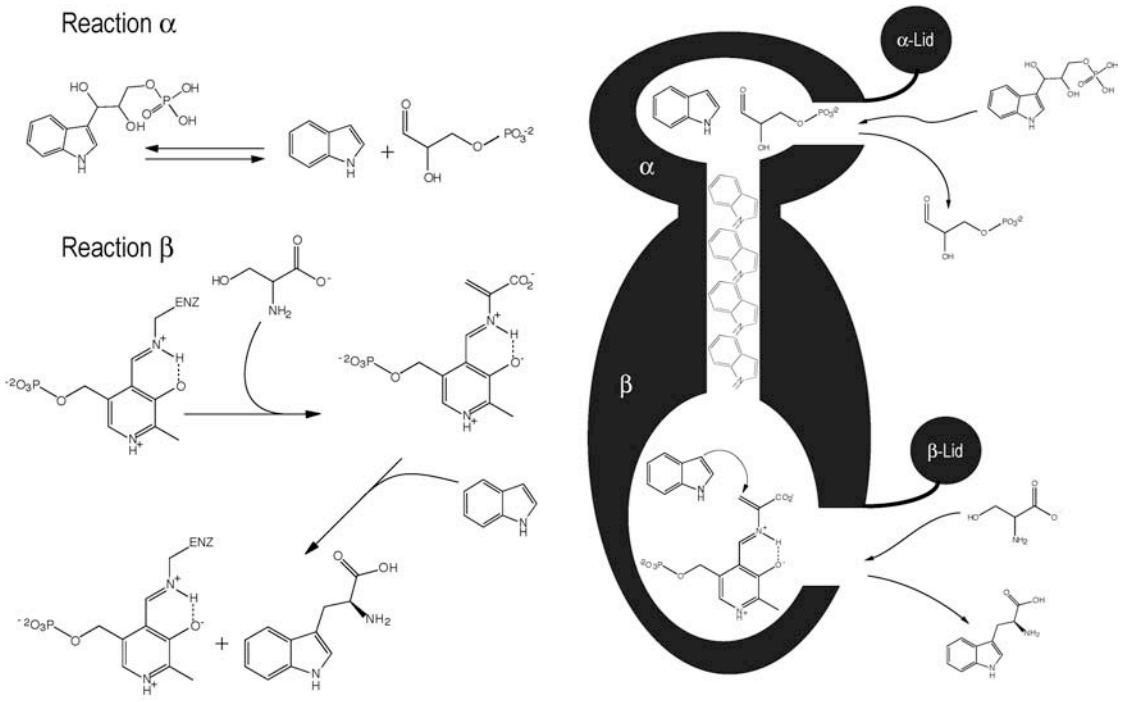


Figure 9 Schematic representation of Tryptophan synthase (α,β) heterodimer and reactions taking place within each subunit.

As already mentioned tryptophan synthase machinery undergoes a sophisticated allosteric ball to guide substrates towards its end products. During the various reaction intermediates α and β catalytic sites are not always accessible to the bulk solvent, nor activated; as a result there is the possibility that the only "reactive" indole in β-subunit active site is that which has transited through the 25 Å tunnel; a fact that would greatly reduce the isotope incorporation efficiency. So, exogenous indole may not be such a straightforward biosynthetic precursor, and therefore the hypothesis should first be verified.

1.1.4.2 PROOF OF PRINCIPLE: REVERSE LABELING STRATEGY

In order to test the validity of our hypothesis –usage of indole as an efficient tryptophan side chain precursor- two labeling schemes where proposed and carried out for XIAP-BIR3 domain.

First, the same ^{15}N uniformly labeled protein batch used for the previous experiments was used to acquire control heteronucler 2D experiments. The second proposed labeling scheme was devised to evaluate indole incorporation into tryptophan side-chain by means of a reverse labeling strategy. Protein production proceeded as in the ^{15}N uniformly labeled case, with the difference that prior to IPTG induction we incorporated ^{14}N indole (50 mg/L) into the media. In this way if *E. coli* assembles tryptophan from this precursor instead of building IGP from the start the isotope present in the aromatic side chain will be ^{14}N, thus unobservable to the NMR experiment.

Unfortunately observation of Trp-NHε spin systems for XIAP-BIR3 free domain is not straightforward and, as portrayed by Fesik and collaborators in their assignment, [10] this is only possible for one tryptophan: W317, the least exposed of the four, found in what could be considered the domain's hydrophobic core.

The reason behind the missing crosspeaks is most likely some sort of dynamic phenomenon within XIAP-BIR3 fold, leading to broad or undetectable correlations for indole NHε in the active site region. Supporting this evidence, Shin and collaborators have analyzed fold stability for several BIR3 truncated versions and determined that minor truncations ($XIAP_{257-347}$ or $XIAP_{253-356}$) show enhanced sensitivity to chaotropic reagents; which suggests that conformational exchange may not solely have functional implications, but could be an artifact arising from XIAP domain dissection.[35, 36] Whatever the reason, addition of Smac N-terminus peptide seems to freeze XIAP-BIR3 conformational sampling and beyond the numerous chemical shift perturbations induced (Figure 5, right) all Trp-NHε crosspeaks become observable. Consequently, if one wishes to analyze indole incorporation by means of ^{15}N-^{1}H HSQC and reverse labeling, this will have to be done in the presence of Smac surrogate peptide.

Under the previous conditions, comparison between ^{15}N-^{1}H HSQCs for control and reverse-labeled proteins (Figure 10) reveals four missing signals for the protein batch grown in ^{14}N indole enriched M9 media. Fesik's assigned W317-NHε cross peak [10] is clearly among the missing signals together with other three signals in the typical Trp-NHε chemical shift region. All in all, it seems clear that ^{14}N-indole has been introduced into XIAP-BIR3's four tryptophans. Interestingly, cross peak disappearance or intensity reduction is circumscribed to the four mentioned resonances and no other peaks seem to be affected, meaning that scrambling or spin dilution effects are probably unimportant.

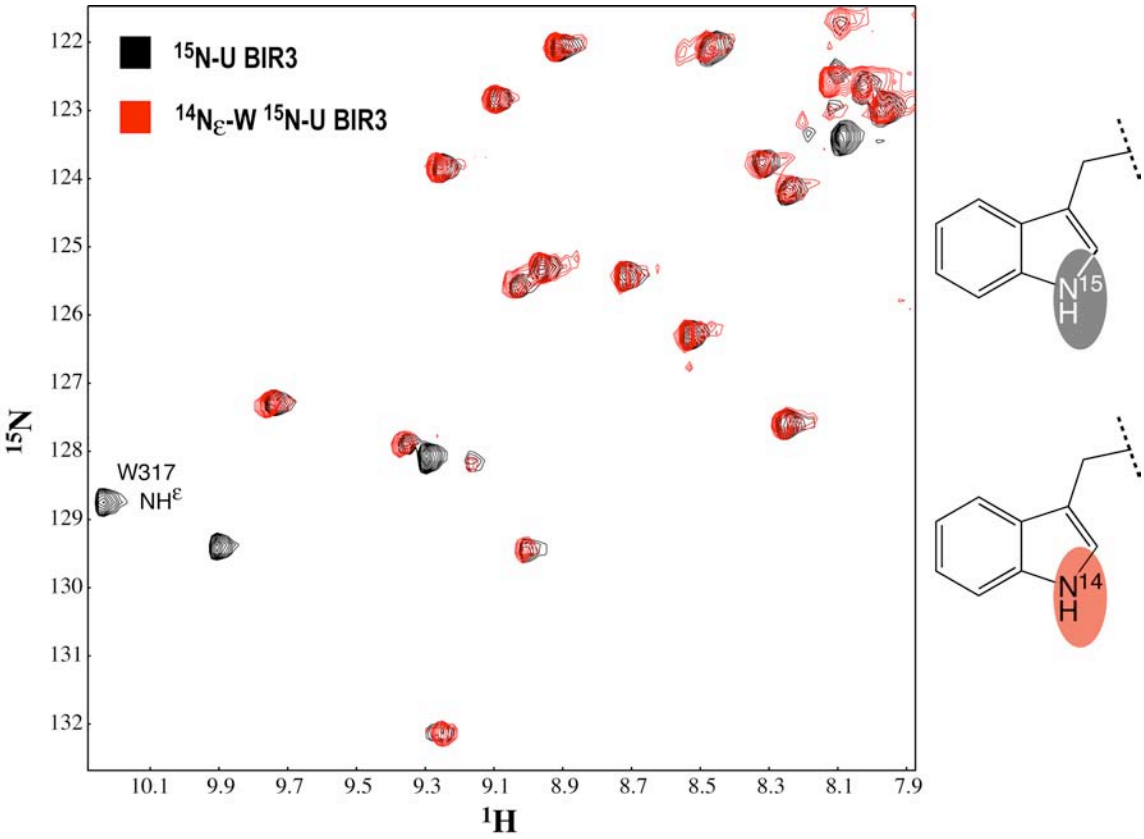


Figure 10 Reverse labeling experiment. ^{15}N-^{1}H HSQC section for ^{15}N-uniformly labeled XIAP-BIR3 domain (Black) and ^{14}Nε-Trp ^{15}N-U XIAP-BIR3 (Red). Both protein samples prepared at 700 μM in the presence of two equivalents of Smac N-terminus heptapeptide.

In conclusion, barely for the cost of ^{15}N uniformly labeled protein we have been able to confirm our hypothesis: indole is an efficient tryptophan side chain precursor and consequently a good means to selectively introduce stable isotopes within this relevant amino acid. We will now be able to attempt incorporation of much more expensive isotopically labeled versions of indole.

1.1.4.3 LABELED PROTEIN PRODUCTION

Once indole incorporation has been demonstrated feasible we decided to pursue incorporation of NMR observable spin probes into tryptophan side chain. For this purpose we explored labeled indole commercial sources, which at the time existed in three flavors: $^{15}N\varepsilon$, $^{13}C_4$ and $^{13}C_2$ indole. Due to the important exchange issues experienced by $^{15}NH\varepsilon$ (see previous sections) it seemed inadequate to try $^{15}N\varepsilon$ indole incorporation, instead we decided to tackle labeling schemes involving ^{13}C with the hope that these problems could be alleviated to some extent by the change in nucleus. Hence, two different protein batches were produced similarly as in the precious sections, but differing in the supplemented labeled indole. In both cases, precursor incorporation was assessed using ^{13}C-^{1}H HMQC NMR experiments.

Protein production proceeded smoothly and yields were not affected by addition of either indole to the media. Judging from heteronuclear ^{13}C-^{1}H experiments (Figure 11) both attempts were equally successful at incorporating indole into Trp side chains; there were however important differences between $^{13}C_2$ and $^{13}C_4$–Trp spectra.

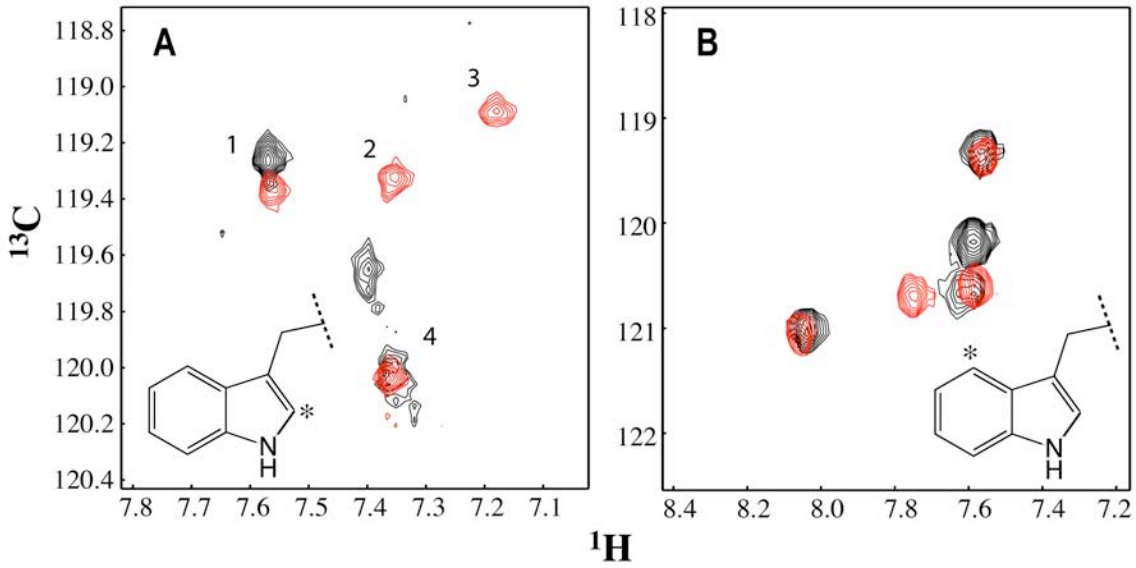


Figure 11 ^{13}C-^{1}H HMQC experiments for ^{13}C-Trp labeled XIAP-BIR3 domain; protein was produced with $^{13}C_2$ indole (A) or $^{13}C_4$ indole (B) as tryptophan biosynthetic precursor. Samples were prepared at 100μM protein concentration, and spectra recorded for the free protein (black), and after addition of 2 equivalents of Smac N-terminus heptamer (red).

