Combined use of NMR and computational tools for fragment based drug discovery targeting protein-protein interactions VEGF protein surface recognition as a case study

Michael Goldflam

Programa de Quìmica Orgànica

Tesi Doctoral

Combined use of NMR and computational tools for fragment based drug discovery targeting protein-protein interactions
VEGF protein surface recognition as a case study

Michael Goldflam

Dirigida i revisada per:

Dr. Ernest Giralt Lledó
(Institut de Recerca Biomèdica Barcelona
Universitat de Barcelona)

Barcelona, 2013

Tesi Doctoral

Combined use of NMR and computational tools for fragment based drug discovery targeting protein-protein interactions
VEGF protein surface recognition as a case study

Michael Goldflam

Departament de Quìmica Orgánica

Facultat de Quìmica

Univesitat de Barcelona

2013

Für meinen Vater

Table of contents

Abbreviations

ACN	Acetonitrile
Asn (N)	Asparagine
Asp (D)	Aspartic acid
CPMG	Carr-Purcell-Meiboom-Gill
CSP	Chemical shift perturbation
Cys (C)	Cysteine
δ	Chemical shift
DMF	Dimethyl formamide
DMSO	Dimethyl sulfoxide
DNA	Desoxyribonucleic acid
ESI	Electron spray ionization
FDA	Food and Drug Administration
Fmoc	9-Fluorenylmethyloxycabronyl
Gln (Q)	Glutamine
HATU	(Dimethylamino)-N,N-dimethyl(3H-[1,2,3]triazolo[4,5-b]pyridin-3-yloxy)methaniminium hexafluorophosphate
HBTU	O-(Benzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum correlation
HTS	High throuput screening
Ile (I)	Isoleucine
Leu (L)	Leucine
Lys (K)	Lysine
MALDI-TOF	Matrix assisted laser desorption ionization time of flight
mRNA	Messenger ribonucleic acid
MS	Mass spectroscopy
MW	Molecular weight
NMR	Nuclear magnetic resonance
Phe (F)	Phenylalanine
Pro (P)	Proline
PPI	Protein-protein interfaces
ppm	Parts per million
RMSD	Root mean square deviation
SAR	Structure activity relationship
STD	Saturation transfer difference
TFA	Tri fluoroacetic acid
Thr (T)	Threonine
tRNA	Transfer ribonucleic acid
Trp (W)	Tryptophan
VEGF	Vascular endothelial growth factor

1 Introduction

1.1 The emergence of protein-protein interfaces as a new class of therapeutic targets

The capacity of proteins to interact rests at the core of biology. From cellular architecture to information transfer and chemical specificity, everything relies on highly precise recognition events, which frequently involve the assembly of two or more proteins. These interactions can occur with high or low affinity, but clearly not in a random disorganized way. Contrary, the interaction of proteins is closely scripted to achieve specific goals as the assembly of subcellular architectures or the relay of signals. Protein-protein interaction is an omnipresent aspect of cellular biochemistry. It includes the mammalian immune response, cell-cell recognition, attachment to the extracellular matrix, signal transduction from the cell surface to the nucleus and transcription itself, which is mediated by a multitude of proteins whose assembly is poorly understood but clearly important.

Given the ubiquitous nature of these relationships, and the knowledge that misregulated protein-protein interactions can lead to disease, it is not surprising that protein-protein interfaces have attracted the attention of scientists both in academia and in the pharmaceutical industry who are interested in producing inhibitors for use as therapeutic agents or biochemical tools. This growing interest occurs in a time when the drug discovery society is in a crisis.

Despite advances in technology, basic science and increased research expenditures, the introduction of drugs has at best remained approximately constant while the attrition rates have risen sharply, especially in late-phase clinical trials.[1] As a result the risk and cost of developing a new drug has increased. This also has implications on society, since increasing prices for new drugs lead to further burden of national healthcare systems. An extreme example is Eculizumab, a monoclonal antibody with annually costs over $400,000 per treatment.[2]

The reason for the decline in productivity is under debate. Some scientists have proposed that the major culprit is the exhaustion of easy targets, so-called low hanging fruits, thereby raising the bar for research success.[3] This may be indeed part of the explanation. In the last decade much effort was directed towards proteins that belong to classes which were already successful drugged. For instance 40% of the most common drugs are beta-blockers, beta-agonists and dopamine or serotonin antagonists that bind to G-protein-coupled receptors, although they constitute probably to less than 10% of the 20,000 genes identified in the human genome.[4]

Targeting protein-protein interfaces and modulation of protein-protein interactions opens in this context a new venue in the development of therapeutics and biochemical tools. As stated above, protein-protein interactions are ubiquitous in the cell. Venkatesan and coworker estimate around

130,000 binary protein interactions present in the human interactosome.[5] Even if only a fraction of these new targets could be manipulated in a therapeutic manner it would still add substantially to the estimated 2,000 to 3,000 classical drug targets.[6]

How this new class of therapeutic targets differ from classical drug targets, how they can be targeted and with what success will be discussed in the next section.

(1)   Pammolli, F.; Magazzini, L.; Riccaboni, M. *Nat Rev Drug Discov* **2011**, *10*, 428.
(2)   *http://www.medicalbillingandcoding.org/blog/the-11-most-expensive-medicines-in-america accessed 12.02.2013*
(3)   Burns LR (Editor), The Business of Healthcare Innovation, Chapter 2. *Cambridge University Press* **2005**,
(4)   Strosberg, A. D.; Nahmias, C. *Biochem Soc Trans* **2007**, *35*, 23.
(5)   Venkatesan, K.; Rual, J. F.; Vazquez, A.; Xin, X.; Goh, K. I.; Yildirim, M. A.; Simonis, N.; Heinzmann, K.; Gebreab, F.; Sahalie, J. M.; Cevik, S.; Barabasi, A. L.; Vidal, M. *Nat Methods* **2009**, *6*, 83.
(6)   Russ, A. P.; Lampel, S. *Drug Discov Today* **2005**, *10*, 1607.

1.2 Protein-protein interfaces and therapeutics

The majority of drug discovery efforts have been focus on so called "druggable" proteins. Druggability is a term to describe proteins which can be manipulated with a therapeutic benefit and can bind a drug with sufficient affinity. In practice these druggable proteins come from distinct families but have been evolved to interact with small molecular metabolites. The binding sites can be pockets on the protein surface or even deep cavities inside of the protein. In both cases the interfaces are relatively small, between 300 to 1000 Å. Protein-protein interfaces have been evolved to bind to their corresponding protein partner. Unlike druggable proteins their interfaces are generally flat and span areas between 1500 to 3000 Å. They often have some inherit flexibility to interact with multiple binding partners.

A good example for these characteristics is VEGF, the protein this thesis is focusing on (Figure 1.2). It interacts with two extracellular receptors. Both share an overlapping interface on VEGF which indicates some flexibility of the biding site.[1,2] The interface is relatively flat, does not show any pockets and covers, for the case of the complex with VEGFR2, an area around 2000 Å.
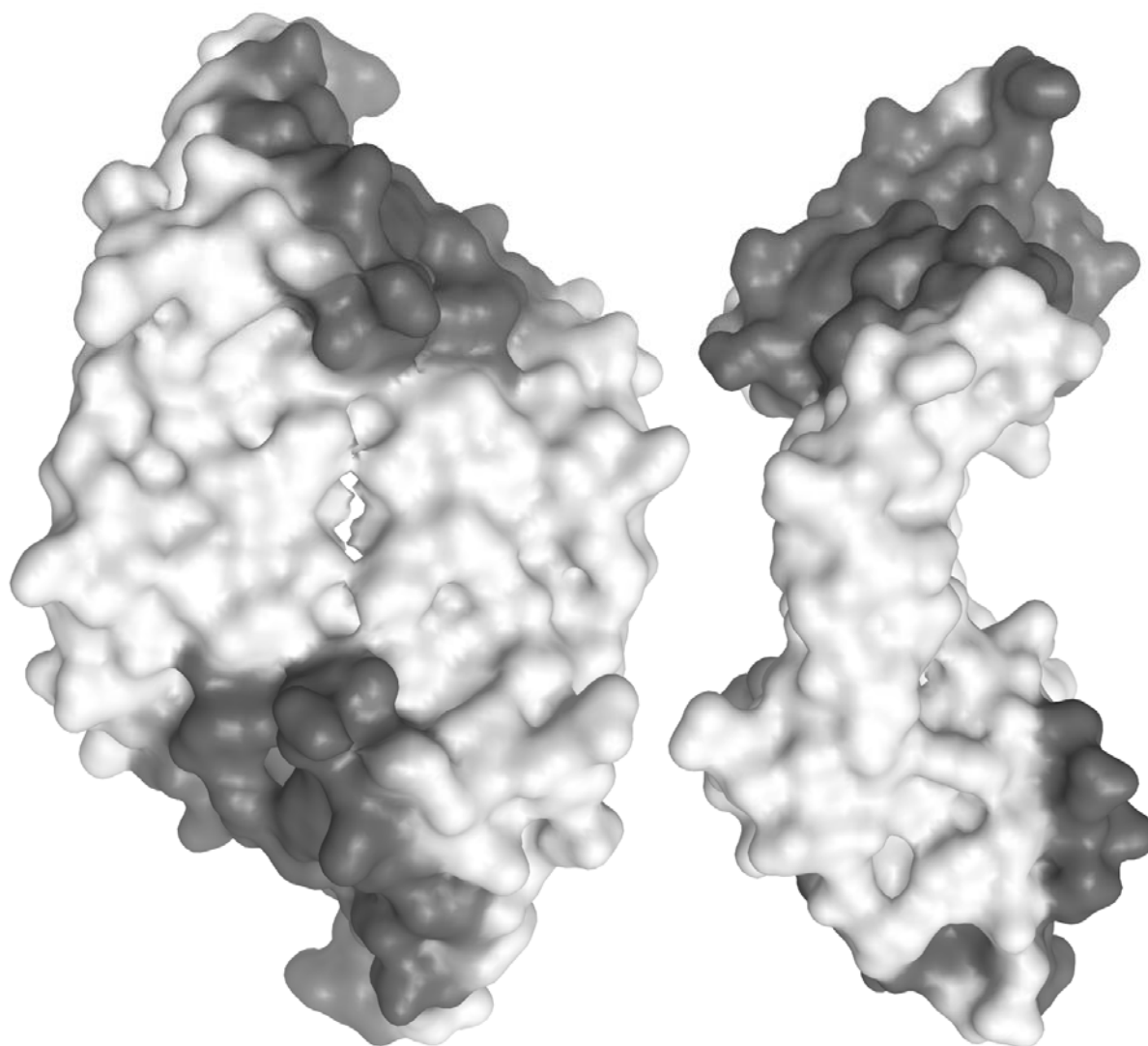
Figure 1.2: Surface representation of the homodimer VEGF from 3V2A. The interface involved in interaction with the receptor KDR/VEGFR-2 is depicted in red.

Clearly protein-protein interfaces differ from small molecule interfaces. The question that arises is how to develop compounds that can bind to these interfaces. As will be discussed in the next paragraphs the use of biologics appeared to be a relatively straightforward approach, although with some intrinsic disadvantages. Contrary protein-protein interfaces like as of VEGF were considered to be undruggable in respect to small molecules, although some recent progress has been made.

(1)   Brozzo, M. S.; Bjelic, S.; Kisko, K.; Schleier, T.; Leppanen, V. M.; Alitalo, K.; Winkler, F. K.; Ballmer-Hofer, K. *Blood* **2012**, *119*, 1781.
(2)   Wiesmann, C.; Fuh, G.; Christinger, H. W.; Eigenbrot, C.; Wells, J. A.; de Vos, A. M. *Cell* **1997**, *91*, 695.

1.2.1 The use of biologics to target PPI

Biologics, or biological drugs, are therapeutics of biological origin. This includes antibodies, interleukins, vaccines, recombinant proteins and aptamers. They are generally biopolymers in the kDa or MDa range with high affinities and specificities towards their target protein. Antibodies in particular enjoy without doubt unique success in the treatment of diseases that would be difficult to tackle with other classes of therapeutics.[1] Examples of FDA-approved biologics that bind at protein-protein interfaces includes, but is not limited to, Humira,[2] Enbrel,[3] Remicade[4] and Avastin[5]. The last was developed to bind to the above mentioned VEGF. All of these therapeutics are among the top selling drugs in the US, with several billions of revenue.[6]

Biologics have some inherent drawbacks despite their great therapeutic and economic success: They are all based on large biopolymers, which the human immune system can recognizes as foreign. All currently marketed antibodies have exhibited some level of immunogenicity.[7] Many biological therapeutics are used for lifelong treatment. Although they are often well tolerated, clinical manifestation can range from local skin reactions at the injection site, pyrexia and influenza-like syndromes to acute anaphylaxis and systemic inflammatory response syndrome, which can be fatal. These side effects can worsen with the progress of the treatment. Of particular concern is that some of the adverse effects of biologics that were recently encountered were not anticipated by the currently available preclinical screening tools and animal models. [8]

The inability of biologics to cross biological barriers in an efficient way poses a further limitation to their general use. As a result they can only be administered by injection and targets inside the cell or beyond the blood-brain barrier are outside of their reach. Recent advances may improve the situation for some biologics, but have yet to proven to be generally applicable. [9]

Finally, biologics are very expensive. Obviously they cannot be synthesized like small molecule drugs but have to be produced by recombinant expression from cell cultures. Nevertheless advocates of biologics emphasize their straightforward development to high affinity and specificity and the lower attrition rate in clinical trials, which should require lower research investment and balance the increased production cost at least to some degree.

This is apparently not the case. Although the vast majority of drugs are small molecules[10], 9 of the 10 most expensive drugs in 2012 are biologics, with annual treatment costing from $400000 to $200,000.[11] This has raised the question of who can afford these drugs as exemplified in chapter 1.**5.3** with the VEGF binding antibody Avastin.

(1)   Brekke, O. H.; Sandlie, I. *Nat Rev Drug Discov* **2003**, *2*, 52.
(2)   Scheinfeld, N. *J Drugs Dermatol* **2003**, *2*, 375.
(3)   Klareskog, L.; Gaubitz, M.; Rodriguez-Valverde, V.; Malaise, M.; Dougados, M.; Wajdula, J. *Ann Rheum Dis* **2006**, *65*, 1578.

(4)   Raza, A.; Candoni, A.; Khan, U.; Lisak, L.; Tahir, S.; Silvestri, F.; Billmeier, J.; Alvi, M. I.; Mumtaz, M.; Gezer, S.; Venugopal, P.; Reddy, P.; Galili, N. *Leuk Lymphoma* **2004**, *45*, 2099.

(5)   Genentec, Avastin Prescribing Information, **October 2006**.

(6)   U.S.Pharmaceutical-Sales . *http://www.drugs.com/stats/top100/sales* **2012,** *accessed 12.02.2013*

(7)   Tabrizi, M. A.; Roskos, L. K. *Drug Discov Today* **2007**, *12*, 540.

(8)   Hansel, T. T.; Kropshofer, H.; Singer, T.; Mitchell, J. A.; George, A. J. *Nat Rev Drug Discov* **2010**, *9*, 325.

(9)   Pardridge, W. M. *J Drug Target* **2010**, *18*, 157.

(10)   Drugbank. *http://www.drugbank.ca/databases, accessed 12.02.2013*

(11)   *http://www.medicalbillingandcoding.org/blog/the-11-most-expensive-medicines-in-america, accessed 12.02.2013*

1.2.2 The use of small molecules to target PPI:

Potentially, small molecules can be cheaper to produce, do not have the risk to raise immunogenicity and can be designed to cross biological barriers. This enables them to target protein-protein interactions that are present inside the cell and the brain, areas which are so far not readily accessible to biologics. Finally they are quite stable and can often be administered as oral drugs. Unfortunately it is under open debate how druggable protein-protein interfaces are for small molecules. The biophysical properties of protein-protein interfaces seem to be contradictory to classical druggability indices and many attempts to develop ligands have failed. Furthermore, protein-protein interfaces possess no natural small molecule interaction partner which could serve as a starting point for drug discovery. This could also be an advantage, since pockets for common metabolites such as ATP are evolutionarily conserved in many protein classes. Drugs that bind in these pockets often face specificity problems.

A concept which might explain why a small molecule, which can cover only a fraction of the interface, is capable to compete with a huge protein is based on the presence of hotspots.

Although protein-protein interfaces are large, only a subset of the involved residues contribute significantly to the free binding energy. These residues are often clustered and located at the centre of the interface. Hotspots could be a sweet spot for small molecules to gain enough free binding energy to compete with the native protein binding partner despite their size.[1]

Recent publications report the successful development of small molecule inhibitors for at least some types of protein-protein interactions. Unfortunately negative results are not published. Certainly it would be very important for the case of protein-protein interfaces to understand not only why we succeed when we do so, but also why we fail in other cases.

Two recent retrospective studies have been made to move beyond case examples and to analyse general principles that rule druggability for protein-protein interfaces and strategies to develop inhibitors. In 2009 Higueruelo et al published the hand-curated TIMBAL database which contains information regarding small molecules (excluding short peptides and) and their target proteins.[2] At

the time of the study the database contained 104 molecules for 18 protein targets. Table 1 summarizes the target protein-protein complexes, the complex type, the discovery techniques and the outcome in terms of ligands and closely related series.

	Complex	Complex type	Therapeutic area	Techniques	No of SM (series)
Extracellular	IL-2 /IL-2Ra	Heterodimer	Immuno-suppressor	Peptidomimetics, Tethering	6 (2) + 247 tethers
	CD80 /CD28	Heterodimer	Immuno-suppressor	Cell based screening, HTS ELISA	4 (2)
	TNFα trimer	Homotrimer	Inflammation	CTGFA with ELISA	2 (1)
	ZipA /FtsZ	**Heterodimer (small peptide)**	Antibacterial	Screening SPR and FPA, SBDD	**21 (7)**
Intra-cellular	Bcl2 or BclXL/Bax or Bak or Bid	**Heterodimer (small peptide)**	Oncology	SBDD, VS-Docking, FPA, HTS, FPA, SAR by NMR	**26 (9)**
	b-Catenin /Tcf4 or Tcf3	Heterodimer	Oncology	SBDD, VS-Docking, NMR, ITC	4 (2)
	c-Myc /Max	Heterodimer binding to DANN	Oncology	Screening FPA	1
	ESX /Sur-2	**Heterodimer (small peptide)**	Oncology	Cell based screening + binding assay	1
	p53 /MDM2	**Heterodimer (small peptide)**	Oncology	SBDD, VS-Docking and FPA, LBDD, VS-Pharmacophore and FPA, Peptidomimetics, Natural products, HTS ThermoFluor, ELISA, SPR, FPA	**16 (7)**
	p53 /S100B	**Heterodimer (small peptide)**	Oncology	SBDD, VS-Docking, Trp Fluorescence assay	**7 (4)**
	XIAP /Casp-9 or SMAC	Heterodimer	Oncology	Peptidomimetics, Natural products, SBDD, VS-Docking, FPA	5 (2)
	UL30(Pol) /UL42	Subunits HSV	Antiviral	HTS, FPA	3 (3)
	E1–E2 /DNA(HPV)	Heterodimer binding to DNA	Antiviral	SAR by NMR	4 (2)
	ToxT dimer	Homodimer	Antimicrobial	HTS phenotypic screen	1
	iNOS dimer	Homodimer	Inflammation Immunology	CombiChem	1
	RGS4 /Gao	Heterodimer	Modulation of GPCRs	Screen FCPI assay	1
	CMR1 /NES	Heterodimer	Antiviral	Cell based screen	2 (1)

Table 1: Summary of the TIMBAL database. Abbreviations: HTS, High-throughput screening; ELISA, Enzyme-linked immunosorbent assay; CTGFA, Combinatorial target-guided fragment assembly, Sunesis; SPR, Surface Plasmon Resonance; FPA, Fluorescence Polarization Assay; SBDD, Structure-based Drug Design; VS, Virtual Screening; SAR, Structure Activity Relationship; NMR, Nuclear Magnetic Resonance. SAR by NMR; ITC, Isothermal titration calorimetry; LBDD, Ligand-Based Drug Design; ThermoFluor; CombiChem, Combinatorial Chemistry; FCPI, Flow Cytometry Protein Interaction assay. For protein abbreviations see original paper. Adapted from[2]

The bold and underlined characters mark trends: the majority of successful campaigns targeted complexes with known structures (standart characters) vs. complexes without known structures (underlined characters). Bold highlights the targets with the major number of reported small molecule inhibitors. All of them have known structural information. Interestingly, high numbers of ligands (bold) correlate with complexes where one of the partners is a peptide (bold), except for the complex without structural information. This would suggest that these types of interfaces are more druggable than the interfaces from globular constituents. Indeed highlighted proteins BCl-XL, MDM2 and S100B have small and deep pockets to accommodate the peptides which can be exploited for small molecules. Contrary inhibition of ZipA to FtsZ is inhibited without penetration of the small

molecule into the protein surface. This indicates that druggability and chances of success cannot be simply limited to the analysis of surface pockets.

In 2010 and 2012 Bourgeas and Basse analyzed a hand-curated database entitled P2PIdb, focusing on orthosteric small molecule inhibitors were both the protein–protein and protein–ligand complexes have been structurally characterized.[3,4] The database contains, at the moment, 14 PPI interactions and 52 small molecule inhibitors. Protein complexes were divided into two classes based on the number of segments incorporated into the interfaces. Class 1 contains protein-peptide complexes that use a lower number of segments, which consist mainly of alpha-helices. The corresponding peptide is generally well ordered, which might be easier to mimic with small molecules or peptide mimetic as indicated by the greater number of inhibitors developed for this type of interface. This analysis correlates to the results from the TIMBAL database. Class 2 contains globular protein-protein interfaces, which consist of a higher number of segments. These interfaces are larger than class 1 interfaces and the involved proteins bind with higher affinities.

A comparison of the P2PIdb complexes against structurally characterized heterodimer complexes showed that the interfaces with known inhibitors are significantly smaller: 530 Å and 760 Å for class 1 and 2, compared to 1000 Å observed for standard heterodimer complexes without known inhibitors. Further differences were a higher density of hydrogen bonds, fewer salt bridges and less charged residues at the interface. Depending on the point of view this could be either interpreted as crude indicator of druggabiliy for protein-protein interfaces or illustrate the bias of our knowledge and technology to target interfaces that are more closely related to classical drug targets.

Shifting the focus from the proteins to the actual inhibitors Higueruelo et al analysed the small molecules from the TIMBAL database with approved drugs, drug-like ligands inside the protein data base (PDB) and typical screening compounds. They concluded that the known protein-protein inhibitors have higher molecular weight, higher lipophilicity, more ring systems, but less rotable bonds and fewer hydrogen bond acceptors and donors. Unfortunately they do not separate the inhibitors for protein-protein and protein-peptide interfaces and also do not analyse the content of the sp3 and stereocenter, which has been proposed to be beneficial for protein-protein inhibitors[5,6]. Nevertheless this analysis suggests that the chemical space of protein-protein inhibitors differs to some extend from known protein ligands and known drugs. But one should remember that this is again a retrospective analysis biased by the technologies that we developed and knowledge that we gained from targeting classical drug targets. Interestingly in a classical drug discovery context the majority of protein-protein inhibitors would be considered as potentially promiscuous and problematic regarding ADMET properties[2].

Many techniques have been used to identify protein-protein inhibitors as reviewed in the TIMBAL database. For three reasons fragment based drug discovery could be a key technology to drive the discovery for this challenging class of proteins forward.[1] First, as can be seen in Table 1, it has already performed well with protein–peptide interfaces and contributed to the discovery to some of the ligand series for these proteins. Second, current protein-protein inhibitors are not very druglike and might give higher attrition rates in clinical trials. Based on an analysis of Murray et al., leads from fragment based drug discovery have improved physicochemical properties. They are in average 62 Da smaller and one log unit less lipophilic compared to HTS leads.[7,8] As discussed in the next chapter fragments make weak but high quality interactions and are evolved gradually into leads which allows constant control of druglike properties. Third, some indications point out that the chemical space of protein-protein inhibitors might be different than for classical targets. Should this be true it will be much more economic to redesign fragment libraries, which contain hundreds or thousands of compounds, rather than HTS libraries, which contain 10^5 to 10^6 compounds.

These arguments highly suggest to further study the performance of fragment based drug discovery when applied to protein-protein interface. This is especially true for interfaces of globular protein complexes.

(1)   Wells, J. A.; McClendon, C. L. *Nature* **2007**, *450*, 1001.
(2)   Higueruelo, A. P.; Schreyer, A.; Bickerton, G. R.; Pitt, W. R.; Groom, C. R.; Blundell, T. L. *Chem Biol Drug Des* **2009**, *74*, 457.
(3)   Bourgeas, R.; Basse, M. J.; Morelli, X.; Roche, P. *PLoS One* **2010**, *5*, e9598.
(4)   Basse, M. J.; Betzi, S.; Bourgeas, R.; Bouzidi, S.; Chetrit, B.; Hamon, V.; Morelli, X.; Roche, P. *Nucleic Acids Res* **2012**.
(5)   CJ, O. C.; Beckmann, H. S.; Spring, D. R. *Chem Soc Rev* **2012**, *41*, 4444.
(6)   Hung, A. W.; Ramek, A.; Wang, Y.; Kaya, T.; Wilson, J. A.; Clemons, P. A.; Young, D. W. *Proc Natl Acad Sci U S A* **2011**, *108*, 6799.
(7)   Murray, C. W.; Verdonk, M. L.; Rees, D. C. *Trends Pharmacol Sci* **2012**, *33*, 224.
(8)   Keseru, G. M.; Makara, G. M. *Nat Rev Drug Discov* **2009**, *8*, 203.

1.3 Fragment based drug discovery

The term fragment based drug discovery describes the search for potential drugs starting from small chemical fragments. The word fragment refers to small organic compounds with molecular weights between 150 to 300 Da. This is in contrast to drugs or compounds tested in HTS which can exceed 500 Da. Fragment based drug discovery is conceptually closely related to the work of Fesik, Hajduk and colleagues in the late 1990s[1,2]. Since then the approach gained momentum and is nowadays widely applied in industry and academia to target conventional protein targets.[3-7]

The success of fragment based drug discovery is based on two rationales. First, the chemical ligand space can be probed much better when a smaller threshold for the maximal molecular weight is

chosen. It has been estimated that the number of chemical compounds which can be imagined until a weight of 500 Da are around 10^{60-200} in contrast to estimated 10^7 compounds for a limit of 300 Da.[3,4] Therefore compound libraries for fragment-based discovery are much smaller than HTS libraries. As by 2012 most fragment libraries in industry and academia range from 500 to 10,000 compounds compared to up to 10^6 member libraries for HTS[3]. The investment and resources to build, maintain and screen such huge libraries are tremendous, while by comparison fragment libraries are quite economical. A better sampling of the chemical space also leads to improved hit rates in fragment-based screening. Researchers from Novartis observed hit rates of 0.001% - 0.151% from HTS with an $IC_{(50)}$ threshold in the μM range and hit rates above or equal to 3% observed in NMR screening of fragments with an affinity threshold in the mM range.[4] This is often summarized as smaller is better (MW of compounds), and that size does not always matter (library size).[5] Nonetheless comparing numbers can never recapitulate the complexity of drug discovery especially when different targets and approaches are compared. An aspect which completely escaped the plain number of library size is the quality of the library as will be discussed later.

The second advantage of fragment based drug discovery is that although fragment hits are weak binders, they must form high quality interactions with the target protein to bind sufficiently for detection. This leads to high ligand efficiency, which can be computed in many ways but is most often expressed as LE, defined as -ΔG in kilocalories per mole divided by the number of heavy (nonhydrogen) atoms.[6] On the other side larger molecules explored in HTS may bind due to numerous suboptimal interactions, which leads to lower LE values. Even if this is not the case Rees et al. estimated that the typical fragment with an affinity of 100 μM forms between -42.8 and -37.8 kJ mol^{-1} of favorable interactions compared to a drug-sized molecule with a 3nM affinity that forms between -68.6 and -63.6 kJ mol^{-1} of favorable interactions. This means that if the fragment is part of this potent drug and still forms the same interactions, it would contribute to over the half of the favorable energy despite being 33,000 times weaker in affinity.[7] The reason for this is that the entropic penalty for binding is in both cases is approximately the same.[8] Therefore although the affinity of fragments is weak they offer very good starting points for elaboration. This is important because successful drugs require not only tight binding (e.g. nM affinity) to the target protein but have to posses certain physicochemical properties as summarized for example in the Lipinsky rule of 5: less than 5 hydrogen bond donors, less than 10 hydrogen bond acceptors, a MW less than 500 Da and an octanol-water partition coefficient not greater than 5. Although this rule is controversial it is obvious that a successful drug requires appropriate ADMET properties. Fragments are generally small soluble compounds. They can be elaborated to higher affinity by adding functional groups or linking fragments while maintaining the desired physicochemical properties. This is challenging for HTS hits, because they are often quite lipophilic and close to the 500 Da limit. Therefore these parameters

have to be optimized without substantial increases in size. Failing in doing so has been proposed as one of the major cause of attrition for small-molecule drugs.[7]

Fragment based drug discovery is not without disadvantages. The intrinsic weak affinity requires biophysical techniques that are sensitive enough to detect these weak binding events. Fortunately there are a multitude of established methods that can be used as described later in this chapter.

The fragment based drug discovery process can be divided in three key stages:

1. The design of the screening library:

Any screening project can only be as good as the employed library. So where do these compounds come from? One possibility is to purchase a prepared library from one of the various international vendors. This approach is quite popular with academic groups and does not require knowledge of library design. Furthermore these libraries are relatively cheap. The drawback is that they probably sample a small chemical space since they might be based only on compounds from a single vendor. The other possibility is to design a proprietary library as most pharmaceutical companies do. The composition of these libraries are undisclosed but several companies published design guidelines, such as Hubbard and coworker at Vernalis.[9,10] They assembled a database with over one million unique chemical compounds from major international suppliers. From this database compounds were removed that had undesired functionalities (e.g. reactive groups, sugars) and did not fulfill druglike requirements (e.g. at least one ring system and presence of at least one desired functional group). The remaining compounds with a molecular weight between 110 and 250 Da were clustered based on their similarity and one compound of the highest populated cluster selected. Finally the designated compounds were evaluated based on possible insolubility and ability for rapid chemical evolution of the fragment scaffold. Although not encoded in the selection criteria the designed library was close to conforming to the rule of 3 (analogous to the rule of 5, but for fragments), which demands a molecular weight under 300 Da, up to three hydrogen bond donors and acceptors and a calculated logP of under 3.

What are the advantages of this library design? Each library compound represents a maximal cluster of similar compounds. Therefore the library covers a huge chemical space. Further the elaboration of hits is straightforward: either related compounds for each hit can be purchased from vendors or their synthesis should be feasible. This library design is entitled "SAR by catalog" and is similar to our own library design as elucidated in chapter 3.3. Other popular designs are the closely related diversity oriented libraries and focused libraries. These are not general libraries but aim on a specific target or classes as kinases, ion-channels, nuclear receptors and maybe in the future protein-protein interfaces.

2. Fragment screening and validation:

Fragment hits have inherent weak affinity in the range of 0.1-10 mM. To detect these weak interactions sensitive biophysical screening methods must be used. Compared to HTS these techniques have lower throughput and require high fragment concentration to detect binding. Various biophysical methods are available for fragment screening. They all come with different intrinsic advantages and disadvantages and are often combined in a screening cascade to validate results or gain additional information. The choice of technique to use at each stage of this cascade is determined by the physical properties of the target protein, the level of throughput required and the equipment available to the researcher. For the initial stage of screening throughput is important. NMR offers a plenitude of ligand based experiments which are robust, fast (if mixtures of compounds are screened) and offer a superior sensitivity. If the protein is relatively robust it can be immobilized on a chip and screened by SPR with very small amounts of protein. Fluorescence-based thermal shift assays are relative rapid, inexpensive and require only an experimentally accessible thermal denaturation transition of the target protein. If soakable protein crystals can be obtained X-ray crystallography can be used to cocrystallize fragment mixtures, which might not have a high throughput but reveals information about the fragment binding mode. Further successful but less applied techniques are mass spectrometry, isothermal calorimetry (ITC), fluorescence resonance energy transfer (FRET) and the related amplified luminescent proximity homogeneous assay (ALPHA), fluorescence polarisation (FP) or fluorescence anisotropy (FA) and, lately, microscale thermophoresis (MST).

All above methods give rise to false positives and negatives. Therefore cross validation with one or several additional techniques is typically performed, either with the whole library or with selected hits. It is also necessary to further characterize the hits. Of central importance is the generation of a binding hypothesis which can be generated by X-Ray crystallography or by ligand and protein based NMR experiments.[17,18] Without this knowledge elaboration of hits can only rely on systematic exploration of large libraries contrary to the rational design of additional interactions, which has better chances of success with less investment. Further information should be available about the affinity to calculate the LE and gain quantitative structure-activity relationships which can be generated by any of the above mentioned techniques. Ideally information about kinetic and thermodynamic of the binding event is available by SPR and respectively ITC.

3. Iterative Elaboration of fragments.

Validated fragment hits are elaborated to improve affinity while maintaining druglike properties. This is done in an iterative process of rational design, synthesis, and exploration of commercial compounds, guided by structural binding information and quantitative affinity data. In practice there are three main approaches to evolving the affinity of fragments:

Fragment growing, merging and linking.[4,19] Fragment growing has proved to be the more straightforward and less problematic approach, since it relies on a single hit as the starting point. Before investing time and resources into growing a fragment, its potency should be optimized by exploring closely related compounds. Then the fragment is modified in positions which give rise to additional interactions predicted by the binding model and the outcome evaluated by one of the biophysical approaches. In this way the effect of each introduced modification is directly assessed and either dismissed or incorporated into future cycle of elaboration. During each step metrics such as LE should be checked to ensure that the final compound arrives in druglike space.

Fragment merging can be applied if two overlapping fragment binding poses are available. If both fragments participate in quality interactions with the protein, the gain of affinity can be substantial when they are combined in a hybrid. This is not as straightforward as it sounds since both compounds have to be merged without disturbance of each other's interactions. This is not always successful and requires a high resolution binding hypothesis of both partners.

Fragment linking relies on connecting two fragments in adjacent binding pockets. If successful it can be the fastest approach to high affinity ligands. For example if two weak 1 mM affinity fragments are linked they form a potent 1 nM affinity compound. The reason for this is that the entropic penalty of binding only has to be paid once for the linked compound. However this is only true if both fragments can be connected with an appropriate linker that does not interfere with the optimal binding geometries and does not introduce excessive conformational flexibility.[11] In praxis it is often impossible to predict, even with high-resolution structures in hand, how a linker will affect binding of two fragments.

As elaborated above NMR offers experiments for relative high throughput screening and information rich experiments for hit validation and characterisation. Often it is used as one of the key technologies to drive fragment based drug discovery forward. NMR will be described in more detail in the next chapter and chapter 3.1 of the results section.

(1)   P. J. Hajduk, G. S., D. G. Nettesheim , E. T. Olejniczak , S. B. Shuker , R. P. Meadows , D. H. Steinman , G. M. Carrera , Jr., P. A. Marcotte , J. Severin , K. Walter , H. Smith , E. Gubbins , R. Simmer , T. F. Holzman , D. W. Morgan , S. K. Davidsen , J. B. Summers, S. W. Fesik *J. Am. Chem. Soc.* **1997**, *119*, 5818.
(2)   Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531.
(3)   Zartler, T. *http://practicalfragments.blogspot.de/2012/09/how-big-is-library.html* **2012**. Accessed 12.02.2013
(4)   Schuffenhauer, A.; Ruedisser, S.; Marzinzik, A. L.; Jahnke, W.; Blommers, M.; Selzer, P.; Jacoby, E. *Curr Top Med Chem* **2005**, *5*, 751.
(5)   Zartler, E. R.; Shapiro, M. J. *Curr Opin Chem Biol* **2005**, *9*, 366.
(6)   Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. *Proc Natl Acad Sci U S A* **1999**, *96*, 9997.
(7)   Murray, C. W.; Rees, D. C. *Nat Chem* **2009**, *1*, 187.
(8)   Finkelstein, A. V.; Janin, J. *Protein Eng* **1989**, *3*, 1.

(9)   Baurin, N.; Aboul-Ela, F.; Barril, X.; Davis, B.; Drysdale, M.; Dymock, B.; Finch, H.; Fromont, C.; Richardson, C.; Simmonite, H.; Hubbard, R. E. *J Chem Inf Comput Sci* **2004**, *44*, 2157.

(10)   Chen, I. J.; Hubbard, R. E. *J Comput Aided Mol Des* **2009**.

(11)   Chung, S.; Parker, J. B.; Bianchet, M.; Amzel, L. M.; Stivers, J. T. *Nat Chem Biol* **2009**, *5*, 407.

1.4 NMR as a tool to study protein ligand interactions

Nuclear magnetic resonance has evolved into a powerful tool for obtaining massive amounts of data on inter- and intra-molecular processes. Use of NMR to detect protein-ligand interactions is widely documented in the literature.[1-3] An advantage of NMR over other techniques is that it provides access to a broad set of experiments that have been optimized for various objectives: determination of affinity and specificity; identification of binding epitopes on the protein and on the ligand; characterization of structural rearrangements induced by binding; turnover of substrates by enzymes and general structural elucidation. Furthermore, since the experiments are performed in solution, physiological or near physiological conditions are possible.

NMR has been an omnipresent technique in the field of fragment based drug discovery since its very beginning. In the late 90s Fesik, Hajduk and colleagues used protein observing experiments in their pioneering work to generate a drug-like compound with nanomolar affinity.[4,5] Nowadays many researchers prefer to use ligand-based NMR experiments to screen fragment-based libraries because they offer higher throughput and require less protein. Subsequently, protein-based NMR experiments are used to validate and characterize promising hits. An exhaustive summary of ligand- and protein-based NMR experiments would go beyond the scope of this introduction. However during the course of my PhD I participated in the conception of a chapter for the book "Protein NMR Techniques" published as part of the "Methods in Molecular Biology" series.[6] The chapter offers a review of the field and detailed explanations of one selected ligand and one selected protein-based experiment which rank in my opinion among the most potent and widely used experiments. Please review chapter 3.1 for a reprint.

(1)   Fielding, L. *Curr Top Med Chem* **2003**, *3*, 39.

(2)   Carlomagno, T. *Annu Rev Biophys Biomol Struct* **2005**, *34*, 245.

(3)   Lepre, C. A.; Moore, J. M.; Peng, J. W. *Chem Rev* **2004**, *104*, 3641.

(4)   P. J. Hajduk, G. S., D. G. Nettesheim , E. T. Olejniczak , S. B. Shuker , R. P. Meadows , D. H. Steinman , G. M. Carrera , Jr., P. A. Marcotte , J. Severin , K. Walter , H. Smith , E. Gubbins , R. Simmer , T. F. Holzman , D. W. Morgan , S. K. Davidsen , J. B. Summers, S. W. Fesik *J. Am. Chem. Soc.* **1997**, *119*, 5818.

(5)   Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531.

(6)   Goldflam, M.; Tarrago, T.; Gairi, M.; Giralt, E. *Methods Mol Biol* **2012**, *831*, 233.

1.5 Vascular endothelial growth factor

1.5.1 VEGF and angiogenesis

Vascular endothelial growth factor (VEGF) is one of the most important regulators of angiogenesis. This term describes a physiological process which involves the growth of new blood vessels from pre-existing vessels. Angiogenesis occur extensively during embryogenesis as well as during general development in the formation of the cardiovascular system. However angiogenesis is not a static process. It is controlled in a dynamic way by pro- and anti-angiogenic factors during the whole adult life span. In the healthy body angiogenesis is involved in wound healing, restoring the blood flow to injured tissue and in females during the monthly reproductive cycle and during pregnancy. Angiogenesis proceeds in a well-regulated cascading fashion (Figure 1.5.1).
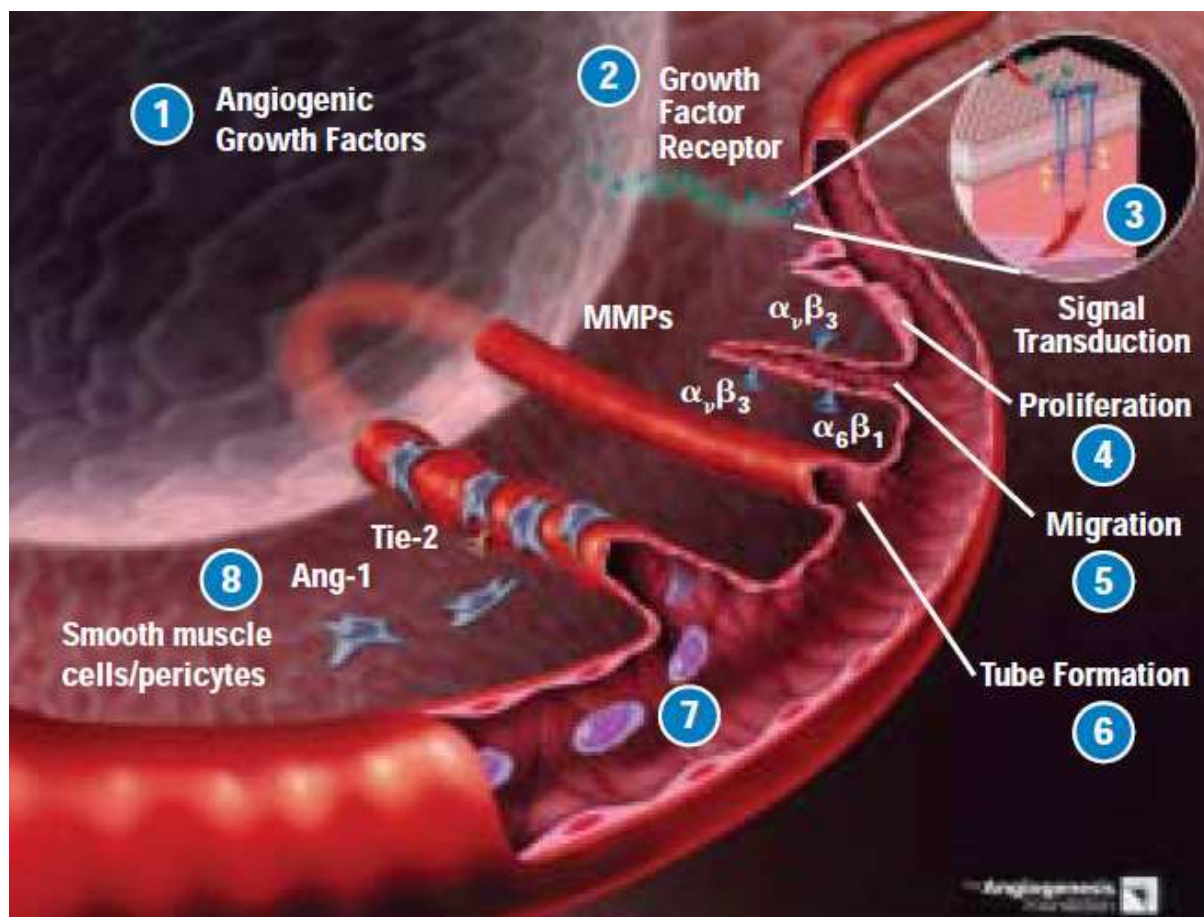
Figure 1.5.1: The angiogenic cascade: (1) Diseased or injured tissue produce and release growth factors that (2) bind to their receptors on endothelial cells, (3) activating signal transduction pathways and (4) stimulating endothelial proliferation, (5) migration, and (6) vascular tube formation. (7) Bone marrow derived endothelial stem cells are mobilized and become incorporated into new blood vessels. (8) Stabilization of the vasculature occurs through the recruitment of smooth muscle cells and pericytes. Adapted from [1]

Angiogenesis is mediated by a wide range of pro- and anti-angiogenic factors. VEGF is one of the key players in angiogenesis since it activates and mediates many steps in the angiogenic cascade in a

direct or indirect manner. It was discovered during the 1980s[2,3] and since then has been extensively studied in different *in vitro* and *in vivo* scenarios. VEGF's gene expression is induced by hypoxia[4,5] and coregulated by paracrine and autocrine growth factors[6,7]. It is well documented that VEGF promotes growth of endothelial cells[8], promotes angiogenesis in various aspects[9], enhances vascular permeability and promotes survival in endothelial cells by inhibition of apoptosis via the PI3 kinase/ Akt pathway[10] and expression of the antiapoptotic proteins Bcl-2, A1[11], XIAP[12] and survivin[13]. Interestingly the pro-survival effects of VEGF are developmentally regulated. It has been shown that VEGF inhibition resulted in extensive apoptotic changes in the vasculature of neonatal, but not adult, mice[13]. Furthermore, a significant VEGF dependence has been demonstrated in newly formed, but not established, vessels inside of tumors[13,14]. These results have implication in the use of anti-angiogenic treatment as discussed later.

(1)   Li, W. *Supplement Contemp Surg* **2003**, *November*.
(2)   Senger DR, G. S., Dvorak AM, Perruzzi CA, Harvey VS, Dvorak HF *Science* **1983**, *219*, 983.
(3)   Ferrara, N.; Henzel, W. J. *Biochem Biophys Res Commun* **1989**, *161*, 851.
(4)   Dor, Y.; Porat, R.; Keshet, E. *Am J Physiol Cell Physiol* **2001**, *280*, C1367.
(5)   Semenza, G. L. *Annu Rev Med* **2003**, *54*, 17.
(6)   Frank S, H. G., Breier G, Longaker MT, Greenhalgh DG,; S, W. *J Biol Chem* **1995**, *270*, J Biol Chem 270:12607.
(7)   Warren, R. S.; Yuan, H.; Matli, M. R.; Ferrara, N.; Donner, D. B. *J Biol Chem* **1996**, *271*, 29483.
(8)   Ferrara, N.; Davis-Smyth, T. *Endocr Rev* **1997**, *18*, 4.
(9)   Mesri, E. A.; Federoff, H. J.; Brownlee, M. *Circ Res* **1995**, *76*, 161.
(10)   Gerber HP, M. A., Kowalski J, Yan M, Keyt BA, Dixit V, Ferrara N *J Biol Chem* **1998**, *273*, 30336.
(11)   Gerber, H. P.; Dixit, V.; Ferrara, N. *J Biol Chem* **1998**, *273*, 13313.
(12)   Tran, J.; Rak, J.; Sheehan, C.; Saibil, S. D.; LaCasse, E.; Korneluk, R. G.; Kerbel, R. S. *Biochem Biophys Res Commun* **1999**, *264*, 781.
(13)   Tran, J.; Master, Z.; Yu, J. L.; Rak, J.; Dumont, D. J.; Kerbel, R. S. *Proc Natl Acad Sci U S A* **2002**, *99*, 4349.

1.5.2 VEGF isoforms and structure

VEGF, also referred to as VEGF-A belongs to a subfamily of the platelet-derived growth factor family of cystine-knot growth factors. Further members of the subfamily are Placenta growth factor (PlGF), VEGF-B, VEGF-C and VEGF-D, which share varying level of homology with VEGF. The broad term "VEGF" (VEGF-A) covers a number of proteins belonging to two families. They result from alternate mRNA splicing of a single VEGF gene containing 8 exons. The two families are referred to their terminal exon 8 splicing site as $VEGF_{xxx}$ for the proximal and $VEGF_{xxxb}$ as the distal spliced variant. Additionally, alternate splicing of exons 6 and 7 alter their heparin binding capability as well as the number of amino acids. In humans the isoforms $VEGF_{121}$, $VEGF_{121b}$, $VEGF_{145}$, $VEGF_{165}$, $VEGF_{165b}$, $VEGF_{189}$ and $VEGF_{206}$ as well as proteolytic isoforms as $VEGF_{110}$, the smallest of its active forms, could

be isolated.[1-3] As described in chapter 3.2, during the course of this work we were using a protein construct very similar to VEGF$_{110}$ which lacks exons 6 and 7. These two exons alter the affinity of VEGF to heparan sulfate proteoglycans and neuropilin co-receptors present on the cell surface. Interactions with both enhance the pro-angiogenic activity of VEGF but have no have significant effect on its affinity towards the main angiogenic VEGF receptors[4]. The terminal exon 8 splice site determines if the protein is pro-angiogenic (VEGF$_{xxx}$) or antiangiogenic (VEGF$_{xxxb}$). Consequently, controlling of VEGF exon 8 splicing has been proposed lately as a new strategy for anti-angiogenic treatment.[2,5] VEGF as well as the other member of the VEGF family can bind to two tyrosine kinase receptors on the cell surface, causing them to dimerize and become activated through transphosphorylation. VEGF interacts with VEGFR-2 (KDR/Flk-1). The activation of this receptor has been associated with proliferation, migration, survival and angiogenesis.[6] VEGF also interacts with VEGFR-1 (Flt-1), but its activation showed less clear biological response. This lead to VEGFR-1 being referred to as a "decoy receptor" and focused research on inhibition of VEGFR-2 mediated signaling. Conversely, recent research showed that this assumption may be wrong and proposed VEGFR-1 signaling as an additional target for angiogenic treatments.[7,8]

VEGF is most commonly expressed as the 165 amino acid isoform, a 45 kD glycoprotein. However, all VEGF isoforms share a common N-terminus receptor binding domain, that is structured as a covalent bonded highly symmetric homodimer (Figure 1.5.2). Each monomer consists of 4 antiparallel ß-sheets and a cysteine knot motif fixed by a network of three intramonomer disulfide bonds. These covalent linkages are quite important to the fold stability as the protein has a considerable lack of a hydrophobic core, as it is very exposed to the solvent. Equally important are the two symmetrical intermonomer disulfide bonds established between cysteines 51 and 60, which bear most of the burden of dimmer stability together with a set of hydrophobic contacts around the receptor-binding epitope. This epitope consists of parts of both dimers and extends over a broad region of the surface. Interestingly, NMR as well as X-ray studies report the epitope as 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.[1]
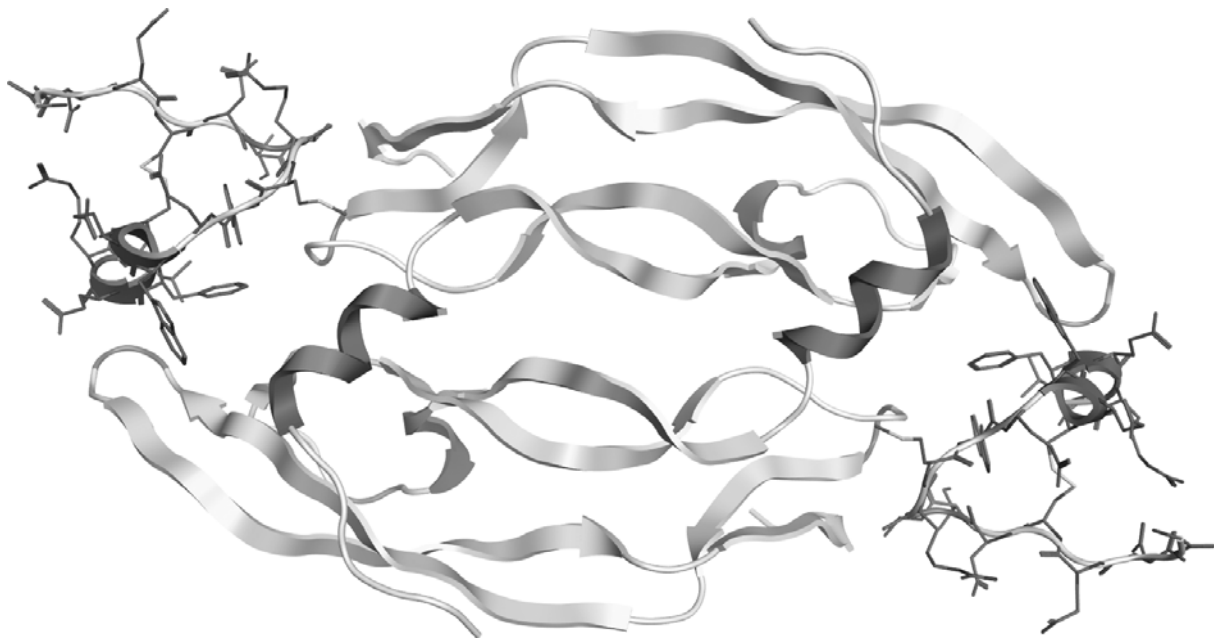
Figure 1.5.2: Solution structure of the VEGF8-109 isoform in complex with a phage-derived peptide antagonist binding to the receptor interface. From 1KAT.

(1)   Muller, Y. A.; Christinger, H. W.; Keyt, B. A.; de Vos, A. M. *Structure* **1997**, *5*, 1325.
(2)   Cebe Suarez, S.; Pieren, M.; Cariolato, L.; Arn, S.; Hoffmann, U.; Bogucki, A.; Manlius, C.; Wood, J.; Ballmer-Hofer, K. *Cell Mol Life Sci* **2006**, *63*, 2067.
(3)   Tischer, E.; Mitchell, R.; Hartman, T.; Silva, M.; Gospodarowicz, D.; Fiddes, J. C.; Abraham, J. A. *J Biol Chem* **1991**, *266*, 11947.
(4)   Ruhrberg, C. *Bioessays* **2003**, *25*, 1052.
(5)   Harper, S. J.; Bates, D. O. *Nat Rev Cancer* **2008**, *8*, 880.
(6)   Ferrara, N. *Endocr Rev* **2004**, *25*, 581.
(7)   Luttun, A.; Tjwa, M.; Moons, L.; Wu, Y.; Angelillo-Scherrer, A.; Liao, F.; Nagy, J. A.; Hooper, A.; Priller, J.; De Klerck, B.; Compernolle, V.; Daci, E.; Bohlen, P.; Dewerchin, M.; Herbert, J. M.; Fava, R.; Matthys, P.; Carmeliet, G.; Collen, D.; Dvorak, H. F.; Hicklin, D. J.; Carmeliet, P. *Nat Med* **2002**, *8*, 831.
(8)   Cao, Y. *Sci Signal* **2009**, *2*, re1.

1.5.3 VEGF, diseases and treatment

Angiogenesis is a highly regulated process in time and space. Misregulation and disorders that interfere with the angiogenic balance lead to pathological angiogenesis. VEGF has been validated as a therapeutic target to treat these pathologic conditions for various reasons: its levels correlate with the extent of angiogenesis, it is the most potent pro-angiogenic protein, the most specific growth factor, and the best characterized one; it stimulates endothelial cell growth, survival and proliferation; and it is over-expressed in a number of pathologic conditions.

One of the most advanced fields in terms of basic science and therapeutic development is the association of VEGF with cancer. One of the earliest discoveries identified VEGF as a vascular permeability factor that is secreted by tumor cell lines.[1] This appears to be true for the vast majority of tumors and lead to the definition of sustained angiogenesis as one of the hallmarks of cancer[2]. The

reason is that the growth of tumors is limited by their supply of nutrients and oxygen, which are distributed by blood vessels. Avascular tumors can not grow beyond 3mm in diameter and are considered benign. A tumor can overcome this limitation by overexpression of VEGF, which initiates the angiogenic cascade, as described above and leads to the generation of blood vessels that supply the tumor.[3] This turns the benign tumor into a malignant tumor which can rapidly grow and spread to other parts of the body by metastasis. Further, blood vessels surrounding the tumor differ from healthy vessels by an abnormal structure and function. Two implications of this are an impairment of drug delivery and that the blood vessels are leaky, which facilities the crossing of tumor cells into the blood stream that precedes metastasis.[4]

Multiple strategies were explored to disrupt VEGF signaling (Figure 1.5.3).
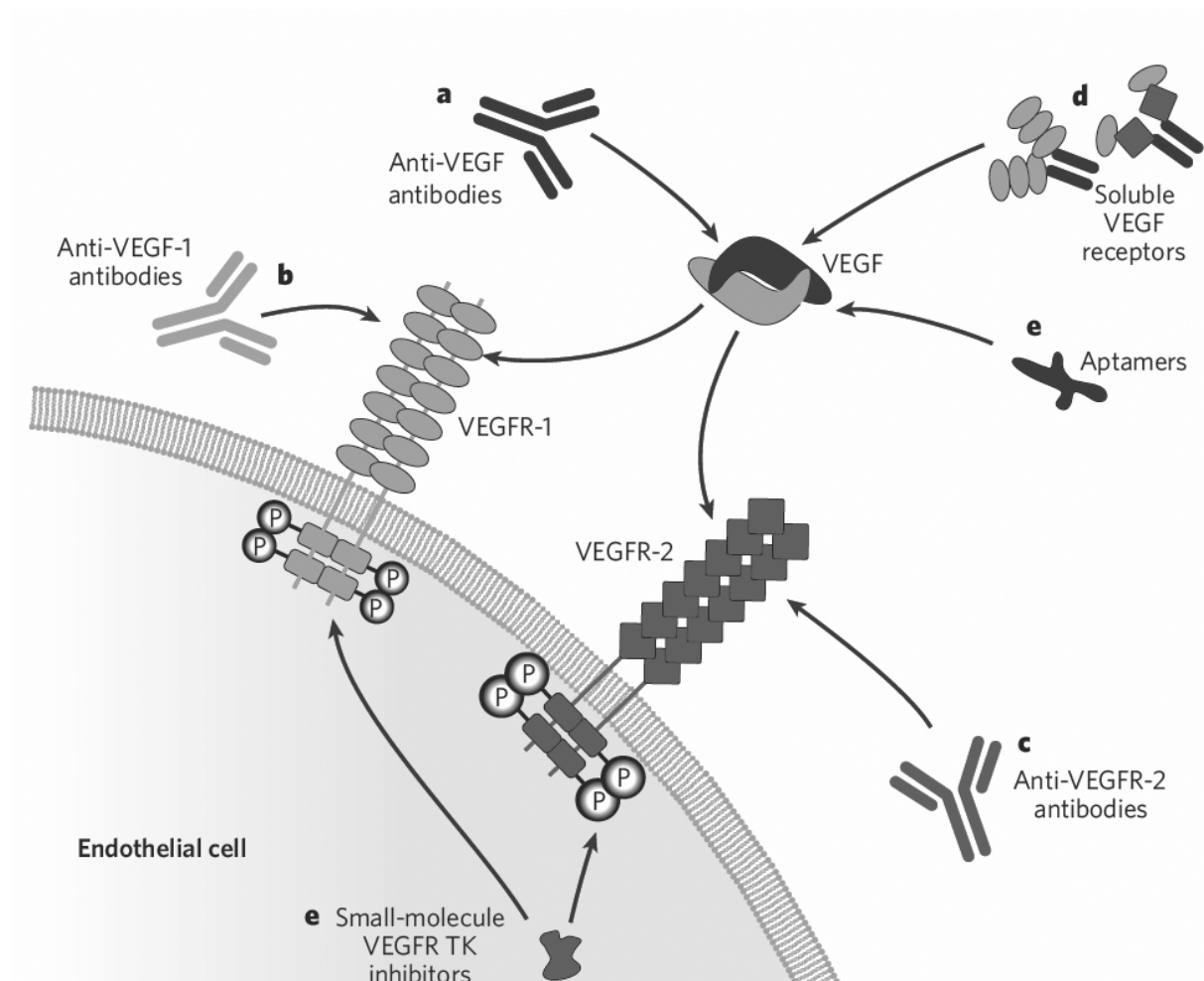
Figure 1.5.3: Strategies to disrupt VEGF signaling in pathological angiogenesis. From[5]

These strategies can be divided into two classes: intracellular approaches based on small molecules that target the kinase domain of the VEGF receptors and extracellular approaches based on biologics that target VEGF itself or the extracellular part of the VEGF receptors. It is important to note that there are no intracellular biological drugs due to on their inability to cross biological barriers.

To date there are three FDA approved second generation kinase inhibitors with anti-VEGF-receptor activity. Although these inhibitors have demonstrated antitumor activity, they are associated with a variety of undesired off-target effects related to their low specificity.[6]

The anti-VEGF signaling biologics are more successful in this regard. The most advanced and top selling drug is Bevacizumab (Avastin), which is approved for the treatment of metastatic colon cancer and certain lung cancers, renal cancers, ovarian cancer and glioblastoma multiforme. Bevacizumab is a scientific and economic success since it was the first FDA approved anti-angiogenesis drug and has multi-billion dollar revenue.[7] Despite this accomplishment, Bevacizumab is very controversial. Contrary to what was expected and shown in preclinical studies in mice, an overall survival benefit in humans could not be observed.[8,9] Instead, the use as an adjuvant to current cytotoxic regimens led to an improved overall survival in previously untreated colorectal cancer as well as improved progression-free survival in previously untreated breast cancer patients.[9] The reasons for these puzzling results are still under debate, but one reoccurring argument is the normalization of the abnormal blood vessels surrounding the tumor which may help drug delivery.[4] A further ambiguity is that in many cases the treatment showed a clear tumor shrinkage but only a relatively small gain of progression-free survival and in most trials no clear overall survival gain.[8]

Interestingly, recent results reported a previously unknown intracellular VEGF survival loop in certain tumor cell lines that could not be disrupted by extracellular biologics but required small molecule based inhibitors to kill the tumor.[10,11]

Bevacizumab is also criticized for its high cost, which varies with nation and cancer type but can be as high as $100,000 a year. As a result, some insurance companies and national health care systems such as the National Institute for Health and Clinical Excellence in the United Kingdom restricted the funding of Bevacizumab based treatments since it may only prolong life and not cure cancer.[12] In my opinion, Becacizumab is an interesting example of the complexity of cancer biology and the development of a treatment and how this process interacts with the interest of big pharma, national health systems and patients.

A second important pathologic condition that is causes by extensive angiogenesis is macular degeneration, the leading cause of vision impairment in elderly people in the developed world. In this disease overexpression of VEGF leads to blood vessel growth which damages the retina. The treatment of macular degeneration was revolutionized with the advent of intravitreous VEGF inhibitors. These state of the art drugs are either monoclonal antibodies (ranibizumab (Lucentis, Novartis, UK), and bevacizumab (Avastin, Roche, UK) or now less frequently pegaptanib (Macugen, Pfizer), an oligonucleotide aptamer.[13] All of these compounds bind directly to VEGF and inhibit its interactions with the corresponding receptors, thus inhibiting angiogenesis. The drawbacks of the

aforementioned drugs is that they are all biologics and have to be applied repeatedly by direct injection into the eye, a process that may pose risks and is inconvenient for the patient.[14]

A strategy that has thus far not been applied in treating pathologic angiogenesis could be based on small-molecule ligands that directly inhibit the protein-protein interactions of VEGF and its receptors. If one could overcome the challenge of developing these ligands they could unite a good specificity with membrane permeability and low production costs. So far very few attempts are reported in the literature to discover non-antibody based ligands that bind to the protein-protein interface of VEGF. With very few exceptions[15,16] these ligands are based on peptides and not small-molecules.[17,18]

(1)   Senger DR, G. S., Dvorak AM, Perruzzi CA, Harvey VS, Dvorak HF *Science* **1983**, *219*, 983.
(2)   Hanahan, D.; Weinberg, R. A. *Cell* **2011**, *144*, 646.
(3)   Hoeben, A.; Landuyt, B.; Highley, M. S.; Wildiers, H.; Van Oosterom, A. T.; De Bruijn, E. A. *Pharmacol Rev* **2004**, *56*, 549.
(4)   Jain, R. K. *Science* **2005**, *307*, 58.
(5)   Ferrara, N.; Kerbel, R. S. *Nature* **2005**, *438*, 967.
(6)   Bhargava, P.; Robinson, M. O. *Curr Oncol Rep* **2011**, *13*, 103.
(7)   *http://www.fiercepharma.com/special-reports/avastin, accessed on 5.12.2012*.
(8)   Hayes, D. F. *JAMA* **2011**, *305*, 506.
(9)   Jain, R. K.; Duda, D. G.; Clark, J. W.; Loeffler, J. S. *Nat Clin Pract Oncol* **2006**, *3*, 24.
(10)   Lee, T. H.; Seng, S.; Sekine, M.; Hinton, C.; Fu, Y.; Avraham, H. K.; Avraham, S. *PLoS Med* **2007**, *4*, e186.
(11)   Lichtenberger, B. M.; Tan, P. K.; Niederleithner, H.; Ferrara, N.; Petzelbauer, P.; Sibilia, M. *Cell* **2010**, *140*, 268.
(12)   Briggs, H. B. N. H. R. **2010**.
(13)   Park, Y. G.; Rhu, H. W.; Kang, S.; Roh, Y. J. *J Ophthalmol* **2012**, *2012*, 637316.
(14)   Aiello, L. P.; Brucker, A. J.; Chang, S.; Cunningham, E. T., Jr.; D'Amico, D. J.; Flynn, H. W., Jr.; Grillone, L. R.; Hutcherson, S.; Liebmann, J. M.; O'Brien, T. P.; Scott, I. U.; Spaide, R. F.; Ta, C.; Trese, M. T. *Retina* **2004**, *24*, S3.
(15)   Roddríguez, R. *PhD thesis* 2006.
(16)   Hajduk, P. J.; Bures, M.; Praestgaard, J.; Fesik, S. W. *J Med Chem* 2000, *43*, 3443.
(17)   Fairbrother, W. J.; Christinger, H. W.; Cochran, A. G.; Fuh, G.; Keenan, C. J.; Quan, C.; Shriver, S. K.; Tom, J. Y.; Wells, J. A.; Cunningham, B. C. *Biochemistry* **1998**, *37*, 17754.
(18)   Pan, B.; Li, B.; Russell, S. J.; Tom, J. Y.; Cochran, A. G.; Fairbrother, W. J. *J Mol Biol* **2002**, *316*, 769.

2 Objectives

In the context of the present thesis we have addressed the following objectives:

1. To use both ligand- and protein-based NMR methods to study the binding of fragment libraries to the surface patch of VEGF involved in binding to its receptors. The VEGF/ VEGFR interaction can be considered as a case study for the general targeting of protein-protein interfaces via fragment screening.

2. To develop tools based on the combination of NMR and computation to address: i) the automatic setup of fragment mixtures; ii) the automatic analysis of NMR based screens; iii) the evolution of very weak binding fragments.

3. To explore the application of mRNA display for the discovery of novel peptidic VEGF ligands.

3. Results

3.1 NMR studies of protein-ligand interactions

Protein-ligand interactions are integral to diverse biological processes. They include the interaction of proteins with signaling molecules, such as neurotransmitters and hormones, or cofactors, as well as antigen recognition and enzyme-substrate interactions. In all of these processes, correct biological functioning of the protein requires that it specifically recognize a ligand at a particular binding area on its surface.

Deep knowledge of these processes and their underlying mechanisms is necessary not only for understanding these events at the molecular level, but also for being able to selectively modulate these interactions to provoke a desired biological response. This can be done by modifying natural compounds or by developing completely new compounds. Both cases offer a nearly unlimited pool of small organic molecules, peptides, carbohydrates or mixtures thereof. Whatever the potential of these compounds to interact with a given protein, recognition itself is steered by the structural orientation of the protein's functional groups. Thus, elucidation of these interactions greatly facilitates selection of appropriate functional groups in an appropriate framework.

Protein-ligand interactions can be studied with several tools, nearly all of which can provide information on binding strength and specificity. This information can be complemented with data acquired by isothermal titration calorimetry (ITC), mass spectrometry (MS), surface plasmon resonance (SPR) and nuclear magnetic resonance (NMR).

Isothermal titration calorimetry records the change in temperature of a protein solution upon titration with a ligand solution in an isolated chamber [1]. It enables determination of thermodynamic parameters, including the free energy (ΔG), enthalpy (ΔH) and entropy (ΔS) of the interaction, and the change in heat capacity (ΔCp).

In mass spectrometry, various techniques are used to ionize compounds or complexes and subsequently analyze their mass-to-charge ratios. Recent developments in MS have facilitated the study of protein-ligand interactions, allowing the detection and characterization of individual conformational states of protein complexes [2]. Owing to the high sensitivity of MS, only minute amounts of sample are needed. The study of hydrogen-deuterium exchange of protein backbone amide hydrogens can give information on the binding epitope of a ligand; however, this is most amenable to higher affinity ligands. Finally, MS is one of the few methods that enable study of complexes in gas phase. Comparison between binding energies in gas-phase and in solution may advance understanding of the forces behind protein-ligand interactions and, more precisely, help establish the role of solvation in molecular recognition at protein surfaces [3].

Surface plasmon resonance probes the interaction between an analyte in solution and a biomolecular recognition element immobilized on a sensor surface [4]. It enables direct determination of the binding kinetics parameters k_{on} and k_{off}, from which thermodynamic parameters can be quantified. If the protein is the immobilized binding partner, then only small amounts are necessary. The main drawback of SPR is that it requires immobilization of one of the binding partners, which may influence the protein-ligand interaction.

Nuclear magnetic resonance has evolved into a powerful tool for obtaining massive amounts of data on inter- and intra-molecular processes. Use of NMR to detect protein-ligand interactions is widely documented in the literature [5-7]. An advantage of NMR over other techniques is that it provides access to a broad set of experiments that have been optimized for various objectives: determination of affinity and specificity; identification of binding epitopes on the protein and on the ligand; characterization of structural rearrangements induced by binding; and turnover of substrates by enzymes. Furthermore, since the experiments are performed in solution, physiological or near physiological conditions are possible. Another advantage of NMR is that it is not limited to high-affinity-systems: it can be applied to study very weak interactions (*i.e.* m*M* range), for which other techniques are often unsuitable [8]. Moreover, for low affinity systems, NMR offers a relatively low incidence of false positive and false negatives compared to other analytical approaches. The main limitation of NMR is its low sensitivity. Also, compared to other techniques, NMR experiments are intrinsically low-throughput. Nevertheless, improved automation, and development of high sensitivity probes (*e.g.* cryoprobes), new pulse sequences, efficient isotopic labeling techniques, and more powerful magnets, have all contributed significantly to minimize these limitations. Nuclear magnetic resonance experiments for protein-ligand interactions fall into two main categories: either studying them form the perspective of the protein or from the perspective of the ligand. In the following section both approaches are overviewed and some typical experiments from each group are analyzed.

3.1.1 Protein observed experiments

Although in some cases monodimensional (1D) ^{1}H-NMR experiments have been used to characterize the protein-ligand binding by following the ^{1}H chemical shifts of specific residues in the protein, most experiments on protein observation entail bidimensional (2D) NMR. Conventional 1D-^{1}H spectra typically cannot resolve the individual proton signals of the protein. This limitation can be overcome by distributing the information along two dimensions and by employing *heteronuclear spectroscopy* (*i.e.* studying magnetically active nuclei other than protons that are present in proteins, such as ^{15}N and ^{13}C). Since the natural abundance of ^{15}N and ^{13}C (0.37% and 1.1%, respectively) is too low for NMR experiments, the protein to be studied must be isotopically labeled, usually, through expression

in *E. coli*. Several efficient labeling schemes are available, and choosing the right one can greatly simplify NMR studies of proteins.

The most widely used labeling method is *uniform labeling* with ^{15}N. For proteins expressed recombinantly in *E. coli* ^{15}N (in the form of an ammonium salt) is added to the expression media. This simple modification provides near quantitative isotopic labeling of the protein: all backbone amides as well the nitrogen containing side chains are labeled with this magnetically active nucleus. Heteronuclear ^{1}H-^{15}N correlation NMR experiments that allow direct observation of J-coupled ^{1}H to ^{15}N nuclei generate spectra containing at least one signal for each amino acid, except proline. Additional signals arise from amides in the side chains. When signal assignment is available, this strategy enables mapping of changes in the protein's backbone amides that are induced by binding of a ligand, and if the 3D structure of the protein is known, then the regions directly involved in the binding process can be easily identified.

Another common labeling scheme for studying protein-ligand interactions is *amino acid specific labeling*, in which the desired amino acid—or a suitable precursor—is added pre-labeled to the expression media and an auxotrophic bacterial strain is used. This is advantageous for large proteins (*i.e.* > 40 kDa), for which it provides far simpler spectra than those obtained with uniform labeling of the backbone. In this context a type of amino acid is selected which is well distributed throughout the protein sequence and which can serve as a probe for changes induced by the ligand-protein interaction. The authors of this review recently used this strategy to map changes induced by ligand binding to POP, an 80 kDa protein, using a ^{15}N-indole selective labeling scheme of the twelve Trp residues in the enzyme [9]. Selective labeling of one type of amino acid can also be attractive for small proteins. According to the Hot Spot theory proposed by Bogan *et al.*, specific amino acids are concentrated in regions of the protein that contribute to interactions with other proteins or ligands [10]. Therefore, one of these amino acids may serve as a site-specific probe. Even if the assignment is incomplete, this scheme can identify ligands that bind to a zone of interest.

For methyl-bearing side chains, ^{13}C labeling provides a very sensitive probe. Due to its mobility and the presence of three degenerated protons, the methyl group generates a high intensity signal in 2D heteronuclear ^{1}H-^{13}C correlation NMR experiments, while being extremely sensitive to environment changes. The advantages and applications of using selectively ^{13}C-labeled methyl groups in the NMR study of large biomolecules have been reviewed by Tugarinov [11].

The key experiment for the study of protein-ligand interactions in ^{15}N labeled target proteins is the ^{1}H-detected 2D-[^{15}N,^{1}H]-HSQC experiment [12,13]. For uniformly ^{15}N-labeled samples, at least one signal per amino acid is observed. The basic experiment comprises four main blocks (Figure 3.1.1-A)
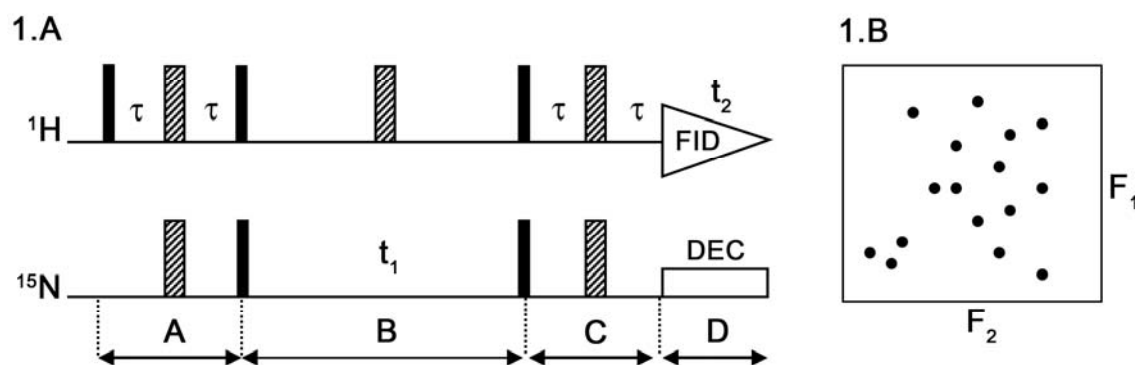
Figure 3.1.1 (A) Basic pulse sequence for the ^{1}H-^{15}N HSQC experiment. The narrow and wide bars depict 90° and 180° pulses, respectively. The delays (τ; equal to $1/[4^1J_{HN}]$) allow magnetization evolution to be transferred between coupled nuclei. ^{15}N magnetization evolves for t_1 and ^{1}H magnetization is directly detected during t_2. Double Fourier transformation along t_1 and t_2 generates a 2D correlation spectrum with frequencies F1 and F2, respectively, as shown in **(B)** Every signal in the spectrum corresponds to one NH group in the protein and gives information on the chemical shifts of an amide nitrogen (F1) and an amide proton (F2) that are directly coupled through the coupling constant $^1J_{HN}$.

Block A comprises an INEPT module [14], whose purpose is to transfer nuclear spin polarization between J-coupled nuclei—in this case, from the more sensitive one, ^{1}H, to the less sensitive one, ^{15}N. Since the scalar coupling constant is adjusted to the $^1J_{HN}$ value (*ca.* 90 to 95 Hz), only magnetization of amide protons is transferred to the adjacent ^{15}N nucleus. During block B a ^{15}N frequency labeling is achieved by incrementing the variable delay t_1, which leads to generation of the indirect dimension of the 2D spectrum (F1 frequency). Block C comprises a reverse INEPT module. Nuclear spin polarization is again transferred—this time, from ^{15}N to ^{1}H. This enables data acquisition in block D, in which ^{1}H magnetization is directly detected during t_2, which corresponds to the F2 frequency in the 2D spectrum (see Figure 3.1.1-B). Both excitation and direct detection of ^{1}H, the nucleus with the higher gyromagnetic ratio, provide a highly sensitive NMR experiment.

Since labile (*e.g.* amide) protons are observed in the experiment, protein NMR must be performed in H_2O, rather than in D_2O. Proton concentration in H_2O (*ca.* 100 M) is usually several orders of magnitude higher than that of the protein (mM range), which implies a wide dynamic range. Thus, the H_2O signal must be strongly attenuated in order to observe the protein protons at a sufficient signal-to-noise ratio in the NMR spectrum. Water suppression is thereby a critical requisite that must be experimentally optimized. Currently, most schemes that provide good water elimination [15] are based on using pulsed field gradients and proton selective pulses that enable manipulation of the H_2O magnetization independently of that of the protein.

Protein observed experiments for studying protein-ligand interactions are very simple: the chemical shifts of the protein signals change upon binding of the ligand. The resulting *chemical shift perturbation* (CSP) provides the basis for detecting binding. Moreover, if signal assignment is

available, the exact location of the interaction on the protein surface can be mapped. Although this approach was pioneered by several authors, including Gerhard Wagner, it is strongly associated with Stephen Fesik and his colleagues at Abbott Laboratories, who coined the term *SAR by NMR* [16] to describe the use of CSP for establishing structure-activity relationships (SAR) in drug discovery.

Interaction of the protein with a ligand affects not only the local magnetic environment of the backbone amides, but also the protein's dynamics. In principle, NMR is well suited for studying protein dynamics, although this approach is still in its infancy for protein-ligand interactions. Smrcka *et al.* studied the intensities of ^{15}N-Trp labeled G-protein βγ subunits in the presence and absence of ligands to gain insight into these subunits ability to interact with diverse molecular partners. They concluded that the wide range of signal intensities corresponding to different Trp residues is related to differences in local mobility, which is the underlying mechanism behind their molecular promiscuity. The experiments done in the presence of a ligand supported this idea, since the intensities of residues close to the ligand decreased upon binding [17].

In addition to binding affinity, binding kinetics are also decisive in CSP experiments. However, since the theory behind this is already covered in the literature [7,18], only a qualitative description of the phenomenon and its impact on CSP are provided here. Depending on the system being studied, the kinetic constants of the binding event can be much faster or much slower than the difference between the chemical shifts of the bound and free states. This leads to a range of behaviors in CSP experiments, whereby increasing amounts of a ligand are titrated into a protein sample and ligand-induced chemical shift changes are subsequently detected.

In the *fast exchange regime* (*see* **Note 1**) the exchange between the bound and free form is faster than the difference in chemical shifts. Only one set of protein signals is visible, and their positions typically shift according to the ratio between the bound and free species. Therefore, the chemical shifts move from the free form of the protein to the position of the bound state, which is reached once the protein sample has been completely saturated with ligand. If the same amount of ligand is used in each titration step, then the chemical shifts will change asymptotically. This can be fit to a mathematical model and used to calculate the affinity (K_D) of the interaction.

In the *slow exchange regime* the situation is reversed: exchange is slower than the difference in chemical shifts between the bound and free states. Therefore, the bound and free states give separate signals. In the course of the titration experiment the signal of the free protein declines while a new signal appears at the position of the bound state, which increases in intensity until becoming the only observable signal at the saturation point.

Fast and slow exchange regimes are not isolated extremes: they are linked by the *intermediate regime*, whereby the rate of exchange between the bound and free states is comparable to the difference in chemical shifts between these two states. Consequently, the behavior is more

complicated, as it entails a mixture of signal shifts, decreasing signals, and newly appearing signals. This results in very broad signals and non-Lorentzian line shapes, which makes analysis very difficult [19]. The equilibrium dissociation constant K_D can be used for quantification of exchange regimes. If a diffusion controlled on rate with $k_{on} \sim 10^8$ M^{-1}s^{-1} is assumed, then k_{off} can be estimated: ligands with K_D < 1-10 nM and $k_{off} \sim$ 0.1-1 s^{-1} will be in the slow regime; ligands with K_D > 10 µM and k_{off} > 10^3 s^{-1} will be in the fast regime; and ligands with values in between these will fall in the intermediate regime. However, these values are only valid if the association is indeed diffusion-controlled.

Chemical shift perturbation experiments are not always easy to interpret, chiefly due to the difficulty in distinguishing between short- and long-distance effects. Short-distance effects are perturbations resulting from the interaction of residues with the ligand. They delineate the binding zone of the ligand. Long-distance effects are perturbations caused by structural rearrangements of the protein under ligand binding. Although detection of long-distance effects may be of interest, they can give misleading information if the ligand interaction zone is unknown. Long distance effects markedly complicate the study of very flexible systems and weak ligands. To overcome this problem, Fesik *et al.* performed CSP studies of FKBP and of Bcl-$_{XL}$ using closely related ligands [20]. Although in both cases all ligands caused massive perturbations, the differences among the perturbations of these related ligands enabled identification of the binding site and the crude orientation of the ligands. More recently, Krishnamoorthy *et al.* addressed this issue by proposing a new way to analyze NMR CSP data in detail [21].

