Tesi Doctoral

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

Ricard A. Rodríguez Mias

UNIVERSITAT DE BARCELONA

Departament de Química Orgànica

Facultat de Química

Universitat de Barcelona

Barcelona, Juliol 2006

2 VEGF ANTAGONIST DESIGN

Formation of the cardiovascular system is a crucial event in embryogenesis as well as during general development of an organism. Early vasculogenesis is characterized by endothelial cell differentiation followed by the formation of a tubular network of cells. This process encompasses several steps, including: endothelial cell activation, cell migration, extra cellular matrix invasion and capillary lumen formation. The newly created vessel network is then stabilized by vessel wall reinforcement and cessation of cell proliferation. Albeit physiological angiogenesis continues in adults, it occurs at a much reduced level and is primarily involved with vasculature maintenance and wound healing.[1]

Angiogenesis is also implicated in several pathologies. It plays an essential role in the growth of most primary tumors as well as their subsequent metastasis. Sufficient nutrients and oxygen are needed by tumors, and growth beyond a certain size requires the elaboration of vascular supply. This is usually done by recruiting neighboring mature vasculature and sprouting of new capillaries that eventually infiltrate the tumor mass.[2, 3]

Physiological or tumor angiogenesis is nonetheless regulated by a variety of proteins and growth factors. Some of which are highly specific for endothelial cells, other proteins target fibroblasts (*e.g.*, bFGF), or have a much broader effect, such as that of metalloproteases (MMPs). But probably one of the most important factors in this list is Vascular Endothelial Growth Factor (VEGF, or VEGF-A); this belongs to the PDFG supergene family that also includes VEGF-B, VEGF-C, VEGF-D and VEGF-E, with which the former shares varying degrees of homology. VEGF is most commonly expressed as its 165 amino acid isoform, a 45 kD homodimeric glycoprotein, but is also found in other splicing (VEGF$_{145}$, VEGF$_{189}$, VEGF$_{121}$, VEGF$_{189}$, VEGF$_{206}$) and proteolytic isoforms such as VEGF$_{110}$, the smallest of its active forms. All these isoforms share a common N-terminus receptor-binding domain that is structured as a covalent antiparallel homodimer (see Figure 30). [4] Each monomer has a central antiparallel four-stranded β-sheet core and a characteristic cysteine knot motif, fixed by a network of three disulfide bonds.
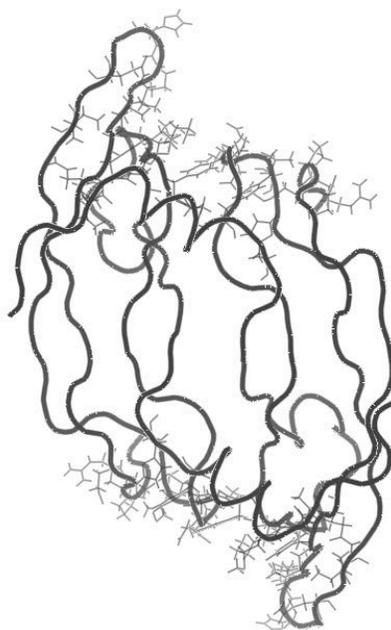

Figure 30 VEGF$_{10-109}$ receptor-binding domain. Side chains for receptor binding epitope residues are shown in lines.

These covalent linkages are quite important to the fold stability as the protein has a considerable lack of hydrophobic core, being very exposed to the solvent. Equally important are the two symmetrical inter-monomer disulfide bonds established between cysteines 51 and 60, which bear most of the burden of dimer stability together with a set of hydrophobic contacts around the receptor-binding epitope. The latter epitope is composed by parts of both monomers and encompasses a broad region of the dimer interface (Figure 30); interestingly, as revealed by both X-RAY and NMR studies, this is a poorly defined region with several accessible conformations and high B-factors. This conformational flexibility probably has important functional implications and allows the protein to interact with multiple receptors.[5]

VEGF gene expression is triggered by several host stimuli such as estrogen, nitric oxide and other growth factors (e.g., bFGF, TNF-α, EGF); but its production is particularly sensitive to oxygen tension, for instance the hypoxic conditions typically found in tumors rapidly upregulate VEGF. As a result, VEGF is over expressed in most tumors and has a direct effect on their development by relieving tumor masses from stress conditions through oxygen and nutrients recruitment, but also acting as a survival factor by enhancing the expression of anti-apoptotic factors such as Bcl-2.
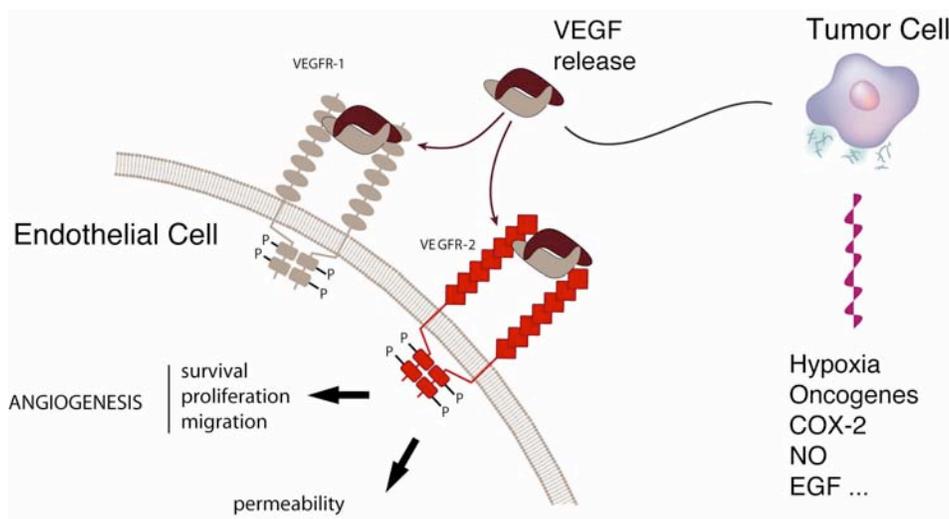

Figure 31 Scheme depicting VEGF-dependent angiogenesis.

VEGF activity is mediated by binding of the protein to different tyrosine kinase membrane receptors, VEGFR-1 (Flt-1), VEGFR-2 (Flk-1) and VEGFR-3. [4] In general, VEGF induces receptor dimerization, thanks to its symmetric nature (see figures Figure 30 and Figure 31), and activation of the receptor's kinase activity through autophosphorylaliton. This in turn induces phosphorylation on a variety of protein substrates, activating a series of intracellular signaling pathways and unchaining several processes common to other growth factors: cell migration, survival and proliferation. However, the peculiarity of these VEGF receptors is its unique ability to transduce these signals in a three-dimensional fashion and synchronize the various angiogenic events to build a new vascular tube. In part, this fine control on the transmitted signal is achieved through the different VEGF isoforms; which thanks to their ability to recruit different co-receptors are able to determine how the growth factor is distributed on the target cells. Another sort of modulation is achieved by the various VEGF receptor types; in this sense most of the load on the vascularization process is typically attributed to VEGFR2, while VEGFR1 receptor is often associated with a down regulation function acting as

decoy receptor; as for VEGFR3 receptors, they are thought to intervene mainly in lymphatic vasculature sprouting.

As a result from the fundamental role of vessel sprouting in tumor growth, antiangiogenic strategies have been proposed to treat cancer; in fact, several key steps in the vascularization process and their effectors have been targeted in the search for effective drugs, including endothelial cell proliferation, signal transduction and matrix metalloproteases.[2] VEGF is particularly interesting as a therapeutic target, not only for its central role in the regulation of angiogenesis, but also due to the fact that it acts directly on genetically stable endothelial rather than on tumor cells. Thus, compounds that inhibit this hormone will be less prone to induce mutation selection on tumor cells and ultimately confer them with drug resistance. Efforts in this direction have provided VEGF inhibitors that are currently being clinically pursued; the list includes several small molecules that target VEGF receptor kinase activity: PTK787, SU11248, etc.; but also various antibodies that aim to block hormone/receptor interaction. Avastin is one of such antibodies, this recognizes VEGF and successfully hinders its receptor-binding interface producing the desired antiangiogenic effects; for which it has recently been approved by the FDA as a first-line therapy for colorectal cancer.[6] However, this is not the only example as there are also antibody therapies targeting VEGFR-2 receptor.

Despite the advantages of antibody therapies its therapeutic use is still limited, primarily as a result of its high cost but also due to the risk of immunological reactions in patients under prolonged treatments. On the other hand, the use of enzymatic inhibitors is also far from perfect, as they are often plagued with side effects stemming from lack of specificity. An alternative strategy, which constitutes the ultimate goal in the present chapter, is the use of small and medium sized synthetic compounds to disrupt VEGF-receptor interaction. Protein-protein interactions are still a major challenge to drug design, but previous structural works on various VEGF complexes indicate there is room for optimism, as its relatively large binding epitope is targetable by medium sized peptides[7] and other organic molecules.[8] In this sense we intent to use an NMR based methodology to screen for inhibitors in very diverse compound libraries: first a set of all D-oligo peptides and later a collection of water plant extracts.

2.1   LIGAND DESIGN AND LIBRARY SET UP

Guidelines for the construction of chemical libraries have already been mentioned in the Introduction. In this context the Lipinski's rules are often brought up to filter compounds included in "lead-like" libraries, which are eventually assayed in traditional HTS programs. An analogous set of rules has been developed to guide the construction of "fragment-like" libraries. [9] The latter are often relatively small but diverse libraries, since their ability to cover large portions of the chemical space stems from the sheer number "theoretical" compounds or combinations one could obtain by linking the various fragments, rather than the actual number of fragments. In any case, often one of the most important considerations in the design of chemical libraries is not fragment or drug-likeness but the particular screening experiments or the follow up strategies that will occur once a hit has been identified in the library.

In our search for novel VEGF-receptor interaction inhibitors we will use two diametrically opposed strategies, each strategy with important implications as to the type of screened compounds and the experiments used in their screening. Our first approach will be inspired on the fragment-based strategy used in the first chapter; however, instead of using the preferred scaffold concept our fragment library will be formed by short oligo-peptides.

The second approach will be a much more conventional lead search. In this case a complex library of plant extracts will be screened for VEGF binding compounds. Eventually, traditional medicinal chemistry strategies will be used to explore the chemical space around hit compounds in order to produce viable lead/s.

2.1.1   FRAGMENT BASED LIBRARY: PEPTIDE LIGANDS

As already stated our first chemical library will participate in a "bottom-up" drug discovery scheme, the novelty, or rather the audacity of our approach, will be the particular compound or fragment selection.

In this sense, oligo-peptides present several interesting features. First of all, they are very modular compounds, easily extensible and very accessible due to their well-established chemistry. But more importantly, with a limited number of building blocks -20 proteinogenic amino acids- it is possible to cover a wide chemical space. In principle, one could reproduce any interaction in nature since the same set of fragments intervenes in all protein-protein interaction.

However, from a "fragment-likeness" standpoint oligo-peptides are probably not the best choice and are noticeably not compliant with the "rule of three". They have at least two rotatable bonds per residue; additionally, amide linkage together with side chain functionality may easily push us beyond the hydrogen bond donor/acceptor rule. The former considerations notwithstanding, dipeptides may still be regarded as the most "reasonable" fragments -within oligo-peptides-, as they usually weight no more than 300 Da and the rest of physicochemical parameters are not too far from the "fragment-like" guidelines; while having enough chemical functions to allow a reasonable number of interactions with the target protein

Another inconvenient typically brought up to dismiss the use of peptides as drugs is their lability to proteases, nevertheless one can easily avoid the issue by using D-amino acids instead of their L homologues. All in all, if one were to build all the possible dipeptides using the 20 proteinogenic amino acids -in their D-configuration- the list would add up to 400. Furthermore, if modifications such as amidation or

acetylation were included in both peptide termini, the list easily reaches a few thousand dipeptides, and given the high economic cost of D-amino acids the approach becomes unfeasible.

Instead, we decided to pursue the synthesis of a smaller set of dipeptides and these were selected according to two different considerations: first using a reported peptide ligand as inspiration, and later attending to diversity criteria.

2.1.1.1 *v107 INSPIRED:*

So far, the most remarkable peptidic VEGF ligands were identified by Fairbrother and collaborators, thanks to a phage display strategy. Their meticulous work was rewarded with three different families of peptides; all of them were able to recognize VEGF's receptor binding epitope, although, as depicted by either NMR or X-Ray structures, each used a slightly different binding mode.[7, 10, 11] v107 is the codename for one of such peptides; its complex with VEGF has been extensively characterized and beyond its structure determination Fairbrother and coworkers also carried out a thorough alanine scanning study that provided invaluable energetic information on the complex. (see Figure 32) Supported by this broad literature we decided to use v107 peptide as mold and inspiration for our dipeptide library. The idea was to dissect this 19mer peptide into all sequentially possible dipeptides, in this way we expected some of the dipeptides would be able to establish, at least partially, similar interactions to the ones performed by v107 in the complex.

Without taking into account cysteine residues and the last four N-termini residues –unimportant to the light of the NMR structure [11] – there are ten different possible dipeptides arising from this dissection. (Figure 33) Two different N-terminus versions were selected for our library: free amino and acetylated N-terminus, the former for its beneficial effects in terms of solubility, whiles the latter to better simulate a peptide environment. For the same reason an amidated C-terminus modification was preferred over the free acid, in addition to the synthetic benefits since the same strategy can be used for whole dipeptide library.
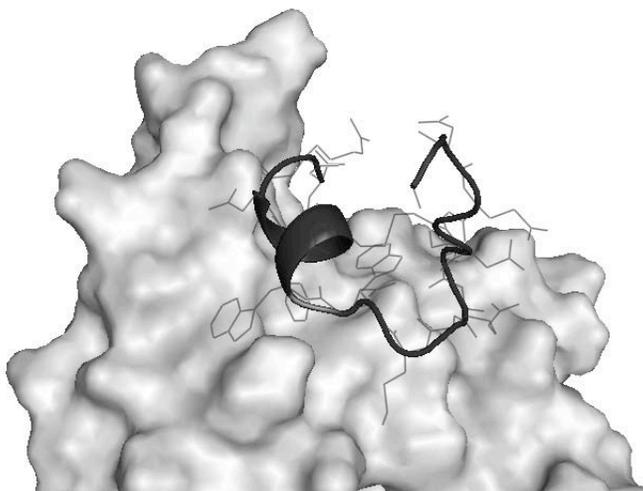

Figure 32 NMR structure for VEGF in complex with v107 peptide.[11]

Nonetheless, the difficulty appears when we impose D configuration to our v107 inspired dipeptide library. The former phage display peptide is by nature an all-L peptide and thus, peptide dissection will unavoidably

lead to a family of L dipeptides that are very different, in terms of recognition properties, to their D counterparts, especially in front a chiral receptor such as our protein. Then, how can we use v107-dissected dipeptides to build a library of D-dipeptides with similar recognitions properties? To solve this issue several strategies have been proposed, but most often the retro-inverso approach is used; here both peptide sequence and amino acid configuration are reverted in such a way that amino acid side-chain orientation is retained.[12]

For this strategy to work there are of course several underlying assumptions: the first one is that interactions should mainly be established through the side chains, as the amide pattern for the retro-inverso peptide does not match those of the L-peptide; thus main chain interaction would not be reproducible by the retro-inverso peptide. Another important consideration is that side chain orientation in retro-inverso and wild type peptides will have the best overlap when in extended conformation and thus this strategy is recommended for β-sheet secondary structures rather than α-helix. In our particular case the above limitations may not be too important; since v107's most relevant interactions are established through hydrophobic side chains particularly with: Phe16, Trp11, Leu19 and Met10 while backbone contributions are nearly nonexistent. On the other hand, D-dipeptides are very short and most likely they will be able to orient side chains so as to mimic α−helix or β−strand conformation without a large energetic penalty.

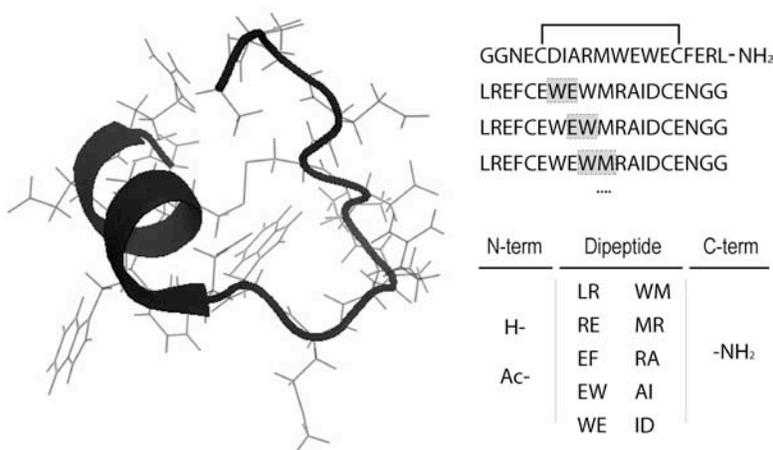 All in all, with these design criteria the total number of v107-inspired dipeptides is 20. (see Figure 33)

Figure 33 Phage display derived peptide v107 displayed in its VEGF bound conformation, and v107 inspired dipeptides. Retro-inverso approach is applied to original v107 sequence, and relevant possible dipeptides selected including various N-terminus modifications (see text for more details). (A) D-Ala, (E) D-Glu, (S) D-Ser, (R) D-Arg, (I) D-alloIle, (W) D-Trp, (M) D-Met, (D) D-Asp.

2.1.1.2 *AMINO ACID REDUCED BASE*

The second approach to reduce the number of dipeptides was somewhat more general, and intended to maximize the explored chemical rather than trying to use available structural information for this purpose.

In order to build this second library of all-D dipeptides a reduced group of 6 amino acids (S, E, A, I, W and R) was selected from the 20 proteinogenic residues. This selection tried to cover the different chemical traits or functionalities present in the whole amino acid set, these are; hydroxyl, acid, aliphatic, aromatic and basic

side chains. We choose from each class a single residue taking into account to their hot spot occurrence [13], secondary structure propensities and cost. For instance, in the particular case of aromatic amino acids tryptophan was chosen over phenylalanine or tyrosine since Bogan and collaborators have found that W is found in 23% of protein hot spots compared to Y and F with 13.3 and 3.1% respectively. The same situation occurs for basic residues, for which arginine was chosen due to its higher hot spot frequency; finally, glutamic and serine were chosen over their counterparts due to much more prosaic reasons: cost and synthetic accessibility. The above reasons are somewhat arbitrary, and one could argue more sophisticated criteria to reduce the number of amino acids; one could use the appearance order throughout evolution, or even similarity distance matrices [14] to come up with a reasonable set of six amino acids. Whichever the selection criteria, by reducing the number of building blocks to 6 means we will have to produce 36 dipeptides. With regard to the termini modifications we decided to use a combination of free N-terminus and amidated C-terminus for the sake of solubility and ease of synthesis. Overall, the above criteria yields the dipeptide list depicted in Figure 34.

N-term	Dipeptide	C-term
	A-X	
	E-X	
H-	S-X	-NH₂
	R-X	
	I-X	
	W-X	

X= A, E, S, R, I, W

Figure 34 All-D dipeptide library constructed using reduced amino-acid base. (A) D-Ala, (E) D-Glu, (S) D-Ser, (R) D-Arg, (I) D-alloIle, (W) D-Trp.

Synthesis of both peptide libraries was kindly performed by Francesc Yraola at Combinatorial Chemistry Unit in the PCB. Compounds were produced and purified as described in the Materials and Methods section by Solid Phase Synthesis and HPLC respectively, to obtain compounds above 95% purity. Library logistics were standard and followed the same guidelines as in the previous chapter. Briefly, this consisted in the preparation of 100 mM DMSO-d$_6$ concentrated stocks; followed by compound identification and solubility confirmation under the same conditions used in binding experiments –namely phosphate buffer– by performing ^{1}H monodimensional experiments.

2.1.2 CONVENTIONAL HTS LIBRARY: TRADITIONAL CHINESE MEDICINAL PLANTS

Our second approach to the discovery of VEGF inhibitors was much more conventional; instead of the fragment discovery approach, the idea was to screen for "lead-like" binders within large collections of molecules. For this "top-down" strategy we chose to screen a set of plant water extracts, these are very complex mixtures and should contain a large number of bioactive chemical entities.

The use of plants to treat ailments is a very antique practice; this knowledge has been gathered through millennia in each particular environment by nearly every human society, from the gatherer-hunters in the Amazonian rainforest to the highly developed Chinese empire. It was not until the 19th century that the

active components (i.e. morphine, salycilic acid, etc) started being isolated paving the road to modern medicinal chemistry and pharmacology. This trend has continued until today, however western medicine has still not unveiled all the hidden power in plants and in many parts of the planet "modern" treatments still coexist with herbal remedies. It is not surprising then, that most pharmaceutical companies regard plants or animals as a very rich source of active chemicals in the treatment of virtually every condition; and consequently include within their drug discovery programs strategies to exploit folkloric medicinal knowledge.