$^{13}C_2$-Trp BIR3 presents three cross peaks in its free spectrum (Figure 11 A), and although this implies that slow dynamics is less of an issue to this ^{13}C-^{1}H spin system, it is still remarkable to see that two signals are rather broad and that the missing cross peak is only observable at 50˚C (see Figure 14). Smac N-terminus

peptide addition is again the antidote to conformational exchange and induces similar sharpening effects on cross peaks –as reported in the ^{15}N-^{1}H HSQC–, along with broad chemical shifts changes particularly for one of the spin systems.

On the other hand, $^{13}C_4$-Trp BIR3 HMQC spectrum clearly shows four sharp slightly more disperse cross peaks; clearly indicating that this labeling scheme is the least prone to suffer from exchange broadening compared to $^{15}N\varepsilon$-Trp or $^{13}C_2$-Trp BIR3. For this reason resonance sharpening upon peptide addition is hardly observed; nonetheless chemical shift changes do occur, and albeit not as dramatic as for $^{13}C_2$ they are mostly limited to one signal.

Altogether, indole has proven a valuable precursor for tryptophan side chain and has allowed us to produce XIAP-BIR3 domain with a variety of probes selectively placed on this important residue. Introduction of ^{13}C does seem to reduce the deleterious exchange effects observed on the indole ^{15}N and allows for an easier ligand-binding monitorization as demonstrated with Smac N-terminus heptapeptide peptide with either ^{13}C labeling. Yet there are several foreseeable drawbacks to these labeling schemes, mainly related with resonance assignment. These issues are intrinsic to labeling selectivity and hence to the lack of observable isotopes (i.e. backbone ^{15}N and ^{13}C) upon which traditional assignment protocols rely; nevertheless we can envisage several alternatives to bypass assignment limitations. Probably the most cumbersome assignment procedure involves tryptophan mutagenesis, similarly to what has been done for other selectively labeled residues; [18] other options include nOe detection with reported ligands –i.e. Smac N-terminus peptide- and subsequent use of published structural data to identify relevant tryptophans. The latter will not always be feasible, but one can also imagine a scheme combining NOESY type experiments with some sort of backbone uniform labeling; backbone resonances could be assigned through conventional methods and identification of backbone-to-side chain nOes could provide assignment information.

In the case of $^{13}C_2$ labeling, with the available data, it seems we can provide a tentative assignment (see Figure 11). Resonances 2 and 3 clearly belong to tryptophans 323 and 310 in BIR3 Smac binding site, as they are the most affected by Smac heptapeptide; thus, the remaining resonances correspond to residues 317 and 275. Assuming a similar exchange behavior for ^{15}N and ^{13}C nucleus it seems reasonable to associate the sharpest heterocorrelation in both experiments to the same residue side-chain: W317 (Figure 10-A, crosspeak 1 and Figure 10).

Once we have demonstrated that selective incorporation of NMR observable nuclei into Trp side chain is both feasible and desirable, during the following sections we will be discussing several potential applications, particularly for ^{13}C schemes.

1.1.5 SELECTIVE ^{13}C-TRP LABELING AND ITS USES

1.1.5.1 SCREENING METHODOLOGY

Our first heteronuclear 2D experiments have already shown that both ^{13}C-labeling schemes are quite powerful at detecting XIAP-BIR3 interactions with its reported ligand Smac N-terminal peptide. So in a way, both schemes retain relevant ligand-binding information while at the same time keeping very simple spectra, both in their absolute number of resonances and extent of ligand-induced changes. Thus classic Chemical Shift Perturbation (CSP) strategies can be successfully implemented with such labeling schemes to

determine binding constants and binding site identification; moreover the latter can in some instances be done without the slow dynamics interference often found in protein binding sites.

But beyond traditional CSP approaches, tryptophan's peculiar features allow us to move a step ahead in our quest towards spectrum simplification. Tryptophan is relatively sparse in protein sequences, hence signal overlap is not likely to occur; on the other hand its striking hot spot fold increase [13] certainly anticipates a relevant function whenever it is found on a protein surface patch. These two properties can be combined into an extremely powerful monodimensional Chemical Shift Perturbation assay.

Selectively introduced ^{13}C nuclei can be used to filter $^{13}C_2$-H or $^{13}C_4$-H tryptophan spin systems from the rest of the protein and background protons; that is to say a monodimensional version of ^{13}C-^{1}H HMQC or HSQC experiment will produce resonances only for those protons attached to a ^{13}C isotope. Such experiments will likely have enough resolution to distinguish between the few Trp present in the protein, while at the same being sensitive enough to monitor binding events. Indeed, such super-simplified experiments have been recorded for our protein, and as shown in Figure 12, 1D ^{13}C filtered ^{1}H spectra for the free and complexed ^{13}C-Trp XIAP-BIR3 are clearly distinguishable.

But, if monodimensional HMQC experiments provide similar information as their 2D counterparts, why is simplification such an important issue? "It's the economy, stupid".[37] Experiments presented on Figure 12 have been recorded for a 100 μM protein sample within roughly 3 minutes and yet provide nearly as much information as a 20 fold longer 2D experiment. Such performance improvement clearly exemplifies the throughput capabilities of 1D ^{13}C filtered ^{13}C decoupled ^{1}H experiments, or conversely, its ability to use far less protein with the same amount of spectrometer time. This should have a huge impact on per-compound assay cost, a crucial factor if one is to perform a massive screening.

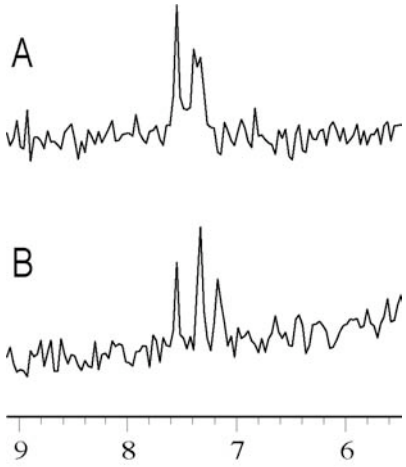


Figure 12 1D ^{13}C filtered ^{13}C decoupled ^{1}H NMR spectra with $^{13}C_2$-Trp labeled BIR3 in the absence (top trace) and presence (bottom trace) of Smac N-terminus 7mer.

1.1.5.2 Indole Proton Deuterium Exchange and Temperature Coefficients

Exchangeable protons, such as amides or Trp-NHε spin systems, undergo chemical exchange with protic solvents like water or ethanol. NMR provides several parameters to characterize rate and extent of this chemical reaction; which can be interpreted in terms of solvent exposure for this proton or even draw

structural conclusions, such as hydrogen bonding involvement or its relative position in the protein fold. Exchange rate is typically measured by following proton signal decay with time in the presence of deuterated protic solvent, and later these data adjusted to an exponential curve. In particular amide proton exchange, obtained from ^{15}N-^{1}H HSQC or ^{1}H NMR experiments, is a very rich source of information to study folding mechanisms, peptide conformational preferences (see chapter 3) and even protein structure.[38]

For XIAP-BIR3's ^{15}N-^{1}H HSQC direct exchange measurement was impaired to some extent, especially for tryptophan's NHε side chains, due to the aforementioned slow protein dynamics and would only be possible in the presence of Smac N-terminus heptapeptide. Interestingly, as we will see in the following paragraphs, $^{13}C_2$-Trp labeling provides various indirect means to measure this chemical exchange.

To $^{13}C_2$-Trp, deuterium exchange in the neighboring NHε is seen as a slight change to its carbon chemical shift and thus an extra cross peak appears for $^{13}C_2$-^{1}H; this isotopic effect provides a means to detect and quantify NHε and N^2Hε populations and its interconversion rate (see Figure 13).

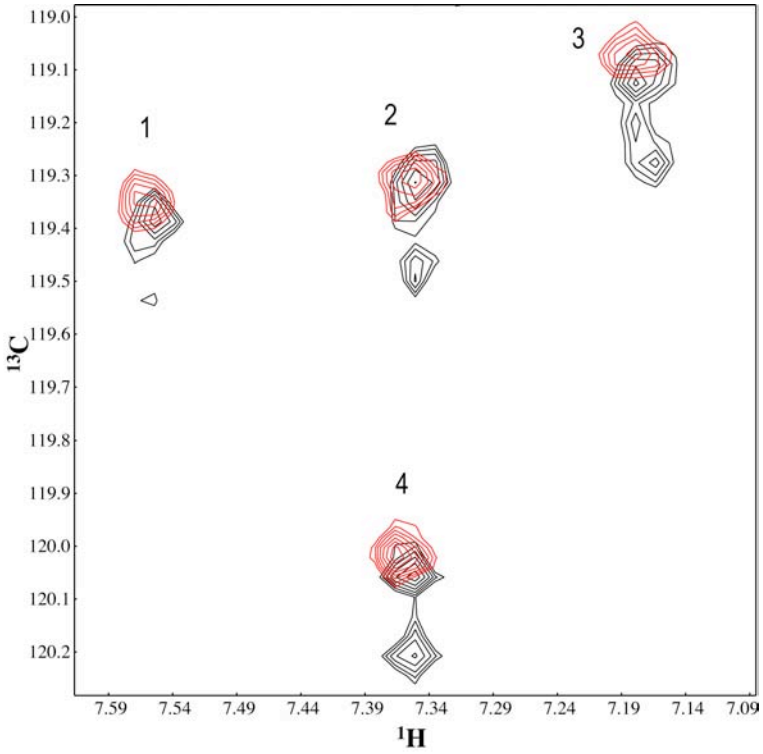


Figure 13 2D ^{1}H-^{13}C HMQC recorded for samples in 100% D_2O buffer (red) and 60% D_2O buffer (black); prepared with 0.2 mM $^{13}C_2$-Trp BIR3 in presence of 0.4 mM Smac N-terminus 7mer.

Addition of 40% H_2O to a $^{13}C_2$-Trp BIR3 / 99% D_2O sample followed by prompt HMQC acquisition results in Figure 13's black experiment. Interestingly, we observe various degrees of exchange for each correlation, meaning that at acquisition time some of the NHε had not reached exchange equilibrium. In fact, peak intensity allows us to establish a qualitative exchange rate order for BIR3 tryptophan side chain NHs: 1<2,3<4 (see Figure 13). According to our tentative assignment, the slowest exchanging NHε belongs to W317, the most buried of the four and as already mentioned in the domain's hydrophobic core;

unsurprisingly this is followed by residues 323 and 310 both experiencing some degree of burial due to interaction with Smac's N-terminus. Finally W275 has the most exposed side chain, a fact that correlates well with Fesik's XIAP-BIR3/Smac N-terminus heptapeptide complex structure. [15]

Another useful parameter related, in a way, to exchange rates is temperature coefficient. This has been widely used to obtain solvent protection information for conformationally flexible or unstructured peptides as an alternative to direct exchange rate measurement. For such systems (see chapter 3 for a broader study) exchange is often too fast to be measured; instead, temperature effects on exchangeable protons' chemical shift are recorded in strong hydrogen bonding solvents (i.e. water, DMSO or methanol). Hydrogen bonded protons will experience increased protection to temperature mediated δ changes; in particular, values below 4 ppb/K [39] (Δ chemical shift per Kelvin) correspond to amides considered protected and thus, likely intervening in an hydrogen bond.