3.1.2 Ligand observed experiments

All ligand observation experiments are based on the difference in NMR parameters between the bound and free states of the ligand. The changes in nuclear Overhauser effects (NOEs) when ligands bind to receptor proteins are especially interesting [22]. Ligands with molecular weight lower than 1,000 u exhibit short correlation times (τ_c) and show only weak positive NOEs, very small negative NOEs, or no NOEs at all, depending on the magnetic field strength and the molecular weight. Proteins, due to their size, show large τ_c, large negative NOEs and highly efficient spin diffusion. Upon binding, the ligand forms a high molecular weight complex with the protein; consequently its properties change, especially its NOE behavior, with the appearance of strong negative NOEs, usually called *transferred NOEs* (trNOEs). The difference in the properties between the bound and free ligand is as large as the difference in molecular weight between the ligand and the protein-ligand complex. Since these differences have a direct impact on the observable NMR parameters of the ligand, several experiments can be used to detect and characterize the binding event. Most ligand observed methods are based on one of the following: assessment of changes in conventional NMR parameters

of the ligand (*e.g.* line widths, chemical shits, relaxation properties, and diffusion); or observation of intermolecular proton magnetization transfer from the protein to the free ligand (via the bound ligand), to distinguish between binding and non-binding ligand molecules.

There are myriad ligand observed experiments currently available, some of which are briefly introduced in the following section.

One of the first reported applications from the first category above entailed using ^{1}H NMR to observe the binding-induced chemical shift changes in certain signals of a ligand upon its interaction with a protein. However, because changes in chemical shifts are small compared to line width changes, experiments based on relaxation rate effects have been more extensively used. One such experiment is the Carr-Purcell-Meiboom-Gill (CMPG) filtered ^{1}H spectrum [23], in which, an R_2 relaxation filter comprising of a train of conveniently spaced 180º pulses is applied prior to data acquisition. Provided that the ligand remains bound to the protein long enough to adopt its relaxation behavior, this method, when adjusted properly, removes signals from the quickly relaxing protons of the bound ligand as well as those of the protein. This procedure is also useful since the degree of the attenuation in ligand signals can be used to rank the affinity of various ligands.

As mentioned above, the sign and size of the NOEs of a small molecule (*i.e.* ligand) will change when binding to a receptor protein. Transient NOE experiments (*see* **Note 2**) such as 2D NOESY [22] can be performed to observe transferred-NOEs to determine conformations of ligands bound to proteins [24]. During the mixing time of the NOESY the NOEs build up to a maximum value, and the difference in build-up rate among transferred-NOES and NOEs from the free ligand is the key point for ligand-binding detection: for binding ligands, trNOE rates range from 50 ms to 100 ms, whereas for non-binding ligands, larger values (200 to 1000 ms) are typical. However, this experiment is less sensitive than other experiments (*e.g.* STD). Nevertheless, its value lies in enabling structure determination of the bound conformation of the ligand in the complex, when intramolecular trNOEs are detected. The intermolecular trNOEs between a ligand and a protein can be used to establish the orientation of the bound ligand in the protein's binding pockets.

The transfer NOE effect can be considered as a precursor to experiments in the second category described above, which are currently very popular. Responses of magnetization transfer experiments are based on exchange-averaged parameters and are affected by many experimental parameters. Saturation Transfer Difference (STD) and Water-Ligand Observed via Gradient Spectroscopy (WaterLOGSY) are among the most important of these experiments. Case study 2 is based on the use of STD NMR; therefore, this experiment is described in more detail.

Saturation Transfer Difference was introduced in 1999 by Bernd Meyer in two seminal papers [25,26]. He described the experiment in studying the interaction of wheat germ agglutinin with saccharides, and reported its potential use for analyzing mixtures of putative ligands. Several other STD experiments

have since been reported, using a broad range of targets, including transmembrane receptors on whole cells.

An ideal sample for an STD experiment comprises a medium to high molecular weight protein plus a low molecular weight ligand in a deuterated buffer. The ligand is in high excess over the protein, to which it binds in fast exchange; these are common conditions for low affinity ligands. A conventional 1D ^{1}H NMR spectrum of this sample will show a combination of broad peaks, corresponding to the protein protons and narrow peaks, corresponding to the ligands protons. The low concentration of the protein and its short T_2 will generate low intensity signals distributed along the entire spectrum as follows: δ 10 to 6 ppm (amide and aromatic protons); δ 6 to 4 ppm (α-protons) and δ 4 to -1 ppm (aliphatic protons). Ligand signals vary strongly by ligand structure, but usually appear at $\delta > 0.8$. This leaves a high-field spectral region occupied exclusively by protein signals.

Saturation Transfer Difference is based on difference spectroscopy, so two sets of experiments, the *on-resonance* and the *off-resonance*, are acquired. In the on-resonance experiment a frequency-selective pulse is repetitively applied to the sample in the aforementioned range in which only protein signals are present (*e.g.* at -1 ppm) to saturate these protons, which are chiefly methyl groups of aliphatic side chains. The magnetic saturation will transfer to protons located in proximity, and then spread over the entire protein due to fast spin-diffusion and fast cross-relaxation mechanisms. This process is observed in the ^{1}H-spectrum as a nearly complete disappearance of protein signals. If a ligand present in the sample then binds to the protein, it will form part of this high molecular weight system, and consequently, will receive part of that magnetic saturation. Interestingly, the degree of saturation received by each ligand proton is not equal, but rather depends on their proximity to the protein. Therefore, this property can be used to determine the binding epitope of the ligand.

The on-resonance experiment must be compared to a reference (*i.e.* the off-resonance experiment), in which no magnetic saturation of the protein is performed, and therefore, no change in signal intensities is observed. Subtracting the on-resonance spectrum from the off-resonance spectrum provides the STD spectrum, in which distinguishing whether or not a ligand binds to the protein is easy, since only the signals of the binding molecule are visible. To minimize the appearance of artifacts in the resulting difference spectrum, the on- and off-resonance experiments must be completely comparable. Because of this, in the off-resonance experiment a train of frequency-selective pulses is applied to a spectral region lacking ligand and protein signals (*e.g.* 40 ppm). Furthermore, both experiments are acquired in an interleaving manner to reduce the impact of equipment instabilities. Figure 3.1.2 is a schematic of the STD pulse scheme. The core of the program comprises three main blocks:
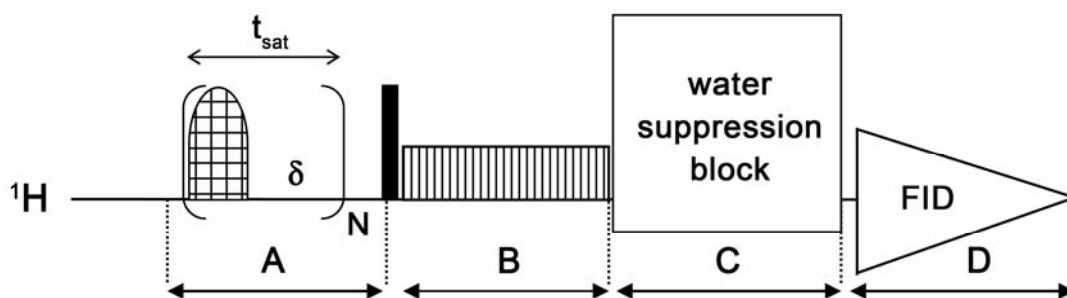
Figure 3.1.2: Basic pulse sequence for an STD experiment. Selective saturation via a train of N selective 90° pulses separated by a delay δ is performed during block A. Protein signals are suppressed during an R_2 relaxation filtering delay (block B). After a module for water suppression (block C), the ^{1}H signal is detected during the FID (block D).

During block A, a train of selective pulses is applied for a total saturation time (t_{sat}). During the on-resonance experiment the selective pulses are applied only to the protein signals at high-field, whereas in the off-resonance experiment the pulses are applied to a region far off-resonance from the protein and ligand signals. During block B, after a hard 90° pulse, a spin lock (R_2 relaxation filter) is applied to remove protein background signals. Block C comprises a water suppression module for samples containing a significant amount of H_2O (*i.e.* > 20%). This block can be omitted when working in D_2O or other organic solvents.

The widespread use of STD NMR stems from its many attractive features:

First, STD is amenable to high molecular weight therapeutic targets. In fact, the larger the target, the more favorable the conditions for the experiment. Magnetic saturation is easily achieved for larger targets, which promotes saturation transfer from the protein to the ligand better than for smaller targets. Saturation Transfer Difference works well for large receptors (> 30 kDa). For masses lower than 10 kDa, special attention must be paid because the R_2 relaxation rate may be insufficient for the intramolecular spreading of the saturation and the intermolecular transfer to the ligand. These cases may demand longer saturation pulse trains, or either addition of viscosity enhancing reagents or use of lower temperatures to slow molecular tumbling.

Secondly, STD experiments can be run with a minute amount of protein, only low µM concentrations are usually chosen. This is true because the ligand is present in molar excess (generally, 100-fold) over the protein. Assuming fast exchange, one molecule of protein can bind to a multitude of ligand molecules during the total saturation time t_{sat} (usually 1 to 3 s). Due to the small R_1 relaxation values for the free ligand state, free ligand molecules conserve the magnetic saturation received by the protein, which leads to a buildup of saturated ligands in the sample imprinted with the information of the binding event. This signal amplification is what makes STD more sensitive than other techniques.

Thirdly, the STD experiment is easy to implement. Optimization of the on-resonance frequency for each protein is important, such that only protein signals are selectively irradiated. Although the

ligand-to-protein ratio, saturation time, and length of the R_2 relaxation filter can all be optimized, in most cases, STD signals will already be observed with standard (default) parameters. Unfavorable kinetics of the ligand exchange may be improved by changing the temperature at which the experiment is performed [27].

Fourth, the *binding epitope* of a ligand (the specific portions of the ligand surface critical for molecular recognition) can be estimated from STD experiments [28] by exploiting the fact that STD signal intensities (I_{STD}) are not equal for the different protons in the ligand. The usual interpretation is that the larger the STD response, the closer the contact between the protein and the ligand. However, the magnitude of the STD signals not only depends on the proximity to the receptor, but also on longitudinal relaxation times (T_1) of the free ligand; thus, the STD response depends both on intermolecular cross-relaxation with the saturated receptor protons and on autorelaxation. Saturation Transfer Difference effects at long saturation times may be misinterpreted for protons in molecules having significantly different T_1 values. To determine the binding epitope without the bias of different relaxation times (T_1), the STD experiment must be performed at different saturation time (t_{Sat}) values [29]. Experimental data are fitted to the STD build-up curves for each proton (having a different T_1) to obtain the slope of the monoexponential equation (STD_{max}) and the saturation rate constant (k_{sat}):

$$STD_{Ampl.} = STD_{max} * \left[1 - \exp(-k_{Sat} * t_{Sat})\right]$$ [eq. 1]

whereby:

$$STD_{Ampl.} = \varepsilon * \eta = \varepsilon * \frac{I_{STD}}{I_0}$$ [eq. 2]

$STD_{Ampl.}$ corresponds to the *STD amplification factor* [28] and is a correction for total ligand concentration to the STD effect; I_{STD} is the intensity of an individual proton in the STD spectrum, and I_0, intensity of the same proton in the reference spectrum; ε is the ligand excess; η is the fraction of I_{STD} from I_0; STD_{max} is the maximal STD intensity achievable with long saturation times and corresponds to the STD intensity in the absence of T_1 bias.

Finally, STD experiments can be used to determine *dissociation constants* if a titration curve is recorded with varying ligand concentrations at the same saturation time [30]. To do this, the STD-amplification factors are first determined as described above, and then plotted against the ligand concentration. For one-site binding models, the curve can be fitted to the following equation:

$$STD_{Ampl.} = \frac{STD_{max} * [L]}{K_D + [L]}$$ [eq. 3]

whereby [L] is the concentration of the ligand; and K_D is the affinity constant of the ligand (relative to the protein).

The range in K_D has been estimated to be from 10^{-8} M to 10^{-3} M, assuming a diffusion-limited on-rate constant (*ca.* $10^8\,s^{-1}\,M^{-1}$). The intrinsic sensitivity of the STD experiment is limited by the efficiency of the signal amplification and the magnetization transfer. The signal amplification depends on the kinetics of the binding process, especially on the off-rate. For $K_D < 10^{-8}$ M, small off-rates cause a low turnover of ligands into saturated ligands; the binding is so tight that saturation transfer from the bound to the free ligand molecules is very inefficient. Additionally, when binding is very weak, the population of the ligand-protein complex is so low that it leads to either weak STD signals or no signals at all.

Despite being extremely utile and versatile, STD suffers from certain limitations. Among these is that the large excess of ligand relative to the protein may promote non-specific binding [7] once the specific binding site has been saturated. Another limitation is that protein saturation is suboptimal in the case of low proton density, local proton deficiency, or molecular motion which compromises the intramolecular ^{1}H-^{1}H dipole interaction network. In such cases, WaterLOGSY may be a more effective experiment than STD [31].

The main difference between STD and WaterLOGSY is the way in which the system receives magnetic saturation. Whereas STD NMR uses direct saturation of the protein, WaterLOGSY applies indirect saturation of the protein, namely, by selective saturation of the bulk water protons (H_2O). Therefore, the transfer magnetization flows from water to protein to ligand. Technically, there are several options to achieve the selective bulk water saturation. Dalvit *et al.* use the selective inversion of the water resonance via the e-PHOGSY scheme [32]. The transfer of magnetization from the water to the protein-bound ligand occurs via labile receptor protons (NH and OH protein protons) situated in the ligand-binding site as well as via remote labile protons in the protein, through spin diffusion. Additionally, direct proton-proton cross-relaxation between the bound ligand and long-lived water molecules within the binding pocket is an effective pathway in the magnetization transfer process. Differential cross-relaxation properties of binding and non-binding molecules with water allow distinguishing between binding and non-binding ligands. Whereas binding molecules interact with the proton spins of inverted water via dipolar interactions, which lead to negative cross-relaxation rates, non-binding molecules yield positive cross-relaxation rates. The result is that signals of non-binding molecules show opposite sign to, and are usually weaker than, the resonances of binding ligands.

3.1.3 Ligand vs. protein observed experiments

Ligand based and protein based approaches have distinct advantages and disadvantages. The former yield information about the strength of the interaction, the binding epitope and the conformation of

the ligand. They can be used to simultaneously screen several compounds for their ability to bind to a protein of interest. Ligand-based experiments have simple requirements. First, they do not require isotopically labeled protein. Secondly, there is no upper limit for protein size, but the difference between protein and ligand has to be substantial enough to result in differential relaxation behaviors. Contrariwise, protein observation experiments are currently only feasible for proteins weighing *ca.* 40 kDa or less. Moreover, they demand considerable amounts of isotopically labeled protein, which must be stable at high concentrations for long periods of time. When signal assignment is available, protein-based experiments may provide more information. Most importantly, they can be used to identify one or several binding sites of the ligand on the protein and indicate zones of structural rearrangement. Furthermore, in these experiments, formation of ligand aggregates cannot be misinterpreted as an interaction. However, the data in protein observation experiments are easiest to interpret when the binding site has been saturated. In the case of low affinity ligands, this point may be beyond the limit of solubility. In conclusion, there are cases for which one approach is better suited than the other. Nevertheless, the full power of NMR to characterize protein-ligand interactions can only be exploited if both approaches are combined. Several examples from academic and industrial drug discovery projects are testament to the great success of a combined approach [7,8,33,34].

3.1.4 Materials

Protein based study on the binding of VEGF to the peptidic ligand P-7i

1. Purified samples of uniformly[15]N-labeled VEGF: 160 μL at 100 μ*M* in 25 m*M* phosphate buffer, pH 7.0 (*see* **Note 3**), 50 m*M* NaCl, 90% H_2O, 10 % D_2O in a 3 mm NMR tube (*see* **Note 4**). VEGF is obtained by recombinant expression as previously described [35].

2. Peptide P-7i:Prepared by using standard solid phase peptide synthesis [3].

3. Bruker Digital Avance 600 MHz spectrometer equipped with a cryoprobe (*see* **Note 5**).

4. Data processing and analysis programs: TopSpin [36], Cara [37] and Origin [38], and the results are visualized using the program MOE [39].

Ligand based study on binding of POP to the ligand baicalin

1. Prolyl oligopeptidase (POP):160 μL of 100 μ*M* POP in 20 m*M* phosphate buffer, pH 7.0 (*see* **Note 3**) in 100 % D_2O in a 3 mm NMR tube (*see* **Note 4**).

2. Baicalin: 160 μL of 500 μM baicalin in 20 m*M* phosphate buffer, pH 7.0 in 100 % D_2O in a 3 mm NMR tube.

3. 160 μL of POP (10 μM) and baicalin (500 μM) in 20 m*M* phosphate buffer, pH 7.0 in 100 % D_2O in a 3 mm NMR tube.

4. 160 µL of POP (20 µM) and baicalin (180 µM) in 20 mM phosphate buffer, pH 7.0 in 100 % D_2O in a 3 mm NMR tube.

5. Bruker Digital Avance 600 MHz NMR spectrometer equipped with a cryoprobe (*see* **Note 5**).

6. Data processing and analysis programs TopSpin and Origin.

3.1.5 Methods

Protein based study on the binding of VEGF to the peptidic ligand P-7i

The following case study describes NMR monitoring of the interaction between vascular endothelial growth factor (VEGF) and the peptidic ligand P-7i, presented here as a representative example of a CSP experiment [3]. The 23 kDa VEGF[11-109] construct used is a truncated version of VEGF[121] which exhibits excellent solubility and stability. Moreover, it is readily labeled with ^{15}N and is a symmetric homodimer, making it highly suited to protein based NMR experiments. Additionally, the signal assignment for this construct is available [35]. P-7i is a nineteen amino acid-long analog of v107 that was discovered by phage display [40]. It differs from v107 by a single mutation: the Ile-7 is D rather than L, which translates to a reduced affinity for VEGF (252 µM) compared to the wild type (1.0 µM). Although P-7i is relatively large (MW = *ca.* 2.3 kDa) for NMR studies of protein-ligand binding, it was selected as an example because it exhibits fast and intermediate exchange behavior and is amenable to CSP studies.

Sample preparation

1. Starting from a stock solution of P-7i in water, prepare six aliquots (two at each of three concentrations) with a total amount of ligand corresponding to 50, 100 or 200 µM in a volume of 160 µL (19, 37 or 75 µg, respectively). Freeze the aliquots in 1.5 mL Eppendorf tubes and lyophilize (Figure 3.1.5-1).

	Ref.	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6
Protein conc.	100 µM	100 µM	100 µM	100 µM	100 µM	100 µM	100 µM
Ligand conc.	0 µM	50 µM	100 µM	200 µM	300 µM	500 µM	700 µM

Ligand increment	50 µM	50 µM	100 µM	100 µM	200 µM	200 µM
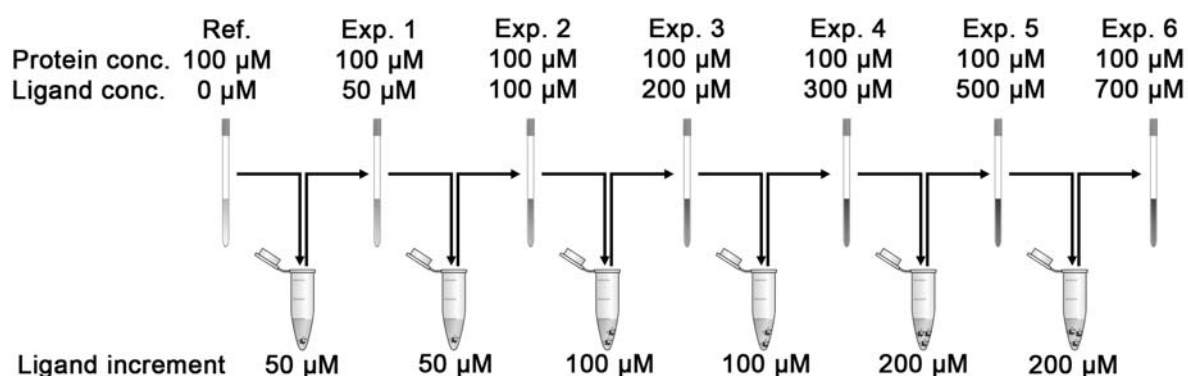
Figure 3.1.5-1: Stepwise addition of P-7i to VEGF: overview of lyophilized ligand aliquots, the respective concentration increments and the total concentration over the course of titration. After the reference experiment on the sample containing only protein, the ligand concentration was increased stepwise by transferring the protein sample to Eppendorf tubes with the denoted amounts of lyophilized P-7i prior to acquisition of the next NMR spectrum. A total of seven spectra were acquired by repeating this procedure: one reference spectrum, plus one spectrum at each of the six titration points.

Recording of CSP NMR spectra

1. Equilibrate the VEGF sample inside the NMR spectrometer for 15 minutes at 318 K before starting the NMR spectra acquisition. This temperature is chosen because the available signal assignment was performed at this temperature [35]. Execute standard NMR procedures: tune the probe, shim the magnetic field, calibrate the length of the 90º pulses for ^{1}H and ^{15}N, and optimize the water suppression.

2. Record a ^{1}H-^{15}N HSQC spectrum using the FAST-HSQC experiment [41], which uses pulsed field gradients and a WATERGATE [42] module to efficiently suppress the water signal. For this, 2048 x 256 complex points with a total of eight transients per increment are used. The total experiment time is *ca.* 40 minutes.

3. Once the spectrum is obtained, remove the sample from the spectrometer and transfer it to the first Eppendorf tube containing the lyophilized ligand (*see* Figure 3.1.5-1). Add 0.5% of DMSO-d$_6$ to ensure ligand solubility (*see* **Note 6**) vortex the sample and centrifuge (5 krcf, 1 min, rt), and transfer all of the liquid to the previously used NMR tube. Introduce the resulting sample, containing the protein and the ligand at the first titration concentration, into the spectrometer. Equilibrate the sample at 318 K for 15 minutes, re-shim the magnetic field and record a new spectrum using the same conditions described in step 2.

4. Repeat step 3 until spectra are recorded for each titration point. By dissolving the lyophilized ligand in the sample incrementally, the concentration of P-7i increases stepwise over the course of the titration from 0 µM to 50 µM, 100 µM, 200 µM, 300 µM, 500 µM and finally, 700 µM (Figure 3.1.5-1).

Data analysis

1. To process the HSQC spectra, increase the number of points in the indirect dimension (F1) from 256 to 512 by linear prediction and then zero fill to 1024 points to yield a 2048 x 1024 matrix. Adjust the phase correction manually and apply a squared sine weighting function in both dimensions. Process all the spectra acquired in the titration experiment identically using Topspin 2.0. Figure 3.1.5-2 shows the seven superimposed spectra acquired from the titration of VEGF with P-7i.
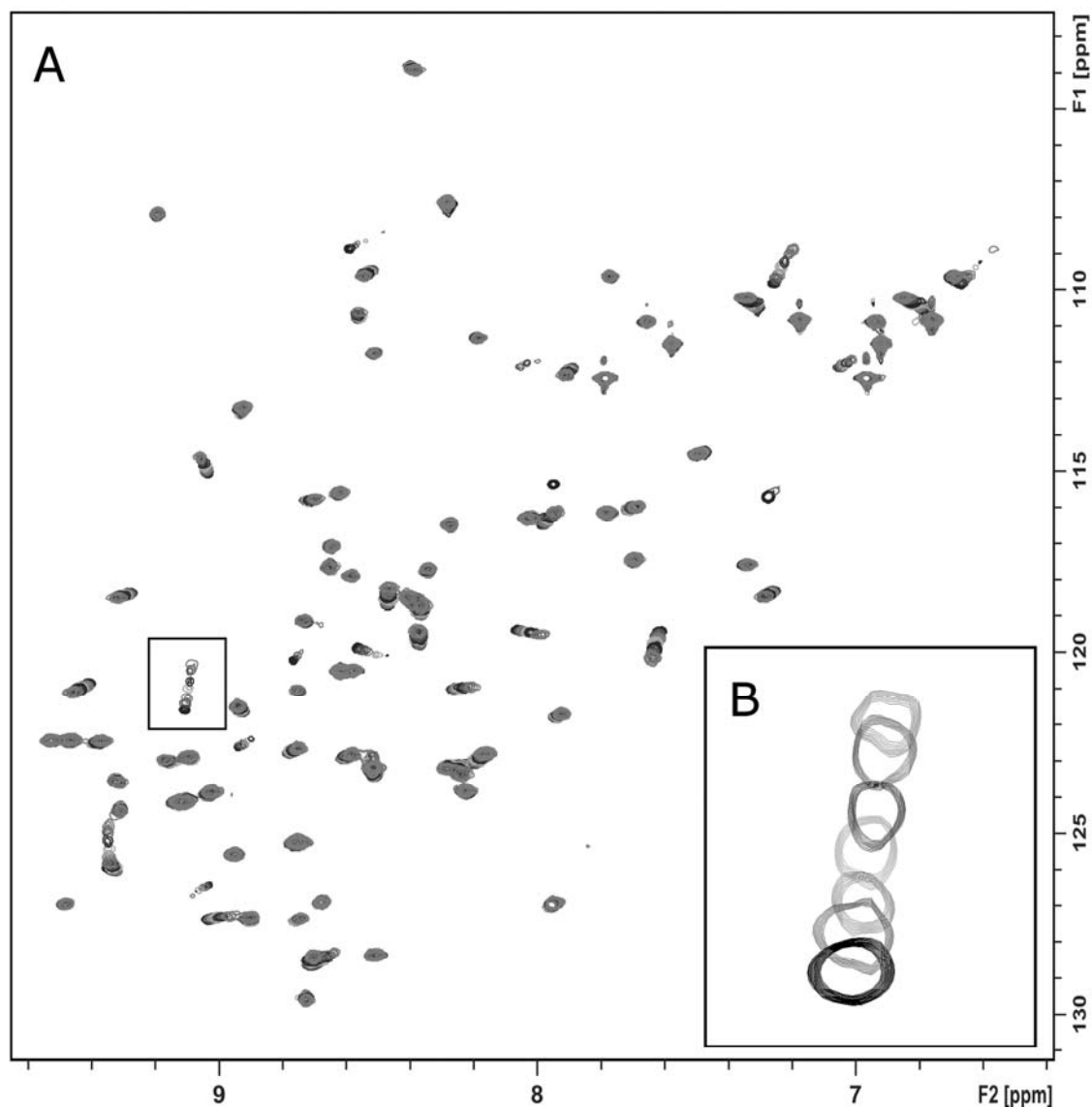
Figure 3.1.5-2: (A) Seven superimposed ^{1}H-^{15}N HSQC spectra of a 100 μM sample of uniformly-^{15}N labeled VEGF$_{11-109}$ at 318 K titrated with 0 to 700 μM of P-7i (600 MHz with cryoprobe). (B) Zoom of Lys48 shifts. (Reproduced from [3] with kind permission from Wiley)

2. Determination of CSP requires peak picking and subsequent assignment of the peaks to the corresponding residues. Use the program Cara to do this for the first and last spectra of the titration (*see* **Note 7**).

3. Extract the relevant data for mapping the binding site by calculating the distance between the position of the reference peak in the spectrum of the protein without ligand and the peak position in the spectrum of the highest ligand concentration. Calculate the distance between two peaks as the difference between the average chemical shift $\Delta\delta_{NH}$ for each peak computed from proton and nitrogen chemical shits (δ_{H} and δ_{N}, respectively), according to the following formula (*see* **Note 8**):

$$\Delta \delta_{NH} = \sqrt{\left(\Delta \delta_H^2 + \left(\frac{\Delta \delta_N}{5} \right)^2 \right)}$$ **[eq. 4]**

The majority of the peaks exhibit fast exchange behavior; however, residues 17, 21, 26, 64 and 104 exhibit intermediate exchange. Signal broadening and a rapid decrease in signal intensity are observed for these residues; however, despite the broad signals, assignment is still possible. Residue 21, located in the ligand-binding zone, exhibits slow exchange behavior. The signal disappears completely after the second titration point. Chemical shift perturbation is not feasible for this behavior (*see* **Note 9**). To identify the binding site, consider as significant only changes greater than the sum of the mean shift and the standard derivation. Figure 3.1.5-3 shows the calculated changes for each residue. The residues with significant changes (10 of 83) are mapped to the 3D-structure of VEGF and depicted in red (Figure 3.1.5-4) using the software MOE.
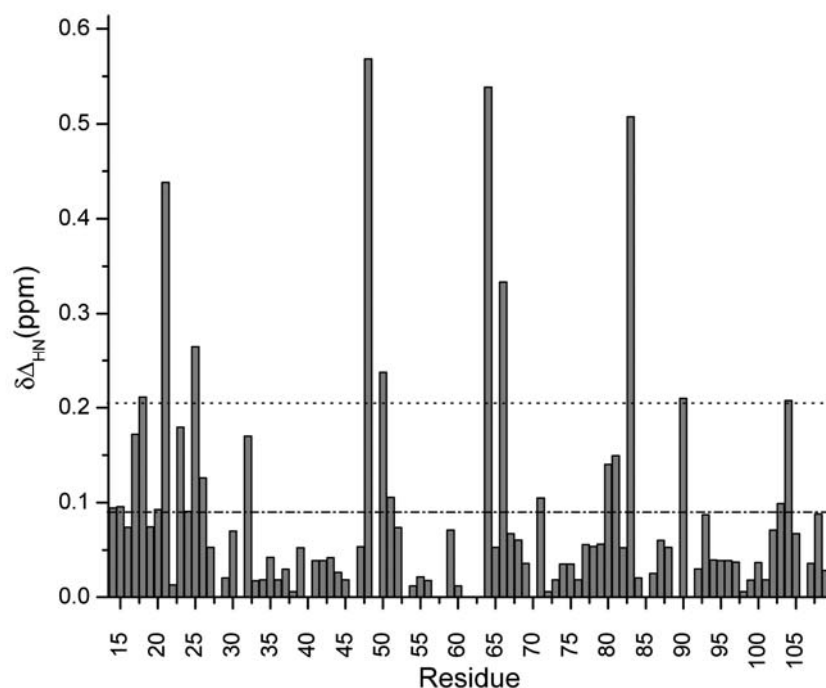
Figure 3.1.5-3: Histogram of the P-7i induced CSP of the backbone amides $\Delta \delta$ for every residue of VEGF$_{11-109}$ observed in the ^{1}H-^{15}N HSQC experiment. The histogram was calculated for the shifts between the reference spectrum (VEGF alone) and the spectrum of the sample with the highest ligand concentration (VEGF plus 700 µM of P-7i). The lower (dashed) horizontal line represents the mean shift, and the upper (dotted) horizontal line, the cutoff for significant changes (mean shift plus one standard derivation).

Figure 3.1.5-4. Surface representation of the homodimer VEGF$_{11-109}$ (PDB:2VPF). Residues encoded in red exhibit significant CSP, thereby indicating the binding zone for the ligand P-7i. (Residues encoded in black show no observable signals)
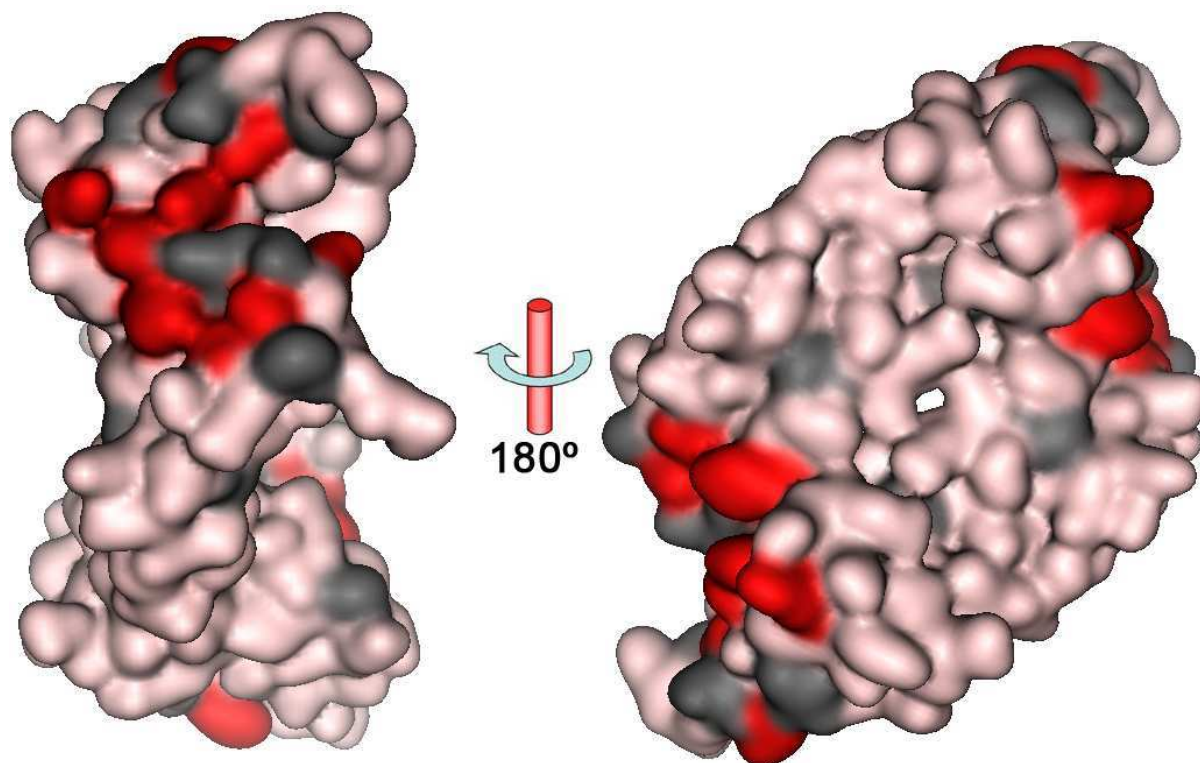
4. Calculate the binding affinity by plotting $\Delta\delta_{NH}$ against the ligand concentration, which requires that $\Delta\delta_{NH}$ is calculated for each of the seven titration points. Although a K_D can be calculated for each residue, this exercise is only performed for some of the residues with a strong shift: those that exhibit the best signal-to-noise ratios and are free from signal overlap. Thus, a K_D is calculated for residues 25, 48, 50 and 66, assuming a model of two independent identical binding sites [3]:

$$\Delta\delta_{NH} = F * \frac{[L_0]-[L]}{[P_0]} = F * \frac{K_D + [L_0] + 2[P_0] - \sqrt{(K_D + [L_0])^2 + 4[P_0](K_D - [L_0] + [P_0])}}{2}$$

[eq. 5]

whereby [P0] is the total protein concentration; [L0], the total ligand concentration; [L], the concentration of unbound ligand in solution; F, a scaling factor; and KD, the affinity to be calculated. The calculated average KD for the VEGF – P-7i system is 252 µM (Figure 3.1.5-5). The fact that the R^2 values are all greater than 0.99 validates this model of the system.
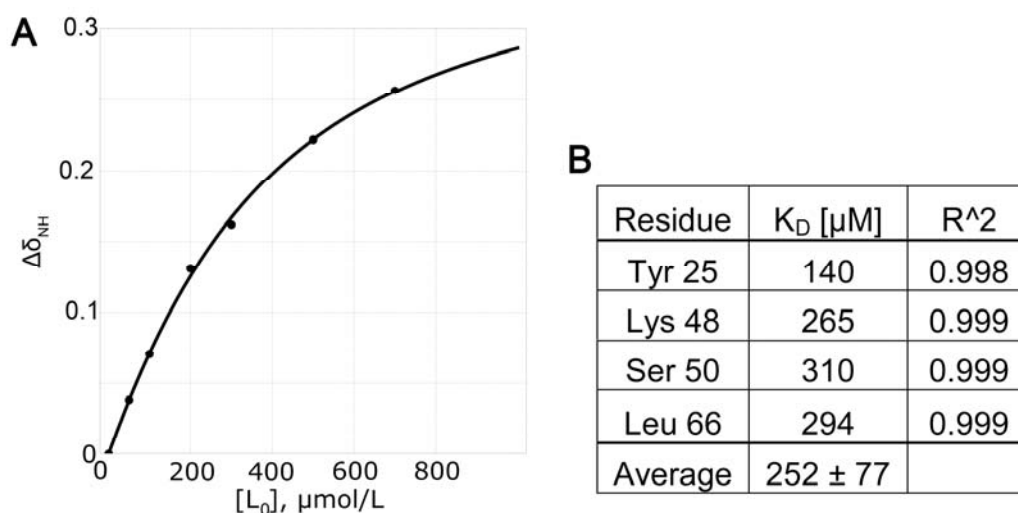
Figure 3.1.5-5. (A) Relative amount of bound ligand plotted against the total ligand concentration and fitted to a model of two independent identical binding sites. Analysis is based on Lys48 peak displacement in the spectra of the VEGF-(**P-7i**) complex. (B) Results of fitting for the four selected residues. (Reproduced from [3] with kind permission from Wiley)

Ligand based study on binding of POP to the ligand baicalin

The following case study was chosen to give a step-by-step explanation of an STD experiment. The example is based on work performed in the authors' laboratory to characterize the interaction between the protease prolyl oligopeptidase (POP) and the flavonoid baicalin [43]. These experiments were designed to confirm baicalin as a ligand of POP and to obtain structure-activity information on POP-baicalin binding, especially on the influence of the baicalin sugar moiety. This entailed recording of a saturation buildup curve.

Based on its molecular weight (80 kDa), POP is not appropriate for simple protein based experiments; however, it is well suited for ligand based experiments. Likewise, baicalin (MW 446 Da), as a relatively small to medium-sized ligand, is ideal for ligand based experiments. Concerning its affinity properties, baicalin is a weak ligand of POP, having an IC_{50} value of 12 µM [43] (*see* **Note 10**).

Optimization of STD parameters and confirmation of baicalin as a POP ligand

1. Equilibrate the sample containing only POP in the NMR spectrometer to 308 K (*see* Note 11) for 15 minutes. Follow the standard procedure: tune the probe, shim the magnetic field, calibrate the length of the 90º pulse and optimize the water suppression (*see* Note 12).

2. To optimize the protein saturation, record a ^{1}H-spectrum of POP to identify promising signals in the aliphatic region. The aliphatic region of POP shows a signal at *ca.* 0.9 ppm that decreases in intensity in the up-field direction without occurrence of any new maxima.

Therefore, the closer to this value the protein is irradiated, the greater the saturation (*see* Note 13).

3. To determine if protein irradiation is affecting any of the ligand's signals, acquire several STD spectra of a sample of baicalin alone, using values of 0, -1 and -2 ppm for the on-resonance frequency. No STD signals are observed in any case. This negative control experiment confirms that no direct irradiation of baicalin or baicalin-aggregates occurres. For the subsequent STD experiments, select an on-resonance frequency of 0 ppm and an off-resonance frequency of 80 ppm (*see* Note 14).

4. Optimize the total saturation time with a sample containing POP (10 µM) and baicalin (500 µM)(*see* Note 15). Acquire STD spectra with different t_{Sat} values (from 1 to 4 s) using the same total experimental time. Superimpose all on-resonance spectra and select the spectrum with the best signal-to-noise ratio as the optimum one (in this case, 2 seconds).

5. Optimize the spin lock filter by testing different mixing times (from 20 to 70 ms). Under optimal conditions the protein signals are completely suppressed, thereby reducing the background noise in the STD experiment, whereas ligand signals are not affected. In this case this is achieved with a spin lock length of 30 ms.

6. Acquire the final STD spectrum with the aforementioned optimized conditions and 2k scans. Process the NMR data by multiplication with an exponential line-broadening function of 0.5 Hz prior to Fourier transformation. The resulting spectrum exhibits clear STD signals (Figure 3.1.5-6) for baicalin, thereby confirming that it binds to POP.
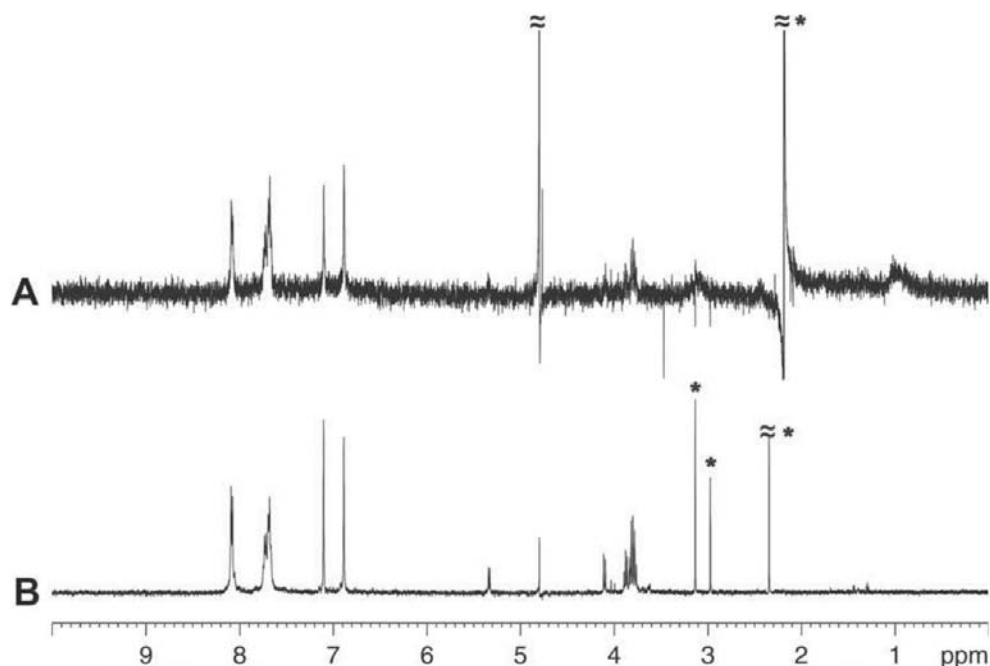

Figure 3.1.5-6: (A) ^{1}H STD spectrum of Baicalin (500 µM) in the presence of POP (10 µM) recorded at 600 MHz and 308 K. The protein signals were suppressed by applying a spin lock filter. (B) [1]H-reference spectrum of baicalin. *Signals arising from sample impurities. (Reproduced from [43] with kind permission from Elsevier)

Identifying the binding epitope on baicalin

A sample containing 20 µM POP and 180 µM baicalin is used to identify the binding epitope (see Note 16). Experiments are performed at 308 K.

1. Obtain data for the saturation buildup curve using the previously optimized parameters (on-resonance irradiation: 0 ppm; 1k scans; 308 K). Record a total of four experiments, using t_{Sat} values of 0.5, 1.0, 1.5 and 2.0 seconds.

2. For each proton, determine the values for $STD_{Ampl.}$ for the different saturation times (Figure 3.1.5-7). Superimpose the STD and off- resonance spectra for each saturation time, measure the difference between the STD and off-resonance signal for a specific proton (as a percentage), and multiply this value by the ligand excess [eq. 2], which was equal to nine (see Note 17).

A

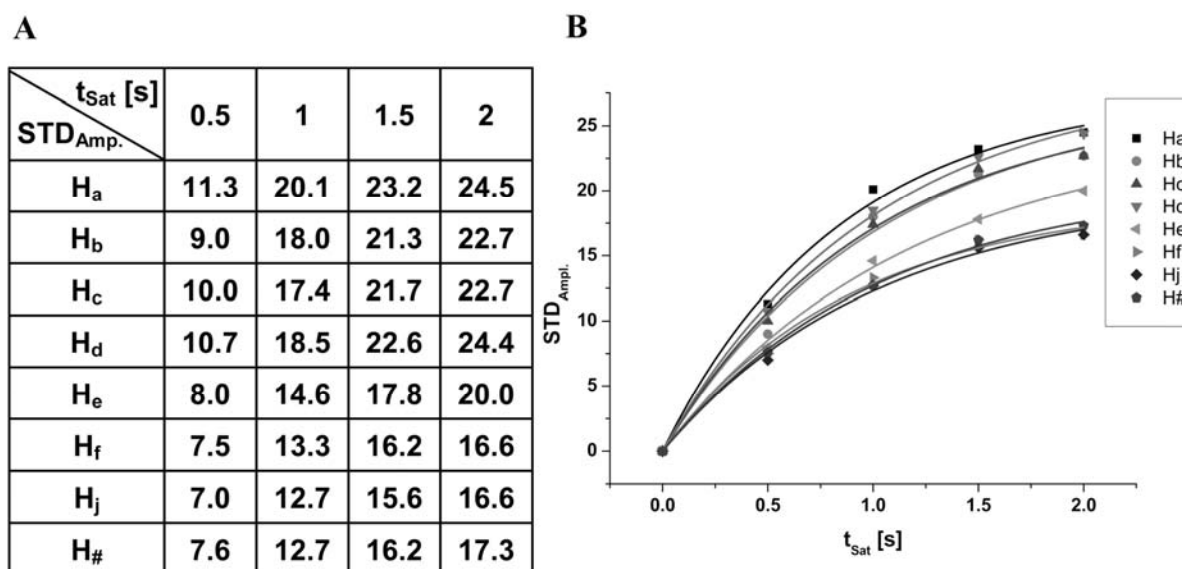| t_{Sat} [s] / $STD_{Amp.}$	0.5	1	1.5	2
H_a	11.3	20.1	23.2	24.5
H_b	9.0	18.0	21.3	22.7
H_c	10.0	17.4	21.7	22.7
H_d	10.7	18.5	22.6	24.4
H_e	8.0	14.6	17.8	20.0
H_f	7.5	13.3	16.2	16.6
H_j	7.0	12.7	15.6	16.6
$H_\#$	7.6	12.7	16.2	17.3

B

Figure 3.1.5-7: (A) Saturation Transfer Difference amplification for individual protons at different saturation times. The data were acquired with a nine-fold excess of baicalin over POP (20 µM). (B) Build-up curves obtained for individual protons in baicalin. The key to the nomenclature for the individual protons is provided in Figure 3.1.5-8. (Adjusted from [43] with kind permission from Elsevier)

3. Determine the binding epitope by fitting the values for $STD_{Ampl.}$ against the saturation times for each proton, according to **Equation 1**. Calculate the initial slope v_0 by multiplying STD_{max} by k_{Sat}. The saturation build-up curves of the protons of baicalin shows different initial slopes. The values were adjusted by setting the proton with the highest initial slope to 100% (Figure 3.1.5-8). These data indicate that the protons in the γ–chromenone and the phenyl ring are in close contact with the protein, whereas those in the saccharide moiety contribute less to binding.

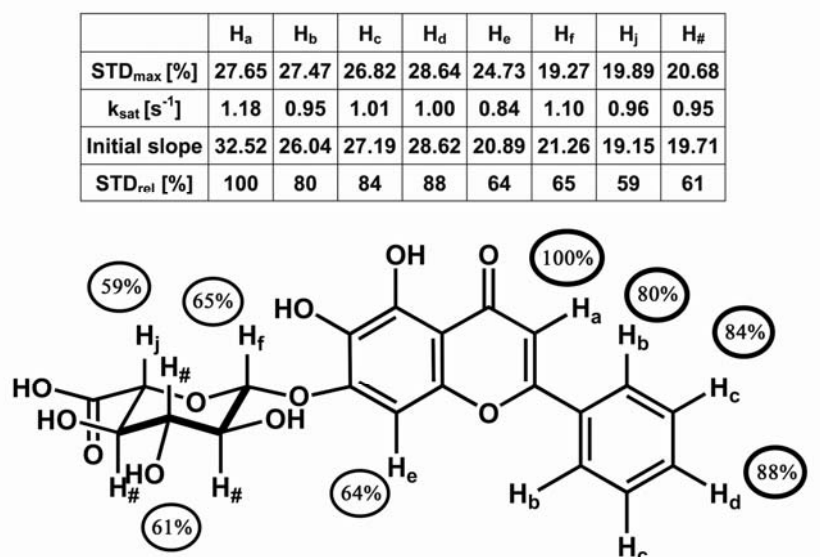|	H_a	H_b	H_c	H_d	H_e	H_f	H_j	$H_\#$
STD_{max} [%]	27.65	27.47	26.82	28.64	24.73	19.27	19.89	20.68
k_{sat} [s^{-1}]	1.18	0.95	1.01	1.00	0.84	1.10	0.96	0.95
Initial slope	32.52	26.04	27.19	28.62	20.89	21.26	19.15	19.71
STD_{rel} [%]	100	80	84	88	64	65	59	61

Figure 3.1.5-8: The STD data were fitted to a monoexponential equation, from which the STD_{max} and the saturation rate constant (k_{sat}) were obtained. The initial slope directly correlates to the proximity of the corresponding proton to the protein and is the product of STD_{max} and k_{sat}. The relative STDs were calculated by setting the proton with the greatest STD effect to 100%. (Adjusted from [43] with kind permission from Elsevier)

3.1.6 Notes

1. Exchange between the free (L) and bound (PL) states of a ligand is considered in this context, assuming that the binding follows a bimolecular association reaction with second-order kinetics: $P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL$; $K_D = \dfrac{[P][L]}{[PL]} = \dfrac{k_{off}}{k_{on}}$
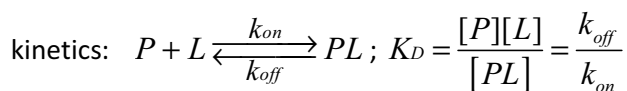
2. In transient NOE experiments a non-equilibrium state is generated via high-frequency pulses, and in a subsequent mixing period, returns to equilibrium by relaxation.

3. A buffer of the desired pH is created by mixing 25 mM solutions of NaH_2PO_4 and Na_2HPO_4 in deuterated water until the final pH is reached.

4. The most widely used NMR tubes are 3 and 5 mm standard tubes and 5 mm Shigemi tubes, which require sample volumes of 160 µL, 600-700 µL and 300 µL, respectively. Voehler *et al.* analyzed the influence of tube type on the sensitivity of HSQC experiments [44]. In general, 5 mm tubes are recommended for cases of abundant, poorly-soluble protein; 3 mm standard or 5 mm Shigemi tubes when the sample is limited; and 3 mm standard tubes for high salt concentrations. Moreover, for titration experiments 3 mm tubes are easier to manipulate than Shigemi tubes. Higher protein concentrations are generally preferable, as they enable shorter acquisition times. One exception to this is the case of a low affinity, poorly-soluble ligand, for which a major excess of ligand must be employed to saturate the binding site, which in turn is only possible at lower protein concentrations.

5. In a cryoprobe the coil and the preamplifier are cooled to decrease thermal noise. The signal-to-noise ratio can be increased by three- to four-fold by using a cryoprobe instead of a conventional probe operating at room temperature. The authors have obtained a signal-to-noise ratio of *ca.* 7000:1 for ^{1}H using a cryoprobe and a standard sample of 0.1% ethylbenzene in CDCl$_{3}$.

6. Ligands soluble at high concentrations (*ca.* 100 mM) in a vehicle such as DMSO-d$_{6}$ can usually be added directly to the sample, without resulting in any significant dilution; however, if this does not apply, then lyophilized aliquots of ligand can be used, as in Case Study 1. Vehicle may be added simply to ensure the solubility of the ligand; however then a control experiment is necessary to evaluate any changes induced by it. If lyophilized aliquots of ligand are used, then a vehicle can also be added to the reference experiment (*i.e.* protein alone) to eliminate any vehicle-induced changes during the course of the titration. Finally, data quality may be improved if the ligand concentration in the sample is controlled. This is easily achieved by adding a reference compound at a known concentration to the sample, such as trimethylsilyl propionate-d$_{4}$ (TSP), and then comparing the signal intensities of the ligand and of the reference in a ^{1}H-spectrum.

7. The authors found the freeware program Cara (available at www.nmr.ch) to be utile and straightforward; nevertheless, other assignment tools (*e.g.* NMRview and Sparky) are equally suitable.

8. This formula is one of the most commonly used ones for calculating the distance between two peaks in the two dimensional plane. Other approaches and their impact on CSP mapping have been reviewed by Schumann *et al.*[45].

9. For the wild type peptide v107, the binding site could not be mapped by observing CSP. This is because the binding kinetics are in the slow exchange regime. All residues that are directly involved in ligand binding or are in close proximity do not shift, but simply decrease in intensity as a new signal appears. In this case an alternative mapping of the binding site for slow exchanging ligands was feasible based on the changes of signal intensities induced by the binding of the ligand.

10. As mentioned in **Section 1.2**, STD NMR enables calculation of K$_{D}$. However, since the IC$_{50}$ of baicalin for POP had previously been determined, this calculation was not performed. Information on the calculation of binding strengths can be found in references [5] and [30].

11. Temperature is another parameter that can be optimized in an STD experiment, as it affects the binding kinetics of the system, the efficiency of protein saturation, and the relaxation rates. In STD signal intensities can be increased by changing the temperature [27]; however, once temperature has been optimized, all other parameters must also be adjusted

12. Before performing the STD experiments, a 1D ^{1}H spectrum with a water suppression module was employed to optimize water suppression. The residual H$_2$O signal was suppressed with an excitation sculpting block [46], which uses a double gradient spin-echo to defocus the H$_2$O resonance. Squared pulses of 2 ms length were used as 180º pulses. Optimized parameters for water suppression in this experiment were then used for the STD experiments.

13. This sample can be also used to verify if the saturation is well spread over the whole protein. This requires that an STD spectrum without spin lock filter is acquired. A decrease in intensity of all protein signals indicates a good distribution, which is the case for most proteins.

14. Some proteins exhibit several up-field shifted signals (*ca.* 1 ppm to -1 ppm) that differ in intensities. In these cases the on-resonance frequency can be optimized by recording a set of STD spectra of a sample containing only protein, without the spinlock filter but using different saturation frequencies. The saturation frequency that provides the strongest decrease between on- and off-resonance spectra is deemed the most favorable (upon confirmation that no direct saturation of ligand signals occurs). Commonly used values for the on-resonance irradiation lie between 0 and -1 ppm, and for the off-resonance, either 40 or 80 ppm.

15. The range of ligand excess in STD-NMR can be very wide. A good starting point is to use a 50-fold molar excess of ligand at a protein concentration of 10 μ*M*. However, the optimal ratio between protein and ligand depends on the system's kinetics. The faster the exchange, the better the signal can be amplified which makes higher protein-to-ligand ratios (100-fold and higher) useful and will ultimately yield stronger STD effects. Independent of the choice of the ratio, the ligand concentration should always be in a range in which aggregation can be excluded.

16. To map a binding epitope, lower ligand-to-protein ratios are typically used. The objective is not to reach the maximal STD amplification, but rather to cover the range of STD$_{Ampl.}$ values in order to calculate the build-up curve. A smaller excess of ligand leads to more pronounced differences between I$_{STD}$ and I$_0$ for the different saturations times than does a larger excess.

17. In this example the STD amplification factor STD$_{Ampl.}$ was used, rather than η, the fraction of I$_{STD}$ from I$_0$. The difference between the two is that STD$_{Ampl.}$ accounts for the ligand excess (**eq. 2**). Since in this example the ligand excess is equal for all saturation times, then it turns into a scaling factor between STD$_{Ampl.}$ and η, which is not required for calculating the binding epitope. STD$_{Ampl.}$ is more commonly used because it enables calculation of K$_D$.

(1)   Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L. N. *Anal Biochem* **1989**, *179*, 131.
(2)   Baldwin, M. A. *Methods Enzymol* **2005**, *402*, 3.

(3)   Dyachenko, A.; Goldflam, M.; Vilaseca, M.; Giralt, E. *Peptide Science* **2010**.
(4)   Englebienne, P.; Hoonacker, A. V.; Verhas, M. *Spectroscopy* **2003**, *17*, 255
(5)   Fielding, L. *Curr Top Med Chem* **2003**, *3*, 39.
(6)   Carlomagno, T. *Annu Rev Biophys Biomol Struct* **2005**, *34*, 245.
(7)   Lepre, C. A.; Moore, J. M.; Peng, J. W. *Chem Rev* **2004**, *104*, 3641.
(8)   Dalvit, C. *Drug Discov Today* **2009**, *14*, 1051.
(9)   Tarrago, T.; Claasen, B.; Kichik, N.; Rodriguez-Mias, R. A.; Gairi, M.; Giralt, E. *Chembiochem* **2009**, *10*, 2736.
(10)   Bogan, A. A.; Thorn, K. S. *J Mol Biol* **1998**, *280*, 1.
(11)   Tugarinov, V.; Kay, L. E. *Chembiochem* **2005**, *6*, 1567.
(12)   Bodenhausen, G.; Ruben, D. J. *Chemical Physics Letters* **1980**, *69*, 185.
(13)   Kay, L.; Keifer, P.; Saarinen, T. *Journal of the American Chemical Society* **1992**, *114*, 10663.
(14)   Morris, G. A.; Freeman, R. *Journal of the American Chemical Society* **1979**, *101*, 760.
(15)   Gang, Z.; William, S. P., *56*, 267.
(16)   Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531.
(17)   Smrcka, A. V.; Kichik, N.; Tarrago, T.; Burroughs, M.; Park, M. S.; Itoga, N. K.; Stern, H. A.; Willardson, B. M.; Giralt, E. *Proc Natl Acad Sci U S A* **2010**, *107*, 639.
(18)   Pellecchia, M. *Chem Biol* **2005**, *12*, 961.
(19)   Reibarkh, M.; Malia, T. J.; Wagner, G. *J Am Chem Soc* **2006**, *128*, 2160.
(20)   Medek, A.; Hajduk, P.; Mack, J.; Fesik, S. *J. Am. Chem. Soc.* **2000**, *122*, 1241.
(21)   Krishnamoorthy, J.; Yu, V. C.; Mok, Y. K. *PLoS One* **2010**, *5*, e8943.
(22)   Neuhaus, D.; Williamson, M. P. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*; 2nd Edition ed.; Wiley: New York, 2000.
(23)   Meiboom, S.; Gill, D. *Review of Scientific Instruments* **1958**, *29*, 688.
(24)   Ni, F.; Scheraga, H. A. *Accounts of Chemical Research* **1994**, *27*, 257.
(25)   Moriz, M.; Bernd, M. *Angewandte Chemie International Edition* **1999**, *38*, 1784.
(26)   Klein, J.; Meinecke, R.; Mayer, M.; Meyer, B. *Journal of the American Chemical Society* **1999**, *121*, 5336.
(27)   Groves, P.; Kover, K. E.; Andre, S.; Bandorowicz-Pikula, J.; Batta, G.; Bruix, M.; Buchet, R.; Canales, A.; Canada, F. J.; Gabius, H. J.; Laurents, D. V.; Naranjo, J. R.; Palczewska, M.; Pikula, S.; Rial, E.; Strzelecka-Kiliszek, A.; Jimenez-Barbero, J. *Magn Reson Chem* **2007**, *45*, 745.
(28)   Mayer, M.; Meyer, B. *J Am Chem Soc* **2001**, *123*, 6108.
(29)   Mayer, M.; James, T. L. *J Am Chem Soc* **2004**, *126*, 4453.
(30)   Meyer, B.; Peters, T. *Angew Chem Int Ed Engl* **2003**, *42*, 864.
(31)   Dalvit, C.; Pevarello, P.; Tato, M.; Veronesi, M.; Vulpetti, A.; Sundstrom, M. *J Biomol NMR* **2000**, *18*, 65.
(32)   Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. *Journal of Biomolecular NMR* **2001**, *21*, 349.
(33)   Pellecchia, M.; Sem, D. S.; Wuthrich, K. *Nat Rev Drug Discov* **2002**, *1*, 211.
(34)   Oltersdorf, T.; Elmore, S. W.; Shoemaker, A. R.; Armstrong, R. C.; Augeri, D. J.; Belli, B. A.; Bruncko, M.; Deckwerth, T. L.; Dinges, J.; Hajduk, P. J.; Joseph, M. K.; Kitada, S.; Korsmeyer, S. J.; Kunzer, A. R.; Letai, A.; Li, C.; Mitten, M. J.; Nettesheim, D. G.; Ng, S.; Nimmer, P. M.; O'Connor, J. M.; Oleksijew, A.; Petros, A. M.; Reed, J. C.; Shen, W.; Tahir, S. K.; Thompson, C. B.; Tomaselli, K. J.; Wang, B.; Wendt, M. D.; Zhang, H.; Fesik, S. W.; Rosenberg, S. H. *Nature* **2005**, *435*, 677.
(35)   Fairbrother, W. J.; Champe, M. A.; Christinger, H. W.; Keyt, B. A.; Starovasnik, M. A. *Protein Sci* **1997**, *6*, 2250.
(36)   Bruker, *http://www.bruker-biospin.com/nmr_software.html* 2007.
(37)   Keller, R. *The Computer Aided Resonance Assignment Tutorial*; first edition ed.; CANTINA Verlag, 2004.
(38)   Origin. *http://www.originlab.com/* 2007.
(39)   Chem. COmp. Group. *http://www.chemcomp.com/index.htm*.
(40)   Pan, B.; Li, B.; Russell, S. J.; Tom, J. Y.; Cochran, A. G.; Fairbrother, W. J. *J Mol Biol* **2002**, *316*, 769.
(41)   Mori, S.; Abeygunawardana, C.; Johnson, M. O.; van Zijl, P. C. *J Magn Reson B* **1995**, *108*, 94.
(42)   Piotto, M.; Saudek, V.; Sklenář, V. *Journal of Biomolecular NMR* **1992**, *2*, 661.
(43)   Tarrago, T.; Kichik, N.; Claasen, B.; Prades, R.; Teixido, M.; Giralt, E. *Bioorg Med Chem* **2008**, *16*, 7516.
(44)   Voehler, M. W.; Collier, G.; Young, J. K.; Stone, M. P.; Germann, M. W. *J Magn Reson* **2006**, *183*, 102.
(45)   Schumann, F. H.; Riepl, H.; Maurer, T.; Gronwald, W.; Neidig, K. P.; Kalbitzer, H. R. *J Biomol NMR* **2007**, *39*, 275.
(46)   Hwang, T. L.; Shaka, A. J. *J. Magn. Reson.* **1995**, *112*, 275

3.2 Production and characterisation of folded recombinant VEGF

The first step in this project was the production of sufficient amounts of folded recombinant VEGF suited for subsequent experiments. A valuable starting point for this was the preliminary work performed by Ricard Rodríguez Mías during his PhD thesis in our group.[1] The expression of VEGF was achieved by a construct developed by Fairbrother and coworker at Genentech.[2] It consists of residues 11-109 of $VEGF_{121}$, the smallest isoform present in humans, as described in chapter 1.5.2. The missing residues were removed to increase solubility of the protein and did not significantly affect the interaction of VEGF with its receptor.

Initially we were only interested in NMR based techniques in this thesis, which require at least for protein based experiments isotopic labeling. We decided to combine an *all*-^{15}N- with a methyl-^{13}C-methionine labeling scheme. The first scheme allowed us to observe signals from the amides of the protein backbone and nitrogen containing side chains with a ^{1}H-^{15}N-HSQC NMR experiment. This was successfully applied for VEGF structure verification and CSP based binding studies with one residue resolution. The second labeling scheme allowed us to observe NMR signals of the methyl groups of the five distinct methionines per monomer. Although this approach provides only five local probes it is valuable for binding studies by CSP since two methionines (M18 and M81) are located in the receptor binding interface. Furthermore, the methyl groups give rise to higher signal intensities compared to amide signals and do not exchange with water, which allows them to work in perdeuterated buffer and at lower protein concentrations.

The VEGF protein was produced in a similar way to Fairbrother and collaborators at Genentech.[2] Briefly, an E. coli strain auxotrophic for methionine was transfected with p6XHisVEGF$_{11-109}$ and pMS421 plasmids. Protein expression was performed in M9 minimal media supplemented with $^{15}NH_4Cl$ as nitrogen source and ^{13}C-metyl-methionine. Expression was followed by His-tag purification and refolding of VEGF to its native dimmeric structure. Since the presence of the His-tag reduces protein solubility and stability, a proteolytic cleavage was performed with the enzyme Genenase I, followed by anion exchange and size exclusion chromatography. The final yield of refolded purified protein was approximately 7 mg/ L starting media.

One important consideration in the production of VEGF was the amount of protein that we needed. The techniques that we were planning to use required relatively high amounts of protein. I expected that an optimization of the expression conditions would improve the yield of protein. This was discussed with Dr. Javier Méndez from the fermentation service of the University of Barcelona. Subsequently we agreed to perform the expression of VEGF in a fermentor at a 10L scale. With this approach we raised the final yields of refolded purified protein to approximately 250 mg /L of starting media.

Further the refolding and purification steps were optimized, which lead to a purity of 97% of refolded dimeric VEGF prior to size exclusion chromatography and allowed us to skip this time-consuming step. The refolded purified protein was exchanged to its final buffer, diluted to the final concentration for the specific application, aliquoted and shock frozen.
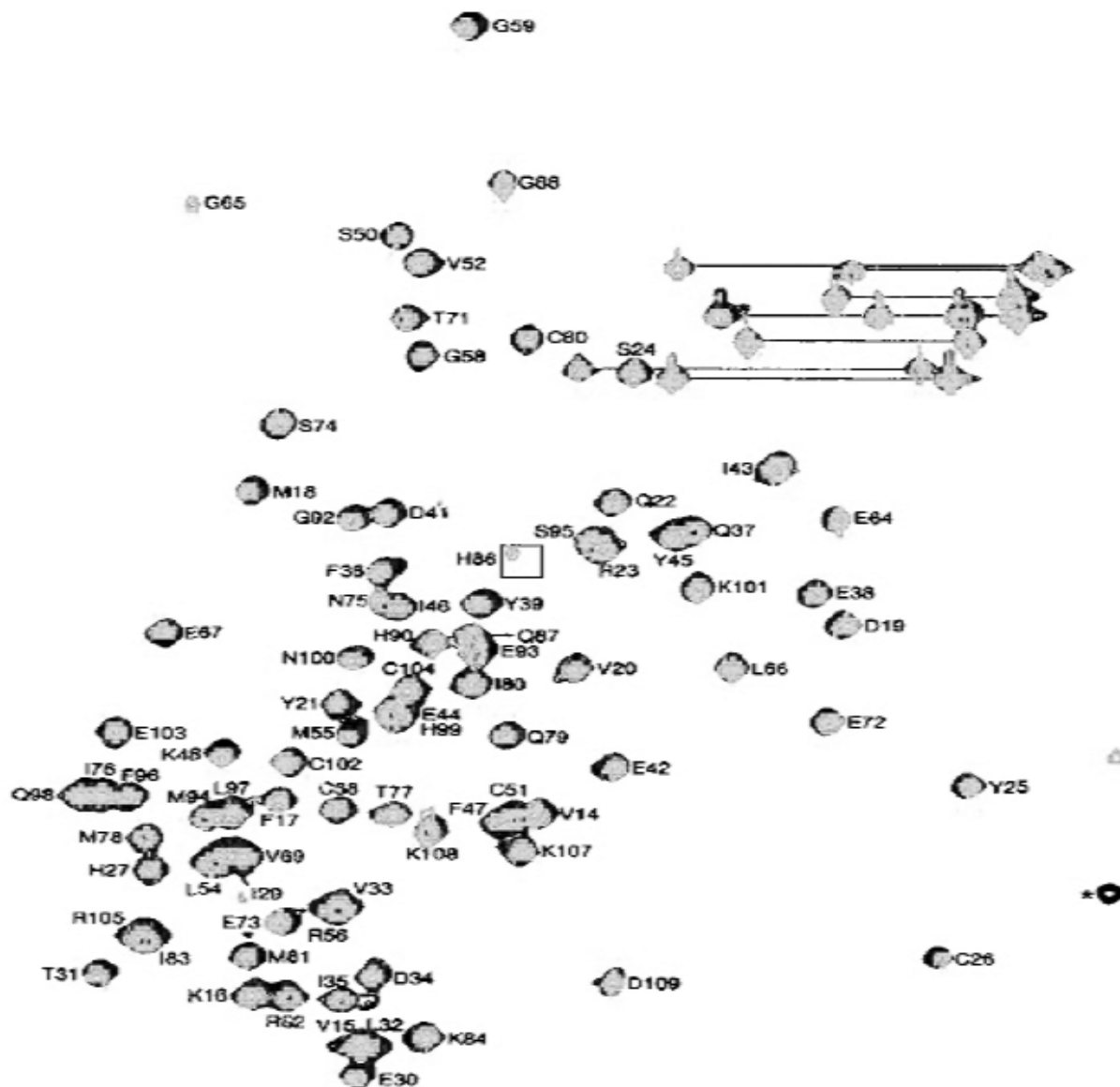

Figure 3.2-1: 15N-1H HSQC spectra acquired of a 100 uM sample of [methyl-^{13}C]-Met-all-^{15}N VEGF$_{11-109}$ at 45 ℃ (grey). To confirm fold integrity the same spectra with assignment is shown from Fairbrother and collaborators (black).

Fold integrity and overall VEGF$_{11-109}$ spectroscopic properties were assessed using the ^{1}H-^{15}N-HSQC NMR experiment (Figure 3.2-1). Again, the preliminary work of Ricard Rodríguez Mías was a valuable guidance.[1] As elaborated above this type of spectrum shows cross peaks corresponding to protons directly bond to isotopic labeled nitrogen. This connection is present in the backbone of each amino acid (excluding proline) and gives rise to one cross peak. Additionally peaks appear if the amino acids asparagine, glutamine, lysine and tryptophan are present in the protein since their side chains

contain additional nitrogen atoms. The recorded spectra for VEGF suggest that the present protein is in a well-folded state since the broad signal dispersion in both dimensions. A closer inspection and comparison to the NMR data used by Fairbrother and collaborators to determine the structure of VEGF showed perfect consistency. It is further noteworthy that the same experiment was repeated after storing the sample for 5 months at 4°C and no significant perturbations were visible.

In order to validate VEGF activity we decided to test its recognition properties against a published ligand. In our lab the peptide based ligand v107 was available which recognizes and occupies most of the receptor binding epitope of VEGF (Figure 1.5.2) For this purpose a ^{1}H-^{15}N-HSQC NMR spectrum of a 100 uM VEGF was recorded at 45°C before and after the addition of two equivalents of v107 (Figure 3.2-2). As elaborated in chapter 3.1 this kind of experiments entitled CSP can be useful in detection of binding events and mapping of the region of this interaction on the protein surface. It was obvious that the addition of v107 affected the ^{1}H-^{15}N HSQC spectrum of VEGF. A closer inspection showed that there were several cross peaks with chemical shifts probably more that can be explained by the proximity of the ligand to the affected backbone in the protein-ligand-complex. Nevertheless this result confirms that the recombinant produced VEGF was able to recognize the ligand.
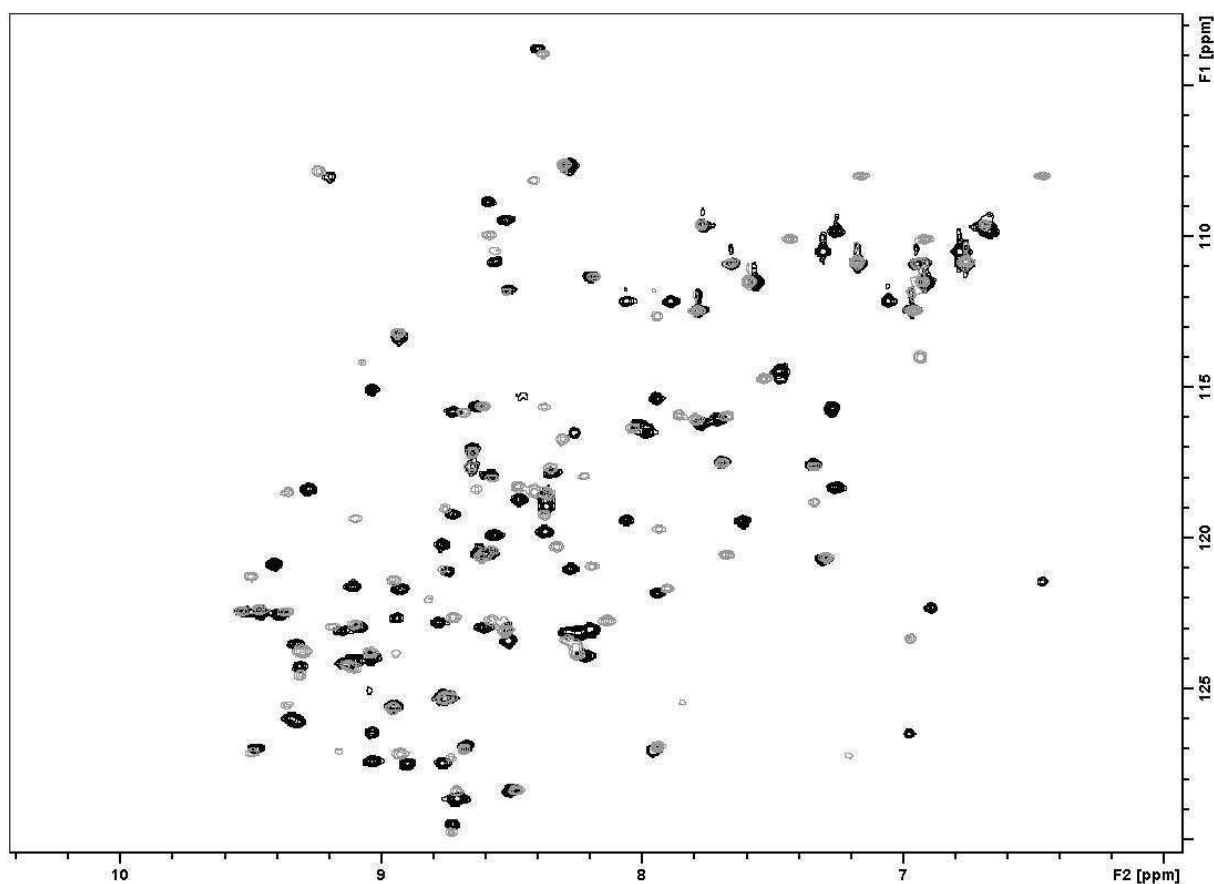

Figure 3.2-2: ^{1}H-^{15}N HSQC spectra of a 100 uM sample of [methyl-^{13}C]-Met- all-^{15}N VEGF$_{11-109}$ at 45 °C before (black) and after the addition of thre e equivalents of v107 (grey).

During the course of this thesis the opportunity emerged to study VEGF with complementary methods. In collaboration with Andrey Dyachenko, one of my closest coworker in Prof. Ernest laboratory MS based approaches were explored to study the interaction of VEGF with peptidic ligands in the gas phase. We also indented to obtain cocrystals of VEGF with putative ligands in collaboration with Dr. Jordi Benach from the synchrotron facility ALBA close to Barcelona. Both approaches do not require isotopic labeling and it is even of disadvantage in the case of MS since it leads to broad peaks due to incomplete labeling. For both collaborations non-labelled VEGF was expressed in LB media and purified it as described above. Figure 3.2-3 shows a MS based characterisation of the final purified labelled and non-labeled protein performed by Andrey Dyachenko.
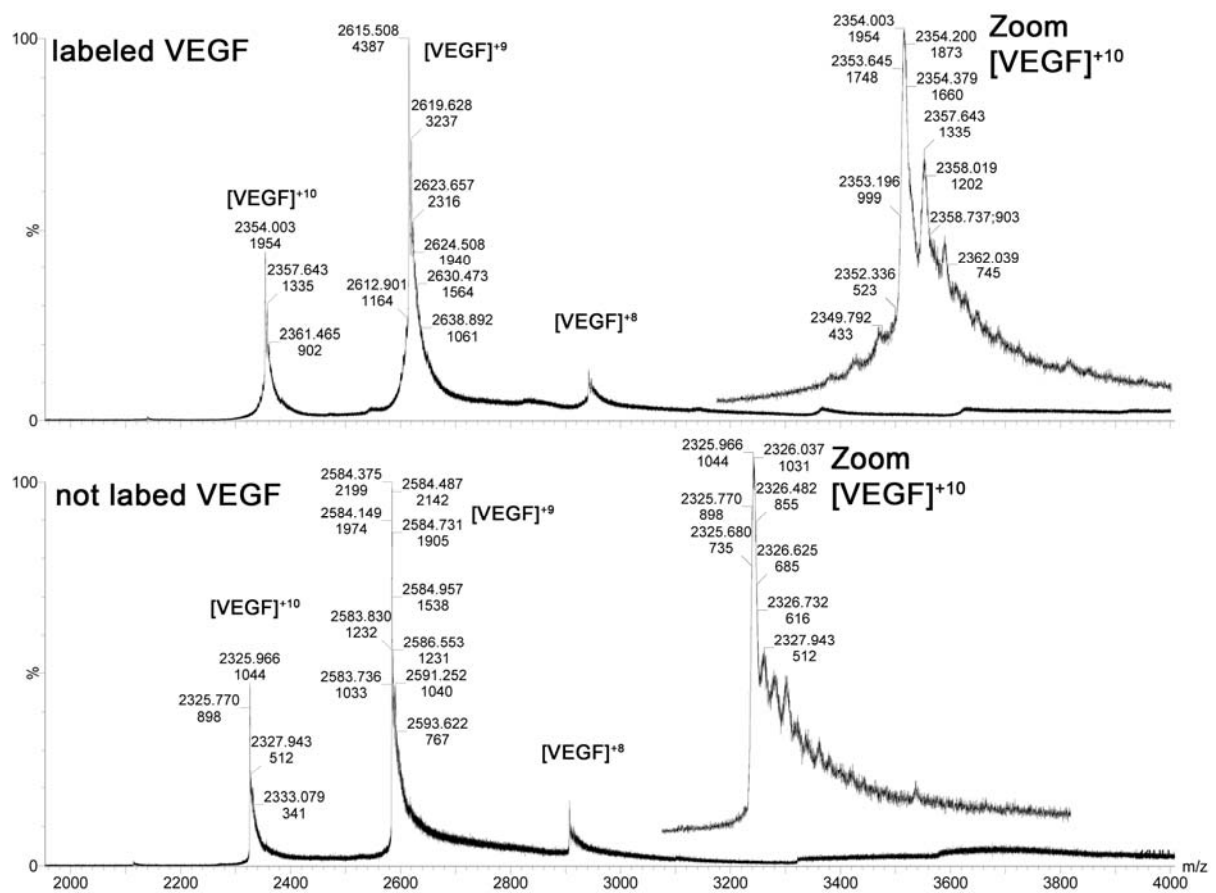

Figure 3.2-3: Characterization of [methyl-^{13}C]-Met- all-^{15}N and nonlabeled VEGF$_{11-109}$ by ESI time-of-flight mass spectrometry

For both samples several charge states could be observed. The labelled protein shows a clear shift to higher m/z values due to the mass gain by the incorporation of ^{15}N and ^{13}C isotopes. The zoom of the +10 charged state shows that the left slope of the peak from the non-labelled protein is much sharper than the slope of the peak from the labelled protein. Both samples show weak slopes with multiple sub-peaks on the right side of the parent peak with results from sodium adducts. As a result,

the main peak of the non-labelled protein was better defined than its labelled counterpart, which facilitated the analysis of the data in this collaboration.

In this chapter we describe the successful expression and purification of VEGF. The whole process was significantly improved by the expression of the protein in a fermentor, which led to an approximately 35 fold increase of yield. It is likely, that this can be contributed to the improved aeration of the fermentor, which allows both faster bacterial growth and higher final culture densities. In my opinion the time was well invested to incorporate this improvement. The expression was so high, that each labeled and non-labeled protein had to be expressed only once during this thesis. Then the frozen bacterial pellets could be harvested and purified whenever more protein was needed. This was convenient as it allowed us to focus on more challenging aspects of this thesis. Finally the use of the fermentor cost only 40 Euro per day but offered much higher yields of protein per liter of starting media. This is important for the case of the labeled protein since $^{15}NH_4Cl$ and ^{13}C-metyl-methionine become expensive when used at gram scales. At the moment the only inconvenience in the production of VEGF is the need to use the enzyme GenenaseI for proteolytic cleavage of the His-Tag. This enzyme has to be ordered from the USA and cost approximately 15 Euro per milligram of VEGF, a considerable expenditure as over 200 mg of VEGF were produced during this thesis.

The structural fold integrity and overall spectroscopic properties of the produced VEGF were assessed using the ^{15}N-^1H HSQC NMR experiment. The protein showed perfect consistency with the published spectra indicating overall fold integrity and the absence of non-native oligomers. The protein behaved very beneficial in the NMR spectrometer. A ^{15}N-^1H HSQC NMR experiment could be recorded in ca. 1 hour at 45°C with a protein concentration of 100 uM.

(1)   Rodriguez Mías, R. PhD thesis 2006.
(2)   Fairbrother, W. J.; Champe, M. A.; Christinger, H. W.; Keyt, B. A.; Starovasnik, M. A. Protein Sci 1997, 6, 2250.

3.3 Design and setup of a proprietary fragment based library

The laboratory of Prof. Giralt had no previous experience in fragment based drug discovery. One of the first milestones for this project was therefore to design and preparation a fragment based library that could subsequently be explored in this project. The key person involved in the design of the library was Prof. Xavier Barril from the University of Barcelona. Prior to his current position, he was a computational scientist at Vernalis (UK), where he gained experience in the analysis and design of fragment libraries. He is also coauthor of a publication that provides a detailed description of design guidelines for fragment based library design.[1] The design of the library was performed in multiple steps. The initial part was designed by Prof. Xavier Barril. During the setup it became apparent that a significant amount of compounds had to be excluded due to solubility issues. This gave me the opportunity, with the support from Prof. Xavier Barril, to learn how to search for substitutes and finally design a second part of the library that was aimed at closing gaps in chemical space. The design of the first part of the library will only be described briefly, the second part in more detail.

(1)   Baurin, N.; Aboul-Ela, F.; Barril, X.; Davis, B.; Drysdale, M.; Dymock, B.; Finch, H.; Fromont, C.; Richardson, C.; Simmonite, H.; Hubbard, R. E. *J Chem Inf Comput Sci* **2004**, *44*, 2157.