Obviously, among the conditions treated by traditional medicine –using a number of plants or herbal mixtures– there is cancer. Some of the preparations with alleged anticancer activity have endured scientific inspection and demonstrated to be active both *in vivo* and *in vitro* [15, 16]; serving as stepping stone towards the isolation of potent anticancer phytochemicals (i.e. Taxol) [17] and in some cases even developed into commercial drugs. Interestingly, some of these antitumoral plants or compounds operate through an antiangiogenic mechanisms, modulating one or various targets involved in the vessel formation process. [16]

Traditional Chinese medicine (TMC) is one of the few folkloric medicinal practices that still persist nowadays; in fact, in China it is an integral part of the health care system and traditional and "western" treatments are often prescribed simultaneously. Part of this success lays on the extensive empirical knowledge collected and organized through thousands of years into various branches, including acupuncture or herbal medicine as some of its most important modalities. In the latter, herbal prescriptions are usually individually tailored cocktail of many herbs, in contrast to western treatments that frequently rely on single active principles; in this way Chinese herbal preparations seek a synergistic effect between the various components termed *fufang* [15]. Within these preparations some plants exert the desired medicinal effect, while others serve as catalysts or minimize toxicity and adverse effects. As a result, thousands of plants have been included into such cocktails and, through trial and error, classified and compiled according to their effects; all in all constituting one of the largest herbal knowledge base on the face of the planet, and a magnificent source of potential pharmaceutical information.

As a group, our initial interest in traditional Chinese medicinal herbs was connected to the discovery of novel protease inhibitors; particularly prolyloligopeptidase (POP) involved in several neurological affections. For this reason, a small subset of herbs (50 aprox.) (Table 2) was selected from the TMC literature due to their recurrent use in cocktails prescribed for conditions affecting the nervous central system. [18] In this way we expected that the knowledge extracted through thousands of years would help us identify POP inhibitors. 3-FABS [19, 20] screening technology was applied and POP specific reagents were developed to allow ^{19}F NMR monitorization of substrate proteolysis or "lack of such" in the presence of inhibitors. In this way, as reported in the work by Tarragó et al., many plant extracts were identified as containing inhibitory activity. Later, using similar methods, Frutos and collaborators tested the same set of plant extracts against HIV-1 protease, which is crucial to AIDS virus assembly [21, 22]; surprisingly, this effort was equally successful even though plants were not picked for this purpose. These observations are a probably a consequence of the complexity and diversity of such plant extracts and surely explains their widespread success in the treatment of multiple diseases.

To the light of the previous observations we decided to carry out a screening for VEGF inhibitors within our herb extracts library. Dried material for roughly 50 plants was extracted with water in a Soxhlet apparatus;

later water was removed by evaporation and lyophilized to obtain a solid residue. Finally concentrated d_6-DMSO stocks were prepared from this residue at 200 mg/mL concentration.

Table 2 List of TMC plants for which water extracts have been included in our library

Library Code	Plant name	Library Code	Plant name
X1	*Medulla Junci*	X23	*Ganoderma*
X2	*Flos daturae metelis*	X24	*Radix rehmanniae*
X3	*Herba hedyotis diffusae*	X25	*Flos inulae*
X4	*Herba lophatheri*	X29	*Herba myrsines africanae*
X5	*Folium mori*	X30	*Radix angelicae sinensis*
X6	*Radix Paeoniae rubra*	X31	*Haematitum*
X7	*Caulis polygoni*	X32	*Radix curcumae*
X8	*Arillus longan*	X33	*Ramullus cinnamoni*
X9	*Cortex cinnamoni*	X34	*Magnetitum*
X10	*Radix et Rhizoma Rhei*	X37	*Radix scrophulariae*
X11	*Rhizoma chuanxiong*	X38	*Radix notoginseng*
X12	*Radix ginseng*	X39	*Os draconis*
X13	*Lapis chloriti*	X40	*Herba thlaspis*
X14	*Herba dendrobii*	X41	*Fructus jujubae*
X15	*Rhizoma polygonati odorati*	X42	*Cortex meliae*
X16	*Rhizoma coptidis*	X43	*Concretio silicea bambusae*
X17	*Carapax trioncys*	X44	*Fructus lycii*
X18	*Poria*	X45	*Poria spirit*
X19	*Radix scutellariae*	X46	*Bombyx batryticatus*
X20	*Fructus gardeniae*	X47	*Semen ziziphi spimosae*
X21	*Radix polygalae*	X48	*Ginkgo biloba*
X22	*Colla corii asini*		

2.2   PROTEIN PRODUCTION AND SYSTEM SET UP

Once compound libraries have been designed and prepared the remaining part of the puzzle is the protein; in the current section we will deal with its production and evaluation of spectroscopic and recognition properties.

Plasmids p6XHisVEGF$_{11-109}$ and pMS421 for VEGF$_{11-109}$ expression were kindly provided by Dr. Wayne Fairbrother at Genentech and protein was produced as described by the authors in their work on the NMR assignment of this construct [23]. Briefly, *E. coli* cells were transformed with both plasmids and cells grown in carbenecillin containing M9 minimal media until optical density at 600nm reached 0.8, to be induced with 1 mM IPTG for 5 hours. The protein was over expressed into inclusion bodies, for this reason a denaturing purification protocol followed by refolding and subsequent purification steps were applied. Harvested cells were dissolved in 6 M guanidine HCl, 0.1 M NaH$_2$PO$_4$, 10 mM Tris, 10 mM 2-mercaptoethanol and the filtered lysate was applied to a Ni affinity column under denaturing purification conditions; the protein was eluted with 8 M urea, 0.1 M NaH$_2$PO$_4$, 10 mM Tris, 0.5 M imidazole, pH 5.9. Protein containing fractions were selected, brought to 1mg/ml concentration and reduced using 20 mM DTT during 3 hours prior to refolding. Refolding proceeded by stepwise removal through dialysis of denaturing and reducing agents from the initial solution (8 M urea, 20 mM cysteine) at 4°C. After His-tag proteolytic cleavage using Genenase I an anion exchange purification step was performed, VEGF containing fractions were concentrated and purified once more on a S75 size exclusion column into the NMR sample buffer (25 mM phosphate buffer pH 7, 50 mM NaCl). We successfully reproduced all expression and purification steps, and in our hands protein yields were excellent, (up to 30mg/L of media) to obtain both unlabeled and ^{15}N uniformly labeled VEGF.

Fold integrity and overall VEGF$_{11-109}$ spectroscopic properties were assessed using ^{15}N-^{1}H HSQC NMR experiments (see Figure 35). In general, spectrum features suggest that we are before a well-folded protein; and indeed, a closer inspection of NMR data allowed us to identify protein amides using the published assignment by Fairbrother and collaborators.[23]

In order to validate VEGF activity we decided to test its recognition properties towards a published ligand. As already mentioned, several peptide ligands for VEGF have been recently developed, and their ability to recognize its receptor binding epitope was demonstrated both in vivo and in vitro.[11] Some of these peptides are very attractive in their binding mode; for instance v107 peptide occupies most of the receptor-binding surface on VEGF (Figure 36) and has been the subject of extensive studies aside from determination of the complex structure. In fact, thanks to this large body of information, v107 has been the inspiration in our designing a dipeptide library and could well serve as binding positive control. Dr. Alberto Adeba from the Servei de Síntesi de Pèptids in the UB-SSCT kindly synthesized peptide v107 using standard solid phase peptide synthesis and an oxidation protocol also optimized by Fairbother and collaborators,[7] later this was titrated into a ^{15}N uniformly labeled VEGF sample and monitored by heteronuclear NMR experiments.(see Figure 35)
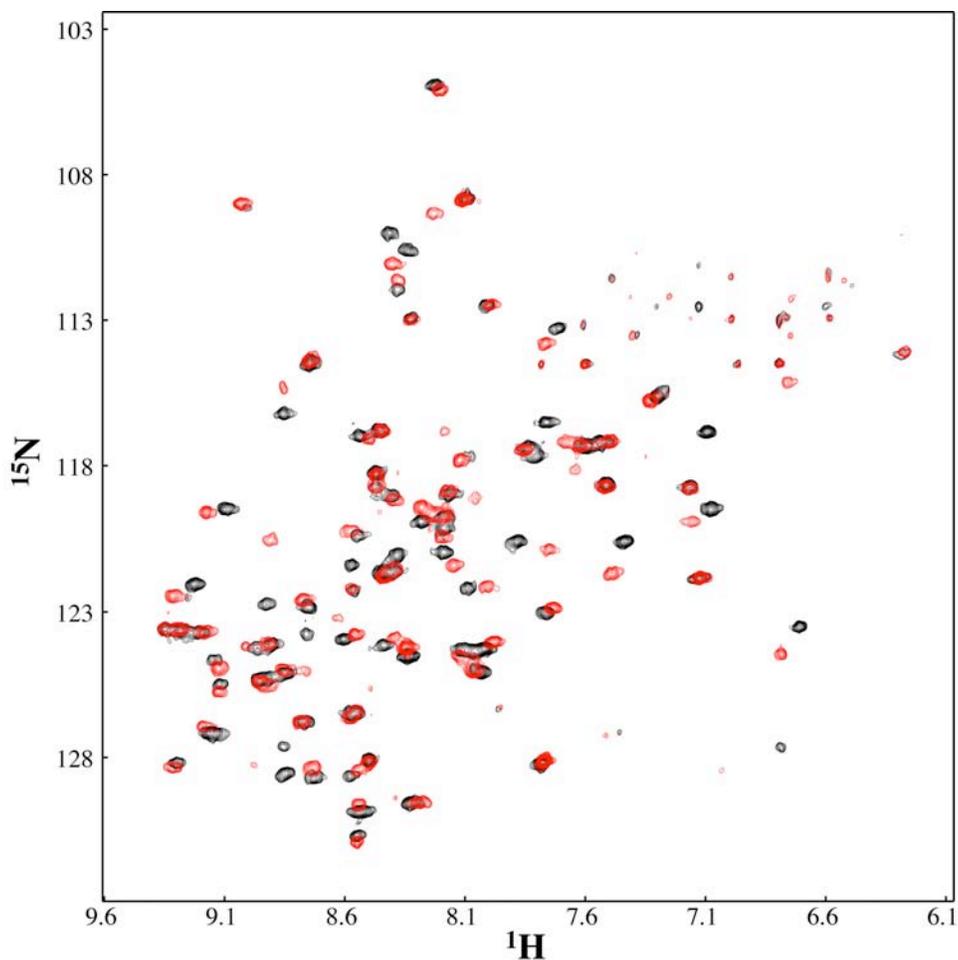
Figure 35 ^{15}N-^{1}H HSQC for uniformly labeled VEGF. Black) 250 µM VEGF sample in 50 mM NaCl and pH 7 25 mM phosphate buffer. Red) Experiment, after addition of 2 equivalents of v107 peptide.

As we see in Figure 35, dramatic changes readily take place upon increasing amounts of v107 peptide; the effects are those typical of a slow exchange ligand –as corresponds to a ligand with $K_D=0.2$µM– consisting on the appearance of a number of extra cross-peaks throughout the spectra. Free and complexed VEGF spectra coexist along the titration, until saturation is reached and all binding sites are occupied by v107 peptide; at which point free protein correlations disappear completely. The extent of changes associated to v107 addition are remarkable; in fact, more than 50% of correlations suffer noticeable changes, a clear evidence of a large rearrangement occurring due to peptide binding. This is not surprising given the conformational divergences and B-factors already observed in VEGF's crystal cell [5], where receptor-binding epitope displays various accessible conformers in the crystalline state. Most likely the presence of v107 deters this conformational sampling by stabilizing the bound state, similar to what we observed for BIR3/Smac N-terminus complex in Chapter 1.

2.2.1   SELECTIVE LABELING

So far we have produced unlabeled or uniformly labeled VEGF$_{11-109}$ and reproduce its binding properties. Indeed, this will be very helpful to test fragments using ligand based NMR methodologies; however, screening of lead-like compounds –such as those present in typical HTS chemical libraries or plant extracts- may not be so easy. Particularly for complex extracts, ligand based NMR experiments may not be able to detect strong binders within them and even if they could, mixtures might be too complex, too difficult to deconvolute or too prone to interferences among mixture components. On the other hand, receptor based experiments, especially those using uniformly labeled protein, may be too expensive for a HTS strategy; moreover, seeing the dramatic effects of v107 binding, CSP structural interpretation may be misleading as a consequence of strong induced fit effects.

In the previous chapter we have dealt with similar issues and demonstrated that selective introduction of NMR stable isotopes in the target protein may alleviate some of the former concerns. In particular, when cleverly introduced into amino acid side chains, one can obtain utterly simple spectra, retain most of the information and at the same time greatly reduce protein-related costs. Moreover, since side chains are less prone to suffer from those long-range CSP effects, they should be more reliable than backbone uniform labeling when extracting structural information.
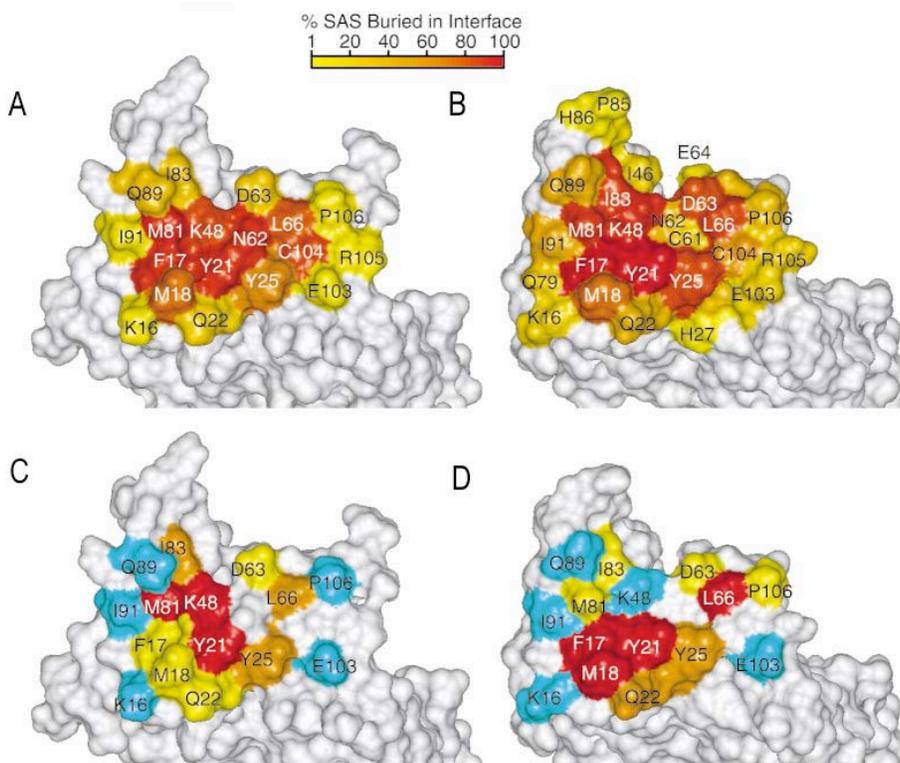

Figure 36 Extracted from Pan et al. (2002)[11]. Diagrams A and B depict VEGF surface burial upon interaction with v107 and Flt1 receptor respectively, residues are colored according to their solvent accessible surface (SAS). Diagrams C and D alanine scanning result: mutated residues and relative IC$_{50}$ values as follows: red, >30-fold increase; orange, 10 to 30-fold increase; yellow, 3 to 10-tenfold increase; cyan,> 3 fold increase.

Dissection of v107-VEGF and Flt1-VEGF interactions has been previously reported thanks to a combination of complex structure resolution and extensive alanine scanning. In both cases VEGF buried surfaces are roughly the same and overlap to a large extent; [11] although due to the size of Flt1 receptor, this spans a larger hormone patch. Despite both complexes being mainly stabilized through hydrophobic contacts and the obvious similarities in the buried surfaces, alanine scanning study suggests that the precise energetics underlying complex formation are slightly different. (Figure 36) In the case of v107 peptide-VEGF complex substitution of M81, K48 and Y21 to alanine has dramatic effects on the complex stability; while for Flt1-VEGF complex the cluster of relevant residues is slightly tilted towards the N-terminus α-helix: F17, M18, and Y21. Nonetheless, there seems to be consensus as to several tyrosines and methionines playing a crucial role in VEGF binding epitope; as a result incorporation of NMR probes on either type of amino acid is very appealing.

Altogether VEGF interactions analysis seems to reinforce the concept of hot spot amino acids, and the preference of certain residues to perform such a function. In particular Trp, Tyr and Arg have already been highlighted previously; indeed, in our case two tyrosines –Y21 and Y25, both at the N-terminus α-helix– sit at the center of the hydrophobic patch buried upon complexation. Both actively participate in the process as their mutation to alanine has a great effect on the complex K_D, especially Y21. With regard to tryptophan – the most frequent hot spot residue [13] – this does not appear on VEGF binding epitope and consequently will prevent our using the previous labeling methodology; nevertheless, this fact does not demean Trp importance in any way since a close inspection of either VEGF complex reveals the presence of crucial tryptophan residues both on v107 peptide and Flt1 receptor.

2.2.1.1 *METHIONINE LABELING*

Two out of the five methionines in VEGF seem to play relevant roles in complex formation and stabilization; they are M81 and M18 and become deeply buried upon interaction with either v107 or Flt1 receptor. Mutation of either methionine has deleterious consequences for both complexes, although M81 seems to have a larger effect on v107-VEGF and M18 to Flt1-VEGF complex. (Figure 37)These evidences are even more striking when considering hot spot amino acid composition according to Bogan and collaborators;[13] in this list methionine is found in less than 3% of hot spots and when its occurrence in proteins is accounted for, we observe one of the lowest fold increases for any amino acids.

Beyond their undeniable privileged location in VEGF, from a practical perspective methionines are also attractive for labeling purposes. First, introduction of stable isotopes into its thioether methyl group is chemically uncomplicated; this can be illustrated by *methyl-^{13}C*-methionine low cost (220 \$/g) compared to uniformly labeled methionine (350 \$/g) or to 6-^{13}C-lysine (2440 \$/g). Furthermore, ^{13}C methyl labeling is very appealing due to its high sensitivity and advantageous relaxation properties. Regarding methionine incorporation into proteins, the issue is easily solved using auxotrophic E.coli strains; such strains have been commercially available for some years as a result of its widespread use in protein crystallography to solve the phase problem. [24]
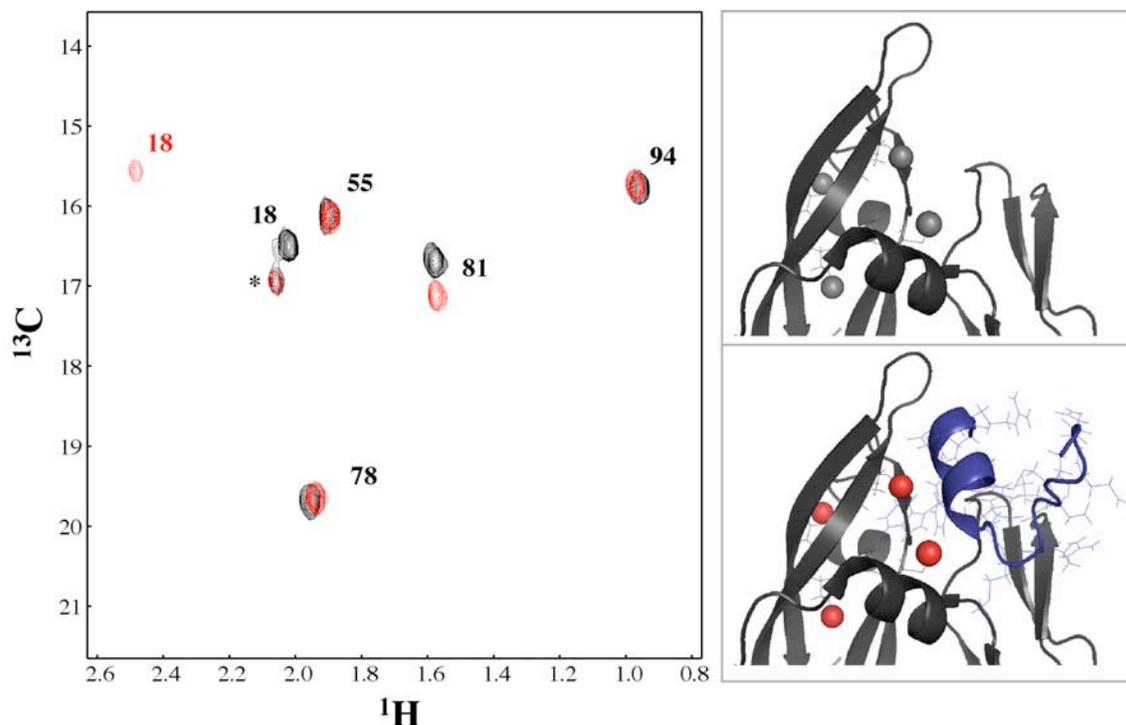
Figure 37 ^{13}C-^{1}H HSQC experiments on *methyl-^{13}C*-Met VEGF$_{11-109}$ and assignment. Black) 250 μM *methyl-^{13}C*-Met VEGF in 50 mM NaCl, pH 7 25 mM phosphate buffer. (Red) after addition of 2 v107 peptide equivalents. Free and v107-complexed VEGF ribbon structures are depicted; methionine methyls are represented as spheres black or red, free or complexed respectively. (*) Crosspeak corresponds to the starting codon methionine from residual VEGF with uncleaved his tag.