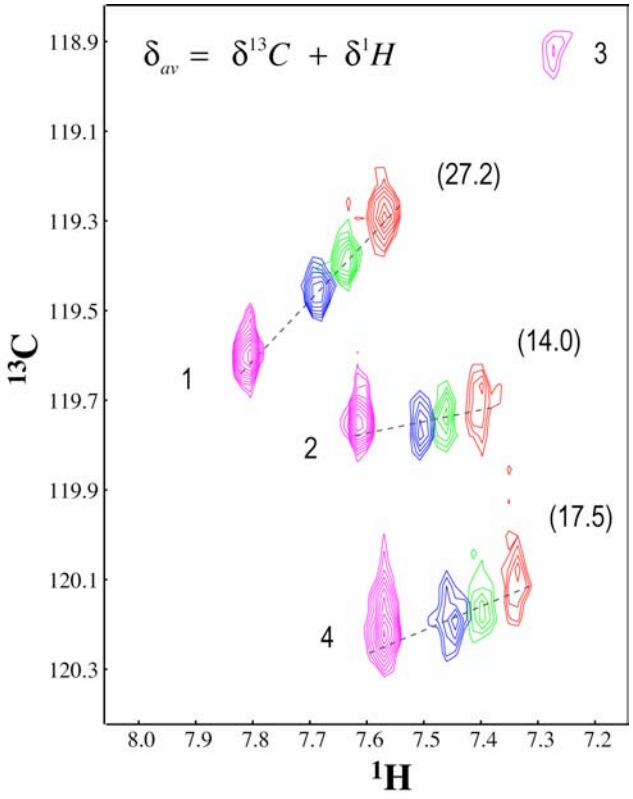


Figure 14 2D ^{1}H-^{13}C HMQC recorded for a 0.2 mM $^{13}C_2$-W BIR3 sample at different temperatures (red: 303 K, green: 308 K, blue: 313 K, purple: 323 K). Temperature coefficients (ppb/K) have been calculated using the average chemical shift value (δ_{av}); they are represented in brackets next to each correlation.

As in the exchange rate case, direct measurement of Trp-NHε temperature coefficient would not be very informative due to our inability to observe most of these NHs, particularly for BIR3 free form. Fortunately, temperature effects on NHε chemical shift are disseminated via through-bond effects to the vicinal $^{13}C_2$-^{1}H, thus a similar chemical shift versus temperature plot is obtained for the C-H system. This exercise has been

carried out for XIAP-BIR3 (Figure 14) and the averaged chemical shift slope, upon temperature increase, calculated for the available $^{13}C_2$-^{1}H cross peaks.

Temperature coefficient interpretation should in principle be analogous to that of amide protons; however, due to the lack of literature with respect to Trp $^{13}C_2$-^{1}H or even NHε coefficients, it is difficult to establish a reliable threshold below which a tryptophan side chain can be considered protected from solvent. Therefore, we are left with their relative solvent protection within XIAP-BIR3. In this context, W317 (Figure 14, crosspeak 1) would not be involved in any hydrogen bond, especially when compared with the other two observable signals (2, 4) -one belonging to W270 and another to a Trp in Smac's binding site-, which are far less affected by temperature. According to XIAP-BIR3 structure [10], only W323 and W270 would be able to hydrogen bond neighboring carbonyl functions, in agreement with our measured data; however, this must be taken with a pinch of salt since Fesik's NMR structure may not be completely reliable when it comes to side chain conformations.

1.1.5.3 OTHER POTENTIAL USES

So far, our proposed labeling schemes have been tested in their ability to identify and characterize small protein-ligand complexes; but eventually, the size of such systems could be expanded through a combination of selective labeling schemes and TROSY type experiments. TROSY was first introduced in a seminal work by Pervushin and collaborators [40]; there the authors achieved mutual compensation of CSA (chemical shift anisotropy) and dipole-dipole relaxation, yielding improved sensitivity and spectral resolution for ^{15}N-^{1}H spin systems. Lately, similar experiments have also been developed by Pervushin and collaborators to tackle aromatic ^{13}C-^{1}H systems.[41, 42] It is in this respect that we believe there could be interesting implementations for our Trp selective labeling schemes, especially $^{13}C_2$-Trp; which due to its relative isolation, could probably deal with very large macromolecules or complexes in combination with perdeuteration.

Dynamic data is yet another type of information one could obtain from tryptophan side chain labeling; in particular, measurement of various relaxation parameters could provide an interesting portrait of protein hot spots dynamics, [43] which is arising more and more interest in the understanding of protein function.

Another active field where our methodology could have an impact is protein folding. Tryptophan, when in protein cores, usually performs relevant duties during folding reactions or fold stabilization; for this reason several authors have used tryptophan probes to monitor such events. [44] Our labeling schemes could be used as an alternative to perform such studies.

1.1.6 SELECTIVE LABELING WITH ^{19}F-TRP

Throughout the previous sections, our understanding of tryptophan biosynthesis has allowed us to introduce various NMR observable isotopes at its very side chain. Yet isotopes have been limited to ^{13}C and ^{15}N, which is in essence an enrichment exercise for certain protein positions in order to enhance their sensitivity to NMR experiments. And, while this approach is very desirable in order to avoid disturbing protein fold and activity, it still relays on ^{15}N-^{1}H or ^{13}C-^{1}H detection that has to compete with an important signal background in the experiment, mainly arising from solvent protons and natural abundance ^{13}C.

Recently, ^{19}F utilization has attracted increasing attention and been proposed to deal with such issues particularly in several HTS NMR experiments. [45] This nucleus displays some especially compelling features: first its sensitivity, stemming from ^{19}F high natural abundance and its large gyromagnetic constant –roughly 80% of γ_H –; also, its wide ppm range and chemical shift sensitivity to subtle changes in the electronic environment have proved extremely useful. Finally, fluorine is not a common atom in macromolecules or even secondary metabolites and will very likely solve NMR background issues, thus reducing protein requirements and overall assay cost.

^{19}F has traditionally been regarded as decent ^{1}H isoster in medicinal chemistry due to its slightly larger Van der Waals radius; but despite of their size similarity, the two nuclei are rather different in their electronic properties. In addition, fluorine has demonstrated to enhance lipophilicity; which on the other hand may be interesting from a medicinal chemist perspective to modulate a compound's protein-binding properties.[46] For these reasons, care should be taken when introducing ^{19}F into proteins as fold distortions may occur; in fact, Duewel and collaborators have reported such phenomena when fluorinated methionine analogs are built into bacteriophage λ lysozyme (LaL).[17]

No matter how close we perceive ^{1}H and ^{19}F to be, the truth is that *E. coli* –or any of the frequently used hosts– does distinguish between natural amino acid and a fluorinated analogs; as a consequence, protein biosynthesis yields are greatly diminished with such amino acids either because cellular machinery recognizes them as exogenous or due to some intrinsic toxicity. Several alternatives to increase incorporation yields have been already mentioned, amongst the most straightforward we find: chemical inhibition of amino acid biosynthesis and the use of auxotrophic hosts.

Kim and collaborators successfully demonstrated the former approach for aromatic amino acids [20] by chemically blocking shikimate pathway via 5-enolpyruvylshikimic acid-3-phosphate synthase inhibition. In the presence of glyphosate (1 g/L), incorporation of fluorinated aromatic residues into proteins proceeded much more efficiently, though not completely, than by simply supplementing the exogenous amino acid to the bacterial media. In the following section we will be exploring a similar approach, only slightly more specific, to introduce ^{19}F-Trp analogs using our model protein XIAP-BIR3.

1.1.6.1 TRYPTOPHAN BIOSYNTHESIS INHIBITION: PROOF OF PRINCIPLE

Indole-3-acrylic acid (IAA) (Figure 16) has long been known to lead to tryptophan starvation in bacteria hampering normal growth rates; and according to several authors, its mode of action combines several simultaneous and opposing targets. Even though it has the ability to prevent Trp repressor to work, and thus enhances expression of enzymes in tryptophan biosynthesis, it also inhibits tryptophanyl-tRNA synthase preventing Trp incorporation in polypeptide chains [47-50]. IAA has also been proposed as an inhibitor of Tryptophan synthase [51] and is believed to interfere with the formation of the α-aminoacrylate intermediate; necessary to initiate the second reaction stage in Tryptophan synthase β-subunit and therefore inducing indole accumulation in bacteria (Figure 9).

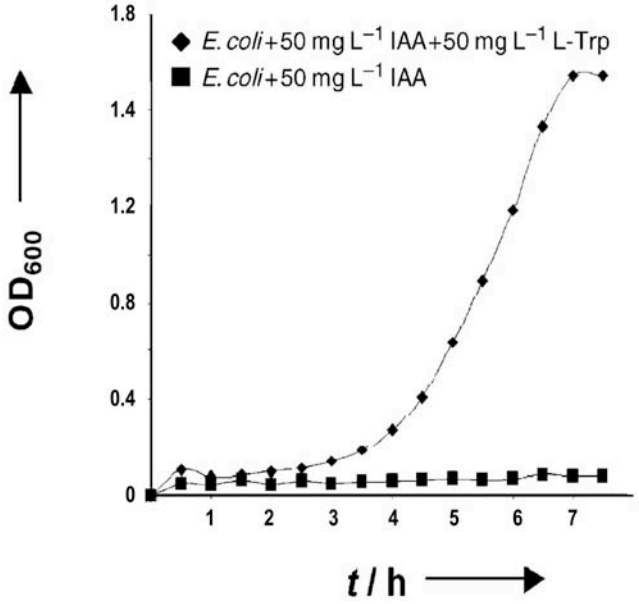


Figure 15 *E. coli* proliferation experiment in the presence of Indole-Acrylic Acid. (■) Culture media supplemented with IAA. (◆) Culture media supplemented with IAA and L-tryptophan.

From the above literature, it seems that whatever the mechanism behind tryptophan starvation, this is limited to this residue and that production of any other amino acid within bacteria will not be affected. However, and given the apparently contradictory mechanisms by which indole-acrylic acid exerts its effect, we decided to perform a proliferation experiment on *E. coli* under IAA effects. Two *E. coli* cultures were grown independently in M9 minimal media; the first culture was solely supplemented with IAA, while in the second tryptophan was also included. Bacterial growth was monitored with time using standard optical density measurement (at 600 nm wavelength) and the result plotted as in Figure 15.

IAA effects on *E. coli* growth are clear and agree with the previous literature: at a concentration of 50 mg/ml, indole acrylic acid seems to induce tryptophan starvation on bacteria and as a result proliferation arrest; cell growth is on the other hand restored by supplying the amino acid to the media, clearly indicating an inhibitory effect on the production of this amino acid.

1.1.6.2 $^{19}F_5$-TRP XIAP-BIR3 PRODUCTION

Once we demonstrated that IAA induced tryptophan biosynthesis inhibition, and conditions for such inhibition were optimized, we decided to pursue protein production where L-tryptophan was supplanted by a fluorinated analog: 5-fluorotrypotphan. (Figure 16)

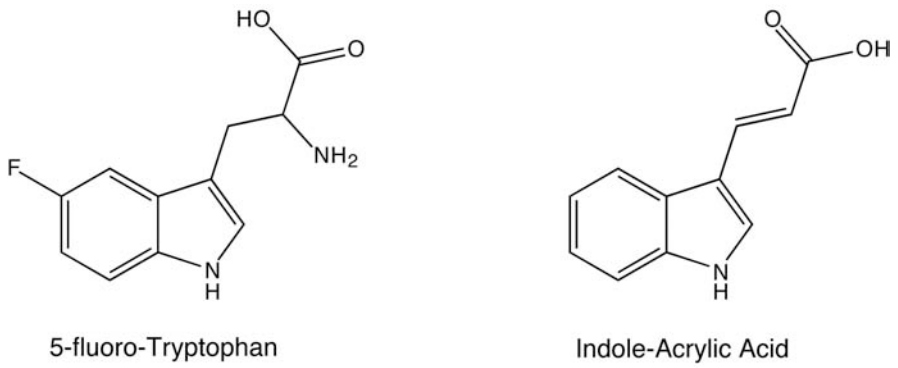


Figure 16 Chemical structures. (Left) 5-fluorotryptophan and (Right) indoleacrylic acid.

BL21 *E. coli* cells, containing XIAP-BIR3 plasmid, where grown in M9 minimal and prior to induction both IAA and fluorinated tryptophan analog were added to the culture. Induction, cell lysis and purification proceeded as described for previous protein batches. Although protein yield was reduced to 50% -compared to unlabeled protein expression-, the sole production of protein was in itself encouraging given the presence of IAA, and is certainly an evidence for either 5-fluoroTrp reversing IAA inhibitory effect or the former functioning as a tryptophan substitute. On the other hand, a protein yield drop was not completely unexpected considering the reported 5-fluoroTrp toxic effects on bacteria [52].

In order to evaluate the extent of fluorotryptophan incorporation, and prior to ^{19}F NMR experiments, we decided to perform protein MALDI-TOF MS analysis. Both for XIAP-BIR3 produced in the presence of 5-fluoroTrp and IAA inhibitor, and for an unlabelled control protein batch.