3.3.1. Design and setup of the first part of the library

The first part of the library consisted of approximately 500 compounds. These compounds were selected from an initial set of over 1.2 million non-redundant compounds collected from different international vendor databases. These compounds were tested against a set of filters to avoid undesired features and to ensure solubility and drug like properties of the fragments. The remaining compounds were clustered based on their similarities and one compound per cluster was selected, that was the best representative for its cluster. With this approach one compound was selected for the library from the 500 most populated clusters leading to the previous mentioned 500 compounds. All selected compounds exhibit a molecular weight between 150 and 300 Da, which entitles them as fragments opposed to the weight of conventional small molecule drugs around 500 Da.

The rational behind the selection procedure was to obtain a library of compounds that cover the chemical space of commercial available compounds in an optimal manner. This design principle is called "SAR by catalog". The idea is that for each active compound one can simply order analogs from the member of the original cluster and thus generate SAR data in a straightforward manner.

The iterative nature of screening programs allows an optimization of the implied procedures to obtain optimal results. However this is only possible if rigid experimental standards are defined. My task was therefore to define these standards and plan the setup of the fragment library. It may seem peculiar to decide on screening conditions before the setup of the fragment library has even started. The setup of the library has to reflect the conditions one would later like to use. With respect to this issue, the discussions with Dr. Marc Martinell, Head of Biophysics at Crystax, a pharmaceutical company involved in structure-based drug discovery were very helpful. As a result I decided to screen the compounds in mixtures with a concentration of 0.5 mM of each fragment and a total amount of 3% DMSO per sample which limits mixture size to 5 to 6 fragments if we prepare them as 100 mM stocks. The ratio between each fragment and the protein was chosen to be 50:1. According to these parameters we defined the workflow presented in figure 3.3.1.
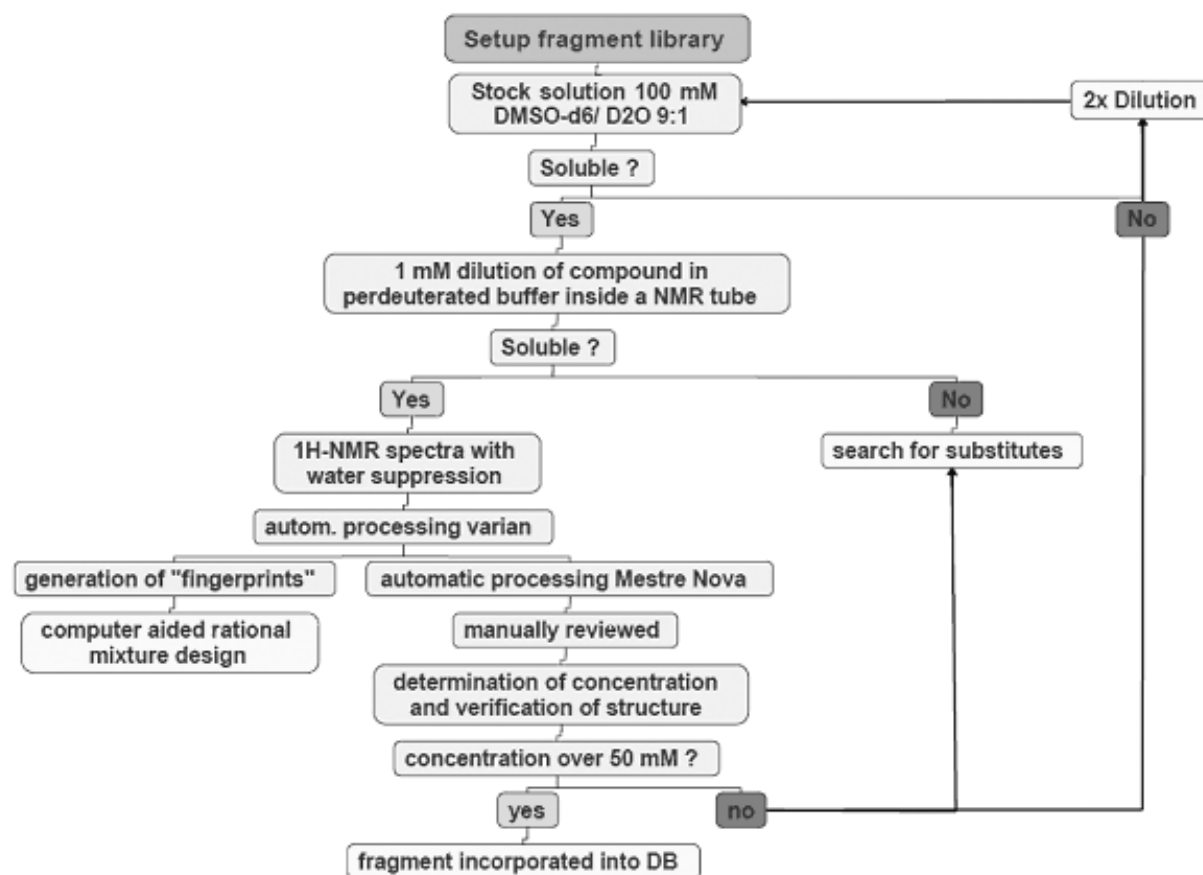
Figure 3.3.1: Workflow for the setup of the fragment library.

Each compound was treated with a specific amount of DMSO-d6/D2O 9:1 to yield a 100 mM stock solution. If the stock passed visible solubility control a NMR sample was prepared. For this purpose the stock solution were diluted with deuterated NMR buffer to a concentration of 1 mM. This was twice the concentration that was used later in the screening and should help to address solubility issues that may arise when the fragments are applied as mixtures and the sample contains therefore a higher load of organic molecules. The NMR sample contained also t-BuOH as a reference to

determine the exact concentration of the stock solution. This was necessary to address solubility issues as well as inaccuracies in the amount of the delivered fragments. NMR samples with absence of precipitation were used to acquire ^{1}H spectra with water suppression at 37 °C. The obtained data was processed in an automatic fashion with Mestre Nova and manually reviewed. If deemed necessary, the phases of the spectra were adjusted and the signals integrated. The signature of the signals was checked if they match the chemical structure of the fragment. The concentration of the compound in the stock solution was determined by comparing intensities of compound signals with the signal of the three methyl groups of t-BuOH.

Compounds that passed both visible solubility inspections and were confirmed to have a stock concentration over 50 mM were incorporated in a database. This database was constructed using the software MOE from Chemical Computing Group[1] and incorporated all available information of the specific compound. Later, the database allowed direct access to cheminformatic tools. Compounds which failed either visible solubility control of the stock solution or showed stock concentrations below 50 mM were diluted by a factor of 2 with additional DMSO-d6/D2O 9:1 to yield a 50 mM stock which was tested again.

Compounds that were not available for purchase initially, failed twice the stock definitions or appeared to be insoluble at 1 mM in NMR buffer were subjected to a similarities search to find substitutes. For this purpose the database of the Maybridge screening collection was downloaded containing over 56.000 compounds.[2] With MOE the database was filtered to meet the criteria for our fragment library and a subset of approximately 7.000 compounds was searched for substitutes, which was performed by calculating MACCS fingerprints. The Tanimoto coefficient was used as a metric for similarity and the entities with the highest similarities were selected as a substitute with a threshold of 70%. Selected compounds were purchased and incorporated in the fragment library.

In summary, 540 compounds were purchased. 71 of this failed to have visible solubility in the NMR buffer at 1 mM. 67 compounds failed visible solubility in the Stock solution or a final stock concentration over 50 mM. 29 compounds matched the defined criteria after the two fold dilution of the stock solution and could be incorporated into the library. Substitutes for 130 compounds were searched and 86 substitutes accepted. The total number of compounds that were present in the MOE database at this point was 402.

NMR based screening, while being capable of detecting very weak interactions, has a relative low throughput. To perform the screening in a reasonable amount of time compounds are screened in mixtures instead of individual compounds. These mixtures can have varying size, but are often randomly generated which may require deconvolution to identify several binding molecules from each other or to distinguish them from non binding molecules. In chapter 3.4.1 we will describe the

development of a novel tool to generate mixtures with minimized overlap that allow the direct assessment of all molecules in a mixture without deconvolution. At this point I will only mention that we completed the setup of the fragment library for NMR based screening by preparing stock solution of mixtures. Each mixture contained 5 fragments at a concentration of 16.6 mM. Originally we were aiming for 20mM, but the presence of some diluted stocks led to reduced concentration.

The mixture stock can be directly used for screening. The addition of 3% of this stock to a protein sample would allow screening the individual fragments at a concentration of ca. 500 uM while keeping the total DMSO content under 3%. The setup was finished by preparing NMR samples of all mixtures as they would be used in a screening exercise to exclude the possibility of cross reactivity or precipitation in the mixed sample. Both behaviours were not observed.

(1)   Chem. Comp. Group. *http://www.chemcomp.com/index.htm*.
(2)   Maybridge *http://www.maybridge.com*.

3.3.2 Design and setup of the second part of the library

During the time of my thesis we decided to expand our fragment library to close gaps in the chemical space and raise the total number of fragments over 500, the lower limit of fragment libraries in the pharmaceutical industry. The second part of the library was designed by me with the guidance of Prof. Xavier Barril around one year after the first part. The process was very similar to what he did for the first part.

As described in figure 3.3.2, compound collections of the four biggest international vendors were united in a database containing over one million compounds.
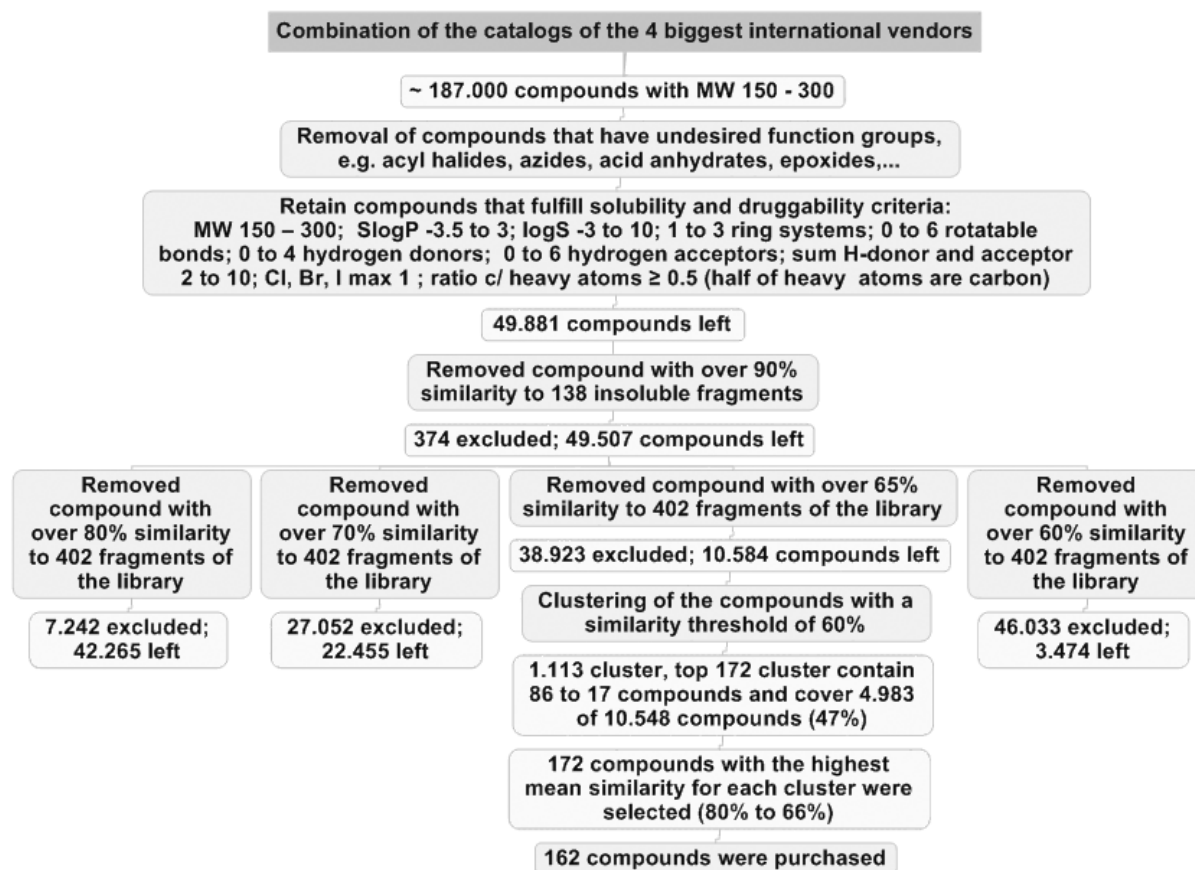
Figure 3.3.2: Computational design of the second part of the library

The database was filtered to remove compounds that had molecular weights outside of the desired range between 150 and 300 Da. The remaining subset of approximately 190 thousand compounds was subjugated to more rigorous selection criteria: Compounds that contained reactive groups, e.g. acyl halides, azides, acid anhydrates or epoxides were excluded. Further only compounds were included that had a predicted solubility over 1 mM (logs >-3) and drug like properties regarding polarity (SlogP) the number of ring systems, the number of rotatable bonds, and the number of hydrogen bond acceptor and donor. Around 50,000 compounds passed these criteria and were used as a basis to search for substitutes for insoluble compounds and to fill gaps in the coverage of the chemical space of our library. Compounds can be compared in many different ways. We used MACCS fingerprints and the Tanimoto coefficient for this purpose. MACCS fingerprints is a set of 166 "questions" (e.g. Are there fewer than 3 oxygens? Is there a S-S bond? Is there a ring of size 4?), that are asked to the molecule and can be answered with yes or no or respectively 1 and 0 and thus form a 166 bit string. Two bit strings A and B that represent two molecules can be simply compared by calculating their Tanimoto similarity, which is defined as the number of features in intersect (A,B) divided by the number of features in union (A,B). A Tanimoto coefficient close to 1 (100%) means both molecules are very similar, coefficients close to 0 (0%) means that they are very different.

To decrease the chance to select insoluble compounds, we decided to exclude molecules that were similar to the 138 insoluble fragments of the first part of the library. After exploring cutoffs between

100% and 85% of similarity we decided to exclude around 370 compounds that had a similarity over 90% to the set of insoluble compounds. To ensure the selection of compounds that cover novel chemical space, we removed compounds with high similarity to the 402 compounds that were already incorporated in the first part our library. Similarity cutoffs of 60%, 65%, 70% and 80% were explored, which lead respectively to the elimination of 46, 39, 27 and 7 thousand from the remaining 50 thousand compounds. We decided to use the similarity cutoff of 65% since this would leave only around 10 thousand compounds.

The remaining 10 thousand compounds were clustered together with a stochastic cluster algorithm.[1] Again, by exploring different similarity cutoffs we choose a value of 60%, which led to the formation of around 1100 clusters. The 172 highest clusters contained 86 to 17 compounds each and together covered around 47% of the 10,000 compounds. From each of these clusters the compounds with the highest mean similarity for all members (between 80 and 66%) was selected for purchasing. The procedure described here was performed with the software MOE.

The setup of the second part of the library was identical to the procedure for the first part of the library. In summary, 161 fragments were purchased. 21 of this failed to have visible solubility in the NMR buffer at 1 mM. 15 compounds failed visible solubility in the stock solution or a final stock concentration over 50 mM. 20 compounds matched the defined criteria after the two fold dilution of the stock solution and could be incorporated into the library. In total 125 compounds were added to the library raising the total number of fragments from 402 to 527, which was the aim of this exercise. As mentioned for the first library, part mixtures were designed of five fragments that showed reduced signal overlap ready for NMR based fragment screening.

(1)   Reynolds, C. H.; Pfahler, L. B. Journal of Chemical Information & Computer Sciences 1998, 38, 305.

3.3.3 Discussion

In this chapter the successful design and setup of a proprietary library was described. Many academic groups do not design their own fragment libraries. Instead they select one offered by various commercial companies, such as Maybdridge, Asinex and Enamine, etc. This is a very cost-effective choice, since these companies offer very competitive discounts for ordering multiple compounds. For instance, Asinex offers a 70% discount for orders over 1,000 compounds compared to orders under 30 compounds. Furthermore, this option does not require any knowledge of and access to software for the library design. In some cases the libraries can even be purchased in pre-plated stock solutions that are ready to use, thus saving time and resources for the setup of the library. But in my opinion, the disadvantage of commercial libraries is the lower quality that they may have concerning the

coverage of chemical space. I assume that they are designed on the basis of the compounds that each company owns. This may be one of the reasons why pharmaceutical companies, despite forced to work cost-efficient design their own fragment libraries and purchase from several vendors. Obviously experts in the pharmaceutical industry are aware of the fact that the success of any screening project can only be as good as the design of the employed library.

The design of our library was based on the experience that Prof. Xavier Barril and his former coworker at Vernalis gained over the years in the field of fragment based drug discovery. Some of the tools that we used in this process are open for debate. For example one could ask if MACCS fingerprints are the best descriptors to portray something as complex as a small molecule and the possible interactions that it could make with a protein. I assume that there are better descriptors, for example fingerprints that recognize that molecules are three dimensional and flexible. However I am not sure if these descriptors are appropriate when the task is to describe and compare hundreds of thousands of compounds. I assume that the fingerprints that we chose offer a good compromise between their descriptive power and the speed of their calculation. But I also have to note that based on my experience, compounds with over 80% similarity appear to be related for my understanding of molecular recognition, contrary compounds under this threshold show increasing changes in their scaffolds which prohibits a similar interaction space. The same discussion could be repeated for the clustering approach. The stochastic cluster algorithm that we used is based on the nearest neighbor principle and computes one possible solution, but not necessary the optimal distribution of compounds into cluster with high internal similarities but distinguishable properties from other cluster. To achieve this, a deterministic cluster algorithm has to be applied, which has exponential growing cost in speed and is therefore not well suited for hundreds of thousands of compounds. Relevant decisions during the design of the second part of the library were the thresholds to exclude compounds that were similar to insoluble or already represented compounds. After visually exploring the results from different similarity thresholds I though a cutoff of 90% selects a reasonable subset that is prone to insolubility issues as the parent compounds. After exploring cutoffs to reject compounds that were already represented by the 402 compounds of the first part of the library, I decided to use a similarity of 65% which eliminated 80% of the compounds. This was a reasonable number, since our aim was to increase the library from around 400 fragments to over 500. This represents an increase of 20% and requires that the other 80% of the chemical space were already covered by the existing library.

Our library was designed following the "SAR by catalog" principle. It focuses on a design which offers a good coverage of the chemical space while enabling a cost-effective exploration of follow up compounds. This was implemented by clustering commercially available compounds and the selection of one representative of the highest populated cluster. Should a representative fragment

be a potential hit, the proximate chemical space can be explored by purchasing analogs from the same cluster that it belongs to. This leads to the generation of SAR data without the need of expensive and time-consuming synthesis.

I think this library design is especially well suited for targeting diverse but classical drug targets.

A slightly different approach is the purely diversity oriented library. This library does not per se enforce easy follow up but focuses on selecting the most diverse set of compounds. To some degree, this leads to the selection of "exotic" compounds from less explored areas of the chemical space. Depending on the target protein this can be a disadvantage, since follow up chemistry is prone to be challenging but offers also the opportunity to explore novel chemical space which might be necessary for challenging targets.

Finally, libraries can be focused, which means they are tailored against a specific set of targets like kinases, ion channels or GPCRs. They exploit the knowledge of native ligands or properties of their binding sites to design a library that only explores the surrounding chemical space. This results in superior hit rates for the target class but will perform poorly against other targets.

Probably the design of a "SAR by catalog" library was appropriate to develop a general tool for the laboratory that can be explored for a multitude of projects. However it might not be the best approach to target a challenging class as protein-protein interfaces.

A striking experience during library design and setup was the high number of compounds that turned out to be insoluble in either their stock solutions of in the screening conditions. From the first part of the library 138 of 540 compounds were insoluble, which represents 26%. Of the second part of the library 36 of 161 compounds were insoluble which represents 22%. In total these were 174 compounds that represent an investment of approximately 4.300 €. Why is this happening and how can this be avoided?

As part of our selection procedure, we computed the solubility of each compound. We included only compounds with a logS value ≥-3, which corresponded to a solubility of at least 1 mM. Apparently this and similar algorithms were not very accurate. This is a general problem; Barurin et al. reported 12% for a proprietary library[1] and Peters et al. 14% of insoluble compounds for a commercial library.[2] These numbers concern the solubility in buffer, not in DMSO stocks. Barurin et al. do not report any insoluble compounds in pure DMSO stocks. If we count only the compounds from our library that fail solubility in buffer we arrive at a very similar value of 13%. I can imagine three possibilities to improve this situation, but as I will discuss they do not offer a general solution: First, use advanced algorithms such as ALOGpS.[3] However this algorithm requires at least some compounds with experimental logs and logP values. Second, to select only compounds with higher logS thresholds. This would probably improve the situation, but limit the diversity and the chemical space of a library,

especially since most fragments are lipophilic. As an example the mean logS value is -4 for the Asinex library from 2011. Only 30% of the library has logS values over -3, and only 11% a logs value over -2.

Third, compounds can be screened at lower concentrations. In the literature a plenitude of different experimental conditions are described that correspond to the requirements of different proteins, techniques and projects. It appears to me, however, that a concentration of 500 uM per fragments is situated relatively in the middle of the field. Lowering this concentration would at least decrease the sensitivity for the proton based NMR experiments that we planed to use.

Contrary to many publications we did not use 100% DMSO stocks, but a mixture 9:1 of DMSO with water, which led to solubility issues for 19% of our compounds. One third of these compounds could be incorporated into the library but as 50 mM stocks contrary to the usual 100mM. This approach "rescued" some of the compounds but at the cost of having different final stock solutions in the library. The use of a 9:1 mixture of DMSO with water was a recommendation of Dr. Marc Martinell. It is based on the high hygroscopicity of pure DMSO, which is reduced in the mixture. As a result the concentration of the stock decreases slower over time and the appearance of precipitates are reduced. Further the concentration of DMSO is more stable. This could have been important when we used protein based NMR approaches to study ligand binding which required the ability to distinguish between ligand and DMSO induced changes (Chapter 3.6).

During the setup of the library a lot of effort was invested in the quality control: Solubility tests of the individual stocks and the single compounds, their chemical identity and purity, and the solubility of compound mixtures in screening conditions. As a result we did not face any obvious solubility or aggregation problems or other surprises during our screening efforts.

On the contrary, sample preparation was straightforward and fast. The briefly mentioned design of mixtures with minimized overlap allowed the direct evaluation of all compounds without the need of deconvolution. And finally, the previous generated data during the quality control proved to be a valuable source of information for the analysis of binding candidates.

In my option our library has a level of high quality and is easy of use.

(1)   Baurin, N.; Aboul-Ela, F.; Barril, X.; Davis, B.; Drysdale, M.; Dymock, B.; Finch, H.; Fromont, C.; Richardson, C.; Simmonite, H.; Hubbard, R. E. *J Chem Inf Comput Sci* **2004**, *44*, 2157.
(2)   Rademacher, C.; Guiard, J.; Kitov, P. I.; Fiege, B.; Dalton, K. P.; Parra, F.; Bundle, D. R.; Peters, T. *Chemistry* **2011**, *17*, 7442.
(3)   Tetko, I. V.; Tanchuk, V. Y.; Kasheva, T. N.; Villa, A. E. *J Chem Inf Comput Sci* **2001**, *41*, 1488.

3.4 Development of optimized methods for fragment based drug discovery

3.4.1 Computer-Aided Design of Fragment Mixtures for NMR-based Screening

One of the milestones of this thesis was the setup and preparation of a fragment based library for NMR based screening. NMR is a relatively slow throughput technique and requires screening of compound mixtures rather than single compounds to increase the overall screening speed. To complete the setup of our library we had to prepare this kind of mixtures.

This is a process that is either performed by randomly mixing compounds - however signal overlap between different compounds is likely to occur and requires subsequent additional experiments for deconvolution - or mixtures can be designed by the trial and error method where single compound NMR spectra are manually combined with the aim to find a solution where at least one signal per compound can be unambiguously assigned. This approach may reduce the possibility of deconvolution but requires a substantial investment of time while only a fraction of possible combinations can be explored when performed by a human. After discussing this issue with Xavier Arroyo, a PhD student in the lab or Prof. Ernest Giralt, we decided that this problem should be solved computationally. The resulting work is presented in the following part of this chapter and was accepted in PLOS one for publication. Xavier Arroyo and I developed this project conceptually. Xavier Arroyo was responsible for programming; I did the wet lab and NMR experiments and, with support of Dr. Miguel Feliz, converted the NMR spectra into fingerprints.

ABSTRACT

Fragment-based drug discovery is widely applied in industrial and academic screening programs. Several screening techniques rely on NMR to detect binding of a fragment to a target. NMR-based methods are among the most sensitive techniques and have the further advantage of yielding a low rate of false positives and negatives. However, NMR is intrinsically slower than other screening techniques; thus, to increase throughput in NMR-based screening, researchers often assay mixtures of fragments, rather than single fragments. Herein we present a fast and straightforward computer-aided method to design mixtures of fragments taken from a library that have minimized NMR signal overlap. This approach enables direct identification of one or several active fragments without the need for deconvolution. Our approach entails encoding of NMR spectra into a computer-readable format that we call a *fingerprint*, and minimizing the global signal overlap with a Monte Carlo algorithm. The scoring function used favors a homogenous distribution of the global signal overlap. The method does not require additional experimental work: the only data required are NMR spectra,

which are generally recorded for each compound as a quality control measure before it is introduced into the library.

INTRODUCTION

Fragment-based drug discovery has emerged in the past decade as a powerful tool for drug development and is now widely applied in academic and industrial screening programs. Its success derives from the structural simplicity and relatively low molecular weight (150 to 300 u) of the fragments, which contrast with the more complex, medium-weight compounds normally employed in high-throughput screening (HTS). Using fragments has three main advantages over using larger compounds: firstly, the chemical space is significantly smaller, and therefore, can be explored more efficiently; secondly, the hit rates are ten to 1000 times higher; and lastly, fragments often show high ligand efficiency, thereby facilitating work to improve their affinity.[1]

Various biophysical techniques such as X-ray crystallography, surface plasmon resonance and NMR have been exploited for fragment screening, where they must provide reliable detection of the mostly weak interactions between fragments and the target, with a low occurrence of false positives and negatives.

NMR-based fragment screening methodologies have become very popular, as they fulfill these requirements excellently. The only disadvantage of NMR compared to other screening methods is its low intrinsic sensitivity. To compensate for this, and to increase throughput, researchers often assay fragment mixtures, rather than single fragments, in NMR-based screening.[2,3] Modern NMR-based screening methodologies rely mainly on ligand observation experiments in which either a conventional NMR parameter of the ligand (*e.g.* relaxation properties), or the intermolecular proton magnetization transfer from the protein to the ligand, is evaluated.[3] In theory, both types of experiments enable direct identification of one or several binding fragments in a mixture that also contains non-binding fragments, yet they do not require deconvolution. The only requirement is that the NMR signals of the fragments in the mixture can be distinguished from one another, so that they can be evaluated separately, a subject which is only marginally covered in the literature.[4,5]

We addressed the NMR signal overlap issue through the following process: firstly, conversion of NMR data for compounds from our in-house fragment library into a meaningful, computer-readable format; secondly, evaluation of different computational algorithms for the task of reducing signal overlap; thirdly, preparation of *in silico* mixtures of fragments taken from our in-house library, based on zero or near-zero signal overlap; next, confirmation that the *in silico* mixtures behave similarly to real (chemically synthesized) fragment mixtures; and finally, testing of the general adaptability of the algorithm.

METHODS

NMR spectrometry and computation

All NMR spectra were acquired on a Varian Inova 500 MHz spectrometer with a 5mm PFG Penta Probe at 37 ºC. S/N for ^{1}H was 815:1 (0.1% ethylbenzene in CDCl$_3$).

All calculations were performed on an SGI$^®$ Altix$^®$ 4700 server (64 cores, 128 GB RAM).

Sample preparation, and generation of fragment *fingerprints*

Stock solutions (100 mM in 9:1 DMSO-d$_6$/D$_2$O) of each compound from our in-house fragment library were prepared, and then inspected visually to confirm solubility. Soluble compounds were further diluted to 1 mM in deuterated buffer (25 mM phosphate, 50 mM NaCl, 11µM *t*-butanol, pH 7.0), and their individual 1D-^{1}H-NMR spectra were recorded using presaturation for water suppression. The purity and identity of each fragment was manually checked in each spectrum. Compound concentration was calculated based on an internal standard (*t*-butanol). NMR data for all fragments that passed quality control were then translated into computer-readable files, called *fingerprints*, by a modified Varian script for automatic processing. The routine for signal integration was modified to integrate a narrow zone around each signal and to create an ASCII file for each NMR spectrum, consisting of the integration range of each signal and the value of the integral. Therefore, the generated file comprises several regions defined by start and end values that mark the spectral regions containing signals. Raw data were adjusted for subsequent calculations by removing regions originating from H$_2$O (4.780-4.530), DMSO (2.754-2.613) and *t*-butanol (1.320-1.130) and by reducing the size of all remaining regions by 50%.

Algorithms

We tested four different types of algorithms: greedy[6,7] and backtracking,[6,7] both of which are deterministic, and simulated annealing (SA)[8] and genetic,[9,10] both of which are stochastic.

As a starting point for all the algorithms we implements the same data structure, consisting in an array with the compounds, which one with its respective fragments. In the case of deterministic algorithms, the solution array was progressively filled, while stocastics was initialized with a random permutation of the compounds.

We used the greedy algorithm and the backtracking algorithm to maximize the number of library fragments that could be used in five-fragment mixtures that would not have any signal overlap. This was accomplished through a scoring function that maximizes the number of fragments (Table 3.4.1-1). The valid criterion for extend partial solutions in this case was the success of adding new fragments into the mixture, such that the process arrives at a solution once no more fragments can be added. To facilitate searching in the greedy algorithm, the fingerprint library was sorted by spectral area coverage at the beginning of the procedure. For the backtracking algorithm, to amplify

screening in the solution space and shorten the time required by the method, the method was executed several times in parallel, using random starting points.

We used the SA algorithm and the genetic algorithm for another scoring function (Table 3.4.1-1): to minimize overlap in each mixture. In the genetic algorithm, an initial population of 1,000 candidates was established with standard conditions of selection, breeding and mutation taxes.[9,10] The evolutionary process was extended over fifty generations. We ran the library through 100 iterations of SA algorithm in parallel with two million cycles, in which the temperature value was near zero. To study scalability with the virtual libraries, the SA algorithm was then run 100 times independently, using each list and the same algorithm conditions. Finally, the effect of temperature effect was tested using values from 0 to 25,000 in a virtual library of 500 fingerprints that contained an equal proportion of aliphatic and aromatic peaks.

Algorithm Type	Scoring Function [a]	Execution Time	Scalable	Optimization
Greedy[1]	$Scoring = N_i$	Very short	Yes	Very low
Backtracking[1]	$Scoring = N_i$	Very long	No	Low
Genetic[2]	$S = \sum_{i=0}^{N} \left((Ni_{ov} / Ni_t) x100 \right)^2$	Short	Yes	Not possible
Simulated Annealing[2]	$S = \sum_{i=0}^{N} \left((Ni_{ov} / Ni_t) x100 \right)^2$	Short	Yes	High

Table 3.4.1-1: Characteristics and performance of the tested algorithms. A**1:** Scoring function based on achieving zero signal overlap (N_i: number of fragments in the mixture). A**2:** Scoring function based on achieving minimal signal overlap (N_{ov}: number of overlapped signals of compound i. N_t: total number of signals of compound i).

Generating virtual fingerprint libraries

We modeled *virtual fingerprints* using parameter values to mimic the fragment fingerprints obtained experimentally from our in-house library. A virtual fingerprint comprises a series of start and end points delimiting various peak regions. The number of peaks per compound, and the position and width of each peak, are each randomly established according to a Gaussian function. In this procedure, a Box-Muller transform is used to generate standard, normally distributed random numbers: $N(\mu,\sigma^2)$. Different mean values and variations are selected in each case. The number of peaks and the peak width follow a normal distribution of *N(8,4)* and *N(0.1,0.03)*, respectively, whereas the peak position is determined using two combined normals [*N(2,1)* and *N(7.5,1.5)*] with different probabilities, producing three different distributions (strongly aliphatic, strongly aromatic, or balanced). The effect of library size on the solution was tested using four different sizes: 500, 1000, 3000 and 5000 fragments. The effect of number of fragments per mixture was also assessed for each distribution-size combination: 5, 8, 10, 15 and 20 fragments.

RESULTS & DISCUSSION

Translating NMR data into fingerprints All computational projects demand careful preparation of the input data, whose quality dictate the quality of the results. Therefore, the first issue we tackled in this project was to translate NMR spectra into a meaningful, computer-readable format. We chose to directly use the NMR spectra that had already been generated in the setup and quality control of our in-house library; thus, we did not have to perform any additional experimental work. The translation was performed using a modified Varian script for signal integration. The script simply integrates, after automatic processing, narrow regions around each signal, and then creates an ASCII file for each spectrum. This file comprises a collection of segments defined by the starting and ending chemical shifts of each integration zone. Thus, each fragment's spectrum is an ensemble of different segments of signals and surrounding space. We refer to the entire ensemble as a *fingerprint*.

The raw data were adjusted for the subsequent calculations in two steps: firstly, signal regions common to all the spectra were defined (*i.e.* signals from H_2O, DMSO, and *t*-butanol [internal standard]), and all segments that overlapped with these regions were partially or completely deleted; secondly, since the size of each segment generated by the script was larger than that required to clearly separate the signals from each other, the size of all segments was reduced by 50%. In fact, the size of the necessary zones varied from 30% to 90% of the size of the zones defined by the script. Therefore, we chose a value of 50%, as it still enables separation of all signals (*i.e.* it is above the full width at half-maximum of two regions that could have been reduced to only 90% of the size determined by the script). Figure 3.4.1-1A shows several overlaid ^{1}H-NMR spectra corresponding to individual fragments; Figure 3.4.1-1B shows the fingerprints of the corresponding compounds, employed to design fragment mixtures that would show nearly zero signal overlap; and Figure 1C shows the actual ^{1}H-NMR spectrum of the mixture of fragments studied in the same NMR tube.
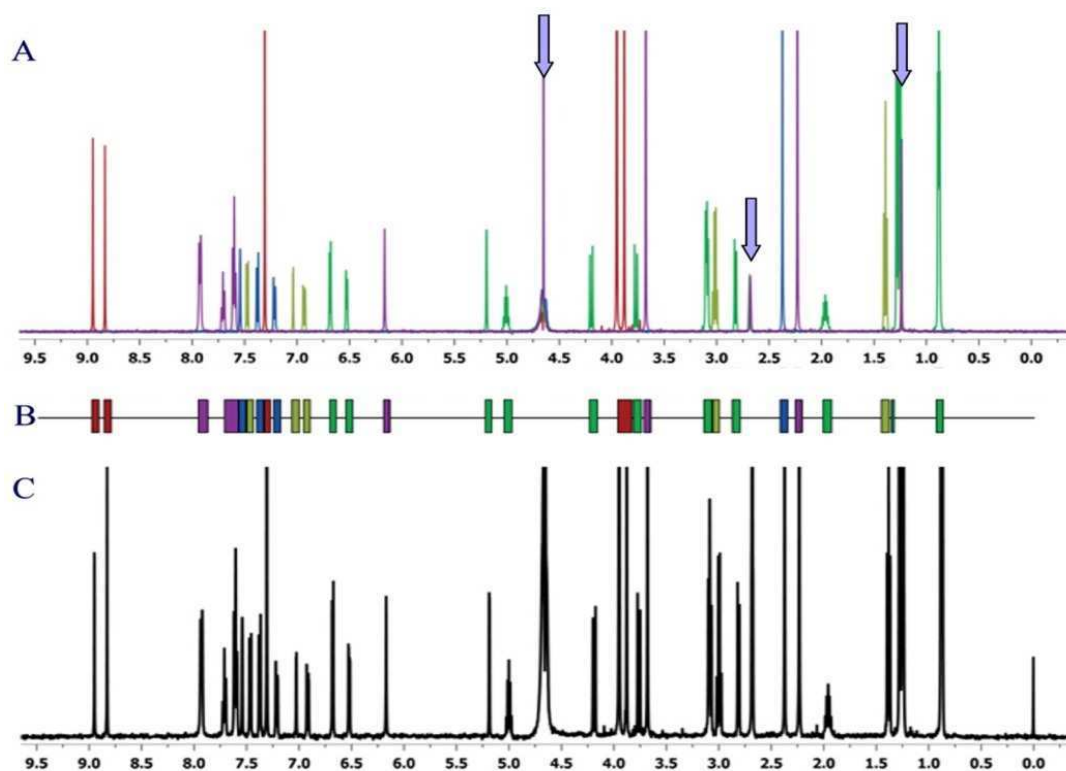
Figure 3.4.1 A: Overlaid 1H-NMR spectra of five different fragments (1 mM in sample buffer: 50 µM phosphate buffer pH 7.0, 50 µm NaCl, 3% DMSO-d6), recorded at 37 °C and 500 MHz. The arrows indicate residual peaks from H2O, DMSO and t-butanol (internal standard). B: Fingerprint of an in silico-designed mixture with zero or near-zero signal overlap. C: 1H-NMR spectrum of the five fragments mixed together (500 uM each) under identical experimental conditions as in 1A (the signal at 0 ppm corresponds to DSS).

Algorithm evaluation

Our task of designing mixtures of a non-redundant pool of fragments is a case of the knapsack problem, one of the typical, non-deterministic polynomial time (NP-complete) problems widely described in the literature.[6,7] To solve this problem, we examined four different algorithms for preparing five-fragment mixtures that would have zero signal overlap. Among the deterministic methods, the greedy algorithm could group only 60% of the compounds into mixtures of five without causing signal overlap; when it grouped the remaining 40% into mixtures of five, these exhibited strong overlap. The backtracking algorithm performed similarly. Owing to their relative simplicity and speed, greedy methods are advantageous over other deterministic algorithms such as backtracking. Backtracking is a refined brute force approach: it systematically searches for a solution to a problem among all available options, which in our case made calculation of a final solution nearly impossible, because the number of compounds demanded long calculation times. However, as a partial solution, the backtracking algorithm was able to group 75% of the compounds into five-fragment mixtures without overlap. Analysis of the remaining 25% of compounds revealed that they comprise difficult compounds that have many signals located in the crowded aliphatic and aromatic regions. In light of this result, we realized that we needed to define a new scoring function based on minimal overlap

for all the fragments from the library, rather than on zero overlap for only some of these fragments. Another problem highlighted by the backtracking algorithm is that the required calculation time grows exponentially with the size of the data set. Therefore, we decided to explore stochastic algorithms, which we expected would be much faster and would enable performing multiple calculations in parallel. Genetic algorithms failed to provide a coherent solution, due to problems in the breeding and mutation steps, which caused a loss of compounds in each evolutionary cycle. Contrariwise, the SA algorithm (Monte Carlo-Metropolis) yielded good results in a reasonable time. It can be run with libraries of up to 1,000 compounds in less than 5 minutes, thereby enabling parallelization of massive independent runs. Unexpectedly, we found that temperature had a negligible effect on the results, as we describe below.

Optimization of mixtures from the in-house fragment library

We performed 100 independent runs of SA for the 342 fragments from our in-house library that passed quality control. One hundred random solutions were calculated in parallel by clustering the fragments into mixtures having the same number of components. The results are shown in Figure 3.4.1-2A. Based on the assumption that each component contributes equally to the final score of the mixture, the random mixtures showed an average signal overlap of 44% per compound. After optimization by SA, the average global signal overlap per compound was reduced to only 2% (an improvement of 42%). The in silico results were confirmed by mixing the appropriate fragments into mixtures whose ^{1}H-NMR spectra were then recorded. A peak list was generated for each mixture and compared to the fingerprint of each fragment. After the 50% reduction in segment size, more than 92% of the peaks coincided with the regions corresponding to the fingerprints (Figure 3.4.1-2B). Based on these results, we concluded that the in silico fragment mixtures corresponded to the real ones.
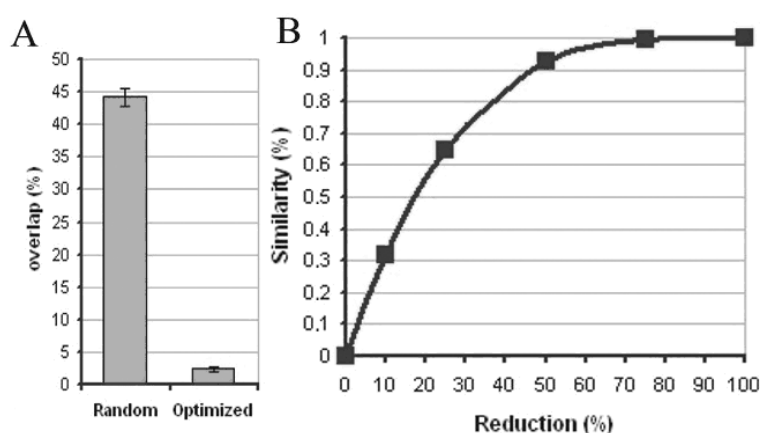

Fig 3.4.1-2: Real library results. A: Comparison of signal overlap between randomly assigned fragment mixtures and simulated annealing (SA)-optimized fragment mixtures. B: Effect of the reduction of the size of the fragments. In x-axis per cert reduction of the size of all segments. In y-axis per cent similarity between the number of peaks in each segment before and after the reduction.

Evaluating algorithm adaptability with virtual libraries

Given the size of our in-house library, we were unable to study scalability and other variables that could affect the SA algorithm. Therefore, we designed a virtual fingerprint-generator to produce *virtual libraries*. A total of twelve different libraries, differing in global size and peak distribution, were generated. Library sizes of 500, 1000, 3000 and 5000 fragments were chosen. Three different distributions were selected, representing libraries whose compounds' NMR signals were either strongly aliphatic, strongly aromatic, or balanced. Figure 3.4.1-3 shows a representative example from one of these virtual libraries.
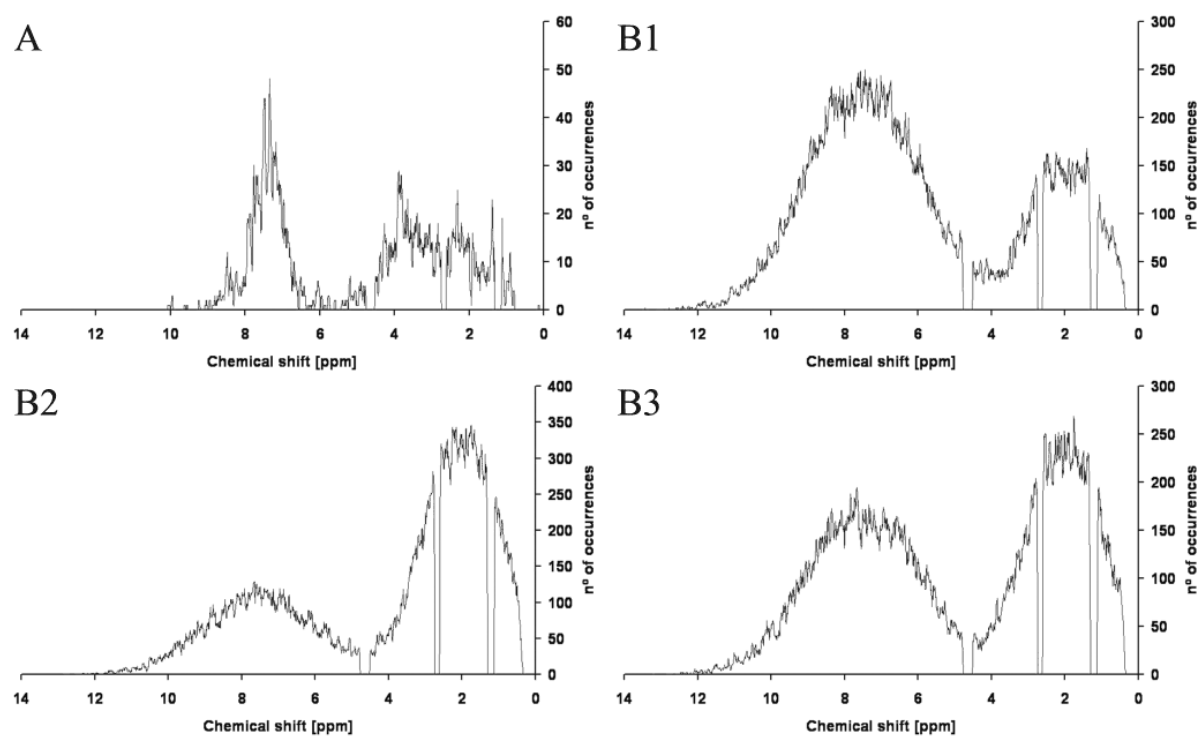
Fig 3.4.1-3 A: 1H-NMR peak distribution among the 342 fragments of the in-house library. B(1-3): Analogous plots for the virtual libraries of 3000 fragments having the following signal density distributions: 50% aromatic, 50% aliphatic (B1); 70% aromatic, 30% aliphatic (B2); and 30% aromatic, 70% aliphatic (B3).

The twelve virtual libraries were then used to test the ability of the SA algorithm to reduce the global signal overlap (compared to that of a random solution) for mixtures of five to twenty fragments (Fig. 3.4.1-4).
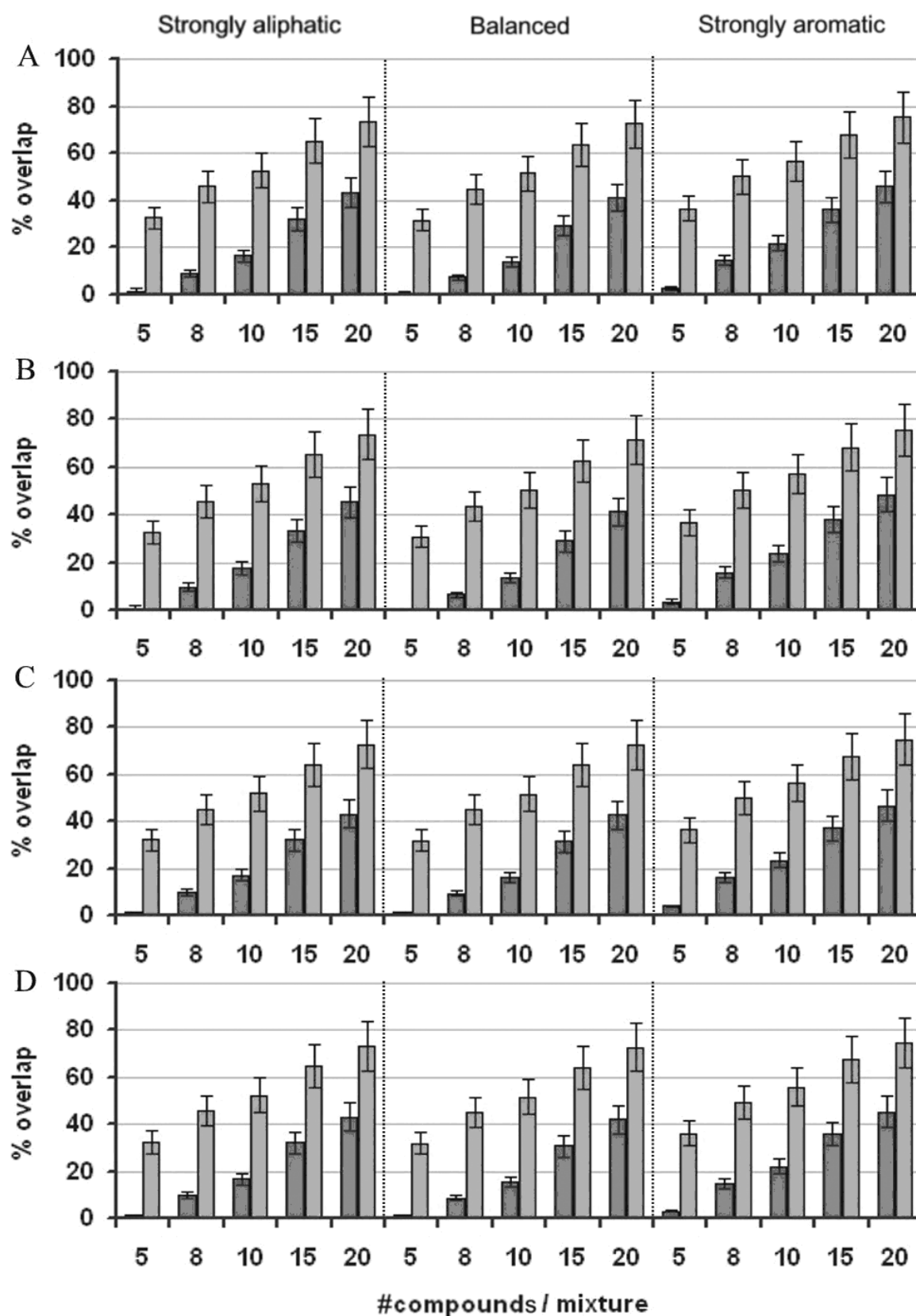
Figure 3.4.1-4: Results for the virtual library of 3,000 fragments. Comparison between randomly constructed (orange) and optimized (blue) libraries. The following parameters and values were tested: library size (A-D: 500, 1000, 3000 and 5000, respectively); peak distribution (strongly aliphatic, strongly aromatic, or balanced); and number of compounds per mixture (five, eight, ten, fifteen or 20). For each set of conditions the SA algorithm was run one hundred times independently.

For each set of library size, peak distribution, and mixture size, the SA algorithm was run independently one hundred times, and the results were compared to those from random clustering of fragments into mixtures. Interestingly, library size and peak distribution had no significant impact on the SA algorithm. In contrast, mixture size affected the global signal overlap for both the random mixtures and the optimized mixtures. Whereas the SA algorithm achieved a global signal overlap close to zero percent for five-fragment mixtures, the overlap increased with increasing mixture size. However, in the random mixtures, for each set of conditions, the SA algorithm still reduced signal overlap by 35%, regardless of the mixture size.

For mixtures of 20 fragments, SA reduced the signal overlap to the level corresponding to a randomly assembled mixture of eight fragments. Although screening of mixtures containing more than eight compounds is not currently common practice, future improvements in NMR sensitivity may enable this for cases in which signal overlap is low enough that deconvolution is not required.

Effect of temperature on SA

Temperature is normally an important variable in the SA algorithm: during optimization it controls the acceptance of uphill moves, thereby avoiding local minima. To determine the best temperature value for the algorithm, the impact of temperature on the capacity to reduce signal overlap was tested with equally distributed libraries of 500 fragments each. The temperature value ranged from 0, at which no uphill moves are allowed, to 25,000, at which nearly 100% of uphill moves are allowed (because this value is larger than a fully simultaneously overlapped score of two compounds). Surprisingly, modifying the temperature had no beneficial effect on the algorithm, independently of the conditions tested (Figure 3.4.1-5). A possible explanation for this observation is that each optimization step in the SA algorithm is not directly linked to the previous one, and consequently, each of these steps can have highly variable effects. Therefore, late uphill moves are translated into high increments in the scoring, which in turn are reflected as a loss of time.
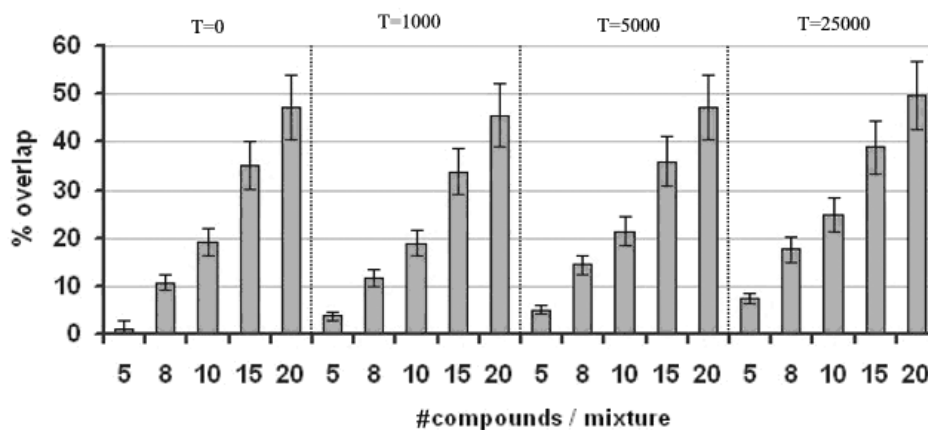

Figure 3.4.1-5: The effect of temperature on signal overlap. The experiment was performed with a virtual library of 500 compounds and peak distribution of 50% aromatic, 50% aliphatic.

CONCLUSIONS

We have devised a powerful method for NMR screening of mixtures of fragments that entails translation of NMR spectra into a computer-readable format that we call a fingerprint. Among the various algorithms that we assessed to solve the problem of signal overlap, the SA algorithm offered the best optimization. As proof of concept, we used this algorithm to design five-fragment mixtures from our in-house library that showed an average signal overlap of only 2%, compared with 44% for randomly designed mixtures of the same size.

We then conceived virtual fragment libraries to evaluate performance of the SA algorithm based on peak distribution (relative aliphatic or aromatic character of the library), scalability (i.e. library size), and temperature. Our results suggest that the method is amenable to libraries of any size or nature. Furthermore, modifying the temperature had no effect on signal overlap.

We anticipate that the described method will improve the efficiency of NMR-based fragment screening by simplifying detection of binding compounds, as it does not require any sophisticated computational hardware and, in the case of compounds whose NMR spectra are already available, does not require any additional experimental work.

(1)   Hajduk, P. J.; Greer, J. *Nat Rev Drug Discov* **2007**, *6*, 211.
(2)   Dalvit, C. *Drug Discov Today* **2009**, *14*, 1051.
(3)   Lepre, C. A.; Moore, J. M.; Peng, J. W. *Chem Rev* **2004**, *104*, 3641.
(4)   Mercier, K. A.; Powers, R. *J Biomol NMR* **2005**, *31*, 243.
(5)   Fejzo, J.; Lepre, C. A.; Peng, J. W.; Bemis, G. W.; Ajay; Murcko, M. A.; Moore, J. M. *Chem Biol* **1999**, *6*, 755.
(6)   Martello, S.; Toth, P. *Knapsack problems : algorithms and computer implementations*; J. Wiley & Sons: Chichester ; New York, 1990.
(7)   Garey, M. R.; Johnson, D. S. *Computers and intractability : a guide to the theory of NP-completeness*; W. H. Freeman: San Francisco, 1979.
(8)   Kirkpatrick S, G. C., Vecchi MP *Science* **1983**, *220*, 671.
(9)   Belda, I.; Madurga, S.; Tarrago, T.; Llora, X.; Giralt, E. *Mol Divers* **2007**, *11*, 7.
(10)   Belda, I.; Madurga, S.; Llora, X.; Martinell, M.; Tarrago, T.; Piqueras, M. G.; Nicolas, E.; Giralt, E. *J Comput Aided Mol Des* **2005**, *19*, 585.

3.4.2 Automatic evaluation of NMR screening data

In the previous chapter we described the development of software to calculate fragment mixtures for NMR based screening. As elaborated above, the approach is straightforward, fast and permits the generation of mixtures with very low signal overlap. This should simplify the detection and evaluation of binding compounds that are contained in these mixtures.

After screening the first part of our own library, we indeed found that no additionally experiments were required for deconvolution since all compound signals could be clearly assigned (chapter 3.5.1). But we also found that even without signal overlap the analysis of NMR screening data is a rather tedious and time consuming process. We therefore wondered if the detection and evaluation of binding compounds had been simplified to a level where software would be able to automatically analyze the data from such mixtures with minimized signal overlap. If feasible this would have greatly reduced the time required for the analysis of NMR screening data.

We therefore decided to develop a tool that could automatically analyze the NMR based screening data that we were about to obtain for the second part of our library. This part consisted of 125 fragments in mixtures of five compounds with very low signal overlap. During the setup of the library [1]H-spectra of the individual compounds and [1]H-spectra with and without CPMG filter with 20 and 400 ms duration were acquired. For the screening identical samples of the mixtures were prepared, but contained an additional 10 uM VEGF. During the screening we acquired [1]H-spectra with and without CPMG filter of 20 and 400 ms durations and an STD spectrum. The nature of these experiments will be discussed in more detail in chapter 3.5.2 In brief, the appearance of STD signals of a fragment indicate its ability to bind to the protein. However, if the fragment contains ^{1}H NMR signals close to the on-resonance frequency (0.09 ppm) that we used during the STD experiment this may lead to false positives. In CPMG spectra the applied filter leads to a decrease of ^{1}H signal intensities. This is especially strong for compounds that bind to proteins. Therefore, if two spectra with CPMG filter are acquired in presence and absence of a protein, ligands exhibit reduced signal intensities in the sample with protein. This effect is very strong for a filter of 400ms. Contrary a filter of 20ms should mainly remove protein ^{1}H signatures from the protein. The integration of mixture generation during the setup, screening and the generation of NMR data is summarized in figure 3.4.2-1.
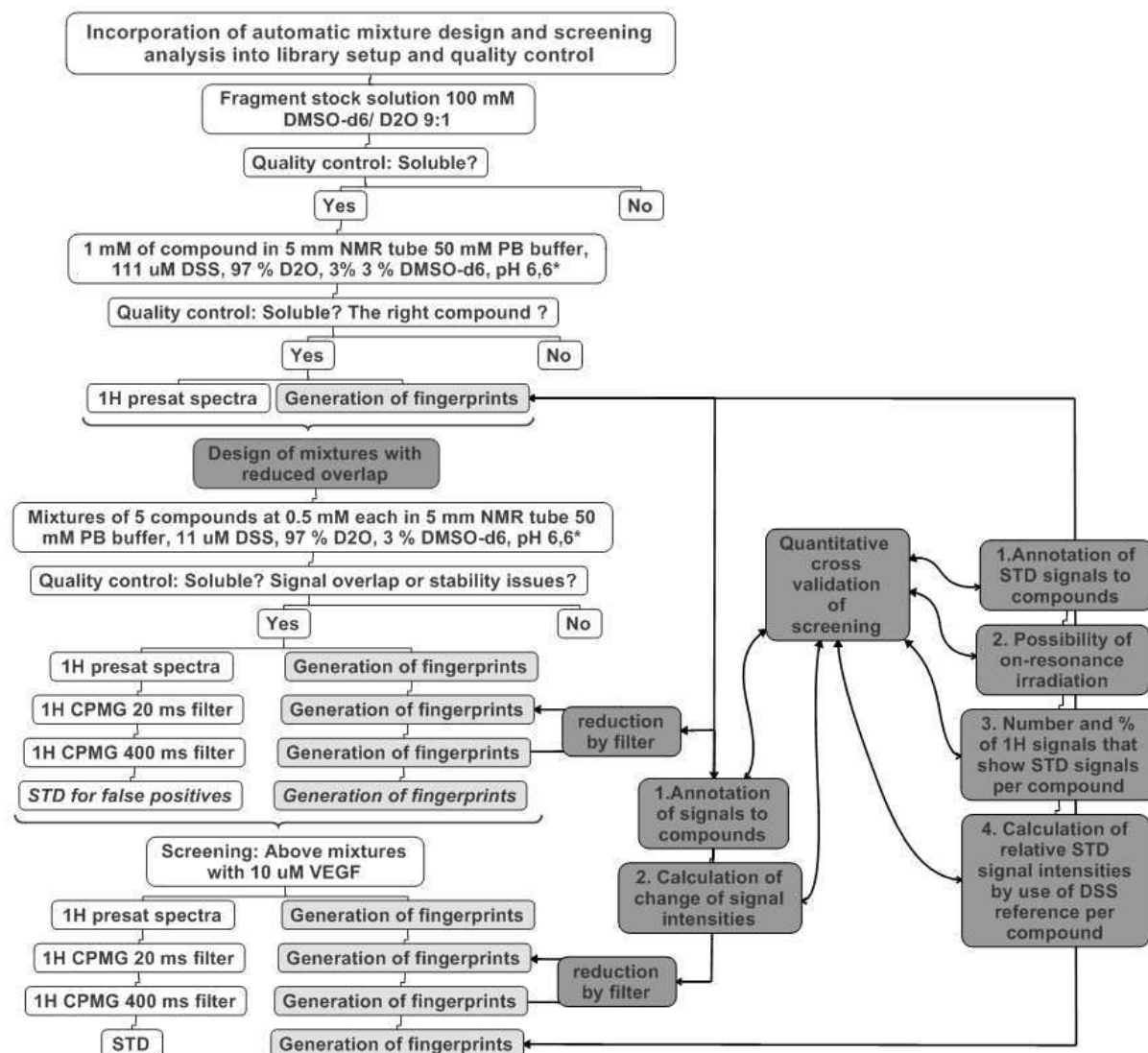
Figure 3.4.2-1: Workflow for the incorporation of library setup, automatic mixture design and analysis of NMR based screening data. Dark grey marks tasks performed automatically by our computational tools. Light grey marks the generation of fingerprints as input for our tools.

A key aspect in the workflow was the generation of fingerprints that were used as computational input. We used a modified automatic Varian script developed by Dr. Miguel Feliz for the mixture design of the first library part. The quality of the generated data was sufficient for this task. However a preliminary review of our screening data showed only small differences in signal intensities between CPMG spectra in absence and presence of protein and STD signals with low S/N ratios. Since the Varian script performed automatic phasing and base line corrections the range or generated errors were in the order of observed effects induced by the protein and would be therefore highly misleading. This problem was solved by the development of a script for MestreNova by their customer support. This allowed the semiautomatic generation of fingerprints. Up to 100 spectra could be loaded and processed in parallel by sub sequential application of fourier transforms, phasing, baseline correction, peak picking and integration. The advantage of this approach was that all spectra could be revised after each step and errors corrected while still being very fast. Finally,

using the MestreNova script the corresponding fingerprints were generated for all spectra. The fingerprints were used as computational input to detect which compounds might bind to VEGF. The tool for this analysis was developed in collaboration with PhD student Xavier Arroyo. The tool first normalized all fingerprints in respect to the DSS signal around 0.0 ppm. The analysis of STD spectra was achieved by the following scheme: The integration ranges in the fingerprint files of individual compounds were used as boundaries and tested if they contained signals in the STD spectrum. This allowed an annotation of how many compound ^{1}H signals gave rise to STD signals. If one of these STD signals was detected under 1.1 ppm, the threshold that we defined for the possible appearance of direct on-resonance irradiation the compounds was annotated as an potential false positive. The intensity of STD signals was assessed by calculating the ratio to the internal DSS signal appearing around 0.0 ppm and then compared to the ratio of the highest ^{1}H signal in the matching integration range of the individual fragment. Since we used identical DSS concentrations for all screening samples and individual compounds this comparison could be used as an internal reference independent of magnetic field changes. Comparing STD signals to parent signals allowed a powerful assessment of the relative binding strength of compounds in the affinity range usually observed in fragment screening. This is not the case if only the STD signal intensity would be considered, since compounds or functional groups with multiple equivalent protons give rise to higher STD signals that overestimate their importance (3.4.2-2). The proper analysis of this is an especially time consuming task if performed manually.
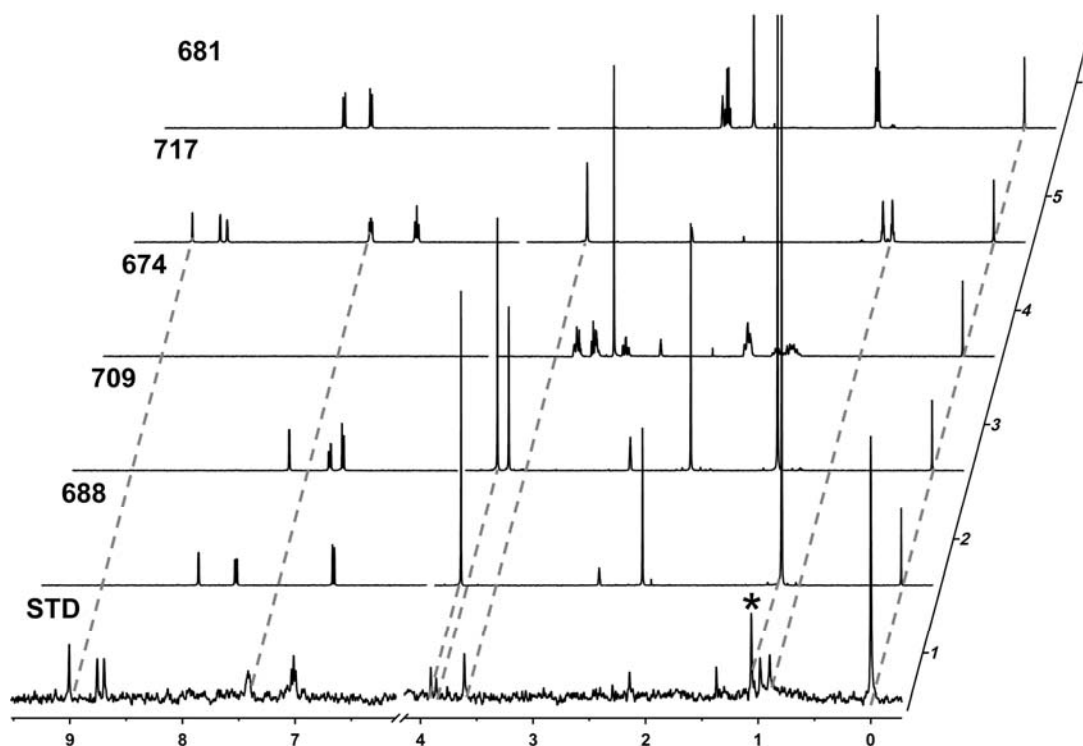

Figure 3.4.2-2: Representative example for mixture 86: STD spectrum in comparison to [1]H-spectra of the five individual compounds. Dotted lines exemplify the annotation of STD signals to compound signals. * marks the STD signal of the highest intensity but low relative intensity compared to the parent signal.

CPMG spectra were analyzed by calculating the intensity decrease for each ^{1}H signal between the 20 ms and 400 ms filtered spectra. This was first performed for the sample with and without protein and then the ratio of both calculated. Ideally an identical decrease (between 400 and 20 ms) leads to a value of 1.0 indicated no interaction; or on the contrary, a complete disappearing of the signals in presence of protein with the 400 ms filter resulting in a value of 0.0 indicating a strong interaction. As mentioned earlier, this behavior was not present for a single compound, as all signal decreases were very modest. We decided therefore to include only ^{1}H signal that we could be tracked automatically in all four CPMG spectra. Finally the signals were assigned to the individual compounds as this was performed for the STD spectra.

The output of the computational analysis is an ASCII file which summarizes the acquired data and the calculated outcome. Table 3.4.2 is an annotated example of this output for the STD results of mixture 86 which is also depicted in figure 3.4.2-2. The analysis showed that the strongest STD signal at 1.061 ppm belongs to compound 688. However this signal corresponds to a high intensity methyl and has therefore a low relative intensity. The best compound appeared to be 717. Is has by a factor of ten higher relative STD signal intensities as can be seen in the summary. Further all "boxes" (integration ranges) are occupied by STD signals. The table shows that the compounds exhibits even more STD signals than boxes. This is caused by the high STD signal intensities which give rise to a triplet in the integration range around 7.0 ppm. If required we could change the code in the future to recognize such cases. The table annotates compound 717 further as a potential false positive. This is based on the signals close to 1.0 and 0.9 ppm which were under the defined threshold. 717 should therefore not be considered or retested with an appropriate on-resonance frequency.

86							
refPeak	0	114.863		Position and intensity of DSS reference			
OnResW		-> The sample may contain false positives at the following positions:					
	0.983	17.136	0.149186				
	0.897	18.756	0.16329				
OUT	0	15	-> All STD signals were assigned to compounds				
IN	15	15					

Mixture number	Fragment ID	STD signals	"Boxes"	STD signal (ppm)	STD signal intensity	relative intensity	Warning?
86	717	10	8	9.006	23.88	0.431724	
86	717	10	8	8.756	17.488	0.333857	
86	717	10	8	8.696	17.613	0.414255	
86	717	10	8	7.416	11.881	0.262652	
86	717	10	8	7.031	12.442	0.180916	
86	717	10	8	7.013	18.48	0.268713	
86	717	10	8	6.994	12.417	0.180552	
86	668	2	6	3.909	12.681	0.028741	
86	709	3	7	3.861	8.105	0.019567	
86	717	10	8	3.61	18.774	0.127842	
86	709	3	7	2.143	10.722	0.026457	
86	709	3	7	1.371	12.978	0.015493	
86	668	2	6	1.061	37.196	0.034736	
86	717	10	8	0.983	17.136	0.224197	onResW
86	717	10	8	0.897	18.756	0.241339	onResW

SUMMARY	Fragment ID	STD signals	"Boxes"	Av. rel. STD intensity	Derivation	Warning?	
86	668	2	6	0.031738	0.002997		
86	674	0	7	0	0		
86	681	0	5	0	0		
86	709	3	7	0.0205056	0.004525		
86	717	10	8	0.266605	0.094893	onResW	

Table 3.4.2: Annotated example of ASCII report of STD results fro mixture 86 (see also figure cb). Interesting features are highlighted in grey: STD signal with highest intensity (1.061 ppm) but low relative intensity. Compound 717 shows high coverage of "boxes" (= integration ranges) and high relative signal intensities. However it is annotated as potential false positive due to on-resonance irradiation.

The ASCII files for the STD and CPMG based experiments were imported to excel for further analysis. Figure 3.4.2-3 shows the plot of the average relative STD signal intensities versus the number of occupied integration ranges (or "boxes"). Compounds with were annotated as potential false positives are depicted as empty squares. The graph can be separated in four quadrants.

Due to the general low affinity of fragments the relative STD signal intensity can be used as rough estimate of their binding affinity. The number of occupied boxes could be interpreted as additional indicator for binding strength, since strong binding fragments require positive interactions of the majority of their atoms. Or it could be used as indicator for the confidence that the fragment is a real binder and the detected STD signal are no artifacts of the experiment, signal overlap of the automatic analysis. Both parameter together cam be use to dissect figure 3.4.2-3 in four quadrants: The upper

right with good affinity / high confidence fragments. The upper left with good affinity / low confidence fragments. The lower right with low affinity / high confidence fragments. And finally fragments that probably do not bind in the lower left.
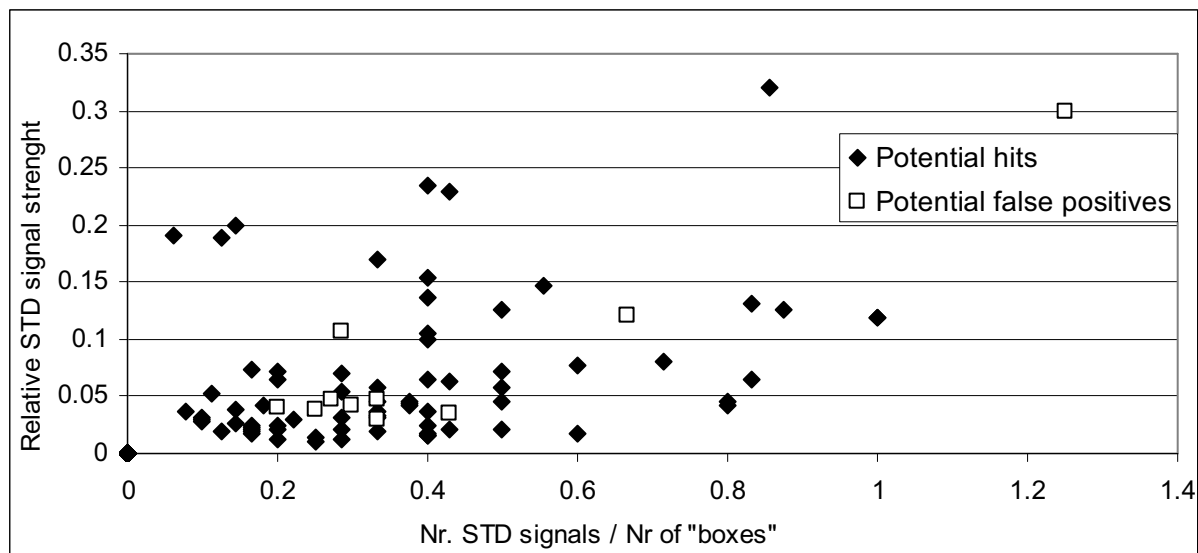

Figure 3.4.2-3: Results of automatic evaluation of STD NMR screening data of 125 fragments.

If NMR is used as screening tool for fragments often several experiments are performed with the same sample to probe for different aspects of fragment binding. In this exercise we were relying on the STD and CPMG experiment. The CPMG experiment has a lower sensitivity than the STD experiment but does not rely on saturation of the protein. Both experiments can and should be used for cross validation. This is a straightforward process once all information have been incorporated into excel. For this purpose we summarized the STD data into a normalized vector product (product = rel. STD intensity *x* occupied boxes). This value was plotted against the normalized decrease in the CPMG experiment. The resulting vales were plotted in figure 3.4.2-4. The correlation between both values is relative low. This is especially true for fragments with an STD vector value under 0.1 and should be considered as noise in the CPMG experiment. This is probably due to the weak interaction between protein and fragments as discussed in chapter 3.5, and the reduced sensitivity of the CPMG experiment. Interestingly the best scored fragment by the STD experiment has a very low score in the CPMG experiment. This fits the annotation of a false positive in the first experiment.
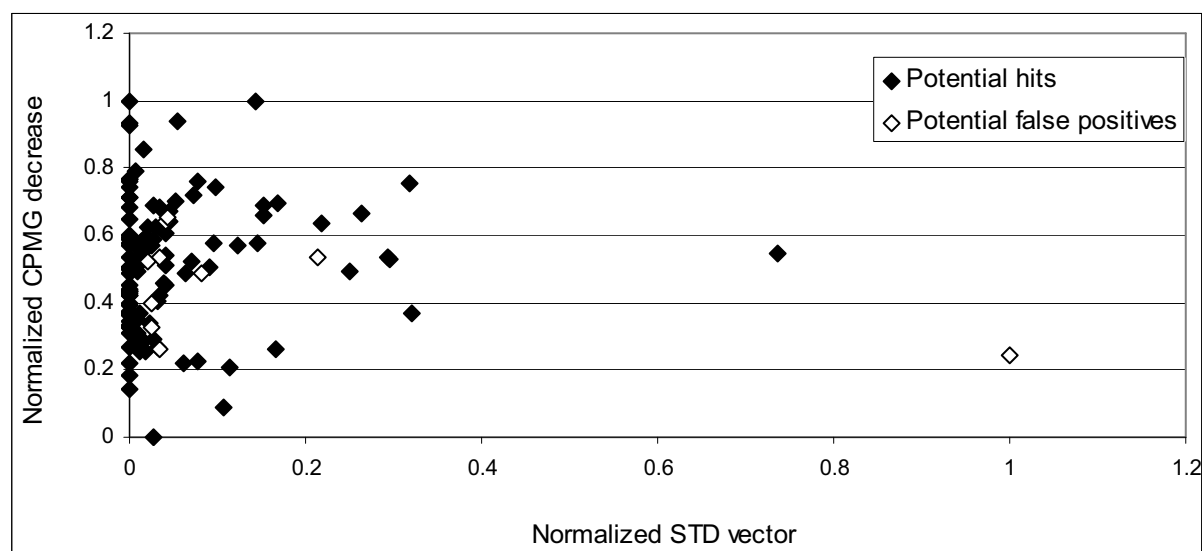
Figure 3.4.2-4: Results of cross validation of STD- and CPMG derived data of 125 fragments.

In this chapter we explored the feasibility of automatic evaluation of NMR screening data. We faced three key issues: 1. The transformation of NMR spectra into a computer readable format. 2. The annotation of NMR signals in mixtures to individual compounds. 3. The computational assessment of ligand binding based on NMR observables.

The first issue was solved with the use of a script for MestreNova that allows the generation of fingerprint files in a semi-automatic fashion. The quality of the transformation was sufficient for STD signals with low intensity. The second issue was simplified by the low signal overlap in our mixtures and small chemical shift differences between compound reference spectra and mixture spectra. This made it straightforward to annotate the mixture signals computationally. Should we screen proteins in significant different conditions this may require more complicated algorithms. Regarding the third issue the fingerprints contained all required information to assess ligand binding. It was only necessary to define in which way the data should be analyzed to allow a quantitative evaluation of each fragment. Based on the nature of the STD experiment we proposed the relative STD signal intensities as estimate for affinity and the number of degree of compound signal that rive rise to STD signals as mixed indicator for affinity and confidence of binding. We also performed the automatic evaluation of CPMG data although the correlation to the STD Data was not very high.

Based on our results the automatic evaluation, al least of mixtures with optimized signal overlap, seems to be highly feasible. We believe that it will facilitates the analysis of NMR based fragment screening data, because it improves the speed and the quality of the analysis compared to a manual analysis. However the capability and limitations of our automatic analysis require further testing. For this purpose it would be useful to have solid NMR data acquired from the fragment screening of classical druggable targets.

3.5 Molecular recognition at the VEGF surface: ligand based approaches

3.5.1 NMR based screening of the first part of the Library

In the previous chapters we described the design and setup of our fragment library. Further we developed mixtures with minimized signal overlap that should facilitate the identification and evaluation of hits. These steps were necessary to start the screening of our proprietary fragment library for compounds that bind to the VEGF protein-protein interface.

As described in chapter 1.3 many techniques are used for fragment screening. NMR is an omnipresent technique in this field since the pioneering work of Fesik, Hajduk and coworkers in the late 90s.[1-3] They applied protein-observing experiments to detect ligand binding. This has the advantage to provide information regarding the binding site of the ligand. Today the majority of researcher applies ligand based methods for screening, because of higher throughput and sensitivity.

We decided to use the STD NMR for screening. This method was introduced in 1999 by Prof. Bernd Meyer in two seminal papers.[4,5] It belongs now to the widest applied NMR experiments for fragment screening in academia and industry. It is described in detail in chapter 3.1.

Simplified the method is based on the intermolecular transfer of magnetization, between a protein and a small molecule. The protein is magnetically saturated in the so called on-resonance experiment and the magnetization transferred to compounds if binding occurs. The final STD spectrum is obtained by interleaving experiments with and without protein saturation (on- and off-resonance experiment). When this two spectra are subtracted only signals from compounds that binds to the protein remain. The advantage of STD based screens is that they are very robust regarding false positives and negatives. The only origin of false negatives are extremely slow exchanging ligands which is generally not the case with fragments. False positives can arise by two means: Either by direct irradiation of ligand signals or by the presence of dynamic aggregates. The second can be excluded by repeating the experiment in the absence of protein. If STD signals are still present then they arise from ligand aggregates. The first source of false negatives arising from ligand irradiation and can be avoided if the frequency of the saturation pulse is set far enough away from the aliphatic signals of the ligand. But departing too much from this region will result in poor protein saturation and therefore reduced sensitivity of the experiment.

Baicalin was previously identified by our laboratory to bind at the receptor-binding surface of VEGF and can therefore serve to optimize the parameters of the STD experiment. For this purpose we used a sample consistent of 500 µM Baicalin and 10 µM VEGF in deuterated buffer. With this sample the same STD experiment was recorded with various resonance frequencies. A frequency of 0.37 ppm gave the highest STD signal. However this value proved to be too close to aliphatic signals of compounds for the screening. Instead a value of 0.09 ppm was used, that appeared to be a good

compromise between a relative low appearance of direct ligand irradiation while maintaining the sensitivity of the experiment.

Using an on-resonance frequency close to 0 ppm had the additional advantage that we could use DSS as a reference compound for the STD experiment. Because the DSS signal was close to 0 ppm we created an on-resonance irradiation artifact that led to an STD signal close to 0 ppm. This is an outcome that we try to avoid under normal circumstances. However at 0 ppm this signal did not interfere with the signals of other compounds and offered a valuable reference signal to phase, align and to compare relative intensities of STD signals.

Further optimization of the STD experiment led to a mixing time of 20 ms for the removal of protein signals in the spectrum and a saturation time of 2.5 s to amplify the STD information. Off resonance frequency of 35 ppm and a relaxation delay time of 2 s was used resulting in a total delay between two on-resonance scans around 11 seconds. The number of scans was set to 1024 for the STD experiment which required about 90 minutes per sample. This relative long experimental time was selected to efficiently use the whole 24 hours with the 9 sample changer that was available and to improve the signal to noise ratio in the experiment.

Originally the experiment was optimized for 37°C instead of 25°C with the idea that this is the physiological relevant temperature. However in practice this led to reproducibility issues with the Baicalin sample and the first mixtures that were screened at this temperature did not show the clear signature of binding compounds. Interestingly changing to 25°C for the optimization and screening improved reproducibility and did show signature of binding compounds in the first mixtures that were explored at this temperature.

Finally we were interested to identify compounds that bind to the protein-protein interface of VEGF. For this purpose we used the peptidic ligand v107 that was previously explored in collaboration with Andrey Dyachenko. The binding zone of this peptide covers at least some part of the interface that is involved in the interaction of VEGF with its receptors. In theory it can be used for a competition STD experiment. If a mixture shows the signature of a potential ligand the peptide can be added to this mixture and the experiment repeated. If the STD signals disappears or decrease this would indicate competition between the peptide and the ligand at the protein-protein interface. If no changes in signal intensities are observed this indicates that the ligand and the peptide do not share a common binding site or allosterism.

