Development of a labeling strategy for the synthesis of fluorescent libraries of peptides and small molecules and their use in fluorescence-based binding assays for the study of protein-protein interactions

DISSERTATION

zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat.)

im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

2009

vorgelegt von

VIVIANE URYGA-POLOWY

1. Gutachter: Prof. Dr. J. Rademann

2. Gutachter: Prof. Dr. M. Bienert

Tag der mündlichen Prüfung   03.12.2009

À Benoît…

« Le bonheur, c'est de continuer
à désirer ce que l'on possède »

Saint-Augustin

Die vorliegende Arbeit wurde unter der Anleitung von

Herrn Prof. Dr. Jörg Rademann

in der Zeit von Oktober 2004 bis März 2009 am Leibniz-Institut für Molekulare Pharmakologie (FMP) im Forschungsverbund Berlin e.V. und am Institut für Chemie und Biochemie der Freie Universität Berlin durchgeführt.

Herrn Prof. Dr. Jörg Rademann danke ich sehr herzlich für die Bereitstellung dieses spannenden und vielseitigen Themas, für die intensive Betreuung, die hervorragenden Arbeitsbedingungen und sein mir entgegengebrachtes Vertrauen.

Danksagung

Table of contents

Symbols and abbreviations

Amino-acids were represented by the one-letter code: A: alanine; C: cysteine; D: aspartic acid; E: glutamic acid; F: phenylalanine; G: glycine; H: histidine; I: isoleucine; K: lysine; L: leucine; M: methionine; N: asparagine; P: proline; Q: glutamine; R: arginine; S: serine; T: threonine; V: valine; W: tryptophan; Y: tyrosine.

AA	amino acid
Ac	acetyl
AcOH	acetic acid
Acm	acetamidomethyl
Ac$_2$O	acetic anhydride
ADMET	absorption, distribution, metabolism, excretion and toxicity
ATR-IR	attenuated total reflexion infra-red
a.u.	arbitrary unit
Boc	*tert*-butyloxycarbonyl
BODIPY	boron-dipyrromethene
BOP	(benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafluoro-phosphate
Bpa	4-benzoyl-phenylalanine
Cbz	carbobenzyloxy
CD	circular dichroism
CF	carboxyfluorescein
CFSE	carboxyfluorescein succinimide ester
CMC	comprehensive medicinal chemistry
d	doublet
Da	dalton
DAD	diode-array detector
DCC	dicyclohexyl carbodiimide
DCE	dichloroethane
DCM	dichloromethane

DIC	diisopropylcarbodiimide
DIPEA	*N,N*-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethylsulfoxyde
DNA	desoxyribonucleic acid
DNS	dansyl
DVB	divinylbenzene
EDC	*N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide
e.g.	exempli gratia
ELISA	enzyme-linked immunosorbent assay
equiv	equivalent
ESI-MS	electrospray ionization mass spectrometry
et al.	et alii / alia
EtOH	ethanol
Et$_2$O	diethylether
EVH1	Ena / VASP homology 1
FCS	fluorescence correlation spectroscopy
FITC	fluorescein isothiocyanate
Fmoc	9-fluorenylmethoxycarbonyl
FP	fluorescence polarization
FRET	fluorescence / Förster resonance energy transfer
FSML	fluorescent small molecules library
FT-ATR	Fourier transformation attenuated total reflexion
g	gram
GFP	green fluorescent protein
GST	glutathione S-transferase
GYF	domain name derived from the amino acid one-letter code for the characteristic tripeptide glycine-tyrosine-phenylalanine
h	hour
HATU	2-(7-aza-1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate

HBTU	2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexa-fluorophosphate
HCl	hydrogen chloride
Hex	hexane
HF	hydrogen fluoride
HFIP	hexafluoroisopropanol
HMPA	hexamethylphosphoramide (Tris(dimethylamino)phosphine oxide)
HMPT	Hexamethylphosphorous triamide (Tris(dimethylamino)phosphine)
HOAt	4-aza-3-hydroxybenzotriazole
HOBt	1-hydroxybenzotriazole
HPLC	high pressure / performance liquid chromatography
HTS	high-throughput screening
IAF	iodoacetamidofluorescein
IC_{50}	half maximal inhibitory concentration
IR	infra-red
ITC	isothermal titration calorimetry
K_D	dissociation constant
K_i	inhibition constant
λ	wavelength
L	liter
LC	liquid chromatography
MAS	magic-angle spinning
MCS	maximum common substructure
MeOH	methanol
MEM	2- methoxyethoxymethyl ether
m	meter
M	$mol \cdot L^{-1}$
min	minute
mP	mili polarisation units
MPLC	medium pressure liquid chromatography
MS	mass spectrometry
MSNT	1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole
MTP	microtiter plate