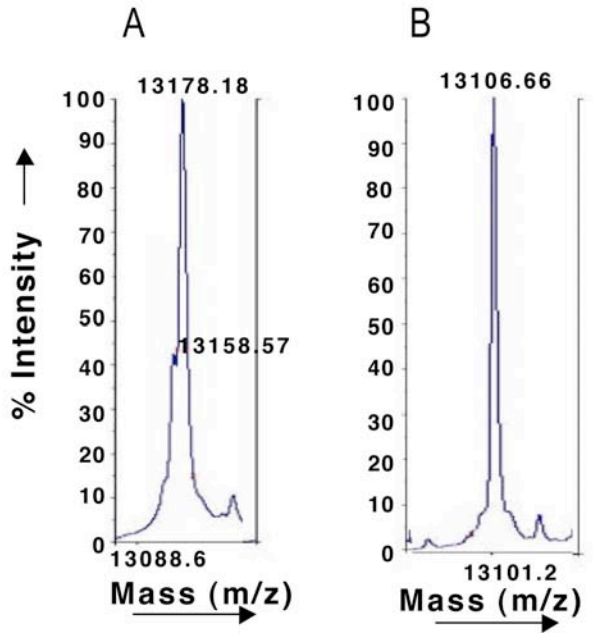


Figure 17 MALDI TOF mass spectra of 5-fluoroTrp BIR3 (a) and of unlabeled BIR3 (b)

MALDI TOF spectrum for the non-labeled XIAP-BIR3 domain displays a major peak at 13106.6 Da. While for 5-fluoroTrp supplemented sample two peaks are obtained, the most intense of them at 13178.1 m/z and the other at 13158.5 m/z with roughly a 40% intensity. Fluorine substitution of all H_5-Trp would translate into an 18 units difference with respect to wild-type protein; interestingly MW difference between the two spectra major peaks is 72 a.m.u, the exact figure one would expect for a 4x^{19}F-Trp substitution. Incomplete ^{19}F-Trp incorporation could explain the lighter peak at 52 a.m.u, with nearly the expected mass of a three fold ^{19}F-Trp substitutions; the error may be attributed to inherent resolution of linear TOF acquisition.

From the data above, it would seem that fluorinated tryptophan incorporation has been quite successful and at least from a qualitative perspective this has occurred for 90% of Trp residues ((0.6*1+0.4*0.75)*100). This incorporation yield is far superior to the strategy where solely fluorinated tryptophan is supplemented [53], but remarkably our approach is also superior to the use of glyphosate as inhibitor with a reported 80% incorporation. [20]

The ultimate ^{19}F-tryptophan incorporation tests were a series of ^{19}F NMR experiments. (Figure 18) As in the MS assays fluorine presence was confirmed; in particular for the free BIR3 domain three signals are clearly observable, two of which being rather broad. The situation is analogous to that of $^{13}C_2$-Trp selective labeling and is possibly due to the same protein dynamics phenomenon; again, addition of Smac N-terminus peptide freezes this conformational rearrangement and induces considerable ^{19}F resonance sharpening, along with important δ perturbations. Two out of the three original signals experience important shifts (0.8 and 0.5 ppm respectively) and a fourth signal appears upon peptide addition.

The observed spectrum changes for 5-fluoroTrp BIR3 are consistent with those observed throughout this chapter and could hence be interpreted similarly. (Figure 18) Analogously, a tentative assignment could be attempted; but in order to obtain a reliable identification one should perform a thorough site-directed mutagenesis study.

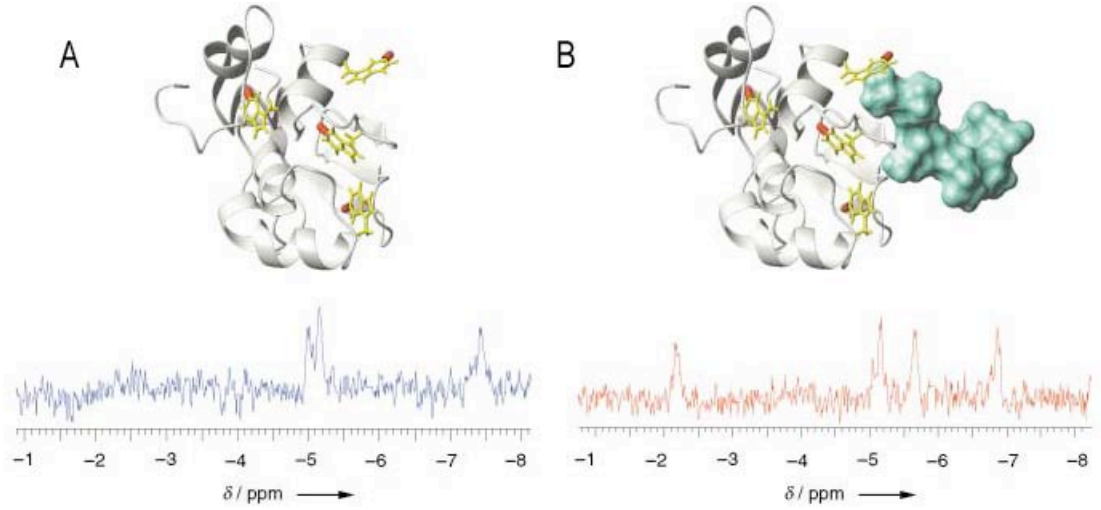


Figure 18 1D ^{19}F NMR spectrum of 5-fluoroTrp XIAP-BIR3 without (A) and with two Smac N-terminus peptide equivalents (B). Spectra were collected at 25 ºC with 0.1 mM protein sample of ^{19}F-Trp in H_2O/D_2O (90:10) phosphate buffer (40 mM, pH 7.5). Ribbon structure of the free and complexed XIAP-BIR3 is depicted on top of each ^{19}F spectrum.

As expected, fluorine is an extraordinarily sensitive probe and in our case is able to monitor Smac/BIR3 interaction with a performance comparable to that of previous ^{13}C-Trp schemes; and although it does not carry as much structural information as $^{13}C_2$-Trp its is superior in terms of signal dispersion and background. For this reasons ^{19}F-Trp labeling should have an important impact on NMR screening methodologies, reporting most of the advantages related to selective isotope incorporation on tryptophan without any of the deleterious background effects. Solvent or other background signals often determine the lower sensitivity thresholds for many such experiments; thus, ^{19}F acquisition will most likely reduce protein requirements and overall assay costs.

Beyond HTS applications, ^{19}F labeling could also contribute to folding and protein dynamics studies; although the range of macromolecules amenable to such studies will probably be somewhat limited due to fluorine intrinsic relaxation properties. Fluorine nuclei in proteins relax predominantly by means of dipole-dipole interactions with surrounding protons and by the chemical shift anisotropy mechanism; both mechanisms are particularly enhanced with correlation time and magnetic field hampering the study of large macromolecules or complexes. [53]

In summary, we have devised a novel approach to selectively incorporate ^{19}F in protein tryptophan side chains, this has demonstrated easier and more efficient than previously reported methods. Our methodology is cheaper than the use of glyphosate (250 Euros/5 g) and requires minor amounts of inexpensive IAA reporting good protein yields. On the other hand, this scheme is able to successfully monitor complexation of our protein with a short peptide and is expected to have a wide range of applications in HTS, protein folding and dynamics.

1.2 NMR SCREENING STRATEGIES APPLIED TO BCL-XL

1.2.1 BCL-XL BIOLOGICAL ROLE AND INHIBITION RELEVANCE

From the previous overview on apoptotic signaling pathway, it is clear that a crucial checkpoint to unleash caspase activity is established on mitochondria.[2, 3] In particular, release of proapoptotic proteins such as cytochrome c is controlled through a sophisticated mechanism by which certain Bcl2 protein family members heterodimerize and translocate to the mitochondria outer membrane (MOM). MOM becomes permeable to the release of cytochrome c and the death signal is transmitted downstream via assemblage of the apoptosome. (Figure 19)

Bax and Bak are the most relevant proapoptotic members in Bcl2 family, those that ultimately dimerize on mitochondria membrane triggering cytochrome c release. However, prior to heterodimerization, a conformational change must occur on these proteins for them to expose important hydrophobic helical domain (BH3 helix). This conformational change is induced by a variety of signaling proteins that also have this BH3 helix in common ('BH3-only' molecules) and are generated upon death stimuli. For instance, Bid truncated peptide is generated upon caspase-8 activation, while other "Bid-like" molecules (NOXA, PUMA) are produced as a response to DNA damage or cytoskeleton disruption through transcriptional control.

On the other hand, several members of the Bcl-2 family assume the opposite role and perform an antiapoptotic function by desensitizing cells to death stimuli. Bcl-Xl and Bcl-2 belong to this group and both display homology to the four BH domains in Bak, interestingly they exhibit affinity towards BH3 domains.

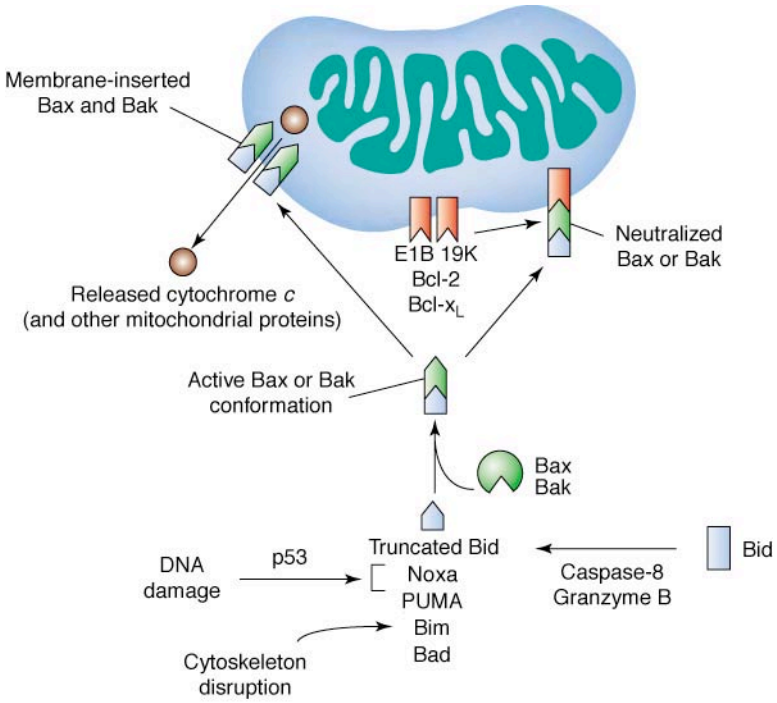


Figure 19 Model for the role of Bcl-2 family in apoptosis. (See text for more details) Extracted from Brenner and col. [2]

This exquisite balance between pro and anti-apoptotic Bcl-2 family members determines mitochondrial pro-apoptotic proteins liberation and ultimately the cellular fate; not surprisingly miss-regulation of such equilibrium has been identified as an important contributor to tumorigenesi and tumor resistance. In fact, a number of the mentioned desensitizing (Bcl-XL, Bcl-2) proteins are found over-expressed in many cancers and operate by capturing pro-apoptotic members of the family via BH3 domains to silence their signals.[9]

The simplest way to neutralize this decreased sensitivity to pro-apoptotic signals consists in increasing the amount of stimuli; and indeed, presence of Bak BH3 peptide has been reported to induce on its own caspase mediated apoptosis on HeLa cells; thanks to its ability to occupy Bcl-2 and Bcl-XL hydrophobic cleft liberating Bak and Bax. Such evidences validate both BCL2 and BCL-XL as cancer therapeutic targets, even more to the light of BCL-XL/BIR3 complex structure solved by Fesik and collaborators; which has fostered various screening and rational design efforts to build non-peptidic BH3 analogs capable of preventing BCL-XL's anti-apoptotic role.