To evaluate this experimentally we used the previous sample of 500 µM baicalin (Kd = 3.7 mM) and 10 µM VEGF. To this sample 40 µM of v107 (Kd = 1 µM) were added, which corresponds to a two fold excess over binding sites. As expected this led to a decrease of the STD signals of baicalin (Figure 3.5.1-1). The decrease was significant only for the two highest STD signals at 7.55 ppm (30%) and 6.75 ppm (60%). Since I was planning to perform the STD experiment during the screening with an

increased scan number over 2.5 fold compared to this preliminary study I assumed the changes were sufficient to be distinguishable against the noise during the screening.
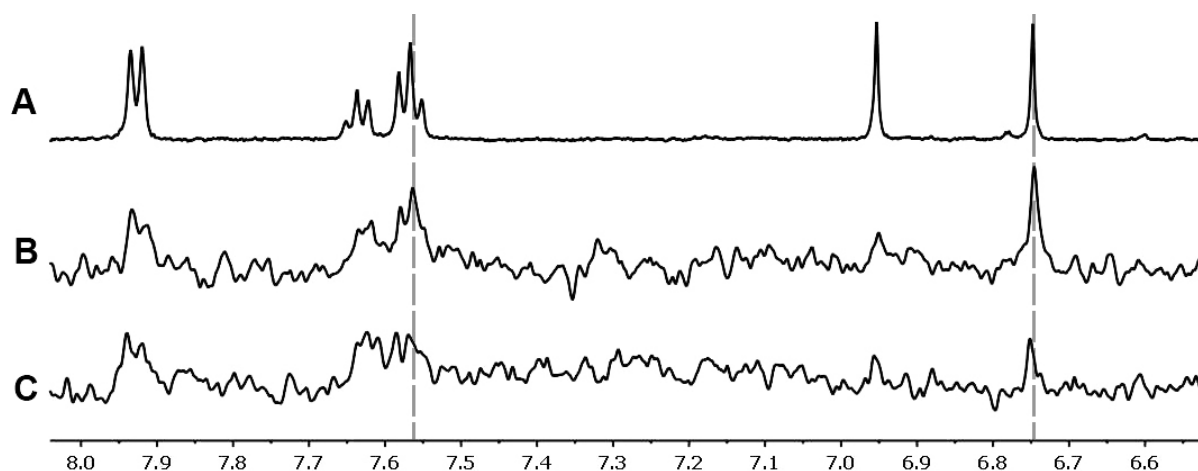

Figure 3.5.1-1: A: Aromatic region of a 1H-spectrum with water suppression of 500 µM Baicalin with 10 µM VEGF at 25℃. B: STD spectrum of the same sample. C: STD spectrum of the same sample after addition of 40 µM v107. The dashed lines mark the drop of intensity of the two highest STD signals after the addition of v107 by 30% (7.55 ppm) and 60% (6.75 ppm). Both STD spectra are recorded with 384 scans.

At this point we optimized sufficiently the screening parameter for the STD experiment and the competition experiment so that we could begin with the screen of the first part of our fragment collection.

As decided during the setup of our library we screened in mixtures of five compounds.

For each mixture a sample was prepared by adding 16.5 µL of each mixture stock to a protein sample resulting in a final concentration of 500 µM per Fragment, 10 µM of VEGF, 11.11 µM of DSS, a DMSO concentration under 3% and a final sample volume of 550 µL. This corresponds to a ratio of 50:1 between fragment and protein or if we consider the protein-protein interface of VEGF of 25:1 due to its dimeric nature. For the screening we used a perdeuterated 50 mM phosphate buffer with sufficient capacity to neutralize pH changes induced by fragments.

All samples were vortexed in their eppendorf tubes and transferred into 5mm NMR tubes. Of each sample one ^{1}H spectrum with presaturation of the water signal was recorded followed by a STD spectrum with the above described parameters. If the spectrum contained STD signals the corresponding sample was transferred back to its original eppendorf tube and v107 stock (22 mM in DMSO) added to a final concentration of 40 µM. The sample was vortexed again and transferred back to its NMR tube. Then the ^{1}H spectrum with presaturation and the STD spectrum were recorded in the presence of the v107 ligand to test for potential competition at the protein-protein interface. This process was performed for all 81 mixtures of the first part of the library containing 402 fragments. Since at least weak STD signals could be observed in nearly all samples the competition experiment was performed for all mixtures. During the screening no precipitation occurred. Neither

any obvious degradation of compounds when [1]H-spectra from the setup of the mixtures where compared with the spectra recorded during the screening.

The analysis of the NMR data was performed with MestreNova, which allows multiple spectra to be processed in parallel. Spectra were treated with an exponential apodization with a line broadening factor of 3.5 Hz. Automatic phase correction was successful applied to ^{1}H spectra, but failed frequently for STD spectra due to low S/N ratio and broad residual protein signals. Manual phase correction was simplified by the use of the artificial STD signal which originated from DSS close to 0 ppm. This made adjustment of PH0 easy and left only PH1 for trial and error adjustment. Without the DSS signal the phasing of both PH0 and PH1 would not be straightforward since the intensity of the STD signals were often close to the noise. After phase correction a baseline correction was applied to all spectra in parallel and the DSS signal referenced to 0.00 ppm. Compound identification and evaluation was performed by comparing the STD spectra of the mixture against the individual compounds. Figure 3.5.1-2 illustrates this process with a representative sample. The first five lines contain the ^{1}H spectra of the five fragments that were present in the mixture. The spectra share common signals at 4.7 ppm, which originates from residual water, 2.7 ppm, which belongs to DMSO and a signal at 1.25 ppm which belongs to *tert*-butanol which was used as an additional reference. It is obvious that all remaining signals that originate from the fragments do not overlap which facilitates the unambiguous identification of active compounds. This result was typical and no additional experiments were necessary since all compounds could de identified directly.

Active compounds were identified and evaluated manually by assessing the following four points: 1. what is the percentage of signals for each compound that show STD effects?

2. Are these STD signals in average "strong", "normal" or "weak"? 3. Could the STD signals be a cause of direct ligand irradiation? 4. Do the STD signals decrease in the competitive experiment? This evaluation would lead to the following results for the sample in figure 3.5.1-2:

The majority of the STD signals correspond to compounds F511 and F225. F511 shows 3 STD signals while the compound had 4 signals in the ^{1}H spectrum. This corresponded to a percentage of 75%. In comparison to the results of the whole screen the STD signal around 2.5 ppm would be evaluated as strong, around 6.2 ppm as normal and at 7.0 ppm as "weak", which averages to "normal". The STD signal from this compound did not decrease in the competition experiment, nor were they close to the on resonance frequency of 0.09 pm. The summary for this compound would be therefore: 75%, "normal", no competition, no direct ligand irradiation. For compound F225 the evaluation would result in 22%, "weak", no competition, no direct ligand irradiation. F147 would be evaluated as 25%, "weak", no competition, but has the possibility of direct ligand irradiation, which makes this fragment a candidate for a false positive. Since even the *tert*-butanol at 1.25 ppm showed a weak STD signal all fragments with ^{1}H signals under this threshold were considered as possible false

candidates that should be excluded or require further testing. Finally the fragments F558 and F479 do not display any STD signals. The results for this mixture were then incorporated into the MOE database that contained the structure and information about each fragment.
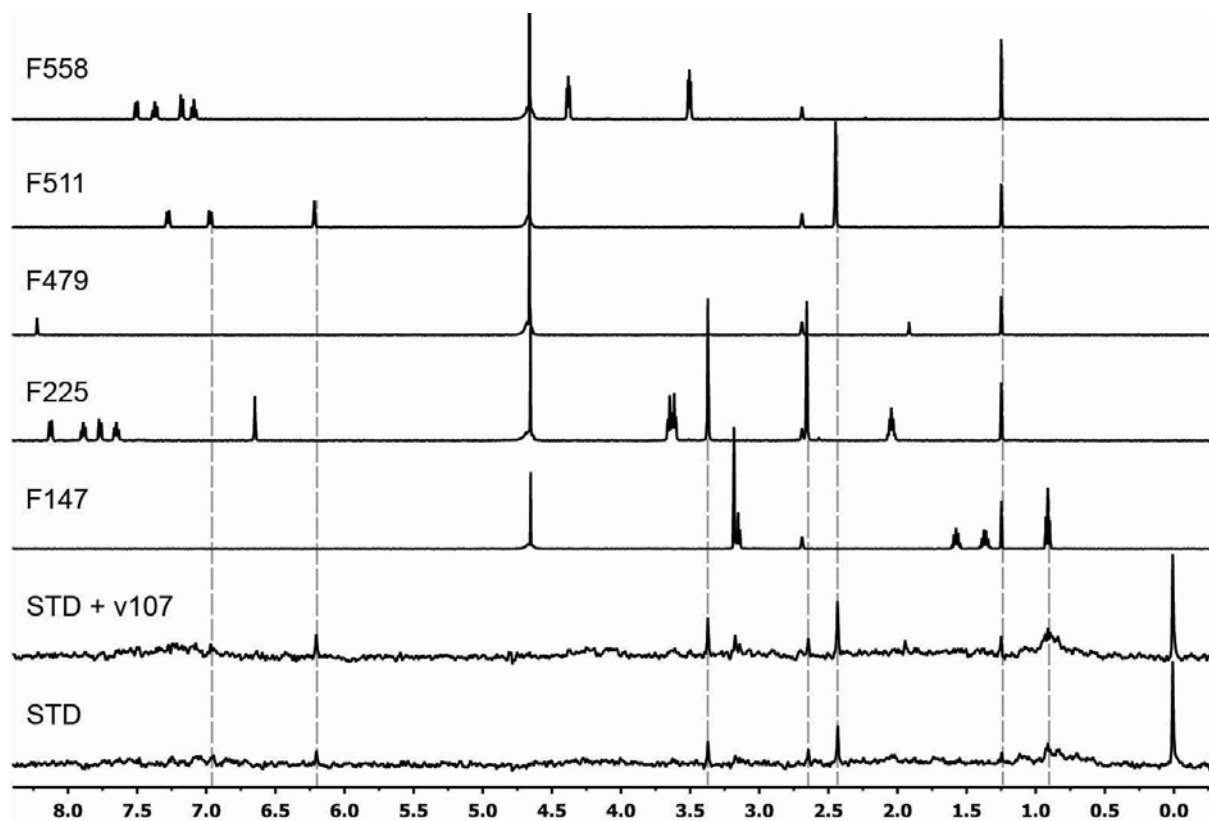

Figure 3.5.1-2: Each line corresponds to a separate NMR experiment. The first 5 lines correspond to ^{1}H spectra of individual fragments which were mixed together and tested at 500 µM in the presence of 10 µM VEGF for potential interaction with the protein. The resulting STD experiment in the last line revealed several STD signals that arise mostly from the compounds F511 and F225. To test whether these compounds bind to the protein-protein-interface of VEGF a competition experiment was performed with peptidic ligand v107. In this case no competition was observed. Rather, the STD signals increased in intensity in the presence of the peptide as can be seen in line 6.

A close comparison of the two STD spectra of figure 3.5.1-2 revealed that the majority of STD signals did not decrease in the presence of peptide v107 but increased. The change ranged from approx. 60% for the signals around 6.2 ppm and 3.4 ppm to 40% for the signal around 2.5 ppm and 20% for the signal around 2.7 ppm. These changes were significant and above the noise of the experiment. Further they could not arise from errors in scaling or sensitivity changes between both experiments since the DSS signal at 0 ppm was used as an internal reference and had the same shape and intensity. Finally the behavior of this mixture appeared not to be an exception but was observed in the majority of samples indicating that this result was not an isolated event or artifact.

The results from the screening for the first part of the library are summarized in figure 3.5.1-3.

From 402 fragments 230 (57%) did not show any STD signals. 172 fragments (43%) did show STD signals. From this, 24 fragments (6%) showed only STD signals for under 30% of their 1H signals. 40

Fragments (10%) showed STD signals for over 30% of their 1H signals but the intensities of these signals were rated as "weak". Finally the 108 compounds (27%) showed STD signals for over 30% of their 1H signals and the intensities of these signals were rated as "normal" or "strong".
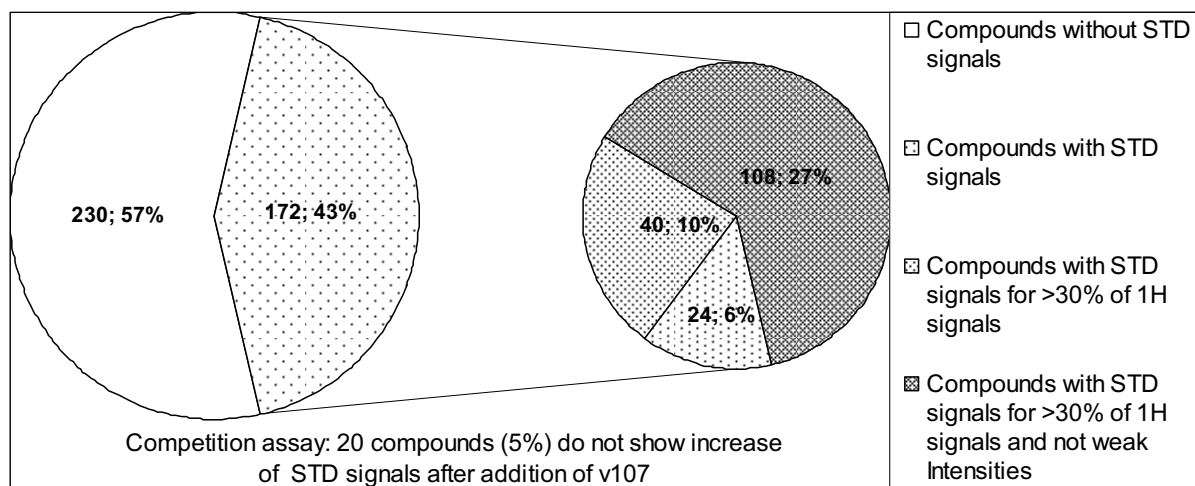

Figure 3.5.1-3: Summary of the screening of the first part of the library.

(1)   Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Science 1996, 274, 1531.
(2)   Hajduk, P. J., Dinges, J., Miknis, G. F., Merlock, M., Middleton, T., Kempf, D. J., Egan, D. A., Walter, K. A., Robins, T. S., Shuker, S. B., Holzman, T. F., and Fesik, S. W. J. Med. Chem. 1997, 40, J. Med. Chem. 40.
(3)   P. J. Hajduk, G. S., D. G. Nettesheim , E. T. Olejniczak , S. B. Shuker , R. P. Meadows , D. H. Steinman , G. M. Carrera , Jr., P. A. Marcotte , J. Severin , K. Walter , H. Smith , E. Gubbins , R. Simmer , T. F. Holzman , D. W. Morgan , S. K. Davidsen , J. B. Summers, S. W. Fesik J. Am. Chem. Soc. 1997, 119, 5818.
        (4)   Klein, J.; Meinecke, R.; Mayer, M.; Meyer, B. Journal of the American Chemical Society 1999, 121, 5336.
(5)   Moriz, M.; Bernd, M. Angewandte Chemie International Edition 1999, 38, 1784.

Discussion

This chapter summarizes the outcome of the screening of the first part of our library. During the setup of the fragment library we already defined how many fragments should be screened per mixture, at what concentration and ratio to the protein. We used the STD NMR experiment to screen our library, which is a reliable and sensitive method widely employed in academia and industry. Baicalin was very useful for the optimization of the experimental conditions. We decided to apply an on-resonance frequency of 0.09 ppm for the irradiation of the protein during the STD experiment. Although values in the low-field direction proved to give higher S/N ratios for the STD experiment this would also lead to a higher frequency of false positives due to direct ligand irradiation. A further advantage was the creation of an STD artifact created by direct irradiation of DSS which guided phasing the STD spectra and proved to be a very useful as an internal reference to monitor the quality of NMR spectra and compare intensities between them.

Originally we were planning to perform NMR screening at 37°C instead of 25°C. However this led to reproducibility issues with the Baicalin sample and in our initial exploration to strongly reduced hit

rates for a number of mixtures. This could be a simple thermodynamic effect, where an increase of temperature leads to an increase of entropic penalty for the binding. A possible additional effect could be related to VEGF. Maybe the protein samples different conformational space or the dynamic behavior changes between these two temperatures that disfavors the binding of fragments at 37°C. Although highly speculative at this point some results presented in chapter 3.6.2 indicate a similar behavior.

We did aim not only to explore the capability of fragments to bind to VEGF, but to identify the subset that is able to bind to the protein-protein interface that is involved in interaction with VEGF's receptors. We proposed for this purpose a competitive experiment which is based on the peptidic ligand v107, which binds with 1µM affinity to the zone of interest.

This approach appeared to work for the Baicalin test case. A concentration of 40 µM of peptide decreased two of Baicalin's STD signals between 30% and 60% which we assumed to be a magnitude of change that can be detected during the screening. However the decrease of intensity was not the same for both signals and the other signals seemed not to be affected at all. This behavior could indicate that Baicalin has additional binding sites that are not competitive, or that the competition between a small molecule and a 2.3 kDa peptide is not as simple as the competition between two small molecules. The peptide spans a relative large interface and possesses a multi modal binding mode with several spatial separated interactions. If only some of these interactions are occupied by a small molecule would this lead to complete competition or is the peptide able to adapt its conformation and capable to bind with the remaining interaction points? To my knowledge this issue is not covered in the literature.

The screening of the 402 fragments in 81 mixtures took around two weeks including the competition experiment. This setup was quite time demanding, but relying on shorter experiments would lead to downtime during the night due to the small sample changer of 9 positions and a decrease of the S/N ratio.

The NMR screening data was analyzed manually by asking four simple questions to each compound, that would allow some degree of quantitative evaluation while detection possible false positives. The first question was addressing the degree of ^{1}H signals of the fragments that give rise to STD signals in the presence of the protein. The fragments that we explored in this library were small compounds. Consequently we believe that although not all ^{1}H signals are required to give STD signals, a significant portion should be in close contact to the protein and give raise to observable STD signals. Therefore compounds that show only a low percentage of STD signals per 1H signals were most probably false positives. Although a high degree of STD signals per ^{1}H signals is an indicator for a relatively potent binding fragment I think this number is predominantly of qualitative value to asses if the compound is a true binder rather than a quantification of its binding strength. The reason for this are that the

total number of ^{1}H signals can be quite different between compounds and that the number of interaction is not related to their strength.

The second question was addressing the average intensity of the STD signals for each fragment. Although high affinity ligands with a very slow off rate give rise to low STD signals they should not be present in this library due to the size of our fragments. In my opinion it is therefore appropriate to roughly rank the affinity of our fragments based on the intensity of their STD signals. For this purpose the NMR spectra were inspected manually and the fragments labelled as "weak", "normal" or "strong" based on the average intensity of its STD signals. This approach allows some degree a quantitative evaluation but has some obvious drawbacks: Since the analysis is performed manually by a human it is quite subjective, the separation in the three categories inaccurate and allows only a relative evaluation. Additionally we only consider the STD signal intensity but do not account for the intensity of the parent ^{1}H signal. This leads to an overestimation of equivalent hydrogen's, as methyl groups, over individual or strongly coupled protons. It would be more accurate to calculate the percentage between STD and ^{1}H parent signals or the STD amplification factor. However this type of analysis would add a substantial amount of time to the already tedious evaluation of screening data if performed manually. We explored this issue further in chapter 3.4.2 and 3.5.2.

The third question was addressing if the STD signals of a fragment could arise from direct irradiation during the on-resonance experiment. Since we used an on-resonance frequency of 0.09 ppm but could observe STD signals for *tert*-butanol at 1.25 ppm we used this value as a threshold to label fragments as potential false positives that require additional confirmation.

The fourth question was addressing if fragments could bind to the repertory interface of VEGF. For this purpose the signal intensities of STD signals were compared in presence and absence of the peptide v107. Surprisingly the majority of STD signals increased in the competition experiment. This appeared not to be an artifact of the experimental conditions nor the analysis, since we could use the DSS signal as an internal reference. This was unexpected, since we anticipated that the STD signals in the competition experiments will either drop, which would indicate a competition, or be unaffected, which would suggest orthogonal binding sites. We do not know why the peptide increased the intensity of the STD signals. If we assume that the peptide and the fragment do not share the same binding zone, this could be either related to positive allosteric feedback or result from an increase of mass of the dimeric protein due to the binding of two units of peptide. This would increase its mass from 23.3 kDa by 12% to 27.9kDa, which may significantly reduce the molecular tumbling and improve the spread of magnetization over the protein and the transfer to the ligand. Further the complex between protein and peptide may have additional proton signals close to the irradiation frequency at 0.09 ppm which might add additional sources of magnetization influx during protein saturation and thus enhance the magnetization transfer to binding compounds.

Assuming that the peptide and fragment share the same binding zone an increase of STD signal intensity is less probable but possible because the protein is dimeric. Therefore the sample may contain a fraction of complex consisting of protein with one unit of peptide and one free interface available for fragment binding. This type of complex may increase STD intensities as discussed above or at least keep them the same range as in the cases with two binding sites available for the fragment. This possibility could have been eliminated by a higher excess of peptide, but could also favor direct interactions between fragments and peptide that could affect the STD signals. Finally as I discussed for the Baicalin sample we do not know exactly how a small fragment competes with a large multi modular binding peptide and how this would affect the STD experiment.

After performing this analysis for the 402 fragments that were screened we came to the following conclusion: The majority, 230 fragments (57%) did not show any STD signals and where considered as nonbinders. 172 fragments (43%) did show STD signals. From these, 24 fragments (6%) showed only STD signals for under 30% of their ^{1}H signals and were considered as potential false positives and therefore the least interesting compounds. 40 Fragments (10%) showed STD signals for over 30% of their 1H signals but the intensities of these signals were rated as "weak" indicating that these compounds could be potential binder but with relative weak affinity. Finally the 108 compounds (27%) showed STD signals for over 30% of their ^{1}H signals and the intensities of these signals were rated as "normal" or "strong". In the competition experiment we observed an increase of STD signal intensities in the presence of the peptide v107 for the majority of fragments. Only 20 compounds (5%) did not follow this trend. However due to the noise of the signals we could not observe a significant decrease of intensities, resulting in only a very vague indication that these compounds may bind to the protein-protein interface.

Originally we were considering the subset of 108 compounds as our most promising hits. Although it may appear good to have so many compounds this is actually problematic since it requires either another screening technique with relative good throughput or a better analysis to prioritize compounds. Further we were surprised of the high hit rate, which would correspond to 27% for the 108 fragments. One of the advantages of fragment base drug discovery are high hit rates. The literature reports examples over a wide range. For example Peters and coworker reported up to 61% when they screened a commercial fragment library against norovirus like particle.[1] This may be an extreme outcome biased by the size of the target. Researcher from Novartis reported a hit rate around or over 3% from multiple screening campaigns.[2] Schultz and coworker reported a rate around 10% for FABP4.[3]

Hajduk and coworker at Abbott reported hit rates between 0.01% to 0.94% for multiple targets.[4] The mot relevant report may come from researcher at Vernalis, that used a similar library design as we and report hit rates ranging from 1% to 7% for non focused libraries.[5]

Compared to these examples our hit rate is remarkably high. This is surprising, because the hit rate from fragment based screening was proposed as a drugability indicator[4] while protein-protein interfaces were reported to be difficult targets while. The explanation for this discrepancy is based on the broad range of screening conditions, protein targets, libraries and screening techniques that were applied. But the most important factor is the individual definition of what is a hit. For the above mentioned examples this varies from thresholds better than 0.5 mM to 5 mM or less specific the mM range or the appearance of signals with clear S/N ratio. When we finished the screening and analysis of this part of the library we were not aware that we used screening condition regarding fragment concentration, fragment to protein ration and experimental time that enables us to detect very weak interactions that most scientists in this field would consider as non-binding compounds.

After establishing this connection we interpreted the outcome of our screening of the first part of our library as a first indication of the low druggability of VEGF.

(1)   Rademacher, C.; Guiard, J.; Kitov, P. I.; Fiege, B.; Dalton, K. P.; Parra, F.; Bundle, D. R.; Peters, T. Chemistry 2011, 17, 7442.
(2)   Schuffenhauer, A.; Ruedisser, S.; Marzinzik, A. L.; Jahnke, W.; Blommers, M.; Selzer, P.; Jacoby, E. Curr Top Med Chem 2005, 5, 751.
(3)   van Dongen, M. J.; Uppenberg, J.; Svensson, S.; Lundback, T.; Akerud, T.; Wikstrom, M.; Schultz, J. J Am Chem Soc 2002, 124, 11874.
(4)   Hajduk, P. J.; Huth, J. R.; Fesik, S. W. J Med Chem 2005, 48, 2518.
(5)   Baurin, N.; Aboul-Ela, F.; Barril, X.; Davis, B.; Drysdale, M.; Dymock, B.; Finch, H.; Fromont, C.; Richardson, C.; Simmonite, H.; Hubbard, R. E. J Chem Inf Comput Sci 2004, 44, 2157.

3.5.2 NMR based screening of the second part of the Library

In this chapter the effort to the screen the second part of our library is described. Chronologically we finished already the screening of the first part before we started to design this second library part. In principle the workflow that we used to design and screen was very similar to how we progressed with first part. But the experience that we gained made us curious to explore improvements to the methodology that we applied previously.

The first modification that we wanted to implement was the use of a second NMR experiment for the screening. It is quite common to perform several NMR experiments with the same sample that complement each other. A good supplementation to the STD experiment is the CPMG experiment.[1] It relies on a very simple pulse sequence that applies after the 90° pulse a train of refocusing 180° pulses. This allows a relaxation decay without spin evolution. This experiment leads therefore to a reduction of ^{1}H signals from fragments that bind to the protein, because they use the fast relaxation behavior while the intensities of non binding compounds are less affected. Both NMR experiments have some intrinsic differences:

The STD experiment has a higher sensitivity than the CPMG experiment, but requires also longer experimental times to obtain significant S/N ratios. Since the CPMG experiment does not require protein saturation no false positives can appear from direct ligand irradiation. This makes the CPMG experiment a good control experiment. The sensitivity of the STD experiment is based on the amplification of the binding information in solution (see chapter 3.1) and therefore very dependent from binding kinetics. It is known that high affinity ligands with slow off rates lead to weak STD signal intensities due to low amplification. This issue is less critical in the CPMG experiment. It may offer therefore a better a better assessment of affinity and the detection of potent ligands that could escape STD detection.

To quantify the signal reduction in the CPMG experiment a reference sample which contains no protein is required. By comparing spectra from both samples the exact decrease of signal intensity can be measured that is caused by the protein. During the setup of the second part of the library we designed again mixtures with minimized signal overlap *in silico* and prepared then the mixtures *in vitro* to test for solubility issues and to record a reference NMR spectrum. In addition to the data acquired for the first library part we also recorded two CPMG spectra with a spin lock filter length of 20 ms and 400 ms. A spin lock filter of 20 ms has only a minimal effect on ^{1}H signal intensities of fragments but should be able to suppress protein signals. The 400 ms filter decreases signal intensities of binding and nonbinding fragments, but the former stronger that in some cases the signals nearly disappear.

The second improvement that we wanted to explore was if and to what degree the analysis of the NMR screening data can be automatized. As described in the previous chapter this work was very tedious and not predestined to be performed by a human being. Instead we envision that a computer could perform the work in a fraction of the time, while being objective and performing a powerful quantitative evaluation that is not possible for a human being in a reasonable amount of time. In the ideal case the computer analysis has identified a subset of the desired size that can then be rechecked manually by a human. AS described in chapter CV we developed a computational tool was that capable of this type of evaluation. Details of the method can be found in the corresponding chapter.

In total we screened 125 fragments in mixtures of five compounds. As for the previous part of the library the fragments were screened in mixtures of five at a concentration of 0.5 mM in the presence or absence of 10 µM VEGF. The results of the analysis are depicted in figure 3.5.2 and are based on a cross validation of the STD. and the CPMG experiments. We calculated the average values of normalized CPMG decrease (0.51) and STD vector (0.08, only for values ≠ 0). Both were used to divide the graph into four sectors. We selected as hits the 13 fragments that had favorable values from both STD and CPMG experiment. Additionally we selected the 6 fragments that had favorable STD values but week CPMG values. This led to the selection of 19 candidates as potential hits and would correspond to a hit rate of 15%.
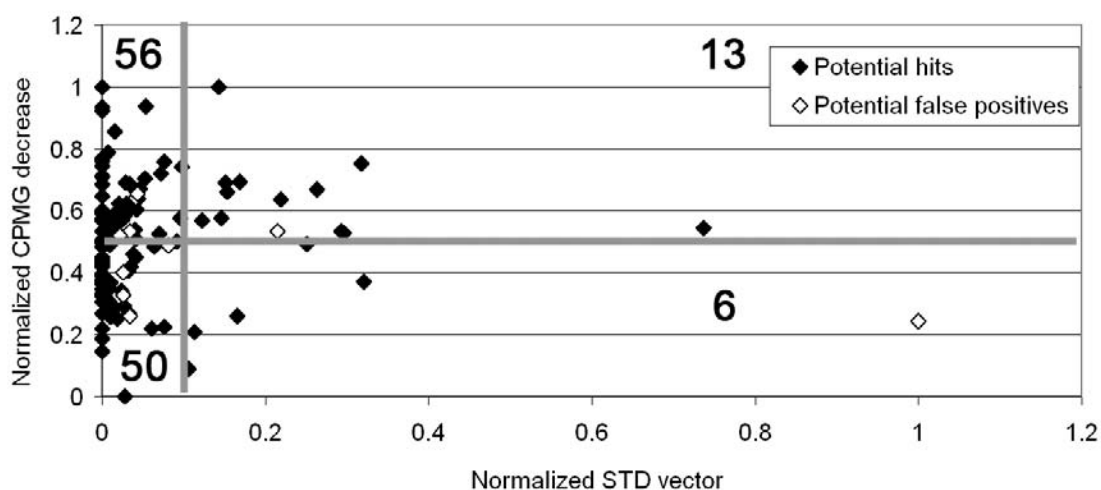
Figure 3.5.2: Results of automatic analysis of NMR based screening data for 125 fragments screened in mixtures of five against VEGF

The experimentally approach was already discussed in chapter 3.5.1. Although the hit rate for this library part was lower compared to the previous 27% it is obvious that slightly lower thresholds for the definition of the sectors would significantly increase the hit rate. Therefore one can argue that both hit rates are qualitative similar. The selected compounds were explored in chapter 3.6. The fragments with the two highest normalized STD vectors could not be confirmed. However several fragments with values between 0.2 and 0.4 could be confirmed.

3.5.3 Explorative study of competitive ^{19}F NMR screening to evaluate ligand binding to the protein-protein interface of VEGF

The screening of our fragment library led to the selection of 108 fragments from the first part of our library. The advanced analysis of the second part of the library allowed an improved quantitative assessment and we selected additional 19 compounds resulting in subset of 119 compounds. It was important to confirm the ability of these fragments to bind to VEGF. Further it was important to identify their binding sites, or at least weather they bind to the protein-protein interface of VEGF or not. Our efforts with the competitive STD experiment were not very conclusive in this context.

In the literature multiple techniques are described to validate fragment binding and to identify their binding sites. NMR offers protein observing experiments as CSP that can identify the binding site of a ligand with a one residue resolution. But these techniques are slow and require significant amounts of protein and were in our opinion not suited for such a large set of compounds. Instead we decided to rely again on ligand based NMR experiments which do not suffer these drawbacks. However these experiments do not per se offer information regarding the binding sites of ligands. This is only possible by a specific label applied to a specific site of the protein. A powerful example is the SLAPSTIC experiment in which the protein is modified with a spin label that suppresses NMR signals of compounds that bind in proximity to the label.[2] In the case of VEGF no selective attachment of a spin label would have been possible without site specific mutagenesis. Otherwise paramagnetic labeling of the protein may result in the modification of residues that are crucial in the protein-protein interface.

Competitive NMR experiments offer an alternative approach to gain information regarding the binding site of ligands. We already explored this approach with the peptidic ligand v107 in the competitive STD experiment. Reasons why this experiment was not very conclusive were discussed in the previous chapter. One possible explanation was that v107 can improve the magnetization transfer from protein to fragment independent of allosteric modulation. During my work in the field of NMR based fragment screening I became very intrigued by the work of Dr. Claudio Dalvit. Amongst other things he proposed the use of ^{19}F labeled "spy molecules" to detect competition for a binding site of interest.[3-5] This kind of experiments has been reported to have very high sensitivity and therefore allow working at low protein, fragment and spy molecule concentrations. The setup and analysis of this experiment is easy since no background signals appear in the ^{19}F NMR spectrum besides the signal of the "spy molecule". Moreover the intensity changes of this signal can be used to rank the affinity of competing fragments. Finally the experiment does not relay on any transfer of magnetization between protein and ligand, which may be important based on our experience with the completive STD experiment.

To explore the use of competitive ^{19}F NMR screening for the evaluation of ligand binding to the protein-protein interface of VEGF we first needed a suitable "spy molecule". The molecule should

bind to the site of interest, the protein-protein interface of VEGF, with an appropriate affinity and a ^{19}F label that allows for high sensitivity. Peptide v107 seemed to be a good template for this purpose. The peptide -VEGF complex is known and shows that the peptide occupies the majority of the protein-protein interface. The peptide has an affinity of 1 µM, which is probably too potent to be displaced by our weak fragments. However the literature describes a series of v107 analogs with reduced affinity.

Further in collaboration Andrey Dyachenko we studied the effect of additional modifications to the v107 template to the affinity of the peptide.[6] It should be straightforward to design an analog of desired affinity with this knowledge. A ^{19}F moiety can be introduced in many ways into v107 peptide. Due to the expertise of peptide chemistry in the laboratory the incorporation of fluorinated amino acid analogs which are readily commercially available should be a straightforward approach. Based on the structural information of the v107 complex we decided on the design of our spy molecule as depicted in figure 3.5.3-1:

B	Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Affinity
	v107	G	G	M	E	C	D	I	A	R	M	W	E	W	E	C	F	E	R	L	1 uM
	spy molecule	G	G	M	E	C	D	I	A	R	M	W	E	W	E	C	F*	E	A	G	152 uM
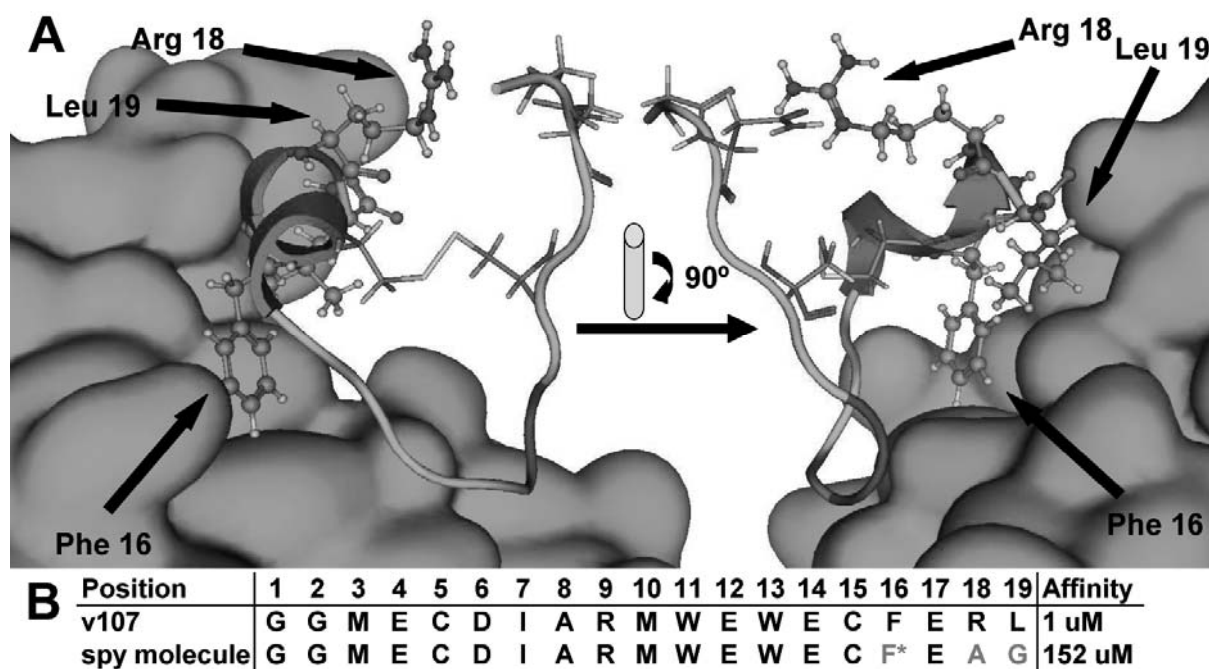
Figure 3.5.3-1: A: Complex between VEGF and peptide v107. Residues suggested for modification are represented as stick & balls. B: Sequence of the template v107 and the spy molecule analogue. F* marks the substitution with para-fluorophenylalanine.

In principle it is of advantage to incorporate a ^{19}F moiety with several equivalent fluorine atoms to increase the S/N ratio in the NMR experiment. But the sensitivity of the experiment is also determined by the magnitude of change that can be observed between the bound and unbound form of the spy molecule. As reported in literature, ^{19}F is strongly influenced by chemical shift anisotropy.[4] This effect is particular strong when the ^{19}F atom binds in proximity to an aromatic residue of the protein. This is true for the para-hydrogen of Phe16, that is located just 2.5 Å over the

centre of Tyr21 of VEGF. Therefore we substituted Phe16 to para-fluorophenylalanine to incorporate our ^{19}F moiety.

We were aiming for an affinity of the spy molecule around 100 μM which is the typical range reported in the literature.[3] The substitution of the para hydrogen is vey subtle but should have an effect on the affinity since Phe16 appears to be crucial for the interaction between peptide and protein.[7] It is difficult to anticipate how strong this effect will be, but we estimated that further modifications were required to decrease the affinity. To do this we substituted two additional residues. Leu19 was substituted by Gly which removed favorable hydrophobic interactions between peptide and protein. Arg18 was substituted by Ala that should destabilize the conformation of the bound peptide without significantly affecting the α-helical structure that orientates the core interacting residues.

Based on this analysis we proposed the peptide sequence in figure 3.5.3-1B for the spy molecule.

The synthesis was performed with the microwave supported standard solid phase peptide synthesis protocol and Fmoc-protected amino acids. The crude was sufficiently pure after cleavage from the resin and ether precipitation to continue without prior purification. The cyclisation was performed by oxidation with air over four days and monitored by MALDITOF-MS. The lyophilized crude was purified by standard RP-HPLC to purities over 95% and yields around 33%.

We used NMR CSP to assess weather the affinity of the spy molecule toward VEGF was in the desired range. For this purpose a sample of 100 μM (methyl[13]C)-Met-all-^{15}N VEGF$_{11-109}$ was titrated in five steps with increasing amounts of peptide up to a final concentration of 400 μM. A ^{1}H-^{15}N-HSQC spectrum was acquired at each step and analyzed as described in chapter 3.1. This computed the affinity of the spy molecule towards VEGF to 152 μM (R2=0.99). This value is slightly over the 100 μM that we were aiming for but well inside the range of affinities of typical spy molecules.

The setup of experimental conditions was performed in multiple steps. First we wanted to identify an appropriate molecule for the use as internal reference. Since the chemical shift range of fluorine covers over 200 ppm it is of advantage to select a reference molecule with a ^{19}F signal in proximity to the spy molecule signal to avoid base line distortions and to maintain a good spectral resolution. TFA is not suited for this purpose, since it is over 34 ppm away from our spy molecule (the TFA signal was assigned to 0.00 ppm). To identify better candidates, we purchased additional compounds and explored non binding fragments from our library that had a similar chemical environment as para-fluorophenylalanine. Two good candidates were 3,5-difluorobenzoic acid (from now on called ref. 1) and a fragment of the library (from now on called ref. 2). Both compounds had signals at 34.9 ppm (ref.1) and 43.2 ppm (ref.2) which were in proximity of the spy molecule signal at 34.6 ppm. This allowed us to reduce the spectral range to 16 ppm.

In literature competitive ^{19}F experiments are often applied in combination with an CPMG filter to amplify the intensity changes between bound and unbound form. To explore the optimal length of the filter a sample of 100 µM spy molecule was tested at different filter lengths and increasing amounts of protein. The results are summarized in table 3.5.3.

VEGF [µM]	CPMG filter length [ms]		
	40	80	160
0	80%	70%	40
2	70%	50%	30%
4	nb	30%	nb

Table 3.5.3: Influence of CPMG filter length and protein concentration on spy molecule signal intensity. The intensity without filter and absence of protein was set to 100%. (Acquisition parameter: spy molecule 100 µM, 64 scans, d1= 3 seconds, CPMG filter length 80 ms, temperature 25℃)

Results showed a strong influence of filter length on the spy molecule signal intensity even in absence of the protein. A filter length of 160 ms suppressed the signal down to 40% that would lead to low S/N ratios and a small margin to detect further the signal decrease due to interaction with the protein. The filter length of 40 ms had a lower affect on the intensity, but unfortunately also in the presence of protein, leading to lower difference and therefore a lower sensitivity of the experiment. 80ms seemed to be the best filter length, reducing the signal intensity to 70 % in absence of protein, and a further decease of 20% for each 2 µM of VEGF that was added. To our knowledge no peptide based spy molecules are reported in the literature. But for small molecule spy molecules filter length of 80, 160 and 320 ms have been applied. Therefore the value that we applied was at the lower edge but not an uncommon value.

To optimize the amount of protein required and to test the capability of the experiment to detect competition we used v107 and two analogs with reduced affinity that were available in the lab. For this purpose we used a sample containing 50 µM of each ref.1, ref.2 and the spy molecule. Working at these low concentrations had the advantage to increase the sensitivity of the competition experiment but required long acquisition times of 2 hours to achieve good S/N ratios. The signal of the spy molecule dropped significantly after adding VEGF to a final concentration of 1.2 µM and 2.0 µM. The later led to a decrease that we considered sufficient for the competition experiment. As a next step we prepared identical samples which contained additionally peptide 1 (v107, Kd= 1 µM), peptide 2 (Kd= 313 µM) and peptide 3 (Kd= 1.8 mM). Peptides 2 and 3 are v107 analogues that contain a substitution with a D-amino acid and were synthesized by Andrey Dyachenko. ^{19}F-NMR spectra were acquired for these samples with the sample conditions as for the previous samples.

For the analysis we use the signal integrals of ref. 1 and 2 and the spy molecule.
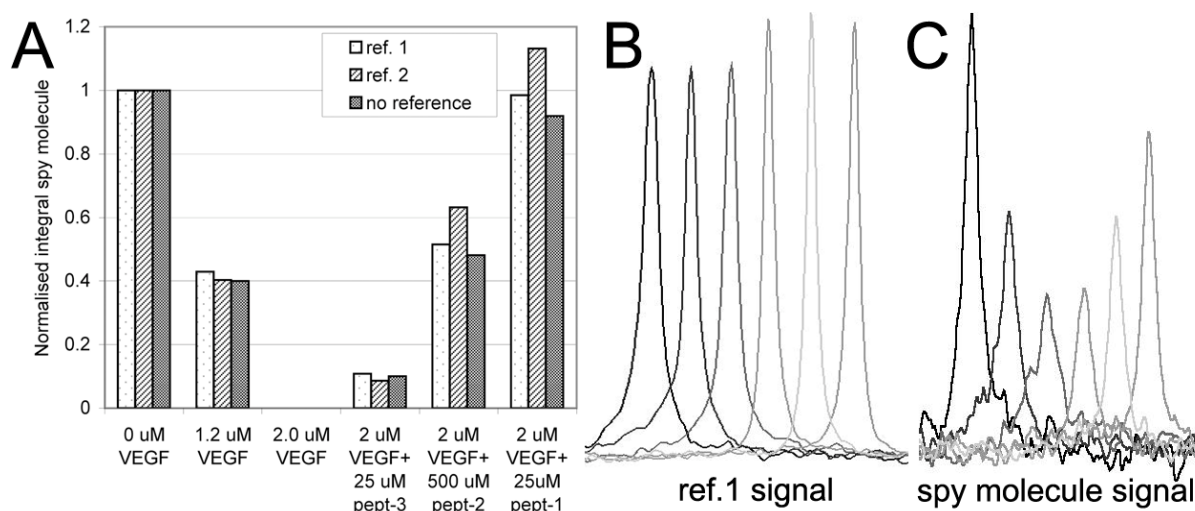
Figure 3.5.3-2: A,B,C: Effect of VEGF concentration and presence of competitive ligands on ^{19}F integral of the spy molecule (50 µM). The order of samples in figures A, B and C is identical. B,C: ^{19}F-NMR signals of ref1 and spy molecule, artificially shifted for better comparison. (Acquisition parameter: 2,5k scans, d1= 2 seconds, CPMG filter length 80 ms, temperature 45℃)

Figure 3.5.3-2, compares the integral of the spy molecule for all samples. All vales were normalized with the values from the sample with and without 2.0 µM VEGF. Further the integral of the spy molecule were either compared directly, or the ratio to one of the two reference integrals. The figure shows that a concentration of 25 µM of a low µM affinity ligand was sufficient to replace the spy molecule completely from the protein. The same concentration of an mM affinity molecule led to a very low degree of displacement which could also be cause experimental noise or differences between the sample composition. Finally, a ligand with a medium µM affinity and a concentration of 500 µM led to a clear displacement. Therefore our experimental setup should be capable of identifying ligands from our library that bind to the protein-protein-interface in this affinity range.

The NMR spectrometer in Barcelona equipped with a ^{19}F capable probe was not outfitted with a sample changer. Since we wanted to test a high number of compounds from our STD screening I joined for a week the NMR unit at the CNIO in Madrid under the guidance of Dr. Ramón Campos-Olivas. Because I wanted to assess around 100 fragments for competition with the spy molecule we decided to test at in mixtures at a concentration of 300 µM. Up to four compounds were pooled together in a fashion that each mixture contained one fragment which promising and the other compounds with less promising STD behavior. After assessing the available time and fast reoptimization of the experimental conditions we decided for relative short experiments of 256 scans, a CPMG filter length of 80ms and a spy molecule concentration of 50µM and 5 µM for VEGF.

To monitor the stability of the assay during the screening we introduced three blank samples consisting of only VEGF and spy molecule (labeled blank in figure 3.5.3-3) and three samples that contained additionally 25 µM of peptide v107. Between these samples we tested the fragment mixtures. The data analysis was time consuming and was performed after I returned from Madrid.

However some compounds of interesting mixtures were tested again individually and at a higher concentration during the end of the screening.

The data analysis was performed by processing the spectra in MestreNova. The integrals of the spy molecule, and the two reference molecules ref. 1 and 2 were calculated and exported to MS Excel for further processing. The two sample marked with an asterisk in figure 3.5.3-3 were defined as references corresponding to 0% for the sample consisting only of spy molecule and protein and 100% for the sample containing additionally 50 µM v107. To assess the degree of competition for the remaining samples the integral of the spy molecule was either compared directly to the reference sample or the internal ratio was calculated between spy molecule and ref. 1 and 2 and then compared to the reference samples. The results are summarized in figure 3.5.3-3.
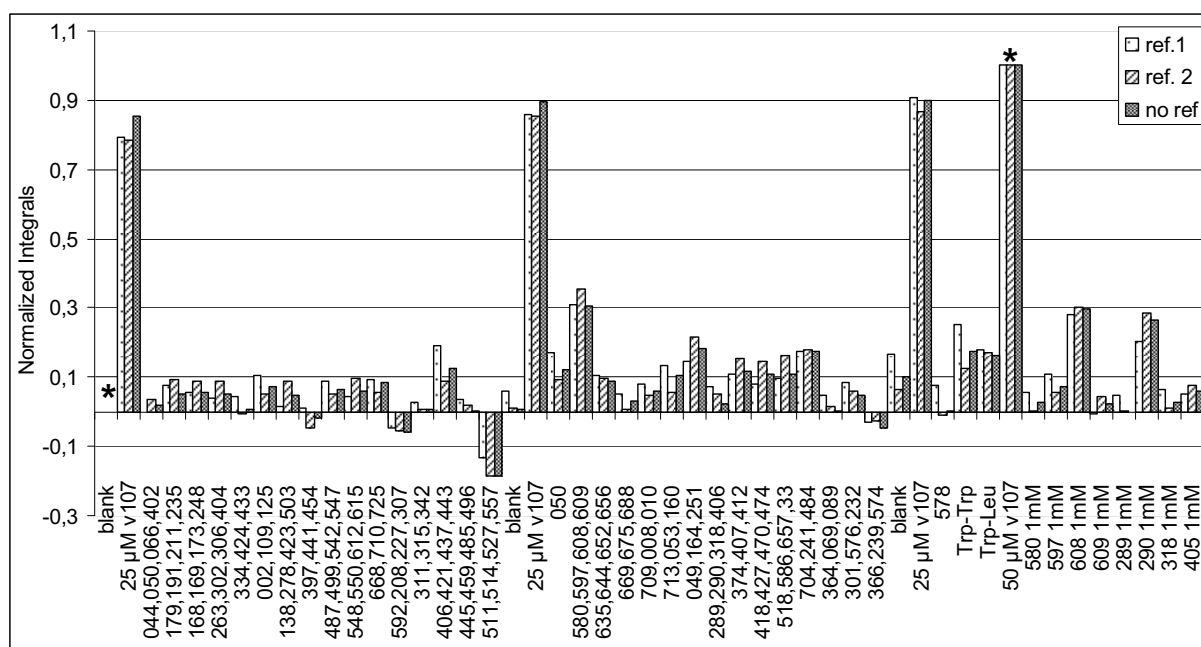
Figure 3.5.3-3: Competitive 19F NMR screening for the identification of fragments that compete with the spy molecule for binding to the protein-protein-interface of VEGF. * marks the two samples that were used to normalize the integral values. Spy molecule integrals were either compared directly, or as internal ratio to ref. 1 and ref. 2. (Acquisition parameters: Concentration spy molecule 50 µM, VEGF 5 µM, fragments 300 µM if not noted otherwise, 256 scans, CPMG filter length 80ms, temperature 25℃)

Discussion

The STD screening of our fragment library described of the previous chapters led to a large subset of approx. one hundred compounds that required validation and evaluation whether they bind to the protein-protein interface of VEGF or not. To asses this issue we explored in this chapter the application of competitive ^{19}F NMR screening. The design of the required spy molecule was based on the v107 template. v107 seemed to be an appropriate choice. It exhibits a relative strong affinity and occupies the majority of the protein-protein interface of VEGF. Modification of the peptide for our purpose should be straight forward, since the structure of the complex is available and the expertise

of peptide chemistry of the laboratory. We choose the para-hydrogen of Phe16 of v107 for the incorporation of a fluorine atom since this position is in direct contact with an aromatic residue of VEGF in the complex. This situation has been described in the literature to strongly improve the sensitivity of the experiment.

Once we decided on a design for the spy molecule its synthesis was straightforward and accomplished with a yield of 33%.

The NMR facility at the Parc Cientific in Barcelona acquired a ^{19}F capable probe directly when we started to work on this project. Unfortunately no person had previous experience with competitive ^{19}F screening. As a first step to derive the optimal experimental conditions we identified possible reference molecules. Then we optimized the length of the CPMG filter. 80 ms appeared to be a good tradeoff between the decrease of the signal in presence of the protein and tolerable reduction of the signal in the absence of the protein. 80 ms is in the range of reported values in the literature, but on the short end. The reason for this is probably that our spy molecule is a peptide with a weight of 2.3 kDa contrary to small molecules that are usually used. Since the peptide had already a significant higher relaxation rate than small molecules we lost significant signal intensity when applying longer filter times. Further the difference of the relaxation rate in the bound and unbound state are smaller than for small molecules making the experiment less sensitive. This appeared to be a drawback that we did not consider in the design of the spy molecule. However we also had no appropriate small molecule available as an alternative.

The sensitivity of the experiment seemed to be satisfactory with the conditions that we explored in Barcelona. For example the v107 peptide analog with an affinity of 313 µM showed a clear competition around 40% at a concentration of 500 µM. Therefore the assay should be capable to detect competitive fragment binding al least in the sub mM range. As the literature of fragment based drug discovery presents various examples where sub mM ligands were identified from the primary screen we were hoping to detect at least one such ligand for VEGF.

Overall, the sensitivity of our experiments was relatively modest compared to examples presented in the literature and it was uncertain to what extend we could detect competitive fragment binding in the mM range.

It was very fortunate that I could join Dr. Ramón Campos-Olivas at the CNIO to perform the competitive screening since he had the facility and knowledge of these experiments that we were missing in Barcelona. Due to the number of compounds that required testing we assayed them as mixtures and had to use conditions that lowered the sensitivity of the assay to increase throughput. Some mixtures that appeared to contain a promising compound were devonvoluted and the compounds retested individually at higher concentrations.

The stability of the assay was monitored by repeated introduction of blank samples without fragments and samples with v107 peptide. The analysis of the experiment showed that there is some degree of variation of 5% to 10% between the reference samples. This could indicate inaccuracy during the sample preparation or a time dependency but occurs probably by the relative S/N ratio that we were forced to work with to test this high number of compounds.

Nearly all fragment mixtures gave a very low competition under 10%. As discussed above this could be the margin of the noise. If so it is surprising that nearly all mixture show positive and not negative competition values as well. This could indicate that the mixtures contain at least one compound that exhibits a very weak competition. If so the affinity of these compounds would be significant weaker than the medium µM affinity range that we tested in our competition experiment in Barcelona. A compromise would be to assume that mixtures with a competition close to 0% are non binder, whether mixtures close to 10% may contain a fragment capable of competition. Nevertheless the affinity of this compound should be very weak. One mixture showed a negative value for competition around 20%. This appears to be too strong to be a result of noise. If not an artifact from sample preparation or aggregation, this could indicate a positive allosteric modulation between compounds of the mixture and the spy molecule. Some mixtures in the middle of the screening show competition values over 10%. However the following blank sample (no fragments) showed also an elevated value. The best scored mixture consisted of fragments 580, 597, 608 and 609. The deconvolution suggested that the majority of competition can be attributed to fragment 608. If this is not an artifact caused by aggregation or other issues this corresponds to a competition of ca 30% at a concentration of 1 mM. Compared to our results with the 107 analogue I would assume that this reflects an affinity in the low mM range rather than the µM range. Given the overall performance of the competitive screening the approx. 100 tested fragments appear to contain no sub-mM affinity ligands that bind to the protein-protein interface of VEGF. Some compounds may bind to other regions of the protein in this affinity range. However since the receptor interface contains the protein hotspot we were expecting that at least some compounds would bind to this area. These results suggest that no or very weak ligands outside of our detection threshold bind to the protein-protein interface of VEGF. This would be an indication of the low druggability of the protein.

(1)   Meiboom, S.; Gill, D. Review of Scientific Instruments 1958, 29, 688.
(2)   Jahnke, W.; Rudisser, S.; Zurini, M. J Am Chem Soc 2001, 123, 3149.
(3)   Dalvit, C.; Fagerness, P. E.; Hadden, D. T.; Sarver, R. W.; Stockman, B. J. J Am Chem Soc 2003, 125, 7696.
(4)   Dalvit, C. Progress in Nuclear Magnetic Resonance Spectroscopy 2007.
(5)   Dalvit, C. Drug Discov Today 2009, 14, 1051.
(6)   Dyachenko, A.; Goldflam, M.; Vilaseca, M.; Giralt, E. Peptide Science 2010.
(7)   Pan, B.; Li, B.; Russell, S. J.; Tom, J. Y.; Cochran, A. G.; Fairbrother, W. J. J Mol Biol 2002, 316, 769.

3.5.4 ^{19}F NMR based screening of the CNIO compound library

During my time at the CNIO in Madrid I used the existing infrastructure, resourses and expertise to screen an additional compound library under guidance of Dr. Ramón Campos-Olivas. The library had a size of 380 compounds and was designed for direct ^{19}F screening. I was curious how this library in combination with ^{19}F methodology would compare to our proprietary library and the STD experiment. The library was already prepared for screening. It was organized as DMSO stocks of mixtures of 8 compounds with non overlapping ^{19}F signals. The mixtures where divided into two classes: CF_3 and CF containing molecules. A sample of each mixture was prepared for screening containing only the fragments at 20 µM for CF_3 and 50 µM for CF containing molecules. Two ^{19}F NMR spectra were acquired for each sample with and without a CPMG filter (400/200ms for CF_3/CF). Subsequently VEGF was added to each sample to a concentration of 1 µM and the two NMR experiments repeated. Putative binders were analyzed by comparing the ^{19}F signal intensities with and without the CPMG filter before and after the addition of protein. This was the case for 34 compounds. To samples containing these putative hits another equivalent was added to examine consistency of the effect. The expected result was observed for 21 compounds that were designated for individual testing (Figure 3.5.4-1).
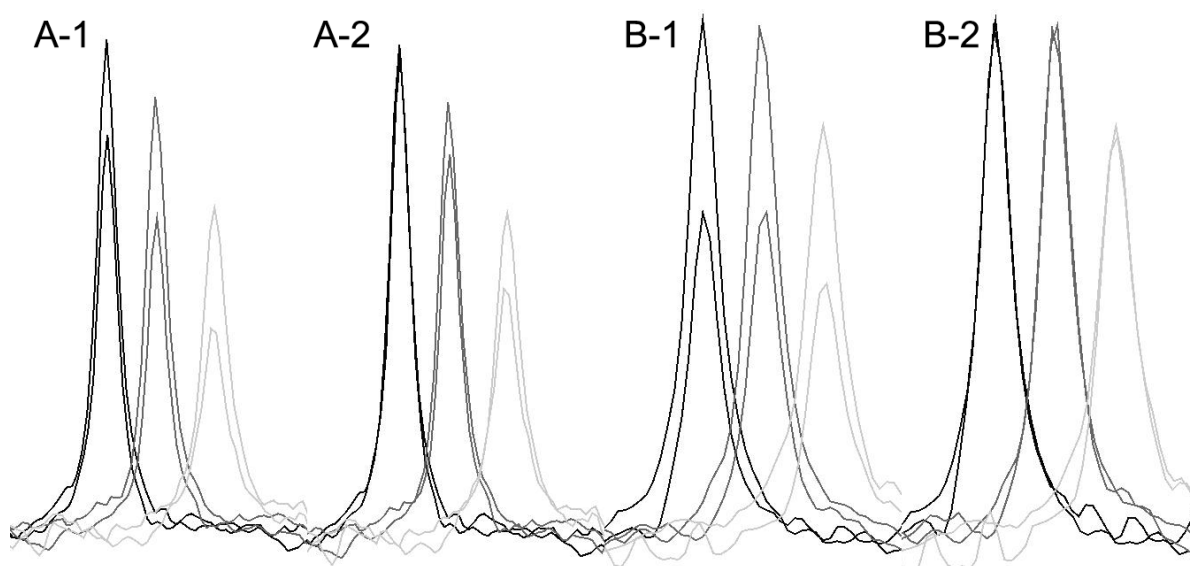
Figure 3.5.4-1: Examples of 19F NMR behavior of a binding (A) and a nonbinding (B) compound from the same mixture. Overlap of six NMR spectra with 0, 1 and 2 µM VEGF (from black to light grey). 19F-signals with identical chemical shift correspond to the sample with identical VEGF concentration but absence (higher intensity) or presence (lower intensity) of a CPMG filter. Spectra in figure A-1 and B-1 were not scaled. Spectra in figures A-2 and B-2 were scaled so that the 19F spectra with and without CPMG filter match in the absence of protein.

Eight of these compounds were readily available and tested immediately after we finished the screening. For this purpose a sample of each compound was prepared at concentrations of 50 µM and 150 µM for CF and CF_3 containing compounds respectively. Again two ^{19}F NMR spectra were acquired as described above in absence and presence of VEGF at concentrations of 1, 2 and 3 µM.

Four compounds did show the expected stepwise ^{19}F signal decrease upon titration with the protein. However the decrease of ^{19}F signal intensity was very modest for all cases and of the magnitude exemplified in figure 3.5.4-1. Based on the knowledge and pervious experience of Dr. Ramón Campos-Olivas this indicates a very low affinity of the identified ligands.

The remaining four compounds did not a dose response behavior and should not be considered as hits. Further five compounds were not available as individual stocks and could not be tested again. The last 8 compounds were tested as above, but with two months of delay. None of these candidates showed the required dose response upon addition of the protein to be confirmed as a hit.
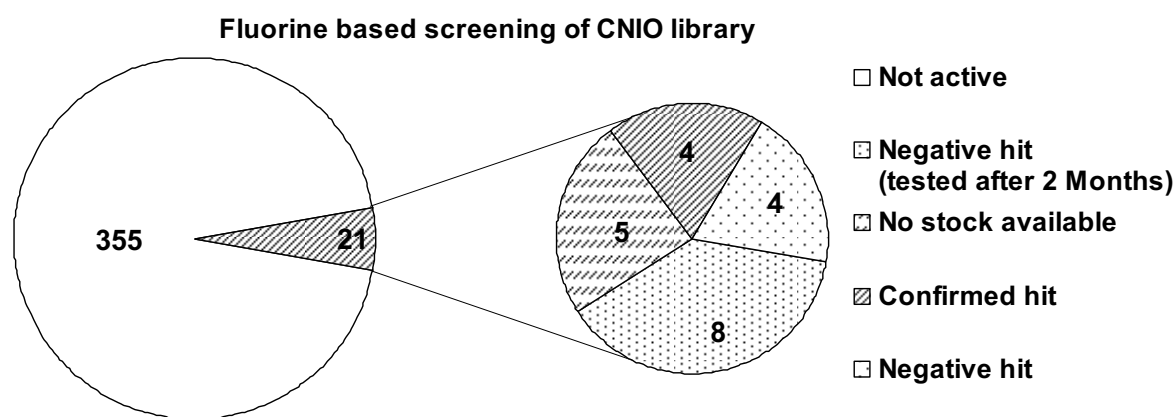
Figure 3.5.4-2: Outcome of the ^{19}F NMR based screening of the CNIO compound library.

Discussion:

In this chapter we screened the compound library of the CNIO with ^{19}F NMR methodology.

This was an interesting opportunity because we could explore an additional library that may cover additional regions of the chemical space to find VEGF ligands. Further I was very curious regarding the advantages and drawbacks of the ^{19}F NMR experiments compared to our STD NMR methodology. This work could be completed in a relative short amount of time due to the well prepared library, good infrastructure and hands on experience of Dr. Ramón Campos-Olivas. An obvious advantage of ^{19}F NMR screening is that the compounds can be tested at low concentrations. This avoids solubility problems and allows the assessment of compounds that are more hydrophobic. These compounds can be screened in larger mixtures, which is facilitated by the large range of ^{19}F chemical shifts that reduce the problem of signal overlap greatly. Although not a concern for us the use of low compound concentration also reduces the required amount of proteins. A very reasonable approach was to divide the compound into mixtures containing CF or CF_3 moieties. Both groups cover different ranges of chemical shifts. Assaying in these groups reduce the required spectral window size of the NMR experiment and therefore avoids artifacts as baseline distortions.

For screening we used very short NMR experiments of just 8 minutes per sample. This is quite outstanding compared to our STD experiments which required up to 2 hours. But one could argue that the STD experiments might have a higher sensitivity at least in some cases.

From around 380 compounds we were able to detect 21 potential hits, which corresponds to a hit rate of 5.6%. These compounds were verified individually. From eight compounds that were tested subsequently four were confirmed as true hits. Eight additional compounds that were tested after two months could not be confirmed. It is possible that the protein was damaged during this time which would explain why all compounds were tested as negative, but this is highly speculative. The remaining five compounds were not available as individual stocks and could not be tested again.

However we purchased some of these compounds and explored them in chapter 3.6.

To conclude the screening of the CNIO library led to a hit rate of 1,1% if we consider only the four confirmed hits. If we assume that there was a problem with the protein after two months we could expect that half of the remaining compounds were true hits as well which would result in a hit rate of 2,6%. In both scenarios the hit rate was significantly smaller than from our STD screening. As STD NMR the ^{19}F screening does not allow quantification of the affinity of the hits. However Dr. Ramón Campos-Olivas had the opinion that based on his experience the identified hits are very weak and have affinities in mM Range. Contrary to the hit rate this seems to be a similar outcome to the fragments that we identified by STD NMR from our own library. Both, the low hit rate and the apparent very low affinity of indentified hits were considered as an indicator for the low druggability of VEGF.

The methodology and implementation of ^{19}F NMR screening appears a very powerful technique. I would assume it has at least in some cases a similar sensitivity as STD NMR while offering additional advantages as I described above. It is certainly much faster and more straightforward to use. What may be of advantage in the context of protein-protein interfaces is that the technique allows screening of compounds that are more hydrophobic. As described in the introduction in chapter CV early evidence seems to indicate that ligands that bind to these surfaces are in average more hydrophobic than the ligands of classical drug targets. In practice this may be of limited use since the low solubility make evaluation and evolution of such compounds difficult. We faced this problem in chapter 3.6 where we tried to identify the binding site of hits from the ^{19}F screening but failed in due to solubility issues.

A drawback of the ^{19}F screening is the limitation to compounds that contain fluorine atoms. This is a general problem, although Zhong[1] and coworker argue that the commercial available chemical fragment space is covered sufficiently by fluorine containing compounds. If this is true it applies primarily to traditional drug targets. Recent literature indicates however that the efficient targeting

of protein-protein interfaces may require the exploration of novel zones of the chemical space. The majority of these compounds will not be, unless especially emphasized, contain fluorine atoms.

(1)   Jordan, J. B.; Poppe, L.; Xia, X.; Cheng, A. C.; Sun, Y.; Michelsen, K.; Eastwood, H.; Schnier, P. D.; Nixey, T.; Zhong, W. J Med Chem 2012, 55, 678.

3.5.5 Complementary computational studies.

Computational approaches have become an integral part of drug discovery efforts. In this project we were relying extensively on cheminformatics for the design of our library. In this chapter I will describe briefly complementary computational studies that were performed in collaboration with Dr. Martin Kotev.

First, we were interested to assess the druggability of VEGF. We were wondering how state of the art methods would evaluate VEGF and if this would reflect our experimental results.

Second, we wanted to explore how modern virtual screening would compare to NMR based efforts in screening our proprietary library for a challenging target as VEGF. Specifically we were wondering to what degree *in vitro* and *in silico* approaches would retrieve the same set of hits and whether this would help us to understand the experimental results and to prioritize compounds.

The prediction of druggabiliy of protein-protein interfaces is an ongoing challenge and active field of research. To our knowledge there is so far no consensus on how to predict their druggability. We used therefore two well-established tools for this purpose. The first is Sitemap from Schrödinger and the second SiteFind from MOE. As input we used the first VEGF unit of the PDB structure 2VPF. The crystal cell contains a total of four VEGF units and has a resolution of 1.93 Å.
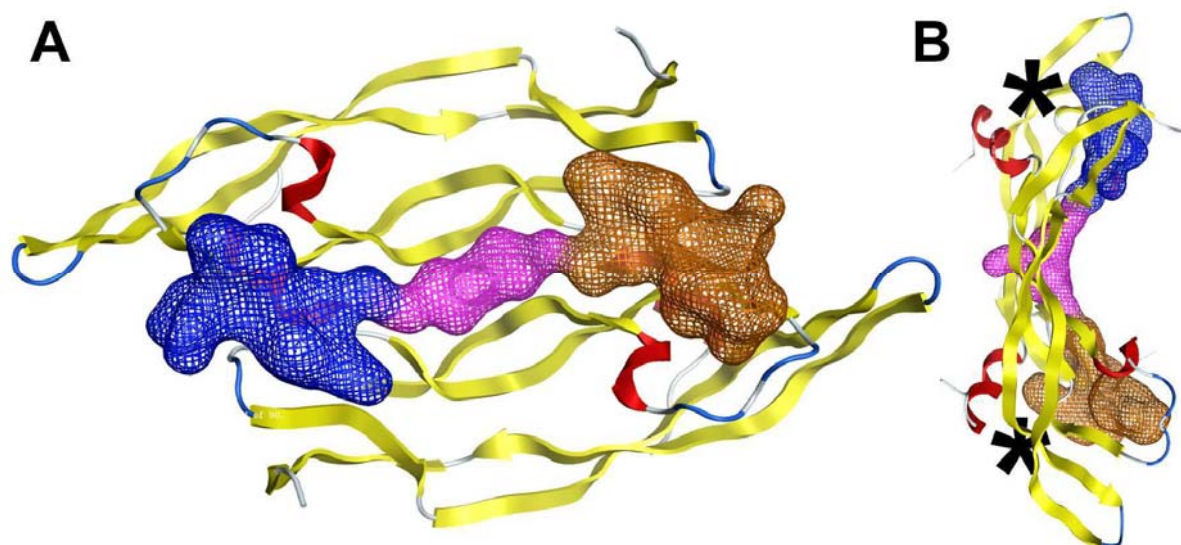
Figure 3.5.5 VEGF structure with detected binding sites (Sitemap)

Pockets	Schrödinger Sitemap		MOE Site Finder	
	Dscore	pocket volume [Å^3]	PLB score	Contacts
1 (orange)	0.965	255	0.40	103
			- 0.20	85
			- 0.87	53
2 (blue)	0.932	220	1.71	142
3 (magenta)	0.717	98	-1.05	85

Table 3.5.5-1: Summary of VEGF binding sites detected by Sitemap and Site Finder.

Both simulations gave very similar results. As can be seen in figure 3.5.5 pockets were detected only on the concave site of the protein. The two best scored pockets by Sitemap were located on both ends of the protein directly behind the zone where the receptor binds and at the interface between the two monomer units. Although the protein is symmetric both pockets did not have the same score. MOE divided the larger of Schrödingers pockets (orange) into smaller sub pockets. As mentioned the crystal cell in this protein has actually four protein units. Although the average RMSD is only between 0.4 to 0.9 Å the changes are localized at the random coils regions away from the center of the protein with local RMSDs over 3 Å for the random coil elements at both ends of the protein. Dependent of the VEGF unit that would be analyzed different definitions of the pockets and scores should be obtained. The result should not be significantly different but indicate some degree of flexibility in the protein. A third pocket which is smaller than the previous ones was detected by both programs at the centre of the protein.

Interestingly the programs did not recognize the protein-protein interface at the end of the convex site (marked with an asterisk) as a druggable zone. We considered this as an additional indication of the low druggability of the protein-protein interface of VEGF.

Both programs also asses the druggability of the detected pockets. The original site map article investigated a set of 63 pockets from 27 proteins and computed pocket scores from 0.2 to 1.4. In the subsequent analysis pockets with a Dscore under 0.87 were considered as undruggable, over ca 1.00 as druggable and values in between as difficult targets.[1] According to this evaluation the pocket in the center of VEGF would be undruggable and the other two pockets difficult targets.

The PLB score of MOE Sitefinder is an interesting concept that is similar to the hot spot concept. A training set was analyzed and the enrichment of specific amino acids at known binding sites compared to their occurrence of these amino acids on the rest of the protein surface. Specific amino acids such as tryptophan were enriched and therefore pocket that contained those scored higher if these residues appear.[2] Unfortunately no comparison of values over a broad set of targets was available to assess where VEGFs pockets score as undruggable, difficult or druggable. Independently the large pocket (blue) has a higher druggability score than the center pocket (magenta) as computed by Sitemap. In this context it is noteworthy that some people consider druggability scores more

reliable for relative assessment of pocket the same protein rather than comparing pockets between different proteins.

Virtual screening is a straightforward method to identify compounds that may bind to a protein. Dr. Martin Kotev was curious how it would perform in the context of protein-protein interfaces and VEGF. Further we were curious to what degree we could reproduce the experimental results that were obtained by NMR screening of our proprietary library. A high reproducibility could provide help to assess the experimental results, prioritize compounds and could ultimately be used to screen additional libraries. A low reproducibility may have value in assessing the challenges in screening protein-protein interfaces.

The Schrödinger software suite was used for docking our library against VEGF. Fragments were prepared with the ligprep module which includes protonation at neutral pH and the generation of possible tautomers and steroisomers if not previously defined. Docking was performed with Glide XP against three protein grids. We used the "undruggable" scored pocket at the center of VEGF(magenta), the pocket with difficult druggability (blue) and generated an additional grid at the protein-protein interface that was not recognized as a pocket.

Pocket	Fragments from first library part	Docking scores	Fragments designated as ligand by NMR (total 108)	Docking scores	Fragments designated as best binder by NMR (total 31)	Docking scores
Magenta	61	-6.8 to -4.4	16 (26%)	-5.8 to -4.4	4 (7%)	-5.5 to -4.4
Blue	71	-7.3 to -4.1	18 (25%)	-6.0 to -4.1	4 (5%)	-5.6 to -4.1
PPI	73	-7.1 to - 3.5	12 (16%)	-4.5 to -3.5	1 (1%)	-3.5

Table 3.5.5-2: Summary of virtual screening and comparison to STD NMR results for the first part of the library. The percentage values refer to the numbers in the first raw that lists the total number of compounds in the top100 scored positions that came from the first library part.(PPI: protein-protein interface)

We docked all fragments that were present in our MOE database including insoluble compounds. The fragments with the best 100 docking scores were retained for further analysis, which are summarized in table 3.5.5-2. To compare in silico with in vitro NMR screening results we retained only the subset of compounds that were tested by STD NMR from the first part of the library. We excluded at this point the second part of the library because the evaluation was not completed yet. For each grid between 61 and 73 compounds were retained from the top 100 scoring fragments. The docking score and range was similar for all grids ranging from around -7 to under -4. At this point no clear difference was obvious between the grids. From the 402 compounds of the first part of the library 108 were considered as potential hits by STD NMR as described in chapter 3.5.1. This corresponds to a percentage of 27%. Although false negatives can appear in STD NMR the probability for this was very low in our experimental setup. Therefore the remaining 73% can be considered as decoys to test

the ability of virtual screening to deliver a candidate set with an enrichment of potential ligands. Depending on the grid between 12 and 18 compounds belonged to the subset of fragments that were designated by NMR as potential ligands. This corresponded to only 26% to 16% and some compounds were selected for several grids. Selected compounds occupied the lower range of docking scores that was previously covered. Therefore the best docking scores were associated with compounds that can be considered with high confidence as non binder. At least some cases could be traced back to the enumeration of stereoisomers that may not be present in the library. If we compare the rate of binding compounds present in the dataset (27%) and the recovery rate after docking the rates are similar or wore in the case of the protein-protein interface. Together these results indicate that the method can not truly distinguish between binding and non binding compounds. One may argue that the set of 108 compounds designated as potential hits by STD NMR may contain false positives and that therefore the fraction of decoys is larger than assumed. Given the weak signal intensities that we obtained this is quite probably, but will not change the results significantly in my opinion.

If we use a very stringent selection of best binder by STD NMR we receive a set of 31 compounds.

Again only low fractions of this set could be recovered by virtual screening. Interestingly the fraction that was recovered belonged to the lowest scored compounds of the original docking range. Should these docking scores have experimental significance their low range would indicate that our most potent fragments should have very weak affinity. Generally there is no direct correspondence between affinity and docking score. But based on the expertise of Dr. Martin Kotev docking scores around -3 and -4 indicate affinities in the mM range.

Comparison of docking scores and recovery rates of NMR designated ligands between the grids (magenta, blue, protein-protein interface) showed some differences. The pocket at the center of the protein (magenta) and the pocket at the interface of the monomers (blue) appear to be quite similar contrary to the prediction by Sitefinder and Sitemap. Interestingly the protein-protein interface that was not detected as pocket by the software had the lowest docking scores and recovery rates for the ligand set of 108 and 31 compounds. If we use this as a druggability index as proposed by Huang et al[3] this indicates the low druggability of the protein-protein interface.

The low correspondence of STD NMR screening and *in silio* results is probably not a surprise. First we do not have a high confidence with all potential hits that were identified by STD NMR. Second, it is known that the accuracy of virtual screening decreases in the case of low affinity molecules as fragments compared to high affinity compounds. This tendency increases probably if the target is a protein-protein interface.

Major challenges in this context may be: to find the appropriate protein conformation that is required for fragment binding, the identification of the right binding site, calculation of accurate scores that separate binder from non binder.

Without addressing this issues virtual fragment screening for protein-protein interfaces is prone to suggest misleading candidates. One possibility to improve the situation may be to implement experimental data as much as possible into the computational efforts. In chapter 3.6 we explored this possibility further.

(1)   Halgren, T. A. J Chem Inf Model 2009, 49, 377.
(2)   Soga, S.; Shirai, H.; Kobori, M.; Hirayama, N. J Chem Inf Model 2007, 47, 400.
(3)   Huang, N.; Jacobson, M. P. PLoS One 2010, 5, e10109.

3.5.6 Concluding remarks

In this section we used several methods to explore the capability of fragment based screening to identify potential ligands for VEGF. Let us briefly recapitulate the results:

1. STD-NMR based screening of our proprietary library led to very high hit rate around 27% and 15%. The affinity of hits was probably very low based on the relative weak STD signal intensities that were obtained.

2. Competitive screening by ^{19}F-NMR of our primary hits did not identify (maybe with one exception) ligands that bind to the protein-protein interface of VEGF. The over 100 tested fragments must therefore bind to other surface patches on the protein or with an affinity that is under our detection threshold.

3. ^{19}F-NMR screening of the CNIO library led to hit rates around 1-3 %: The affinity of the identified compounds appears to be very weak based on the expertise of Dr. Ramón Campos-Olivas.

4. Two state of the art tools for binding site prediction and druggability prediction did not recognize the protein-protein interface of VEGF as duggable.

5. Virtual screening showed a low correlation to experimental results obtained by NMR based screening. The range of docking scores between -3 and -4 that was obtained for the most promising fragments would indicate very low affinities.

Since the hit rate of fragment based screening has been proposed as an indicator for the druggability of proteins,[1] this summary raises the question how the results correlate to each other and what conclusions they allow regarding the druggability of VEGF.

Let me propose a hypothesis to explain our results and put them into perspective:

Druggable proteins have a surface with at least one binding site that allows tight binding of small molecules. If one would plot the frequency to find a ligand of a given affinity this would result in a graph similar to figure 3.5.6-A. For druggable targets the size, number and quality of small molecule binding pockets governs the hit rate observed in fragment based screening. The hit rate is further dependent on the definition of what is considered as a hit. For druggable proteins the screening conditions and hit definitions are generally stringent (high cut off) allowing only the observation and selection of relative potent fragments. For example Schultz et al[2] reported a hit rate around 10% for a definition of screening behavior that corresponded to an affinity of at least 0.5 mM.

Contrary to druggable targets protein-protein interfaces lack cavities or surface patches that offer high quality interaction. Further their intrinsic flexibility and lower spatial confinement allows more promiscuous interaction modes. This can be extrapolated to the behavior depicted in figure 3.5.6-B, which is defined by a low probability to bind ligands with high affinity, but increased probability to bind ligands weakly which can be detected if appropriate screening conditions and hit definitions are selected. I believe that VEGF is an example for this behavior. We observed high hit rates from our library because of the sensitive screening conditions that were applied and the low threshold for the acceptance of hits.
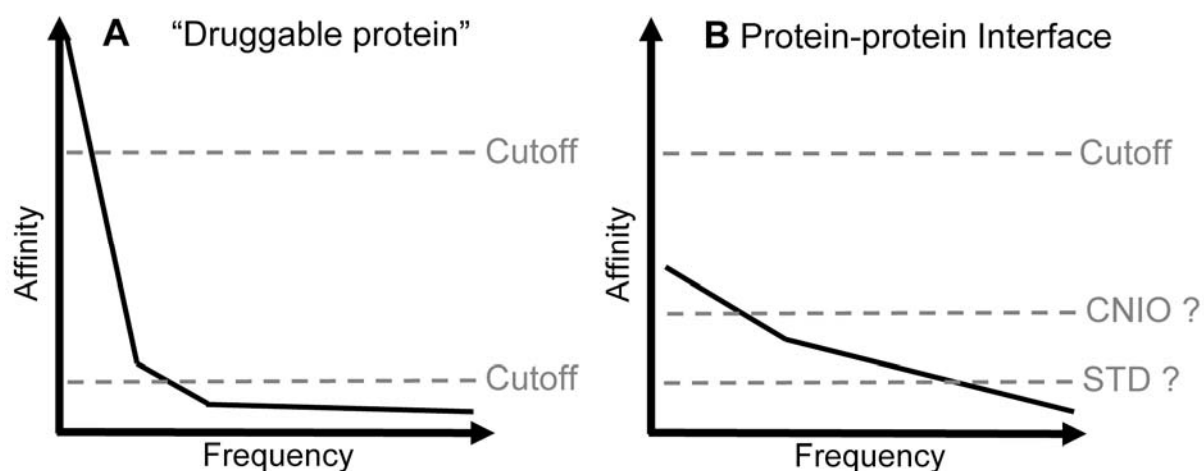
Figure 3.5.6: Hypothesis regarding fragment screening hit rates observed for druggable proteins and protein-protein interfaces. The hit rate depends on the frequency of identifying a ligand with a specific affinity and the definition of hit rate.

However this would require that the identified hits have very low affinities. Based on a comparison of the STD signals intensities obtained for our fragments to spectra reported in the literature[3] I would assume that the majority of our compounds have affinities over 10 mM. We believe this estimation is also valid for the compounds from the CNIO library. Further it may be in part responsible for the low correlation of virtual screening to experimental results and the low docking scores. Finally the absence of strong CSP for our initial fragments reported in the next chapter indicates that their affinity is significant lower than the testes low mM concentrations. Together these observations would match the behavior described in figure CV-B. This model would also explain the difference of

hit rates observed by our STD screening (27%, 15%) and the ^{19}F-screening (1-3%) if we assume that the later had a lower sensitivity, or less favorable screening conditions to detect binding in the range of 10 mM. Certainly the use 8 minutes per sample compared to 90 minutes for the STD experiments lowered the sensitivity of the ^{19}F-NMR screening.