Auxothrophic E.coli strain B834 cells were transformed with plasmids p6XHisVEGF$_{11-109}$ and pMS421. VEGF$_{11-109}$ was produced as in the previous sections, except that labeled methionine was supplemented into the minimal media. Despite the use of auxotrophic cells protein production was not greatly affected and we obtained similar yields as for uniform ^{15}N labeling.

Once over expressed and purified, ^{13}C-^{1}H HSQC experiments were acquired on the selectively labeled protein. (Figure 37) Protein spectrum displays 6 correlations compared to the expected 5; the weaker extra peak at (2.1 ppm ^{1}H, 17 ppm ^{13}C) can be explained by the presence of residual protein with uncleaved his-tag; and indeed, in protein preparations where digestion was thorough the signal is not present. Addition of v107 and its effects were also monitored with heteronuclear experiments; this titration produces stepwise appearance of 2 new signals as well as waning of two signals from the original set. Spectrum assignment using published data [11] reveals that the most affected signals by peptide addition correspond to methyl groups in methionines 18 and 81; this is not surprising since these are in direct contact with the v017 peptide. The largest change occurs for M18, probably as a result of important contacts with Trp13 in v107 peptide and the effect of its aromatic ring currents. On the other hand, changes to M81's methyl chemical shift are mainly restricted to ^{13}C chemical shift. Finally, binding effects on the rest of methionines are practically negligible.

Considering the extent of changes in ^{15}N-^{1}H HSQC and those observed in ^{13}C-^{1}H, it seems methionine selective labeling is less prone to suffer from CSP overinterpretation as a result from long-range conformational changes. This is clearly illustrated by M94 or M78, barely affected by the peptide on the ^{13}C-^{1}H experiment –as they are not in contact with v107–, whereas their amides suffer important changes on the ^{15}N-^{1}H HSQC due to the large conformational rearrangement occurring concomitant to the binding.

So far the results obtained using *methyl-^{13}C*-methionine selective incorporation are very encouraging. The scheme is efficient, cost effective and achieves great spectrum simplicity, and although we have only dealt with 2D experiments, its monodimensional versions (^{3}C-filtered-decoupled-^{1}H NMR- will likely be equally sensitive to ligand binding and retain nearly as much of the structural information.

Certainly such 1D experiment (Figure 38) has enough resolution for the 5 methionines in VEGF; moreover, thanks to the methyl high sensitivity we can reduce the protein amount considerably. In fact, 30 μM samples still provide an excellent signal to noise ratio with very short acquisition times: 7 minutes; making an excellent HTS screening assay.
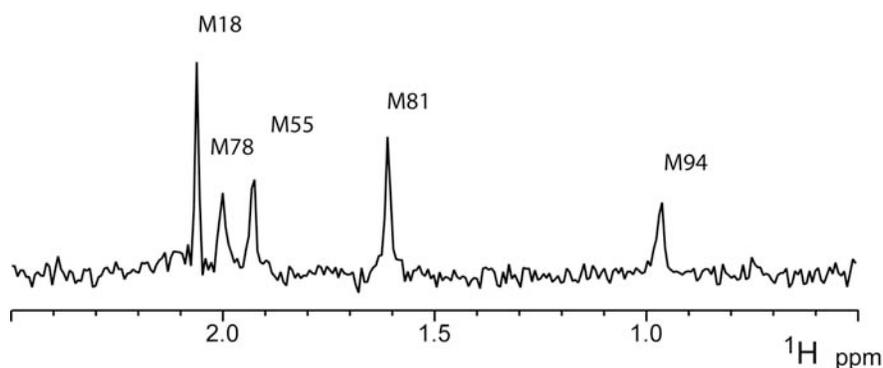

Figure 38 ^{13}C-filtered-decoupled-^{1}H NMR experiment for a *methyl-^{13}C*-Met VEGF 30 μM sample in 50 mM NaCl, pH 7 25 mM phosphate buffer. Assignment for ^{13}C filtered methyl resonances is shown.

2.3   NMR SCREENING

Once the details for protein production have been sorted out, during the following sections we shall cover the screening process for the earlier compounds. The assay strategy, as already mentioned, will depend on the nature of each chemical library; briefly ligand based experiments such as STD and WaterLOGSY will be used to test our fragment libraries –dipeptides- while our methionine selective labeling scheme will be used on a CSP approach to assay the collection of plant water extracts.

2.3.1   PEPTIDE LIBRARIES

10 μM VEGF samples were prepared in $D_2O:H_2O$ (9:1) 25 mM phosphate buffer at pH 7. Mixtures of four dipeptides, at a 500 μM each, were added from the d_6-DMSO stocks; subsequently, STD and WaterLOGSY experiments were performed for each sample until all compounds were finally assayed. Pulse sequences for both ligand detected experiments were implemented and optimized on a test sample composed of Lactate Dehydrogenase (20 μM), and two weak ligands: NAD+ (1 mM).(Figure 39) This sample has been widely studied by London and collaborators and reported to produce important cross-saturation and transferred-nOe effects; for this reason constitutes a reliable test bench for this set of experiments. [25]

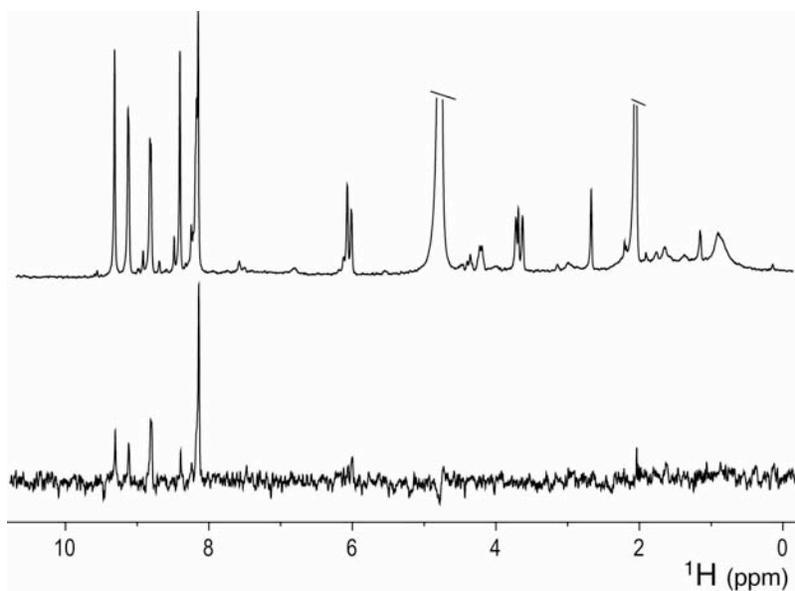Figure 39 ^{1}H (top) and STD (bottom) experiments on a 20 μM Lactate Dehydrogenase sample in the presence of 1 mM NAD+. STD experiment was acquired with 1024 scans, using 50ms Gaussian shaped pulses during 2s for saturation at 0 ppm and a 20 ms spinlock.

2.3.1.1 *DIPEPTIDE LIBRARIES: V107 INSPIRED & BROAD LIBRARY*

Cross-saturation experiments were applied to the screening of our dipeptide libraries and despite our exploration of various experimental parameters: protein and ligand concentration, number of scans or saturation frequency, we did not obtain any satisfactory result. (Figure 40) Apparently none of the tested dipeptides seems to bind VEGF in the STD or WaterLOGSY affinity range.
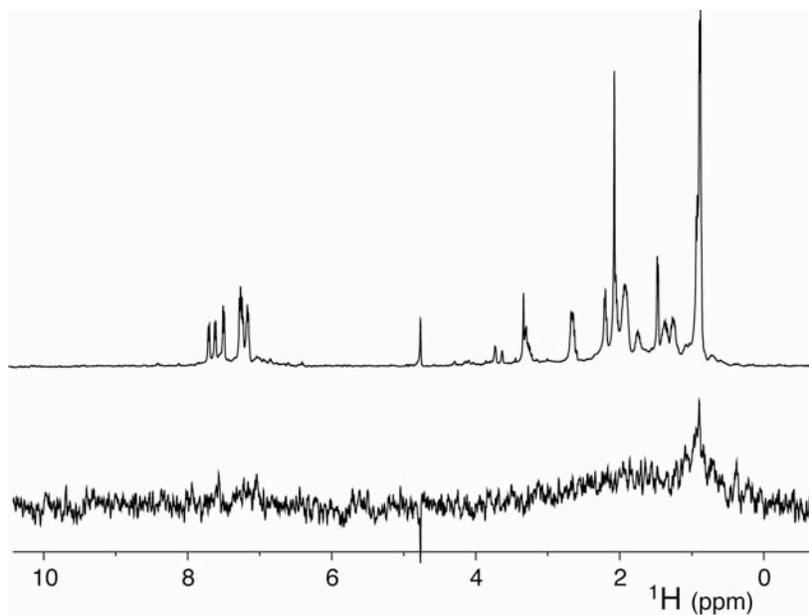
Figure 40 [11]H (top) and STD (bottom) experiments on a 20 μM VEGF sample in the presence of 4 D-dipeptides (EW, WE, AI, Ac-ID) 1mM each. STD experiments was acquired with 1024 scans, using 50ms Gaussian shaped pulses during 2s for saturation at 0 ppm and a 20 ms spinlock.

Given the chemical diversity of the tested libraries, these results may be due to dipeptide intrinsic limitations. Some authors suggest in this regard that molecular rigidity is crucial for the screening of weak binders, since entropic cost often outweighs the enthalpic component for small flexible molecules, hampering binding. Others could attribute our failure to the relative lack of chemical functions in dipeptides rendering molecules unable to interact efficiently with VEGF. Or one could summarize it by saying that peptide oligomers easily fall outside the "fragment-like" class and thus are not suited for this type of screening.

In this sense, to further pursue a fragment-based strategy, we should reconsider our library and probably build this using preferred scaffolds or even some sort of peptidomimetics with fragment-like properties. In the latter case we could propose $\alpha-\beta$ dehydropeptide oligomers; these are interesting peptide analogs for they have restricted flexibility, they are resistant to proteases and are accessible through solid phase supported chemistry. [26]
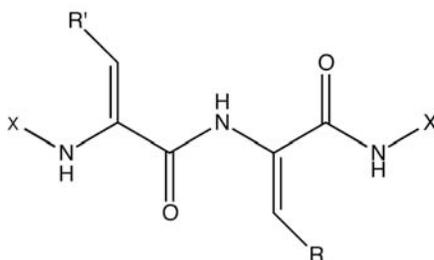

Figure 41 $\alpha-\beta$ Dehydropeptide oligomers as restricted analogs of dipeptides. X corresponds to termini modifications such as amide, acetyl or carboxylic groups.

2.3.2   ORGANIC SOLVENTS

Despite the initial distress due to the lack of positive binders in our dipeptide libraries, a somewhat serendipitous observation was made during the screening process. Measurements of VEGF ^{15}N-^{1}H HSQCs in the presence of high amounts of various dipeptides revealed perturbations, although minor, to various amides. Follow up control experiments, with the sole addition of DMSO, demonstrated that these shifts were induced by the vehicle organic solvent rather than by the presence of dipeptides.

In fact, this type of observations is not novel and several authors have reported residual affinities of organic solvents towards protein surfaces. As an example, Dalvit and coworkers utilized several NMR experiments, originally devoted to study water solvation effects on proteins, to characterize and optimize the binding of various DMSO related compounds to FKPB12. [27]

Several other authors have described similar observations with a number of proteins; interestingly in all cases, organic molecules do not seem to bind randomly onto the protein surfaces but tend to accommodate into biologically relevant spots. [27-29] This pattern probably exposes certain intrinsic properties of protein hot spots such as its dual hydrophobic/hydrophilic character. From a practical point of view, these observations could serve as a diagnostic tool for protein hot spots. For instance, one could use organic solvent HSQC titrations to pin down relevant sites on protein surfaces in advance.

Similar evidences have been provided by X-Ray crystallography, oftentimes the addition of co-solvents during crystallization results in the presence of such additives in the final structure, precisely on hot spots. And very much like in the case of NMR, several authors have proposed schemes to take advantage of such observations to diagnose relevant protein sites; that is by solving the structures for different protein crystals previously soaked in organic solvents. [30, 31] Whichever the biophysical tool, identification of organic solvent binding sites may have interesting applications in the drug discovery field; especially in identification of protein hot spots to guide *in silico* screening, or even to inspire the design of novel ligands.

With this background in mind we decided to carry out a comparative study performing different ^{15}N-^{1}H HSQC titrations with various water-miscible solvents; this served a three-fold purpose. First a practical purpose; as we have seen compound libraries are typically prepared as concentrated stocks in DMSO, and while this is logistically very convenient, this could induce undesired competition artifacts. It is interesting then to find an organic solvent, which has minor or no effects over VEGF, to be used as vehicle for the chemical compounds in our libraries. Secondly, these experiments may allow us to identify unreported binding sites on VEGF, which could be very useful in the design of VEGF inhibitors. Finally, as seen in other works, these organic solvents could serve as "early" leads and characterization of their binding properties can be a good inspiration for the design of novel chemical libraries.

^{15}N-^{1}H HSQCs experiments were performed for VEGF in the presence of increasing amounts of dimethylsulfoxide, acetonitrile, isopropanol, and dimethylformamide. (Figure 42) Following amide identification on free VEGF and in the presence of organic solvent, an average δHN was used to evaluate the magnitude of chemical sift perturbations throughout the sequence; this averaged value takes into account the relative sizes of ^{1}H and ^{15}N spectroscopic window. (Equation 5)

The analysis of the various titrations reveals similar behaviors for the organic solvents tested; VEGF seems to have residual affinity towards all these molecules. The observed perturbations are however not widespread on VEGF surface nor randomly localized; instead they seem to follow a pattern which suggests

that this is not a consequence of a denaturalization process but a binding event on relevant surface spots. (Figure 43)

$$\Delta\delta HNav = \sqrt{\frac{\left(\delta H_f - \delta H_s\right)^2 + \left(\dfrac{\delta N_f - \delta N_s}{5}\right)^2}{2}}$$

Equation 5 Chemical shift perturbation for average δHN; calculated using amide ^{1}H and ^{15}N chemical shifts in the free (f) experiment and in the presence of organic solvents (s).[32]
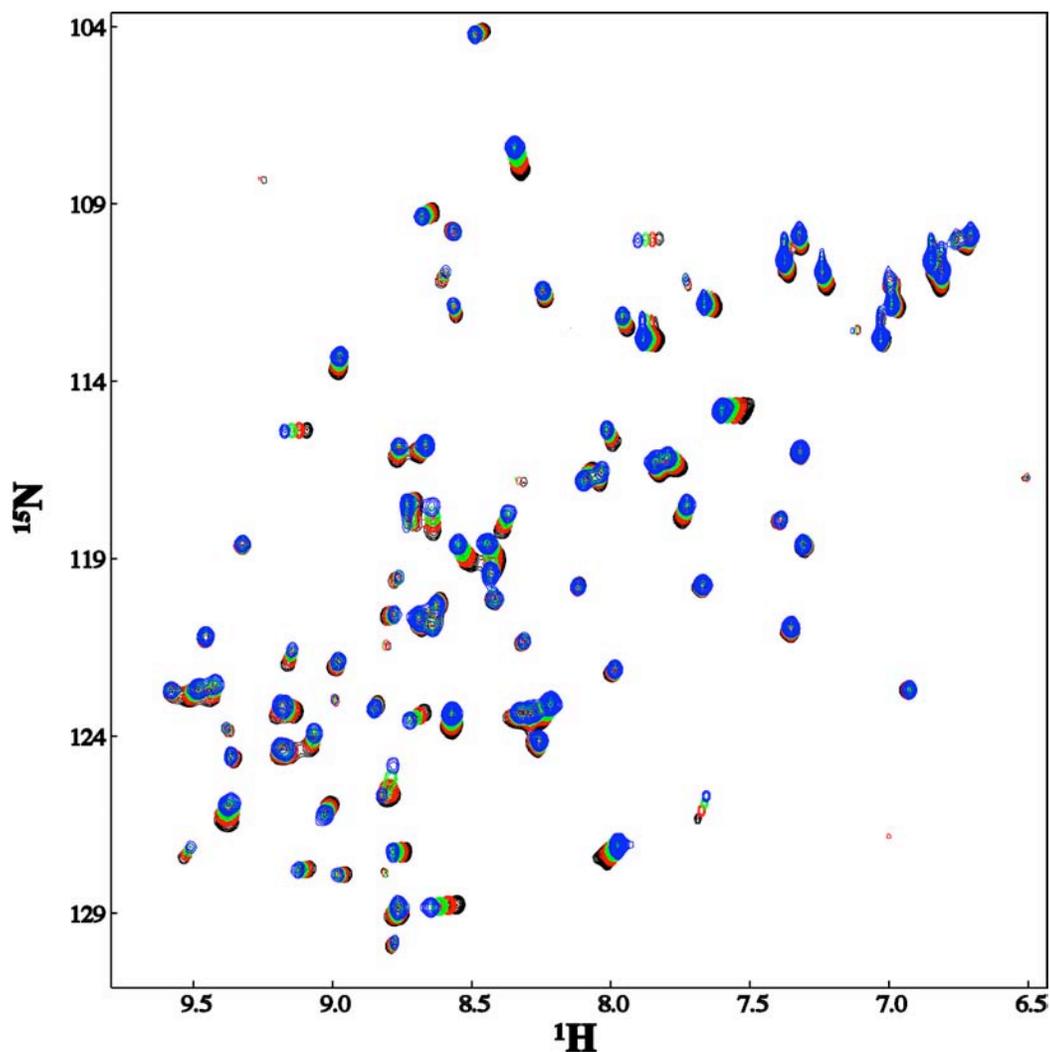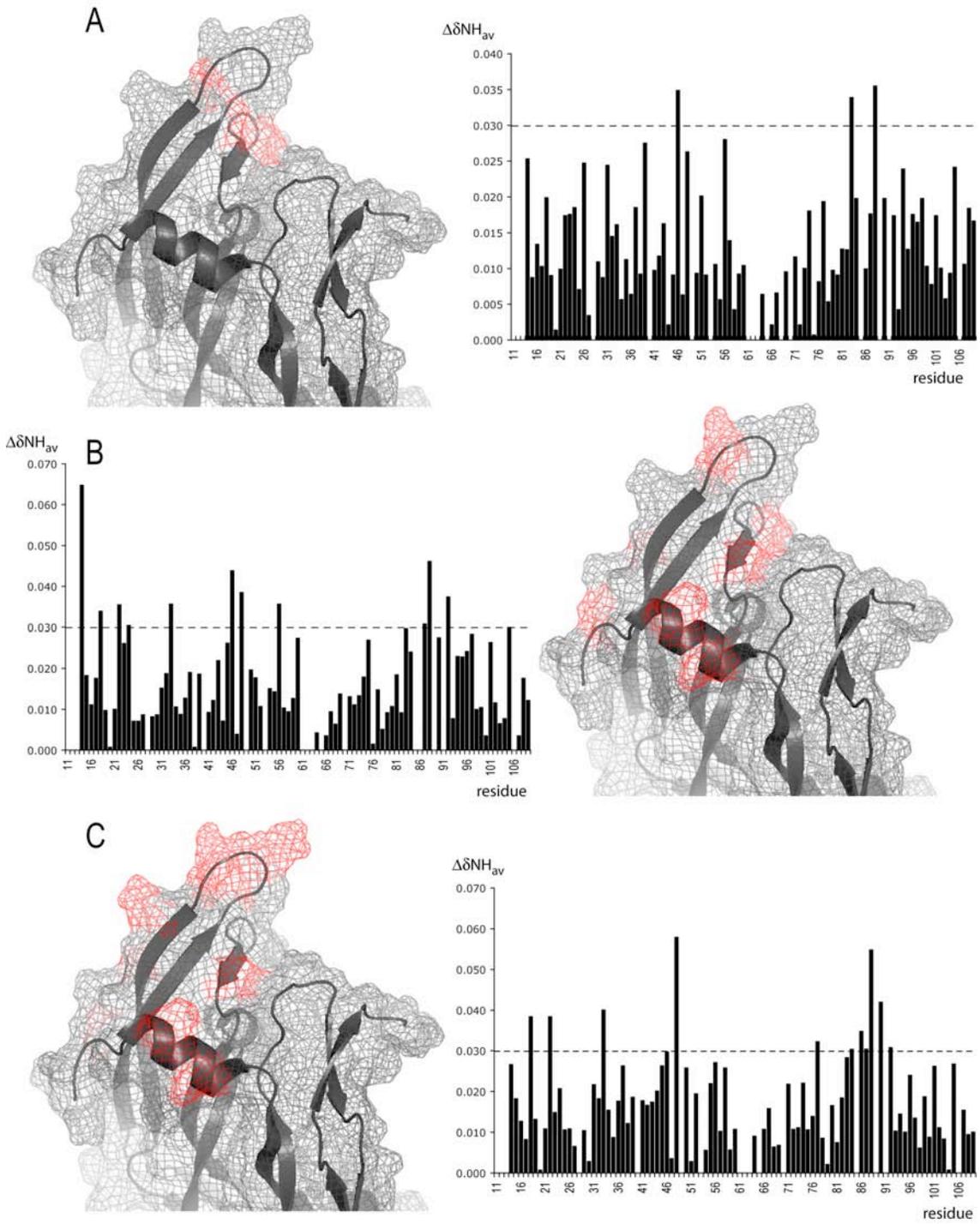