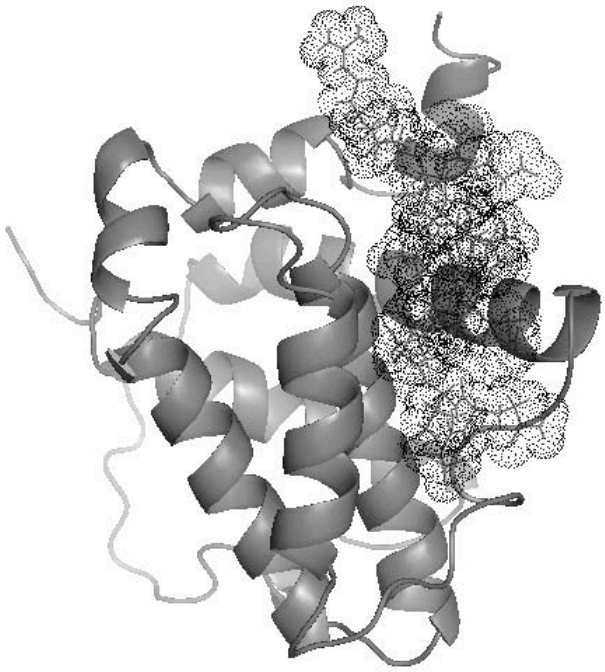


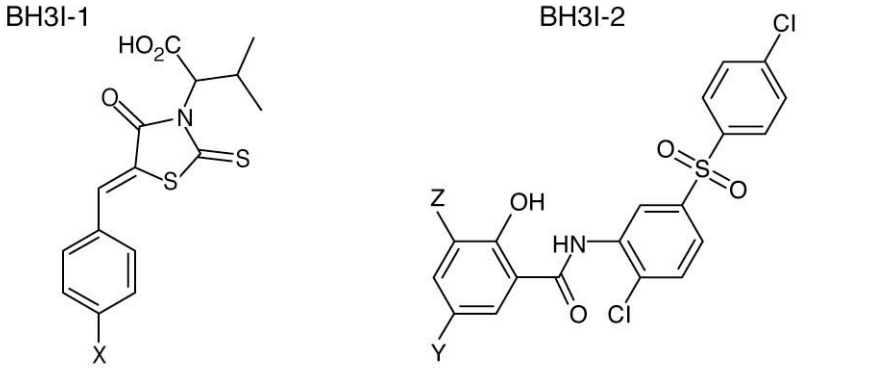
Figure 20 Bcl-XL (ribbon) in complex with BH3 Bak 16mer peptide (sticks and dotted surface).[54]

BH3 peptide binds to Bcl-XL through a cleft situated between BH1, BH2 and BH3 Bcl-XL helix domains, this interaction is mainly dominated by hydrophobic contacts although alanine scanning [54] has also revealed several important electrostatic and hydrogen bonding interactions. (Figure 20) Comparison between free and BH3 complexed structures highlights significant conformational changes that occur upon peptide binding, especially involving exposure of the large binding hydrophobic patch.

Degterev and collaborators [55, 56] have also reported similar rearrangements for Bcl-XL in the presence of several small molecular weight ligands; they would apparently also be capable of stabilizing the former "open cleft" conformation. The previous small molecules were identified in a fluorescence polarization assay for their ability to compete with BH3 peptide binding to Bcl-XL; later they demonstrated to induce in vivo apoptosis on a tumor cell panel through a mechanism consistent with Bcl-XL heterodimerization inhibition.

[56] Initial attempts to characterize by NMR the complex between Bcl-XL and BH3I, one of the molecules reported by Degterev, were unsuccessful due to its affinity regime –fast or intermediate exchange- that often translates in a limited number of experimental restraints. Another difficulty was related to the mentioned Bcl-XL plasticity; large conformational changes were observed upon BH3I-1 and BH3I-2 binding, thus traditional Chemical Shift Perturbation (CSP) analyses would not provide a reliable picture of their binding sites. Instead, a differential CSP [55, 57] strategy was applied, and a battery of ^{15}N-^{1}H HSQC titrations performed for various BH3I-1 and BH3I-2 close analogs (see Figure 21). The differences in their protein perturbation profiles were introduced as restraints into an *in silico* docking protocol to obtain complex structures. Both types of inhibitors seem to be placed in the same BH3 peptide groove and although they overlap to some extent, their binding mode is slightly different. According to the authors, BH3I-2 type inhibitors would cover a longer stretch of the cleft establishing energetically relevant interactions with various aromatic and aliphatic residues.

Nonetheless, this has not been the only successful story in trying to obtain BH3 peptide surrogates.[58, 59] Huang et al. reported the first small MW Bcl-XL inhibitor as a result of a computational screening and similar strategy proved equally fruitful to Wang et al.[60, 61] Other authors have used so-called “rational design” and obtained helix mimics to perform Bak-BH3 peptide job. In this sense, Hamilton and collaborators [62-64] have successfully employed oligo-amide foldamers and other scaffolds to spatially locate chemical groups as found in the inhibitor peptide, later they have tested their ability to disrupt Bcl-Xl- Bak interaction both in cellular and biophysical assays. Similarly, Schepartz and collaborators used the “protein graphting” approach with the same end; there, instead of foldamers a restricted peptidic scaffold was chemically decorated to mimic BH3 peptide.[65]



	K_i by FP	K_i by NMR		K_i by FP
BH3I-1: X=Br;	2.4±0.2 μM	7.8±0.9 μM	**BH3I-2**: Y=Cl; Z=Br;	4.1±0.4 μM
BH3I-1': X=Cl;	3.6±0.3 μM	8.4±0.5 μM	**BH3I-2'**: Y=Cl; Z=I;	3.3±0.3 μM
BH3I-1'': X=H;	12.5±0.9 μM	13.0±0.7 μM	**BH3I-2''**: Y=I; Z=I;	6.4±1.2 μM
BH3I-1''': X=N(CH$_3$)$_2$;	n.d.	15.6±0.7 μM		

Figure 21 Chemical structures for BH3I-1 and BH3I-2 inhibitors [56] and analogs used for Differential CSP strategy. Measured affinities for each analog are also presented.

All in all, Bcl-XL BH3 binding surface is not the typical flat protein-protein interface; instead, it is a flexible and quite promiscuous hydrophobic groove that has proved targetable to organic molecules, peptides and mini-proteins. The previous relatively successful stories are encouraging and the nature of protein surface, together with the ligands identified so far, suggest that Bcl-XL may not only be an interesting biological target but one suitable for fragment based screening. In fact, a recent work by Hajduk and collaborators mention Bcl-Xl as a particularly druggable system to the light of its high success rate in several screening efforts. [66]

1.2.2 BCL-XL

^{15}N uniformly labeled Bcl-XL was produced as described by Fesik et al. [67] and supplied by courtesy of Reed's laboratory at the Burnham Institute. The construct was not full length Bcl-XL, since the putative transmembrane domain (residues 45 through 84,) was removed to improve solubility and overall behavior. Nonetheless, and according to the published data, the polypeptidic chain retains its anti-apoptotic properties in vivo and is able to recognize BH3-Bak peptide. The latter was demonstrated by means of NMR ^{15}N-^{1}H HSQC experiments; these were recorded on an appropriately labeled Bcl-XL sample with and without Bak-BH3 peptide.

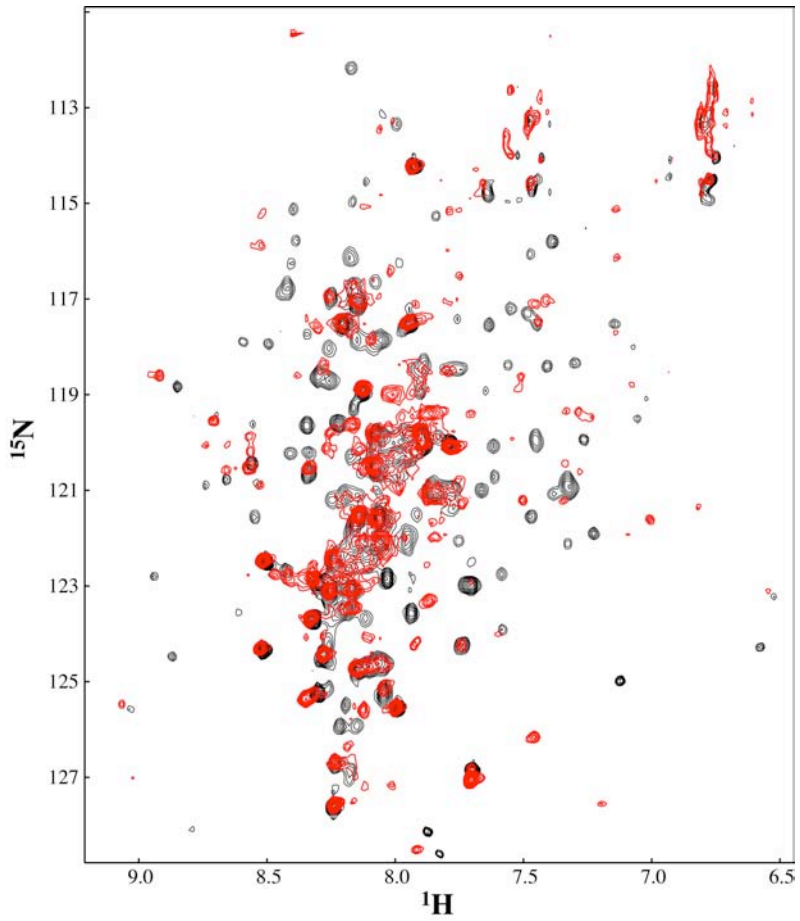


Figure 22 ^{15}N-^{1}H HSQC for ^{15}N uniformly labeled Bcl-XL. Black spectrum was recorded for a 300 μM protein sample; red spectrum was acquired after addition of one equivalent of Bak-BH3.

Heterocorrelation spectrum for the free protein confirms the construct's large helical content, as expected from the published spectroscopic data. Later, we observed that Bak-BH3 peptide addition has a widespread effect on the protein spectrum inducing numerous shifts, which are also consistent with its published affinity and binding mode. (Figure 22)

In conclusion, the protein target has been successfully validated both in terms of fold and recognition properties; unfortunately due to important overlapping issues –understandably, given the helical protein nature- assignment of ligand induced perturbations cannot be carried out thoroughly using published protein assignments. This will be an important limitation during the upcoming sections.

1.2.3 FRAGMENT LEAD DISCOVERY AND LIBRARY COMPOSITION

As already mentioned, NMR experiments can be wisely incorporated into fragment lead discovery approaches taking full advantage of their wide range of detected affinity ranges in ligands, throughput properties and structural information; but more importantly, dodging the particular pitfalls of each experiment. [68, 69]

This idea can be implemented in various ways and include other techniques than NMR, however during the internship in Dr Pellecchia's laboratory we applied a strategy that relied solely on NMR. Experiments were organized as follows: in a first stage the throughput capabilities of cross-saturation experiments such as STD or WaterLOGSY were used to screen the bulk of a fragment chemical library. Once positive binders are identified, the second part of the process consisted on their binding validation and the extraction of structural information to be used in the design of tighter hit analogs, namely by means of ^{15}N-^{1}H heteronuclear and/or competition STD or WaterLOGSY experiments, among others.

Primary fragment screenings are often successful for a wide range of protein targets; in fact, it is not uncommon to identify various hits from a single screening round. This is especially desirable when their binding modes are slightly different, as it is a good source of inspiration to improve their binding. In the best cases, primary screening may even provide hits with different binding modes or targeting distinct protein sites; which for design purposes is a very valuable scenario as a fragment-linking strategy can be applied. Prarticularly when these sites are adjacent and positive compounds sit next to each other on the protein surface.

In order to diagnose the above situations one can map the binding sites for the various positive fragments by means of ^{15}N-^{1}H HSQC CSP studies, alternatively one can also establish whether compounds share the same binding site using competition cross-saturation experiments (STD or WaterLOGSY). But probably the most informative experiment, not only diagnostic wise but for design purposes, is ILOE.[70] In this experiment inter-ligand nOes, mediated by formation of the ternary complex, are detected and can be used to determine the relative orientation and distance between the two fragments on the protein surface. Later, this information is used to design a convenient linker that will connect the two fragments yielding a dual ligand, hopefully with enhanced binding. Interestingly this information is extracted without the need of labeled protein or assignment and with the cost advantages of ligand based experiments.

ILOE based fragment linking strategy as outlined above [71] offers obvious benefits over schemes relying heavily on receptor-based experiments, for which structural analysis requires prior knowledge on protein assignment and structure; moreover binding site mapping is often fuzzy and unreliable due to distal

conformational changes (see chapter 1.1.4.2 and chapter 2). For these reasons, whenever different fragments are identified in a primary screening, those hits with vicinal binding sites will be prioritized and a chemical linkage devised on the basis of ILOE experiment.

With this general strategy in mind, an all-purpose fragment library was designed and the particular chemical compounds selected, purchased and assembled for screening. The general criteria used to select the chemicals conforming a fragment library have already been mentioned ("rule of three"), in our case we followed somewhat more stringent guidelines by setting molecular weight and lipophilicity upper thresholds at 250 Daltons and 2 for LogP respectively. Molecules were selected from a family of scaffolds similar to those employed in the SHAPES approach; [72, 73] the underlying idea is to select compounds with chemical features commonly found in commercial drugs and filter them according to our molecular weight and lipophilicity thresholds.