Should this hypothesis be true it would have implication on fragment based drug discovery in the context of protein-protein interfaces. Compared to HTS screening fragment based drug discovery offers higher hit rates but identifies (although efficient) weaker ligands. Both behaviors might be amplified when targeting protein-protein interfaces: The hit rates might be higher (if detection and selection thresholds are lowered) but the affinity of identified compounds lower.

The work in this chapter is mainly based on the exploration of our library, which was designed after the "SAR by catalog" principle. Although we do not have this information regarding the CNIO library, its compounds seems to be slightly larger and hydrophobic compared to our fragments but cover a similar chemical space. Recent publications have suggested that targeting of protein-protein-interfaces may require the exploration of novel chemical space, in particular compounds with sp3 enrichment. [4,5] Certainly it would be of interest to explore libraries of such compounds and compare their ability to bind to VEGF to our results.

(1)   Hajduk, P. J.; Huth, J. R.; Fesik, S. W. J Med Chem 2005, 48, 2518.
(2)   van Dongen, M. J.; Uppenberg, J.; Svensson, S.; Lundback, T.; Akerud, T.; Wikstrom, M.; Schultz, J. J Am Chem Soc 2002, 124, 11874.
(3)   Maurer, T.; Garrenton, L. S.; Oh, A.; Pitts, K.; Anderson, D. J.; Skelton, N. J.; Fauber, B. P.; Pan, B.; Malek, S.; Stokoe, D.; Ludlam, M. J.; Bowman, K. K.; Wu, J.; Giannetti, A. M.; Starovasnik, M. A.; Mellman, I.; Jackson, P. K.; Rudolph, J.; Wang, W.; Fang, G. Proc Natl Acad Sci U S A 2012, 109, 5299.
(4)   Connor, O. C.; Beckmann, H. S.; Spring, D. R. *Chem Soc Rev* **2012**, *41*, 4444.
(5)   Hung, A. W.; Ramek, A.; Wang, Y.; Kaya, T.; Wilson, J. A.; Clemons, P. A.; Young, D. W. *Proc Natl Acad Sci U S A* **2011**, *108*, 6799.

3.6 Molecular recognition at the VEGF surface: protein structure based approaches

Based on the work outlined in the previous chapters we identified a group of fragments that bind to VEGF. They were either recognized by STD NMR from the first and second part of our proprietary library or by ^{19}F-screening of the CNIO library. Probably not all of these compounds are real ligands and have quite certainly very low affinity. Further, we do not know their binding sites. Some fragments may bind to the protein-protein interface of VEGF as they show a week competition with the ^{19}F-labeled spy molecule that recognizes this zone. Finally we performed virtual screening of the available fragments, although there seemed to be a low correlation between *in vitro* and *in silico* results. Taken together all the presented results indicate that VEGF is an undruggable protein, at least with the methodology that we applied so far.

A pharmaceutical company would probably not continue such a project as the low druggability of the protein and the absence of promising hits would result in a low chance to develop a lead compound. However since the targeting of protein-protein interfaces and their druggability are currently under debate we thought it worth wise to continue with this project.

At this point we considered our options on how to proceed.

In general there are two ways to improve the potency of our compounds. The first approach relies on "brute force" where one needs to test multiple analogs of potential binding compounds to improve potency. However this requires huge chemical libraries and capabilities to test these compounds, resources that we did not possess at the IRB Barcelona. The second approach relies on the rational modification of compounds to improve their potency. Here, significantly fewer resources are needed, but the knowledge of the binding mode of the protein-fragment complex is critical. In collaboration with Dr. Jordi Benach from the ALBA synchrotron facility we pursued this goal with X-ray crystallography. Although protein crystals of VEGF were obtained with a reasonable resolution, no cocrystals with fragments could be obtained. NMR is an alternative approach to X-ray crystallography. A multitude of pulse sequences are available that differ in the kind and amount of information they offer and the resources they require. We decided to rely on CSP as this technique served us well for determining affinity and binding sites of peptides binding to VEGF as described in chapter 3.1. Further, CSP is a medium throughput method which allows testing of a relatively large group of compounds, an important consideration due to the large set of potential compounds but limited confidence we have. The drawback of CSP is that it can describe the binding site only with a one residue resolution. This may be sufficient to asses if a fragment binds to the protein-protein-

interface of VEGF but not to determine the exact binding pose which is necessary for the rational modification of the fragment to improve its potency. As described later in this chapter we proposed a strong correlation between NMR data and induced docking to overcome this gap.

3.6.1 First round of elaboration

As noted above we had a multitude of compounds that have the potential to bind to VEGF. Due to the inconsistent experimental behavior of these compounds we had no clear ligand candidates. We decided therefore to combine the information from different approaches as much as possible to help us prioritize compounds with higher confidence. For example, a compound was prioritized that showed relatively good STD effects in the NMR screening, a potential competition in the ^{19}F screening and had a favorable docking score by virtual screening for a relevant grid.

For the first round of assessment we selected with this approach 64 fragments from our proprietary library. Further we selected 12 compounds from the CNIO ^{19}F library, four which had been confirmed in the single compound assay and eight, which failed this assay although it was not completely clear if this was based on the inability of the compounds to bind or deterioration of the protein.

The selected compounds were tested for their maximal solubility by visual inspection of stock dilutions with NMR buffer. Unfortunately the CNIO compounds appeared to be highly insoluble and only 5 compounds could be used for further tests as they exhibited a solubility of at least 0.5 mM. The remaining compounds were assayed at concentrations under their maximal solubility for binding to VEGF. This was performed by preparing samples of 100 uM VEGF with 0.5 to 4 mM of each fragment depending on solubility. Of each sample, a ^{1}H-^{15}N-HSQC NMR spectra was recorded at 45°C. Addition of the fragment stock to the protein sample resulted also in the addition of approx. 1% to 5% of DMSO. Initial experiments showed that the CSP induced by DMSO was the same order or higher than changes induced by the fragments. To remove the DMSO bias each NMR spectrum had to be compared to a reference spectrum that contained the same concentration of DMSO. Special care was required to ensure that both samples were as identical as possible e.g. age of protein stock, buffer, experimental conditions and processing since changes induced by the fragment were close to the experimental noise.

The 69 NMR spectra of fragment containing samples were visually compared to reference spectra for the occurrence of visible peak shifts. This was the case for 18 compounds which were analyzed further. ^{15}N and ^{1}H chemical shifts were determined for all backbone signals and for each residue the normalized CSP calculated.

The assessment of fragments was simplified by plotting the total number of backbone residues that exhibited a normalized CSP over 0.01 ppm versus the subset of residues that were located inside or

within 3 Å of the protein-protein interface of VEGF (Figure 3.6.1-1). The latter was possible due to the

PDB structure 2VPF and the NMR backbone assignment published by Pan et al.[1]

The fragments showed on average a CSP over 0.01 ppm for 4.5 residues from which 1.7 were located

inside or close to the receptor binding interface. Both values were used to divide the plot into four

quadrants. Quadrant A contained six compounds that demonstrated significant CSP for the total

number and the residues at the interface. Quadrant B enclosed three fragments that showed an

inferior number of CSP but mostly located at the receptor interface. Quadrant C contained three

compounds with a relative high number of CSP but only a minority was located at the protein-protein

interface. Quadrant D enclosed the remaining six compounds that had inferior values in both

categories.
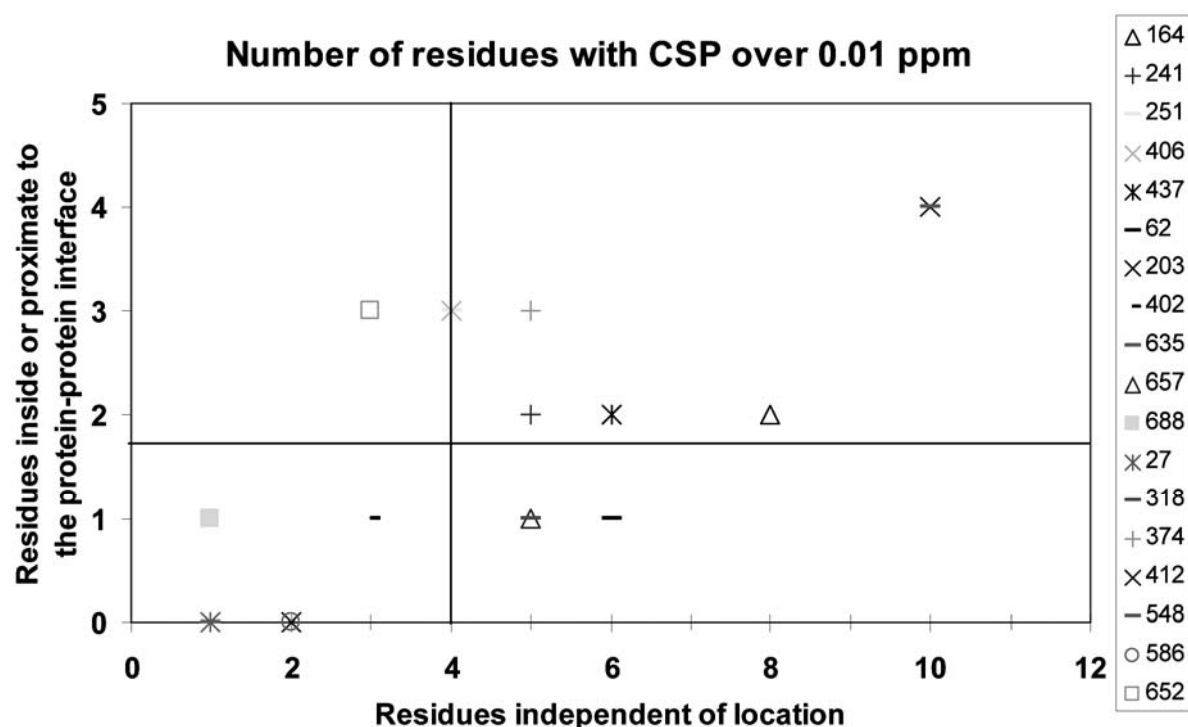
Figure 3.6.1-1: Analysis of 18 fragments that showed visible CSP in a 1H-15N-HSQC experiment
when compared to reference samples. The number of residues with a CSP over 0.01 ppm was plotted
along the x-axis for the total number of residues and the y-axis for residues in or close to the protein-
protein interface of VEGF.

Fragments were prioritized according to the alphabetical order of their quadrants. Compounds from

quadrant D were discarded. The binding epitope on the protein surface was determined for all

remaining 12 compounds. This was achieved by selecting residues that exhibited significant CSP,

which was defined as perturbations greater than the sum of average CSP and standard derivation for

all residues. The first VEGF unit of PDB entry 2VPF was used as template and all residues that showed

significant perturbations were depicted bold and in black. The resulting distribution of black protein

surface could be divided into three categories: First, a localized surface patch on or near the protein-

protein interface. Second, a localized surface patch away from the protein-protein interface. Third, multiple not connected surface patches. Four examples are listed in figure 3.6.1-2.
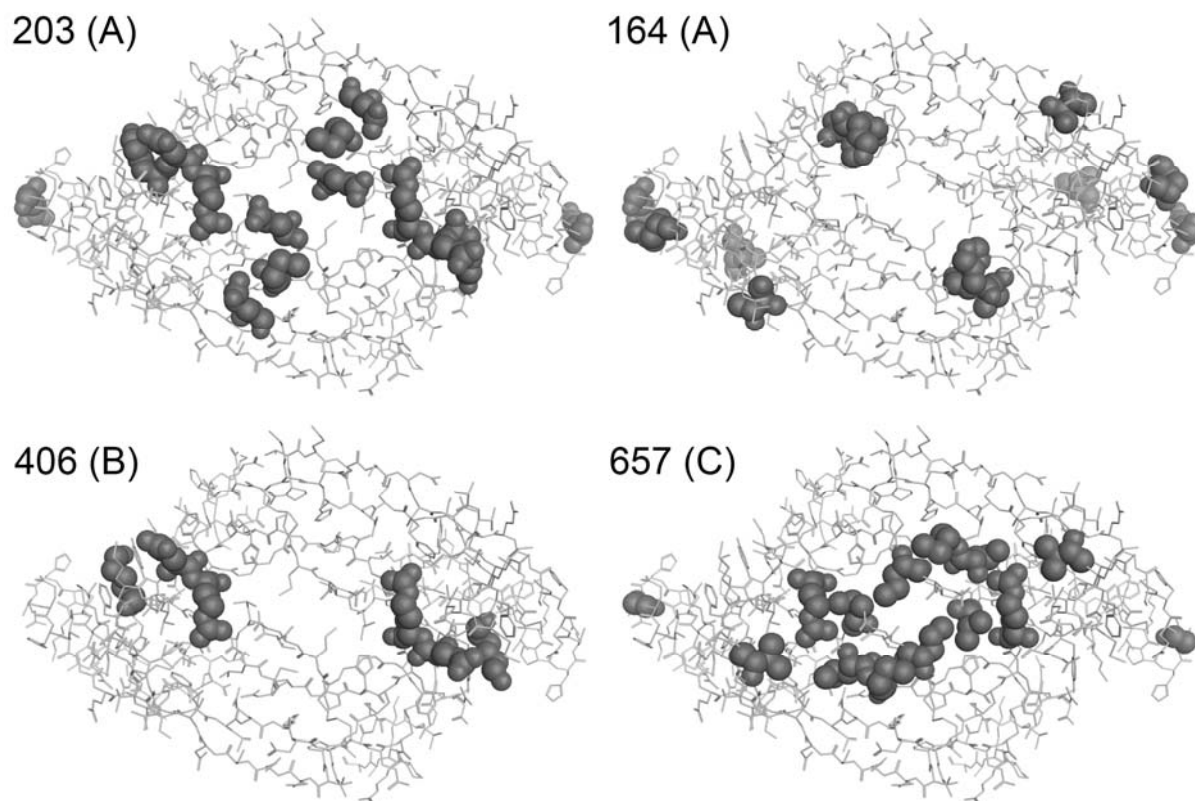
Figure 3.6.1-2: Example for typical distribution of significant CSP. The number annotates the fragment number, the letter the previously associated quadrant. For fragment 203 and 164 only non overlapping residues were depicted.

With this approach we were able to gain information regarding the binding site of the fragments with a one residue resolution. The selected compounds, residues with significant CSP and information regarding their binding site are summarized in table 3.6.1.
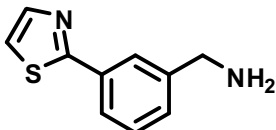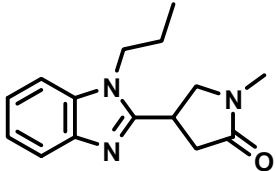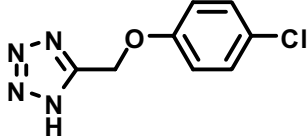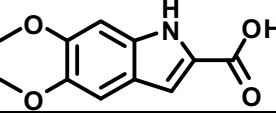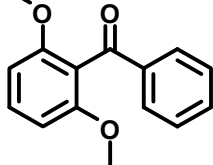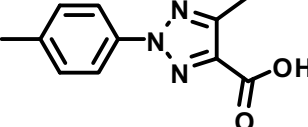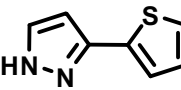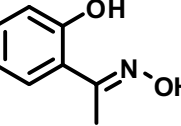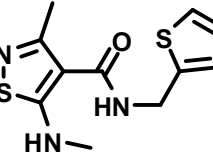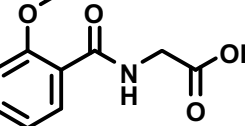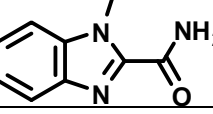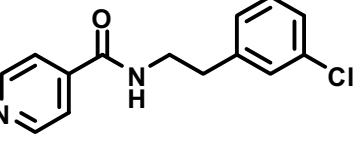
ID	Structure	Conc.	Quadrant	Protein residue ID`s with significant CSP	Putative binding site	PELE docking score
548		4 mM	A	73, 100, (54, 29), **48**, 90, 30, <u>24</u>, 43, <u>27, 26</u>	No clear/ several binding sites	-4.2
203		1 mM	A	**21, 22**, 98, <u>23,</u> (47, 51), (15, 32), 87, 55, 30	close to PPI, additional binding site?	-5.8
164		4 mM	A	43, <u>26,</u> 77, (33, 56), 90, <u>27,</u> 87, (15, 32)	No clear binding site	-4.2
374		1 mM	A	<u>23,</u> **48**, <u>26,</u> (15, 32)	maybe close to PPI	-5.7
437		2 mM	A	31, (44, 99), 84, **48**, 87, **22**	maybe binding to flexible loop	-3.8
241		3 mM	A	<u>26,</u> **48**, (15, 32), 27, 87	no clear binding site	-6.2
406		4 mM	B	<u>23,</u> **48, 22**	close to PPI	-4.2
251		4 mM	B	**22**, 30, **21**, <u>23</u>	close to PPI	-5.3
652		2 mM	C	**48**, (47, 51)	maybe binding to interface between monomers	-5.9
062		3 mM	C	77, 90, 87, 82, 98, <u>26</u>	binding to flexible loop	-
318		1 mM	C	**81**, 30, 100, 96, 55	Center of the protein?	-4,4
657		3 mM	C	<u>23,</u> 55, 88, (15, 32),(54, 29)	center of the protein	-

Table 3.6.1: 12 fragments from quadrant A,B and C that were characterized for their ability to bind to VEGF by ^{1}H-^{15}N-HSQC NMR with a protein concentration of 100 µM. Conc. lists the fragment concentration during the CSP experiment. Bold and underlined residue ID's mark residues closer than 3Å and respectively 6 Å from the protein-protein interface. Residues in brackets mark a signal overlap of two residues which should be considered with caution.

Unfortunately the table does not contain any fluorine bearing molecules from the CNIO library. They exhibited no or a weaker CSP than the fragments from our own library. This does not necessarily indicate a weaker affinity since only lower concentrations could be tested.

The one residue resolution of the binding mode provided by CSP was not sufficient to determine the exact position of the ligand which an atomic resolution that was required to rationally predict the modifications that can be applied to increase the potency of the compounds. For this purpose we sent fragments that showed a localized coherent CSP at the protein-protein interface to our collaborators for X-Ray crystallography. In the meantime we were trying to approach this problem in parallel computationally. Dr. Martin Kotev, a shared post doctoral scientist between the Prof. Giralt and the Prof. Guallar laboratory at the Barcelona super computing center used the Protein Energy Landscape Exploration (PELE) software for this purpose.[2] It is a novel stochastic Monte Carlo approach to predict protein structures. We performed the structure prediction in the presence and translation of our fragments and used it as an induced fit docking program. Briefly each fragment was docked with the Schrödinger suit to the area that showed significant CSP. A 10 Å radius around the docked conformation was explored with the PELE methodology. The resulting trajectory was clustered and representative member of each cluster redocked with Glide XP. The outcome was a list of 10 to 100 protein-ligand complexes with varying docking scores for each ligand.

Up to three different complexes for each fragment that showed the best docking scores and which corresponded to the experimentally observed CSP were selected. The best docking score for each fragment are listed in table CV.

(1)   Pan, B.; Li, B.; Russell, S. J.; Tom, J. Y.; Cochran, A. G.; Fairbrother, W. J. *J Mol Biol* **2002**, *316*, 769.
(2)   Kenneth W. Borrelli, A. V., Raul Alcantara, and Victor Guallar *J. Chem. Theory Comput* **2005**, pp 1304.

3.6.2 Second round of elaboration

In the second round of assessment we wanted to improve the potency of our compounds.

We assumed that the previous selected binding poses contained the true conformations that are involved in the interaction of the fragment with VEGF. To identify modifications that improve their potency the binding poses were inspected and the characteristics and orientation of surrounding binding sites were noted. Then the CAS database was queried for commercial available analogs that

displayed additional desired features. With this approach approx. 500 compounds were selected. These compounds were docked with Glide XP to 21 protein grids. 16 grids corresponded to one to three conformations for each fragment from the PELE exploration. The remaining five grids corresponded to binding pockets of the native protein conformation from PDB entry 2VPF as listed in table 3.5.5-1 in chapter3.5.5. Since the protein is not completely symmetric, both, the magenta and the blue pockets were used. Additionally we created two grids for the protein-protein interfaces at both sides of the protein. Several compounds could be identified that exhibited improved docking scores compared to the values of the original fragments. The improvement ranged from a very modest increase to up to 2.8 units.

In total 64 compounds were selected as promising candidates for purchase. Unfortunately the selected compounds from the CAS database were distributed by a large number of vendors , many of which required high minimum order volumes resulting in excesive compound prices and shipping costs. As a result we were able to purchase only 14 compounds. To test additional compounds we explored libraries that were already available in the laboratory. The majority of compounds belonged to the Maybridge library of approved drug molecules and natural products. In total, approx. 2.000 molecules were docked to the previous selected 21 protein grids leading to over 40.000 complexes. The data analysis was performed by selecting 100 complexes with the highest docking score and by clustering them based on the similarity of structure of the ligands with MOE. This led to the generation of 30 clusters with up to 15 members. Clusters were then rated according to their member size, the range of docking score and ligand efficiency. The individual docking poses of highly rated clusters were inspected manually to confirm that ligands were either identical or similar and recognized the same binding site on the protein surface. Compounds were selected for experimental validation that belonged either to high ranked clusters. Or were close analogs of the initial fragments in order to determine the relevance of present functional groups. With this approach 29 molecules were added to the initially 14 purchased compounds resulting in a total of 43 compound candidates.

The 43 candidates were tested for their capability to bind to VEGF by ^{1}H-^{15}N-HSQC NMR. As previously described, the maximal solubility of the compounds was determined and concentrations under these values used for the binding assay. In the case of compounds from the Maybridge library, only concentrations of 0.5 mM could be tested due to the low stock concentrations. The previous results indicated that all compounds bind very weak, and that the observed CSP was close to the noise of the experiment. Therefore we searched for adjustments to improve the sensitivity of the methodology. Two modifications to the previous process were introduced. First, the protein concentration as reduced from 100 to 50 uM. This increases the excess of ligand versus protein and should lead to larger CSP since we used protein observing NMR experiments. Second the five ^{13}C-methyl methionine signals were used for principal assessment of compound binding to VEGF. The

methyl groups of methionine exhibit a higher flexibility than the mostly rigid protein backbone and should be more receptive for ligand induced changes. Further this experiment could be executed in a fraction of the time of the backbone experiment. Otherwise the decrease in protein concentration would have reduced the throughput significantly. The methionine signals were observed with a ^{13}C-^{1}H HSQC pulse sequence. The experiment was optimized by folding the five [methyl-^{13}C]-methionine signals into a small spectral range. This permitted to acquire an NMR spectrum with a 50 uM protein sample in approx. 15 minutes. As described for the ^{15}N-^{1}H HSQC experiment all spectra were visually compared to reference spectra with equal DMSO concentration. If a CSP was visible the normalized value was calculated for each methionine signal and free [methyl-^{13}C]-methionone that was present in each sample as a reference. This analysis was performed for 22 compounds and is summarized in figure 3.6.2-1
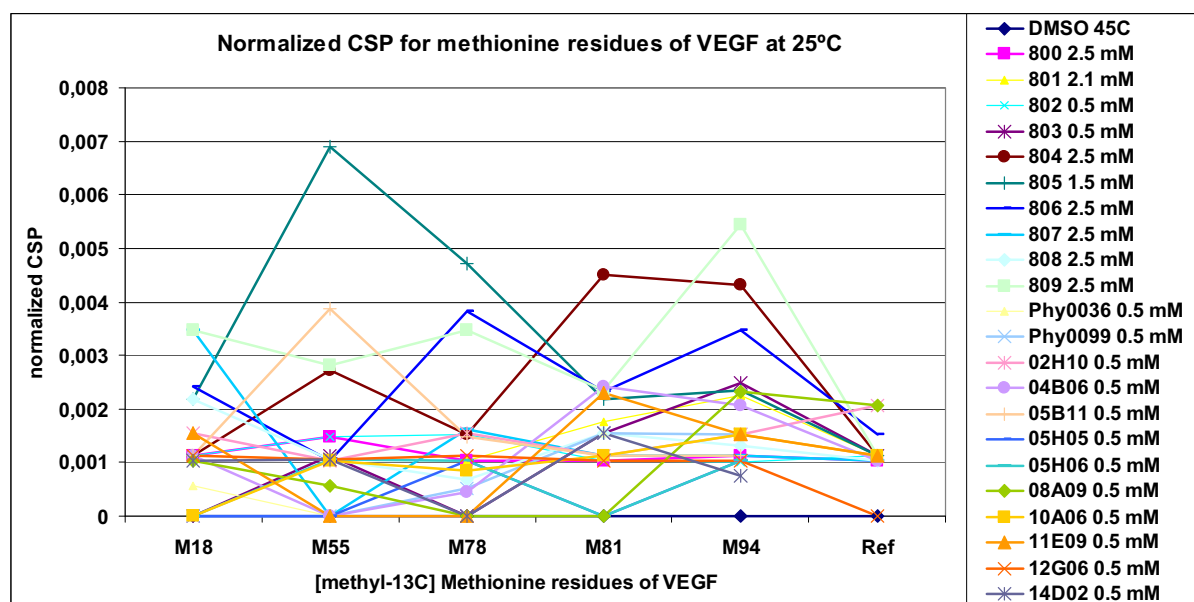
Figure 3.6.2-1: Normalized ^{1}H-^{13}C-HSQC CSP at 45℃ of [methyl-^{13}C]-methionine signals present in the protein and free [methyl-^{13}C]-methionine as reference.

The observed CSP was lower than we anticipated. To gain additional information we performed ^{1}H-^{15}N-HSQC experiments for 14 samples with an emphasis on compounds that showed normalized CSP values over 0.002 ppm. The data was analyzed as in the previous round.

The results are summarized in table 3.6.2-1.

ID	Conc.	Highest CSP of M55 M73 M94	Highest CSP of M18 and M81	Protein residue ID`s with significant CSP in ^{1}H-^{15}N -HSQC	Maximal CSP in ^{1}H-^{15}N –HSQC (Av. significant CSP)
800	2.5 mM	0,001	0,001	(47, 51), 26, 87, 46, 79, 82, 27, 90, 71	0.0282 (0.0185)
801	2.1 mM	0,002	0,002	(47, 51), 26, 87, 24	0.0241 (0.0191)
802	0.5 mM	0,002	0,001	30, 67, (47, 51), 87, 24, 46, 93, **18**, 52, 90, (44, 99), (47, 51), 69, **66**, 75, 92, (44 ,99)	0.0175 (0.0127)
803	0.5 mM	0,002	0,002	75, 26, 87, 46, 23, 31, 67, (15, 32), 73, **17**, 24, 58, (47, 51), 30, 59, 95	0.0241 (0.0129)
804	2.5 mM	0,004	0,005	(47, 51), 87, 75, (47, 51), 26, 52, 46, 14, 100	0.0175 (0.0166)
805	1.5 mM	0,007	0,002	24, 23, 52, 75, **21**, *16*, 55, 87, **22**, 30, 71	0.0235 (0.0176)
806	2.5 mM	0,004	0,002	87, (47, 51), 90, 26, 52, 39, 27	0.0247 (0.0184)
807	2.5 mM	0,002	0,003	nd	nd
808	2.5 mM	0,001	0,002	59, 87, 92, 14, 31, **48**, 67, 73, 24, 52, 94, (47, 51), 30, 69, 79, 90	0.0175 (0.0117)
809	2.5 mM	0,005	0,003	52, **48**, 30, 24, 90, 87, 92, 71, 43, 88, 93, 46, 36, 67, **18**, 73, 37, 19, 39, 95, 79	0.0175 (0.0130)
Phy0036	0.5 mM	0,002	0,001	nd	nd
Phy0099	0.5 mM	0,002	0,002	88, (47, 51), 26, 87, 102, 75, 59	0.0175 (0.0164)
02H10	0.5 mM	0,002	0,002	79, (33, 56), (44, 99),77, 88	
04B06	0.5 mM	0,002	0,002	(47, 51), 75, 79, 87, 96, (47, 51)	0.0234 (0.0193)
05B11	0.5 mM	0,004	0,001	nd	nd
05H05	0.5 mM	0,001	0,000	nd	nd
05H06	0.5 mM	0,001	0,000	87, 16, 23, (47, 51), 31, 24	0.0247 (0.0181)
08A09	0.5 mM	0,002	0,001	(47, 51), 75, 73, 52, 30, 67, **21**, (33, 56), 26, 79, (47, 51), (44, 99)	0.0235 (0.0179)
10A06	0.5 mM	0,002	0,001	nd	nd
11E09	0.5 mM	0,002	0,002	nd	nd
12G06	0.5 mM	0,001	0,001	nd	nd
14D02	0.5 mM	0,001	0,002	75, 14, 23, 87, 16, (47, 51), 67, 97, **66**	0.0323 (0.0173)

Table 3.6.2-1: Summary of ^{1}H-^{13}C- and ^{1}H-^{15}N-CSP behavior of selected compounds at 45℃. CSP behavior that is stronger than the average is highlighted in grey. Conc. lists the fragment concentration during the CSP experiments. Bold and underlined residue ID's mark residues closer than 3Å and respectively 6 Å from the protein-protein interface. Residues in brackets mark a signal overlap of two residues which should be considered with caution. Nd = not determined

As can be seen from the table, the behavior from the [methyl-^{13}C]-methionine and the backbone amide CSP did not appear to converge into a common picture as anticipated. Compounds that show a normalized CSP in the [methyl-^{13}C]-methionine experiment over 0.002 ppm (grey) don't show often a maximal CSP over 0.02 ppm for the backbone amides (grey). Compounds that show a relative strong CSP of methionines M18 and M81 which are located in the protein-protein-interface do not have necessary significant CSP of residues in the same area. Compared to the previous round the number of significant CSP and their magnitude seemed to be slightly increased. However their colocalization is relatively low. These results may serve to eliminate candidates that have a weak behavior in both experiments but do not correlate sufficiently to identify a potent compound.

After considering this issue we decided to retest a subset of the compounds at 25°C. So far all our protein based NMR experiments were performed at 45°C since the signal assignment and NMR experiments for PDB structure 1KAT were performed at this temperature and the protein showed excellent stability and NMR properties. In contrast, at 25°C many backbone signals broaden, overlap or disappear completely and the remaining signals require significantly increased acquisition times. Initially we retested the compounds at a single concentration with the ^{1}H-^{13}C-HSQC experiment. The results for the CSP and intensity changes of the five methionines and the reference signal are plotted in figure 3.6.2-2:
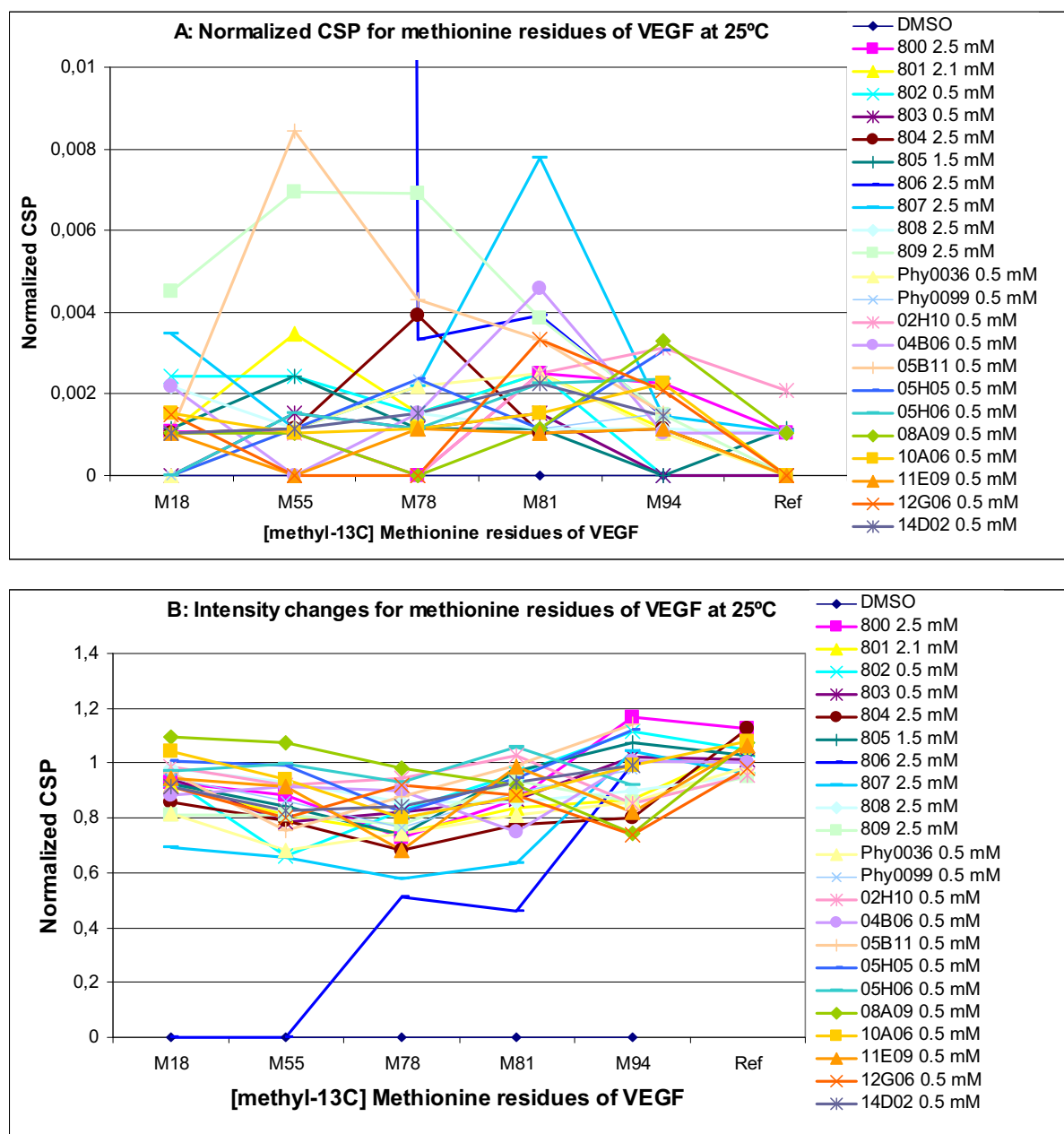
Figure 3.6.2-2: Normalized ^{1}H-^{13}C-HSQC CSP and signal intensity changes at 25°C for [methyl-^{13}C]-methionine signals present in the protein and free [methyl-^{13}C]-methionine as reference when exposed to selected compounds.

Generally the CSP values seemed to have increased slightly. But compounds that showed the strongest shifts at 45°C as 804 and 805 showed relative weak shifts at 25°C. Compound 809 retained its strong shifts, but the affected methionine changed from M94 to M55 and M78. A similar behavior at both temperatures was only evident for compound 05B11, that showed the strongest CSP at M55 which nearly doubled at 25°C. Some compounds gained significant CSP for both M81 and M18, which was the case for 807, 04B06 and 12G06. Interestingly all three compounds showed the strongest perturbation for M81 and the second strongest perturbation for M18 which are located in proximity

at the protein-protein interface. This trend was not visible at 45°C. Surprisingly, the signals of M18 and M55 could not be observed for compound 806 which could indicate a slow exchange behavior. We therefore decided to perform ^{1}H-^{15}N-HSQC experiments to see how the CSP for the amide backbone correlates with the data at 45°C and with the CSP of the methionine signals. Figure 3.6.2-3 compares the CSP for compound 806 at both temperatures.
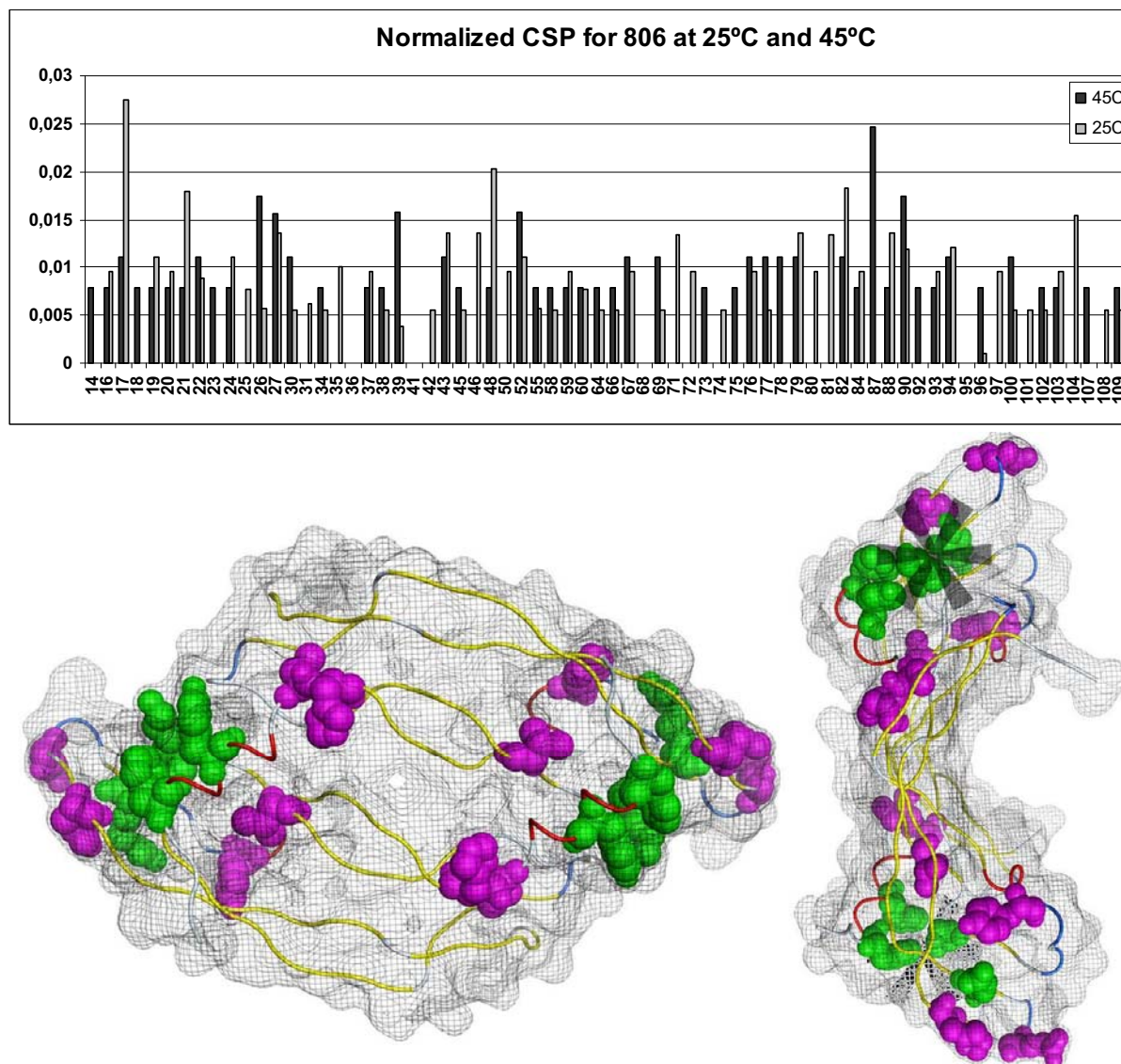
Figure 3.6.2-3. A: ^{1}H-^{15}N-HSQC CSP induced by compound 806 (2.5 mM) for non-overlapping residues of VEGF at 50 µM at 25°C and 45°C. B: Residues with significant pe rturbations are plotted on protein structure. Green and magenta residues mark significant CSP at 25°C and 45°C respectively. The asterisk in the right figure marks the protein-protein interface which contains M18 and M81. See table 3.6.2-2 for structure of fragment 806

The perturbations appear to be slightly stronger at 45°C. More importantly the residues that show a significant perturbation change drastically. At 45°C residues 87, 90, 26, 27, 39 and 52 show significant shifts. At 25°C this changes to residues 17, 82, 21, 48, all which are immediate neighbors in proximity

of M18 and M81 and are located in the protein-protein interface. In contrast, at 45°C the significant perturbations are dispersed over the protein.

The observed differences could be based on several phenomena: 1. Procedural errors during sample preparation and analysis. 2. Binding behavior is altered by significant modulation of free enthalpy by temperature change (ΔG= H- TΔS). 3. The conformational sampling or dynamics of the protein differ at the two temperatures modulating therefore molecular recognition of ligands. Due to time limitations we could not explore this issue further.

Independent of the theoretical considerations the results suggested a modified workflow to retest several compounds and their analogs may be of value: A sample of 50 uM VEGF was titrated in four or five steps with compound and compared to reference samples exposed to the same concentration of DMSO. During this assay the concentrations of compounds were monitored by [1]H-experiments to ensure that the assumed concentration was present in the sample. Tested compounds were grouped into scaffold families based on their structures. The compounds were then rated according to the CSP behavior and the changes of signal intensity during the titration. The highest rating was assigned to compounds that showed a dose response behavior and relatively strong CSP or signal intensity decrease. The lowest rating was assigned to compounds that exhibited the inverse behavior. Compounds with weak CSP but dose response behavior were rated over molecules with the opposing characteristics. If the CSP showed a dose response behavior, the data was fitted to a mathematical model assuming two independent identical binding sites[1]:

$$\Delta\delta_{CH} = F * \frac{[L_0]-[L]}{[P_0]} = F * \frac{K_D + [L_0] + 2[P_0] - \sqrt{(K_D + [L_0])^2 + 4[P_0](K_D - [L_0] + [P_0])}}{2}$$

whereby $\Delta\delta N_{CH}$ is the normalized CSP; $[P_0]$ is the total protein concentration; $[L_0]$, the total ligand concentration; $[L]$, the concentration of unbound ligand in solution; F, a scaling factor; and K_D, the affinity to be calculated. Figure 3.6.2.4 shows the representative behavior from 4 compounds.
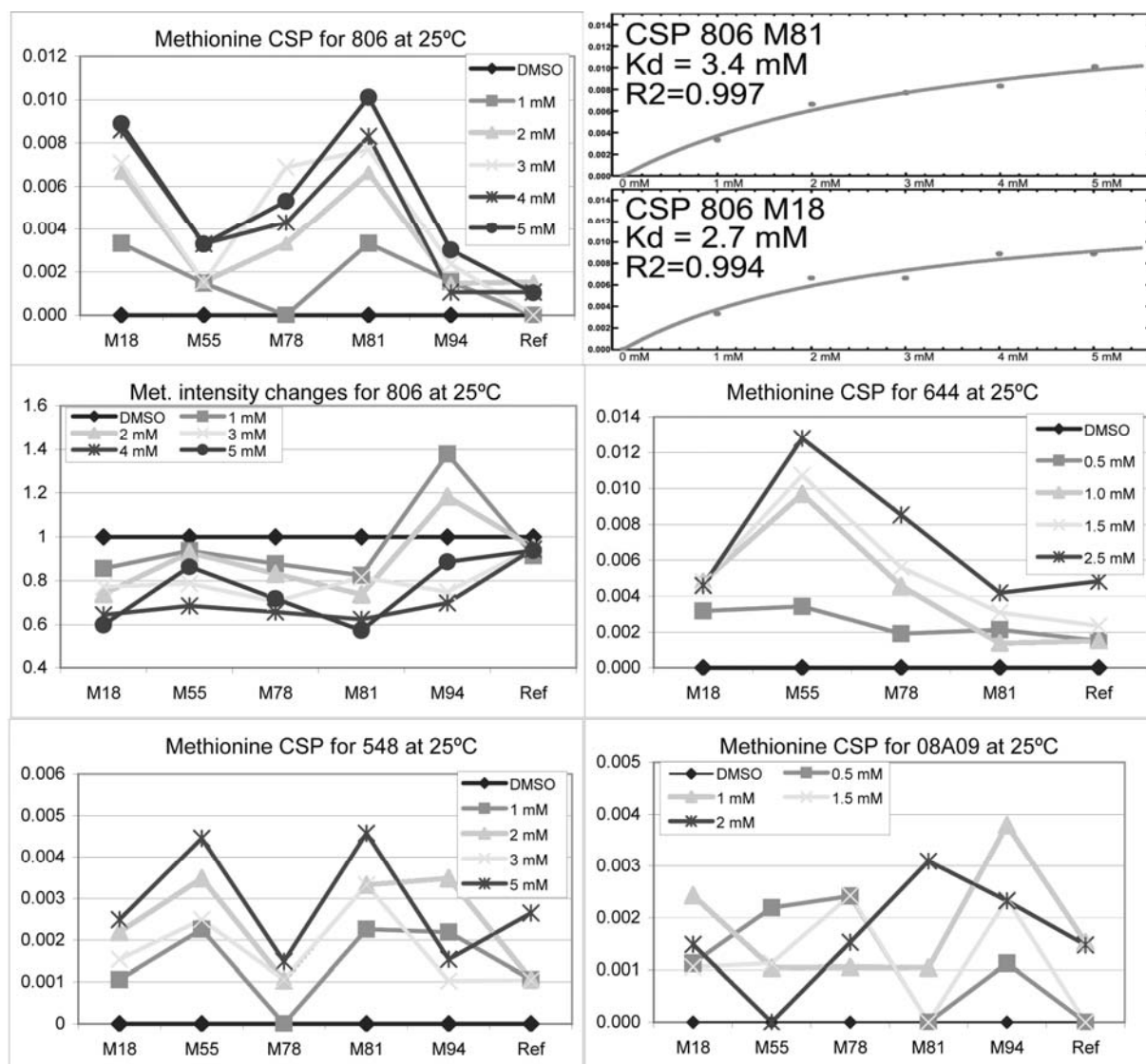
Figure 3.6.2-4: Representative examples of ^{1}H-^{13}C-HSQC CSP behavior for the five [methyl-^{13}C]-methionines of VEGF (50 μM) and a reference when titrated with selected compounds. The two graphs (top right) depict the mathematical analysis for M18 and M81 of 806.

Compound 806 showed relatively strong CSP behavior for M18 and M81 while the intensities of both signals decrease with increasing concentrations. In the initial one concentration assay M18 and M55 were not visible due to a lower S/N ratio. M18 and M81 show a reasonable dose response behavior that translated to an estimated affinity around 3 mM. Not all compounds that induce strong CSP affect M18 and M81. Compound 644 for example exhibits the biggest changes for M55. In general, perturbation of M18 and M81, often in parallel, was more common than perturbation of the remaining methionines. This could either indicate the higher propensity of the hotspot at the protein-protein interface to bind small compounds or may be a result from our previous selection effort which was targeting this area of the protein surface.

Compound 548 exhibited a weak CSP. But the perturbations could follow a dose response curve and would translate to affinities around 1mM. This behavior, low CSP and a relative good affinity could be

expected if the compound does not bind in proximity of the methionines. Or could just be an artifact or noise. Therefore we ranked compounds with this behavior lower as compounds as 806 and 644. The majority of compounds showed a CSP behavior as 08A09 where the magnitude of CSP is low and does not follow a dose response. Compounds with this behavior were rated as non-binder. Compound 08A09 is a representative example of compounds with a high density of hydroxyl groups that were selected during our computational exercises from our in lab libraries containing approved drugs and natural products. The majority of these compounds did not show any propensity to bind to VEGF. As we learned later, this type of compounds are very challenging for docking and are frequent false positives. As a consequence we tested several compounds that did exhibited lower docking scores and density of hydroxyl groups. Several of these compounds such as 04B06 tested in the initial ^{1}H-^{13}C-HSQC assay did show propensity to bind to VEGF.

Tested compounds with similar features were grouped into scaffold family and ranked upon their CSP behavior. In some cases the rating between family members was straight forward, in other cases not due to discrepancies between the quality of dose response, quantity of CSP, signal intensity changes and estimated affinity. In our opinion the data is not solid enough for a quantitative assessment, but should allow a rough qualitative ranking.

Figure 3.6.2-5 presents the data for the diphenylamine family of which 04B06 was identified as first candidate.
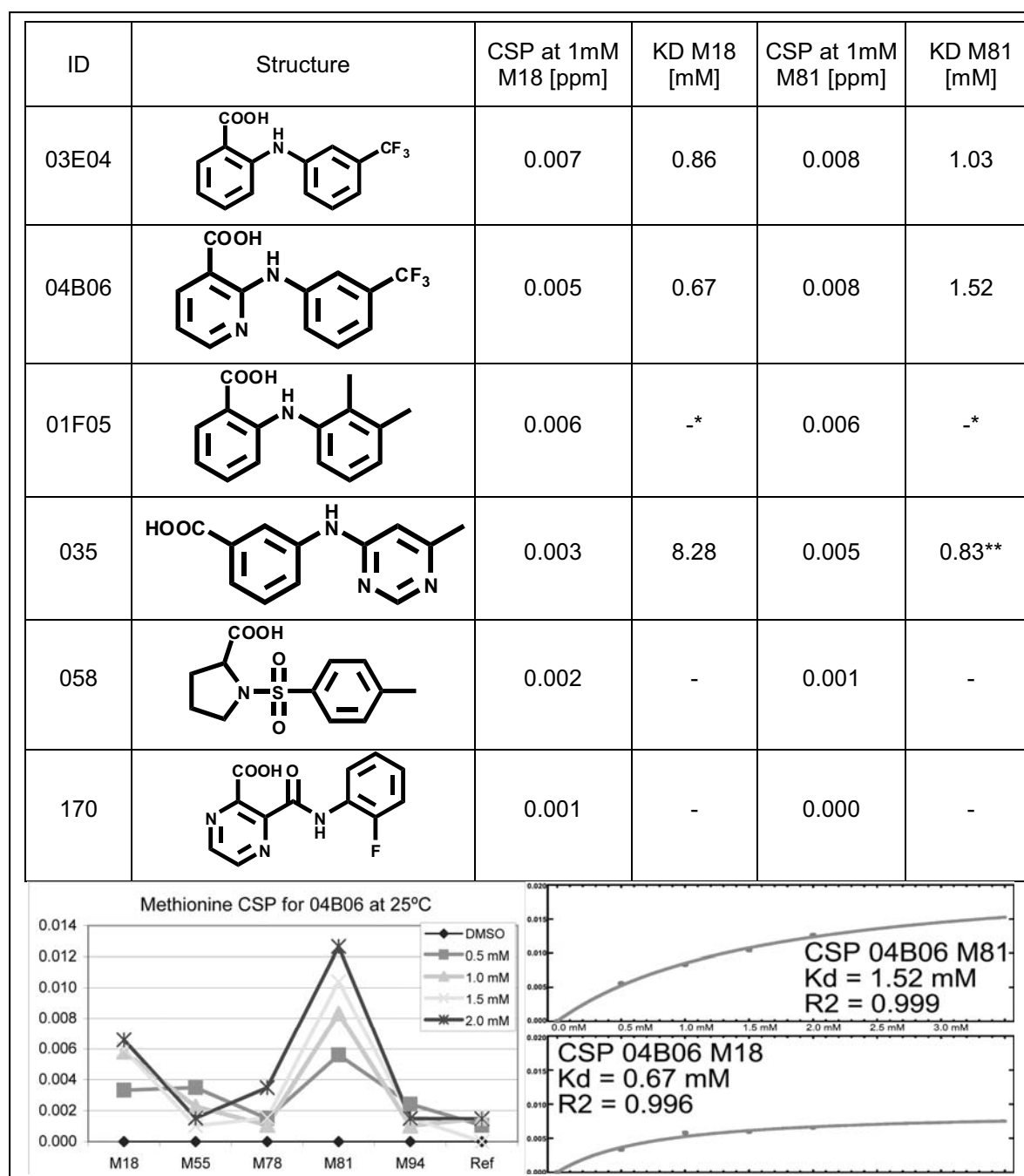
ID	Structure	CSP at 1mM M18 [ppm]	KD M18 [mM]	CSP at 1mM M81 [ppm]	KD M81 [mM]
03E04		0.007	0.86	0.008	1.03
04B06		0.005	0.67	0.008	1.52
01F05		0.006	-*	0.006	-*
035		0.003	8.28	0.005	0.83**
058		0.002	-	0.001	-
170		0.001	-	0.000	-

Figure 3.6.2-5: Summary of compounds from the diphenylamine scaffold family and their binding behavior assessed by ^{1}H-^{13}C-HSQC CSP. Details of CSP behavior and mathematical binding analysis for compound of 04B06. (* only soluble to 1 mM no dose response curve acquired.** R2 value is bad, no clear dose response.)

In total we identified in the second round of assessment five scaffold families and two individual promising compounds that are listed in table 3.6.2-2.
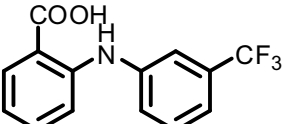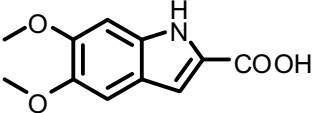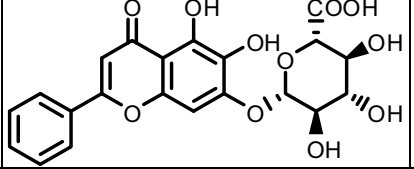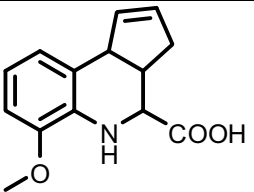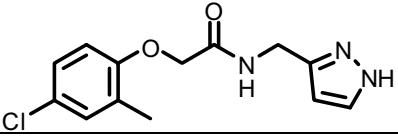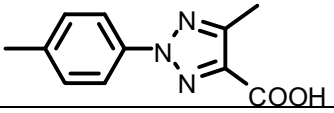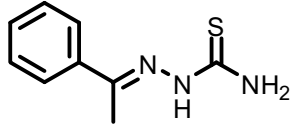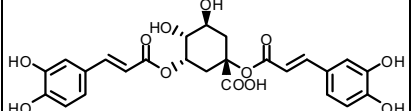
Scaffold	Best candidate	Structure	Highest CSP at 1mM (Methionine)	Kd for this residue, (R2)
diphenylamine	03E04		0.003 (M81)	3.4 mM (0.997)
indole	374D		0.003 (M81)	20.0 (-) , for other residues around 2 mM
baicalin	Baicalin		0.015 (M81)	10.7mM (0.999), 4 mM for intensity decrease but with bad R2
tetrahydroquinoline	806		0.003 (M81& M18)	3.4 mM (0.997)
linked arom.rings	644		0.009 (M55)	1.6 mM (0.987), 4 mM for another residue
biaryl	241		0.002 (M81)	9.4 mM (0.980)
-	578		no CSP but intensity decrease (M78)	2.1 mM (0.993 for Intensity decrease)
-	Phyto0036		0.004(M78)	0.5 mM (0.950)

Table 3.6.2-2: Scaffold families and individual compounds of interest identified in the second round of assessment. Stereoconfiguration is depicted if known.

The majority of compounds show [methyl-^{13}C]-methionine CSP around 0.003 ppm at 1mM concentration and affinity values in the low mM range. Exceptional is the affinity of Phyto0036 however the R2 value and the magnitude of CSP is average making this value questionable. Baicalin exhibits the strongest CSP of the table. Although this indicates a good potency the calculated affinity values were over 10 mM, the highest value of the table. Baicalin was identified as a VEGF ligand by Dr. Ricard Roddríguez Mias, whose work predates our studies. Dr. Ricard Roddríguez Mias obtained affinity values between 4.9 and 6.8 mM for Baicalin. However he did not eliminated the influence of the vehicle DMSO which may explain the difference to the values obtained in our experiments. Baicalin was already applied for the setup of the NMR screening conditions. Although we too were considering exploring Baicalin and its analogues previously, we waited until stage. Our decision was

based against the backdrop of the apparent low affinity of baicalin, together with its relatively high molecular mass. Further there were issues with the possibility of aggregates that may bias binding assays. Exploring baicalin analogues we were not able to obtain higher affinity compounds of a similar scaffold.

The best candidate of each scaffold family and the two compounds 578 and phyto0036 were further analyzed by acquisition of a ^{1}H-^{15}N-HSQC spectrum to characterize the binding site on the protein with a one residue resolution as performed for compound 806. To gain a description of the binding complex with atomic resolution we docked the candidates as described previously first with Glide close to the putative binding site and explored then the resulting protein-compound complex with the PELE methodology as described above. Again we selected up to three complexes per compound that should describe the putative binding conformation(s). A retrospective analysis of this exercise in the previous round showed that several compounds that had a reasonable prediction to bind to the protein, failed to do so. Although this may have several underlying causes, our concern was that one of the main factors may be the selection of the wrong complexes from the ensemble generated by PELE. So far, we have based our selection on the docking score of the complexes. However, we knew from the previous examples that our confidence in the docking score should not be very high. Further, the score compares complexes with different protein conformations. Certainly not all of these conformations could be observed in solution and certainly some of them are exited estates. Ideally, the docking score should be corrected by the energy of the exited state to obtain real values or a ranking of different ligands and binding poses. Since this is not possible the docking score may be highly misleading as it is. We decided to approach this problem with the following proposition: 1. Only GRIDS are considered that are in accordance to the observed CSP of the ^{1}H-^{15}N-HSQC spectrum. 2. Grids are ranked according to their cluster size and their docking score with emphasis on the first. 3. The 1 - 3 top ranked grids were selected that explain the ranking of affinity for the scaffold families (e.g. can the ranking be explained with the binding mode). 4. If several grids were selected we favored the selection of grids with different ligands positions to cover different binding scenarios. An example for this is presented for compound 806 in figure 3.6.2-6.
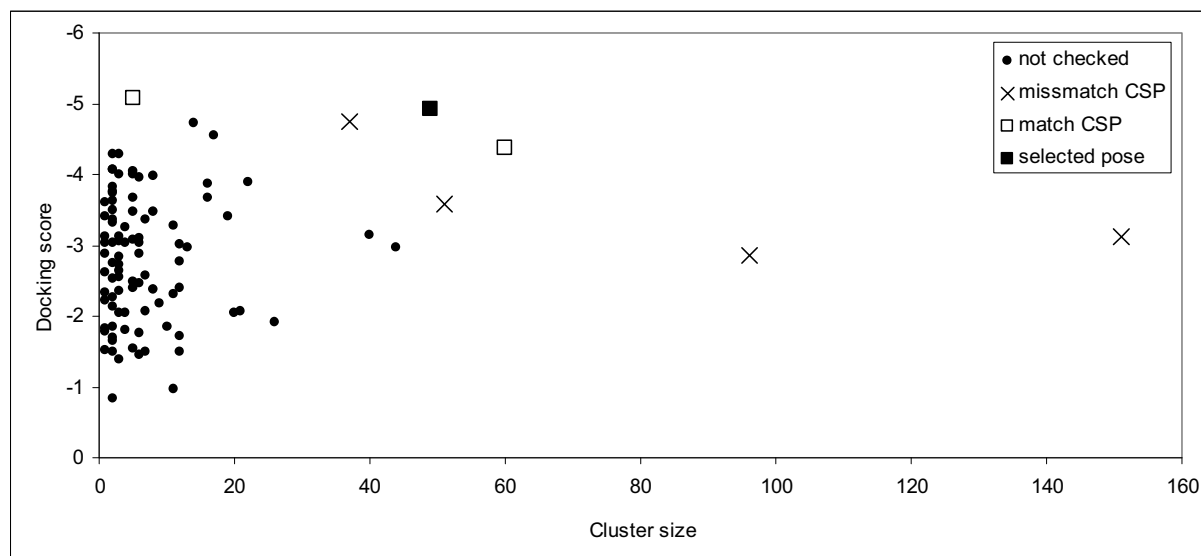
Figure 3.6.2-6: Cluster analysis of PELE trajectory and redocking of representative snapshots for compound 806. Grids with high cluster size and docking scores were checked visually and an appropriate grid selected.

For the case of 806 the highest populated cluster showed a ligand binding pose that did not correlate with CSP binding data. Of the three binding modes that were in agreement with the CSP we selected the one which was rated second in terms of both docking score and cluster size. At this stage, we did not select an additional grid, although we could have selected the highest populated binding pose that matched the CSP signature. Further the tetrahydroquinoline family did not have enough members to discharge specific grids. Therefore, the selected grid was our best proposal for the binding mode of 806. It is depicted in figure 3.6.2-7. As may be noted the original structure (2VPF) and the selected grid have a high similarity with an RMSD of 1.1 Å. But in the selected protein conformation the residues Phe17 and Met18 change the orientation of their side chains to open a transient pocket which is not present in the crystal structure. Residues Phe17, Tyr 21 and Lys 48 (depicted in red) exhibited significant perturbations in the ^{1}H-^{15}N- HSQC experiment and are in immediate proximity to the ligand. The same is true for the side chains of M18 and M81 which exhibited significant perturbations in the ^{1}H-^{13}C- HSQC experiment.
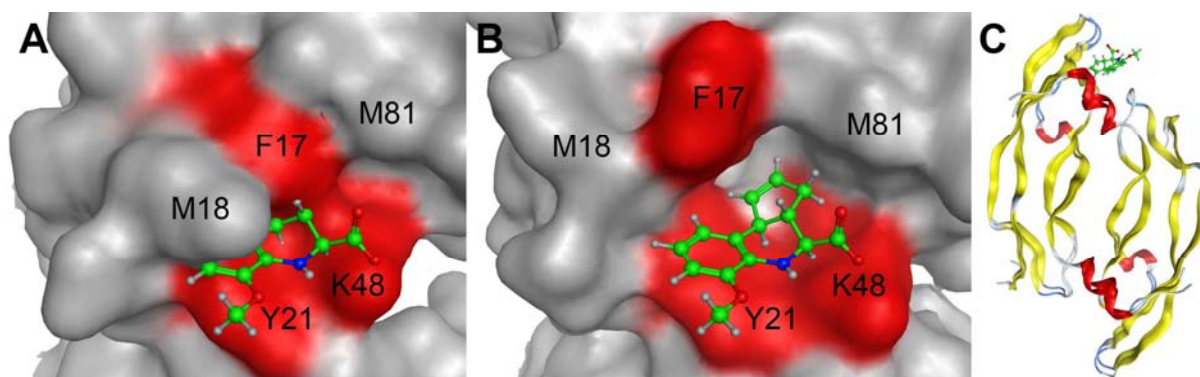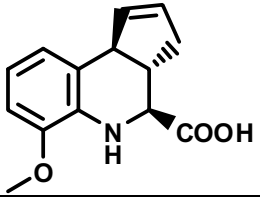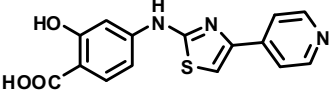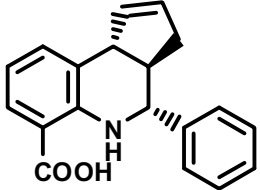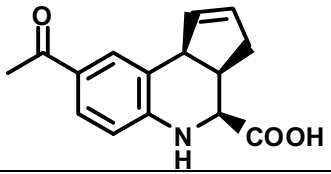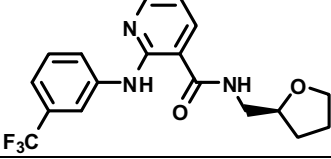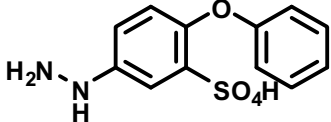
Figure 3.6-2-7: Selected ligand pose from PELE simulation overlapped with native X-ray structure (A) and PELE generated protein conformation (B). C: Backbone comparison of both structures with indication of the bound ligand (green).

(1)   Goldflam, M.; Tarrago, T.; Gairi, M.; Giralt, E. *Methods Mol Biol* **2012**, *831*, 233.

3.6.3 Third round of elaboration

To increase the affinity of our compounds and to test our ability to do so we explored analogs in a third round of assessment. As a starting point we used 11 grids that were generated as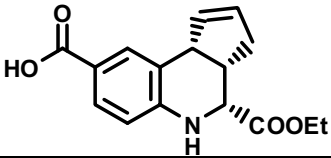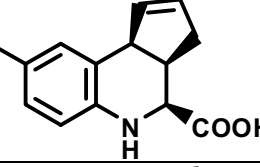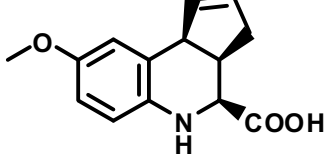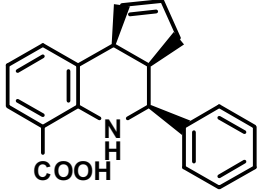 putative binding poses for the scaffold families of diphenylamine, indole, tetrahydroquinoline, biarly and the linked aromatic rings. The remaining families were discharged.

Due to the problems that arose by selecting compounds from the scifinder CAS database we limited ourselves to the libraries of three international vendors, Asinex, Life Chemicals and Enamine which led to a merged library of 200k compounds after limiting us to a molecular weight under 400 Dalton and a log S value better than -4. To simplify the data analysis we selected a subset of approx. 200 compounds that showed a MACCS Fingerprint similarity of at least 60% to one of our candidate compounds. This subset was docked with Glide XP to the 11 selected protein conformations that were previously selected as putative binding conformations. We purchased in total 34 compounds for the five scaffold families. These compounds were tested for their ability to bind to VEGF as described in the previous round: First by monitoring their CSP behavior by observing the [methyl-^{13}C]-methionine signals during a titration and then after ranking similar compounds by performing a ^{1}H-^{15}N- HSQC experiment for the best candidate to gain additional information about the binding site. The computational predictions and the experimental result are summarized as an example for compounds 806 in table 3.6.3-1 ordered after the docking scores. The table contains the top docked compounds that we were able to purchase.

ID	Structure	Possible isomer?	Docking score	Exp. ranking	CSP at 1 mM (residue)	Kd of this residue by CSP(R2)	Kd of this residue by Int. change(R2)
806		no	-4,99	parent comp.	0.003 (M81)	3.4 mM (0.997)	No fit
843		-	-6,08	insolub.			
(834)		no	-5,69	-			
831		yes	-5,68	1a	0.016 (M81)	5.7 mM (0.999)	0.7 mM (0.994)
817		-	-5,31	non binder			
841		-	-5,13	1b	0.010 (M18)	5.0 mM (0.999)	0.4 mM (0.999)
820		yes	-5,09	non binder			
832		yes	-5,00	2	0.009 (M81)	8.3 mM (0.999)	5.9 mM (0.997)
822		yes	-4,99	3	0.007 (M81)	0.4 mM (0.998)*	-
834		yes	-4,88	1c	0.020** (M81)	0.3 mM (0.917)	0.1 mM (0.997)
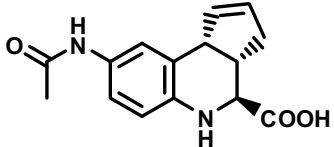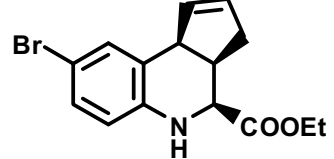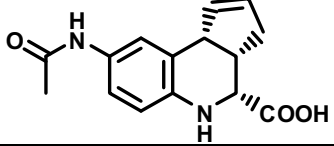
(830)	*structure*	no	-4,73	-			
814	*structure*	yes	-4,11	non binder			
830	*structure*	yes	-3,81	2-3?	0.009 (M81)	3.3 mM (0.996)	1.8 mM (0.920)

Table 3.6.3-1: Summary of purchased 806 follow-ups compounds, ordered after docking score and their experimentally validation by ^{1}H-^{13}C-HSQC CSP at 25℃. (*only four data points, **esti mation, only soluble until 0.5 mM)

In total we ordered ten compounds of two types as follow-ups for 806. The first type had the same tetrahydroquinoline scaffold as 806 and the second belonged to different scaffolds. Compounds of the second type had a high similarity to one of the other candidate compounds but were ranked very high for the selected protein conformation for 806. Problematic was the missing information of the correct stereoconfiguration of 806. It has three stereocenters which leads to a total of 16 possible stereoisomers. The company which sold the compound did not supply us with any detailed information about the stereo configuration but was claiming that the compound showed only one peak in standard RP-HPLC conditions which could be confirmed. This indicates that the compound may be present as a racemic mixture. We used for the initial docking exercise of compound 806 the configuration depicted in table 3.6.3-1. We decided to use this relative flat structure since it should have a lower steric repulsion than more folded configurations and could be a possible product of a thermodynamic reaction control. Further, the shape of this conformer is a good average of the possible configurations. We discovered later that this assumption was wrong since compound 806 and analogs are synthesized by an aza-Diels-Alder reaction which favors kinetically the formation of the two enantiomers which exhibit the endo orientation of the cyclopentene and carboxyl group. [1]

Since we were not aware of this, we generated all possible stereoconfigurations for the approx. 200 compounds that we used for docking to the selected protein grids from the five scaffold families. Interestingly the majority of 806 analogs with top docking scores that were purchased from this grid exhibited one of the two possible stereoconfigurations. The only exceptions were the compounds 834 and 830 which are listed first with their compound ID in brackets marking the wrong stereoconfiguration with the higher docking score and later in the list the compound ID without brackets marking the first possible enantiomer with the higher docking score.

For the example of the tetrahydroquinoline family the majority of purchased compounds showed improved binding to VEGF compared to the original compound. This was assessed by the magnitude of CSP during the titration experiment by observing the [methyl-^{13}C]-methionine signals and the estimation of affinity which was based on the dose-response behavior of CSP and significant intensity changes which could be observed for the majority of compounds. Experimental ranking of similar compounds based on the magnitude of induced CSP at 1 mM was very similar to the order predicted by the docking score. For these cases the experimental behavior could be explained by the obtained docking poses (Figure 3.6.3-1), which predict no favorable interactions for the methoxy-group of 806 (both for the used and allowed configurations) contrary to the additional hydrogen bond present between the carbonyl oxygen of 832 and Glu 22.
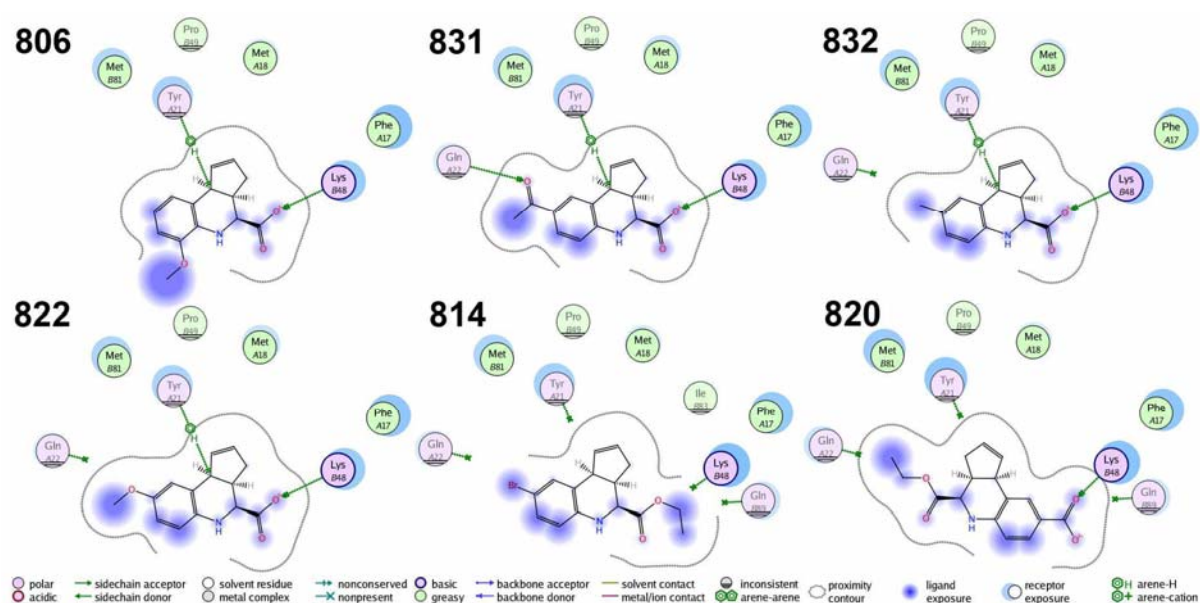

Figure 3.6.3-1: Ligand interactogramm for members of the tetrahydroquinoline family from table 3.6.3-1. Generated with MOE.

Compounds 832 and 822 have a functional group at the same position as 831 but can not form this favorable interaction. Compound 814 does not posses a free negative charged carboxyl-group, which disrupts the interaction to Lys48 present in the complex of 806 and resulted in non binding of this compound. Compound 820 contains an additional carboxyl group which can form this interaction, but only in a flipped binding mode and for the opposite enantiomer of the previous configurations. Since this compound does not bind is appears that this configuration and binding mode are wrong.

In our opinion one of the most promising compounds from the new selection is therefore 832. It exhibits a five fold increase of CSP at 1 mM compared to 806. Its affinity is estimated to 5.7 mM (1.7x decrease) based on CSP and 0.7 mM (5x increase) based on the changes of signal intensities. This discrepancy can be observed for several compounds. Possible explanations include: appearance of non-specific interactions; the incapacity of the [methyl-^{13}C]-methionine experiment or our data analysis to yield a correct description of the interaction, the use of racemic mixtures instead of pure
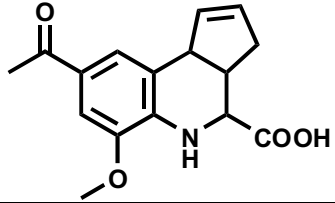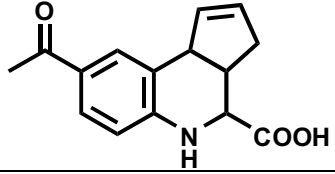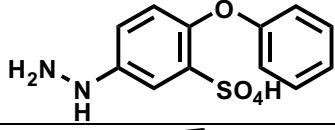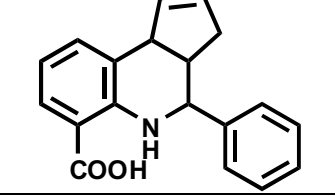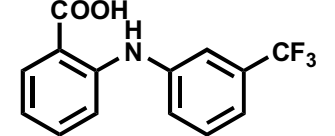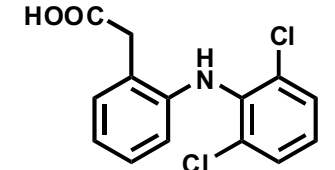
individual compounds. Although we can not rule out the first case, the localized CSP of the ^{1}H-^{15}N-HSQC experiments and the correlation between magnitude of CSP between 806, 831, 832, 814 which matches the docking prediction let us assume the observed discrepancy has other origins. Further, both the magnitude of CSP at 1 mM and the affinity based on intensity changes increase by approx. five fold making an affinity value under 1 mM more probable than a decrease of affinity compared to 806. It would be interesting to compare affinity values obtained by observing [methyl-^{13}C]-methionine signals to values obtained by observing the 1H-^{15}N-amide backbone signals. Unfortunately acquisition of this information would require one day per compound at 25°C.

Compound 834 had only a solubility of 0.5 mM but showed promising CSP and affinity values. It may have a better potency than 831, however, with a lower confidence. Since its docking pose suggests a slightly different mode of interaction we decided to separate it into a different scaffold family.

Sine the investigated compounds are very similar to the parent compound 806, the described improvements could only be based on serendipity. The docking exercise suggested to us also compounds from completely different scaffold families. One of the two tested compounds did exhibit a very promising behavior in the [methyl-^{13}C]-methionine assay. This opens up the exploration of a new scaffold family. Further this supports in our opinion the feasibility and practical value of our computational approach, in particular when compared to low success rates in the initial screening outlined in chapter 3.5 and during the first round of elaboration described in chapter 3.6.1.

Table 3.6.3-2 summarizes the results of promising candidates for all five in this round explored scaffold families. The best candidates were obtained from the already described tetrahydroquinoline scaffold family if we rate the compounds by magnitude of CSP and include Kd by intensity decrease, which could only be observed for these compounds. The second best family is the diphenylamine family with compound 836 exhibiting approx. four fold increase of CSP but exhibiting a lower gain considering the affinity estimation. The biphenyl family was of particular interest to us. We learned while our work was already in progress that Fesik and coworker reported a single small molecule based VEGF ligand in 2000.[2] This compound is a functionalized biphenyl and is depicted in the last row of table 3.6.3-2. Fesik and coworker reported an affinity of 0.5 mM for this compound. Unfortunately the article states only that this value was obtained by NMR at 45°C, but no data nor the accuracy of this value were provided. We could not purchase this compound due to its high price, but were able to test the analogue 845 which is missing only the bromine group. Our assessment for this compound gave an affinity value of 3.7 mM and a magnitude of CSP which was significant but lower than that for the best compounds from the previous scaffold families. An affinity around 3.7 mM appears to be over seven times lower than the value reported by Fesik for his compound. Although it contains an additional bromine group and the experiment is performed at 45°C, the difference between 0.5 mM and 3.7 mM appears to be quite large. This could indicate that either

Fesik´s affinity assessment is quite optimistic, or that we underestimate the affinity of our ligands. Compound 845 and 827 have similar affinity values and magnitude of CSP but different structural fictionalizations. A compound which unites these features would be a good candidate to improve the potency further. Although we tested several candidates for the indole scaffold family no significant gain was observed for the magnitude of CSP at 1mM. The estimated affinity of the new compounds suggests a slight improvement. However without a significant CSP, this value should be considered with caution. At this point we would assume that the original compound 374D was either a false positive, that the estimated binding site was wrong or that the site does not offer suitable opportunities to improve the potency of the scaffold family. Finally we did not purchase any new candidates for the linked aromatic ring families. On one side the highest CSP was observed for M55 which is located in the center of the protein indicating that the compound might bind in this area. On the other side it was very difficult to find promising candidates in the vendors libraries that we explored during the course of this study. Therefore, due to our budget limitations we decided on purchasing compounds for the other families.

Original scaffold family	ID	Structure	tested follow-ups	CSP at 1 mM (residue)	Kd of this residue by CSP(R2)	Kd of this residue by Int. change(R2)
tetrahydro-quinoline	806		9	0.003 (M81)	3.4 mM (0.997)	-
	831			0.016 (M81)	5.7 mM (0.999)	0.7 mM (0.994)
	841			0.010 (M18)	5.0 mM (0.999)	0.4 mM (0.999)
	834			0.020** (M81)	0.3 mM (0.917)	0.1 mM (0.997)
diphenyl-amine	03E04		12	0.003 (M81)	3.4 mM (0.997)	-
	836			0.013 (M81)	2.5 mM (0.998)	-
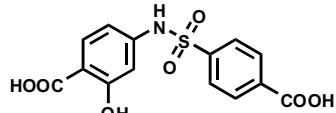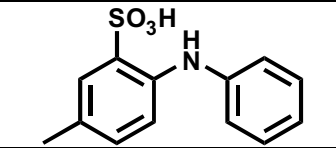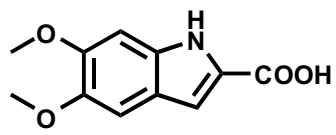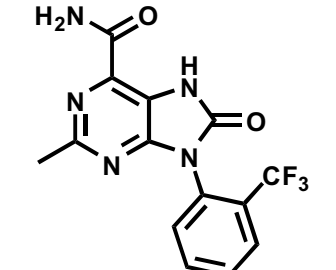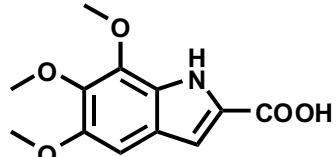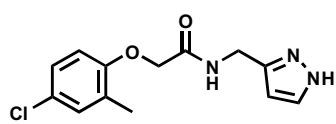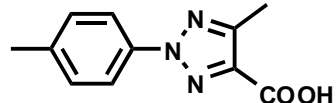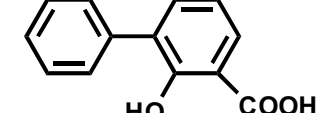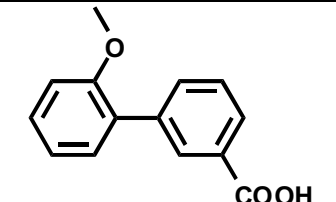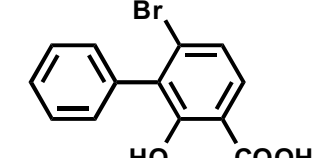
	839*			0.006 (M81)	2.2 mM (0.998)	-
	840**			0.004 (M18)	6.3 mM (0.999)	-
indole	374D		6	0.003 (M81)	20.0 (-), for other residues around 2 mM	-
	818			0.004 (M81)	1.6 mM (0.993)	-
	821			0.005 (M81)	2.6 mM (0.999)	-
linked arom. Rings	644		-	0.009 (M55)	1.6 mM (0.987), 4 mM for another residue	-
biaryl	241		7	0.002 (M81)	9.4 mM (0.980)	-
	845			0.01 (M81)	3.7 mM (0.999)	-
	827			0.009 (M81)	3.7 mM (0.996)	-
Fesik	-		-	-	0.5 mM (?)	-

Table 3.6.3-2: Summary of the five scaffold families that were explored in this round. For each family the first entry lists the best candidate from the previous round. The additional entries list promising compounds discovered in this round. (* could also belong to the biaryl family, since no backbone data was acquired, ** could also bind in proximity of 841 since no backbone data was acquired).