Figure 42 ^{15}N-^{1}H HSQC experiments for a 250 μM VEGF sample upon addition of increasing amounts of dioxane: free (black), 5% v/v (red), 10% v/v (green), 15% v/v (blue).
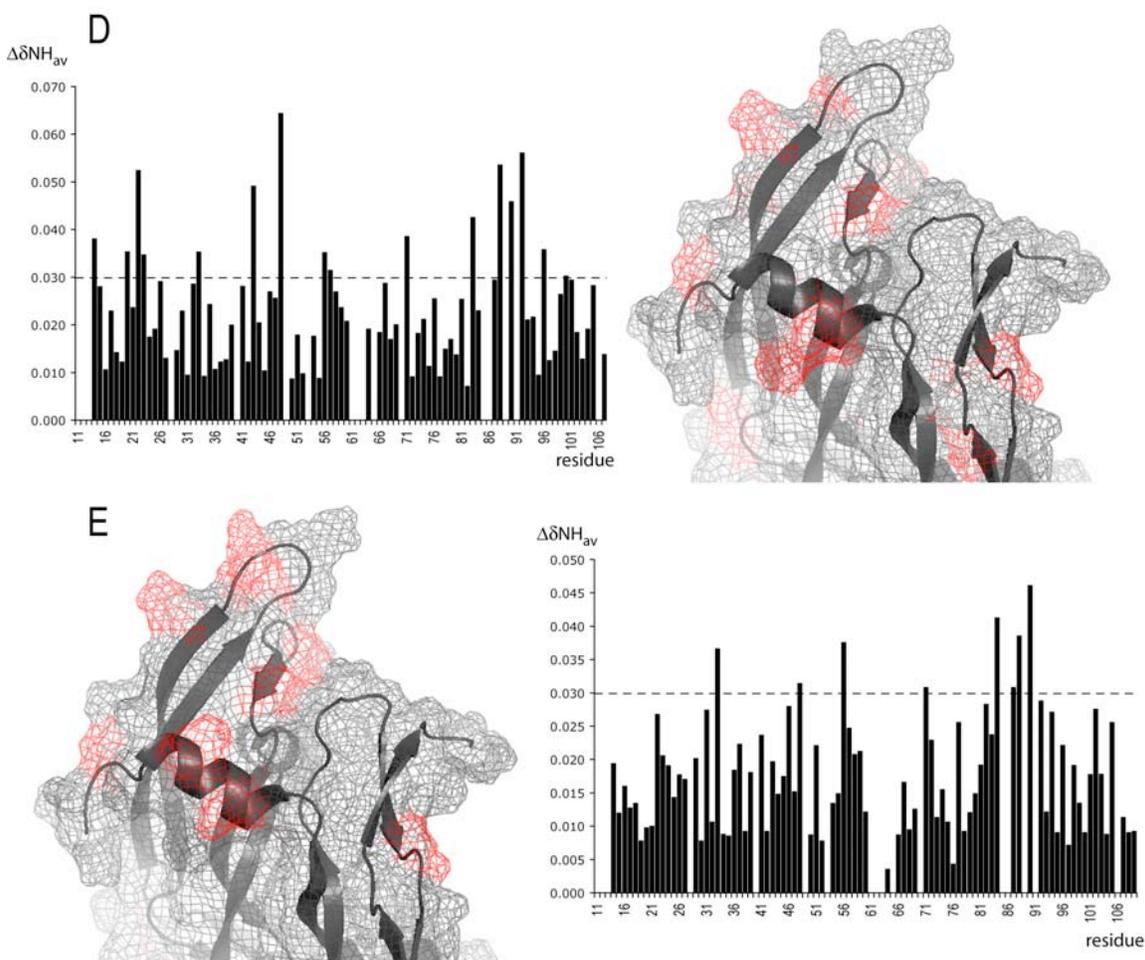
Figure 43 Chemical shift perturbation analysis for VEGF in the presence of various organic solvents at 5% v/v concentration: (A) DMSO, (B) DMF, (C) dioxane, (D) iPrOH and (E) ACN. ^{15}N-^{1}H HSQC experiments were carried out with a 250 μM protein. CSP for protein amides was calculated using Equation 5 and plotted vs. sequence. Amides suffering changes above the $\Delta\delta_{HN}$=0.03 ppm threshold have been colored in red on VEGF receptor binding interface.

In general dioxane and isopropanol seem to induce the larger changes on VEGF amides, followed by dimethylformamide and acetonitrile; finally, according to the data, DMSO is the worst binder of the series and at 5% v/v concentration it is barely able to induce changes above the 0.03 ppm threshold on protein amides. (Figure 43) This answers one of the questions previously raised; DMSO would seem to be the least of evils, if one were to choose among various solvents to prepare concentrated stocks for a chemical library.

Interestingly, chemical perturbation analysis both at a sequence level and on the protein structure plot reveals certain consensus on the clusters of disturbed amino acids. The first and most common cluster throughout the different titrations is placed in the vicinity of residues Q89 and H90, at a reportedly important β-strand.[11] Amino acids close to I46 and K48 also tend to be affected by most organic solvents, and although sequence-wise they fall far away from Q89, it is clear on the 3D structure that they belong to the same β-sheet structural motif. Consequentially, it is possible that the perturbations observed for this two

clusters are coupled, and that these effects might be transmitted at a relative long distance through the β-sheet.

A third remarkable cluster is located around the N-terminus α-helix, particularly around M18 and Q22. It is also quite close to I46 in the 3D structure, but in this case the two clusters do not belong to the same secondary structure motif and thus a long-range effect seems less likely. Moreover, and despite the reported flexibility of VEGF receptor binding epitope, it would be surprising to see such long-range effects induced by small organic solvents. For this reason, the broad effects on the protein spectrum may be better explained by the coexistence of multiple solvent binding sites, which according to our chemical shift perturbation data could be placed around M18, I46 and Q89 respectively.

Interestingly the presence of these adjacent sub-pockets correlates fairly well with the structural dissection of VEGF peptide ligands. [11] On the one hand phage-display peptide v107, obtains most of its interaction energy from contacts with K48, Y21 and M81 residues. While binding of Flt1 receptor occurs manly through contacts established to a sub-site located at the N-terminus helix: F17, M18 and Y21 amino acids. Finally another class of phage display peptides -i.e. v108- has a similar binding mode as Abastin (monoclonal antibody), mainly involving a set backbone-backbone contacts with the exposed 6th β strand. this would correspond to our third sub-site around Q89, but also includes some hydrophobic interactions with some residues in the vicinity K48.
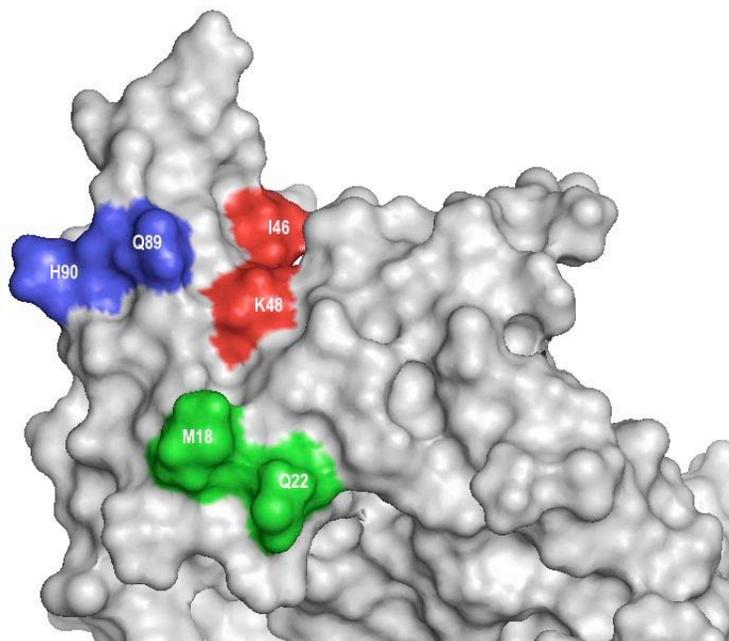

Figure 44 VEGF receptor binding epitope with our proposed organic solvent binding subsites colored in green, red and blue.

All in all, this study is consistent with the prior knowledge on VEGF structure and function. In this sense solvent titrations have been able to identify a set of amino acids critical for the binding to Flt-1 receptor and to other peptide ligands. Also, although no novel site has been identified, this technology has a great potential for the preliminary identification of attractive protein surface patches. Likewise, we anticipate

interesting uses to assess protein "druggability" [33, 34] which should help to determine whether a given target is suitable for HTS; for those proteins displaying affinity towards small organic molecules will likely be susceptible to the binding of bigger molecules.

2.3.3   TRADITIONAL CHINESE MEDICINE PLANT EXTRACTS

As already outlined, after our previous screening attempts with small, diverse and relatively simple chemical libraries, we will now pursue a radically different yet complementary strategy; assaying large complex libraries, in the form of traditional Chinese medicine plant water extracts.

These mixtures are chemically very rich, both in diversity and absolute numbers, and will presumably contain good VEGF binders. Good binders however, are not always detected by means of ligand-based NMR experiments, moreover extract complexity and heterogeneity would probably render this type of experiments nearly useless when the need for deconvolution arises. Thus, a cost effective receptor-based NMR experiment, such as the one described in prior sections, will be very helpful to screen this kind of libraries; since they provide a wide affinity detection range, retain useful structural information and have interesting throughput properties.

2.3.3.1 *Preliminary considerations on DMSO binding*

As a result of how libraries have been constructed, organic solvent will unavoidably accompany each addition of compound or extract to a sample. Again, similar to what we have observed for ^{15}N-^{1}H VEGF's spectra, we can anticipate solvent binding effects on methionines ε-CH$_3$; in fact these spin systems may even be more sensitive to solvent interactions than main-chain amides. And although solvent binding is very weak – in the molar affinity range- this does occurs on VEGF's binding epitope and probably has implications concerning the lower affinity threshold of CSP experiments; for this reason we shall evaluate the effect of DMSO on *methyl*-^{13}C-Met VEGF spectra.

A series of ^{13}C-^{1}H HSQC experiments and ^{13}C-filtered-^{13}C-decoupled ^{1}H NMR experiments were performed for samples containing *methyl*-^{13}C-Met VEGF and increasing amounts of DMSO. (Figure 45) This study was carried out under two different sample preparations in order to evaluate two typical experimental conditions: a high protein (250 μM) concentration as in reported ^{15}N-^{1}H HSQC titrations, and a low concentration (30 μM) similar to the used in the screening 1D experiment versions.

As expected, changes do occur upon solvent addition and are mostly focused on M18, M94 and M81 side chains; and although perturbation of these residues agrees with our knowledge on VEGF binding epitope, it is most interesting to note several discrepancies with respect to DMSO effects on ^{15}N-^{1}H HSQC protein spectra. M18 ε-CH$_3$ for instance, is particularly affected by DMSO while its amide is not; this may suggest direct participation of this side chain in the interaction with DMSO and underscores the importance of spin probe locations to monitor interaction events. In this sense, main chain probes, although very useful thanks their uniform distribution, may lead to binding site overestimation especially considering that most interactions are established through side-chains. (Figure 45)
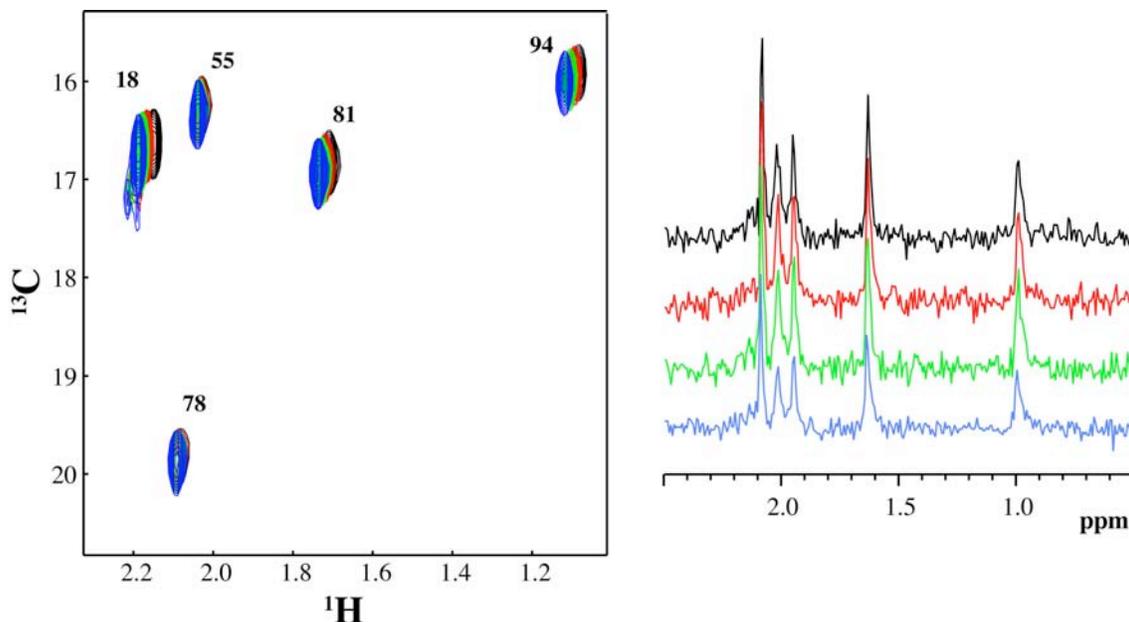
Figure 45 DMSO effects on *methyl-13C*-Met VEGF. (Left) DMSO Titration performed on a 250μM VEGF sample monitored with successive ^{13}C-^{1}H HSQC experiments: (black) free protein,(red) 10% v/v DMSO, (green) 15%, (blue) 20%. (Right) DMSO Titration performed on a 30uM VEGF sample monitored with successive ^{13}C-filtered-^{13}C-decoupled ^{1}H experiments: (black) free protein, (red) 1% v/v DMSO, (green) 2%v/v, (blue) 3% v/v.

DMSO effect on methionine 94 is much more puzzling, and although perturbation on its amide cannot be observed either, its changes on the *methyl-13C*-Met are undeniable. According to the NMR structure M94 sits in the vicinity of VEGF binding site; however, its side chain is not oriented towards v107. Moreover it does not seem particularly exposed to the solvent and should not be prone to participate in any binding. As a result we may probably attribute this perturbation effect to a long-range conformational effect linked to DMSO binding; it is indeed a caution call suggesting that *methyl-13C*-Met selective labeling is not completely devoid of such artifacts.

In addition to the differential analysis of DMSO effects on ^{15}N-^{1}H and ^{13}C-^{1}H HSQC experiments, these experiments corroborate that solvent interaction is indeed very weak, and that it should be safe to perform experiments below 10% (v/v) with 250 μM VEGF samples and below 3% (v/v) with 30 μM protein. At these concentrations its effects on methionines ε-CH_3 are practically imperceptible.

2.3.3.2   *SCREENING OF TRADITIONAL CHINESE MEDICINE PLANTS USING METHYL-13C-MET VEGF*

Once the library preparation logistics were through, we proceeded with the extract screening as follows: diluted VEGF samples (30 μM) were prepared in 9:1 D_2O:H_2O phosphate buffer, and a ^{13}C-filtered-^{13}C-decoupled ^{1}H experiment was measured for the free protein. A second experiment was recorded upon addition of a mixture of three different plant extracts (3μL each to a 500μl sample). The differences between the two experiments were used to diagnose the presence of VEGF binders in the extract mixture. Despite

the low protein concentration, methionine's ε-CH₃ high sensitivity, the total experimental set up for each sample does not surpass 20 minutes; this allowed us to assay 43 plant water extracts in a very short time.
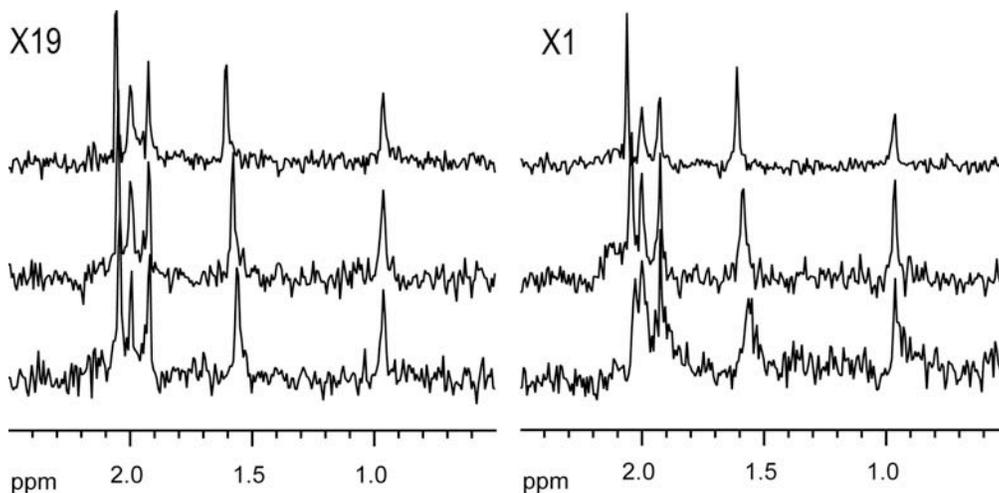

Figure 46 ^{13}C-filtered-^{13}C-decoupled ^{1}H experiments for 30μM *methyl-^{13}C*-Met VEGF samples with increasing amounts of *Radix scutellariae* (X19) and *Medulla Junci* (X1) extracts. (Top) free protein.(middle) plus 3μl extract d₆-DMSO stock. (Bottom) plus 6μl extract stock.

Unfortunately, using the above methodology, we were unable to detect the presence any tight binder within our extract library and not a single experiment displayed the dramatic perturbations that one would expect for such molecules –namely resonance broadening, splitting or large shit perturbations. A close inspection of the date however, did reveal that a two extract mixtures (X1-X3 and X16-X19) were indeed able to induce minor shifts to methionine methyls.

In order to validate the previous observations subsequent extract additions were performed, and indeed a dose response effect was encountered in both cases. Such effects diagnose the presence of some binder in the mixture and are not in any case false positives linked to solvent binding; such artifacts can be ruled out thanks to the previous control experiments, but also due to the direction of chemical shift perturbations: in both cases, as opposed to DMSO titrations, M81 and M18 were shifted towards higher fields.

Next, we proceeded to deconvolute the positive mixtures and individual extracts were titrated against VEGF; after which, extracts X1 and X19 were identified as responsible for the perturbations in each mixture. (Figure 46) In both cases the observed shifts are typical for a ligand in the fast exchange –a weak binder- and despite this suggest a lower affinity than we initially expected, interestingly perturbations mainly occur to M18 and M81. These correspond to methionines located in VEGF receptor binding interface, also those most affected upon v107 peptide binding, suggesting the active compounds within X1 and X19 extracts could potentially inhibit VEGF/VEGFR interaction.

2.3.3.3 ACTIVE COMPOUND ISOLATION AND CHARACTERIZATION

Usually the problem with libraries from natural sources is not having positive screening results, but rather identifying which compound among the hundreds in the mixture is responsible for the response. Plenty of examples in the literature rely on cell activity-based purification schemes to fractionate extracts and isolate

bioactive chemicals,[35] but often the process ends in a *cul-de-sac* when active compounds are found in so very little amounts that their characterization cannot be accomplished.
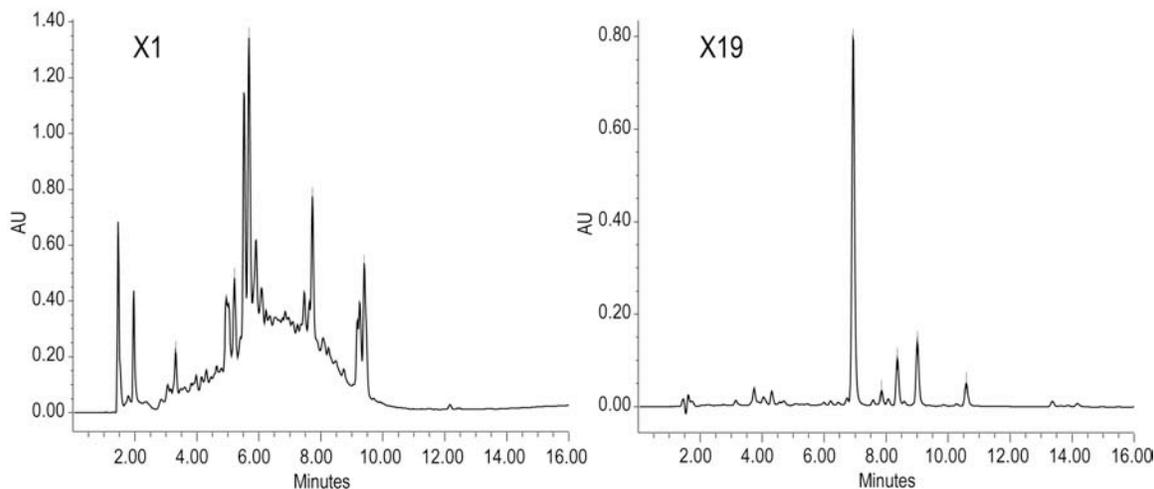

Figure 47 Chromatographic profiles for *Medulla Junci* (X1) and *Radix scutellariae* (X19) water extracts. HPLC analyses were carried out using 15 minute H_2O/ACN gradients (from 0 to 100% ACN).