Roughly 4-5 compounds per scaffold (Figure 23) were selected and purchased. In terms of logistics, each chemical was given a code, weighted and dissolved in d_6-DMSO at 100 mM as concentrated stocks. Later a ^{1}H NMR spectrum was acquired for each compound under typical NMR assay conditions: 0.5 mM compound in a 50 mM NaCl, and 25 mM phosphate buffer at pH 7. This typical library set up is very convenient, as sample preparation and screening are easily performed. On the other hand, ^{1}H NMR experiments performed for all compounds are very useful as reference spectra, but also in order to verify compound integrity and solubility under experimental conditions; albeit compound solubility should have been addressed to some extent by LogP filter.

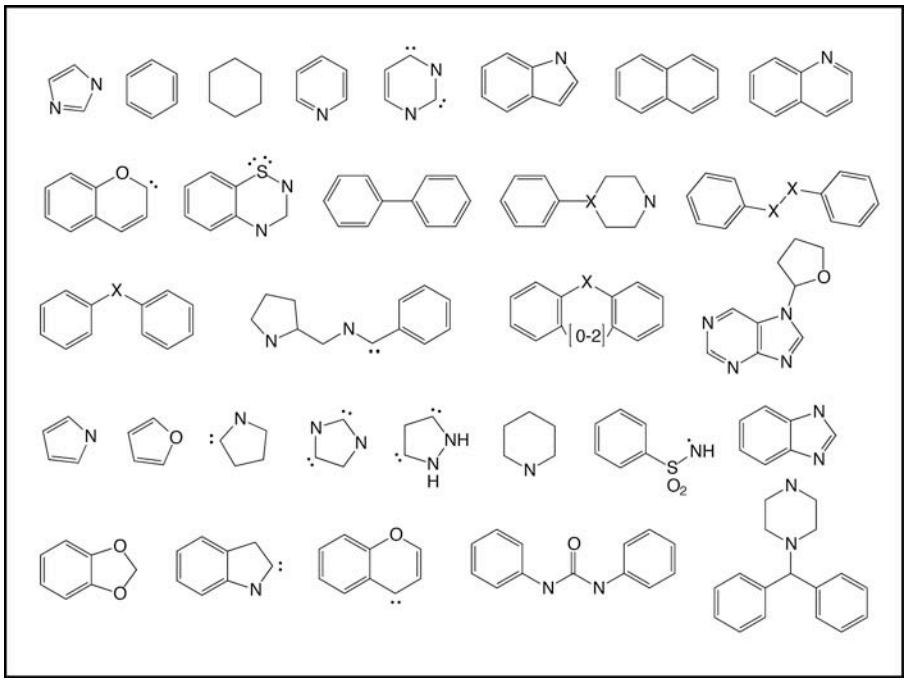


Figure 23 Molecular frameworks used for SHAPES library selection. [72] Attachment points for side chains are indicated by single electrons or lone pairs. 'X' represents a C, N, O or S atom.

1.2.4 SCREENING

Bcl-XL is a widely accepted therapeutic target and in our hands has demonstrated to recognize its natural ligand suggesting this could be a suitable protein for fragment based lead discovery. Once we have set up an all-purpose fragment library, observing well-established guidelines, we will now report the results of its primary screening using WaterLOGSY experiments.

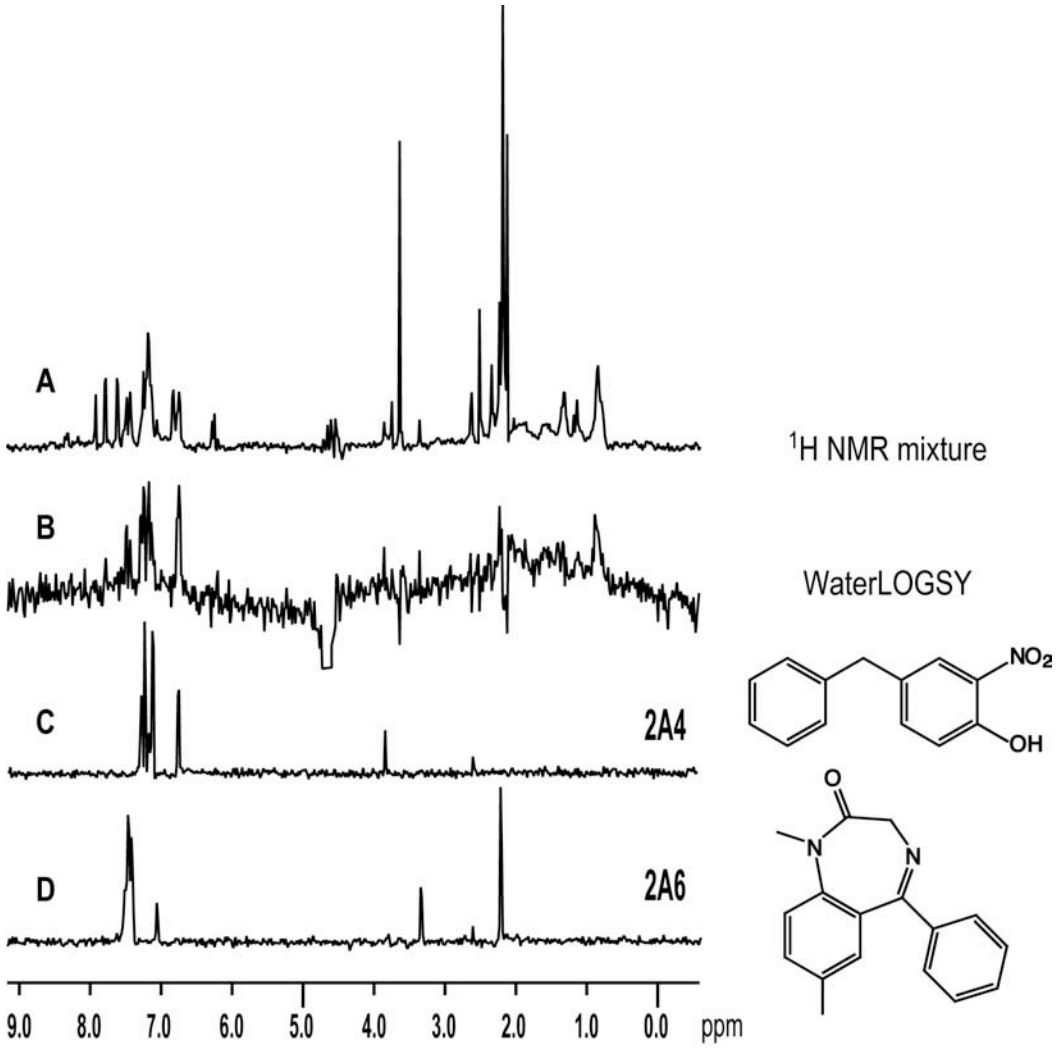


Figure 24 Screening results for sample containing compounds: 1E11, 1E12, 2A1-2A8. A) Watergate ^{1}H experiment for compound mixture in the presence of Bcl-XL. B) Positive WaterLOGSY experiment. C) control ^{1}H NMR experiments for compound 2A4 and D) compound 2A6.

Samples containing 10 μM Bcl-XL together with mixtures of ten compounds (0.5 mM per compound, 25 mM pH 7 phosphate buffer and 50 mM NaCl) were prepared for all chemicals in our fragment library. Next, a WaterLOGSY experiment was performed on every sample and one sample identified to contain positive compounds; this mixture was made up by codename compounds: 1E11, 1E12, 2A1-2A8. Deconvolution proceeded by comparing positive WaterLOGSY experiment with ^{1}H NMR experiments for free compounds;

and as we see in Figure 24 cross-saturation signals in the positive experiment belong to compounds 2A4 and 2A6; 4-benzyl-2-nitrophenol and (Z)-1,7-dimethyl-5-phenyl-1H-benzo[e][1,4]diazepin-2(3H)-one respectively.

Later, we validated 2A4 and 2A6 binding by performing the same WaterLOGSY experiment for the individual compounds in the presence of Bcl-XL, and although we observed residual cross-saturation signals for 2A6 the effect was far greater for benzyl-nitrophenol (2A4) compound.

Altogether, primary screening experiments indicate that both 2A4 and 2A6 bind to Bcl-XL although the latter probably does so with lesser affinity. Interestingly, a BH3I-1 analog ((E)-2-(5-(4-fluorobenzylidene)-4-oxo-2-thioxothiazolidin-3-yl)-3-methylbutanoic acid Figure 25) also experiences cross-saturation a similar WaterLOGSY experiment upon Bcl-XL addition. This is not surprising given the μM affinities reported by Degterev and collaborators;[56] these are within the typically detectable affinity range for cross-saturation experiments. The above primary screening was quite profitable, two novel Bcl-XL binding scaffolds were identified and a BH3-I analog has also been evaluated with positive results.

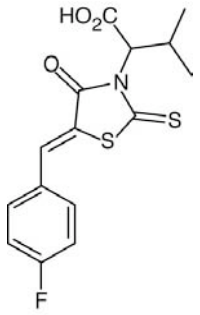


Figure 25 BH3I-1 fluorinated analog: ((E)-2-(5-(4-fluorobenzylidene)-4-oxo-2-thioxothiazolidin-3-yl)-3-methylbutanoic acid.

1.2.5 SCREENING FOLLOW UP

A closer inspection to the positive fragments identified in the previous section (2A4 and 2A6) reveals that they both share a common diphenylmethane motif. This is not surprising, and as a matter of fact diphenylmethane and biphenyl scaffolds are the most recurrent hits in NMR screening programs; some authors suggest this is the result of protein binding sites intrinsic properties.[73] In the current section we will examine in detail binding for 2A4, the simplest of the two hits and the one displaying the better cross-saturation effect, so as to extract information to eventually improve binding.

First we performed a classic CSP study to validate 2A4 binding. For this purpose successive 2A4 additions were performed on a ^{15}N uniformly labeled Bcl-XL and monitored with ^{15}N-^{1}H HSQC experiments. As a result, a large number of cross-peaks become affected upon addition of diphenylmethane compound and, although most of these changes occur in the fast exchange regime, some signals experience significant broadening. However, accurate binding site mapping is not possible due to the large number of affected residues and the difficulty to identify resonances with the published assignment. [55] The latter probably as a result of Bcl-XL's crowded spectrum, typical of highly helical proteins. Even so, some amides could be identified including A168 and R165 in Bcl-XL-BH1 domain; both are remarkably affected by 2A4 addition despite of their distance to the active groove. (See Figure 26) This is consistent with a large rearrangement

occurring on BH3 BH2 and BH1 domains to expose Bcl-XL's hydrophobic cleft; a phenomenon that appears both upon addition of Bak-BH3 peptide, but also in response to the small inhibitors reported by Degterev and collaborators. [56]

A direct consequence of the poor assignment is that our understanding on 2A4 binding is rather limited and for now we will not be able to determine whether 2A4 binding-mode resembles that of Degterev's BH3I-1 compound, BH3I-2 or none of the former. Still, beyond detection of the binding event, ^{15}N-^{1}H HSQC titration provides some very interesting piece of information: 2A4's binding constant. [69]

In order to estimate 2A4 binding constant one must first define an averaged chemical shift difference for amide cross peaks: $\Delta\delta HN_{av}$. (See Equation 3) This value takes into account both ^{1}H and ^{15}N chemical shifts (weighted according to the spectral window of each nucleus) and allows us to reliably quantify chemical shift perturbations on the 2D heteronuclear experiment.