We acquired ^{1}H-^{15}N-HSQC spectra at 25°C for the best novel compounds of each scaffold family to characterize their binding behavior with a one residue resolution. Sample preparation and data analysis was performed as described in the previous round. The results are summarized in table 3.6.3-3.
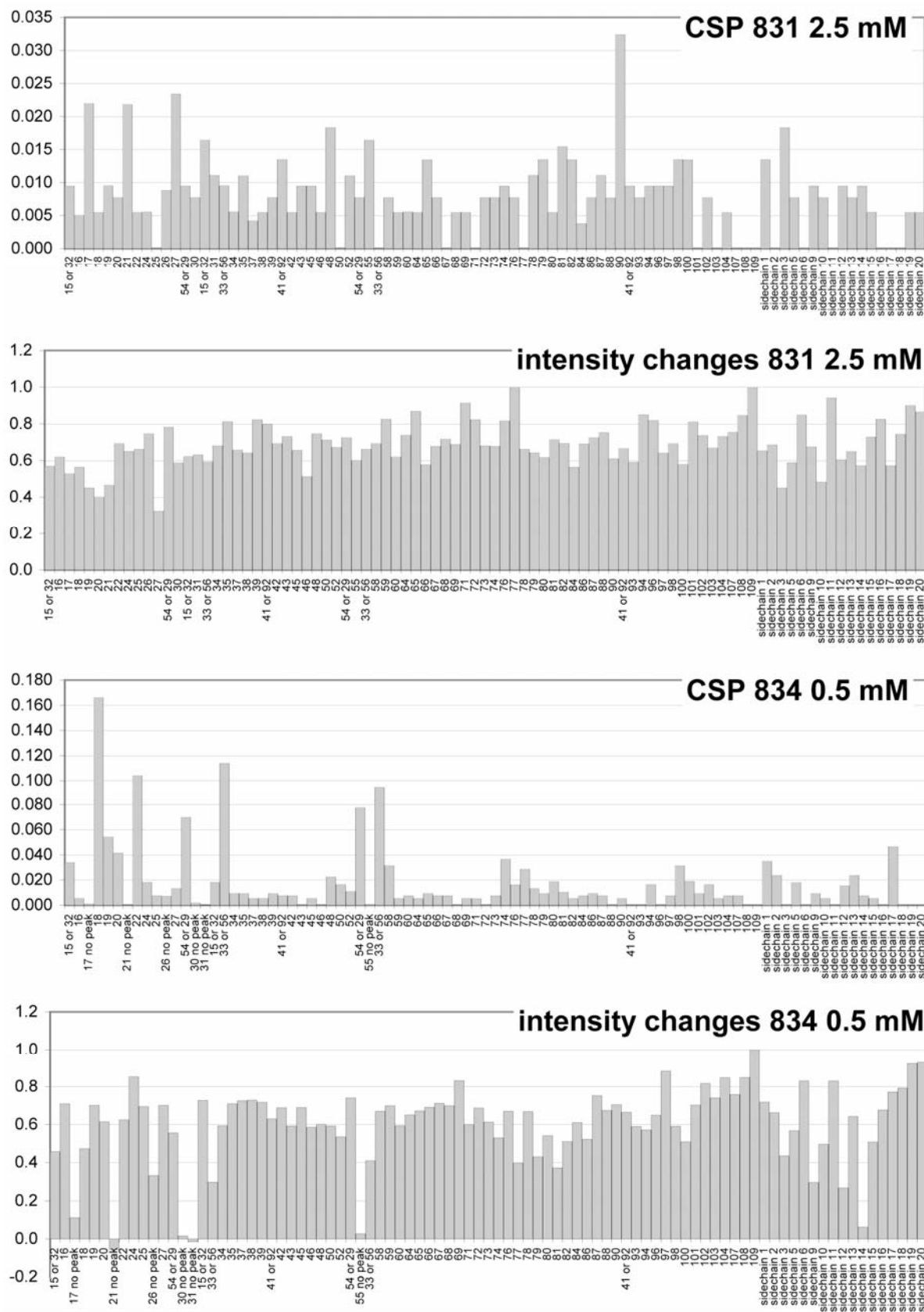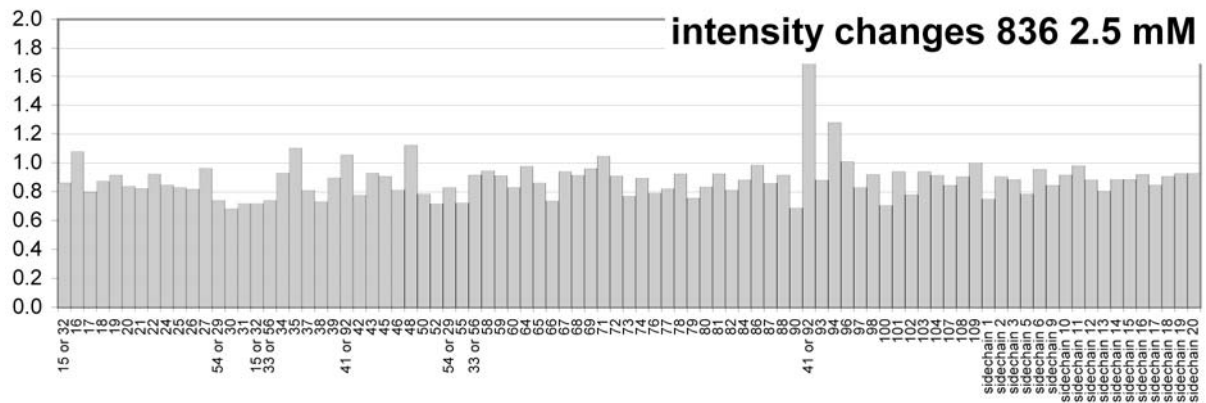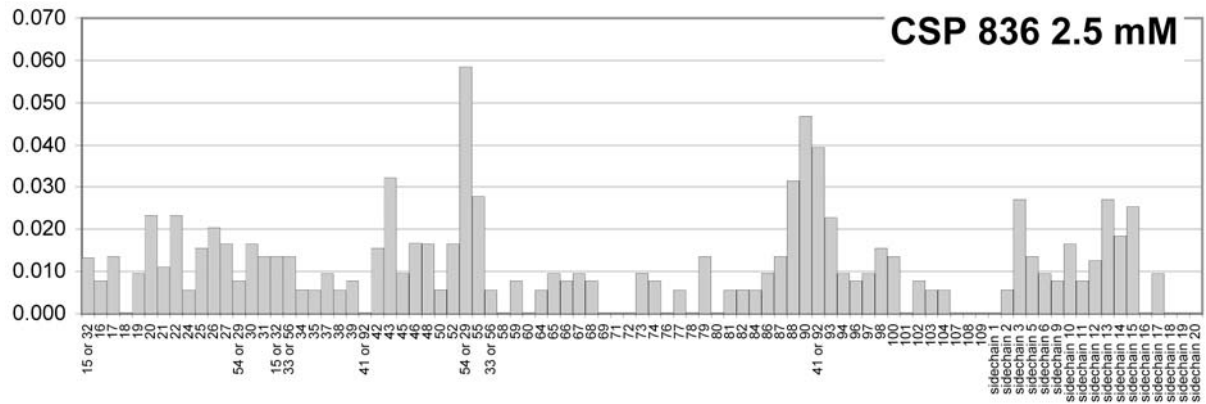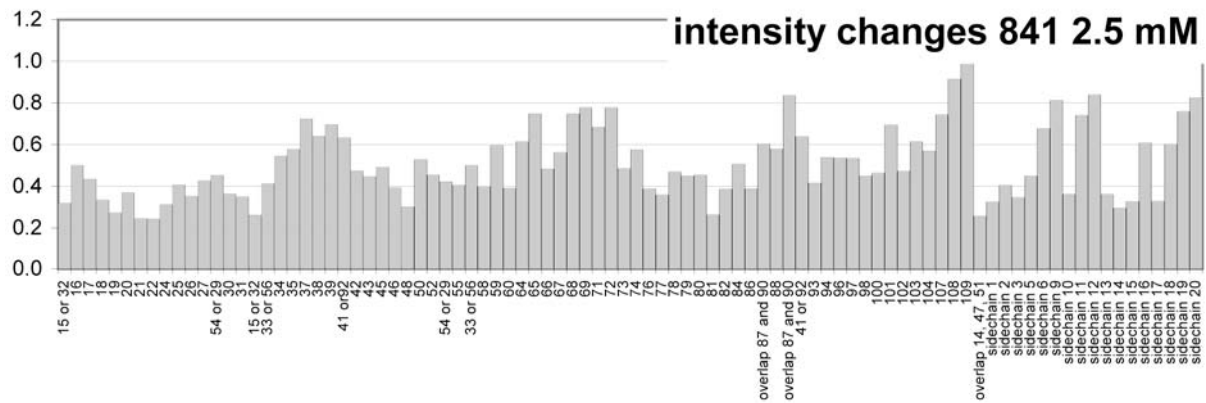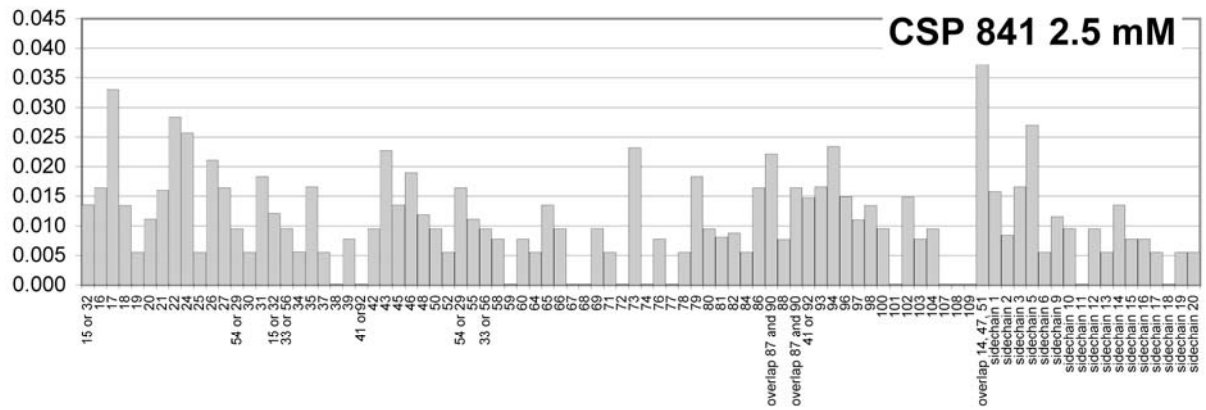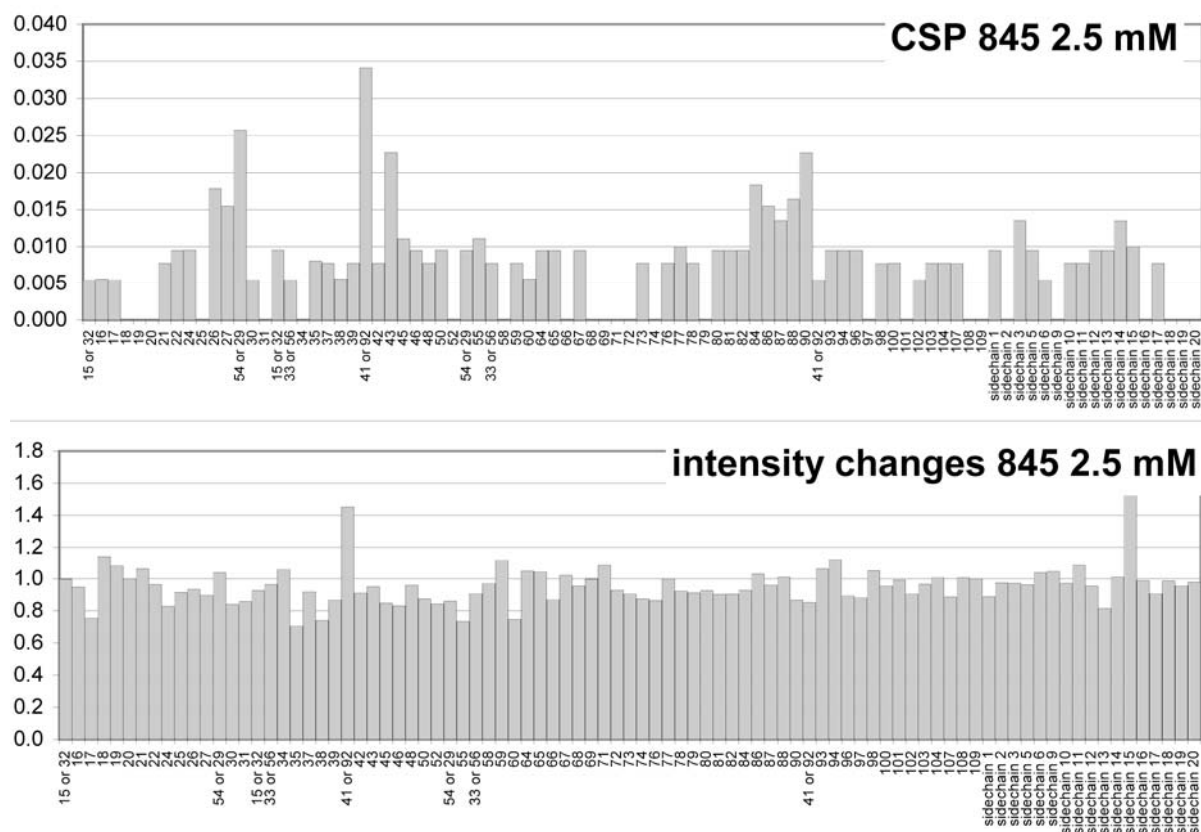
Table 3.6.3-3: Analysis of ^{1}H-^{15}N-HSQC spectra of 50 µM VEGF at 25℃ for CSP and intensity changes induced by promising compounds from table 3.6.3-2. Residues 27 and 90 overlap in the spectra for 831 and the real CSP is probably smaller than calculated.

All investigated compounds exhibit CSP that were significantly stronger than the parent compounds and show distinctive binding sites that match the areas of the parent compounds. Compound 841 binds despite the different scaffold to the same area as the other compounds of the tetrahydroquinoline family as it was described by our computational prediction.

The perturbations of the 1H-15N HSQC spectrum were especially striking for compound 834 although it was tested only at a concentration of 0.5 mM compared to 2.5 mM for the other compounds. The spectrum is depicted in figure 3.6.3-2 but the experiment should ideally be replicated to obtain exact values since the sample was incubated for a longer time than usually.
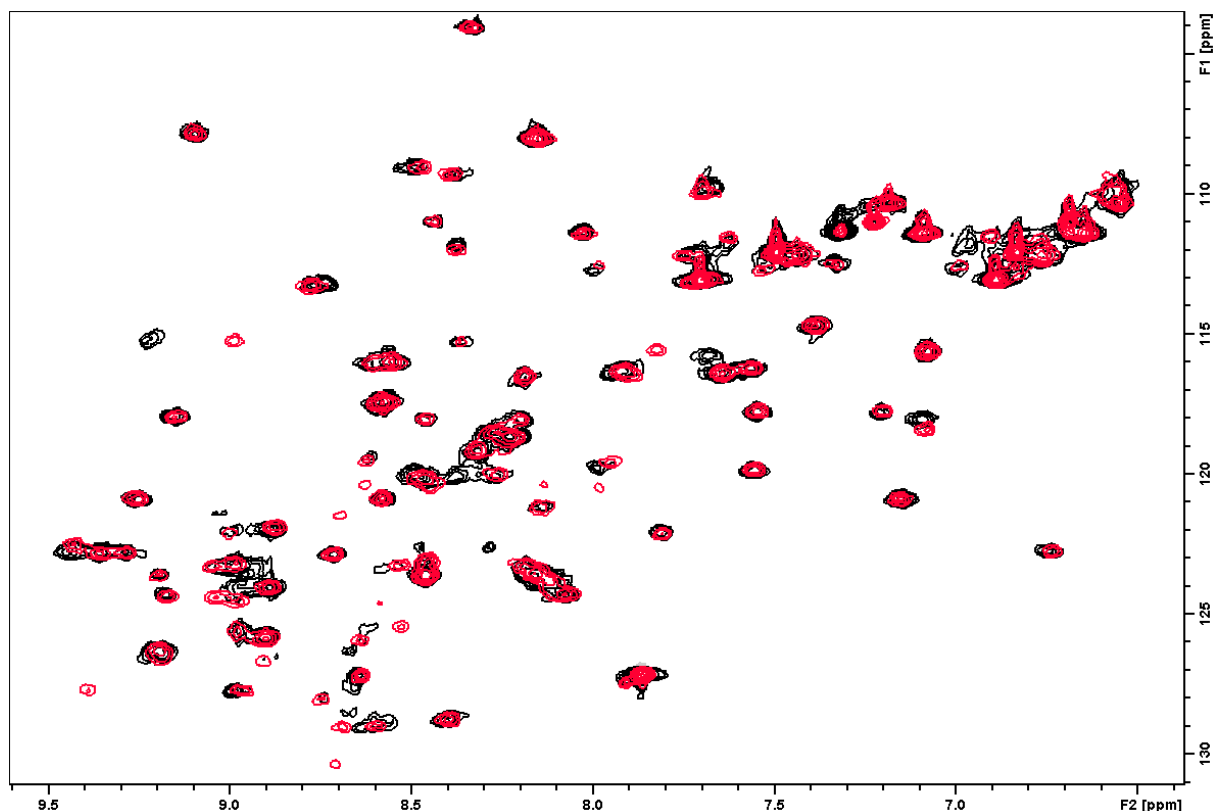
Fig: 3.6.3-2: ^{1}H-^{15}N-HSQC spectra recorded at 25℃ of 50 μM VEGF and 2.5% DMSO in the absence (red) and presence (back) of 0.5 mM 834.

The highest CSP for 834 was detected for the backbone amide of M18. Many signals that are located in the proximity of this residue and on the protein-protein interface disappeared, as it was the case for F19 and Y21. Further the ^{1}H-^{15}N signal that was labeled as side chain 14 disappeared completely. This signal is also strongly affected by the peptidic ligand v107 which binds to the protein-protein interface and we believe that this signal is the side chain amine of K48.

The strong CSP as well as the disappearance of signals indicated that the affinity of 834 is in the order of the concentration that was tested. Previously we estimated the affinity to 0.1 and 0.3 mM by the [methyl-^{13}C]-methionine titration assay which matches this observation. Considering the low number of only 22 heave atoms and assuming an affinity of 0.3 mM this would translate to ligand efficiency[3] of 0.22 as determined by:

$$\Delta G = -RT\ln(K_d) \text{ and } LE = \Delta G / HAC$$

With ΔG= Free energy (kcal), R= molar gas constant, T= temperature (K), HAC (heavy atom count LE= ligand efficiency.

Generally LE values under 0.2 are considered poor, over 0.4 as high and in between as satisfactory in the context of classical drug targets. Consequently, depending on the error of the affinity estimation, our LE value could change. Since we tested only the racemic mixture of 834 so far this value could

also be significantly higher if only one of the two possible enantiomers exhibits affinity to VEGF. Even if this is not the case, this is a very good LE in the context of protein-protein-interfaces.

Indeed, if one could find a similar potent compound in a proximal binding site and would successfully link them, this would lead to a low nM affinity ligand with a molecular weight close to 500 Da. We believe that a fragment linking scenario, or perhaps merging scenario would be risky with the present data, but could be feasible in the future.

In figure 3.6.3-3 the distinct binding areas of our compound families are depicted. The surface that is in direct contact with the receptor VEGF-R2 was colored in red. We have a fairly high confidence with respect to compound pairs 841, 834 and 836, 845. One could imagine either to merge the individual compounds of each pair, or the two pairs could be linked. Although the distance is relatively long, it spans over the receptor binding interface, additional interactions that increase the potency of the inhibitor could be made. Compounds 644 and 578 require additional validation. Although they do not bind to the receptor interface, they could be useful as affinity anchors if linked to 841 or 834. We have only a limited confidence in compound 818 however and one should find a higher affinity compound to validate this binding site first. Finally, Baicalin is located at the biding site as proposed by Dr. Ricard Roddríguez Mias. Placing this compound on the receptor interface but at a different position may be a worthwhile option to consider.

Although the majority of the data would require additional validation, we believe it offers a positive perspective for future fragment linking or merging scenarios.
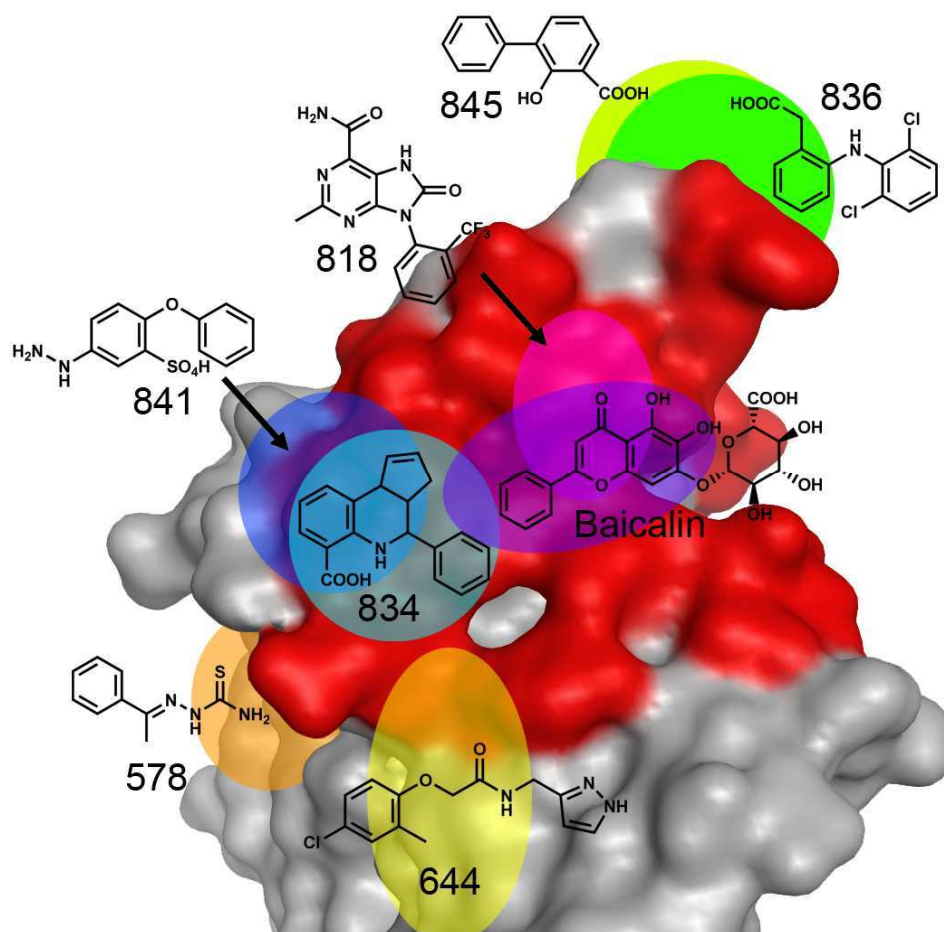
Figure 3.6.3-3. VEGF surface representation from 3V2A. The surface in direct contact with receptor VEGF-R2 is colored in red. The oval zones mark the putative binding sites of compounds.

We wanted to obtain an exact description of the binding modes for compounds 831 and 834 and explore the influence of their stereoconfigurations. For this purpose the two possible enantiomers of both compounds (Figure 3.6.3-4) were docked with Glide to the binding areas indicated by ^{1}H-^{15}N-HSQC CSP. The four grids were used as starting points for a PELE simulation as explained in the previous rounds. The trajectories were clustered and one representative redocked with Glide to the individual protein conformation. These computational exercises were again performed by Dr. Martin Kotev. We analyzed the data as described above by plotting cluster size versus the docking score of the representative snapshots (figure 3.6.3-4). Snapshots with high cluster size and docking score were inspected manually for mismatches regarding CSP and QSAR. For example the second most populated cluster of RSS-831 could not explain the increase of affinity by the presence of the keto group and was therefore discharged.

For compound 831 the SRR configuration seems to have a higher potency than the RSS configuration. The first has higher docking scores and the two most populated clusters have actually a near identical binding pose and more members. Unfortunately, we do not have further evidence, whether the obtained differences between SRR and RRS would be significant or not.

For compound 834 the RSS configuration seemed to be more potent. Since the highest populated cluster of SRR mismatched with the CSP information, the first possible cluster was significantly smaller than the RSS cluster. Compared to 831 the difference of the two enantiomers seemed to be more evident, but again we did not know if it is meaningful. It would be interesting to obtain pure enantiomers of both compounds to assess this issue.
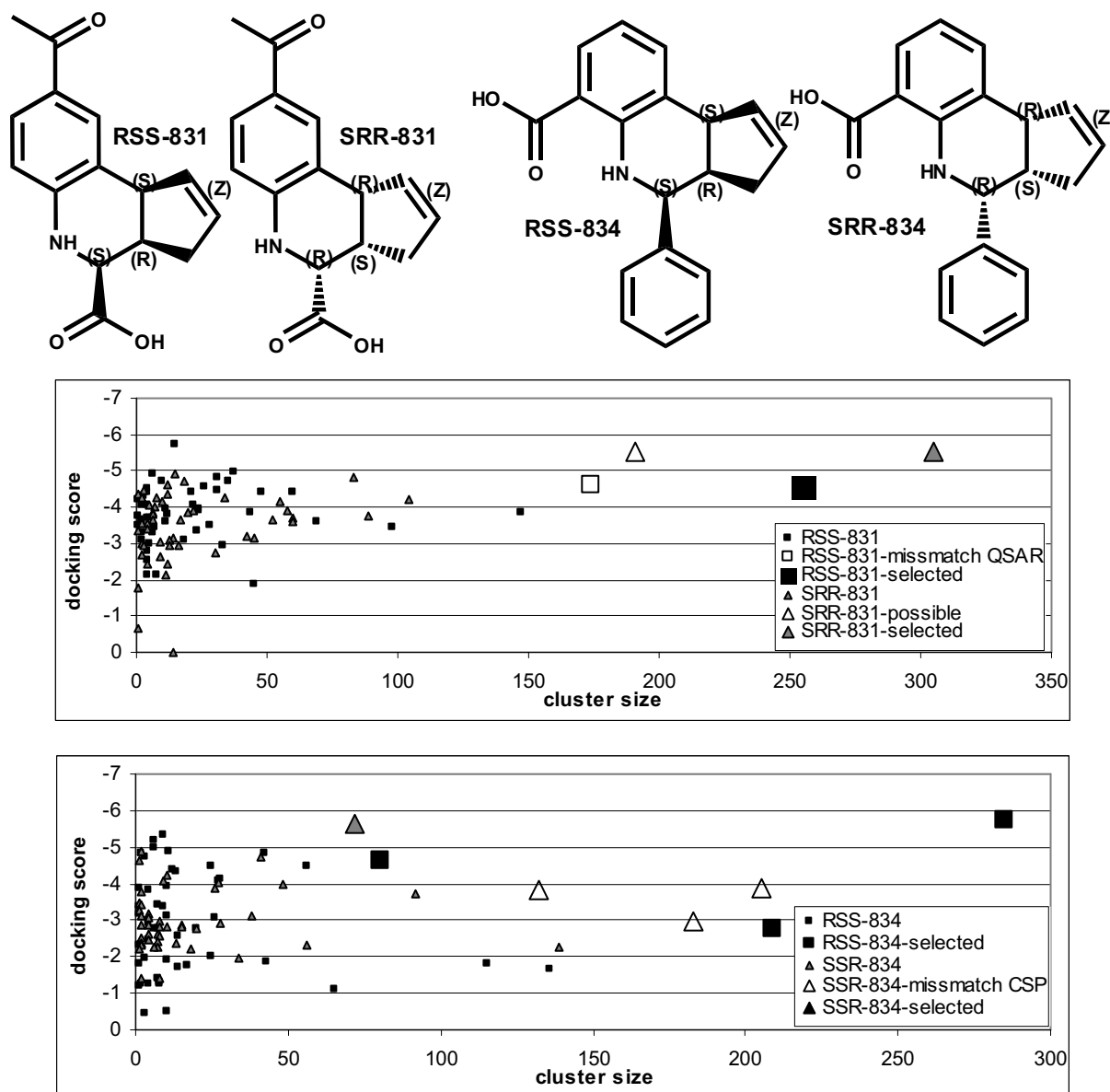
Figure 3.6.3-4 Structure of possible enantiomers of 831 and 834 and Analysis of PELE trajectory for all four enantiomers.
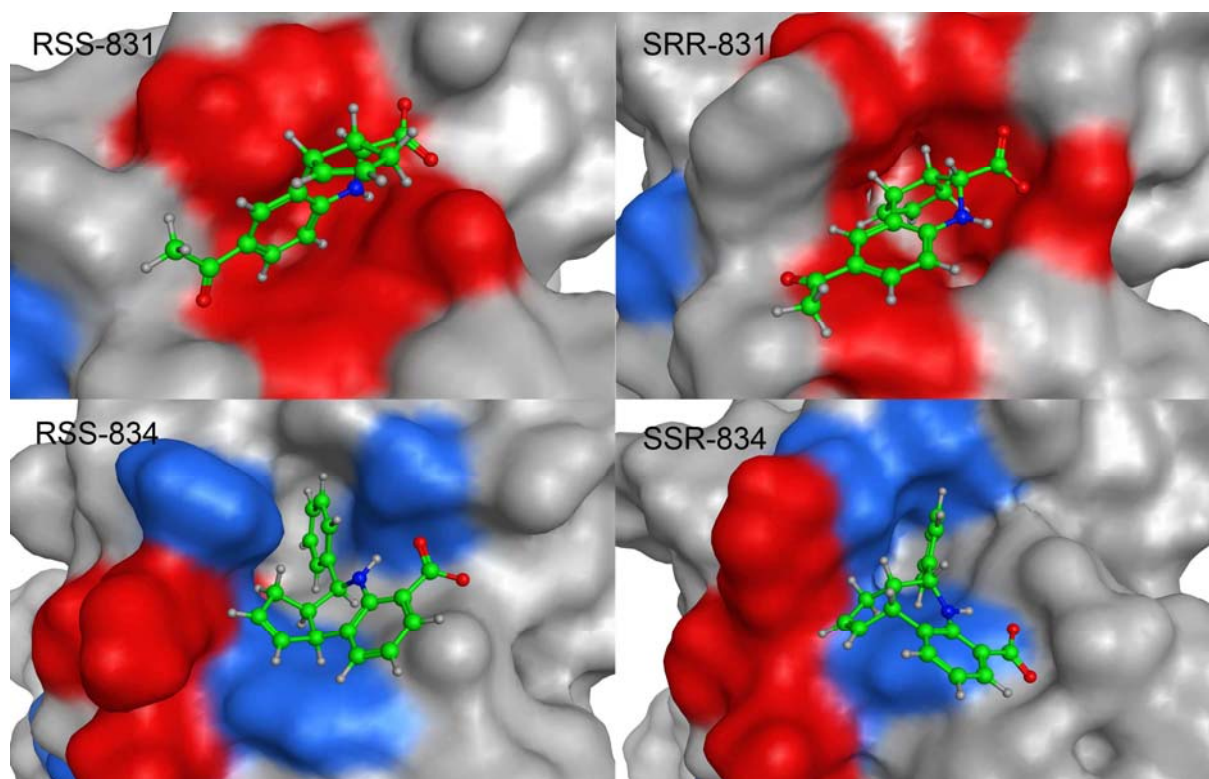
Figure 3.6.3-5 Docking poses of the highest populated cluster family for both enantiomers of 831 and 834. Colored surface depicts significant CSP (red) and strong signal intensity decrease (blue).

Visual inspection of the selected binding modes showed that the majority of experimentally detected perturbations were in proximity to the bound ligands as can be seen in figure 3.6.3-5. The transient pocket that was formed by rotation of the Phe17 side chain reappeared in all selected binding poses although the depth of the resulting pocked varied. The shallowest pocket is present for the RSS-831 enantiomer while the 834 enantiomers perturbates deeper into the protein. The selected docking poses could serve as basis for the selection of novel compounds for the confirmation of the binding mode and further improvements of potency.

(1)   Saupe, J.; Roske, Y.; Schillinger, C.; Kamdem, N.; Radetzki, S.; Diehl, A.; Oschkinat, H.; Krause, G.; Heinemann, U.; Rademann, J. *ChemMedChem* **2011**, *6*, 1411.
    (2)   Hajduk, P. J.; Bures, M.; Praestgaard, J.; Fesik, S. W. *J Med Chem* **2000**, *43*, 3443.
    (3)   Hopkins, A. L.; Groom, C. R.; Alex, A. *Drug Discov Today* **2004**, *9*, 430.

3.6.4 Concluding remarks post screening results

This chapter documents the results of two years of intensive lab and complex analytical work. A work, that was especially demanding in the early stages when it was very difficult to make any conclusions given the fuzzy and inconsistent nature of the data. Considerable time was invested to explore various experimental approaches and to brainstorm ways to validate the compounds consistently while improving their potency. Summarizing this type of work and the results in a form that may be easy to understand for people who were not involved in this process was particularly challenging.

In this chapter, we faced three key issues:

1. Experimental validation of our compounds regarding binding site and affinity.
2. Analysis and interpretation of fuzzy and inconsistent data.
3. Improvement of our compounds.

The very low affinity of our starting compounds, around or over 10 mM, was the main reason. Working, characterization and prediction of such weak interactions is prone to be problematic. These issues are especially severe in the context of fragment based drug discovery and protein-protein interfaces.

VEGF is a very popular drug target. Apparently, Fesik and colleagues at Abbott had a fragment based VEGF project in 2000 but never published anything besides the single compound mentioned earlier.[1] Gellman et all published 2008 a HTS compatible fluorescence assay for the discovery of small molecule inhibitors of VEGF but no discovery of novel compounds is reported.[2] Also, it appears that Genentech explored the possibility of small molecule based VEGF inhibitors, but, again, no published data is available to validate and to allow others to build on their results. In fact, no outline of how scientists approach and evaluate these type of issues exists to our knowledge at the present.

Although the results from our last round of elaboration have some remaining inconsistencies, they do show a significant improvement to the previous rounds. Certainly serendipity had some share on this improvement. On the other hand, the increasing rate of binding compounds per elaboration round and the match of prediction and experimental results strongly indicate that the methodology and approach that we used was a significant driving force behind this process improvement.

Our results indicate the presence of a transient pocket which is required for binding of compounds from the tetrahydroquinoline family. This pocket is to our knowledge not present in published crystal structures of VEGF. The importance of transient pockets and their implications for drug discovery, also in the also in the context of protein-protein interfaces are increasingly recognized by scientist in the field.[3-6]

To this end, the consolidated version of our final workflow is summarized in figure 3.6.4.
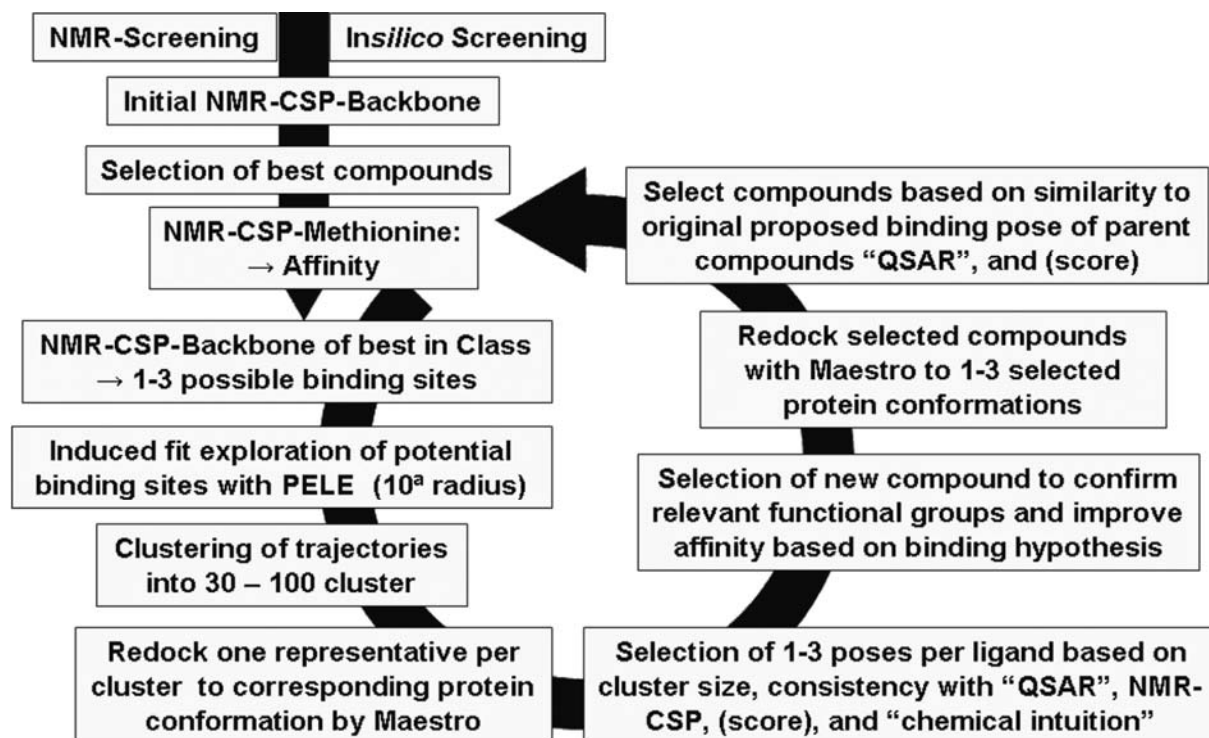
Fig. 3.6.4: Workflow for the elaboration of fragments for binding to VEGF by combination of NMR and induced fit docking with the PELE algorithm.

Obviously a future avenue to explore would be to validate this approach even further with the ultimate goal of establishing a general methodology for the challenging elaboration of fragments in the context of protein-protein interfaces.

(1)   Hajduk, P. J.; Bures, M.; Praestgaard, J.; Fesik, S. W. *J Med Chem* **2000**, *43*, 3443.
(2)   Peterson, K. J.; Sadowsky, J. D.; Scheef, E. A.; Pal, S.; Kourentzi, K. D.; Willson, R. C.; Bresnick, E. H.; Sheibani, N.; Gellman, S. H. *Anal Biochem* **2008**, *378*, 8.
(3)   Patschull, A. O.; Gooptu, B.; Ashford, P.; Daviter, T.; Nobeli, I. *PLoS One* **2012**, *7*, e36612.
(4)   Eyrisch, S.; Medina-Franco, J. L.; Helms, V. *J Mol Model* **2012**, *18*, 2031.
(5)   Metz, A.; Pfleger, C.; Kopitz, H.; Pfeiffer-Marek, S.; Baringhaus, K. H.; Gohlke, H. *J Chem Inf Model* **2012**, *52*, 120.
(6)   Wells, J. A.; McClendon, C. L. *Nature* **2007**, *450*, 1001.

3.7 Preliminary exploration of mRNA display to identify peptide VEGF ligands

So far, all our efforts were concentrated on fragment based strategies to identify small molecules that bind to the protein-protein interface of VEGF. Consistent with many publications, our results indicate that it is very challenging to target a flexible flat protein-protein interface with a small molecule. During my PhD thesis, I joined Prof. Hiroaki Suga's laboratory in Tokyo for a short stay. I used this opportunity to explore a novel mRNA display technology[1-3] that was developed in his laboratory, which is used to search for medium sized peptides that might bind to VEGF. This may advance the development of VEGF ligands and also facilitate our understanding of how to target such flat interfaces.

mRNA display offers certain advantages over *in vivo* display techniques such as phage, bacterial or yeast display. 1. Increased libraries as large as 10^15 members. These large libraries increase the probability of selecting rare sequences and improving diversity. 2. Absence of undesired selection pressure, e.g. poor expression or rapid degradation of some sequences. 3. Application of *in vitro* mutagenesis and recombination techniques. 4. The use and display of unnatural amino acid containing peptides.

For the selection, we used "Random Peptide Integrated Discovery"(RaPID) technology, which was developed in the laboratory of Prof. Suga. This technology is based on the reprogramming of the genetic code and the expression of cyclic peptides that contain non-natural amino acids that often exhibit increased *in vivo* stability and affinity.

During my stay in Prof. Suga's laboratory, I completed four selections with two different libraries against VEGF. The work for all selections can be divided into four similar parts: A: Preparation of flexizymes (ribozymes that catalyses aminoacylation of t-RNAs) and charged t-RNAs. B: Immobilization of VEGF. C: Selection against VEGF. D: Sequencing and identification of putative binder. Subsequently, I will briefly summarize the individual parts.

(1)   Goto, Y.; Suga, H. Methods Mol Biol 2012, 848, 465.
(2)   Morimoto, J.; Hayashi, Y.; Suga, H. Angew Chem Int Ed Engl 2012, 51, 3423.
(3)   Goto, Y.; Katoh, T.; Suga, H. Nat Protoc 2011, 6, 779.

3.7.1 Preparation of flexizymes and charged t-RNAs:

All non-natural amino acids that were used during the selection had to be charged priory to selected t-RNAS. For the first selection, we wanted to use Cl-acetyl-phenylalanine instead of methionine as the initiating amino acid. For this purpose, we amplified by PCR and purified by SDS-PAGE flexizymes. Further, we amplified by PCR and purified by SDS-PAGE t-RNA which exhibited the initiating codon for methionine. Incubation of this t-RNA with flexizyme and activated Cl-acetyl-phenylalanine after purification yielded t-RNA encoding for methionine but charged with Cl-acetyl-phenylalanine. This process is called genetic code reprogramming and allows for the expression of peptides with non-natural-amino acids (Figure 3.7.1-A).
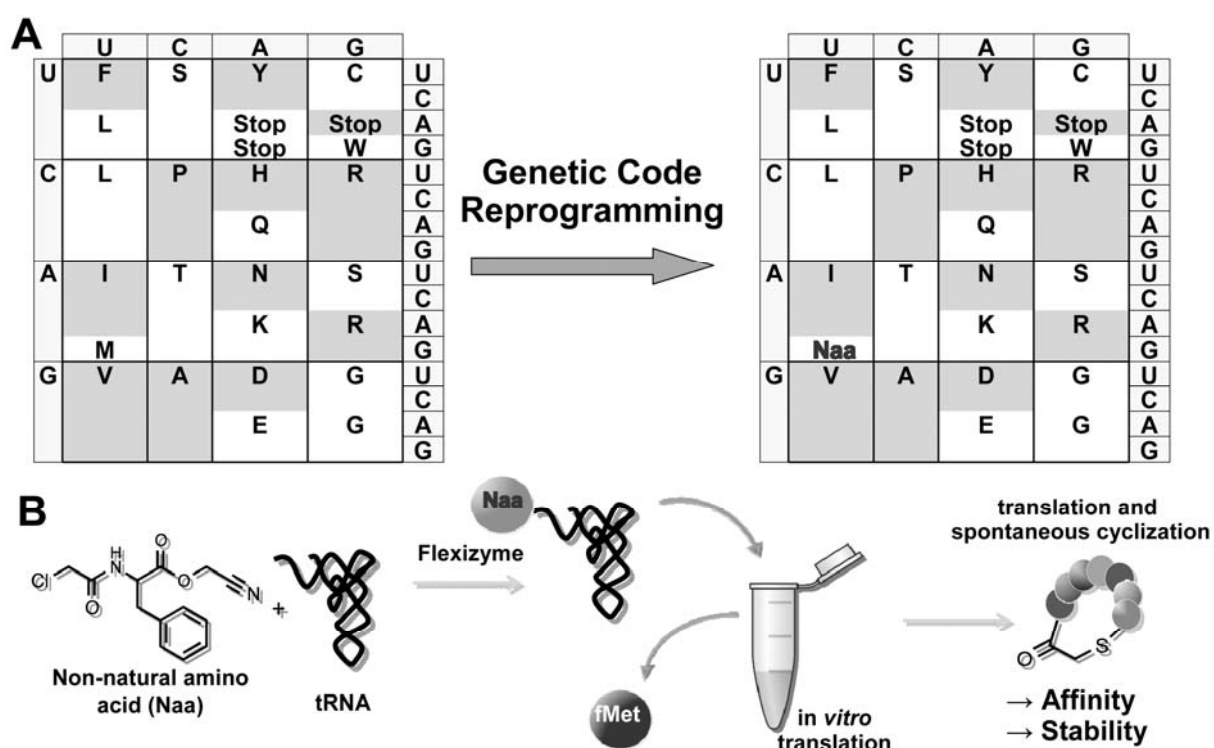
Figure 3.7.1: A: Reprogramming of the genetic code to incorporate non-natural amino acids. B: Reprogramming of Met to Cl-Acetylated amino acids leads to spontaneous formation of cyclic thioether from translated cystein containing peptides.

The purity and functionality of this charged t-RNA was tested with an *in vitro* translation system in the absence of methionine and a peptide encoding mRNA. As a positive control, the same *in vitro* translation system with methionine charged t-RNA, instead of our mischarged t-RNA, was used. This resulted in the synthesis of the corresponding peptide as detected by MALDI-MS. In the case of our mischarged t-RNA, the peptide detected by MALDI-MS differs from the peptide expressed in the presence of methionine by exactly the mass corresponding to the replacement of methionine by Cl-acetyl-phenylalanine and the subsequent elimination of HCl which results from the nucleophilic attack of a N-terminal cysteine with the introduced the Cl-acetyl-group. The resulting peptide is

therefore cyclized by a thioether bond, which gives the peptide higher rigidity and protease stability (Figure 3.7.1-B). Since this is exactly the class of peptides that we want to explore during the selection, this result confirms that the quality of our charged t-RNA was sufficient.

For the first selection, we charged and tested methionine-encoding t-RNA with L-Cl-acetyl-phenylalanine and D-Cl-acetyl-phenylalanine. For the second selection, we charged and tested methionine-encoding t-RNA with L- and D-Cl-acetyl-tryptophan as well as phenylalanine encoding t-RNA with L-N-Methyl-phenylalanine, leucine encoding t-RNA with N-Methyl-serine, isoleucine encoding t-RNA with N-Methyl-glycine and alanine encoding t-RNA with N-Methyl-Alanine. This heavy genetic reprogramming should increase rigidity, *in vivo* stability and membrane permeability of the modified peptides and represents the current limit of non-natural-amino acid containing peptides in Prof. Suga's lab.

This completed the preparation of flexizymes and charged t-RNAs and we proceeded to the next step.

3.7.2 Immobilization of VEGF

A successful selection strategy requires defined conditions for immobilization of the protein. For immobilization, the medium of choice in Prof. Suga's lab is nanobeads, small beads with a magnetic core and a variety of surface functionalizations to bind to the protein. The two most utilized nanobeads bear either a streptavidin modification to bind biotinylated proteins or are coated with a Ni^{2+} surface to bind His-tagged proteins. The streptavidin-biotin system is superior in terms of binding strength. However, biotinylation of VEGF seemed to not be a good choice because of the high number of carboxy or amine groups, also present in the receptor binding interface, which would lead to heavy modification of the protein. Since VEGF is expressed with a His-Tag, we choose nanobeads coated with a Ni^{2+} surface. The drawback of this choice is that binding of the His_6-tag to Ni^{2+} is weak enough to be reversible at room temperature. Therefore, selections have to be performed at 4°C to reduce leakage of immobilized protein into the supernatant. In spite of this drawback, this immobilization is the most used strategy for selections in Prof Suga's lab.

The next step was to determine under which conditions the beads could be saturated with protein. Without saturation, free binding sites on the beads lead to the selection of bead binding peptides.

After protein refolding, we used a VEGF solution from Barcelona that contained besides the native dimer, mono- and trimer. This was necessary because we could not obtain the pure his-tagged dimer since, after shipping to Tokyo, the protein precipitated during dialysis prior to anion exchange chromatography. This problem was not present with newly expressed protein, however, due to issues during the refolding process, we preferred to work with the VEGF solution from Barcelona. Despite this problem, we achieve saturation of the nanobeads with the native VEGF dimer by

exploiting the increased affinity of the two his-tags of the dimer over the monomer through repeated incubation and washing cycles of the beads with fresh protein solutions.

3.7.3 Selection against VEGF

The first selection was performed with an mRNA library encoding for peptides of varying length between 5 and 15 random amino acids, one initial methionine (which is reprogrammed to a Cl-acetyl bearing amino acid) and a C-terminal cysteine (to form the cyclic thioether) followed by a glycine-serine repeat region as flexible linker. The random region was encoded as NNK codons, restricting the last nucleotide to U or G. This does not exclude amino acids from the codon table but reduces the degeneracy of the genetic code and excludes two stop codons. The first selection was further executed with two identical sub-libraries and the only difference that one selection was performed with Cl-acetyl-L-Phenylalanine (from now on called NNK-L) and the other with Cl-acetyl-D-Phenylalanine (from now on called NNK-D) as the initial amino acid (Figure 3.7.3-1). Based on the knowledge of the lab, modification of this one stereo center is sufficient to enforce different backbone conformations and therefore explore a different chemical space.
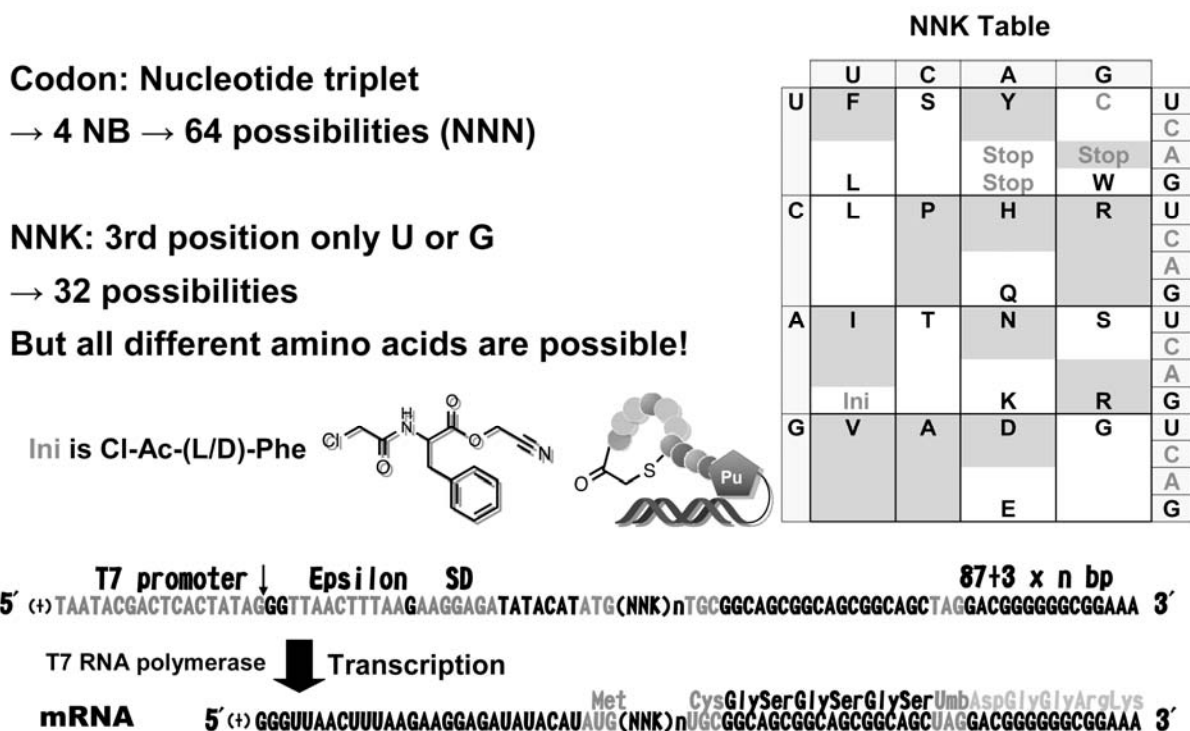
Figure 3.7.3-1: Overview of the two NNK selections: Removal of U and G nucleotides in the third codon position decreases the degeneracy of the genetic code. The mRNA library encodes for a random region of varying length, which is flanked by the methionine initiator codon and a cysteine codon followed by Gly Ser repeats as flexible linker to puromycin which links after translation to the mRNA.

The second selection was also executed in two sub-libraries with L or D Cl-acetyl-Tyrosine as the initiating amino acid and with an mRNA library consisting of NNU codons. This means that the last nucleotide is fixed to U, which removes the degeneracy of the genetic code and also restricts the

encoded amino acids to 16 excluding Trp, Gln, Glu and Lys. Further the amino acids Phe, Leu, Ile and Ala were reprogrammed to N-methylated amino acids as described above (Figure 3.7.3-2).

Figure 3.7.3-1: Overview of the two NNU selections: Removal of U, G and G nucleotides in the third codon position limits the genetic code to 16 amino acids. The mRNA library encodes for a random region of varying length, which is flanked by the methionine initiator codon and a cysteine codon followed by Gly Ser repeats as flexible linker to puromycin which links after translation to the mRNA.

For the NNU-D selection a random region between 8 and 12 amino acids was explore while for the NNU-L selection a random region of 13 to 15 amino acids was used. For the second selection, I choose to explore different length regions separately since smaller peptides generally have better membrane permeability but are rarely selected against longer peptides.

Beside these differences, selection for the NNK and NNU libraries was performed following the same workflow (Figure 3.7.3-3):
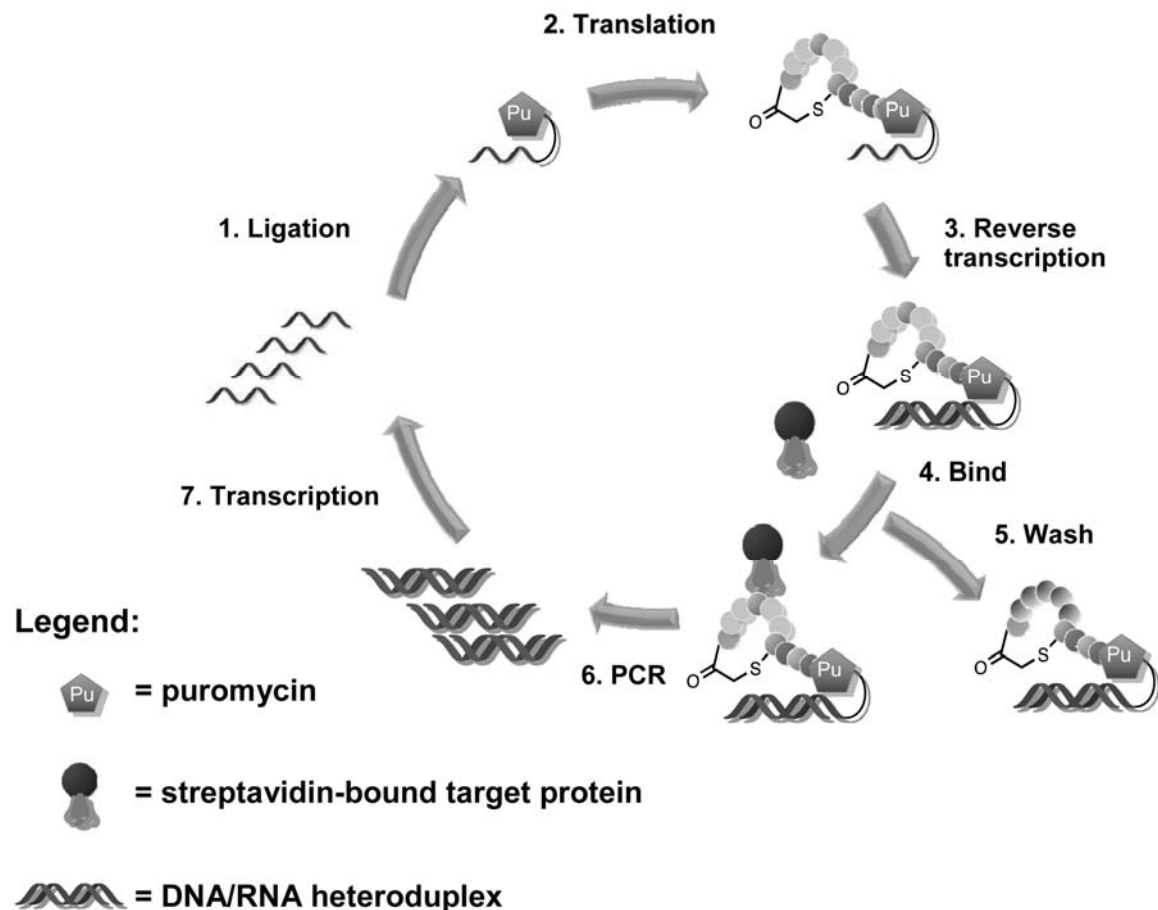
Figure 3.7.3-3: Simplified workflow for the selection of non-natural amino acid containing cyclic peptides with mRNA display.

1: Ligation: The mRNA sequences are enzymatically linked to puromycin.

2: Translation: The mRNA sequences are transcribed to their corresponding peptides with a reconstituted *in vitro* translation system. The system lacked the reprogrammed amino acids (as methionine) but contained the non-natural amino acid bearing t-RNAs. Therefore the mRNA sequences are translated to peptides containing non-natural amino acids as the Cl-acetylated amino acids in the initiator position. The translation system also lacked release factors which led to stalling of translation once the linker region was translated. This allowed time for incorporation of puromicin which resembles the 3′ end of an aminoacylated t-RNA and therefore created a covalent bond between the peptide and its encoding mRNA. After giving this step time to finish, the mRNA-peptide conjugate was released from the ribosome by complexation of magnesium. In solution, the engineered N-terminal cysteine reacted spontaneously with the C-terminal Cl-Acetyl group to form a stable thioether bond leading to the formation of the macrocyclic peptide, the class of structure that we wanted to explore as putative ligands.

3: Reverse transcription: The mRNA sequence was reverse transcribed via polymerase chain reaction to its corresponding cDNA. cDNA and its originating mRNA form a heteroduplex. This was necessary

to amplify the genetic information later and to avoid the selection of mRNA aptamers which may bind to the target protein.

4 and 5: Binding and washing: The solution containing the conjugated peptide-DNA/RNA heteroduplex was diluted with a blocking solution containing bovine serum albuminum and incubated with empty nanobeads. After incubation, the supernatant was recovered and the process was repeated several times. This step is called preclear and was necessary to remove bead binding peptides. Generally, after three preclears, the peptide containing solution was incubated with the VEGF saturated nanobeads . After 30 minutes the supernatant was removed and the beads washed several times with buffer.

6: PCR: The washed beads should now be covered with VEGF and bound to it the peptides conjugated to the DNA/RNA heteroduplex. The mixture was heated for a short time to allow for dissociation. Then the DNA was amplified by PCR. In principle this amplified DNA should only encode for peptides with affinity to the target protein.

7: Transcription: In the last step the amplified DNA was transcribed to an mRNA pool. Compared to the previous round, this pool should be enriched in sequences encoding for peptides that bind to the target protein.

The selection cycle is repeated several times. During this time the progress is monitored by quantitative real-time polymerase chain reaction. Of special interest was the fraction of DNA that was recovered from the empty beads (negative selection; bead binder) as well as the beads saturated with the target protein (positive selection; putative ligands) as compared to the amount of DNA that was present at the start of the binding and washing step. Successful selection is indicated by increasing amounts of recovered DNA from the saturated beads while the amount of DNA recovered from the empty beads stays the same or decreases. For the NNK-D selection, 10 rounds were performed, while one fewer round was performed for the NNK-L selection, because one round (4) had to be repeated. Both selections showed a drop in recovered DNA in round 7. After troubleshooting we decided to continue since the next round showed normal recovery rates again. Excluding this round, the selection profile seemed to be very favorable with an overall increase in recovery rates for the positive selection, which finally significantly exceed the negative selection recovery (Figure 3.7.3-4).
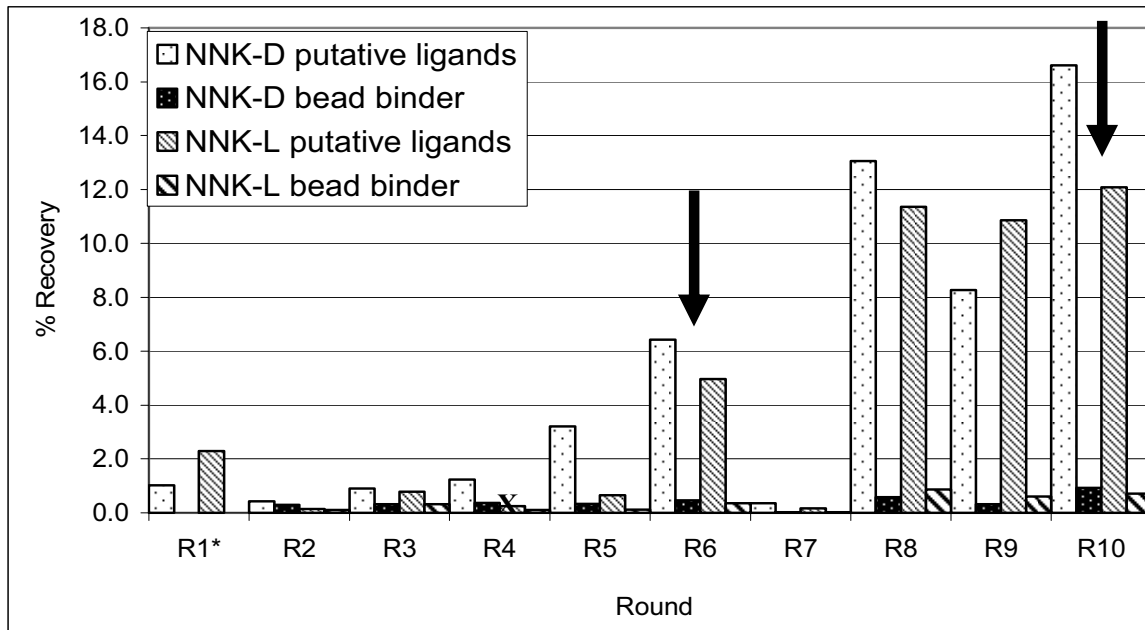
Figure 3.7.3-4: Profile for the two NNK selections monitored by recovered DNA from VEGF coated (putative ligands) and empty beads (bead binder). The arrows mark DNA pools that were used for sequencing. *In the first round no DNA was recovered from empty beads. **X** marks a round for the NNK-L library that was repeated.

For the NNU-L and NNU-D selection, 8 rounds were performed. In round 2, the recovery of the negative selection was higher than the recovery of the positive selection. To avoid the selection of bead binder the number of preclears was increased in the following rounds. As for the NNK selections, the NNU profiles seemed to be favorable with high positive recovery rates at the end of the selection and lower negative recovery rates (Figure 3.7.3-5). The drop of recovery rates in round 7 is probably caused by the massive preclearing. This is supported by the high increase of positive recovery rates for the next round.
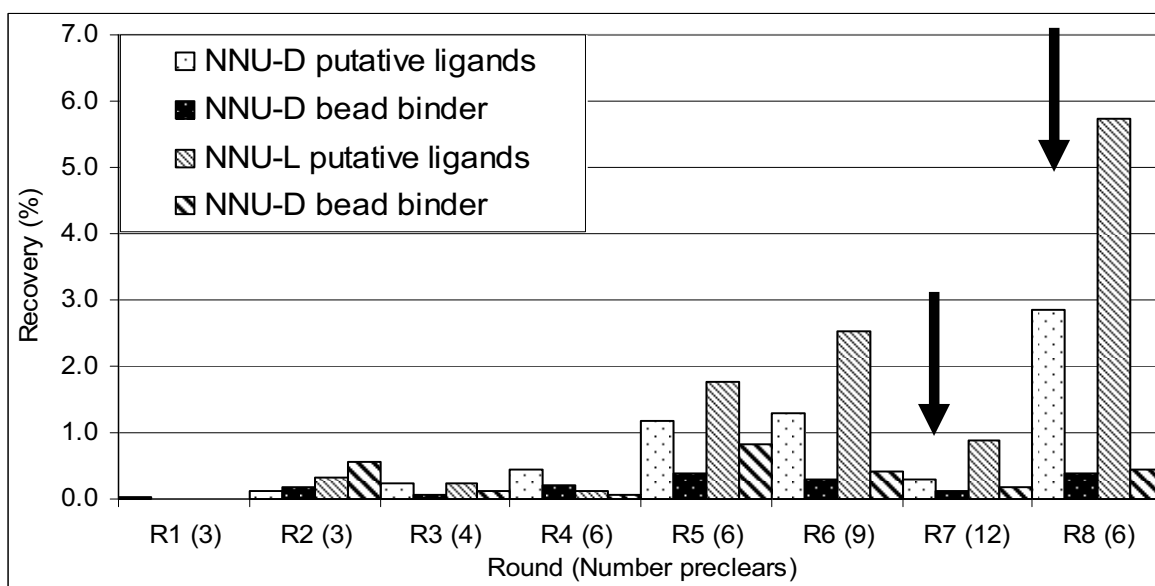
Figure 3.7.3-5: Profile for the two NNU selections monitored by recovered DNA from VEGF coated (putative ligands) and empty beads (bead binder). The arrows mark DNA pools that were used for sequencing. In the first round no DNA was recovered from empty beads

3.7.4 Sequencing to identify putative ligands.

After all selections reached satisfactory stages, we decided to identify the putative peptides.

For this the DNA from the enriched pool of the two last selection rounds (for NNK round 6 and 10) was ligated into a vector encoding for ampicillin resistance. The vector was introduced by heat-shock into *e.coli* cells grown on agar plates containing ampicillin. Separated colonies were picked from the plates and the region of the vector that contained the peptide encoding insert was amplified by PCR. The purified PCR product was sent for sequencing. The data was then analyzed with CLC Sequence Viewer 6.[1] The alignment of sequences was performed with standard parameters, although this may not be the method of choice for small cyclic peptides.

Both NNK and NNU selections did not show sequence convergence but instead showed the enrichment of specific motives. Although not the usual case, similar outcomes could be observed in other successful selection projects in Prof. Suga's laboratory.

The display of the NNK-D library led to the selection of peptides with a random region size of 14 amino acids which are structured as one big macro cycle (Thioether bond between initial methionine (now Phe) and the final cysteine codon (Figure 3.7.4-1).
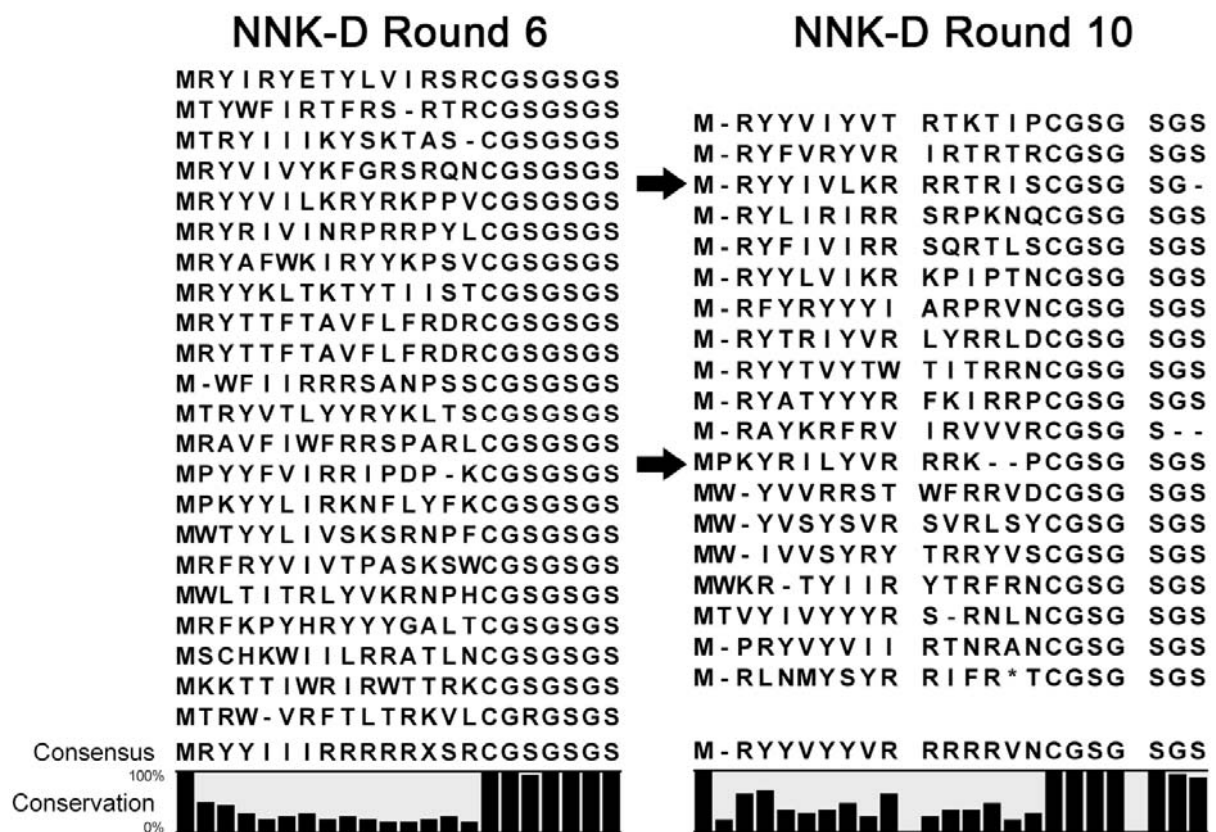
Figure 3.7.4-1: Sequencing results for round 6 and 10 of the NNK-D selection with Cl-Acletyl-D-Phe initiator. – marks a gap based on sequence alignment. Arrows mark sequences selected for the clone assay.

The C-terminal portion is dominated by a motive defined by Arg followed by two positions occupied by either Tyr or Phe. This motive is followed by a hydrophobic region consistent of Leu, Ile and Val.

The N-terminal region is less conserved, however, Arg, Thr and Ile and Val are frequently occurring amino acids.

The display of the NNK-L library led again to the selection of peptides with a random region size of 14 amino acids (Figure 3.7.4-2).

NNK-L Round 6

```
MRCLKLILLSTASSPCGSGSGS
MRCHIRFRYPSIQYNCGSGSGS
MRCYHLVIRLRPSERCGSGSGS
MNCRTYYQIRIKRYTCGSGSGS
MSCGYVVYTWRIPTVCGSGSGS
MSCTYIRLTYTATYPCGSGSGS
MLCANRIFRSRTIYVCGSGSGS
MWCKTYSNQRHR--PCGSGSGS
MHNCFRISVRIKYVDCGSGSGS
MKNYCVRRRVRYRIGCGSGSGS
MKRVICKFR--RTLTCGSGSGS
MRARYCRRKYVVTWLCGSGSGS
MPRRYCRLRVIRISSCGSGSGS
MRFSNCLTFAYRIPKCGSGSGS
M--TRSCIYVRFRLNCGSGSGS
MKRNKRQCIIMRRYYCGSGSGS
MKKIQVYRYRLMRYCCGSGSGS
MMIIRRKRYLKLCSYCGSGSGS
MFRYIIVYGIEPRDSCGSGSGS
MIVKRSRMIYVILPCCGSGSGS
MTLILYRNTWGRRTCCGSGSGS
MRSYSRRGRTWTYHCCGSGSGS
MYLSFRRSNWVTYCKCGSGSGS
```
Consensus MRCYYRRRRRRRYCCGSGSGS
100%
Conservation
0%

NNK-L Round 10

```
MRCRLVRYRRPII-VCGSGSGS
MRCYRAYRIRLKRISCGSGSGS
MRCYYAI-FRTKEIDCGSGSGS
MRCKLRIIFRLNRQNCGSGSGS
MRCIRRRVWIIRTLPCGSGSGS
MRCRRRIKISYYYLACGSGSGS
MRCYYVRIKRTTRNGCGSGSGS
MRCVRTRLVRLRYAICGSGSGS
MRCIVRIRIKRRQSACGRGSGS
MRCIVRIRIKRRQSACGRGSGS
MKCRRNKRYVLILTRCGSGSGS
MNCRWYYVPNTQPRFCGSGSGS
MNCFRWYQRPSARLICGSGSGS
MRRWCLSRRVIKLRNCGSGSGS
MRNRICVLRIRPYWRCGSGSGS
MRRYRCYKVYTFTVKCGSGSGS
MRWKRCVYIRIVSRNCGSGSGS
MLGYKCYVIVKSNYSCGSGSGS
MLGYKCYVIVKSNYSCGSGSGS
MRKIRYRYRVYYISCCGSGSGS
```
MRCYRXYRIRXRRXXCGSGSGS

Figure 3.7.4-2: Sequencing results for round 6 and 10 of the NNK-L selection with Cl-Acletyl-L-Phe initiator. – marks a gap based on sequence alignment. Arrows mark sequences selected for the clone assay.

Contrary to the NNK-D selection, an additional cysteine was conserved either at position 3 or 6 from the C-terminus. Since the formation of the thioether bond is kinetically controlled, this led to the formation of a three or six member macrocycle followed by a linear peptide chain, as compared to the structures selected in the NNK-D selection. The sequences inside the macro cycle are dominated by Arg or Lys, therefore a positive charge seemed to be required at this position. The linear peptide is less conserved, but again dominated by Arg, Ile and Tyr.

The display of the NNU-D library led to the selection of peptides with a random region size of 12 amino acids (3.7.4-3). In round 7, as well as in round 8, two scaffolds are present: A macro cycle of the maximal length which shows an N-methyl Ser (Leu codon) at the C-terminal followed by a Val and Tyr rich motive which also exhibits Arg towards the less conserved N-terminal region. A small macro

cycle formed by the appearance of addition cysteines in position 3 to 6 and followed by a Val, Arg and Thr rich motive.

NNU-D Round 7   NNU-D Round 8

```
                                    MR - CCRV - VVVRYPPCGSGSGS
                                    MR - CT - V - YVVRVPPCGSGSGS
                                    MS - CVYVVRVTTYR - CGSGSGS
MRCVYVRVSV - PRYCGSGSGS            MPRCSYRVRVTRV - - CGSGSGS
MRCVRVRVYV - RSLCGSGSGS            MRSCTYVRRVT - - VSCGSGSGS
M - - YCYYRYRGPDPCGSGSGS           MRSCTYVRRVT - - VSCGSGSGS
MRSYCVRVVY - RVNCGSGSGS            MRRFCLFRRVVR - - PCGSGSGS
MRRP - VYSCVVTVPCGSGSGS            MTRRTCVT - VTRY - - CGSGSGS
MRRYYVYYYRGTI - CGSGSGS            MR - YRRYRVVVRYS - CGSGSGS
MPYVYVVRVR - TRYCGSGSGS            MLVYVVYRRRVRY - - CGSGSGS
MPD - - - - RTYPRSVCGSGSGS         MLVRYVY - - YVRVYLCGSGSGS
                                    MRVYYVVR - - TYYTRCGSGSGS
                                    MPLYYVV - - VVSYRACGSGSGS
                                    MTP - - VYRYVVRYVFCGSGSGS
Consensus  MRXXYVXXYX - XXXCGSGSGS  MR - CTVVRRVVRY - - CGSGSGS
```

Consensus MRXXYVXXYX - XXXCGSGSGS

Conservation 100% / 0%

Figure 3.7.4-3: Sequencing results for round 7 and 8 of the NNU-D selection with Cl-Acletyl-D-Trp initiator. – marks a gap based on sequence alignment. Grey background mark codons that were reprogrammed: L-> N-methyl Ser; F-> N-methyl-Phe; A-> N-methyl-Ala.

The display of the NNU-L library led to the selection of peptides with a random region size of 15 amino acids (Figure 3.7.4-4). In nearly all sequences, an additional cysteine was selected. The position of this additional cysteine differs between the c-terminal part leading to macro cycles between 3 to 8 amino acids or directly before the N-terminus leading to a macro cycle of 16 amino acids. Both scaffolds show often a tipple Val motive, while other frequently occurring amino acids are Arg and Tyr.

NNU-L Round 7

```
MRCYRYYYVTVSISRD--CGSGSGS
MRCYRTVYVR--YNRVIPCGSGSGS
MRCISRVYVR--YTRSAVCGSGSGS
MRCPYRRYVYV-YNATT-CGSGSGS
MSC--TYYVRYSRNSVSTCGSGSGS
MSCVNVYVVYV--NSIRPCGSGSGS
MLCGSIYRYVVVYYSVT--CGSGSGS
MGCTVVHVRYGNYN--TRCGSGSGS
MFTCRVT-VRVRYSRDT-CGSGSGS
MGVVRTRRGYVYVN--VCCGSGSGS
MVVYR--YSRNTYSYCRVCGSGSGS
MRVVYVRCRRVTYTYI--CGSGSGS
MVV--SYHLVRYVTVTRCCGSGSGS
MVRHPRYYNV--YVRVSCCGSGSGS
MDPNRCYRVVV--VRRASCGSGSGS
MRCXRVYYVRV-YNRXX-CGSGSGS
```

NNU-L Round 8

```
MPCRN--RYYVSVVRGTNCGSGSGS
MRCYY--VVRVRNIPYSDCGSGSGS
MPCSSRARRRR--TYYGPCGSGSGS
MTCRRSCT--VVYVYSNTCGSGSGS
MPRCVRTRRYRYVVYS--CGSGSGS
MP--RVCRRGVYGYYVACCGSGSGS
M--RLCPRRVYRVVTVTYCGSGSGS
M--RLCPRRYVVVFVNGYCGSGSGS
MRYRRCTVSVSRPPRY--CGSGSGS
MVYSRCYVRRDYPSY--RCGSGSGS
M--RNRCVYVVYRRSVRDCGSGSGS
M--RPYRSGCTYVVVVNYCGSGSGS
MNYGT--NYCYVVVRVVLCGSGSGS
MTVRRRGSRRYVV--YTCCGSGSGS
MVVRV--SRRYVVRSVSCCGSGSGS
MVVVR--SRYYRVRYNPCCGSGSGS
MVVVT--VSRSRVVTYPCCGSGSGS
MX-RR--RRRVXVVYVTCCGSGSGS
```
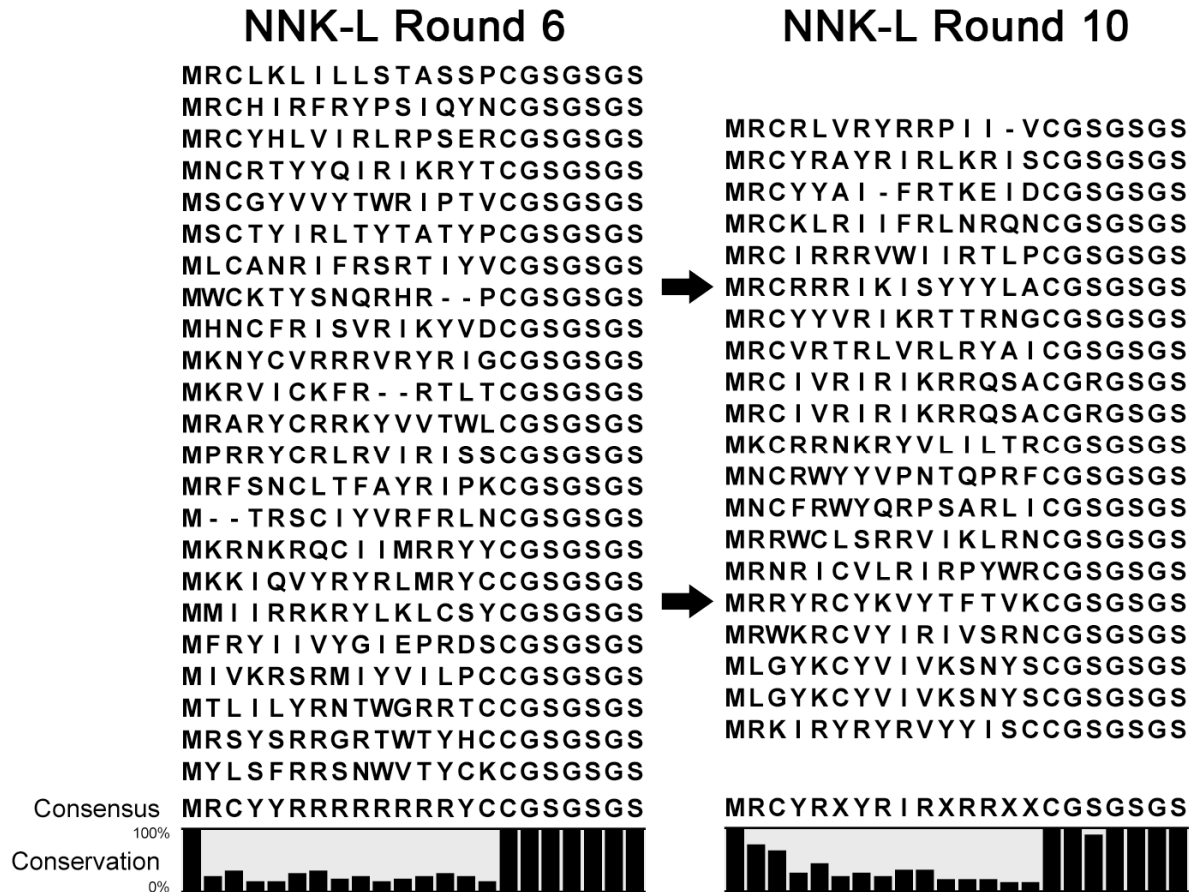
Consensus
Conservation

Figure 3.7.4-4: Sequencing results for round 7 and 8 of the NNU-L selection with Cl-Acletyl-L-Trp initiator. – marks a gap based on sequence alignment. Grey background mark codons that were reprogrammed: L-> N-methyl Ser; F-> N-methyl-Phe; A-> N-methyl-Ala; I-> N-methyl-Gly.