Our situation is slightly different and although we will still have to isolate the active component, the effects observed on ^{13}C-decoupled-^{13}Cfiltered ^1H experiments indicate that, albeit being a weak binder, the active compound will likely be present in the extract at very high concentrations. Thus, given the circumstances, the advantage of identifying a weak binder is that it will likely be easier to purify.

In a first instance, in order to assess extract complexity, we decided to carry out a standard HPLC chromatographic characterization of the positive X19 and X1. The result was particularly striking for X19, which judging from the profile at 220 nm, (Figure 47) contains few well-resolved compounds and one of its components accounts for nearly 67% of the extract. On the contrary, extract from *Medulla Junci* (X1) is a much more complex mixture and presents many overlapped peaks spread throughout a wide elution range and without the presence of main component. Besides its greater complexity, the availability of *Medulla Junci* (X1) extract was rather limited (few milligrams) compared to *Radix scutellariae*, for which grams of extracted material were available. For these reasons we decided to pursue X19 purification and isolation of its VEGF binding component.

Preparative reverse phase HPLC was used to fractionate X19; in particular its major component was isolated and its mass established to be 447 (m/z) using ESI-MS. Later, we were able to confirm (Figure 48) with the same set of ^{13}C-filtered-^{13}C-decoupled ^1H titration experiments that this 446 Da molecule induced similar effects on VEGF spectra as the whole extract.
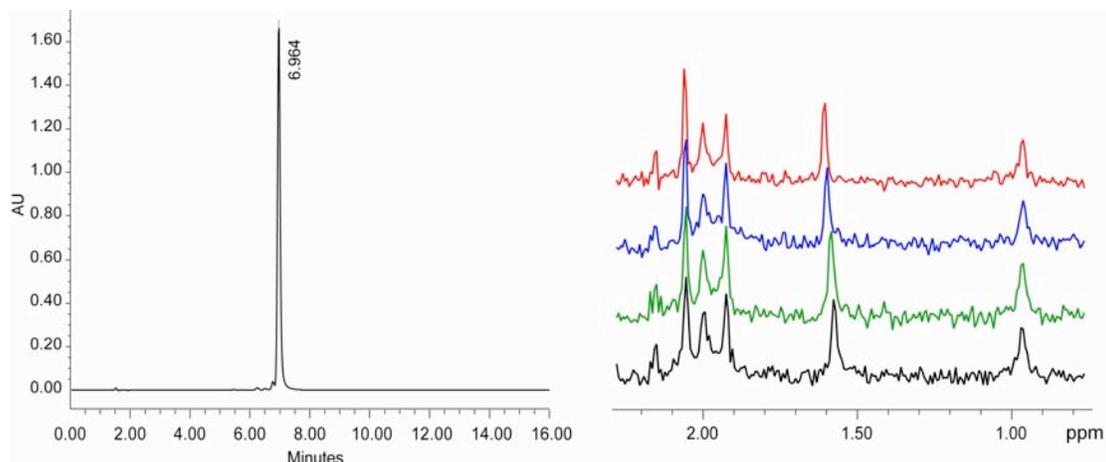
Figure 48 (Left) HPLC profile for X19 purified fractions containing extract's main component. HPLC purifications and analyses were carried out using 15:85-45:55 (H_2O:ACN) gradients. (Right) titration experiments with X19 main component using ^{13}C-filtered-^{13}C-decoupled ^{1}H experiments.

Next, we attempted characterization of the isolated molecule using standard NMR methodologies. 2D ^{1}H homocorrelation (COSY, NOESY) and ^{13}C-^{1}H heterocorrelation experiments (HSQC, HMBC) were used to assign and identify resonances from the isolated molecule. (Figure 49) NMR spectroscopy data suggests the presence of aromatic and sugar moieties; and, together with MS data, allowed us to identify the active compound as Baicalin.

Baicalein and its glucuronic derivative Baicalin are major constituents in *Radix scutellariae* –the source plant for X19 extract- and belong to the type of compounds known as flavonoids. These are probably the most abundant family of phytochemicals present in our daily diet; not only this, but flavonoid-rich foods are currently regarded as helpful in the prevention of several conditions: heart diseases or chronic inflammation, etc. [36, 37]

Besides its utilization in CNS affections, for which the plant was initially selected, *Radix scutellariae* has been widely used in folkloric Asian medicines and is present in many herbal preparations to treat tumors, hepatitis and leukemia. Such traditional knowledge has attracted much interest and consequently these two flavonoids have been inspected for their antitumoral effects; in this sense, both baicalin and baicalein have been associated with various properties: cytotoxicity, pro-apoptotic or antiangiogenic activities among others. [38-40] However, there is no actual consensus on their mechanism of action to produce such diverse effects. Several authors propose they induce oxidative stress on tumoral cells and activate the mitochondrial apoptosis pathway through caspase-3. [41] Others, highlight baicalin mediated inhibition of MMP-2 (Matrix Metallo Proteases) as the reason for their antiproliferative and antiangiogenic activities [42] and some authors even suggest these compounds may even have the ability to interfere in protein-protein interactions, particularly by disrupting the interaction between several chemokines and their receptors [43] to produce anti-inflammatory effects. There is however, to our knowledge, no evidence on their ability to affect VEGF-VEGFR interaction, and given the previous results this is a plausible mechanism, although surely not the only one considering their affinities.
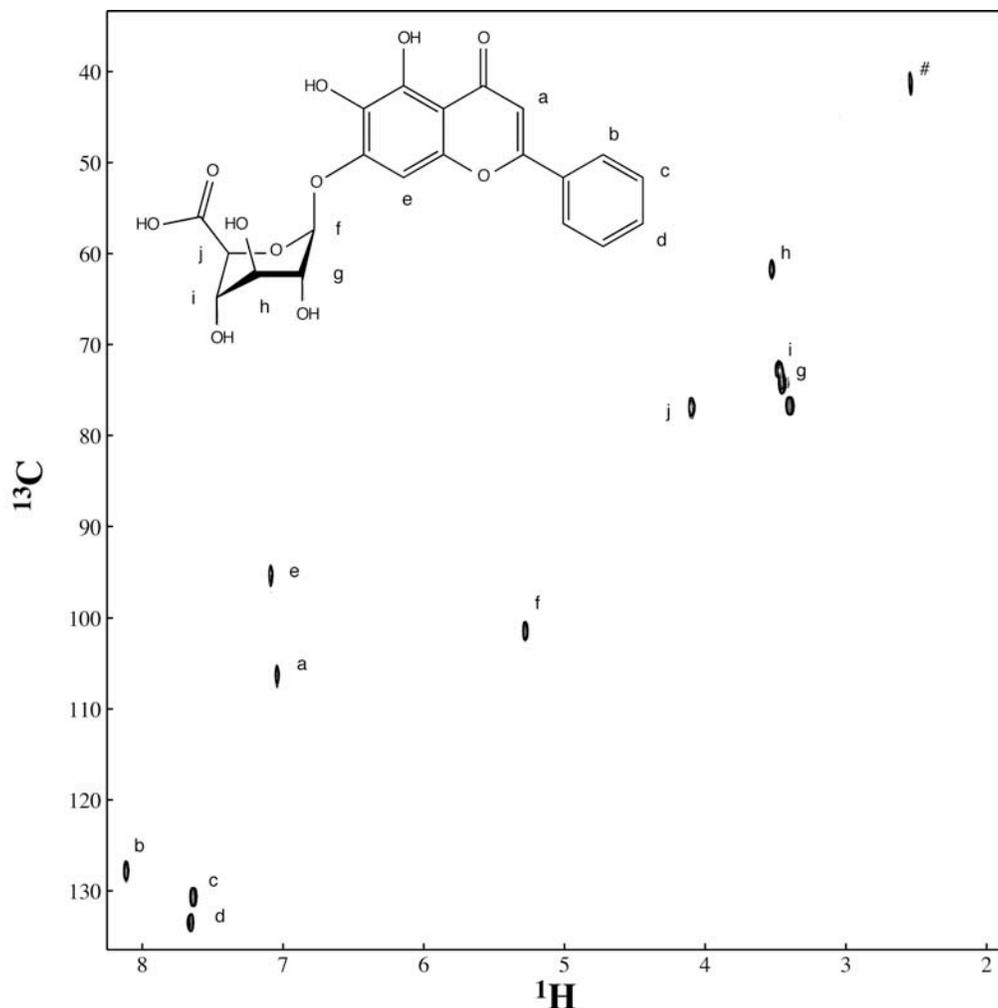
Figure 49 ^{13}C-^{1}H HSQC experiment for 446 Da molecule isolated from X19 extract and dissolved in d$_6$-DMSO. Correlations have been assigned on the molecule and (#) corresponds to solvent signal.

2.3.3.4 BAICALIN BINDING

Once isolated and characterized, both by mass spectrometry and NMR, we decided to carry out a series of NMR experiments to further understand how baicalin binds to VEGF. As a first approach we performed a chemical shift perturbation study using ^{15}N uniformly labeled protein, this experiments should provide information complementary to the methionine selective labeling spectra, along with a wider perspective on the baicalin binding epitope. In the latter, baicalin induced changes were circumscribed to methionine 81 and 18 side chains, but when the same analysis was carried out using ^{15}N-^{1}H experiments perturbations extend to a larger region of the protein (Figure 50 and Figure 55); the most remarkable being found around residues 85 through 96, together with isolated amides close to amino acids 46-48 and 18-22 at the N-terminus α-helix. We again observe differences between CSP analyses using uniform ^{15}N or ^{13}C methionine labeling; particularly for methionine 81, which contrary to the observations using *methyl*-^{13}C-Met labeling it is barely affected in ^{15}N-^{1}H experiments. As in previous DMSO titrations this is probably due to several reasons; first the limited perspective provided by *methyl*-^{13}C-Met labeling but also as a consequence of

isotope location and the propensity of backbone probes to overestimate binding surfaces due to long-range conformational effects.

The same CSP studies on *methyl-^{13}C*-Met and ^{15}N uniformly labeled samples allowed us to estimate baicalin-binding affinity. As described in the first chapter this was carried out by fitting the chemical shift changes to equation in Figure 50 and producing K_D values of 4.9 mM and 6.8 mM for ^{13}C-filtered-^{13}C-decoupled ^{1}H and ^{15}N-^{1}H HSQC titrations respectively. Which, despite being performed on two different samples and protein concentrations the estimate affinities are in the same range.

From a ligand perspective, STD experiments were performed for baicalin; these were successful as we see in Figure 51, and 10 μM VEGF clearly induces cross-saturation effects on several baicalin aromatic resonances, in contrast to control experiments without protein. Relative quantification of transferred saturation allows us to define the molecule's binding epitope in similar fashion as Meyer and collaborators have reported.[44, 45] Distribution of the cross-saturation effects seems to indicate that the most buried part of the molecule is the phenyl group while the least saturated and consequently the most exposed part is the glucuronic moiety, not completely unexpected given its hydrophilicity. Likewise, this would explain the weak saturation suffered by proton (e), vicinal to the glucuronic fragment.

We have already mentioned the difficulties linked to defining accurate binding sites for flexible proteins. Consequently determining whether two ligands share the same binding site is difficult relying solely on CSP, in these cases competition experiments are much more reliable and will provide the answer to the question: Do v107 and baicalin share the same binding site on VEGF?
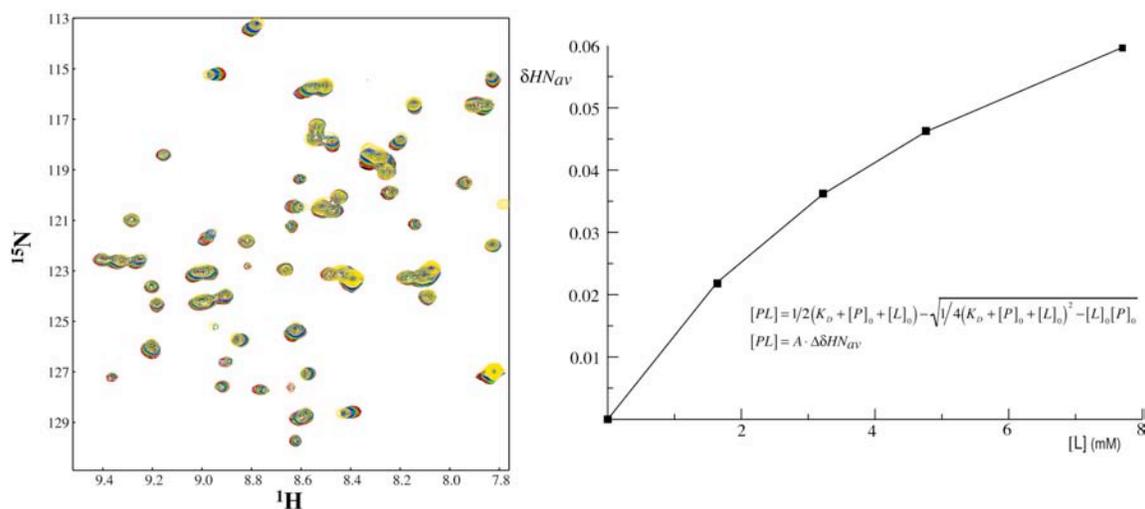

$$[PL] = 1/2 (K_D + [P]_0 + [L]_0) - \sqrt{1/4 (K_D + [P]_0 + [L]_0)^2 - [L]_0 [P]_0}$$
$$[PL] = A \cdot \Delta \delta HN_{av}$$

Figure 50 CSP study performed using ^{15}N-uniformly labeled VEGF. A series of ^{15}N-^{1}H HSQCs were recorded for a 375 μM protein sample with increasing amounts of baicalin. (black) free protein, (red) 1.69 mM, (green) 3.22 mM, (blue) 4.76 mM and (yellow) 7.7 mM baicalin. On the right, chemical shift differences for an affected amide, with respect to the free protein have been plotted versus baicalin concentration and non-linearly fitted to the depicted equation. K_D has been in this way determined to be 6.8mM with a 0.999 correlation factor and a 0.165 value for A.
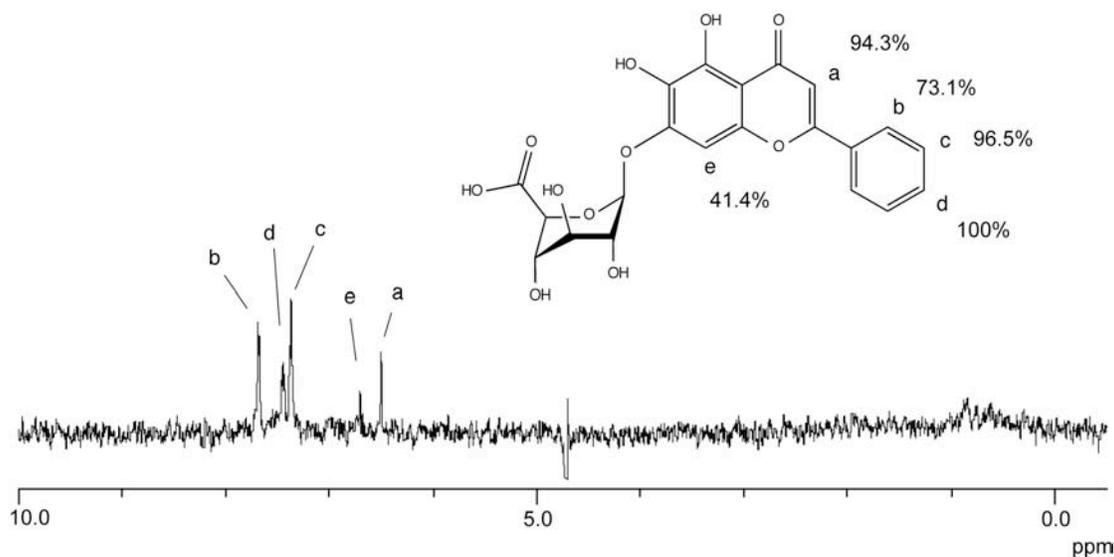
Figure 51 STD experiment for a 1 mM baicalin sample in the presence of 10 μM VEGF. Saturation was carried out at 0 ppm and 10ppm, for on and off-resonance scans respectively, with a train of 50 ms shaped (Gaussian) pulses during 2s prior to acquisition. Water suppression was achieved with a Watergate scheme and subtraction, between on and off-resonance experiments, was performed internally through phase cycling for 1024 scans. Cross-saturation effects on spectrum have been assigned to baicalin protons and normalized relative to the most saturated proton (d).
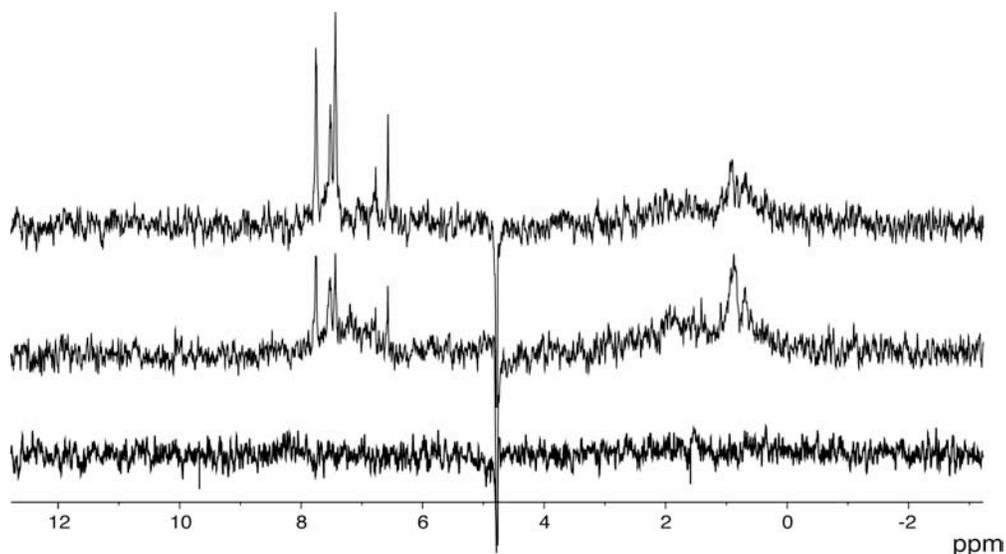

Figure 52 Competition STD experiments for baicalin. Increasing amounts of v107 peptide compete with baicalin binding and render a negative experiment. Experiments were performed as in Figure 51 for a 1 mM baicalin, 10 μM VEGF sample containing (top) 0 μM, (middle) 50 μM and (bottom) 100 μM v107 peptide.

After successful acquisition of an STD experiment using a baicalin/VEGF sample, increasing amounts of v107 peptide were added to the mixture and subsequent STD experiments recorded. These experiments

reveal a stepwise reduction in baicalin saturation effect, which becomes undetectable at a 100 μM peptide concentration. This is a clear evidence that both v107 peptide and baicalin compete for the same binding site, and given v107 affinity [7], it is not surprising to see that one 10th of the peptide concentration displaces baicalin from its binding site on VEGF.

In addition to the relevant structural information provided by the competition STD experiment, this experimental set up may be potentially very useful. Competition assays offer a way to elude the intrinsic affinity limitations of ligand based screening experiments by using a weak ligand as a binding probe, in this way it is possible to assay tight ligands that on their own do not suffer the cross-saturation effect. This may be particularly useful once a good lead molecule has been identified and one wishes to screen an array of chemical modifications in cost effective manner.