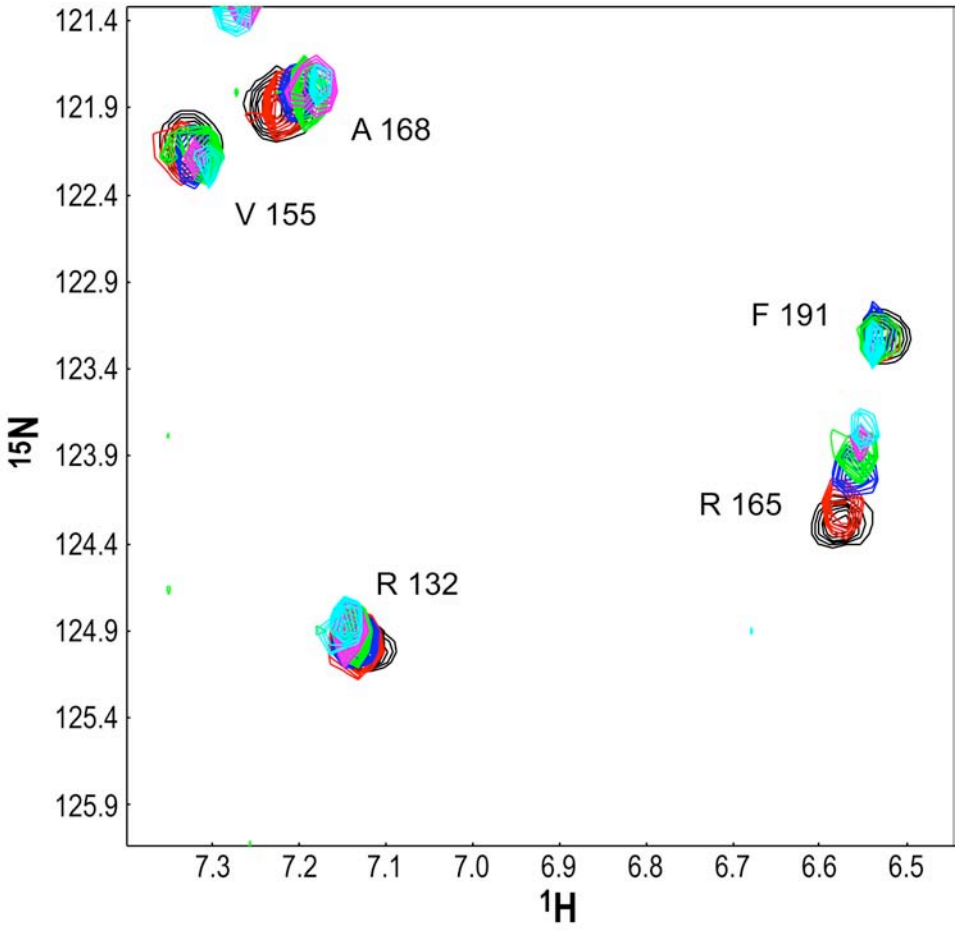


Figure 26 Bcl-XL ^{15}N-^{1}H HSQC experiments in the presence of increasing amounts of 2A4: (black) free Bcl-XL, (red) 1 equivalent 2A4 (blue) 5 eq. (green) 10 eq. (magenta) 15eq. (cyan) 20 eq.

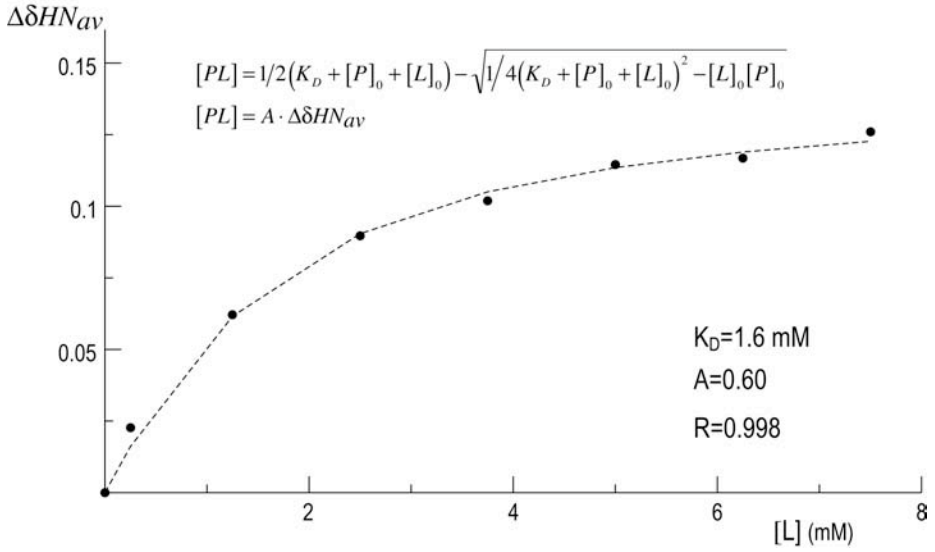


Figure 27 Non-linear curve fitting of $\Delta\delta HN_{av}$(ppm) versus compound concentration (mM) for an amide peak shifted upon 2A4 addition. $\Delta\delta HN_{av}$ for each HSQC experiment are calculated according to Equation 3 and fitted to the presented equation with the resulting K_D, A and R correlation.

$$\Delta\delta HN_{av} = \sqrt{\tfrac{1}{2}\left(\left(\Delta\delta H\right)^2 + \left(\tfrac{1}{5}\Delta\delta N\right)^2\right)}$$

Equation 3 Averaged chemical shift difference for ^{15}N-^{1}H heteronuclear correlations.

$$[PL] = 1/2\left(K_D + [P]_0 + [L]_0\right) - \sqrt{1/4\left(K_D + [P]_0 + [L]_0\right)^2 - [L]_0[P]_0}$$
$$[PL] = A \cdot \Delta\delta HN_{av}$$
$$[P]_0 \cdot A = \left[\Delta\delta HN_{av}\right]_{bound}$$

Equation 4 Used in non-linear curve fitting for $\Delta\delta HN_{av}$ vs ligand concentration plot.

Later, one should determine this averaged chemical shift difference with respect to the free protein spectrum for all titration experiments and desired amides. $\Delta\delta HN_{av}$ values would be plotted versus ligand concentration and non-linearly fitted to Equation 4 in order to obtain K_D.

This exercise was successfully carried out for 2A4 titration and the resulting estimated dissociation constant is 1.6 mM, with a very good correlation coefficient. (See Figure 27) As expected from the observations in ^{15}N-^{1}H HSQC and WaterLOGSY experiments, 2A4 is a weak binder in the high-μM to low-mM range.

1.2.5.1 Competition Experiment and Interligand nOE

Due to the limited structural information regarding 2A4 binding on Bcl-XL, it remains unclear whether this compound shares the same binding site as BH3I-1 type compounds. For this reason a series of

WaterLOGSY experiments were performed in the presence of both compounds: diphenylmethane compound (2A4) and Degterev's fluorinated analog (Figure 25).

Interestingly, addition of the slightly tighter binder (BH3I-1) does not affect cross-saturation effects observed on 2A4, in fact, such effects were observable for both compounds overall suggesting that they do not compete for the same site on the protein surface. This was later confirmed by acquiring NOESY experiments for the ternary mixture. The resulting experiment shows intense negative nOes for both compounds, which is a typical behavior for high MW molecules and a clear confirmation that both compounds are undergoing exchange between bound and free states. The presence of such transferred nOes (trNOE) is very useful to determine bioactive conformations for small flexible molecules, and some authors recommend these experiments as a reliable screening method.[69] But more interestingly, besides intra-molecular nOes we can also identify interligand cross peaks between a 2A4 aromatic and aliphatic protons in BH3I-1 analog; these are also negative nOes and correspond to contacts established on the ternary complex (ILOEs).

ILOEs are very informative in the context of a ternary complex as they demonstrate that two given molecules target different protein binding sites and, should it be the case, one could rule out the possibility of protein-mediated spin-diffusion effects by using QUIET-NOESY experiment. [74] Furthermore, they afford information regarding the relative spatial orientation of the ligands sitting on adjacent sites; in our case 2A4 and BH3I-1 would be facing each other with their methyl and orto-phenolic protons respectively as depicted in Figure 28.

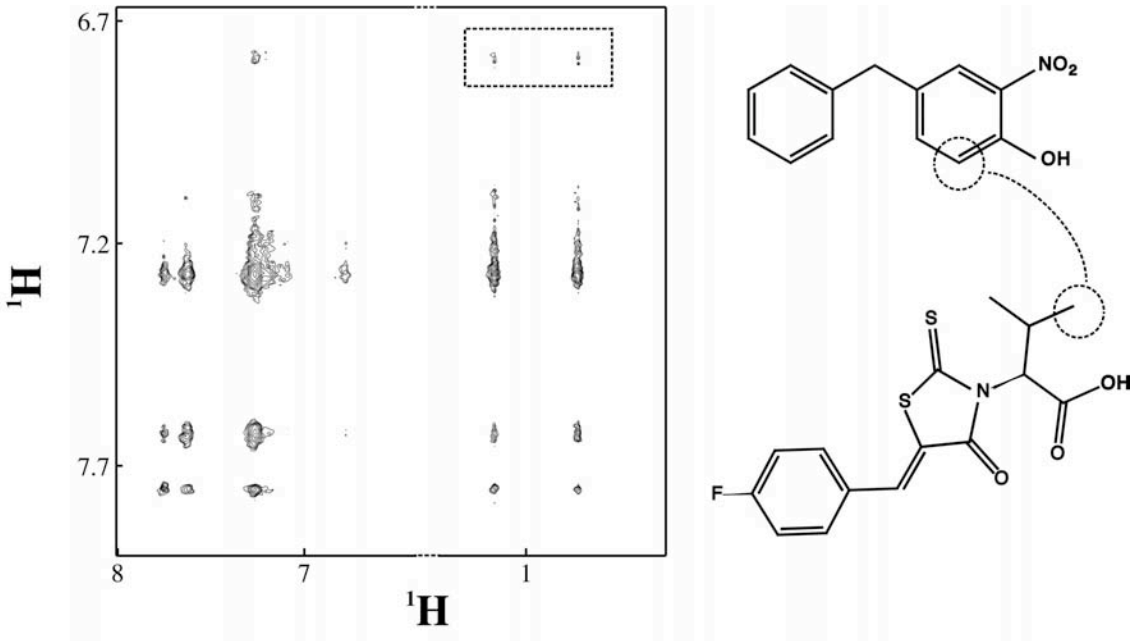


Figure 28 NOESY (τ_m=600 ms) experiment section for 20 μM Bcl-XL sample in the presence of 0.5 μM 2A4 and BH3I-1 fluorinated analog. Relevant iLOES are depicted on spectrum and chemical structures (dotted line).

1.2.6 BILIGAND DESIGN AND FUTURE PROSPECTS

Once the relative positions of Bcl-XL binding fragments have been established one has to devise a way to link them; we will consider various possibilities within this last section. But prior to the design, from a practical perspective, it is important to remind us that chemistry involved in fragment linking should be kept as simple as possible, since a robust chemistry will allow us to explore important parameters such as linker length and chemical nature, before addressing its thorough optimization in later stages.

Following this "keep it simple motto", we should look at 2A4 and BH3-I analogs containing, close to the resonances experiencing iLOEs, chemical functions that can be used in an eventual chemical linkage. Functions such as COOH, OH and NH_2 are particularly interesting; they can easily produce esters and amides under mild conditions, with a well-established scalable chemistry, thanks to the availability of solid phase synthesis strategies and resin-attached reagents.

Analogs with anchoring chemical functions should in principle be tested for their ability to recognize Bcl-XL with the same WaterLOGSY/NOESY experiments, and once binding analogs have been identified a suitable linker is chosen according to the chemical functions involved. Next, in order to determine the best linker length one could use information from *in-silico* docking methods or ILOE build up curve information; however, a more pragmatic approach consists in the synthesis of a small collection of dual ligands with variable length linkers.

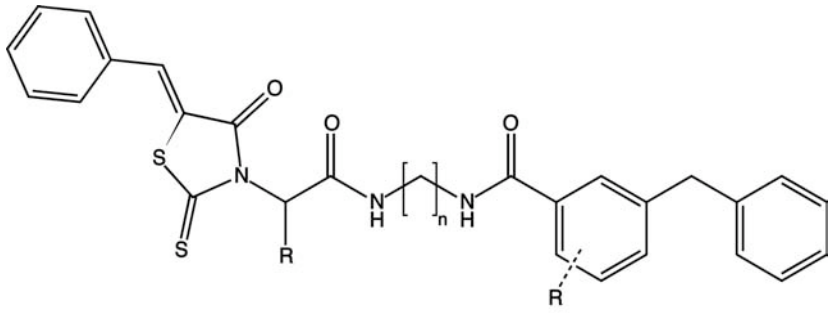


Figure 29 Proposed design for a Bcl-XL dual ligand based on ILOE experiments and using suitable 2A4 and BH3-I' analogs.

With the above considerations, BH3-I 1 compound could use its carboxylic moiety to anchor the amino group of a linker, while a carboxylic containing 2A4 analog could also be reached through a peptide linkage as depicted in Figure 29. In this design all reagents are commercially available and a simple linker exploration can be achieved with limited chemical hassle.