In both NNU-D and –L selections, several sequences contain the Leu, Ile, Ala, and Phe codon, which were mischarged with the previous listed N-methylated amino acids. The presence of these codons in the sequencing data confirms that the quality of the *in vitro* translation system and non-natural charged t-RNAs was sufficient for the genetic reprogramming to obtain the desired peptides. The candidate with the highest incorporation of non-natural amino acids belongs to round 8 of NNU-D, which exhibit three N-Methylated amino acids out of the 12 random positions. This is close to the current record in Prof. Suga's lab.

A possibility of gaining further confidence with selected peptide sequences as putative ligands is the clone assay. In this assay, the DNA from a specific colony with known sequence is PCR amplified and then reverse transcribed to obtain its mRNA which can be used for another round of selection. This selection round is nearly identical to the previous performed selections rounds with the only difference being that the DNA pool now consists of one specific known sequence and not a library of sequences. This selection round can be performed for several sequences in parallel followed by a comparison of the DNA recovery rates for the selection against empty beads and target-protein saturated beads for the different sequences. The best candidate is characterized by the highest positive and the lowest negative recovery rate. This would indicate high affinity towards the target and low affinity against the beads. For time restrictions, I could only perform clone assays for 2 selected sequences from each NNK-L and NNK-D selection. As described above, we had to use a mix of VEGF monomer and trimer together with the native dimer for these selections. Despite our ability so enrich the nanobeads with the VEGF dimer, we wanted to ensure that the selected peptide sequences recognize the native protein form. For this reason we performed a selection of each of the

sequences against the empty beads, the target protein saturated beads and beads saturated with VEGF trimer. Additionally, the selection against the VEGF dimer saturated beads was performed in duplicate: One time with and and time without preincubation of the peptide containing solution with histag free VEGF whose structure was confirmed by NMR spectroscopy. The results for all four sequences were similar (Figure 3.7.4-5):

D11-3:**MRRYRCYKVYTFTVKCGSGSGS** L11-10: **MPKYRILYVRRRK - -PCGSGSGS**
D11-7:**MRCRRRIKISYYYLACGSGSGS** L11-G: **M-RYYIVLKRRRTRISCGSGSG-**
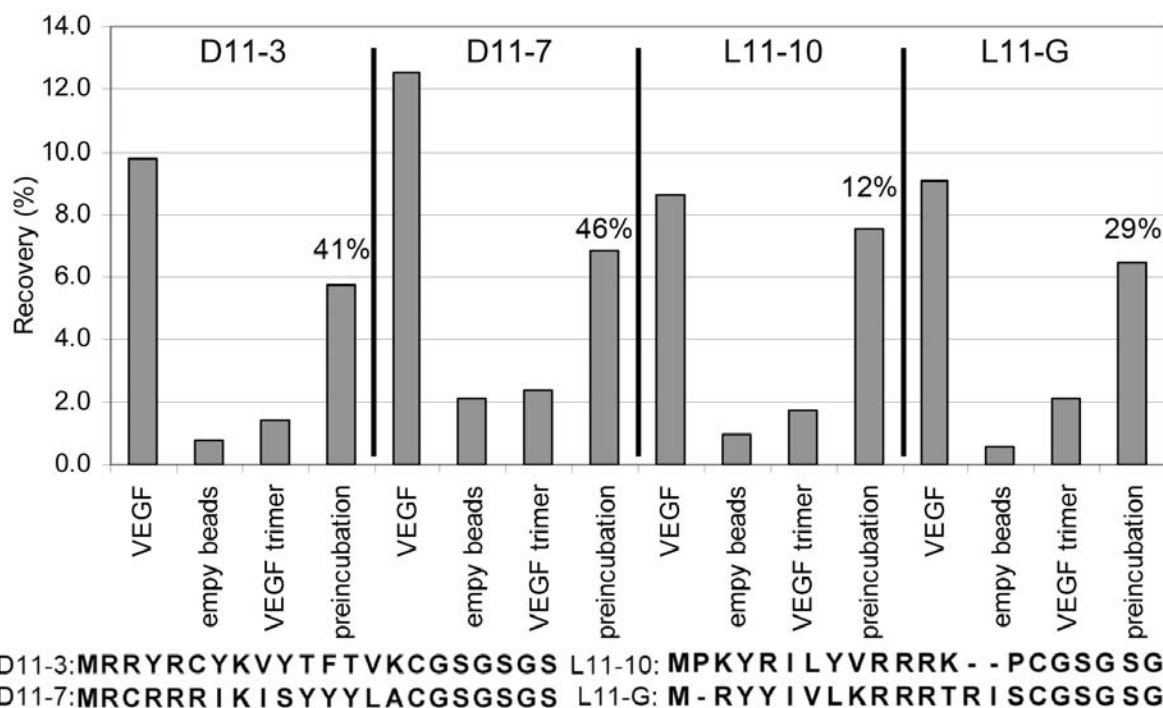
Figure 3.7.4-5: Clone assay for two selected sequences of each NNK selection. Each bar corresponds to the % of recovered DNA from (left to right), VEGF saturated beads, empty beads, VEGF trimer saturated beads and VEGF saturated beads with preincubation of the mRNA pool with soluble native VEGF. The percentage value describes the decrease in recovered DNA when the protein immobilized on beads was competing with well-folded VEGF in solution (preincubation).

Generally high recovery rates were observed for the target protein saturated beads and low recovery rates for the empty beads. The beads saturated with VEGF trimer also showed low recovery rates, which were similar to the empty beads. In all cases, preincubation with soluble histag free VEGF (structure confirmed by ^{1}H-^{15}N-HSQC) led to reduction of DNA recovery from the target protein saturated beads. This indicates that the soluble native VEGF competes with the immobilized VEGF and that therefore the peptides recognize the native protein form. Based on the numbers of the recovery rates, peptide sequence D7 might be the best candidate since the positive recovery rate is the highest and reduction by soluble histag free VEGF was the most accentuated. However, the recovery rate for the negative beads and the VEGF trimer is also higher compared to the other sequence making it difficult to define a best candidate.

(1)   CLC sequence viewer, 2013, http://www.clcbio.com/products/clc-sequence-viewer/

3.7.5 Synthesis of peptides.

Back in Barcelona, our aim was so synthesize promising peptide candidates and validate them for binding to VEGF (Table 3.7.5). The synthesis was performed in two groups. First, we decided to focus on the four peptides that were explored in the clone assay from the NNK libraries because the results seemed very promising to Prof. Suga. We also chose one peptide from the NNU library that showed the highest incorporation of non-natural amino acids to asses our ability to synthesize these heavily modified peptides. In the second group we decided to focus on sequences that were different from the candidates of the first group and did not contain N-methylated amino acids since their synthesis proved to be work intensive. Further all candidates of this group were labelled with carboxyfluorescein attached to the side chain of a C-terminal lysine. If the peptide was structured as a large macrocycle, an additional beta-alanine was incorporated between the sequenced amino acids and the fluorescein labelled lysine.

ID	Sequence	Yield (%)	Purity (%)	Expected mass (charge state)	Found mass (charge state)
D11-3	*f*PKYRILYVRRRKP**C**GS	15.43	>95	726.7451 (Z=3)	*726.7451(z=3)*
D11-7	*f*RYYIVLKRRRTRIS**C**GS	0.52	>95	771.7678 (z=3)	*771.7676 (z=3)*
L11-10	FRRYR**C**YKVYTFTVKC*GS	0.32	>95	796.4020 (z=3)	*796.4005 (z=3)*
L11-G	FR**C**RRRIKISYYYLAC*GS	2.3	>95	765.3989 (z=3)	*765.3976(z=3)*
TX	*w*RR(F)**C**(S)(F)RRVVRPC*GS	3.2	>95	722.0480 (z=3)	722.0485 (z=3)
D6-F	*f*RYAFWKIRYYKPSV**C**a*K-fluorescein	<0.5	>95	908.4355 (z=3)	908.4360(z=3)
D6-L	*f*RYTTFTAVFLFRDR**C**Ga*K-fluorescein	<0.5	>95	899.4185(z=3)	899.4181(z=3)
D6-2	*f*WTYYLIVSKSRNPH**C**a*K-fluorescein	<0.5	87	870.7358(z=3)	870.7344(z=3)
D6-H	*f*KKTTIWRIRWTTRK**C**a*K-fluorescein	<0.5	80	907.4694(z=3)	907.4668(z=3)
L6-3	FRSYSRRGRTWTYH**C**SK-fluorescein	<0.5	>95	647.7888(z=4)	647.7870(z=4)

Table 3.7.5: Overview of synthesised peptides from the first and second group. Minor letters indicate non-natural amino acids: f= D-Phe; w=D-Trp; a= beta-alanine; bold C indicates the location of the thioetherbond to the N-terminal N-Acetyl group. C* indicates the use of ACM side chain protection groups. Expected mass and found mass correspond to monoisotopic distribution detected my ESI-MS.

All peptides were synthesized automatically by SPPS using the standard Fmoc strategy and microwave support. An exception was the N-methyl bearing region of peptide TX which was synthesized manually by Dr. Meritxell Teixido and Cristina García.

According to a protocol from Japan the N-terminal amine of all peptides was capped on resin with Cl-acetly-NHS. Fluorescein labelling of the second group was achieved by the use of MMT-protected Lysine at the C-terminus. Sidechain deprotection was achieved after capping of the N-terminus in a mild acidic conditions. Subsequently the free ε-amino group was coupled on resin with NHS-fluorescein to achieve the labeling. Cleavage from the resin was performed under standard conditions over 3 hours. The solution containing the cleaved peptides was directly diluted and the pH adjusted to basic conditions to allow the formation of the thioether between the Cl-Acetyl mojecty

and cysteine, which was followed by MALDITOF-MS. After completion of the reaction the solution was freeze dried and finally the peptides purified by HPLC. This work was performed for several peptides by Esther Zurita.

We obtained very different yields for our peptides. Peptide D11-3 was obtained with a relatively good yield of 15%. The yield of the other peptides was much lower. This was especially true for the peptides of the second group. The crude of the peptides showed multiple side products which very similar retention times. Peptide D6-2 required 10 purification steps to achieve a purity of 87%. The synthetic challenge of these peptides was confirmed in attempts made by Dr. Christopher Hipolito, my collaborator in Tokyo. The sources of these problems were probably: 1. Peptide sequences with high incorporation of Arg and beta-branched amino acids. 2. Multiple conformations of the macrocycle. 3. Use of 5-/-6-Carboxyfluorescein mixed isomers. Although these problems only allowed us to obtain low amounts of relatively pure peptides, it was sufficient for explorative binding assays.

3.7.6 Preliminary validation of peptides.

Assaying of peptides was performed in the first instance by fluorescence polarisation. For the first group we were relying on an assay published by Gellman et al.[1]. The required labelled tracer (Kd= 38 nM), an analogue of peptide v107, was synthesized by Ester Zurita. The five synthesized peptides were incubated over a range from 30 uM to 3 nM with 150 nM of VEGF and 10 nM of the tracer. We used peptide v107 (Kd= 1 uM) as a control. The results are summarised in figure 3.7.6-1.
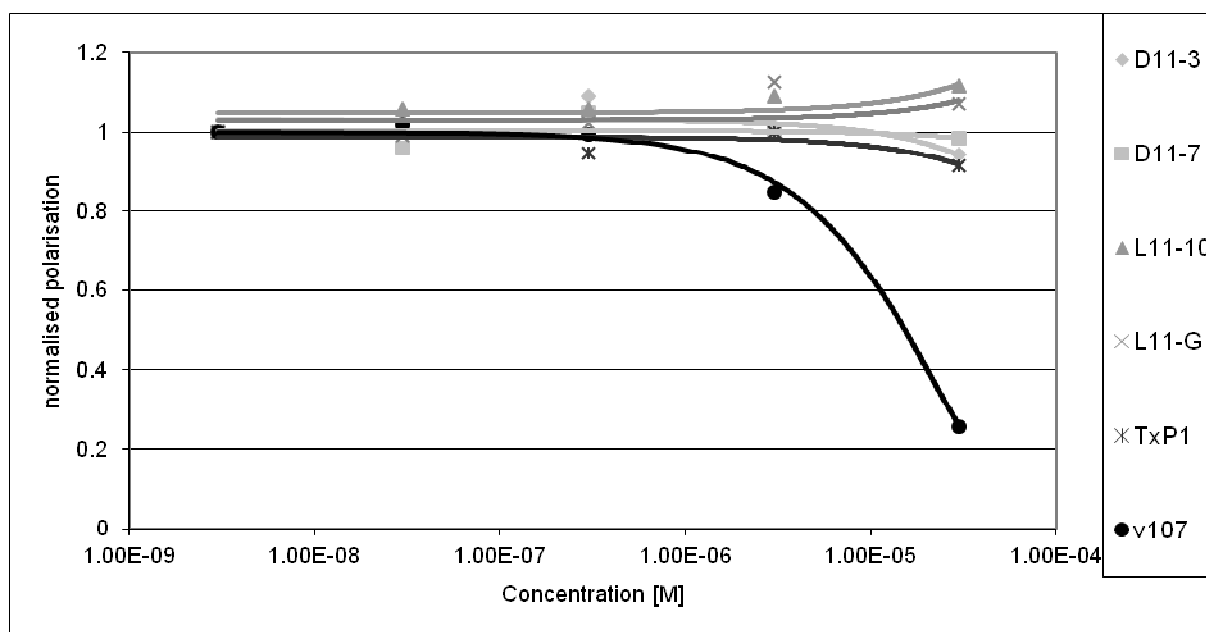

Figure 3.7.6-1: Fluorescence polarisation binding assay for synthesised peptides of the first group and v107 as positive control. The assay is based on the use of 10 nM tracer, which binds to the protein-protein interface of VEGF and 150 nM of protein.

The positive control showed the expected behaviour of increasing polarisation values with increasing concentrations of v107 indicating displacement of the tracer. The mathematical fit of this behaviour with Graph Pad Prism[2] led to an apparent affinity value of 1.6 uM. This value is only slightly higher than the value reported in the literature (Kd= 1 uM) and demonstrates that the assay works properly for our purpose. The synthesised peptides did not exhibit increasing polarisation values in the range of tested concentrations. Assaying higher peptide concentrations was not possible since the peptides began to affect the tracer in the absence of protein. These results indicate that either the peptides do either not bind with a sufficient affinity to the protein-protein interface to replace the tracer or that they bind to other surface areas of the protein.

The second group of peptides was labelled with carboxyfluorescein to assess their affinity independently from their binding site. Since we had only a limited amount of VEGF available during the time of this assay, we used a one-point concentration setup. Each labelled peptide and free carboxyfluorescein, used as a reference, were incubated either with or without 2 uM of VEGF. The polarisation values were measured and the effect of the presence of protein was calculated (Table 3.7.6-1). Carboxyfluorescein was used as a reference and the value in the absence of the protein was defined as 20. The peptides exhibited in the absence of the protein increased to values of around 100. This is probably due to their increased molecular weight, but could also indicate aggregation. The assay did not detect a significant increase in polarisation in the presence of the protein. Since we used an excess of protein, at least strong binding peptides should exhibit a significant polarisation increase. One may suspect that the use of carboxyfluorescein as a reference may decrease the ability of the assay to properly detect the interaction of such peptides.

Peptide (40 nM)	L6-3	D6-H	D6-2	D6-L	D6-F	CF
FP in absence of VEGF	100	83	98	82	109	20
FP in presence of 2 uM VEGF	109	79	111	92	113	32
change of polarization	1.09	0.95	1.13	1.12	1.04	1.45

Table 3.7.6-1: Fluorescence polarisation values acquired for labelled peptides and free carboxyfluorescein (CF) in absence and presence of protein.

My collaborator Dr. Christopher Hipolito pointed out that fluorescence polarization, although applied routinely in the lab of Prof. Suga, failed in several project to detect sub uM interactions, which could be confirmed with SPR and bioassays. This led SPR to be established as the method of choice for the characterization of mRNA display derived peptides with their target proteins. Therefore, we decided to test our peptides with additional biophysical methods. Due to time constrains we did not use SPR but relied again on protein based NMR experiments.

Peptides D11-3, D11-7, L11-10 and TX showed good solubility in pure water. They were tested by ^{1}H-^{15}N-HSQC NMR CSP for binding to VEGF at concentrations of 300 and 200 uM in the case of

peptide TX. Contrary to the behavior in pure water, peptide L11-10 showed strong precipitation and no protein signals could be observed. Some precipitate appeared for D11-7 while nearly no precipitation occured for peptides D11-3 and TX. The resulting NMR spectra showed CSP for peptide D11-3 and D11-7 (figure 3.7.6-2), but not for TX. However the results required further confirmation due to the appearance of precipitate.
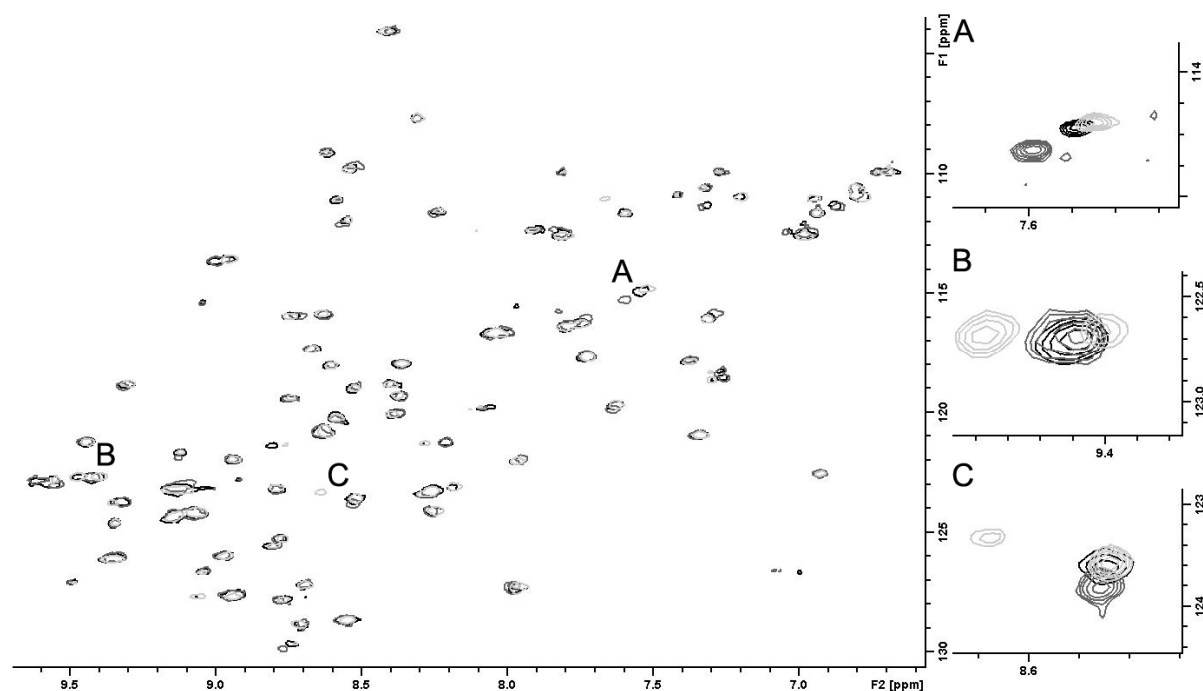
Figure 3.7.6-2: ^{1}H-^{15}N-HSQC spectra of 50 uM VEGF and 5% DMSO at 45℃ in absence of peptide (light grey), 300 uM D11-7 (dark grey) and 300 uM D11-3 (black). The D11-7 sample contained precipitation after retrieving it from the magnet.

The solubility of peptides L11-10 and L11-G was determined to be only 40 uM and 100 uM respectively in the presence of protein. Due to the low solubility, we used VEGF concentrations of 25 uM for the subsequent 1H-15N-HSQC experiment. Due to the length of the experiments no reference spectrum could be acquired. The comparison of both spectra showed perturbation and intensity changes of some peaks (Figure 3.7.6-3). Without this and due to the low solubility, it is not clear if one, both, or neither peptide binds to the protein.
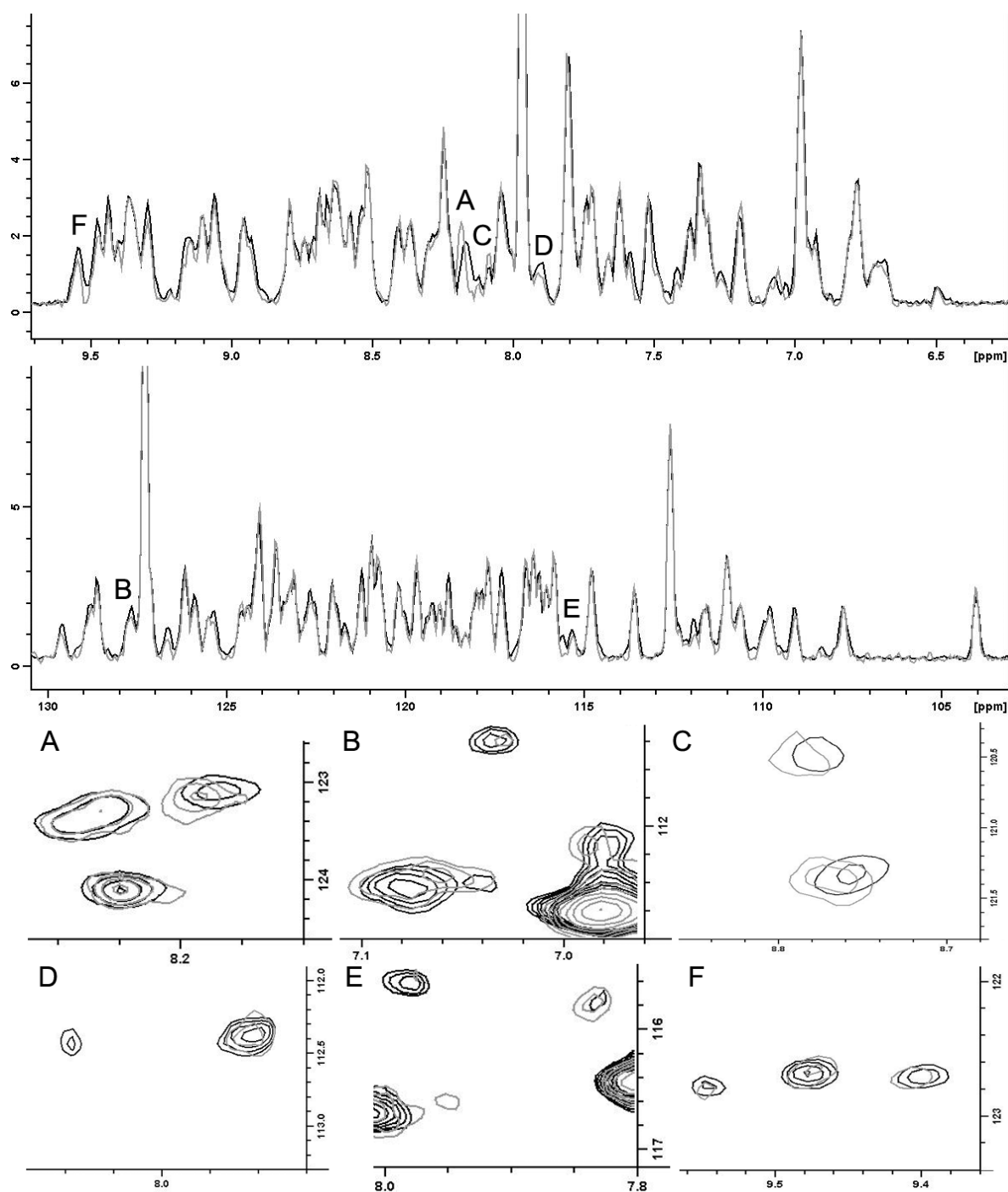
Figure 3.7.6-3: ^{1}H-^{15}N-HSQC spectra of 25 uM VEGF and 5% DMSO at 45°C in presence of 40 uM L11-10 (black) or 100 uM L11-G (grey). First two pictures contain the ^{1}H and ^{15}N 1D-projections, A-F mark differences between both spectra and zooms of regions in the 2D-spectra.

We wanted to retest some peptides together with the second group with was labelled with carboxyfluorescein. For this purpose we tested the minimal solubility in exactly the same buffer that was used for the protein during the NMR experiments. After measuring the minimal solubility, we titrated samples containing 50 uM of VEGF with increasing concentrations of peptide while monitoring the CSP behaviour of the [methyl-^{13}C]-methionine groups. Surprisingly the majority of the

samples showed the appearance of precipitate at or under the previously determined minimal solubility in the absence of protein (Table 3.7.6-2).

ID	D11-3	D11-7	D6-F	D6-L	D6-2	D6-H2	L6-3
Minimal solubility buffer (uM)	not re-tested	75	13	147	75	37.5	28
Maximal solubility in presence of VEGF (uM)	50	75	not tested	no	30	22.5	27

Table 3.7.6-2: Minimal and maximal solubility of peptides in absence and presence of 50 uM VEGF.

The titration could only be finished for peptide D6-L which only showed very weak CSP and intensity changes at a concentration of approx. 150 uM. The majority of the remaining peptides showed very weak CSP for at least one methionine signal but at much lower concentrations. Since the concentration was except for D6-L equal or lower than the protein concentration and because the protein is a dimer, it was difficult to predict if the chemical shift was significant and, if it was, then whether the peptides may bind with strong or weak affinity.

3.7.7 Recapitulation and concluding remarks

To date, we do not have a clear indication of whether our mRNA display selection was successful or not.

An early issue was that we were forced to use the refolded protein solution that was brought from Barcelona. Although we managed to enrich the native dimer on the beads there was also an impurity of trimer. Nevertheless, there were no further problems during the selection. The sequencing results showed enrichment of specific motives, rather than sequence convergence. Although this is unusual, it has been observed before in the Suga lab. The outcome of the clone assay was very positive with respect to the DNA recovery rates and the competition between bound protein and well folded-his tag free VEGF in solution. Later would indicate the selection of peptides that recognize the native protein stucture. The fluorescence polarization assays that we used did not indicate a strong affinity of our peptides towards VEGF. According to our collaborators this issue also occurred in several successful selections and therefore additional methods should be exploited. We decided to use protein based NMR experiments. Although we could observe some peptide-induced perturbations, their magnitude was very week. A major issue was the low solubility of the peptides that seemed to decrease significantly in the presences of protein. This could indicate some interaction of protein and peptides, if one assumes that the complex has a reduced solubility. Protein-based NMR experiments may not be suited for this scenario since we require excess ligand compared to protein. This was not possible for our ligands. Therefore, the obtained results do not allow for a final conclusion of

whether the peptides bind strongly, weakly, or not at all to the protein. An appropriate alternative may be SPR which the member of Prof. Suga's lab rely heavily on.

(1)   Peterson, K. J.; Sadowsky, J. D.; Scheef, E. A.; Pal, S.; Kourentzi, K. D.; Willson, R. C.; Bresnick, E. H.; Sheibani, N.; Gellman, S. H. Anal Biochem 2008, 378, 8.
(2)   Prism Graph Pad, 2013, http://www.graphpad.com/.

4 Materials and Methods

4.1 Protein production

Expression of VEGF

Competent B834 (DE3) E. coli cells were transformed with p6XHisVEGF11-109 and pMS421 plasmids and plated on solid LB/Agar media. 5ml of overnight carbenicillin/spectinomycin containing LB rich media were inoculated into 0.5 to 1 L M9 minimal media containing 80 mg/L Met(13C)methionine and 1 g/L 15NH4Cl as nitrogen source (both from Cambridge Isotopes). Growth was monitored using optical density at 595 nm until a value of 0.8 and then protein expression induced with 1 mM IPTG for 4 to 6 hours at 37°C. Large scale expression was performed in the Protein expression service of the University of Barcelona at a 10 L scale with a Biostat B fermentor (Sartorius). In this case a preculture of 0.5 L of M9 minimal media was used and protein expression induced at an optical density of 1.2 and harvested at an OD around 6.0. During the fermentation the pH was kept sonstant at 7.0 and the medium saturated with oxygen by purging with air at 10L/ min.

In the case of non-labeled VEGF, that was desired for MS and X-ray crystallography standard LB media was used instead of M9 minimal media.

Lysis and Histag purification

Harvested cells were dissolved in 6 M guanidine HCl, 0.1 M NaH2P04, 10 mM Tris, 10 mM 2-mercaptoethanol, 10 mM imidazole pH 8 and then filtered. The lysate was purified by FPLC (ÄTKA explorer) using HisTrap HP 5mL columns precharged with Ni2+. The immobilized protein was washed with lysis buffer and consequently 8 M urea, 0.1 M NaH2P04, 10 mM Tris, 5 mM DTT, pH 8.0; 8 M urea, 0.1 M NaH2P04, 10 mM Tris, 5 mM DTT, pH 6.3; 8 M urea, 0.1 M NaH2P04, 10 mM Tris, 5 mM DTT, pH 5.9, 0.1 M imidazole. The protein is eluted with 8 M urea, 0.1 M NaH2P04, 20 mM Tris HCl, 5 mM DTT, 0.6 M imidazole, pH 5.9. Fractions are analyzed by SDS-PAGE and VEGF containing fractions are selected and pooled together.

Refolding and enzymatic digestion

Protein concentration was estimated by A_{280} (with ε_{280} = 0.47 mL $\cdot$ mg^{-1} $\cdot$ cm^{-1}) and diluted to 1 mg/ml. The pH was adjusted to 8.0, DTT added to a final concentration of 20 mM and reduction allowed to proceed for 3 hours, at room temperature with gentle stirring in the dark. Refolding was proceeded by stepwise removal through dialysis (MWCO 6000-8000 Da) of urea during three days and cysteine in the consequent three days at 4°C from the initial solution (20mM Tris HCl, 8M urea, 25mM cysteines at pH 8.4). Refolding yield was assed by non-reducing SDS-PAGE. His-tag proteolytic cleavage was performed using 100:1 (VEGF:Genenase I) ratio in 20mM Tris HCl buffer at pH 8.4, 200

mM NaCl and 1mM EDTA overnight at room temperature. Digested protein was dialyzed into 20mM Tris HCl pH 7.5 at 4°C.

Anion exchange and size exclusion

Purification was performed with a HP-Q sepharose 5 mL column using a binding buffer of 20mM Tris HCl pH 7.5 buffer at 20°C. The Protein was eluted by raising NaCl gradient up to 1M. Protein eluates are analyzed by non-reducing SDS-PAGE and VEGF containing fractions pooled together and concentrated by ultrafiltration prior to S75 size exclusion purification. S75 size exclusion was performed using 25mM phosphate buffer pH 7, 50mM NaCl. Pure protein fractions are checked by non-reducing SDS-PAGE. After scaling protein expression up and optimization of the previous purification steps no size exclusion was performed after anion exchange as the protein eluate showed sufficient purity.

Sample preparation for NMR

VEGF concentration was adjusted to around 200 uM and the buffer exchange performed with a PD-10 (GE Healthcare) desalting column utilizing the gravity protocol. VEGF for screening purpose was exchanged to 100% deuterated 50 mM phosphate buffer, 50mM NaCl at pH 7. VEGF for HSQC based experiments was exchanged to 10% deuterated 25 mM phosphate buffer, 50mM NaCl at pH 7 containing 0.02% NaN$_3$. All stocks were shockfrozen and stored at -20 °C.

4.2 Cheminformatics

4.2.1 Selection of appropriate fragments

Libraries from up to nine international vendors were downloaded in the sdf format and merged into a single database. Only compounds with a MW between 150 and 305 Da were retained. The remaining compounds were filtered with a MOE (Chemical computing Group) script developed by Prof. Xavier Barril with the following descriptors:

```
local Fdesc =[
//      [min,'descriptor_name',max],
        [150,'Weight',305],
        [-3.5,'SlogP',3.0],
        [-3.0,'logS',100.0], // Solubility at 1mM
        [1,'opr_nring',3], // At least 1 ring (Baurin et al; SHAPES cores
are rings)
        [0,'opr_nrot',6],  // nu. rotatable bonds (Oprea definition)
        [0,'lip_don',4],  // nu. donors (Lipinski definition)
        [0,'lip_acc',6],  // nu. acceptors (Lipinski definition)
// the filters below are combinations of descriptors
        [2,['don_acc','lip_don','add','lip_acc'],8],  // Donor+Acceptors
//      [1,['O+N','a_nO','add','a_nN'],100],    // O+N >= 1 NOT NECESSARY
```

```
          [0,['Cl+Br+I','a_nCl','add','a_nBr','add','a_nI'],1], // Cl+Br+I <=
1
          [0.50,['C_ratio','a_nC','div','a_heavy'],1.00] // C_ratio >= 0.5
(C_ratio = #C/#Atoms)
        ];
local Fsmi = [
//      ['smiles_string',max_nu_occurrences,'name (for reference only)'],
// ZINC (JCIM 2005, 45, 177-182): Remove all molecules containing an
// atom other than: H,C,N,O,F,S,P,Cl,Br or I
        ['[#M]',1,'Inorganic'], // Metals
        ['[#S]',1,'Inorganic'], // Semi-Metals (B,Si,Ge,As,Sb,Te)
// R01-R31 taken from JCICS 1999, 39, 897-902
        ['[Cl][#6X4]',1,'R01'], // Alkyl Halides (extended)
        ['[Br][#6X4]',1,'R01'], // Alkyl Halides (extended)
        ['[I][#6X4]',1,'R01'], // Alkyl Halides (extended)
        ['[Cl][#6r0]',1,'R01'], // Alkyl Halides (extended)
        ['[Br][#6r0]',1,'R01'], // Alkyl Halides (extended)
        ['[I][#6r0]',1,'R01'], // Alkyl Halides (extended)
        ['[#G7]C(=O)',1,'R02'], // Acid Halides
        ['[#G7]C(=S)',1,'R02'], // Acid Halides
        ['[#G7]S(=O)',1,'R02'], // Acid Halides
        ['O=CN=[N+]=[N-]',1,'R03'], // Carbazide
        ['COS(=O)O[#6]',1,'R04'], // Sulphate ester
        ['COS(=O)=O[#6]',1,'R05'], // Sulphonate esters
        ['C(=O)OC(=O)',1,'R06'], // Acid anhidride
        ['OO',1,'R07'], // Peroxides
        ['C(=O)[Or0]c1c(F)c(F)c(F)c(F)c1(F)',1,'R08'], // Pentafluorophenyl
esters
        ['C(=O)[Or0]c1ccc([N+1](=O)[O-1])cc1',1,'R09'], // Paranitrophenyl
esters
        ['C(=O)Onnn',1,'R10'], // Esters of HOBT
        ['N=S=O',1,'R11'], // Isocyanates
        ['N=S=S',1,'R11'], // Isothiocyanates
        ['OS(=O)(=O)C(F)(F)F',1,'R12'], // Triflates
        ['P(=S)(S)S',1,'R13'], // Lawesson's reagents and derivatives
        ['NP(=O)(N)N',1,'R14'], // Phosphoramides
        ['cN=[N+1]=[N-1]',1,'R15'], // Aromatic Azides
        ['C(=O)[C;r0][#7+1]',1,'R16'], // Beta carbonil quaternary nitrogen
        ['[Nr0][Nr0]C(=O)',1,'R17'], // Acylhydrazide
        ['[C+1]',1,'R18'], // Quaternary C
        ['[#G7+1]',1,'R18'], // Quaternary Halogen
        ['[P+1]',1,'R18'], // Quaternary P
        ['[S+1]',1,'R18'], // Quaternary S
        ['[C-1]',1,'R18'], // Anionic C (XB extension)
          ['[n+1]',1,'R18'], // Cationic Aromatic nitrogen (XB extension)
        ['C=P',1,'R19'], // Phosphoranes
//      ['[Cl]C([C;r0])=N',1,'R20'], // Chloramidines (extended R01 covers
it)
        ['[ND2](=O)',1,'R21'], // Nitroso
//      ['P[#G7]',1,'R22'], // P halide (extended R01 covers it)
//      ['S[#G7]',1,'R22'], // S halide (extended R01 covers it)
        ['N=C=N',1,'R23'], // Carbodiimide
//      ['[N+1]#[C-1]',1,'R24'], // Isonitrile (extended R18 covers it)
        ['C(=O)N(C(=O))OC(=O)',1,'R25'], // Triacyloximes
        ['N#CC[OH]',1,'R26'], // Cyanohydrins
        ['N#CC(=O)',1,'R27'], // Acyl cyanides
        ['S(=O)(=O)C#N',1,'R28'], // Sulfonyl cyanides
        ['P(OCC)(OCC)(=O)C#N',1,'R29'], // Cyanophosphonates
        ['[Nr0]=[Nr0]C#N',1,'R30'], // Azocyanamides
        ['[Nr0]=[Nr0]CC=O',1,'R31'], // Azoalkanals
// R32-R37 taken from JCAMD 2002, 16, 311-323
        ['C1OC1',1,'R32'], // Epoxide
        ['C1SC1',1,'R33'], // Thioepoxide (extension of R32)
```

```
        ['C1NC1',1,'R34'], // Aziridine (extension of R32)
        ['[CH]=O',1,'R35'], // Aldehyde
        ['C(=O)[C;r0](=O)',1,'R36'], // 1,2 dicarbonyl (acyclic)
        ['[N+1](=O)[O-1]',1,'R37'], // Nitro
// R38-49 taken from DDT 1997, 2, 382-384 (See also DDT 2003, 8, 86-96)
        ['[#G7]c1ncccn1',1,'R38'], // Halopyrimidines
        ['[CX4]C(=[#7r0][#6])[#6]',1,'R39'], // Imine
        ['[CX4]C(=[#7Hr0])[#6]',1,'R39'], // Imine
        ['[CX4]C(=O)C([#G7])([#G7])[#G7]',1,'R40'], // Perhaloketones
//      ['[CX4]C(=O)O[#6]',1,'R41'], // Aliphatic esters (too agresive?)
        ['[CX4][C;r0](=O)[#6]',1,'R42'], // Aliphatic ketones (acyclic)
        ['[CX4][CH]=[CH]C(=O)[#6]',1,'R43'], // Michael Acceptors
        ['[CX4]C(=O)S[#6]',1,'R44'], // Thioesters
        ['[CX4]P(=O)(=O)O[#6]',1,'R45'], // Phosphonate esters
        ['[#7X3]-[#8X2]',1,'R46'], // -N(*)-O- single bond (modified)
        ['[#7X3]-[#7X3]*',1,'R47'], // -N(*)-N(*)- single bond (modified)
        ['[#7X3]-[SX2]',1,'R48'], // N-S(reduced) single bond
        ['SS',1,'R49'], // S-S
        ['[SX2]-[OX2]',1,'R50'] // S(reduced)-O single bond
      ];
```

4.2.2 Library design

Fragments with desired properties were assigned to clusters based on their MACCS key fingerprint similarity. From each cluster the candidate closest to the centroid was selected for purchasing. Both steps were performed with scripst developed by Prof. Xavier Barril in MOE.

Compounds for the second part of our fragment library were rejected if they exhibited a similarity over the threshold to compounds of the first part of the library. This was computed with the above script and while our first library was used as reference set.

The chemical space around potential hits from our screening exercise was investigated by calculating the similarity of commercially available compounds to our fragment hits. This task was performed by a MOE script provided by Prof. Xavier Barril.

4.3 automatic mixture design

NMR spectrometry and computation

All NMR spectra were acquired on a Varian Inova 500 MHz spectrometer with a 5mm PFG Penta Probe at 37 °C. S/N for ^{1}H was 815:1 (0.1% ethylbenzene in CDCl$_3$).

All calculations were performed on an SGI® Altix® 4700 server (64 cores, 128 GB RAM).

Sample preparation, and generation of fragment fingerprints

Stock solutions (100 mM in 9:1 DMSO-d$_6$/D$_2$O) of each compound from our in-house fragment library were prepared, and then inspected visually to confirm solubility. Soluble compounds were further diluted to 1 mM in deuterated buffer (25 mM phosphate, 50 mM NaCl, 11µM *t*-butanol, pH 7.0), and their individual 1D-^{1}H-NMR spectra were recorded using presaturation for water suppression. The purity and identity of each fragment was manually checked in each spectrum. Compound concentration was calculated based on an internal standard (*t*-butanol). NMR data for all fragments

that passed quality control were then translated into computer-readable files, called *fingerprints*, by a modified Varian script for automatic processing. The routine for signal integration was modified to integrate a narrow zone around each signal and to create an ASCII file for each NMR spectrum, consisting of the integration range of each signal and the value of the integral. Therefore, the generated file comprises several regions defined by start and end values that mark the spectral regions containing signals. Raw data were adjusted for subsequent calculations by removing regions originating from H_2O (4.780-4.530), DMSO (2.754-2.613) and *t*-butanol (1.320-1.130) and by reducing the size of all remaining regions by 50%.

The mixtures of the second part of our fragment library we processed by MestreNova (Mestre) instead of the modified Varian script to generate fingerprint files. This allowed the parallel processing of over 100 spectra while visual inspection was possible weather the fully automatic processing of the Varian script led in some cases to baseline distortions and phasing problems. In detail the processing was performed by: 1. Fourier transformation with four times zero filling and exponential apodisation of 0.5 Hz; 2. Automatic phasing of all spectra and visual inspection; 3. Parallel automatic baseline correction with Bernstein Polynominal Fit (Order = 3), visual inspection; 4. Assignment of DSS signal to 0.00 ppm; 5. Automatic peak picking with noise factor 25 (a very low cutt off), max Peaks 100, only positive peaks; 6. Automatic signal integration with sum method, autodetection with algorithm peak picking (uses predefined peaks), J Max = 16 Hz, peak width factor = 1.00, Minimum Area = 1%, peak width factor =1; 7. Parallel normalization of DSS integral to 1 mM of ^{1}H.

4.4 NMR spectroscopy and data processing

4.4.1 NMR experiments for the setup and the screening of our libraries:

NMR spectra were acquired on a Varian Inova 500 MHz spectrometer with a 5mm PFG Penta Probe at 25 °C. S/N for ^{1}H was 815:1 (0.1% ethylbenzene in CDCl3). Fragment concentration was adjusted to 1 mM for quality control and 0.5 mM for screening purpose. In both cases a 100 % deuterated buffer consisting of 50 mM phosphate buffer and 50 mM of NaCl at pH 6.6* was used. ^{1}H spectra were recorded with preset water suppression and 32 increments.

CPMG spin lock filter length of approx. 20 ms and 400 ms were used. For STD experiments a pulse sequence with water suppression by the watergate technique was selected and between 256 and 2048 increments chosen. The length of the mixing time to remove residual protein signals was set to 10 or 20 ms. The saturation time was set to 2.5 s and the relaxation delay to 2 s. On and off resonance frequency for VEGF were -0.09 and 35 ppm. The total delay between two on-resonance scans was around 11 s. The experiments were either conducted at 37°C or 25°C.

4.4.2 ^{19}F-NMR experiments for the setup and the screening of our libraries:

All ^{19}F-NMR experiments for screening were performed in the Spectroscopy and NMR Unit of CNIO in Madrid on a 700 MHz NMR spectrometer equipped with a dual fluorine/proton probe.

4.4.3 Protein based NMR experiments

HSQC experiments were conducted on a Bruker Digital Avance 600MHz equipped with a cryoprobe. Protein concentration was adjusted to 100 uM. Samples were measured at 45 °C. 15N-1H HSQC Spectra were acquired using the fhsqcf3gpph pulse sequence with 128x2048 complex points with a total of 8 transients per increment. 1H-13C-HSQC-spectra were recorded with the hsqcetgpsisp2 pulse sequence with 128x512 to 96x256 complex points with 8 transients per increment.

Data processing was performed with Topspin 2.0. and 3.0. Briefly the f2 domain of the datasets was increased to a factor of two by linear prediction and then the f1 and f2 domains zero-filled by another factor of two. A qsine function was used for line broadening.

4.4.4 Analysis of CSP NMR data:

For the data acquired from ^{1}H-^{15}N-HSQC and ^{1}H-^{13}C-HSQC normalized chemical shift was calculated as:

$$\Delta\delta_{NH} = \sqrt{\left(\Delta\delta_H^2 + \left(\frac{\Delta\delta_N}{5}\right)^2\right)} \; ; \; \Delta\delta_{CH} = \sqrt{\left(\Delta\delta_H^2 + \left(\frac{\Delta\delta_C}{4}\right)^2\right)}$$

Binding sites from ^{1}H-^{15}N-HSQC were defined on the basis of significant CSP. This was defined as normalized CSP ≥ (Av. normalized CSP + STD).

For assessment of the affinity range of fragments the curve for Δδ versus total fragment concentration was fitted to the analytical model assuming two independent binding sites:

$$\frac{[L_0]-[L]}{[P_0]} = \frac{K_D + [L_0] + 2[P_0] - \sqrt{(K_D + [L_0])^2 + 4[P_0](K_D - [L_0] + [P_0])}}{2}$$

whereby [P0] is the total protein concentration; [L0], the total ligand concentration; [L], the concentration of unbound ligand in solution; F, a scaling factor; and Kd, the affinity to be calculated where [P0] is the total protein concentration and [L] is the concentration of unbound ligand in solution.

4.5 General peptide synthesis

Peptide synthesis was performed using a standard solid state peptide synthesis (SSPS) protocol, with Fmoc-protected amino acids (Iris Biotech GmbH, Gaurstrasse, d-95615, Marktredwitz, Germany). Other reagents, if not otherwise stated, were purchased from Carlo Erba Reagenti SpA, Strada Rivoltana, I-20090, Rodano. Dimethylformamide (DMF) and Dichloromethane (DCM) solvents were used; Piperidine/DMF 20%/80% (v/v) was used as a deprotection agent with 1mM of

N-Hydroxybenzotriazole (HOBt) in order to prevent the racemization of Cys; TBTU (Iris Biotech) as a coupling activator, and N,N-Diisopropylethylamine (DIEA, Fluca Chemie GmbH, CH-9471, Buchs, Switzerland) as an activator base. As solid support, Aminomethyl ChemMatrix resin (Matrix Innovation, 1450 City Councillors Suite 230, Montreal (Quebec), Canada), previously functionalized by Fmoc-Rink-Amid Linker (Iris Biotech) was used. The synthesis itself was carried out with the CEM Liberty Automated Microwave peptide Synthesizer (CEM Corporation, 3100 Smith Farm Rd, Matthews, North Carolina, United States) following standard procedures. The scale of synthesized peptides was 100 umol.

Lyophilized crude peptides were purified using Waters HPLC semi preparative instrument, equipped with Waters Symmetry C18 5μM 30x100 mm preparative column and freeze dried to yield the pure peptide.

4.5.1 Special issues of v107 analogues

Deprotection and cleavage was performed with TFA:H2O:ethanedithiol: triisopropylsilane 94:2.5:2.5:1 for up to 3 h under shaking.

Oxidation of Cys and formation of the Cys5-Cys15 disulfide bond were performed using Potassium Ferricyanide $K_3Fe(CN)_6$ (Sigma-Aldrich Co Ltd, Gillingham, Dorset, SP8 4XT, UK). The peptide was dissolved in 5mM NH_4HCO_3 aqueous buffer in a concentration of approx. 20 mg/ml, resulting in 2 liters of solution. 1ml of 20 mM $K_3Fe(CN)_6$ was added every 30 minutes for 2 hours. The reaction was monitored by Matrix-Assisted Laser Desorption Ionization (MALDI) TOF MS. After 2 hours, excess solvent was removed with a rotary evaporator and the sample was freeze-dried.

MALDI TOF MS experiments were performed on an Applied Biosystems 4700 Proteomics Analyzer instrument (AB, 850 Lincoln Centre Drive Foster City, CA 94404 USA). For the data analysis Data Explorer 4.5 was used.

4.5.2 Special issues Synthesis of mRNA display derived peptides.

Chloroacetylation of free N-terminus was performed with a solution of 2.5 mL of 0.2 M Cl-Ac-NHS dissolved in NMP and incubated under shaking for 40 minutes. Afterwards the beads were washed 5x with DMF.

Fluorescein conjugation of (MMT)-Lysine functionalized peptides was performed after Chloroacetylation and prior to cleavage form the resin. For this purpose the resin was incubated 10 times with 1% TFA, 1% TIS in CH2Cl2 for 10 min to remove MMT protection. Then the resin was washed subsequently with CH_2Cl_2, NMP, 10% DIPEA in NMP and finally 0.1M NHS-Fluorescein in 10% DIPEA in NMP for one hour.

Deprotection and cleavage was performed with 925:25:25:25 TFA:TriIsopropylsilane:Ethanedithiol:H2O for up to 3 h under shaking. Subsequently TFA was evaporated under nitrogen stream. The resulting crude was dissolved in 2.5 mL DMSO, 0.1% TFA and then diluted with 1000 mL Acetonitrile/H2O, 0.1% TFA. The pH was adjusted to 9.5 with Triethylamine and the solution stirred for 1 h at room temperature. Then excess solvent was removed with a rotary evaporator and the remaining solution was freeze-dried.

4.5.3 Summary of synthesized peptides

Spy molecule	$NH_2G-G-M-E-C-D-I-A-R-M-W-E-W-E-C-F^*-E-A-G-CONH_2$ (C and C linked via S—S disulfide bridge)		
Formula	C94H129FN26O30S3	MW (g/mol)	2218.38
Yield (%)	32.8	Purity (%)	>95
HPLC	G0100t8 (C18 4.6mmx100mm)	tR	4.8
Expected mass (charge state)	m/z 1109.4335 (z=2)	Found mass (charge state)	m/z 1109.4354 (z=2)
D11-3	$f-P-K-Y-R-I-L-Y-V-R-R-K-P-C-G-S-CONH_2$ (f linked via C=O to C)		
Formula	C100H160N32O21S	MW (g/mol)	2178.61
Yield (%)	15.43	Purity (%)	>95
HPLC	G0100t8 (C18 4.6mmx100mm)	tR	4.2
Expected mass (charge state)	726.7451 (Z=3)	Found mass (charge state)	726.7451(z=3)
D11-7	$f-R-Y-Y-I-V-L-K-R-R-R-T-R-I-S-C-G-S-CONH_2$ (f linked via C=O to C)		
Formula	C103H169N35O24S	MW (g/mol)	2313.73
Yield (%)	15.43	Purity (%)	>95
HPLC	G0100t8 (C18 4.6mmx100mm)	tR	4.1
Expected mass (charge state)	771.7678 (z=3)	Found mass (charge state)	771.7676 (z=3)
L11-10	$F-R-R-Y-R-C-Y-K-V-Y-T-F-T-V-K-C-G-S-CONH_2$ (C6 linked via C=O to C; C16 has ACM)		
Formula	C109H163N31O26S2	MW (g/mol)	2387.78
Yield (%)	0.32	Purity (%)	>95
HPLC	G070t8 (C18 4.6mmx100mm)	tR	4.9
Expected mass (charge state)	796.4020 (z=3)	Found mass (charge state)	796.4005 (z=3)

L11-G				
	F—R-CH-R—R—I—K—I—S—Y—Y—Y—L—A—C—G—S-CO (ACM on C; O on CH; NH₂ on CO)			
Formula	C103H160N32O24S2	MW (g/mol)	2294.7	
Yield (%)	2.3	Purity (%)	>95	
HPLC	G5100t8 (C18 4.6mmx100mm)	tR	4.1	
Expected mass (charge state)	765.3989 (z=3)	Found mass (charge state)	765.3976(z=3)	
TX				
	w—R—R—F—C—S—F—R—R—V—V—R—P—C—G—S-CO (Me on F, C, S; ACM on C; O; NH₂ on CO)			
Formula	C96H150N34O20S2	MW (g/mol)	2164.56	
Yield (%)	3.2	Purity (%)	>95	
HPLC	G1050t8 (C18 4.6mmx100mm)	tR	4.4	
Expected mass (charge state)	722.0480 (z=3)	Found mass (charge state)	722.0485 (z=3)	
D6-F				
	f—R—Y—A—F—W—K—I—R—Y—Y—K—P—S—V—C—A*-K—CONH₂ (CF on K; O)			
Formula	C137H175N29O29S	MW (g/mol)	2722.28	
Yield (%)	<0.5	Purity (%)	>95	
HPLC (column)	G2050t15 (C4 4.6mmx150mm)	tR	7.7	
Expected mass (charge state)	908.4355 (z=3)	Found mass (charge state)	908.4360(z=3)	
D6-L				
	f—R—Y—T—T—F—T—A—V—F—L—F—R—D—R—C—G-A*-K—CONH₂ (CF on K; O)			
Formula	C130H170N30O32S	MW (g/mol)	2696.99	
Yield (%)	<0.	Purity (%)	>95	
HPLC	G0100t18 (C18 4.6mmx100mm)	tR	4.9	
Expected mass (charge state)	899.4185(z=3)	Found mass (charge state)	899.4181(z=3)	

D6-2		CF (structure: f—W—T—Y—Y—L—I—V—S—K—S—R—N—P—H—C—A*—K-CONH₂ with C linked to A* and a ketone/carbonyl group =O)		
Formula	C127H164N28O31S	MW (g/mol)	2609.18	
Yield (%)	<0.5	Purity (%)	87	
HPLC	G2050t15 (C4 4.6mmx150mm)	tR	10	
Expected mass (charge state)	870.7358(z=3)	Found mass (charge state)	870.7344(z=3)	
D6-H		CF (structure: f—K—K—T—T—I—W—R—I—R—W—T—T—R—K—C—A*—K-CONH₂ with =O)		
Formula	C130H186N34O29S	MW (g/mol)	2721.14	
Yield (%)	<0.5	Purity (%)	80	
HPLC	G2080t8 (Phenomenex 4.6mmx 150 mm)	tR	3.9	
Expected mass (charge state)	907.4694(z=3)	Found mass (charge state)	907.4668(z=3)	
L6-3		CF (structure: F—R—S—Y—S—R—R—G—R—T—W—T—Y—H—C—S—K-CONH₂ with =O)		
Formula	C119H154N34O31S	MW (g/mol)	2588.77	
Yield (%)	<0.5	Purity (%)	>95	
HPLC	G1040t8 (C18 4.6mmx100mm)	tR	5.5	
Expected mass (charge state)	647.7888(z=4)	Found mass (charge state)	647.7870(z=4)	

Minor letters indicate D-amino acid e.g. f= D-Phe; w=D-Trp; A*= beta-alanine; F* para-fluoro-phenyl-alanine; ACM = S-acetamidomethylside side chain protection group; CF = Carboxyfluorescein.

4.6 mRNA display

Protocol for mRNA display from the laboratory of Prof. Hiroaki Suga for the NNK selection. All reagents were provided by the lab and the majority of enzymes as well as the cell free expression system expressed my lab members.

4.6.1 Protocol for the first round:
Ligation
1. Run the ligation reaction at r.t. for 30 min at 200 µL scale.

(x 2.2)

H2O	70 µL	(154 µL)
10x T4 RNA ligase buffer	20 µL	(44 µL)
DMSO (final 20%)	40 µL	(88 µL)
7.5 µM linker(f.c. 1.5 µM)	40 µL	(88 µL)
10 µM mRNA(f.c. 1 µM)	20 µL	(44 µL)
(mRNA from library with approx 200 pmol sequences, ⌒10^{14})		
T4 RNA ligase (home made)	10 µL	(22 µL)
Total	200 µL	(440 µL)

2. Add 200 ul of 0.6 M NaCl 10mM EDTA (pH 7.5) and do the P/C/I & C/I extraction. (Phenol/Choroform/ Isoamyl Alcohol)
   Add 800 ul of EtOH and centrifuge (13000 rpm r.t. 15min) & 70 % EtOH wash.
   Air dry.
   Dissolve in 30ul of water (~6 uM).

Translation

3. Translate the Pu-conjugated mRNA pool for 30 min @ 37°C.

   Translation mix
Water	16.3 µL (68.9 µL)
Hv7 SolA (10.9%)	16.4 µL (36.1 µL)
Hv7 SolB (10.2%)	15.3 µL (33.7 µL)
5 mM 19 aa (-Met) (f.c. 500 mM)	15 µL (33 µL)
Total	78 µL (171.6 µL)

   (Hv7 SolA & B are a reconstituted cell free protein expression system called PURE)

Translation mix	78 uL
250 µM ^{ClAc}F-tRNA^{ini} (f.c. 50 uM)	42 µL
6 µM mRNA fusion (f.c. 1.2 uM)	30 µL
Total	150 µL

4. Incubate 12 min @ RT to enhance the display efficiency.

5. Disrupt the ribosome structure by adding 15 µL of 200 mM EDTA (pH 7.5). Incubate 30 min @ 37°C for the cyclization.
   (A) Take 1 uL and dilute it in 99 uL of water.

Desalting

6. Add SephadexG25 suspension to 1 mL syringe up to the top(3,000 rpm, 1 min). Add 1xTBS (tris buffered saline) up to the top and rotate at 3000 rpm for 1 min (x2). Divide the sample into two and add each of translation product to the syringe. Collect the eluate with 1.5 mL tube (3000 rpm, 1 min).

Preclear (removing PURE proteins)

7. Add 165 µL of 2xBlocking solution to 165 uL of the eluate.
   2xBlocking solution: 1 M NaCl, 0.2% acBSA, 5 µM tRNA in 1xTBS.

8. Pre-clear the conjugate pool by incubating the pool with 150 µL of Dynatalon slurry for 30 min at 4°C. Take sup, add it to 30 uL of Dynatalon and rotate at 4°C for 30 min. (B) Take 1 µL and dilute it in 99 µL water for the RT-PCR (f.c. ~ 4nM).

Target-resin preparation

9. Saturate 15 μL resin slurry (x 2) with his-tagged VEGF by incubating 3x with 30 uL VEGF solution for 20 min @ 4°C with rotation and subsequential removal of supernatant. Wash the resin with 150 μL ice-cold 1xTBS three times (in the last washing step, whole resin suspension is transferred to a new tube and then sup is removed).

Selection

10. Add 300 uL of peptide-mRNA fusion to the resin and let the pre-cleared pool bind to the target-resin by incubating for 1h at 4°C.

11. Wash the resin with ice cold 3xTBS (800 μL each). In each step, take sup, resuspend the resin with TBS and transfer to the other tube and take sup again.

Reverse transcription (RT)

12. Add 8 uL of water and 32 uL of RT mix and incubate 1hr at 42

    *RTmix

Water	27 uL	(54 uL)
MMLV 5xRT buffer	12 uL	(24 uL)
5 mM each dNTP (f.c. 0.5 mM)	6 uL	(12 uL)
100 uM CGS3an13.R39 (f.c. 2.5 uM)	1.5 uL	(3 uL)
RNase in	0.2 uL	(0.4 uL)
MMLV RTase (+H) (1/40)	1.5 uL	(3 uL)
Total	48 uL	(96 uL)

    Add 4 uL of RT mix to 1 uL of (A) and (B) and incubate 1hr at 42

Real time PCR (RT-PCR)

13. Add 19 uL of RT-PCR mix to 1 uL of RT product of (A) and (B) and 0.1 fmol/uL RT-PCR STD (2x10^7).
    To the 40 uL of resin-slurry, add 200 uL of PCR mix. Incubate at 95   for 5 min and transfer the sup to new tube. Add 600 uL of PCR mix and 6 uL of Taq. Take 20 uL and add 0.2 uL of 1/10000 SYBR Green to it.
    Run RT-PCR.

    Aliquot the PCR mixture into 50 uL each.
    Run PCR (cross point + 4 cycles. Check by agarose and run more cycles.)

PCR mix (-Taq)		
Water	820 uL	(1600 uL)
10x PCR buffer	100 μL	(200 uL)
250 mM MgCl2	10 μL	(20 uL)
5 mM dNTP	50 μL	(100 uL)
25 μM T7g10M.F48 (f.c. 0.25 uM)	10 μL	(20 uL)
25 μM GS3an13.R39 (f.c. 0.25 uM)	10 μL	(20 uL)
Total	1000 μL	(2000 uL)

    RT-PCR mix
    200 μL PCR mix (-Taq) + 3 μL Taq + 2 μL 1/10000 SYBR (freshly prepared)

    SYBR
    100 times diluted from the stock in water.

To the 840 μL of PCR product, add 84 μL of 3M NaCl
P/C/I, C/I
add 2 vol. of EtOH and centrifuge
Air dry
Dissolve in 84 μL of 50 mM KCl.

<u>**Transcription**</u>

14. Transcription.

		(X2.2)
H2O	6.3 μL	(13.86)
10X T7 buffer	2 μL	(4.4)
100mM DTT	2 μL	(4.4)
250mM MgCl2	1.6 μL	(3.52)
2N KOH	0.22 μL	(0.484)
25mM NTPs	3 μL	(6.6)
Rnase inhibitor	0.1 μL	(0.22)
T7 pol.	0.8 μL	(1.76)
Total	16	(35.2)
DNA		4 μL
Total		20 μL

Incubate at 37 ˚C for 2h.
Add 20 uL of 0.6 M NaCl, 50 mM EDTA.
P/C/I, C/I
Add 32 uL of iPrOH and cf 13000 rpm at rt. for 5 min.
Dissolve the ppt in 40 μLof 0.3 M NaCl.
Add 32 uL of iPrOH and cf 13000 rpm at rt. for 5 min.
Wash the ppt with 70 % EtOH
Air dry, 5 min
Dissolve in 20 μL of water
Adjust the mRNA concentration to 10 μM.

4.6.2 Protocol for the second round and later rounds

<u>**Ligation**</u>

1. Run the ligation reaction at r.t. for 30 min at 20 μL scale.

		(x 2.2)
H2O	7 μL	(15.4 μL)
10x T4 RNA ligase buffer	2 μL	(4.4 μL)
DMSO (final 20%)	4 μL	(8.8 μL)
7.5 μM linker(f.c. 1.5 μM)	4 μL	(8.8 μL)
T4 RNA ligase (home made)	1 μL	(2.2 μL)
Total	18 μL	(39.6 μL)
10 μM mRNA(f.c. 1 μM)	2 μL	
Total	20 μL	

2. Add 20 ul of water and 40 uL of 0.6 M NaCl 10mM EDTA (pH 8.0) and do the P/C/I & C/I extraction.
   Add 160 ul of EtOH and c.f. (13000 rpm r.t. 15min) & 70 % EtOH wash.
   Air dry.
   Dissolve in 4 ul of water (~5 uM).

Translation

3. Translate the puromycin-conjugated mRNA pool for 30 min @ 37°C.

Water	0.54 uL	(1.19 uL)
Hv7 SolA (10.9 %)	0.55 µL	(1.21 µL)
Hv7 SolB (10.2 %)	0.51 µL	(1.12 µL)
5 mM 19 aa (-Met)	0.5 µL	(1.1 µL)
Total	2.1 µL	(4.62 µL)

^{ClAc}Tyr-tRNA^{ini} (final 50 uM)	1.4 µL
5 µM mRNA fusion (final 1.5 uM)	1.5 µL
Total	5 µL

4. Incubate 12 min @ RT to enhance the display efficiency.

5. Disrupt the ribosome structure by adding 1 µL of 100 mM EDTA (pH 7.5). Incubate 30 min @ 37°C for the cyclization.

Reverse transcription

6. Add RT mix 3.43 ul and incubate 1hr at 42°C

*RTmix		(x2.2)
PremixRT	1.68 ul	(3.696)
150 mM Mg(OAc)2 (f.c. 16 mM)	1 ul	(2.2)
0.2 N KOH	0.5 ul	(1.1)
RTase (–H)	0.25 ul	(0.55)

*premixRT
5mM dNTPs 5ul (f.c. 0.45 mM) + 100 uM CGS3an13.R39 1.68 ul(f.c. 1.8 uM) +250mM Tris-HCl (pH 8.3) 10 uL (f.c. 45 mM)

7. Add 21 µL 1xTBS to the RT mixture. Now you have 30 µL mRNA-peptide fusion solution. (A) Take 0.5 µL and dilute it in 499.5 µL water for the RT-PCR (f.c.~ 0.25nM).

Desalting

8. Run GPC using ~700 µL Sephadex G-25 pre-equilibrated in 1xTBS. (3000 rpm, 3 min) (B) Take 0.5 µL and dilute it in 499.5 µL water for the RT-PCR (f.c.~ 0.25nM).

9. Add 30 µL of 2xBlocking solution.
   Blocking solution: 1 M NaCl, 0.2% acBSA, 5 µM tRNA in 1xTBS.

Pre-clearance

10. Pre-clear the conjugate pool three times by incubating the pool with 5 µL of resin slurry for 30 min at 4°C (use 2.5 µL for the second and third pre-clearance). Now the volume must be around 60 µL. (C) Take 0.5 µL and dilute it in 499.5 µL water for the RT-PCR (f.c. ~ 0.25nM).

Preparation of target-resin

11. Saturate 1 µL resin slurry (x 2) with his-tagged VEGF by incubating 3x with 3 uL VEGF solution for 20 min @ 4°C with rotation and subsequential removal of supernatant. Wash the resin with 10 µL ice-cold 1xTBS three times (in the last washing step, whole resin suspension is transferred to a new tube and then sup is removed).

Selection

12. Divide the 60 ul of solution into 30 ul x 2. Let the pre-cleared pool bind to the target-resin and resin (no-target) by incubating the solution with resin for 30 min at 4°C.

13. Wash the resin with ice cold 1xTBS-T (x2) and 1xTBS (100 µL each). In each step, take sup, resuspend the resin with TBS and transfer to the other tube and take sup again.

RT-PCR&PCR

14. PCR amplify the selected sequences. Prepare PCR mix (-Taq) first, and take 300 µL of that to make RT-PCR mix. Use 100 µL to elute peptide-fusions from the resin by heating for 5 min. Take the sup, take 1uL and mix it with 19 uL of RT-PCR mix ((D)negative selection, (E)positive selection) and add 0.75 µL Taq. After getting RT-PCR result, the left-over (100 µL, 2x 50 µL) was used for the amplification. The PCR cycle is (crosspoint).

PCR mix (-Taq)

10x PCR buffer	98 µL	
250 mM MgCl2 (f.c. 2.5 mM)	9 µL	
5 mM dNTP (f.c. 0.25 mM)		45 µL
25 µM T7g10M.F48 (0.25 uM)	9 µL	
25 µM CGS3an13.R39 (0.25 uM)		9 µL
H2O	730 µL	
Total		900 µL

RT-PCR mix
300 µL PCR mix (-Taq) + 4.5 µL Taq + 3 µL 1/10000 SYBR (freshly prepared)
19 µL RT-PCR mix + 1 µL DNA sample (after RT (A), after GPC (B), after pre-clearance(C), sup after selection((F)negative selection, (G)positive selection) and STD 2x10^7 for the quantification
sample No. 1. STD 1x10^6, 2. Template (-), 3-9. L(A)-(G), 10-16. D(A)-(G)

To the 100 µL of PCR product, add 10 µL of 3M NaCl
P/C/I, C/I
Add 220 µL of EtOH and c.f.
Air dry
Dissolve in 10 µL of 50 mM KCl.

Transcription

15. Transcription.

		(x2.2)
H2O	6.3 µL	(13.86)
10X T7 buffer	2 µL	(4.4)
100mM DTT	2 µL	(4.4)
250mM MgCl2	1.6 µL	(3.52)
2N KOH	0.22 µL	(0.484)
25mM NTPs	3 µL	(6.6)
Rnase inhibitor	0.1 µL	(0.22)
T7 pol.	0.8 µL	(1.76)
Total	16 uL	(35.2)

DNA	4 µL	
Total	20 µL	

37 ˚C, 2h
Add 20 uL of water and 40 uL of 0.6 M NaCl, 50 mM EDTA.
P/C/I, C/I
Add 64 uL of iPrOH and cf 13000 rpm at rt. for 5 min.
Dissolve the ppt in 40 µLof 0.3 M NaCl.
Add 32 uL of iPrOH and cf 13000 rpm at rt. for 5 min.
Wash the ppt with 70 % EtOH
Air dry, 5 min
Dissolve in 20 µL of water
Adjust the mRNA concentration to 10 µM.

4.6.3 Preparation of Flexizymes (dFx or eFx)

Preparation of template DNA for in vitro transcription
1. Prepare PCR master mix on ice.

Water	276 uL
10x PCR buffer	33 uL
250 mM MgCl2	3.3 uL
5 mMeach dNTPs	16.5 uL
Taq DNA polymerase	2.48 uL
Total	330 uL

Keep the prepared master mix on ice.

2. Make extension reaction mixture

PCR master mix	100 uL
200 uM Fx-5'.F36	0.5 uL
200 uM eFx.R45 (or dFx.R46)	0.5 uL

3. Set extension reaction
95°C 1 min
50°C 1 min ⎤ 5 cycles
72°C 1 min
Use 1uL of the resulting solution for the next PCR reaction. Store the rest in -20°C.

4. Make PCR reaction mixture

PCR master mix	200 uL
Extension reaction	1 uL
100 uM T7ex5.F22	1 uL
100 uM eFx.R18 (or dFx.R19)	1 uL

5. Set PCR reaction
95°C 40 sec
50°C 40 sec ⎤ 12 cycles
72°C 40 sec

6. Check the amplification by agarose gel electrophoresis.

7. Phenol extraction
Add 200 uL of phenol/chroloform/isoamyl alcohol solution, then mix the sample severely.
 Centrifuge at 13000 rpm for 5min. Recover the aqueous layer.

Add 200 uL of chroloform/isoamyl alcohol solution, then mix the sample severely.
Centrifuge at 13000 rpm for 2min. Recover the aqueous layer.

8. Ethanol precipitation
Add 1/10 vol. of 3M NaCl and 2.2 vol. of 100% EtOH, then mix well.
Centrifuge the sample at 25°C at 13000 rpm for 15 min.
Discard the supernatant.
Add 200 uL of 70% EtOH and centrifuge the tube for 3 min, and then discard the supernatant.
Open the lid and cover the tube with clean paper towel, and stand for 5-10min to dry up the pellet.
Resuspend the resulting DNA pellet with 20 uL of H2O.

<u>in vitro transcription</u>
1. Prepare the reaction mixture for in vitro transcription
Prepare the following transcription reaction mixture.

H2O	69 uL
10x T7 buffer	20 uL
100 mM DTT	20 uL
250 mM MgCl2	24 uL
25 mM each NTPs	40 uL
2 M KOH	3 uL
DNA template (PCR product)	20 uL
T7 RNA polymerase	4 uL _
Total	200 uL

2. in vitro transcription reaction

Incubate the reaction mixture in the incubator at 37°C for over 3 hours (ideally for overnight).
After incubation, add 4 uL of 100 mM MnCl2 and 1 uL of DNase, then mix the sample well.
Stand the sample at 37°C for additional 30min (~1 hour).

3. Isopropanol precipitation
Add 15 uL of 0.5 M EDTA (pH 8.0) and 20 uL of 3 M NaCl into the sample.
Add 200 uL of isopropanol into the resulting sample.
Centrifuge at 25°C at 13000 rpm for 5 min. Discard the supernatant.
Add 100 uL of 70% EtOH and centrifuge for 3 min. Discard the supernatant.
Open the lid and cover the tube with a clean paper towel, and stand for 5-10min to dry up the pellet.
Resuspend the resulting pellet into 20 uL of H2O.

<u>Purification of the resulting RNA</u>
1. PAGE purification
Make 12% denaturing PAGE.

13.3% acrylamide (19:1), 6.7 M urea	18 mL
5x TBE	2 mL
10% APS	200 uL
TEMED	15 uL

Add 20 uL of 2x RNA loading buffer into 20 uL of RNA solution, then heat the mixture at 95°C for 1 min.
Load the RNA sample onto the gel and run at 230 V for 60 min.
Transfer the gel onto a TLC plate. Visualize RNA bands by a UV hand lamp in the dark room.
Mark shape of the band, then cut the gel containing objective RNA.
Crush the gel piece finely, then add over 1 mL of 0.3 M NaCl. Gently shake the tube at room temperature for 1-2 hours.

Filtrate the resulting mixture by 0.45 um syringe filter.
Estimate volume of the recovered RNA solution.
Add the twice volume of EtOH and mix well.
Centrifuge at 25°C at 13000 rpm for 5 min. Discard the supernatant.
Add 100 uL of 70% EtOH and centrifuge for 3 min. Discard the supernatant.
Open the lid and cover the tube with a clean paper towel, and stand for 5-10min to dry up the pellet.
Resuspend the resulting pellet into 10 uL of H2O.

2. Determination of RNA concentration
Take 0.5 ul of RNA solution and dilute with 4.5 uL of H2O.
Measure the absorption at 260 nm, and calculate the concentration of the resulting RNA solution.
(Observed absorption at 260 nm) * 40 = (ug/ml concentration of the RNA solution)
??? uM = [(obtained ug/ml) * 1000] / [(length of RNA)*340]
(1 OD = 40 ug/mL, dFx = 46 mer, eFx = 45 mer)
Generally, you can obtain 300~900 uM of RNA solution.
Dilute the RNA solution to 250 uM and store it at -20°C.

4.6.4 Aminoacylation of tRNA by flexizymes and determination of acylation efficiency

Mix the solutions as follows.

H2O	3 uL
500 mM HEPES-KOH (pH 7.5) 1 uL	
200 uM Flexizyme	1 uL
200 uM tRNA	1 uL

Heat the resulting mixture at 95°C for 2 min, then stand it at room temperature for 5 min to facilitate RNA folding (mainly folding of tRNA).
Add 2 uL of 3M MgCl2 and mix well.
Put the tube onto ice, and add 2 uL of 25 mM amino acid substrate (DMSO solution). Mix the reaction mixture well.
Incubate the reaction mixture on ice for 1-72 hours (The proper reaction time varies depending on the amino acids).
2.EtOH precipitation of aminoacyl-tRNA
After reaction, add 40 uL of 0.3M AcONa and 100 uL of EtOH, then mix well.
Centrifuge at 25°C at 13000 rpm for 15 min. Discard the supernatant.
Add 50 uL of 70% EtOH, then centrifuge for 2 min. Discard the supernatant.
Open the lid and cover the tube with clean paper towel, and stand for 5-10min to dry up the pellet.
Keep the tube including RNA pellet on ice.
3. PAGE purification of aminoacyl-tRNA as described for felxizymes.

5 Conclusions

1. Previously reported procedures for VEGF *E.coli* expression were scaled-up without problems to yield both unlabelled and isotopically labelled protein. Isotopically labelled protein was required for protein-based NMR experiments whereas the use of unlabeled VEGF was highly advantageous for mass spectrometry experiments. Currently the only bottleneck in large scale production is the price for the required enzyme to remove the His-tag.

2. Prediction of fragment water solubility is an unsolved problem and has negative impact on fragment library design and setup.

3. A method has been developed for the automatic mixture generation for NMR-based screening that allows complete assessment of all fragments without mixture deconvolution.

4. A method has been developed that allows the automatic evaluation of NMR-based screening data from STD and CPMG experiments in a quantitative manner, improving both speed and quality of the analysis.

5. Screening of our proprietary fragment library resulted in a high hit rate but yielded very low affinity compounds. This was the first indication that the protein-protein interface of VEGF is hardly druggable. In this scenario, the automatic quantitative analysis of screening data proved to be a valuable tool.

6. Screening of a ^{19}F containing compound library by NMR resulted in a lower hit rate compared to our fragment library but yielded compounds of similar weak affinity. These results were a second manifestation of the poor druggability of VEGF.

7. We have established a ^{19}F-NMR based competition assay ground on the use of a peptide spy molecule. The absence of clear competitors was a further indication of the low druggability of VEGF.

8. Rigid docking has proved to be not adequate to predict fragment binding to protein-protein interfaces, at least in the case of VEGF.

9. Both Sitemap (Schrödinger) and Sitefinder (MOE) did not recognize the protein-protein interface of VEGF as druggable.

10. The use of protein-based NMR methods to characterize the interaction of VEGF with low affinity ligands (>10mM) has proven to be very difficult. Working at reduced temperatures and protein concentrations was required to overcome this problem.

11. A first assessment of compounds by [methyl-^{13}C]-methionine CSP significantly increased throughput and allowed the estimation of affinities.

12. The combination of PELE strongly correlated with NMR derived binding parameters appears to be a promising approach to evolve very weak ligands even in the context of protein-protein interfaces.

13. Fragment-like compounds from at least three scaffold families that bind with high uM to low mM affinities to VEGF have been identified. Some of these compounds bind to the protein-protein interface of VEGF and are promising candidates for further development.

14. The application of mRNA display led to the preliminary discovery of peptide-based VEGF ligands but confirmation of binding by NMR and FP is still pending.

6 Resumen en español

6.1 Introducción

6.1.1 Interfaces proteína-proteína como objetivos terapéuticos

La capacidad de las proteínas para actuar con otras proteínas radica en el centro de la biología. La presencia de estas interacciones ha atraído la atención de científicos para el desarrollo de inhibidores o instrumentos bioquímicos. Este interés creciente ocurre en un tiempo en que la sociedad basada en el descubrimiento de fármacos está en una crisis. A pesar de los avances en la tecnología, la ciencia básica y el aumento de los gastos en investigación, la introducción de fármacos ha bajado o en el mejor de los casos se ha mantenido constante mientras la proporción de fármacos que fracasan, sobre todo en las últimas etapas de los ensayos clínicos ha aumentado.[1] Por consiguiente el riesgo y coste para desarrollar un nuevo fármaco se ha visto incrementado. La razón de la disminución en la productividad está sometida a debate. Algunos científicos han propuesto que el motivo principal sea la falta de nuevos objetivos sencillos, conocidos también como "low hanging fruits", haciendo que sea un reto más difícil para los científicos.

El diseño de fármacos y la modulación de las interfaces proteína-proteína abren en este contexto nuevas dianas terapéuticas. Venkatesan y colaboradores estiman alrededor de 130,000 interacciones binarias entre proteínas presentes en el interactoma humano.[2] Incluso si sólo una fracción de estos nuevos objetivos pudiera ser manipulada de forma terapéutica esto añadiría un número considerable a los aproximadamente 2,000 a 3,000 dianas terapéuticas clásicas.[3]

Pero las interfaces proteína-proteína son diferentes a las clásicas proteínas "druggable". Estas proteínas "druggable" vienen de distintas familias, pero han sido desarrolladas para actuar recíprocamente con pequeños metabolitos moleculares. Sus zonas de unión pueden ser huecos sobre la superficie de la proteína o cavidades profundas dentro de la misma. En ambos casos las interfaces son relativamente pequeñas, entre 300 y 1000 Å.

Las interfaces proteína-proteína han sido desarrollados para unir-se a su correspondiente proteína asociada. A diferencia de las proteínas "druggable" sus interfaces son generalmente planas y presentan áreas de entre 1500 a 3000 Å. A menudo éstas tienen flexibilidad intrínseca para unir-se recíprocamente con múltiples proteínas asociadas.

Un buen ejemplo de ello es VEGF, la proteína de esta tesis (Figura 1.2). Esta protína actúa con dos receptores extracelulares. Ambos comparten una misma interfaz en la superficie de VEGF que indica alguna flexibilidad.[4,5] La interfaz es relativamente plana, no muestra ningún hueco ni cavidades, en el caso del complejo con VEGFR2, presenta una área de alrededor de 2000 Å.
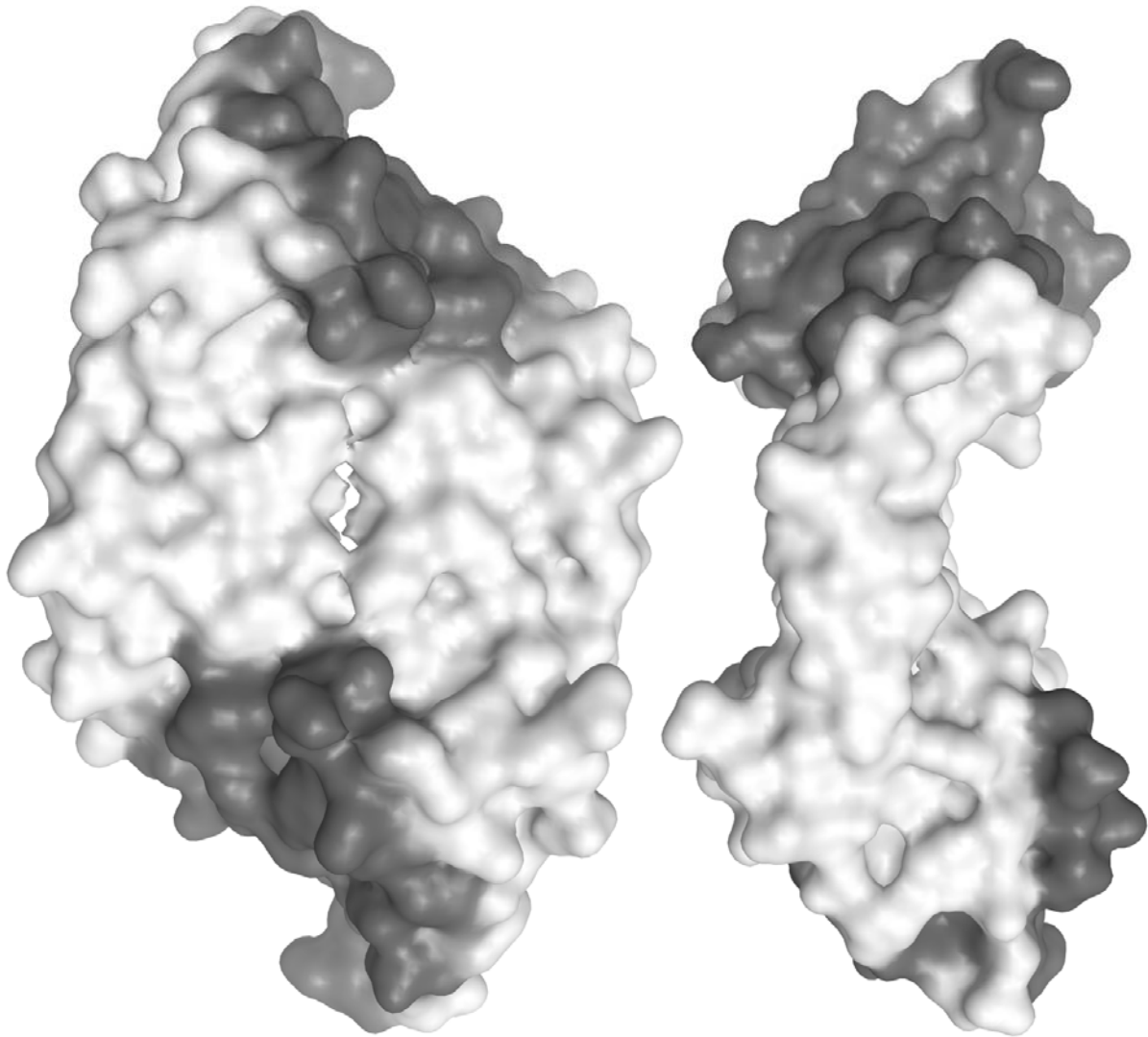
Figura 1.2: Representación superficial del homodimero VEGF de 3V2A.La interfaz involucrada en la interacción con el receptor KDR/VEGFR-2 esta representada en gris.