MW	microwave
m/z	mass to charge ratio
NEt$_3$	triethylamine
n.d.	not determined
NMR	nuclear magnetic resonance
PAM	4-hydroxymethyl-phenylacetamidomethyl
Pbf	2,2,4,6,7-pentamethyl-dihydrobenzofuran-5-sulfonyl
PBS	phosphate buffered saline
PDB	protein database
PE	polyethylene
PEG	polyethylene glycol
PERQ	P, E, R, and Q amino acid rich with GYF domain protein
PG	protecting group
Pmc	2,2,5,7,8-pentamethyl-chromane-6-sulfonyl
PP	polypropylene
PPII	polyproline type II
ppm	parts per million
PRD	proline-rich sequence recognition / proline recognition domain
PRS	proline rich sequence
PS	polystyrene
pyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
QSAR	quantitative structure-activity relationship
RFU	relative fluorescence unit
rpm	revolutions per minute
RT	room temperature
s	singulet
SAR	structure-activity relationship
Sec	second
SEC	size-exclusion chromatography
SH3	Src homology 3
SIM	single ion monitoring
SPOS	solid phase organic synthesis

SPPS	solid phase peptide synthesis
t	triplet
TAMRA	tetramethylaminorhodamine
TATU	2-(7-aza-1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetra-fluoroborate
TBAF	tetrabutyl ammonium fluoride
TBDMS	*tert*-butyl dimethylsilyl
TBTU	2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoro-borate
tert-Bu	*tert*-butyl
TEA	triethylamine
TES	triethylsilyl
THF	tetrahydrofurane
TIC	total ion chromatogram
TIS	triisopropyl silane
TFA	trifluoro acetic acid
TFE	trifluoroethanol
TLC	thin layer chromatography
TRITC	tetramethylrhodamine isothiocyanate
Trt	trityl
TSA	thermal shift assay
UV	ultra violet
v/v	volume/volume
WDI	world drug index
WT	wild-type
WW	domain name derived from the one-letter code for the two conserved tryptophan residues
°C	Celsius degree

I. INTRODUCTION

Scientific progress is driven by human curiosity, and by a constant need of people to modify their environment and improve their well-being. Thus, the human body has been extensively studied. However, a real understanding of its mechanisms is recent and all major advances, basis for current medical practice, have been developed in the 19th and 20th centuries.

In the last twenty-five years, genome sequencing opened a new chapter in medicinal research. A lot of data have been generated: the whole human genome could be sequenced, but as data analysis did not progress as fast, a large part of the human proteome still remains undefined. Therefore, cooperation between medical scientists, biologists, and chemists is necessary to develop new analytical tools for the elucidation of interactions governing the organism. In particular, protein-protein interactions are an interesting target, due to their implication in many metabolic pathways and DNA processes.

However, unlike the study of enzymes, which have known substrates that can be used as starting point for the development of inhibitors, the study of protein-protein interactions is complex. In particular, the precise localization of the interacting areas is difficult. In this perspective, the development of new assays targeting the study of protein-protein interactions is a great challenge. In particular, fluorescence-based experiments involving libraries of labeled compounds have been addressed in this dissertation project.

II. GENERAL PART

II.1. Basics in medicinal chemistry

II.1.1. The process of drug discovery

Medicinal chemistry is the part of chemistry dealing with the design and synthesis of new molecules useful as diagnostic or therapeutic tools. An interdisciplinary approach combining medicinal chemistry, pharmacology, and biology, is important for a successful elucidation of the structure and function of unknown proteins, as well as the generation of active compounds.

II.1.1.1. Nature, the first source of "drugs"