2.3.3.5 OTHER FLAVONOIDS

As a result from our screening methodology we have isolated and identified a VEGF binder and although baicalin is weaker than the ligands initially intended, it seems a reasonable design starting point. In the following section we will explore the chemical space around baicalin in order to get a better understanding on the features that allow baicalin to bind VEGF and eventually identify better binders. We will pursue this exploration using the previously NMR experiments, initially ^{13}C-filtered-^{13}C-decoupled ^{1}H spectra followed by STD and CSP experiments.
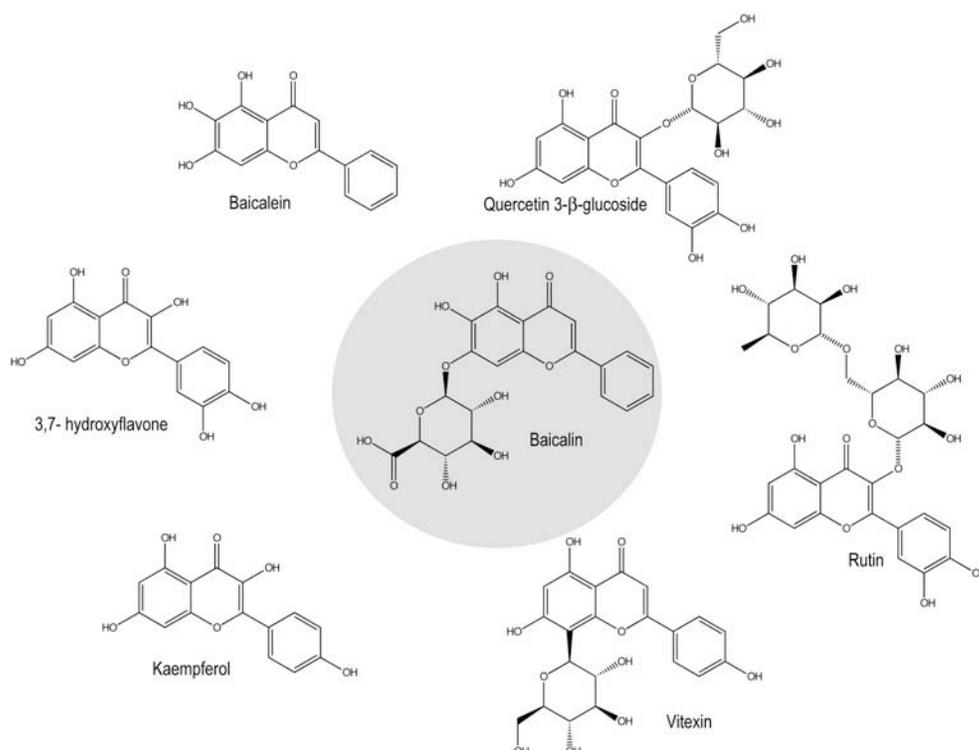

Figure 53 Other assayed flavones.

Our first selection of baicalin-related compounds consisted on a small set which conserved the flavone scaffold with minor variations on the number and position of several aromatic hydroxyls. (Figure 53) In this

way we intended to dissect the molecule into its most relevant parts and evaluate the importance of the glucuronic sugar moieties in the binding, specially considering that STD data suggests that the latter is largely exposed to the solvent. Unfortunately, in this first attempt all tested flavones (baicalein, 3,7-hydroxyflavone and kaempferol) turned out being not soluble enough to perform our assays. As a result, we decided to pick another small collection of compounds, but in this case to improve their solubility profile these were selected from the commercially available glycosylated flavones.

The second group of glycosylated flavones was composed of quercetin-3-β-glucoside, rutin hydrate and vitexin (Figure 53), they all share the common flavone scaffold but they mainly diverge in the location and type of their glycosidic moieties, along with minor differences on several hydroxyl groups. Molecules were assayed using our simple ^{13}C-filtered-^{13}C-decoupled ^{1}H methodology, with differing outcomes for each compound. Compound addition only perturbed methionine signals for rutin and quercetin-3-β-glucoside, albeit the latter did so to a larger extent. In fact, titration and K_D estimation suggests that quercetin-3-β-glucoside is even tighter a binder than baicalin with a 1.9 mM dissociation constant (Figure 54); while rutin on the other hand, is a weaker binder even when compared to baicalin. But may be more interesting than this modest affinity gain is the utter lack of binding observed for vitexin; even more so considering that the glycosidic moiety in baicalin sits very close to where it is in vitexin. Vitexin lack of binding may be in part due to the linkage between flavone and sugar in vitexin; which in contrast to the typical glycosidic bond it is established directly to the "anomeric" carbon, certainly reducing the mobility of the sugar and probably hindering flavone scaffold interactions.
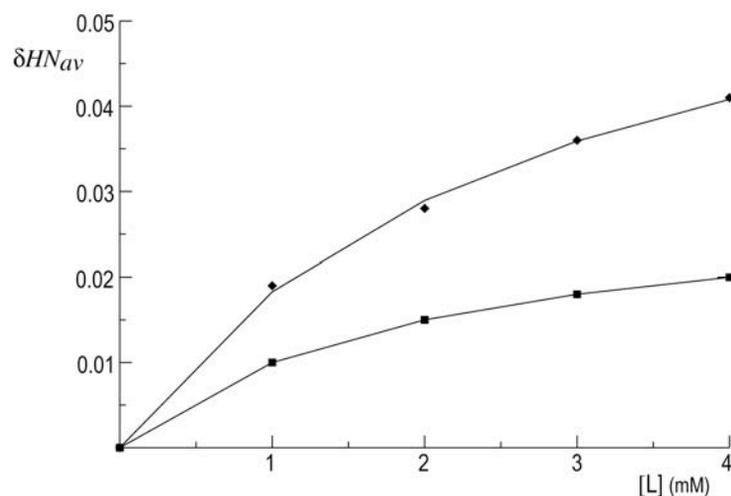

Figure 54 Quercetin-3-β-glucoside titration using ^{13}C-filtered-^{13}C-decoupled ^{1}H experiments and a 30 μM *methyl-^{13}C*-Met labeled VEGF sample. Chemical shift differences for thioether methyl groups in M81 (◆) and M18 (■) have been plotted versus compound concentration and non-linearly fitted. (See Figure 50) K_D values were 1.9 and 2.7 mM for M81 and M18 curves respectively.

^{15}N-^{1}H Chemical shift perturbations analysis for quercetine-3-β-glucoside shows three different clusters of amino acids particularly affected throughout the titration (Figure 55); these are centered around I35, F47 and G65 and are located in the vicinity M18 and M81. But despite such proximity, again, observations in

[13]C-filtered-^{13}C-decoupled ^{1}H experiment do not seem to fully agree with [15]N-^{1}H experiment since M18 and M81 amides are not especially perturbed.

With regard to STD, quercetine-3-β-glucoside experiments were far less successful than for baicalin; and although we were able to detect cross-saturation effect for a 10 μM VEGF and 1 mM compound sample, so were we in absence of protein. Such lack of a neat negative control is distressing and causes STD derived information to be unreliable. This type of behavior has been linked to compounds with poor solubility or aggregation issues [46, 47]; in fact, these notions were confirmed when we measured NOESY spectra for free quercetine-3-β-glucoside in water and the resulting experiment displayed negative nOes, as corresponds to a high MW compound. As a result, epitope mapping or competition STD experiments for quercetine-3-β-glucoside will not be reliable. However and despite the lack of a clean negative control, we did perform the competition STD experiment for quercetin-3-β-glucoside and v107 peptide; the results were even more puzzling when saturation effects on quercetin-3-β-glucoside diminished and eventually disappeared, with increasing amounts of peptide. After all, and with due caution, it may well be that at least part of the observed cross-saturation is not completely artifactual.

2.3.3.6 STRUCTURAL DISQUISITIONS

So far we have been successful at identifying two interesting phytochemicals with the ability to bind VEGF's receptor binding surface. Quercetin-3-β-glucoside and baicalin have a common flavone scaffold but mainly differ in the type and location of glycosidic moiety and to a lesser extent on the position of various hydroxyl groups. Moreover the available data suggests that in both molecules the flavone scaffold is largely responsible for the interaction, being particularly true for baicalin as demonstrated by STD binding epitope information. Thus, assuming that both molecules share a common binding mode we should in principle be able to perform a comparative analysis of their perturbations on the protein spectrum and use this information to determine location and relative orientation of these molecules on VEGF surface.

Differential analysis of quercetin-3-β-glucoside and baicalin effects on protein ^{15}N-^{1}H HSQCs shows several interesting trends. (Figure 55) A first observation is that we require twice as much baicalin to produce similar changes as quercetine's derivative; of course, this is consistent with the fact that the latter compound is roughly two times better binder than baicalin. Regarding the sequence dependent perturbations, it is interesting to note that both compounds affect a common cluster of amino acids centered around F47 together with some residues close to N-terminus α1-helix like M81 and R23. Beyond these similarities most of the changes produced by baicalin seem to concentrate around residues 86-93 located at 5th and 6th β-strands and their interconnecting loop. Conversely, quercetin-3-β-glucoside specific perturbations are primarily located on VEGF's opposed side, primarily around F36 and the loop connecting to the 2nd β-strand. Yet another set of quercetin-3-β-glucoside specifically affected amides corresponds to those around G65, at the loop connecting 3rd and 4th β-strands in the midst of the cysteine knot motif.
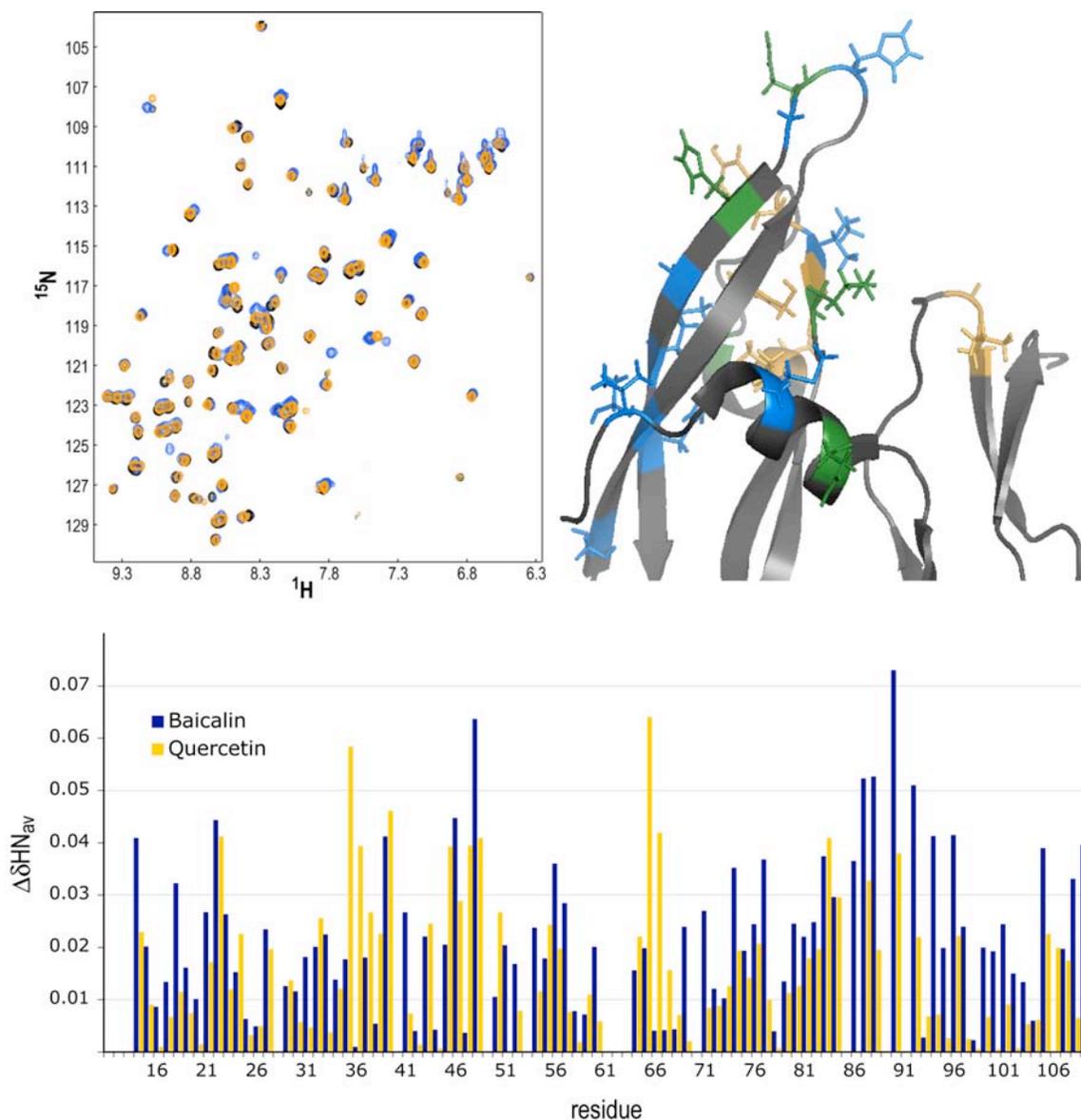
Figure 55 Differential chemical shift perturbation analysis for baicalin and quercetin-3-β-glucoside. (Top-left) ^{15}N-^{1}H HSQC experiments for 250 μM sample of ^{15}N-uniformly labeled VEGF: (black) free, (blue) after addition of 7.7 mM baicalin and (yellow) after addition of 4.4 mM quercetin-3-β-glucoside. (Bottom) Chemical shift perturbations induced on VEGF amides ($\Delta\delta HN_{av}$) by (blue) baicalin or (yellow) quercetin-3-β-glucoside plotted with respect to protein sequence. (Top-right) cartoon structure for VEGF receptor binding epitope, residues with the most affected δHN_{av} (>0.03 ppm) are represented as sticks and colored depending on the particular compound inducing the chemical shift changes: (blue) baicalin, (yellow) quercetin-3-β-glucoside and (green) both.
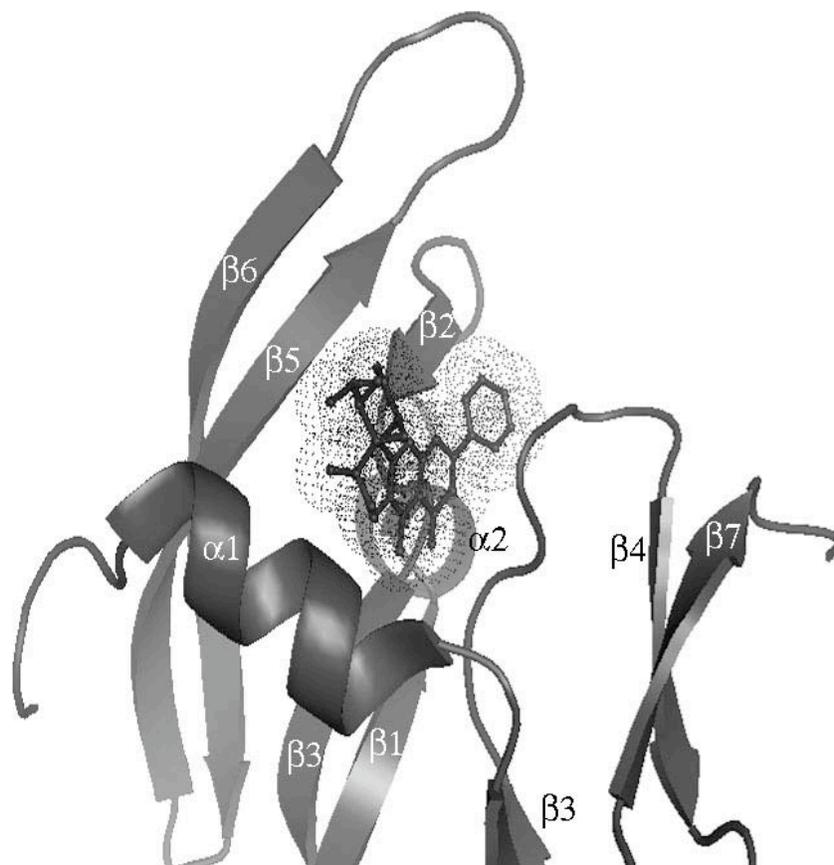
Figure 56 Model depicting Baicalin binding to VEGF. VEGF receptor binding surface is represented as cartoon and baicalin molecule as a ball and stick model.

With the information harvested so far we will dare to propose a model for the binding of flavonoids to VEGF. (Figure 56) Under the assumption that the most important difference between baicalin and quercetin-3-β-glucoside lays in the sugar position and that the flavone skeleton participates in both complexes with similar set of interactions, it is only reasonable to place their common scaffold on the region equally affected by the two compounds: around residue 47. F47 sits at the end of the 2nd β-strand (β2) right in the middle of the dimer interface; which thanks to its flexibility should be able to accommodate flavonoids. As to the orientation of baicalin's glucuronic moiety, differential effects on α1-helix, β5 and β6 indicate that this is probably accommodated between these secondary structure motifs. Quercetin-3-β-glucoside on the other, probably locates its sugar moiety towards the opposite face of the hormone inducing important perturbations to G65 loop and α2-helix in order to accommodate such a bulky group. Finally, the information extracted from baicalin STD epitope mapping (Figure 51) allows us to determine which side of the flavone scaffold is more deeply buried within VEGF; as depicted in Figure 56 the ortho proton to the glucuronic moiety seems the most exposed to the solvent, while the carbonyl function in the flavone scaffold will most likely be involved in some sort of hydrogen bond towards the interior of the protein.

2.3.4   FUTURE LIGANDS AND PROSPECTS

Overall our strategy has allowed us to identify a molecular scaffold that is able to bind VEGF receptor binding interface. Two differently glycosylated versions of this flavonoid skeleton have been identified with low millimolar dissociation constants and through a combination of NMR experiments we have proposed a plausible binding mode for this molecular family. The next natural step would be to further explore the chemical space around the previous hits in order to identify better ligands; for this, one could turn to the best of the two molecules - quercetin-3-β-glucoside- and devise a strategy to produce large a large number of analogs. However, from a pragmatic standpoint, baicalin is a much more appealing molecule thanks to its glucuronic moiety. This is particularly suitable for chemical modifications and through ester or amide chemistry it should be possible to efficiently explore a wide variety of modifications.

With this philosophy one can envisage oligopeptide baicalin derivatives where a peptide is attached to the glucuronic moiety; (Figure 57) or even use combinatorial chemistry approaches, together with the use of competition STD experiments, to optimize such peptide fragments. Conversely, one could also use the proposed binding model or information from *in silico* docking to design tighter baicalin-peptide chimeras inspired on known peptide ligands.[7, 11] In this sense one of the families of peptides described by Fairbrother and collaborators seems extremely suitable for the purpose; v108 peptide targets a neighboring site on the exposed 6th β-strand of VEGF receptor binding epitope. Thus, it should be possible to use a fragment linking approach to design hybrid v108-baicalin molecules. (Figure 57)
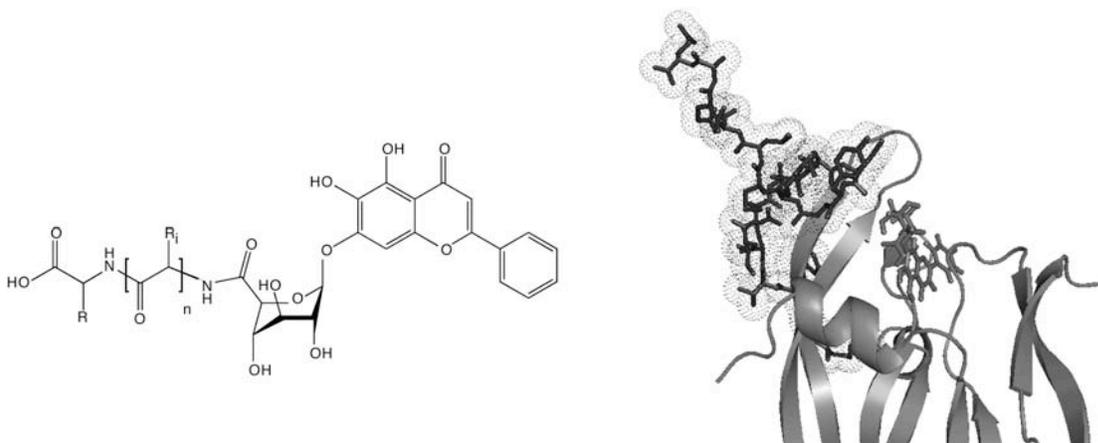

Figure 57 (Left) General formula for oligopeptide baicalin derivatives. (Right) Cartoon model for VEGF receptor binding epitope depicting binding of v108 peptide (sticks and dots) [7, 11] and binding of baicalin (sticks) on an adjacent site.

2.3.4.1   *PHYTOCHEMICALS AND PLANT EXTRACTS WITH REPORTED ANTIANGIOGENIC ACTIVITY*

Given the relative success of the above exploration of plant extracts, a possible future direction could include screening of extracts with reported antiangiogenic properties, rather than screening of general libraries or even collections chosen with other therapeutic targets in mind. In this sense, several reviews report plants with such activities and in several instances their mode of action has been attributed to the inhibition of VEGF signaling pathway. [15] Several of these plants are frequently encountered in Asian dietetic plants, such as soy or tea, and could explain several epidemiologic studies, which suggest lower

incidence of cancer in populations with such dietary habits. These evidences have also spurred the development of dietary supplements with chemoprotective activities or nutraceuticals [48] and bestowed Chinese food therapy, another ancient folkloric practice, with an aura of credibility.

In this context we have carried out some preliminary assays on some of the most abundant components in green tea: (-)-epigallocatechin (EGC) and (-)-epigallocatechin gallate (EGCG). (Figure 58) They belong to the flavan family of compounds and although they lack the carbonyl moiety and the 2,3-unsaturated ketone, in a way they are related to flavones. These two catechins are particularly interesting for they are thought to be largely responsible for the beneficial effects of green tea. They have been found to inhibit new vessel formation by shutting down the VEGF signaling pathway although there is no clear consensus as to how this effect is achieved. Some evidences point towards an inhibitory effect on the Flt1 and KDR receptor kinase activity, while other authors suggest catechins are able to interfere in the binding between VEGF and its receptors. [49-52] With such bibliographic background we were encouraged to test both EGCG and EGC on VEGF with the same set of NMR experiments used the flavonoids above.