Las interfaces proteína-proteína se diferencian claramente de las pequeñas interfaces de las moléculas. La pregunta que surge es como desarrollar compuestos que pueden unirse a estas interfaces. El empleo terapéutico de compuestos biológicos que se unen a las interfaces proteína-proteína está relativamente avanzado. La lista de estos compuestos biológicos aprobados por la FDA incluye, Humira, Enbrel, Remicade y Avastin. El último fue desarrollado para unirse a VEGF. Toda estos fármacos están incluidos entre las medicinas de mayor venta en EEUU, y suponen varios miles de millones de ingresos.[6]

Pero estos compuestos biológicos tiene algunas desventajas inherentes a pesar de su gran éxito terapéutico y económico: 1) Están basados en biopolímeros grandes, que el sistema inmunológico humano puede reconocer. 2) Su inhabilidad para cruzar barreras biológicas de un modo eficiente plantea una limitación para su uso general. 3) Finalmente, éstos son muy caros, porque no pueden

ser sintetizados como pequeños fármacos moleculares, por lo que deben ser producidos mediante expresión recombinante. Aunque la gran mayoría de fármacos sean pequeñas moléculas [7], 9 de los 10 fármacos más caros en 2012 son compuestos biológicos, con un coste anual del tratamiento de entre 200,000 a 400,000 dólares.[8] Esto ha llevado a cuestionarse si la sociedad puede permitirse este tipo de fármacos.

Potencialmente, los fármacos pequeños pueden presentar una producción más barata, no tienen el riesgo de provocar inmunogenicidad y pueden ser diseñados para cruzar barreras biológicas. Además son bastante estables y pueden ser administrados por vía oral. Lamentablemente existe un debate abierto sobre como las interfaces proteína-proteína pueden ser moduladas por pequeñas moléculas. Las propiedades biofísicas de las interfaces proteína-proteína parecen ser contradictorias con respecto a los índices clásicos de "druggability" y muchas tentativas para desarrollar ligandos han fallado. Además, las interfaces proteína-proteína no poseen ningún metabolito asociado a su interacción que pudiera servir como punto de partida para el descubrimiento de fármacos. Publicaciones recientes describen la exitosa obtención de pequeños compuestos que se unen al menos a algunos tipos de interfaces proteína-proteína. La mayoría de estos compuestos descritos se unen a interfaces proteína-proteína donde uno de los componentes es un péptido. Esto sugeriría que este tipo de interfaces sean más "druggable" que las interfaces entre componentes globulares.

Los pequeños inhibidores de las interacciones proteína-proteína descritos se diferencian de sus respectivos que se unen a las dianas terapéuticas clásicas. Estos inhibidores tienen generalmente un peso molecular más alto, mayor lipofilicidad, más sistemas cíclicos, pero menos enlaces rotatorios y menos aceptores y donadores de puentes de hidrógeno. Interesantemente en un contexto de descubrimiento clásico de fármacos la mayoría de estos inhibidores proteína-proteína sería considerado como potencialmente no específico y problemático en cuanto a propiedades ADMET[9].

Aunque muchas tecnologías han sido usadas para identificar inhibidores de interfaces proteína-proteína, el descubrimiento de fármacos basado en fragmentos podría ser una herramienta clave.[10] 1) Anteriormente, esto ya ha funcionado bien con interfaces del tipo péptido-proteína. 2) Además esto puede dar lugar al desarrollo de inhibidores de interfaces proteína-proteína con propiedades fisicoquímicas mejoradas. 3) Esto permite una exploración muy eficiente del espacio químico. Estos argumentos sugieren estudiar la capacidad de descubrimiento de fármacos basado en fragmentos cuando se aplicada a la interfaz proteína-proteína, sobre todo en el caso de interfaces basadas en componentes globulares, como VEGF y sus receptores.

(1)   Pammolli, F.; Magazzini, L.; Riccaboni, M. Nat Rev Drug Discov 2011, 10, 428.
(2)   Venkatesan, K.; Rual, J. F.; Vazquez, A.; Xin, X.; Goh, K. I.; Yildirim, M. A.; Simonis, N.; Heinzmann, K.; Gebreab, F.; Sahalie, J. M.; Cevik, S.; Barabasi, A. L.; Vidal, M. Nat Methods 2009, 6, 83.
(3)   Russ, A. P.; Lampel, S. Drug Discov Today 2005, 10, 1607.
(4)   Brozzo, M. S.; Bjelic, S.; Kisko, K.; Schleier, T.; Leppanen, V. M.; Alitalo, K.; Winkler, F. K.; Ballmer-Hofer, K. Blood 2012, 119, 1781.
(5)   Wiesmann, C.; Fuh, G.; Christinger, H. W.; Eigenbrot, C.; Wells, J. A.; de Vos, A. M. Cell 1997, 91, 695.
(6)   http://www.drugs.com/stats/top100/sales 2012, *Consultado el 12.02.2013*
(7)   Bank, D. http://www.drugbank.ca/databases 2012, *Consultado el 12.02.2013*
(8)   coding, m. b. http://www.medicalbillingandcoding.org/blog/the-11-most-expensive-medicines-in-america/ 2012, *Consultado el 12.02.2013*
(9)   Higueruelo, A. P.; Schreyer, A.; Bickerton, G. R.; Pitt, W. R.; Groom, C. R.; Blundell, T. L. Chem Biol Drug Des 2009, 74, 457.
(10)   Wells, J. A.; McClendon, C. L. Nature 2007, 450, 1001.

6.1.2 Descubrimiento de fármacos basado en fragmentos

El descubrimiento de fármacos basado en fragmentos se describe como la búsqueda iniciada a partir de pequeños fragmentos químicos. Entendemos por fragmento aquellos compuestos orgánicos con pesos moleculares entre 150 y 300 Da. El éxito del descubrimiento de fármacos basado en fragmentos está fundado en dos causas razonadas. 1) El espacio químico del ligando puede ser explorado mucho mejor cuanto más pequeño es el umbral del peso molecular máximo, que conduce a proporciones de "hits" más elevadas comparado con HTS. 2) Aunque los "hits" descubiertos mediante este método sean de afinidad débil, éstos deberían formar interacciones de alta calidad con la proteína usada como "target" las cuales pueden ser detectadas. Esto los hace buenos puntos de partida para el desarrollo de fármacos.

El descubrimiento de fármacos basado en fragmentos también presenta desventajas. Las afinidades bajas requieren técnicas biofísicas bastante sensibles para detectar este tipo de uniones débiles. El descubrimiento de fármacos basado en fragmentos se puede dividir en tres etapas claves: 1) El diseño de la quimioteca. 2) El cribado de fragmentos y su validación. 3) La optimización iterativa de fragmentos. La RMN ha sido un método omnipresente en el campo del descubrimiento de fármacos basado en fragmentos desde el trabajo del pionero Fesik a finales de la década de los 90.[1,2] Esto ofrece un gran variedad de experimentos de RMN especializados que pueden ser usados en las diferentes etapas del proceso de descubrimiento de fármacos.

(1)   P. J. Hajduk, G. S., D. G. Nettesheim , E. T. Olejniczak , S. B. Shuker , R. P. Meadows , D. H. Steinman , G. M. Carrera , Jr., P. A. Marcotte , J. Severin , K. Walter , H. Smith , E. Gubbins , R. Simmer , T. F. Holzman , D. W. Morgan , S. K. Davidsen , J. B. Summers, S. W. Fesik J. Am. Chem. Soc. 1997, 119, 5818.
(2)   Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Science 1996, 274, 1531.

6.1.3 Factor de crecimiento endotelial vascular en enfermedad y tratamiento

VEGF es la proteína pro-angiogénica más potente. Aunque la angiogénesis es un proceso normal y vital en el organismo, se han descrito varias condiciones patológicas asociados con el sobreexpresión de VEGF. La asociación de VEGF con el cáncer es uno de los campos más avanzados en términos de ciencia básica y de desarrollo. El crecimiento de los tumores está limitado por su suministro de substancias nutritivas y oxígeno, distribuido por vasos sanguíneos. Así los tumores que carecen de vásculos no pueden crecer más de 3mm de diámetro. Sin embargo los tumores puede vencer esta limitación gracias de la sobreexpresión de VEGF, que conduce a la generación de vasos sanguíneos que abastecen al tumor [1] y permite a las células del tumor extenderse a otras partes del cuerpo mediante metástasis**.**

Una segunda condición patológica e importante es la degeneración macular causada por angiogenésis extensa, que es la causa principal de daño de visión en la gente de avanzada edad en el mundo desarrollado. En esta enfermedad la sobreexpresión de VEGF conduce al crecimiento de vasos sanguíneos que dañan la retina. Múltiples estrategias han sido exploradas para interrumpir la cascada de eventos celulares que desencadena VEGF (Figura 1.5.3).
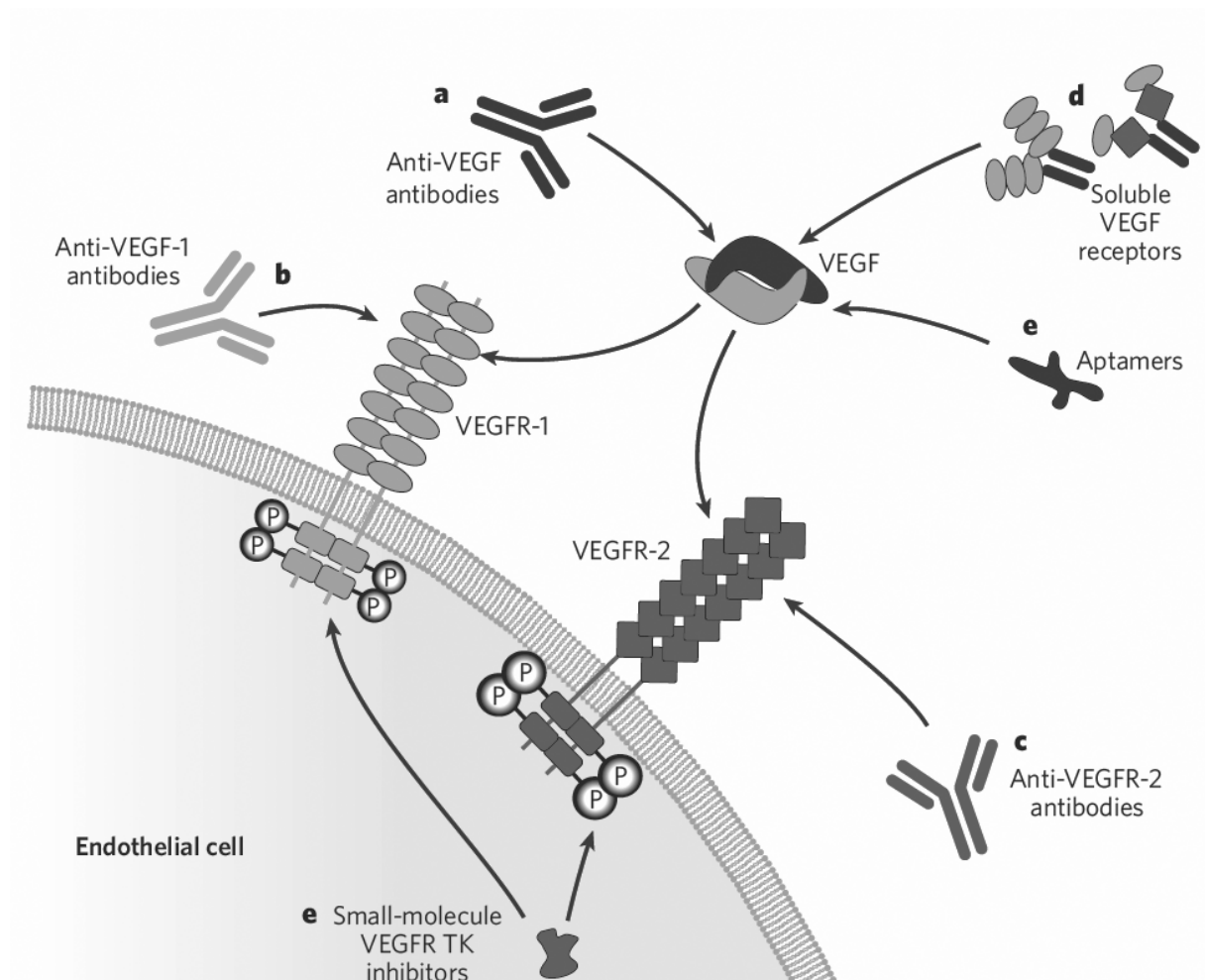
Figura 1.5.3: Estrategias de interrumpir angiogenésis patológico que desencadena de VEGF.[2]

Estas estrategias se dividen en dos clases:

1.) Estrategias intracelulares basados moléculas pequeños que inhiben el dominio quinase de los receptores de VEGF. En la actualidad hay tres fármacos de segunda generación aprobados por la FDA, con actividad contra el receptor de VEGF. Aunque estos inhibidores han demostrado actividad contra el tumor, debido a su baja especificad, provocan una variedad de efectos indeseados.[3]

2.) Estrategias extracelulares basados en compuestos biológicos que se unen a VEGF o las partes extracelulares de sus receptores. Estos compuestos, aprobado por la FDA para el tratamiento de algunos tipos específicos de cáncer, han sido un éxito económico y suponen varios miles de millones de ingresos.[4]

Sin embargo, algunas de estos tratamientos, cómo Avastin, son muy polémicos. Algunos estudios no han probado una ventaja en supervivencia de población tratada.[5,6] También ha sido criticado por su alto coste de tratamiento, que puede alcanzar 100,000 dólares al año. Por esto, algunos seguros y sistemas de asistencia médica nacionales, restringieron el financiamiento de estos tratamientos, que sólo pueden prolongar la vida pero no curar el cáncer.[7]

Una estrategia que hasta ahora no ha sido aplicada en el tratamiento de angiogénesis patológico podría ser el uso de fármacos pequeños que inhiban directamente las interacciones proteína-proteína entre VEGF y sus receptores. Si se pudiera vencer el problema de desarrollar estos inhibidores, podrían constituir un tratamiento de buena especificidad con alta permeabilidad de membranas y bajos costes de producción. Hasta ahora, en la literatura se han descrito pocos trabajos de desarrollo de ligandos de naturaleza no basada en anticuerpos, que pueden unirse a la interfaz de proteína-proteína de VEGF.

Con muy pocas excepciones[8,9] estos ligandos están basados en péptidos y no en moléculas pequeñas.[8,9]

(1)   Hoeben, A.; Landuyt, B.; Highley, M. S.; Wildiers, H.; Van Oosterom, A. T.; De Bruijn, E. A. *Pharmacol Rev* **2004**, *56*, 549.
(2)   Ferrara, N.; Kerbel, R. S. *Nature* **2005**, *438*, 967.
(3)   Bhargava, P.; Robinson, M. O. *Curr Oncol Rep* **2011**, *13*, 103.
(4)   *http://www.fiercepharma.com/special-reports/avastin*, *Consultado el 12.02.2013*
(5)   Hayes, D. F. *JAMA* **2011**, *305*, 506.
(6)   Jain, R. K.; Duda, D. G.; Clark, J. W.; Loeffler, J. S. *Nat Clin Pract Oncol* **2006**, *3*, 24.
(7)   Briggs, H. B. N. H. R. **2010**.
(8)   Fairbrother, W. J.; Christinger, H. W.; Cochran, A. G.; Fuh, G.; Keenan, C. J.; Quan, C.; Shriver, S. K.; Tom, J. Y.; Wells, J. A.; Cunningham, B. C. *Biochemistry* **1998**, *37*, 17754.
(9)   Pan, B.; Li, B.; Russell, S. J.; Tom, J. Y.; Cochran, A. G.; Fairbrother, W. J. *J Mol Biol* **2002**, *316*, 769.

6.2 Objetivos

En el contexto de la presente tesis hemos abordado los siguientes objetivos:

1. El uso de métodos de RMN, basados tanto en la observación del ligando como en la observación de proteína, para estudiar la unión de los compuestos de una quimioteca a la zona de VEGF involucrada en la unión a sus receptores. La interacción VEGF/VEGFR puede ser considerada como un caso de estudio para la evaluación de las interfaces proteína-proteína mediante cribado de fragmentos.

2. Desarrollar herramientas basadas en la combinación de RMN y métodos computacionales para abordar: i) un sistema automático de diseño de mezclas de fragmentos; ii) el análisis automático de datos procedentes de cribados basados en RMN; iii) la evolución de fragmentos con muy baja afinidad.

3. Explorar el uso de técnicas de "mRNA display" para el descubrimiento de nuevos ligandos peptídicos para VEGF.

6.3 Resultados

Los resultados de esta tesis se pueden dividir en cuatro secciones. 6.3.1 desarrollo de métodos y resultados de pre-cribado, 6.3.2 resultados de cribado, 6.3.3 resultados post- cribado, 6.3.4 mRNA display. Los resultados para cada sección se resumen, brevemente, a continuación.

6.3.1 Desarrollo de método y resultados pre-cribado

6.3.1.1 Estudios de interacciones ligando-proteína por RMN

La primera sección proporciona una descripción general del estudio de interacciones ligando-proteína por RMN. La sección está dividida según dos categorías principales de experimentos: aquellos basados en la observación de las señales de la proteína y aquellos basados en la observación de las señales del ligando. La siguiente sección explica dos estudios experimentales concretos. El primero de estos trata la interacción entre el factor de crecimiento endotelial vascular (VEGF) y un ligando peptídico, e incluye un protocolo detallado de experimentos de perturbación de desplazamientos químicos (CSP). En el segundo se estudian la interacción entre la enzima prololigopeptidasa (POP) y una pequeña molécula utilizando el método de diferencia de transferencia de saturación de ligando (STD) e ilustra la utilidad de la RMN para confirmar interacciones e identificar el epítopo de unión de un ligando.

6.3.1.2 Producción y caracterización de VEGF recombinante

En este capítulo se describe la expresión y purificación de VEGF. La proteína primero fue expresada con marcaje en todas las posiciones ^{15}N y metil-^{13}C-metionina para los estudios de RMN. Luego se expresó la proteína sin marcar para estudios de MS y cristalografía de rayos X. Todo el proceso para la expresión de la proteína fue mejorado considerablemente, respecto la llevada a cabo en incubador, usando un fermentador, que condujo a un rendimiento 35 veces superior. Es probable que esto pueda ser debido a una mejor aireación del fermentador, que permite tanto un crecimiento bacteriano más rápido como un aumento en la densidad final del cultivo. El rendimiento de la expresión fue tan alto que tanto la proteína marcada como la no marcada solo tuvieron que ser expresadas una sola vez durante esta tesis. Esto permitió enfocar aspectos más desafiantes. El empleo del fermentador cuesta sólo 40 euros por día y además permite rendimientos mucho más altos de proteína por litro de medio de partida. Esto era importante en el caso de la expresión con marcaje ya que tanto el ^{15}NH$_4$Cl como la ^{13}C metil metionina son caros. A partir de este paso la limitación en la producción de VEGF es la necesidad de usar la enzima Genenasa I para el lugar de escisión proteolítica del His-Tag. Esta enzima se debe pedir a EEUU y cuesta aproximadamente 15 euros por miligramo de VEGF, un gasto considerable teniendo en cuenta que más de 200 mg de VEGF

fueron producidos durante esta tesis. La integridad estructural y propiedades espectroscópicas de la VEGF producida fueron evaluadas usando el experimento de RMN ^{1}H-^{15}N HSQC. La proteína mostró una perfecta consistencia con los espectros publicados, indicando la integridad de pliegue y la ausencia de estructuras oligoméricas no nativas. Las características espectroscópicas de la proteína facilitaron el estudio con RMN. Fue posible llevar a cabo el experimento de RMN ^{1}H-^{15}N HSQC registrando 1 hora a 45°C con una concentración de proteína de 100 uM.
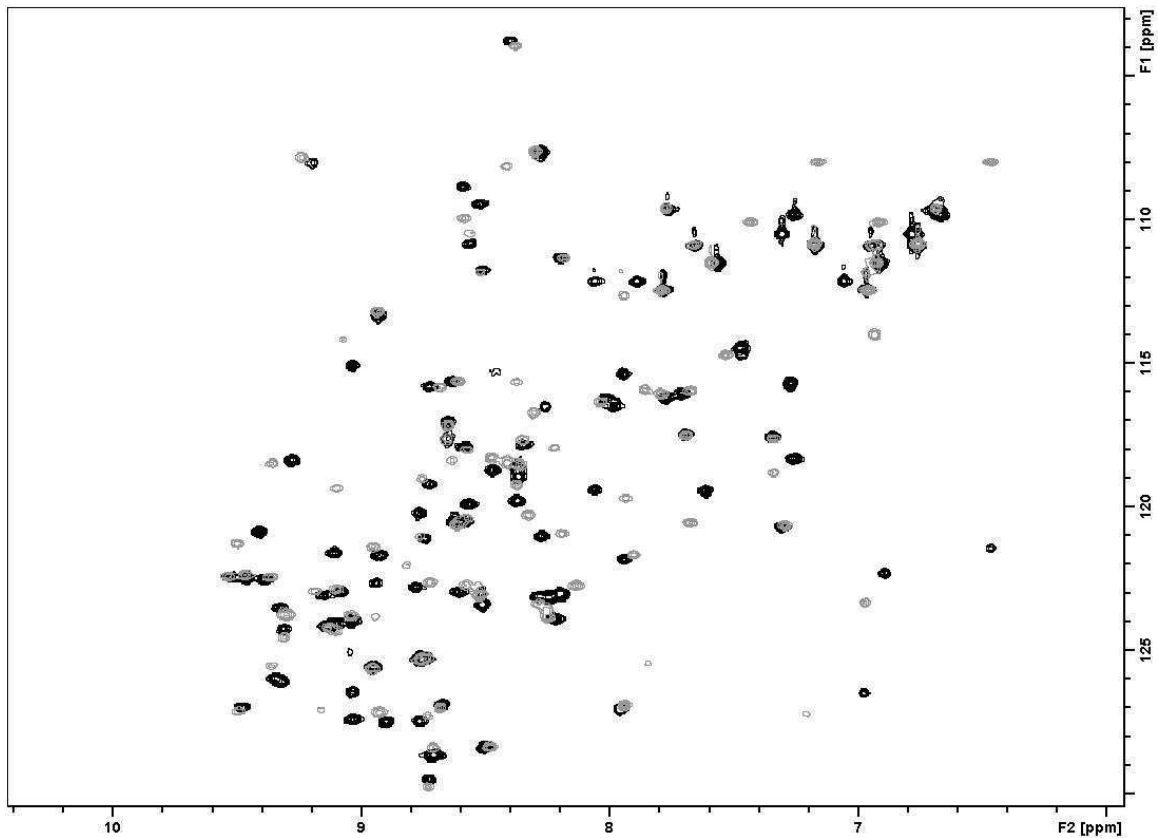
Figura 3.2-2: Espectro de ^{15}N-^{1}H HSQC de una muestra de 100 uM de [metilo-13C] metionina – ^{15}N VEGF11-109 a 45°C antes (negro) y después de la adición de tres equivalentes de v107 (gris)

La calidad de la proteína también fue evaluada por espectrometría de masas ESI-time-of-flight.
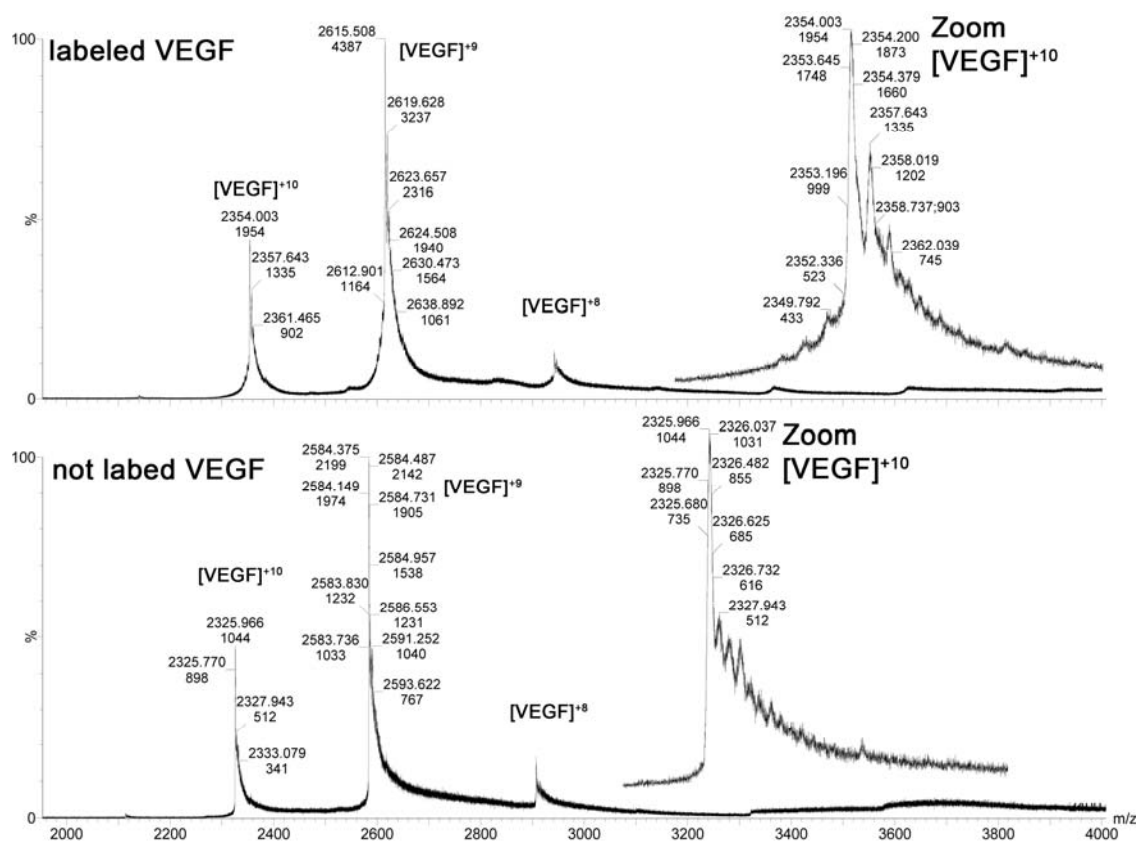
Figura 3.2-3: Caracterización de [metilo-13C] metionina – 15N VEGF11-109 por espectrometría de masas ESI-time-of-flight

6.3.1.3 Diseño y preparación de una biblioteca de fragmentos

El laboratorio del Prof. Dr. Giralt no tenía ninguna experiencia anterior en el descubrimiento de fármacos basado en el cribado de fragmentos. Así, uno de los primeros objetivos para este proyecto era diseñar y preparar una biblioteca basada en fragmentos, que posteriormente podría ser evaluada. Muchos grupos académicos no diseñan sus propias bibliotecas de fragmentos, sino que seleccionan una entre las que ofrecen las distintas empresas comerciales como Maybdridge, Asinex y Enamine, etc. Estas bibliotecas son más económicas, pero probablemente el espacio químico explorado es más limitado. El diseño de nuestra biblioteca se basó en la experiencia del Prof. Xavier Barril, un ex-empleado de Vernalis. En resumen, se montó una base de datos con más de un millón de compuestos químicos únicos de proveedores internacionales. De estos, se eliminaron aquellos que tenían funcionalidades indeseadas (por ejemplo, grupos reactivos, azúcares) y los que no cumplían las exigencias para ser posibles fármacos (por ejemplo, que al menos tengan un sistema cíclico y también como mínimo un grupo funcional deseado).

Los compuestos restantes con un peso molecular entre 110 y 305 Da fueron agrupados según su semejanza y se seleccionó un compuesto de los grupos más poblados. El diseño de este tipo de biblioteca es conocido como " SAR por catálogo " y tiene las ventajas siguientes:

Cada compuesto de la biblioteca representa un grupo de compuestos similares. Por lo tanto la biblioteca cubre un espacio químico muy grande. Además la obtención de *leads* a partir de *hits* es sencilla: para cada fragmento evaluado con buena afinidad (ligando) para la proteína, se pueden comprar compuestos similares a los proveedores anteriormente comentados. La preparación de la biblioteca fue llevada a cabo en dos pasos, con el proceso laboral idéntico descrito en la figura 3.3.1.
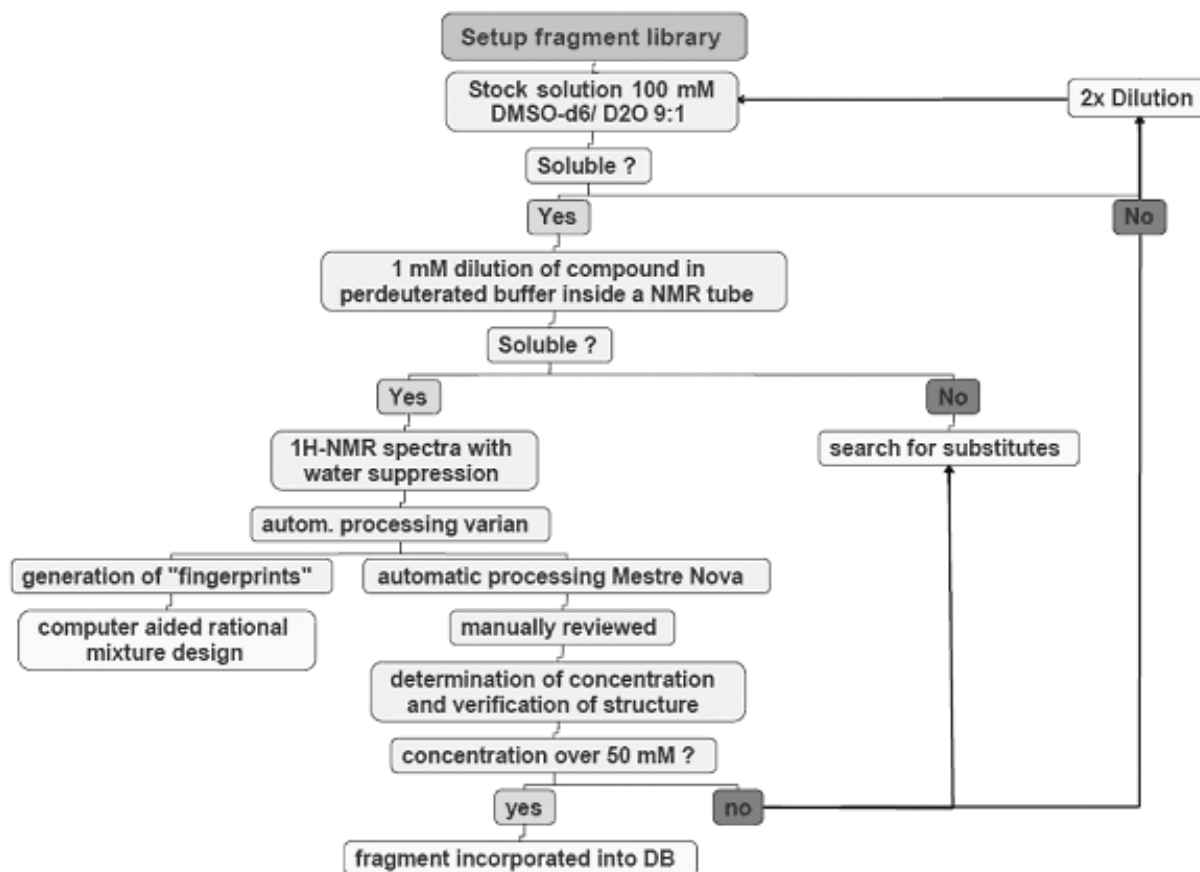

Figura 3.3.1: Diagrama de trabajo para la preparación de la biblioteca de fragmentos.

Una experiencia asombrosa durante el diseño y preparación de la biblioteca fue el alto número de compuestos que resultaron ser insolubles en las disoluciones stock o en las condiciones del cribado. Este fue el caso para 174 de los 701 fragmentos comprados, con un coste aproximado de 4.300 euros. Este problema apareció aunque se hicieron cálculos previos para predecir la solubilidad de todos los compuestos. En base a esto, vimos que el poder de predicción computacional no es muy alto. Durante la preparación de la biblioteca se invirtió mucho esfuerzo en el control de calidad, como en pruebas de solubilidad, identidad y pureza de las alícuotas de los compuestos individuales, y la solubilidad de las mezclas de compuestos en las condiciones de cribado. Por consiguiente no afrontamos ningún problema de solubilidad o agregación evidente u otras complicaciones durante el cribado.

6.3.1.4 Diseño asistido por ordenador de mezclas de fragmentos para el cribado en RMN

La RMN es una técnica relativamente lenta y requiere el cribado de mezclas de compuestos para aumentar la velocidad de cribado total. Para completar nuestra biblioteca tuvimos que preparar estas mezclas. Desarrollamos para este propósito un método novedoso, rápido, sencillo y automatizado para diseñar las mezclas de fragmentos de nuestra biblioteca con un solapamiento entre las señales de RMN mínimo. Nuestra aproximación implica la codificación de los espectros de RMN en un formato legible para el ordenador que llamamos huellas digitales. Entre varios algoritmos que evaluamos para solucionar el problema del solapamiento de señales, el algoritmo SA ofreció la mejor optimización. Como prueba de concepto, usamos este algoritmo para diseñar mezclas de cinco fragmentos de nuestra biblioteca, obteniendo un solapamiento de señales promedio de sólo el 2 %, comparado con el 44 % para las mezclas del mismo tamaño diseñadas al azar. Entonces generamos bibliotecas de fragmento virtuales para evaluar el funcionamiento del algoritmo SA según la distribución de picos, el escalado y la temperatura. Nuestros resultados sugieren que el método es apto para bibliotecas de cualquier tamaño o diversidad química. Esperamos que el método descrito mejore la eficacia del cribado por RMN basado en fragmentos, simplificando la detección de ligandos, puesto que no requiere hardware sofisticado y, en el caso de compuestos cuyos espectros RMN ya estén disponibles, no requiere ningún trabajo experimental adicional. La figura 3.4.1 muestra un ejemplo de una mezcla que contiene cinco fragmentos sin solapamiento de señales.
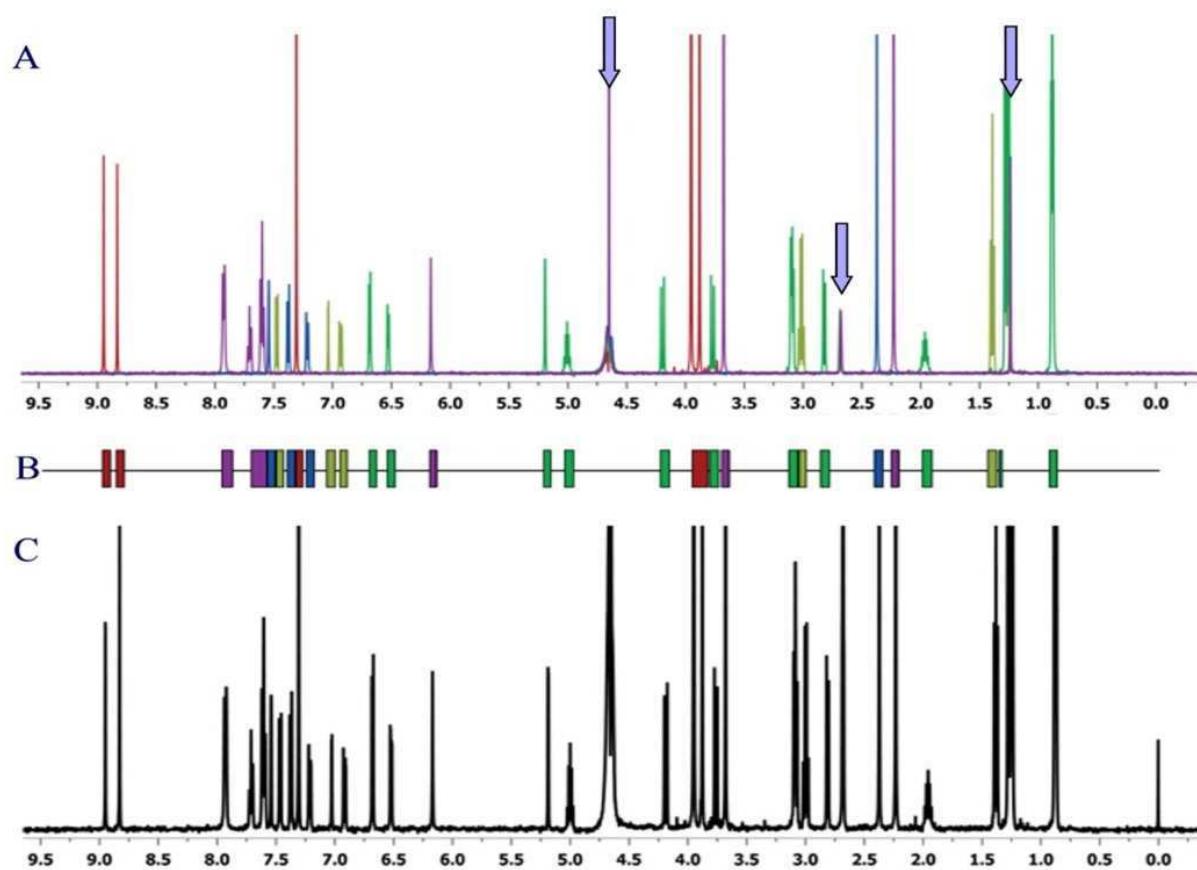
Figura 3.4.1: A: Superimposición de los espectros de ^{1}H-RMN de cinco fragmentos diferentes (1 mM de cada compuesto, 50 mM tampón fosfato, pH 7.0, 50 mM NaCl, 3 % DMSO-d6), registrado a 37ºC y 500 MHz. Las flechas indican los picos residuales de H2O, DMSO y t-butanol (el estándar interno). B: La huella digital de una mezcla diseñada in silico con solapamiento de señal cero o cerca de cero. C: Espectro de ^{1}H-RMN de los cinco fragmentos mezclados juntos (500 uM cada uno) en condiciones experimentales idénticas a 1A (la señal en 0 ppm corresponde a DSS).

6.3.1.5 Evaluación automatizada de los datos del cribado por RMN

El método descrito en el capítulo anterior permitió la generación de mezclas con solapamiento de señales muy bajo. Confirmamos que esto simplificó la detección y la evaluación de ligandos durante el cribado de nuestra biblioteca. Aún así, encontramos que incluso sin solapamiento de señales el análisis de los datos de cribado por RMN es un proceso bastante tedioso y largo. En este capítulo exploramos, por lo tanto, la viabilidad de evaluación automática de los datos de cribado por RMN. Afrontamos tres problemas clave: 1. La transformación de los espectros de RMN en un formato legible por ordenador. 2. La asignación de cada señal de la muestra al compuesto correspondiente. 3. La evaluación computacional de los datos de las interacciones proteína-ligandos obtenidos por RMN. La primera cuestión fue solucionada con el empleo de un *script* para MestreNova que permite la generación de archivos que contienen los datos de las huellas digitales de manera semiautomática. La calidad de la digitalización de la información era suficiente para las señales de STD con intensidad baja. La segunda cuestión fue simplificada por el bajo grado de solapamiento de señales en nuestras mezclas y pequeñas diferencias entre los desplazamientos químicos de los compuestos de referencia y los de las mezclas. Esto facilitó mucho la asignación computacional de las señales de cada mezcla. En este proceso se usaron condiciones normalizadas (misma temperatura, pH, tampón, etc.), de no haber sido así, habría sido necesario recurrir a algoritmos más complicados. En cuanto al tercero, al estar toda la información necesaria para evaluar la unión de los ligandos contenida en las huellas digitales, sólo fue necesario definir de qué modo los datos se analizarían para permitir una evaluación cuantitativa de cada fragmento. Basado en la naturaleza del experimento STD, como estimación para la afinidad propusimos las intensidades relativas de las señales de STD y como indicador, tanto para la afinidad como para la confianza en la unión, la fracción de señales observadas en el espectro de STD respecto el total teórico (figura 3.4.2-3). También se realizó la evaluación automática de los datos de CPMG aunque la correlación con los resultados de STD no fuera muy alta (figura 3.4.2-4).
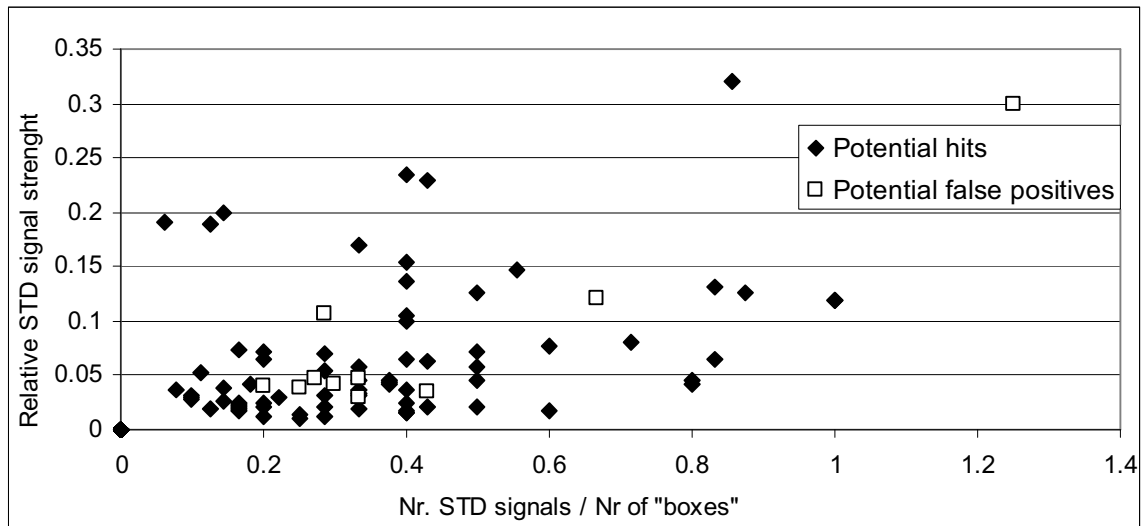
Figura 3.4.2-3: Resultados de la evaluación automática de los datos de cribado por STD RMN de 125 fragmentos.
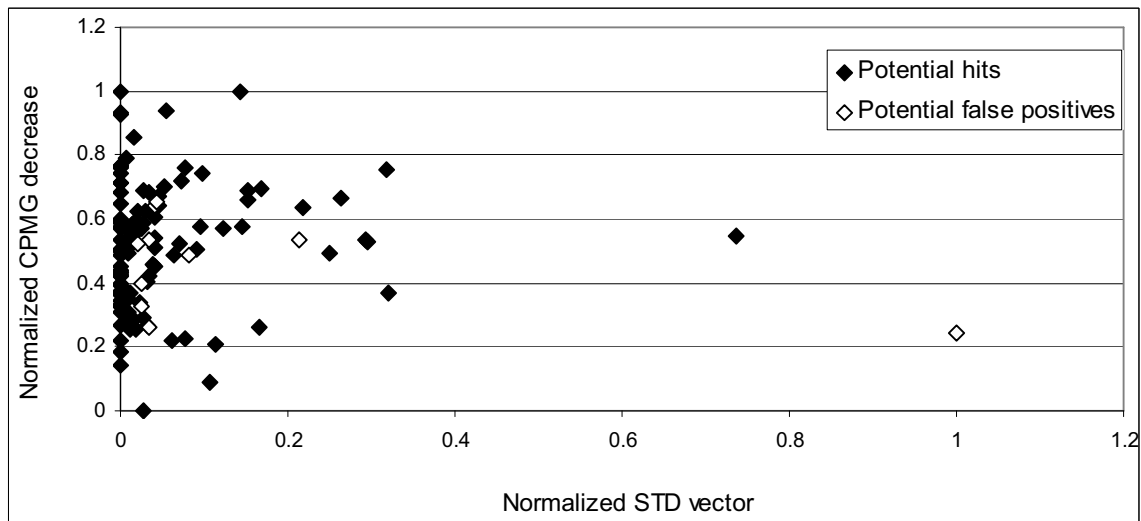

Figura 3.4.2-4: Los resultados de la validación de los datos cruzados de STD y CPMG de 125 fragmentos.

Basándonos en nuestros resultados, la evaluación automática, al menos con mezclas con solapamiento de señales optimizado, parece ser sumamente factible. Creemos que esto va a facilitar el análisis de datos del cribado de fragmentos por RMN, porqué esto aumenta la velocidad y la calidad del análisis comparando con un análisis manual. Sin embargo, la capacidad y las limitaciones de nuestro análisis automático requieren más pruebas. Por esta razón, sería útil tener datos sólidos de RMN adquiridos del cribado de fragmentos con dianas típicamente *druggable*.

6.3.2 Cribado

En esta sección usemos varios métodos para explorar la capacidad del cribado basado en fragmentos para identificar ligandos potenciales para VEGF:

1. En total 527 fragmentos de nuestra biblioteca fueron cribados mediante STD y algunos también con CPMG. Observamos altas proporciones de *hits* en los cribados, entre el 15 y el 27 %. La afinidad de los *hits* probablemente era baja, puesto que en el experimento STD había una relación señal-ruido baja y en el de CPMG una baja disminución de señal en presencia de proteína.

2. Desarrollamos una molécula espía basada en el péptido v107, que contiene un átomo ^{19}F, permitiendo la detección de ligandos unidos a la interfaz proteína-proteína de VEGF con el receptor (VEGFR). El cribado competitivo de nuestros *hits* con ^{19}F-RMN no identificó ligandos que se unieran con fuerza en la interfaz proteína-proteína de VEGF. Entonces, los más de 100 fragmentos probados deben unirse a otras superficies de la proteína o con una afinidad por debajo de nuestro umbral de detección.

3. Se obtuvieron a proporciones de *hits* de alrededor 1-3 % en el cribado de la biblioteca del CNIO mediante ^{19}F-RMN, pero en afinidades comparables a las obtenidas con los ligandos de nuestra biblioteca. Estos resultados se basaron en la pequeña influencia de la concentración de proteína sobre las intensidades de señal de ^{19}F.

4. Dos técnicas computacionales, para la predicción del lugar de unión y la predicción de la *druggability,* no reconocieron la interfaz proteína-proteína de la VEGF como posible diana.

5. El cribado virtual mostró una correlación baja con los resultados experimentales del cribado mediante RMN. El rango de valores de *docking* obtenidos para los fragmentos más prometedores fue entre -3 y -4, indicando muy bajas afinidades.

En general, la mayoría de los resultados indican que VEGF no es una proteína con muchas posibilidades para ser modulada por fármacos. La única excepción es la elevada proporción de *hits* obtenidos del cribado por STD ya que la proporción de *hits* obtenidos por RMN han sido propuestos como indicador experimental de la facilidad de las proteínas como diana de fármacos. Para explicar estos resultados y ponerlos en el contexto propusimos la hipótesis siguiente: las proteínas susceptibles de ser moduladas por fármacos tienen al menos una superficie que permite la unión fuerte de pequeñas moléculas. Si uno trazara la frecuencia para encontrar un ligando con una afinidad dada, se obtendría un gráfico similar a la figura 3.5.6-izquierda. Para estas dianas, el tamaño, la calidad y el número de huecos que permitan la unión de moléculas pequeñas esta en relación con la proporción de *hits* observada en el cribado basado en fragmentos. Además, la proporción de *hits*

encontrados es totalmente dependiente de la definición que se tome para *hit*. Para proteínas susceptibles de ser moduladas, las condiciones de cribado y la definición de *hit* son generalmente rigurosas (elevado valor de corte), sólo permitiendo la observación y la selección de fragmentos relativamente potentes. En este contexto la literatura relata proporciones de *hits* alrededor del 10 % para unos valores de cribado que corresponderían a una afinidad de al menos 0.5 mM, propio de una proteína susceptible de ser modulada.
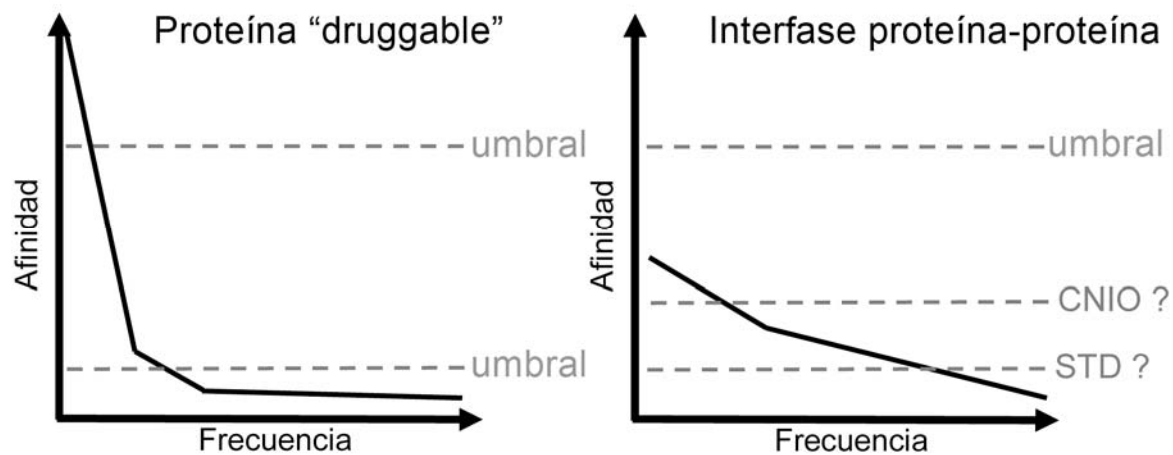
Figura 3.5.6: Hipótesis de proporciones de *hits* obtenidos por cribado de fragmentos para proteínas susceptibles de ser moduladas e interfases proteína-proteína. La proporción de *hits* depende de la frecuencia de identificación de ligandos con una afinidad específica y la definición de *hit*.

Contrariamente a las interfases proteína-proteína susceptibles de ser moduladas, la carencia de cavidades o superficiales dificulta una interacción de alta calidad. Además, su flexibilidad intrínseca y el bajo confinamiento espacial de las superficies permiten modos de interacción más promiscuos. Esto puede ser visto en el comportamiento representado en la figura 3.5.6-derecha, que muestra una baja probabilidad para unir ligandos con alta afinidad, pero la elevada para unir ligandos débilmente. Éstos últimos pueden ser descubiertos en condiciones de cribado y definiciones de *hit* apropiados. Creemos que VEGF es un ejemplo para este comportamiento. La alta proporción de *hits* obtenidos en el cribado por STD son consecuencia de las condiciones de cribado y el bajo umbral para la aceptación de *hits*. Basándonos en esto, la afinidad de los ligandos encontrados debe ser baja. Esto se correlaciona con las intensidades relativas de STD bajas, la ausencia de múltiples competidores de la molécula espía ^{19}F y la afinidad aparentemente baja de los *hits* obtenidos del cribado con ^{19}F-RMN. Estimamos que la mayoría de los *hits* tienen afinidades menores a 10 mM.
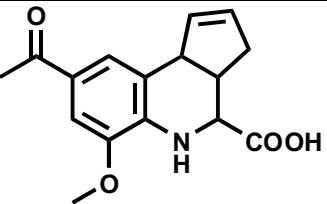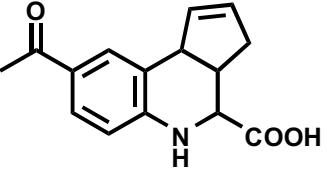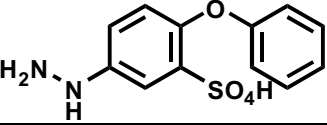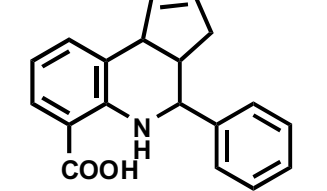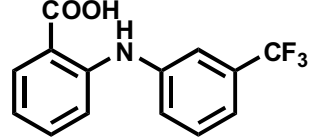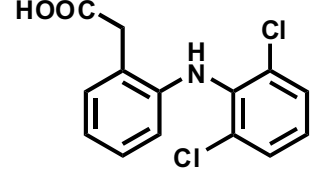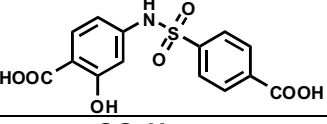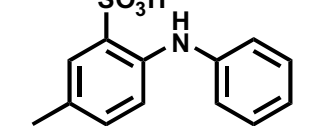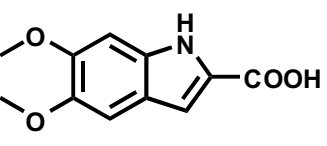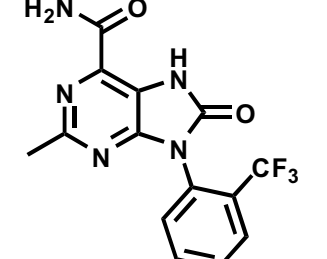
6.3.3 Resultados post-cribado

En esta sección se exploró la posibilidad de caracterizar y elaborar fragmentos que se unan a VEGF. Se perseguía una aproximación racional que confiaba en la combinación de datos de RMN, sacados de experimentos basados en la observación de proteína, con un protocolo de *docking* por ajuste inducido basado en el algoritmo PELE. Al ponerlo en práctica, afrontamos tres problemas clave:

1. Validación experimental de nuestros compuestos en cuanto a lugar de unión y afinidad.
2. Análisis e interpretación de datos confusos e incoherentes.
3. Mejora de nuestros compuestos.

La validación experimental fue realizada mediante CSP de ^{1}H-^{13}C-HSQC y ^{1}H-^{15}N-HSQC. El primero fue usado para una evaluación rápida, para saber si los fragmentos se unían a la proteína y estimar un rango de afinidad. El segundo fue usado para determinar los sitios de unión de los fragmentos en la superficie de la proteína con una resolución a escala de residuo. Los experimentos fueron realizados a 25°C porque la unión a temperaturas más altas daba lugar a más inconsistencias. Para aumentar la sensibilidad del experimento disminuimos la concentración de proteína para ser capaces de trabajar a proporciones de proteína-ligando más altas, ya que quedaba limitada por la solubilidad de los fragmentos.

Las representaciones de las uniones fueron generadas por *docking* de fragmentos con Glide XP cerca del sitio de unión indicado por RMN. Entonces, el algoritmo PELE se utilizó para generar diferentes conformaciones de la proteína en presencia del ligando. Más de tres representaciones de unión, que mostraban buenos resultados de unión (de acuerdo con los datos de CSP y los datos disponibles de QSAR), fueron seleccionadas de una población mucho mayor. La representación de unión seleccionada se utilizó para el *docking* de nuevos compuestos para mejorar la afinidad. Mediante la repetición de este método después de tres ciclos descubrimos varias familias de compuestos que se unen a VEGF (ver tabla 3.6.3-2).

Familia de esqueleto	ID	Estructura	CSP a 1 mM (residuo)	Kd de este residuo por CSP(R2)	Kd de este residuo por cambio de intensidad (R2)
tetrahidro-quinolina	806		0.003 (M81)	3.4 mM (0.997)	-
	831		0.016 (M81)	5.7 mM (0.999)	0.7 mM (0.994)
	841		0.010 (M18)	5.0 mM (0.999)	0.4 mM (0.999)
	834		0.020** (M81)	0.3 mM (0.917)	0.1 mM (0.997)
difenil-amina	03E04		0.003 (M81)	3.4 mM (0.997)	-
	836		0.013 (M81)	2.5 mM (0.998)	-
	839*		0.006 (M81)	2.2 mM (0.998)	-
	840**		0.004 (M18)	6.3 mM (0.999)	-
indol	374D		0.003 (M81)	20.0 mM (-), para otros residuos 2 mM	-
	818		0.004 (M81)	1.6 mM (0.993)	-
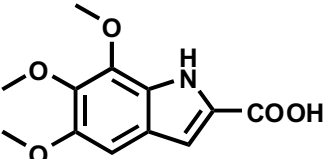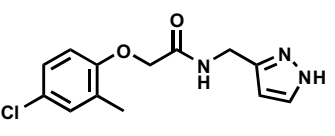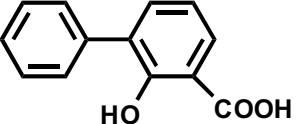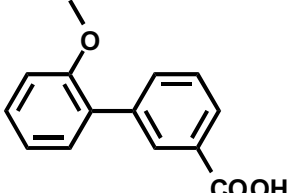
	821		0.005 (M81)	2.6 mM (0.999)	-
Anillos aromáticos unidos	644		0.009 (M55)	1.6 mM (0.987), 4 mM para otros residuos	-
biaril	845		0.01 (M81)	3.7 mM (0.999)	-
	827		0.009 (M81)	3.7 mM (0.996)	-

Tabla 3.6.3-2: Resumen de las cinco familias de esqueletos que fueron explorados en el último ciclo. Para cada familia la primera entrada muestra el mejor candidato del ciclo anterior. Las entradas adicionales listan los compuestos prometedores hallados en este ciclo. (* También podría pertenecer a la familia de biarilo, ya que no se adquirieron datos de la cadena principal de la proteína, ** también podría unirse en la proximidad de 841, ya que no se adquirieron datos de la cadena principal de la proteína).

En la figura 3.6.3-3 están representadas las distintas zonas de unión de las familias de nuestros compuestos. Se muestra en rojo la superficie que está en contacto directo con el receptor de VEGF-R2. Tenemos un alto grado de confianza respecto a los pares de compuestos 841, 834 y 836, 845 y sus zonas de unión. Uno podría imaginar fusionar los compuestos individuales de cada par, o los dos pares. Aunque la distancia entre los fragmentos es relativamente larga, se extiende por encima de la interfase de unión al receptor. Se podrían promover interacciones adicionales que aumenten la potencia del inhibidor. Los compuestos 644 y 578 requieren validación adicional. A pesar de que los fragmentos 841 y 834 no se unen a la interfase del receptor, podrían ser útiles como anclajes en la proteína para mejorar la afinidad. Como perspectiva de futuro, los datos de unión presentados pueden servir para la fusión de fragmentos.
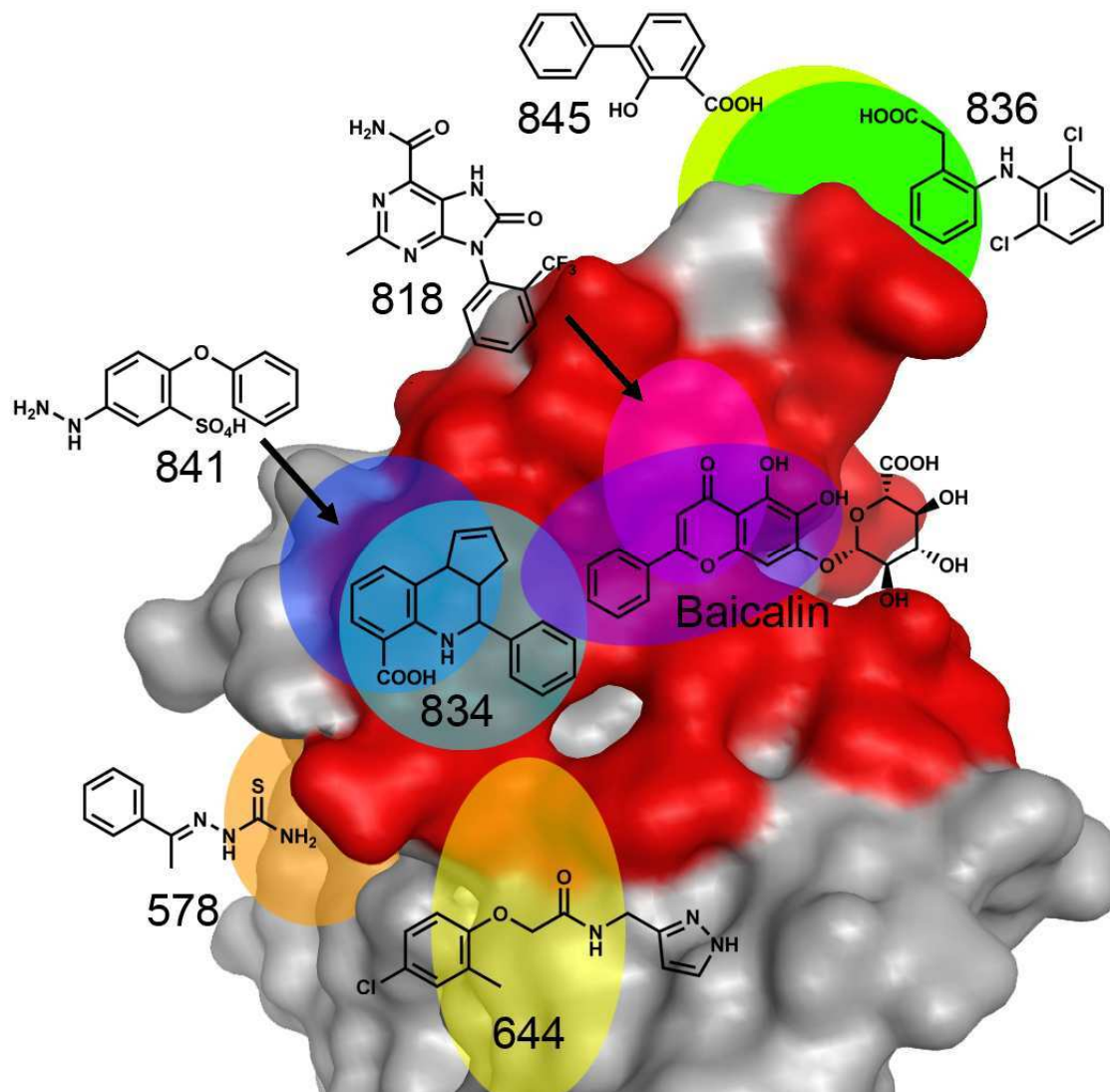
Figura 3.6.3-3: Representación de de la superficie VEGF (3V2A). La superficie de contacto con el receptor VEGF-R2 se muestra en rojo. Las zonas esféricas coloreadas marcan los sitios potenciales de unión de los compuestos.

6.3.4 Exploración preliminar de mRNA display para identificar ligandos peptídicos de VEGF.

Hasta ahora, todos los esfuerzos se concentraron en estrategias basadas en fragmentos para identificar pequeñas moléculas que se unieran a la interfase proteína-proteína de VEGF. En este capítulo hemos explorado una tecnología novedosa, *mRNA display,* para buscar péptidos de tamaño mediano que podrían unirse a VEGF.

Para la selección se utilizó *Random Peptide Integrated Discovery* (RaPID), la cual se desarrolló en el laboratorio del Prof. Dr. Suga. Esta tecnología se basa en la reprogramación del código genético y la expresión de péptidos cíclicos que contienen aminoácidos no naturales, los cuales a menudo presentan una mayor estabilidad *in vivo* y mayor afinidad.

En total se realizaron cuatro selecciones, dos con la biblioteca de mRNA NNK y dos con la NNU. Una reprogramación genética se utilizó para incorporar amino ácidos no naturales reactivos que forman marco-péptidos cíclicos mediante la formación de un enlace tioéter. La biblioteca NNU contenía la reprogramación adicional de cuatro aminoácidos a N-metil aminoácidos.

Un problema encontrado durante este trabajo fue la utilización de una solución de proteínas que contenía impurezas, además del dímero nativo. Sin embargo, no hubo problemas adicionales durante la selección.

Se realizaron un máximo de 11 y 8 ciclos de selección para la biblioteca NNK y la NNU, respectivamente. Todas las selecciones mostraron un buen perfil, con creciente tasa de recuperación de *mRNA* de las bolitas con VEGF inmovilizado (ligandos potenciales), y tasas de recuperación de bolitas vacías significativamente menores (ligandos de bolitas). Esto se puede observar en el perfil de selección de NNU en la figura 3.7.3-5.
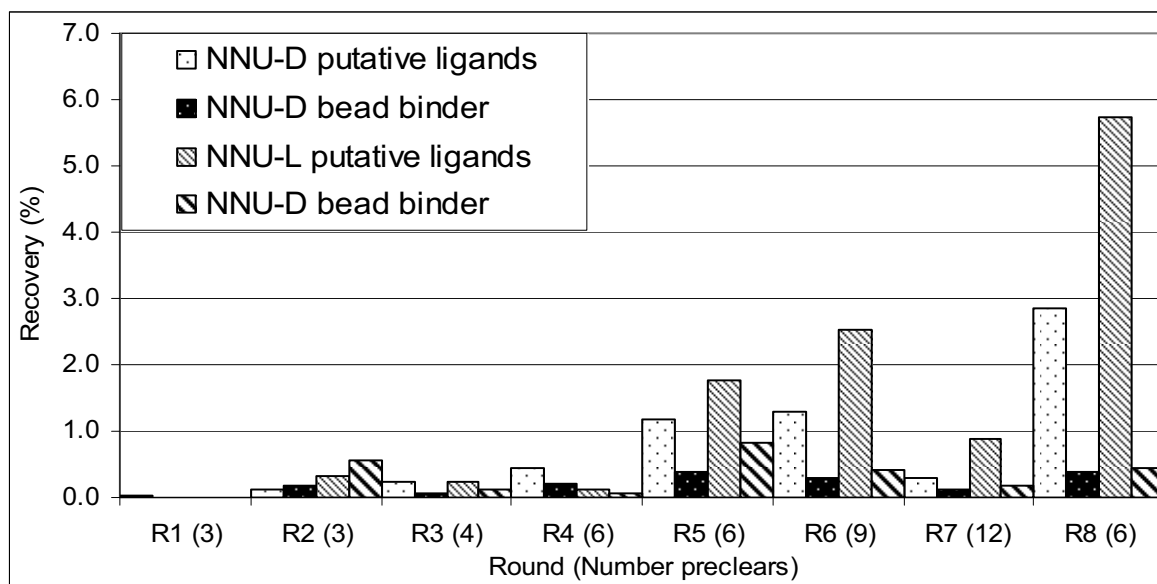

Figura 3.7.3-5: Perfil de las dos selecciones de NNU supervisados por el *DNA* recuperado de bolitas recubiertas de proteína (ligandos de VEGF) y las bolitas vacías (ligandos de bolitas). Las flechas marcan los *DNA pools* que fueron utilizados para la secuenciación.

Para las bibliotecas de NNK, los ciclos 6 y 11 fueron secuenciados, y para la biblioteca NNU, los ciclos 7 y 8. Los resultados de la secuenciación mostraron un enriquecimiento de motivos específicos, en lugar de obtener una convergencia en las secuencias. Aunque esto es inusual, se ha observado antes en el laboratorio del Prof. Dr. Suga.

En las dos selecciones, NNU-D y -L , varias secuencias contienen el codón para Leu, Ile, Ala, y Phe modificado con aminoácidos *N*-metilados. La presencia de estos codones en la secuencia de datos confirma que la calidad del sistema de traducción *in vitro* y los *t-RNAs* no naturales cargados eran suficientes para la reprogramación genética para obtener los péptidos deseados. El candidato con la

mayor incorporación de aminoácidos no naturales pertenece al ciclo 8 de la biblioteca NNU-D, que presenta tres aminoácidos *N*-metilados de las 12 posiciones totales. Esto se acerca a los mejores resultados obtenidos en el laboratorio del Prof. Dr. Suga.
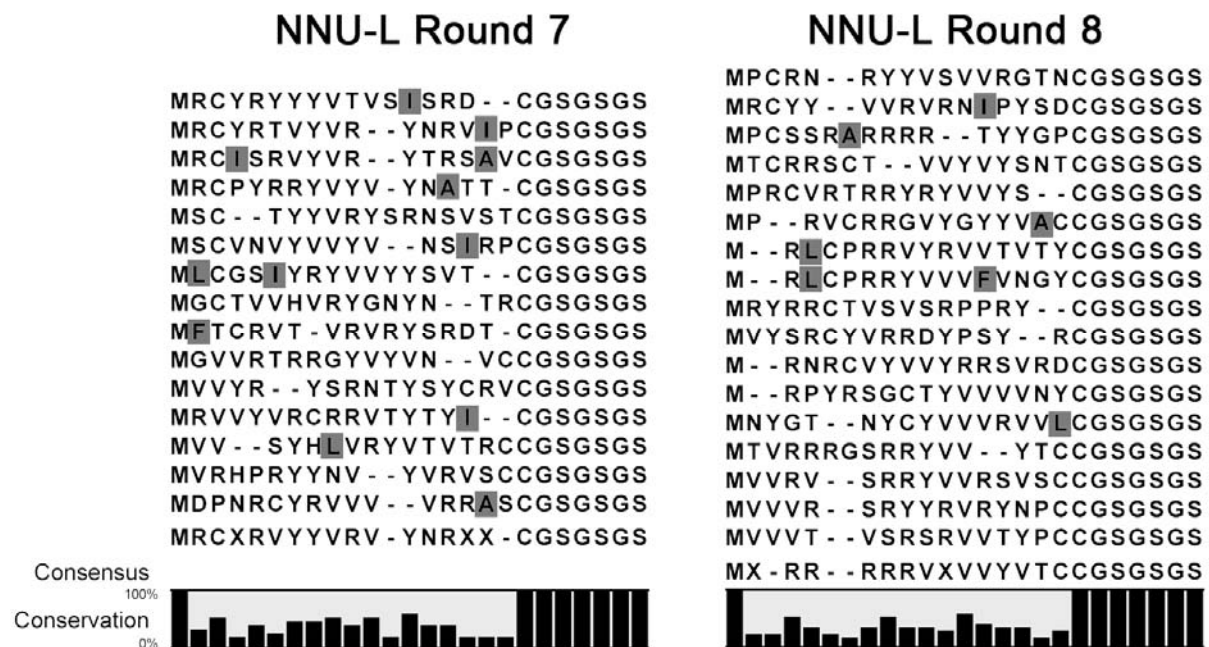
NNU-L Round 7

```
MRCYRYYYVTVSISRD- -CGSGSGS
MRCYRTVYVR- -YNRVIPCGSGSGS
MRCISRVYVR- -YTRSAVCGSGSGS
MRCPYRRYVYV-YNATT-CGSGSGS
MSC- -TYYVRYSRNSVSTCGSGSGS
MSCVNVYVVYV- -NSIRPCGSGSGS
MLCGSIYRYVVYYSVT- -CGSGSGS
MGCTVVHVRYGNYN- -TRCGSGSGS
MFTCRVT-VRVRYSRDT-CGSGSGS
MGVVRTRRGYVYVN- -VCCGSGSGS
MVVYR- -YSRNTYSYCRVCGSGSGS
MRVVYVRCRRVTYTYI- -CGSGSGS
MVV- -SYHLVRYVTVTRCCGSGSGS
MVRHPRYYNV- -YVRVSCCGSGSGS
MDPNRCYRVVV- -VRRASCGSGSGS
MRCXRVYYVRV-YNRXX-CGSGSGS
```

NNU-L Round 8

```
MPCRN- -RYYVSVVRGTNCGSGSGS
MRCYY- -VVRVRNIPYSDCGSGSGS
MPCSSRARRRR- -TYYGPCGSGSGS
MTCRRSCT- -VVYVYSNTCGSGSGS
MPRCVRTRRYRYVVYS- -CGSGSGS
MP- -RVCRRGVYGYYVACCGSGSGS
M- -RLCPRRVYRVVTVTYCGSGSGS
M- -RLCPRRYVVVFVNGYCGSGSGS
MRYRRCTVSVSRPPRY- -CGSGSGS
MVYSRCYVRRDYPSY- -RCGSGSGS
M- -RNRCVYVVVYRRSVRDCGSGSGS
M- -RPYRSGCTYVVVVNYCGSGSGS
MNYGT- -NYCYVVVRVVLCGSGSGS
MTVRRRGSRRYVV- -YTCCGSGSGS
MVVRV- -SRRYVVRSVSCCGSGSGS
MVVVR- -SRYYRVRYNPCCGSGSGS
MVVVT- -VSRSRVVTYPCCGSGSGS
MX-RR- -RRRVXVVYVTCCGSGSGS
```

Consensus 100%
Conservation 0%

Figura 3.7.4-4: Resultados de la secuenciación para los ciclos 7 y 8 para la selección NNU con el iniciador Cl-Acletil-L-Trp. Los cuadrados oscurecidos señalan los codones que fueron programados L-> N-methyl Ser; F-> N-methyl-Phe; A-> N-methyl-Ala; I-> N-methyl-Gly.

Para tener más confianza en estos resultados, hemos realizado un ensayo de clonación que permite la evaluación de las secuencias específicas. El resultado fue muy positivo respecto a las tasas de recuperación de mRNA. También lo fue para la competencia entre VEGF bien plegada sin HisTag en solución y proteína unida a las bolitas por los péptidos. Esto indica la selección de péptidos que reconocen la proteína nativa.

Esta fue una confirmación importante teniendo en cuenta que el stock de proteínas que usamos contenía impurezas.
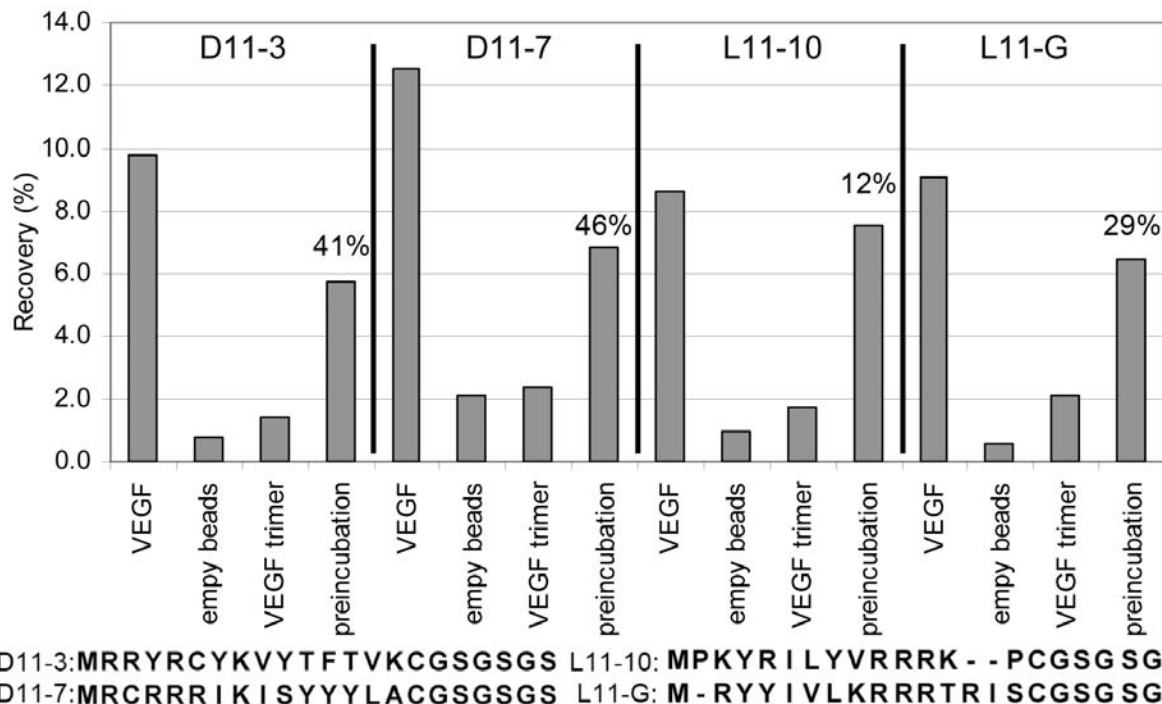
Figura 3.7.4-5: Ensayo clon para dos secuencias seleccionadas de cada selección NNK. El valor de porcentaje describe la disminución de *DNA* recuperado cuando la proteína inmovilizada sobre las bolitas competía con VEGF bien plegada en solución.

Aunque la síntesis de péptidos seleccionados plantea un desafío sintético, pueden ser obtenidas cantidades pequeñas para los estudios preliminares de unión. Se utilizó para este fin la polarización de fluorescencia y experimentos de RMN basados en observación de proteína.

Los ensayos de polarización de fluorescencia no indicaron una fuerte afinidad de los péptidos por VEGF.

En los experimentos de RMN basados en proteína se observaron algunas perturbaciones inducida por el péptido pero con magnitudes débiles. Una cuestión importante era la baja solubilidad de los péptidos, que disminuía aún más, de manera significativa, en presencia de proteína. Esto podría indicar cierta interacción de la proteína y los péptidos, si se supone que el complejo tiene una solubilidad reducida. Debido a esto, los experimentos de RMN basados en proteína pueden no ser adecuados, ya que requieren un exceso de ligando respecto la proteína. Esto no fue posible para nuestros ligandos. Una alternativa para estudios futuros podría ser la evaluación mediante *surface plasmon resonante,* la cual se usa ampliamente en laboratorio del Prof. Dr. Sugas.

6.4 Conclusiones

1. El escalado de métodos previamente descritos ha permitido la expresión de VEGF a gran escala, tanto marcada isotópicamente como no marcada. La proteína marcada se requiere para los estudios de RMN mientras el uso de VEGF no marcada resulta más conveniente para los estudios basados en espectrometría de masas. Actualmente el único inconveniente es el elevado precio de la enzima usada para eliminar el HisTag.

2. La predicción de la solubilidad en algunos de los fragmentos es un problema no resuelto que tiene un gran impacto negativo en el diseño y preparación de la quimioteca

3. Se ha desarrollado un método automático para la generación de mezclas para cribado basado en RMN, que permite la evaluación completa de todos los fragmentos sin deconvolución de la mezcla.

4. Se ha desarrollado un método que permite la evaluación automática de datos procedentes de experimentos de cribado basados en STD-RMN ó CPMG-RMN de manera cuantitativa, mejorando tanto la velocidad como la calidad del análisis.

5. El cribado de nuestra quimioteca de fragmentos condujo a una alta proporción de "hits", pero de afinidad muy baja. Esto fué el primer signo de que la interfaz proteína-proteína de VEGF tiene mucha dificultad para interaccionar con ligandos de bajo peso molecular. En este contexto, el análisis automático cuantitativo de cribado de datos demostró ser una herramienta muy útil.

6. El cribado de una quimioteca de compuestos fluorados mediante RMN de ^{19}F condujo a la identificaron de "hits" en una proporción inferior a la observada utilizando nuestra quimioteca de fragmentos aunque nuevamente, todos los "hits" presentaron una baja afinidad por VEGF. Estos resultados constituyen una segunda indicación de la baja "druggability" de VEGF.

7. Se ha puesto a punto por primera vez un ensayo de competición para la unión a VEGF basado en el uso de RMN ^{19}F y un péptidos fluorado "como espía". La no detección de competidores potentes apunta, nuevamente, a una baja "druggability" de VEGF.

8. El docking rígido se ha relevado como una aproximación no adecuada para el cribado virtual de fragmentos capaces de unirse a interfaces proteína-proteína, al menos en el caso de VEGF.

9. Ninguno de los dos programas Sitemap (Schrödinger) y Sitefinder (MOE) fue capaz de clasificar como "druggable" la interfaz proteína-proteína de VEGF.

10. El empleo de métodos de RMN basados en la proteína para caracterizar la interacción de VEGF con ligandos de afinidad bajos (>10mM) ha demostrado ser muy difícil. Solo se obtuvieron buenos resultados trabajando a bajas temperaturas y a concentraciones de proteína bajas.

11. Una primera evaluación de compuestos utilizando experimentos de [metilo-13C]-metionina CSP mejoró considerablemente los resultados y permitió la estimación de afinidades.

12. El uso combinado de PELE con información sobre la unión ligando-proteína obtenida mediante experimentos de RMN parece ser un enfoque prometedor para la evaluación química de ligandos muy débiles incluso en el contexto de interfaces proteína-proteína.

13. Se han identificado tres familias estructurales de fragmentos capaces de unirse a VEGF con afinidades en un rango uM – mM. Algunos de estos compuestos se unen a la interfaz proteína-proteína de VEGF y constituyen un buen punto de partida para estudios futuros.

14. El uso de técnicas de "mRNA display" condujo al descubrimiento de manera preliminar, de varios péptidos candidatos a unirse a VEGF pero la confirmación de estos resultados mediante RMN y FP está todavía pendiente.