Plants have already been used for a long time either as medicines, as perfumes, or as poisons. However, the extraction, isolation, and identification of the enclosed active compounds began only in the nineteenth century. Later, these natural molecules were modified and analogs were synthesized to increase their activity, their selectivity, and to reduce their side effects.
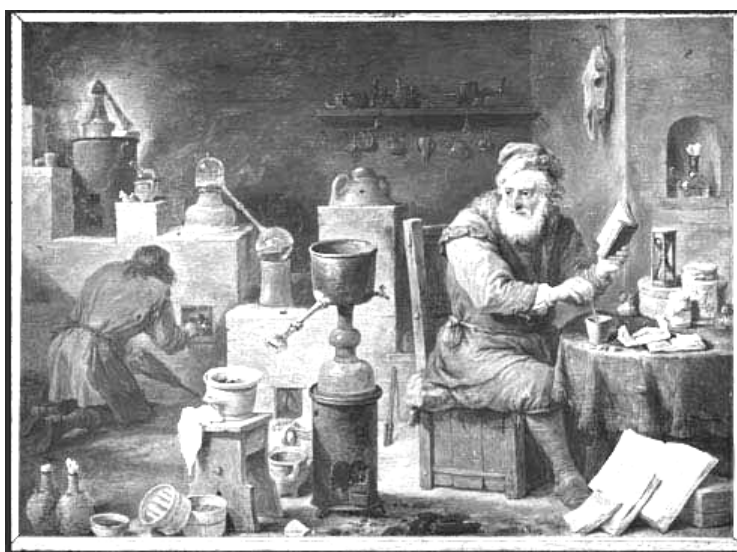
Figure 1: An alchemist in the medieval times: the precursor of the modern chemist

II.1.1.2. Rational drug development

The first rational development of a synthetic drug was achieved by P. Ehrlich and S. Hata. Starting from *Atoxyl*, they developed a more efficient drug for the treatment of syphilis: *Arsphenamine* (Scheme 1) [1,2]. This molecule was the result of a whole process of synthesizing analogous compounds, evaluating their beneficial and toxic properties in biological screenings [3], and selecting the appropriate candidate for a new drug.

Scheme 1: From Atoxyl to Arsphenamine: the first rational development of a synthetic drug (1910).

II.1.1.3. Modern drug discovery

Nowadays, two main strategies are used for the development of a new drug: either the search for completely new molecules, or the modification and optimization of already existing active compounds. Anyway, drug discovery is a long, tedious and expensive process.

Figure 2: Modern drug discovery. The process of d*e novo* drug discovery is long (10-17 years) and has a quite low probability of success (under 10%). In contrast, the strategy of drug repositioning can lead faster (within 3-12 years) to new drugs with minimized risks [4].

De novo drug discovery (Figure 2, upper part) is a very long process. After the target has been identified, different techniques such as molecular modeling, traditional or combinatorial chemistry and HTS (high-throughput screening) can be used to find a lead structure. Then, this structure is modified to improve its activity and selectivity by traditional medicinal chemistry and rational drug design. The ADMET (absorption, distribution, metabolism, excretion and toxicity) properties are also optimized and the molecule is finally sent to clinical trials (phase I and II). In the best case, the procedure is successful and the molecule can be registered by the different responsible organizations. Unfortunately, the probability of success of such a complex process is very low.

Therefore, pharmaceutical companies often privilege "drug repositioning" (Figure 2, lower part). It enables to circumvent the major problem of *de novo* drug discovery, i.e. extreme time consumption. Indeed, all steps of lead discovery are eliminated. Furthermore, the final success is more likely, thanks to a better knowledge of the pharmacokinetics and safety profile of the molecule.

II.1.2. Some important properties of lead molecules

II.1.2.1. Bioavailability

To be suitable as a lead for oral administration, a molecule has to fulfill bioavailability criteria: this means that a major part of the ingested drug should be found again in the general circulation. To avoid unnecessary synthetic work by producing non-bioavailable and thus unadapted substances, this question should be already addressed at the beginning of drug design. As a helpful tool, Lipinski proposed in 1997 the "rule of five" (Figure 3) [4-6], derived from the analysis of orally administrated drugs.

✓ molecular mass < 500

✓ log $P_{octanol/water}$ < 5

✓ number of hydrogen bond acceptors < 10

✓ number of hydrogen bond donors < 5

Figure 3: Lipinski's "rule of five". A substance following at least three of these rules has a good drug-like character.

II.1.2.2.   Solubility

Most of the drugs are administrated either orally or by intravenous injection; they have thus to be hydrophilic enough to be soluble in water or blood. On the other hand, molecules have also to be lipophilic enough to pass through membranes. The difficulty in the development of a new drug is the accurate adjustment of its solubility properties to establish a good balance between these two opposite characteristics that are hydrophilicity and lipophilicity.

The solubility of a molecule can be modified by changing some functional groups: water solubility can e.g. be increased by introducing polar groups such as alcohols, amines, amides, carboxylic acids, sulphonic acids, or phosphates. Lipophilicity can be improved by replacing polar groups by less polar such as methyl or iodine.

For a drug, the solubility and adsorption properties can also be modified at a later stage, by adjusting the formulation.

II.1.2.3.   Stability

Another important aspect to consider in drug development is the stability of the target molecule within the organism: a drug should be stable enough to reach its active site before being metabolized, but not too much, to avoid accumulation in the organism due to insufficient excretion.