According to ^{13}C-filtered-^{13}C-decoupled ^{1}H experiments EGCG is the better binder of the two, inducing changes on methionines 18 and 81; and although we did not carry out a complete titration with EGCG, its affinity seems to be in the range of quercetin-3-β-glucoside. And much as it happens for the latter solubility issues did also arise; thus STD and tr-nOe data were unreliable and revealed some tendency to aggregation as well. As for ^{15}N-^{1}H CSP experiments, these were also feasible and as we see in Figure 59 several cross-peaks suffer significant broadening or shift upon EGCG addition. The pattern of affected amides is very similar to that of quercetin-3-β-glucoside, in fact the same effects on loop residues 65 and 66, together with 36 in α-helix, seem to indicate that probably catechin resembles flavone's binding mode. According to this EGCG would position gallate moiety in a similar manner as the glycosidic fragment in quercetin-3-β-glucoside, while the flavone carbonyl could be substituted in EGCG by the ester function
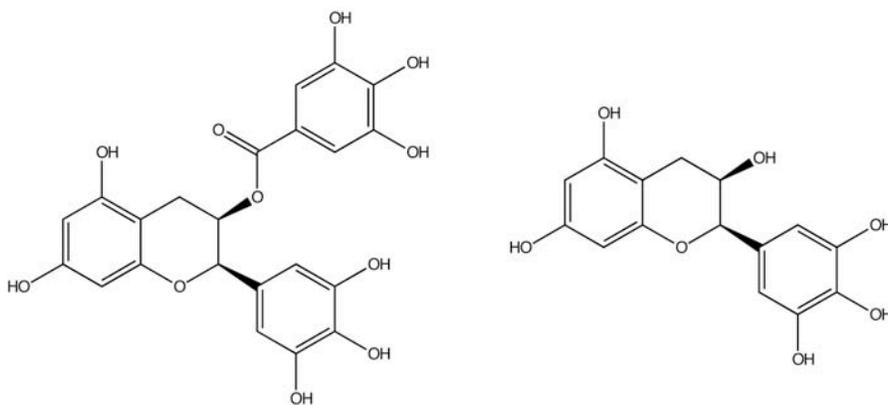

Figure 58 Structures for (-)Epigallocatechin (right) and (-)epigallocatechin gallate (left). They are abundant in green tea and have been attributed antiangiogenic properties.

Interestingly the lack of this ester group has dramatic effects on EGC affinity, which to the light of both ^{13}C-^{1}H and ^{15}N-^{1}H CSP experiments barely binds VEGF, as opposed to its gallate derivatives. Such evidences agree with the in vivo studies for these compounds in which EGCC is more efficient at inhibiting formation of new vasculature. [50]

Differences in the affinity of EGCG and EGC are even more intriguing when we consider another observation made throughout our NMR screening assays; this is the lability of the EGCG ester bond under our sample conditions. Apparently this ester bond hydrolyses under aqueous conditions at neutral pH and release of gallic acid occurs already several hours after sample preparation. This probably has implications on the bioavailability of active EGCG, since this reaction will likely take place in vivo, even more so considering that tea is administered orally. Interestingly, to the light of these observations we can suggest modifications to enhance compound stability and probably activity, namely substituting the ester by a more stable amide group or even replace the oxygen with a methylene.
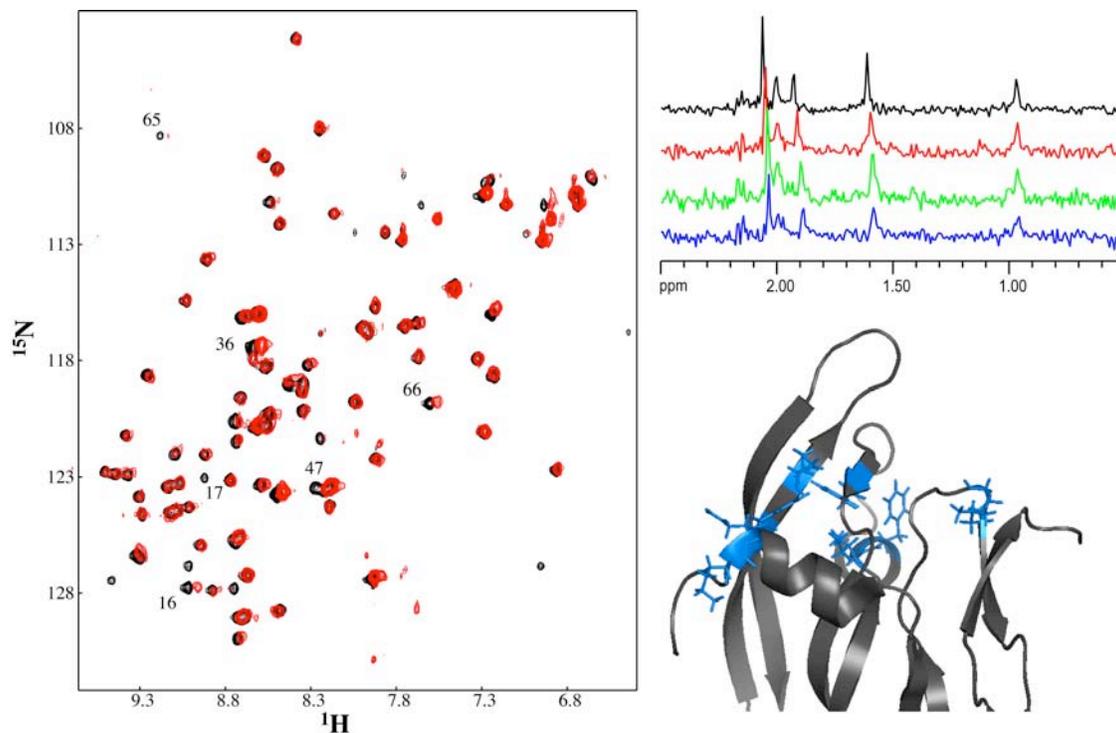

Figure 59 EGCG binding on VEGF. (Left) ^{15}N-^{1}H HSQC experiments for a 200 μM ^{15}N uniformly labeled VEGF: free protein (black) and after addition of 2 mM EGCG (red); assignment for the most remarkably affected amides is provided on the spectra. (Right-top) ^{13}C-decoupled-^{13}C-filtered ^{1}H experiments for a 30 mM *methyl-^{13}C*-Met VEGF sample and different concentrations of EGCG: 0 mM (black), 1 mM (red), 2 mM (green) and 3 mM (blue). (Right-bottom) The most affected residues in ^{15}N-^{1}H HSQC experiments upon EGCG addition have been highlighted (stick-blue) on the VEGF receptor binding epitope cartoon structure.

2.4 CROSS-SATURATION TROSY HSQC

Throughout the first and second chapters we have dealt with various protein complexes and their observation by NMR; besides their therapeutic relevance, these protein complexes seem to share an annoying feature for the NMR spectroscopist: protein plasticity. This observation seems to be the rule, rather than the exception, in most protein complexes and could be an example of the "principle of economy" so common in nature; thanks to plasticity, proteins may interact with multiple ligands and complex interaction networks can be built with a limited number of members. In fact, this is such a common feature among proteins that some authors suggest it is possible to identify protein hot spots only by diagnosing supple parts on protein surfaces. [53-55]

VEGF is not an exception and in fact a wide range of ligands can be accommodated on its receptor binding epitope thanks the mentioned flexibility. These are very different molecules and include peptides, phytochemicals and even the utterly small organic solvent molecules; and albeit their diverse size and nature, all these compounds seem to perturb large portions of VEGF's surface upon binding. Such floppiness and its concomitant promiscuity is in a way desirable, as it raises our success odds when screening for ligands; but on the other hand, this may difficult the collection of structural information, particularly for weak protein-ligand complexes and may even jeopardizes the outcome of structure based drug design.

In order to deal with floppiness-related inconveniencies one can choose from a list of NMR strategies and experiments developed through the years: selective labeling schemes, differential CSP approaches and a variety of nOe based experiments used to characterize protein-ligand complexes much more accurately. Nevertheless, probably one of the most interesting experiments is cross-saturation TROSY; [56] this method takes advantage of its ability to deal with very large systems and the distance dependant nature of saturation transfer to map the large protein-protein complex interfaces in a very accurate manner.

Cross-saturation TROSY has proven very useful for large protein complexes and tight ligands; but to our knowledge there is no report on its application to weak binders, such as those identified for VEFG during the previous sections. For such complexes, we can anticipate very low efficiencies for the cross-saturation phenomenon, nonetheless characterization of these interactions may not be impossible sin experiments as e-PHOGSY or nOe-pumping [57, 58] rely on similar magnetization transfer mechanisms and are widely used with weak binders.

In the current section we intend to set up the Cross-saturation TROSY experiment and explore its ability to deal with VEGF and its complexes with weak ligands. In a first phase we will produce an appropriately labeled protein batch and later the experiment will be implemented with the help of the v107peptide. Finally we will use the experiment to characterize VEGF complexes with flavonoids and/or organic solvents.

2.4.1   PROTEIN: ^{2}H-^{15}N-VEGF

The concept behind cross-saturation TROSY has already been introduced and in fact, in some regard consists on reversed version of a Saturation Transfer Difference experiment. Here, saturation is performed on the ligand resonances rather than the receptor (although for large macromolecular complexes it is difficult to tell ligand from receptor) and contrary to STD experiment, the saturation effect is observed on the "receptor". Such observation is performed thanks to HSQC experiments, a typical receptor-based technique.

The experiment proceeds through acquisition of two consecutive ^{15}N-^{1}H hetero-correlation spectra as depicted in Figure 60. The first serves as control while in the second experiment a saturation period precedes TROSY-HSQC pulse sequence. During this period, saturation is spread throughout the "ligand" and transferred to close by protons in the "receptor". The latter is finally observed by means of a hetero-correlation experiment and the transferred saturation on the protein-ligand interface appears as an intensity decrease for amides in this region. Subtraction of control and saturated HSQC experiments provides an easy way to pin down cross-saturated amides.
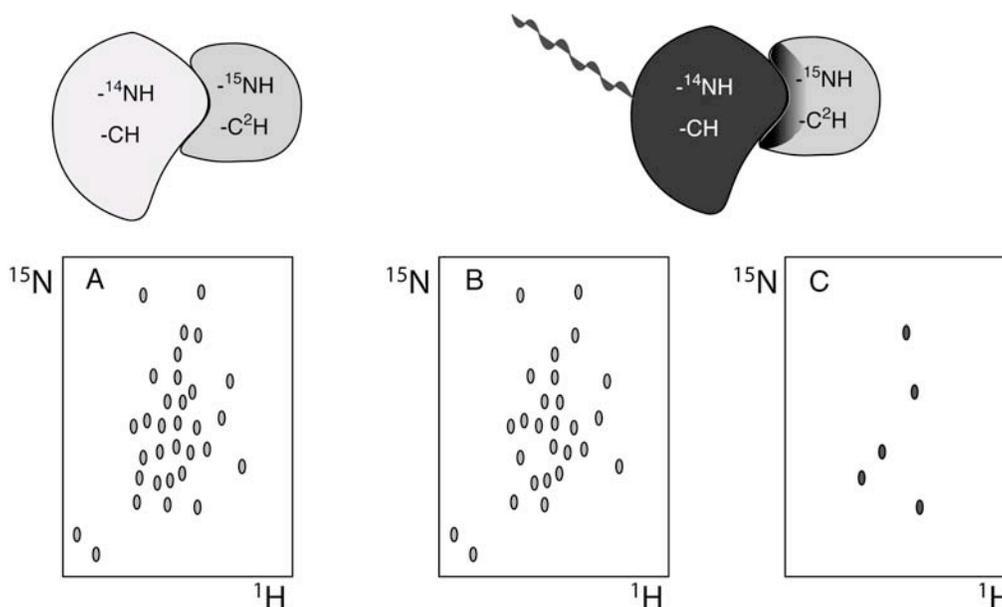

Figure 60 Schematic representation of ^{15}N-^{1}H cross-saturation TROSY-HSQC experiment for protein-protein complex, labeling requirements for the different components are depicted. A) control TROSY-HSQC experiment for ^{15}N-^{2}H labeled protein. B) Experiment with on-resonance irradiation on the aliphatic protons of the unlabeled partner. C) Spectrum produced by subtracting off and on-resonance experiments.

The nature of the experiment requires a particular labeling scheme for the various parts of the complex. Regarding the observed protein, this has to be ^{15}N uniformly labeled to allow the acquisition of HSQC experiment. Also, in order to achieve selectivity in the saturation of the macromolecular partner, all non-exchangeable protons in the ^{15}N labelad protein must be substituted by deuterons. As for the irradiated protein partner, this has no especial need for stable isotopes as long as it has aliphatic protons.

In order to meet the above conditions, VEGF was over-expressed in *E. coli* with a slightly modified protocol where minimal medium was prepared with D$_2$O and containing ^{2}H-glucose and ^{15}NH$_4$Cl as carbon and nitrogen sources respectively. Purification and refolding proceeded normally and incorporation of the various isotopes did not affect significantly the overall yield; finally protein fold and percentage of deuteration were checked by MALDI-TOF mass spectrometry and NMR experiments (1D ^{1}H and ^{15}N-^{1}H HSQC spectra).

It is clear from the monodimensional NMR experiment (Figure 61) and MALDI-TOF results that deuteron proton substitution was efficiently achieved. MS spectrometric data for the deuterated protein batch indicates that its molecular weight is 12344.5 Da, compared to previous ^{15}N uniformly labeled VEGF, with

11758 Da per monomer. Assuming quantitative incorporation of ^{15}N and ^{2}H into VEGF, the expected values for ^{15}N-^{2}H uniformly labeled VEGF would be 12361.5 Da whereas for ^{15}N-U VEGF we would expect 11758 Da per monomer. From this data it seems that neither isotope is incorporated quantitatively, probably as a result of impurities in the isotope sources. If we account for the ^{15}N incomplete incorporation, which results in a 3 Da correction with respect to the expected value, there is still a 14 Da difference between the theoretical and experimental MW values for ^{15}N-^{2}H VEGF; and considering a total of 603 non-exchangeable protons our yield of deuterium incorporation in the protein is 97.7%. Quite a reasonable figure given the 98% isotopic purity in the glucose used for the minimal media.
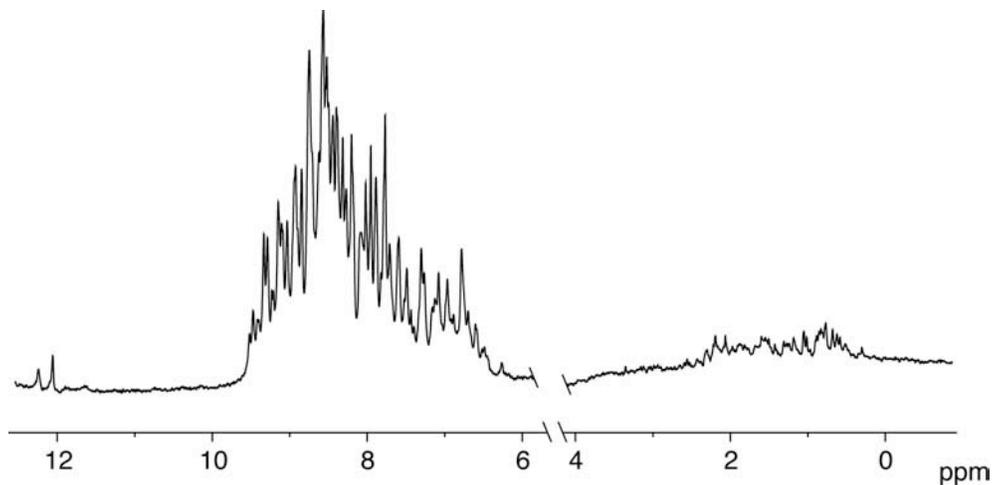

Figure 61 ^{1}H Watergate NMR spectrum for a ^{15}N-^{2}H uniformly labeled VEGF sample.
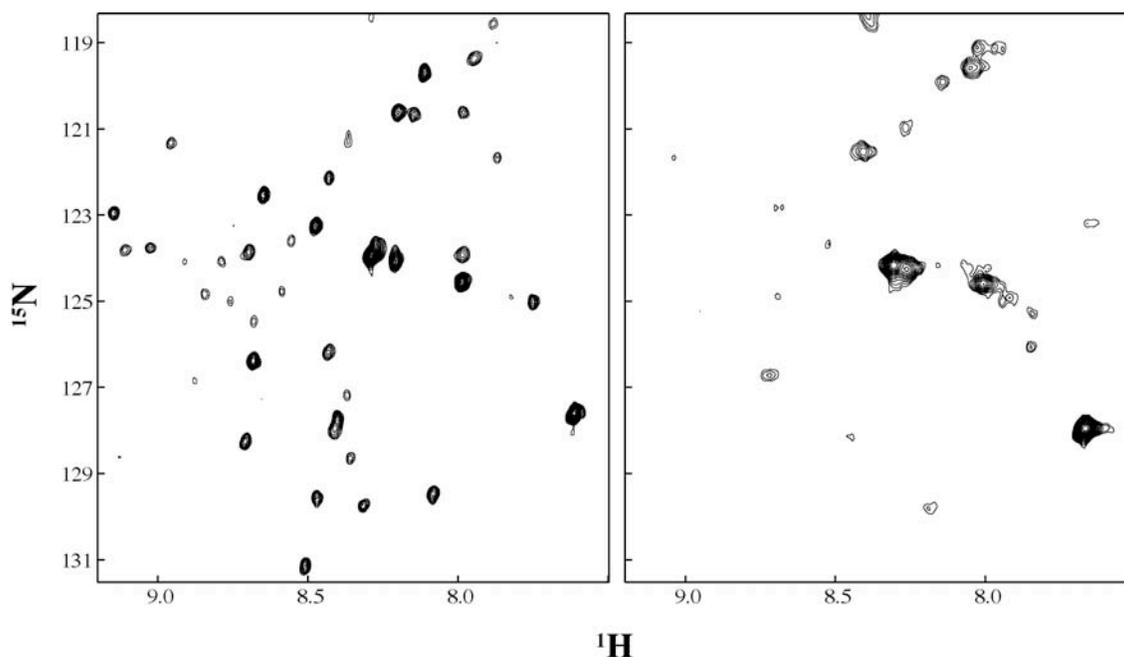

Figure 62 Portions of ^{15}N-^{1}H TROSY-HSQC experiments for uniformly ^{15}N-^{2}H labeled VEGF at 10°C (right) and ^{15}N-VEGF at 20°C.

2D heteronuclear NMR experiments also suggest that deuterium incorporation occurred satisfactorily and that this protein batch was well folded. Interestingly, the quality of spectra for the perdeuterated protein was impressively high both for the standard and TROSY HSQC experiments. In both experiments we obtained good resolution and practically no signal overlap thanks to a considerable cross-peak sharpening. The improvements were so remarkable as to allow us to record good quality experiments at room temperature, where non-deuterated VEGF batches produce poor spectra. (Figure 62) This is consistent with previous reports by Fairbrother and collaborators when working with $VEGF_{10-109}$ construct [23], the authors mention some difficulties when working at high protein concentrations (above 0.5 mM) that they attribute to a micro-aggregation phenomenon. As a consequence of micro-aggregation the protein probably displays a higher apparent MW and thus adverse relaxation properties; these effects are partly alleviated by a combination of perdeuteration and TROSY-type experiments yielding the observed spectrum quality improvements.

2.4.2   EXPERIMENT

After successfully producing ^{15}N-^{2}H VEGF we decided to set up cross-saturation TROSY pulse program and optimize several of its parameters with our model complex: v107-VEGF. This complex is not very large, at least compared to previously reported systems [56]; nonetheless due to its affinity, v107-VEGF complex is still a valuable test bench.
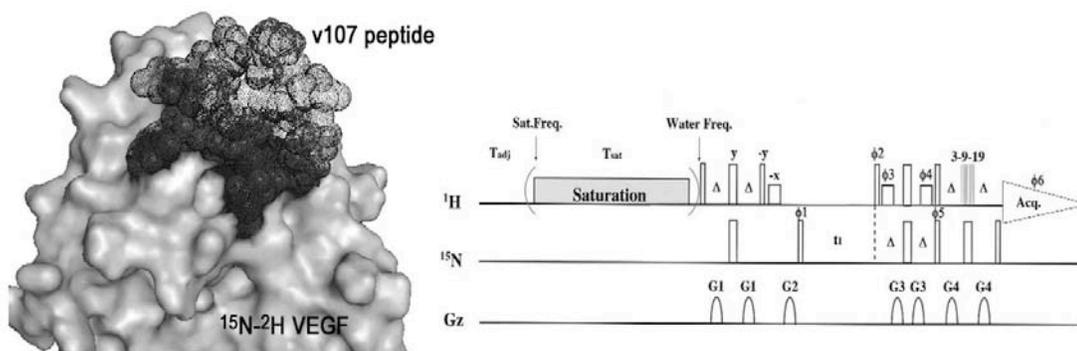

Figure 63 (Left) Model complex composed by ^{15}N-^{2}H uniformly labeled VEGF (grey surface) and phage display peptide v107 (dotted black surface). Pulse sequence for cross-saturation TROSY-HSQC experiment as reported by Takahashi and collaborators [56].