1.3 BIBLIOGRAPHY

1. Danial, N.N. and S.J. Korsmeyer, *Cell death: Critical control points.* Cell, 2004. **116**(2): p. 205-219.
2. Kaufmann, S.H. and M.O. Hengartner, *Programmed cell death: alive and well in the new millennium.* Trends In Cell Biology, 2001. **11**(12): p. 526-534.
3. Brenner, C. and G. Kroemer, *Apoptosis - Mitochondria - the death signal integrators.* Science, 2000. **289**(5482): p. 1150-1151.
4. Deveraux, Q.L., et al., *X-linked IAP is a direct inhibitor of cell-death proteases.* Nature, 1997. **388**(6639): p. 300-304.
5. Huang, Y.H., et al., *Structural basis of caspase inhibition by XIAP: Differential roles of the linker versus the BIR domain.* Cell, 2001. **104**(5): p. 781-790.
6. Riedl, S.J., et al., *Structural basis for the inhibition of caspase-3 by XIAP.* Cell, 2001. **104**(5): p. 791-800.
7. Chai, J.J., et al., *Structural and biochemical basis of apoptotic activation by Smac/DIABLO.* Nature, 2000. **406**(6798): p. 855-862.
8. LaCasse, E.C., et al., *The inhibitors of apoptosis (IAPs) and their emerging role in cancer.* Oncogene, 1998. **17**(25): p. 3247-3259.
9. Fesik, S.W., *Promoting apoptosis as a strategy for cancer drug discovery.* Nature Reviews Cancer, 2005. **5**(11): p. 876-885.
10. Sun, C.H., et al., *NMR structure and mutagenesis of the third Bir domain of the inhibitor of apoptosis protein XIAP.* Journal Of Biological Chemistry, 2000. **275**(43): p. 33777-33781.
11. D'Silva, L., et al., *Monitoring the effects of antagonists on protein-protein interactions with NMR spectroscopy.* J Am Chem Soc, 2005. **127**(38): p. 13220-6.
12. Arac, D., T. Murphy, and J. Rizo, *Facile detection of protein-protein interactions by one-dimensional NMR spectroscopy.* Biochemistry, 2003. **42**(10): p. 2774-80.
13. Bogan, A.A. and K.S. Thorn, *Anatomy of hot spots in protein interfaces.* J Mol Biol, 1998. **280**(1): p. 1-9.
14. Tsai, C.J., et al., *Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect.* Protein Science, 1997. **6**(1): p. 53-64.
15. Liu, Z.H., et al., *Structural basis for binding of Smac/DIABLO to the XIAP BIR3 domain.* Nature, 2000. **408**(6815): p. 1004-1008.
16. Kegg, *Kyoto Encyclopedia of Genes and Genomes.* http://www.genome.jp/kegg/.
17. Duewel, H., et al., *Incorporation of trifluoromethionine into a phage lysozyme: implications and a new marker for use in protein 19F NMR.* Biochemistry, 1997. **36**(11): p. 3404-16.
18. Duewel, H.S., et al., *Elucidation of solvent exposure, side-chain reactivity, and steric demands of the trifluoromethionine residue in a recombinant protein.* Biochemistry, 2001. **40**(44): p. 13167-76.
19. Eichler, J.F., et al., *Biosynthetic incorporation of fluorohistidine into proteins in E. coli: a new probe of macromolecular structure.* Chembiochem, 2005. **6**(12): p. 2170-3.
20. Kim, H.W., et al., *The specific incorporation of labelled aromatic amino acids into proteins through growth of bacteria in the presence of glyphosate. Application to fluorotryptophan labelling to the H(+)-ATPase of Escherichia coli and NMR studies.* FEBS Lett, 1990. **272**(1-2): p. 34-6.
21. Gardner, K.H., M.K. Rosen, and L.E. Kay, *Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR.* Biochemistry, 1997. **36**(6): p. 1389-401.
22. Goto, N.K., et al., *A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins.* J Biomol NMR, 1999. **13**(4): p. 369-74.
23. Lee, A.L., J.L. Urbauer, and A.J. Wand, *Improved labeling strategy for 13C relaxation measurements of methyl groups in proteins.* J Biomol NMR, 1997. **9**(4): p. 437-40.
24. Rajesh, S., et al., *A novel method for the biosynthesis of deuterated proteins with selective protonation at the aromatic rings of Phe, Tyr and Trp.* J Biomol NMR, 2003. **27**(1): p. 81-6.

25. Kigawa, T., Y. Muto, and S. Yokoyama, *Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis.* J Biomol NMR, 1995. **6**(2): p. 129-34.
26. Spirin, A.S., et al., *A continuous cell-free translation system capable of producing polypeptides in high yield.* Science, 1988. **242**(4882): p. 1162-4.
27. Yabuki, T., et al., *Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis.* J Biomol NMR, 1998. **11**(3): p. 295-306.
28. Chong, S., et al., *Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step.* Nucleic Acids Res, 1998. **26**(22): p. 5109-15.
29. Dawson, P.E. and S.B. Kent, *Synthesis of native proteins by chemical ligation.* Annu Rev Biochem, 2000. **69**: p. 923-60.
30. Xu, R., et al., *Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies.* Proc Natl Acad Sci U S A, 1999. **96**(2): p. 388-93.
31. Chin, J.W., et al., *An expanded eukaryotic genetic code.* Science, 2003. **301**(5635): p. 964-967.
32. Chin, J.W., et al., *Progress toward an expanded eukaryotic genetic code.* Chemistry & Biology, 2003. **10**(6): p. 511-519.
33. Pan, P., E. Woehl, and M.F. Dunn, *Protein architecture, dynamics and allostery in tryptophan synthase channeling.* Trends Biochem Sci, 1997. **22**(1): p. 22-7.
34. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.
35. Scott, F.L., et al., *XIAP inhibits caspase-3 and -7 using two binding sites: evolutionarily conserved mechanism of IAPs.* Embo J, 2005. **24**(3): p. 645-55.
36. Shin, H., et al., *The BIR domain of IAP-like protein 2 is conformationally unstable: implications for caspase inhibition.* Biochem J, 2005. **385**(Pt 1): p. 1-10.
37. Clinton, B., *Successful presidential campaign.* 1992.
38. Woodward, C., N. Carulla, and G. Barany, *Native state hydrogen-exchange analysis of protein folding and protein motional domains*, in *Energetics Of Biological Macromolecules, Pt E*. 2004. p. 379-400.
39. Kessler, H., *Conformation and Biological Activity of Cyclic Peptides.* Angewandte Chemie International Edition in English, 1982. **21**(7): p. 512-523.
40. Pervushin, K., et al., *Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution.* Proc Natl Acad Sci U S A, 1997. **94**(23): p. 12366-71.
41. Ollerenshaw, J.E., V. Tugarinov, and L.E. Kay, *Methyl TROSY: explanation and experimental verification.* Magnetic Resonance In Chemistry, 2003. **41**(10): p. 843-852.
42. Pervushin, K., et al., *Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in C-13-labeled proteins.* Journal Of The American Chemical Society, 1998. **120**(25): p. 6394-6400.
43. Palmer, A.G., 3rd, *NMR characterization of the dynamics of biomacromolecules.* Chem Rev, 2004. **104**(8): p. 3623-40.
44. Bann, J.G., et al., *Real-time and equilibrium (19)F-NMR studies reveal the role of domain-domain interactions in the folding of the chaperone PapD.* Proc Natl Acad Sci U S A, 2002. **99**(2): p. 709-14.
45. Dalvit, C., et al., *A general NMR method for rapid, efficient, and reliable biochemical screening.* J Am Chem Soc, 2003. **125**(47): p. 14620-5.
46. Bohm, H.J., et al., *Fluorine in medicinal chemistry.* Chembiochem, 2004. **5**(5): p. 637-43.
47. Khodursky, A.B., et al., *DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in Escherichia coli.* Proc Natl Acad Sci U S A, 2000. **97**(22): p. 12170-5.

48. Borden, K.L., P. Beckmann, and A.N. Lane, *Determination of the orientations of tryptophan analogues bound to the trp repressor and the relationship to activation.* Eur J Biochem, 1991. **202**(2): p. 459-70.
49. Ramesh, V., et al., *NMR studies of the mode of binding of corepressors and inducers to Escherichia coli trp repressor.* Eur J Biochem, 1996. **235**(3): p. 804-13.
50. Marmorstein, R.Q. and P.B. Sigler, *Stereochemical effects of L-tryptophan and its analogues on trp repressor's affinity for operator-DNA.* J Biol Chem, 1989. **264**(16): p. 9149-54.
51. Matchett, W.H., *Inhibition of tryptophan synthetase by indoleacrylic acid.* J Bacteriol, 1972. **110**(1): p. 146-54.
52. Browne, D.T., G.L. Kenyon, and G.D. Hegeman, *Incorporation Of Monofluorotryptophans Into Protein During Growth Of Escherichia-Coli.* Biochemical And Biophysical Research Communications, 1970. **39**(1): p. 13-&.
53. Gerig, J.T., *Fluorine NMR of Proteins.* Prog Nuc Magn Spectr, 1994. **26**: p. 293-370.
54. Sattler, M., et al., *Structure of Bcl-x(L)-Bak peptide complex: Recognition between regulators of apoptosis.* Science, 1997. **275**(5302): p. 983-986.
55. Lugovskoy, A.A., et al., *A novel approach for characterizing protein ligand complexes: molecular basis for specificity of small-molecule Bcl-2 inhibitors.* J Am Chem Soc, 2002. **124**(7): p. 1234-40.
56. Degterev, A., et al., *Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL.* Nat Cell Biol, 2001. **3**(2): p. 173-82.
57. Medek, A., et al., *The use of differential chemical shifts for determining the binding site location and orientation of protein-bound ligands.* Journal Of The American Chemical Society, 2000. **122**(6): p. 1241-1242.
58. Yin, H., et al., *Terephthalamide derivatives as mimetics of helical peptides: disruption of the Bcl-x(L)/Bak interaction.* J Am Chem Soc, 2005. **127**(15): p. 5463-8.
59. Rutledge, S.E., J.W. Chin, and A. Schepartz, *A view to a kill: ligands for Bcl-2 family proteins.* Curr Opin Chem Biol, 2002. **6**(4): p. 479-85.
60. Wang, J.L., et al., *Structure-based discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells.* Proc Natl Acad Sci U S A, 2000. **97**(13): p. 7124-9.
61. Enyedy, I.J., et al., *Discovery of small-molecule inhibitors of Bcl-2 through structure-based computer screening.* J Med Chem, 2001. **44**(25): p. 4313-24.
62. Kutzki, O., et al., *Development of a potent Bcl-x(L) antagonist based on alpha-helix mimicry.* J Am Chem Soc, 2002. **124**(40): p. 11838-9.
63. Yin, H., et al., *Terphenyl-Based Bak BH3 alpha-helical proteomimetics as low-molecular-weight antagonists of Bcl-xL.* J Am Chem Soc, 2005. **127**(29): p. 10191-6.
64. Ernst, J.T., et al., *Design and application of an alpha-helix-mimetic scaffold based on an oligoamide-foldamer strategy: antagonism of the Bak BH3/Bcl-xL complex.* Angew Chem Int Ed Engl, 2003. **42**(5): p. 535-9.
65. Chin, J.W. and A. Schepartz, *Design and Evolution of a Miniature Bcl-2 Binding Protein.* Angew Chem Int Ed Engl, 2001. **40**(20): p. 3806-3809.
66. Hajduk, P.J., J.R. Huth, and S.W. Fesik, *Druggability indices for protein targets derived from NMR-based screening data.* J Med Chem, 2005. **48**(7): p. 2518-25.
67. Muchmore, S.W., et al., *X-ray and NMR structure of human Bcl-x(L), an inhibitor of programmed cell death.* Nature, 1996. **381**(6580): p. 335-341.
68. Pellecchia, M., D.S. Sem, and K. Wuthrich, *NMR in drug discovery.* Nat Rev Drug Discov, 2002. **1**(3): p. 211-9.
69. Meyer, B. and T. Peters, *NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors.* Angewandte Chemie-International Edition, 2003. **42**(8): p. 864-890.
70. London, R.E., *Theoretical analysis of the inter-ligand overhauser effect: a new approach for mapping structural relationships of macromolecular ligands.* J Magn Reson, 1999. **141**(2): p. 301-11.

71. Becattini, B. and M. Pellecchia, *SAR by ILOEs: an NMR-based approach to reverse chemical genetics.* Chemistry, 2006. **12**(10): p. 2658-62.
72. Fejzo, J., et al., *The SHAPES strategy: an NMR-based approach for lead generation in drug discovery.* Chem Biol, 1999. **6**(10): p. 755-69.
73. Hajduk, P.J., et al., *Privileged molecules for protein binding identified from NMR-based screening.* J Med Chem, 2000. **43**(18): p. 3443-7.
74. Becattini, B. and M. Pellecchia, *SAR by ILOEs: An NMR-Based Approach to Reverse Chemical Genetics.* Chemistry, 2005.