Figure 63 depicts the experiment's pulse sequence, for which parameters such as temperature, recycling delay or saturation scheme (shaped pulse, time, etc.) will have to be explored. Of utmost importance will be setting up the experiment so that on and off-resonance experiments are acquired simultaneously, or what is the same, in an interleaved fashion. This will provide reliability to the end subtraction spectrum, and given the duration of the experiment, will avoid possible artifacts stemming from spectrometer instabilities over long periods of time; being particularly helpful when the efficiency of cross-saturation between receptor and ligand is low.

Aside from the practical details, this exercise with a relatively small system will be a good preliminary study before moving to harder systems, such as those involving even smaller and weaker ligands. It will be very

interesting to see the saturation efficiency within v107, but more importantly the amount of saturation transferred to the protein; especially since the reduced size of our model complex will probably hamper the saturation and its diffusion within the complex thus affecting the overall sensitivity of the experiment.

On the other hand this low saturation efficiency may not be completely disadvantageous as it might allow us to reduce the amount of D_2O in sample preparations, reducing the required amount of protein for the experiment. In the seminal use of cross-saturation to map FB-Fc interaction,[56] Takahasi and collaborators used a mixture of 1:10 $H2O:D_2O$ in their sample preparation, this was conceived to further reduce the number of protons in the protein and achieve near-complete isolation for amide protons so that saturation could only originate from the partner's aliphatic protons and not being relayed from a neighboring-amide. Were they not to increase amide deuteration the enormous size and high efficiency of saturation diffusion in their complex could induce overestimation of the FB binding surface.

2.4.2.1 V107

Implementation of cross-saturation TROSY-HSQC experiment pulse sequence was performed using a 250 μM ^{15}N-^{2}H VEGF sample containing 1 mM v107 peptide. In particular several variations of the pulse sequence were set up incorporating different saturation schemes, either using continuous wave irradiation or various shaped pulses trains. With these sequences we explored values for saturation time, power and recycling delay by saturating at various frequencies on v107 peptide aliphatic protons.

Although we did observe differences in their efficiency, overall the various saturation schemes provide similar results given a reasonable recycling delay and saturation times. For instance, a 2s continuous wave irradiation produced the experiment depicted in Figure 64; for which saturation above 5% of the original cross-peak intensity were detected for some twenty amides. Assignment of the most saturated amides –those above the 10% saturation threshold- shows that, as expected, they belong to VEGF peptide binding surface and in fact are located very close to the irradiated peptide resonances. This suggests that saturation does not spread evenly throughout the peptide and remains localized around the irradiated resonances probably due to v107-VEGF complex size. The lack of thorough saturation on the peptide is an obvious inconvenient if one wants to get an accurate picture of VEGF's ligand binding interface; however being able to selectively saturate individual peptide resonances and later observe the cross-saturation effect only on the closest protein amides may provide interesting nOe-type information; which could eventually be used to guide *in silico* docking. In this sense, exploration of different saturation frequencies on v107 protons: β and α proton regions, provided remarkably different saturation profiles.

The scenario for amides suffering from 5 to 10% saturation was much more difficult to interpret structurally; these residues are distributed all over VEGF without a preferred location around v107 binding site and in some cases even include residues in the β-strand hydrophobic core. Interestingly, these amino acids tend to be valines, leucines or isoleucines or are otherwise sequentially located to one such residue, indicating that this saturation effects could stem from incomplete VEGF perdeuteration combined with saturation of these residual protons in the protein. In order to evaluate the existence and extent of such residual saturation we decided to perform control experiments using a similar VEGF sample but lacking v107 peptide. As suspected, experiments reveal the existence of certain residual saturation, however, under the same conditions used for Figure 64 amide saturation in control experiments remained below the 10% threshold, which is reassuring considering our previous structural analysis.

With the presented observations so far, it seems that our experimental set up is barely able to differentiate between artifactual VEGF residual saturation and v107-mediated saturation. This is particularly distressing since we regarded v107-VEGF complex as a test bench to later undertake two predictably harder systems: organic solvents or flavonoids. Furthermore, residual saturation is a direct consequence of a defficient VEGF deuteration, which can only be tackled by using purer ^{2}H sources and careful protein production. As a result, the yield of deuterium substitution represents a noise threshold that will have to be surpassed by ligand-mediated saturation and unfortunately modulating experimental conditions in order to enhance the cross-saturation effect will not allow us to identify or suppress such artifactual saturation. In fact, acquiring the same experiments at lower temperatures (10ºC) affects sample viscosity and increases the overall complex correlation time producing cross-saturation effects up to 40% for several amides; nonetheless the same enhancement occurs for the undesired residual saturation.
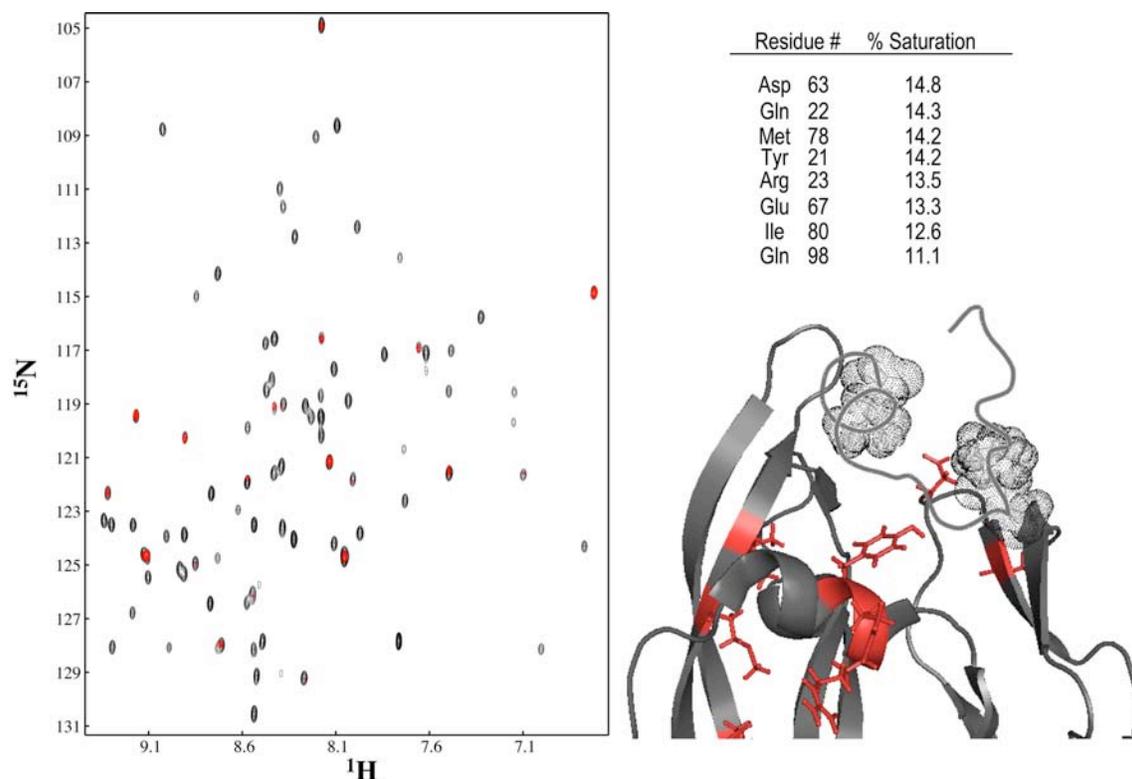

Residue #	% Saturation
Asp 63	14.8
Gln 22	14.3
Met 78	14.2
Tyr 21	14.2
Arg 23	13.5
Glu 67	13.3
Ile 80	12.6
Gln 98	11.1

Figure 64 (Left) Cross-saturation TROSY-HSQC experiments at 40°C for VEGF-v107 complex (250 μM ^{15}N-^{2}H VEGF sample and 1 mM v107 peptide). (black) Off-resonance experiment and (red) spectrum resulting from subtraction of the off-resonance and on-resonance -2s continuous wave irradiation at 0.8 ppm - experiments. (Right-top) List of assigned amides suffering saturation greater than 10% of its original intensity, these have been plotted in red on v107-VEGF complex structure (Right-bottom) and their side chains depicted as sticks. Van der Wals surface is shown for the aliphatic amino acids in v107, irradiated during the on-resonance experiment, the rest of the peptide is depicted as a grey ribbon.

2.4.3 LIGANDS: ORGANIC SOLVENTS AND FLAVONOIDS

Despite its apparent limitations we decided to carry out the cross-saturation experiment for ^{15}N-^{2}H VEGF in the presence of two weak binders identified and characterized throughout this chapter: these are isopropanol and quercetin-3-β-glucoside.

We prepared protein samples in the presence of both ligands, along with a control sample with ^{15}N-^{2}H VEGF alone. Later we measured cross-saturation TROSY experiments at 10 ºC with 2s saturation periods and long recycling delays for all samples. (Figure 65) In all experiments we could detect a significant amount of saturation, which resulted in decent subtraction experiments; unfortunately the amount of protein saturation in control experiments was in the same order. Consequently structural interpretation of these experiments was severely hampered and as anticipated estimation of VEGF-ligand interface could not be carried out, once more underscoring the limitations of our experimental set up.
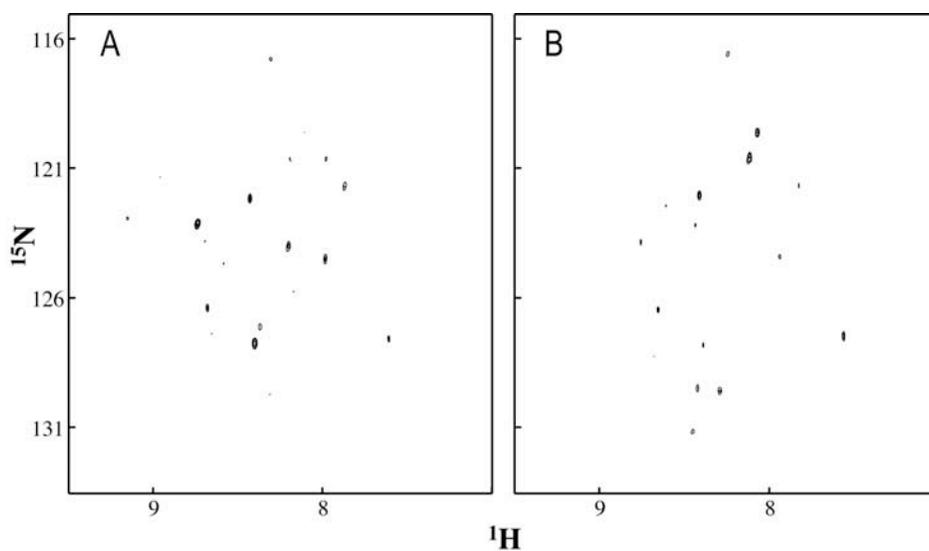

Figure 65 Cross-saturation TROSY-HSQC experiments for a 250 µM ^{15}N-^{2}H VEGF protein sample and 5% (v/v) isopropanol (B). Experiments have been acquired at 10°C and the resulting spectrum has been produced by subtracting off and on resonance -2s continuous wave irradiation on isopropanol resonances - experiments. Control experiments were carried out with the protein alone. (A)

2.5 BIBLIOGRAPHY

1.   Ferrara, N., *Role of vascular endothelial growth factor in Physiologic and Pathologic angiogenesis: Therapeutic implications.* Seminars In Oncology, 2002. **29**(6): p. 10-14.
2.   Ferrara, N. and R.S. Kerbel, *Angiogenesis as a therapeutic target.* Nature, 2005. **438**(7070): p. 967-974.
3.   Ferrara, N., *Vascular endothelial growth factor as a target for anticancer therapy.* Oncologist, 2004. **9 Suppl 1**: p. 2-10.
4.   Ferrara, N., H.P. Gerber, and J. LeCouter, *The biology of VEGF and its receptors.* Nature Medicine, 2003. **9**(6): p. 669-676.
5.   Muller, Y.A., et al., *The crystal structure of vascular endothelial growth factor (VEGF) refined to 1.93 A resolution: multiple copy flexibility and receptor binding.* Structure, 1997. **5**(10): p. 1325-38.
6.   Ferrara, N., et al., *Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer.* Nature Reviews Drug Discovery, 2004. **3**(5): p. 391-400.
7.   Fairbrother, W.J., et al., *Novel peptides selected to bind vascular endothelial growth factor target the receptor-binding site.* Biochemistry, 1998. **37**(51): p. 17754-64.
8.   Hajduk, P.J., et al., *Privileged molecules for protein binding identified from NMR-based screening.* J Med Chem, 2000. **43**(18): p. 3443-7.
9.   Gill, A., A. Cleasby, and H. Jhoti, *The discovery of novel protein kinase inhibitors by using fragment-based high-throughput x-ray crystallography.* Chembiochem, 2005. **6**(3): p. 506-12.
10.   Wiesmann, C., et al., *Crystal structure of the complex between VEGF and a receptor-blocking peptide.* Biochemistry, 1998. **37**(51): p. 17765-72.
11.   Pan, B., et al., *Solution structure of a phage-derived peptide antagonist in complex with vascular endothelial growth factor.* J Mol Biol, 2002. **316**(3): p. 769-87.
12.   Merrifield, R.B., et al., *Retro and retroenantio analogs of cecropin-melittin hybrids.* Proc Natl Acad Sci U S A, 1995. **92**(8): p. 3449-53.
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.   Yampolsky, L.Y. and A. Stoltzfus, *The exchangeability of amino acids in proteins.* Genetics, 2005. **170**(4): p. 1459-1472.
15.   Fan, T.P., et al., *Angiogenesis: from plants to blood vessels.* Trends Pharmacol Sci, 2006. **27**(6): p. 297-309.
16.   Yance, D.R., Jr. and S.M. Sagar, *Targeting angiogenesis with integrative cancer therapies.* Integr Cancer Ther, 2006. **5**(1): p. 9-29.
17.   Wani, M.C., et al., *Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia.* J Am Chem Soc, 1971. **93**(9): p. 2325-7.
18.   Liu, C., A. Tseng, and S. Yang, *Chinese Herbal Medicine*. 2005: CRC Press. 904.
19.   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.
20.   Dalvit, C., et al., *Reliable high-throughput functional screening with 3-FABS.* Drug Discov Today, 2004. **9**(14): p. 595-602.
21.   Frutos, S., T. Tarrago, and E. Giralt, *A fast and robust 19F NMR-based method for finding new HIV-1 protease inhibitors.* Bioorg Med Chem Lett, 2006. **16**(10): p. 2677-81.
22.   Tarrago, T., et al., *Identification by 19F NMR of traditional Chinese medicinal plants possessing prolyl oligopeptidase inhibitory activity.* Chembiochem, 2006. **7**(5): p. 827-33.
23.   Fairbrother, W.J., et al., *1H, 13C, and 15N backbone assignment and secondary structure of the receptor-binding domain of vascular endothelial growth factor.* Protein Sci, 1997. **6**(10): p. 2250-60.
24.   Hendrickson, W.A., *Determination of macromolecular structures from anomalous diffraction of synchrotron radiation.* Science, 1991. **254**(5028): p. 51-8.

25. 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.

26. Coghlan, P.A. and C.J. Easton, *beta-nitro-alpha-amino acids as latent alpha,beta-dehydro-alpha-amino acid residues in peptides.* Tetrahedron Letters, 1999. **40**(25): p. 4745-4748.

27. Dalvit, C., et al., *Use of organic solvents and small molecules for locating binding sites on proteins in solutions.* J Biomol NMR, 1999. **14**(1): p. 23-32.

28. Byerly, D.W., C.A. McElroy, and M.P. Foster, *Mapping the surface of Escherichia coli peptide deformylase by NMR with organic solvents.* Protein Sci, 2002. **11**(7): p. 1850-3.

29. Liepinsh, E. and G. Otting, *Organic solvents identify specific ligand binding sites on protein surfaces.* Nat Biotechnol, 1997. **15**(3): p. 264-8.

30. English, A.C., et al., *Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol.* Proteins, 1999. **37**(4): p. 628-40.

31. Mattos, C., et al., *Multiple Solvent Crystal Structures: Probing Binding Sites, Plasticity and Hydration.* Journal of Molecular Biology, 2006. **357**(5): p. 1471.

32. Grzesiek, S., et al., *The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase.* Nature Structural Biology, 1996. **3**(4): p. 340-345.

33. Hajduk, P.J., J.R. Huth, and C. Tse, *Predicting protein druggability.* Drug Discov Today, 2005. **10**(23-24): p. 1675-82.

34. 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.

35. Newman, D.J. and G.M. Cragg, *Marine natural products and related compounds in clinical and advanced preclinical trials.* J Nat Prod, 2004. **67**(8): p. 1216-38.

36. Tabak, C., et al., *Chronic obstructive pulmonary disease and intake of catechins, flavonols, and flavones - The MORGEN Study.* American Journal Of Respiratory And Critical Care Medicine, 2001. **164**(1): p. 61-64.

37. Martinez-Dominguez, E., R. de la Puerta, and V. Ruiz-Gutierrez, *Protective effects upon experimental inflammation models of a polyphenol-supplemented virgin olive oil diet.* Inflammation Research, 2001. **50**(2): p. 102-106.

38. Matsuzaki, Y., et al., *Cell death induced by baicalein in human hepatocellular carcinoma cell lines.* Japanese Journal Of Cancer Research, 1996. **87**(2): p. 170-177.

39. Kumagai, T., D. Heber, and H.P. Koeffler, *Scutellaria and its major components, baicalin and baicalein: Effect against a variety of human cancers.* Blood, 2003. **102**(11): p. 260B-260B.

40. Sonoda, M., et al., *Cytotoxic activities of flavonoids from two Scutellaria plants in Chinese medicine.* Journal Of Ethnopharmacology, 2004. **91**(1): p. 65-68.

41. Ueda, S., et al., *Baicalin induces apoptosis via mitochondrial pathway as prooxidant.* Molecular Immunology, 2002. **38**(10): p. 781-791.

42. Liu, J.J., et al., *Baicalein and baicalin are potent inhibitors of angiogenesis: Inhibition of endothelial cell proliferation, migration and differentiation.* Int J Cancer, 2003. **106**(4): p. 559-65.

43. Shen, Y.C., et al., *Mechanisms in mediating the anti-inflammatory effects of baicalin and baicalein in human leukocytes.* European Journal Of Pharmacology, 2003. **465**(1-2): p. 171-181.

44. Mayer, M. and B. Meyer, *Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor.* J Am Chem Soc, 2001. **123**(25): p. 6108-17.

45. Meyer, B., et al., *Saturation transfer difference NMR spectroscopy for identifying ligand epitopes and binding specificities.* Ernst Schering Res Found Workshop, 2004(44): p. 149-67.

46. McGovern, S.L., et al., *A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening.* J Med Chem, 2002. **45**(8): p. 1712-22.

47. Widmer, H. and W. Jahnke, *Protein NMR in biomedical research.* Cell Mol Life Sci, 2004. **61**(5): p. 580-99.

48.   Morse, M.A. and G.D. Stoner, *Cancer chemoprevention: principles and prospects.* Carcinogenesis, 1993. **14**(9): p. 1737-46.
49.   Kojima-Yuasa, A., et al., *Green tea extract inhibits angiogenesis of human umbilical vein endothelial cells through reduction of expression of VEGF receptors.* Life Sci, 2003. **73**(10): p. 1299-313.
50.   Lamy, S., D. Gingras, and R. Beliveau, *Green tea catechins inhibit vascular endothelial growth factor receptor phosphorylation.* Cancer Res, 2002. **62**(2): p. 381-5.
51.   Tang, F.Y., N. Nguyen, and M. Meydani, *Green tea catechins inhibit VEGF-induced angiogenesis in vitro through suppression of VE-cadherin phosphorylation and inactivation of Akt molecule.* Int J Cancer, 2003. **106**(6): p. 871-8.
52.   Lee, S.J., et al., *1,2,3,4,6-Penta-O-galloyl-beta-D-glucose blocks endothelial cell growth and tube formation through inhibition of VEGF binding to VEGF receptor.* Cancer Lett, 2004. **208**(1): p. 89-94.
53.   Luque, I., S.A. Leavitt, and E. Freire, *The linkage between protein folding and functional cooperativity: two sides of the same coin?* Annu Rev Biophys Biomol Struct, 2002. **31**: p. 235-56.
54.   Luque, I. and E. Freire, *Structural stability of binding sites: consequences for binding affinity and allosteric effects.* Proteins, 2000. **Suppl 4**: p. 63-71.
55.   Carlson, H.A., *Protein flexibility and drug design: how to hit a moving target.* Current Opinion In Chemical Biology, 2002. **6**(4): p. 447-452.
56.   Takahashi, H., et al., *A novel NMR method for determining the interfaces of large protein-protein complexes.* Nat Struct Biol, 2000. **7**(3): p. 220-3.
57.   Bertini, I., et al., *ePHOGSY experiments on a paramagnetic protein: location of the catalytic water molecule in the heme crevice of the oxidized form of horse heart cytochrome c.* FEBS Lett, 1997. **415**(1): p. 45-8.
58.   Chen, A. and M.J. Shapiro, *NOE pumping: A novel NMR technique for identification of compounds with binding affinity to macromolecules.* Journal Of The American Chemical Society, 1998. **120**(39): p. 10258-10259.