The stability of a molecule in the organism can also be modulated by incorporating specific groups in the structure and using modified dosage forms.

II.1.2.4.   Structure

The action of a drug is due to its interaction with a target domain, as explained by the lock-and-key principle [7] (Figure 4). Molecules with different absolute configurations have different three-dimensional structures and are thus unequally binding to the target domain. In addition, the configuration is also important because of the stereoselective nature of many metabolic processes [8].

Another relevant structural aspect is the rigidity of the molecule. Although high conformational flexibility enables a better fitting to the target domain, the resulting affinity is lower than the affinity of an equivalent rigid system. Selectivity is also

enhanced by a rigid system, because binding to similar targets after conformational adaptation is avoided.

Figure 4: Lock-and-key principle. The structure of a lead compound has to be complementary to this of its target domain.

II.1.3.   Design and generation of lead molecules

II.1.3.1.   Structure-activity and quantitative structure-activity relationships

SAR (structure-activity relationship) is based on the principle that molecules structurally similar to biologically active compounds are themselves active. In practice, a series of structurally similar compounds is studied to determine which part of the molecule is important for the activity. The length and multiplicity of the side-chains are varied, rings are added or suppressed, and the saturation grade is modified. The collected information is then used to find new active compounds, with an improved potency or reduced side-effects.

This approach can be improved by introducing numerical parameters for the physicochemical properties. QSAR (quantitative SAR) relates quantitatively the changes in the molecular descriptors to the variations of biological activity and is therefore a very helpful tool to direct the chemist in the choice of analogs to synthesize.

II.1.3.2.   Structure-based drug design

Computers enable 3D visualization of interactions between molecules using molecular modeling tools. Atoms are basically described as point charges with an associated mass. Interactions between neighboring atoms are described by electrostatic and van der Waals forces. All resulting parameters enable the calculation of the potential function and thus the determination of local energy minima. Different softwares can be

used for molecular modeling (e.g. MOE, Sybyl). If an NMR or crystal structure of the protein is available, potential ligands can be docked into the binding pocket or on the protein surface.

Modeling simplifies the analysis of the binding mode and is a source of inspiration for the selection of further ligands to be synthesized and tested.

II.1.3.3. Combinatorial chemistry

Combinatorial chemistry was developed to satisfy the growing demand of compound libraries for HTS. This technique enables simultaneous synthesis of a large number of molecules from a limited number of building blocks – a performance which cannot be realized by traditional organic synthesis.

Different apparatus enable to carry out reactions either in solution, or on solid support. In both cases, however, the reactions have to be specific, easy to carry out and to automate, and deliver products in high yields and purities. Furthermore, the building blocks should be readily available and chemically diverse, to cover a broad range of interaction possibilities with the receptor. Different strategies can be used to generate libraries.
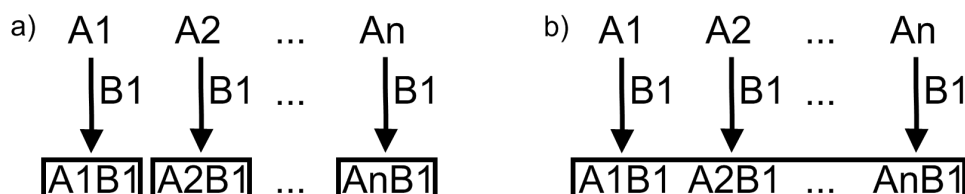
Figure 5: Different strategies in combinatorial chemistry. Either isolated pure compounds (a) or mixtures of compounds (b) can be obtained.

Parallel synthesis consists in using different building blocks that undergo the same reaction, at the same time, but in separate vessels. At the end of the process, each compound can be isolated (Figure 5a). Specific equipments are available for synthesis at different scales: from a volume of some milliliters in glass vessels to some microliters in MTP, or even lower volumes (picoliter range) with spot synthesis on surfaces [9].

Another method for combinatorial synthesis in solution is the generation of libraries of mixtures (Figure 5b): each well contains a mixture of compounds, their number

depending on the exact procedure. These mixtures are directly tested without separation, but a deconvolution step is necessary at the end of the process to identify the active compounds. It is also assumed that the different compounds have no interaction with each other, and do not interfere in the binding with the target.

On solid support, a broadly applied synthesis way is the "mix-and-split" technique developed by Furka in 1988 [10-12] (Figure 6). It is based on the "one-bead-one-compound" concept [11,13]: combinatorial bead libraries contain single beads displaying only one type of compound although there may be up to 1013 copies of the same compound on a single 100 µm diameter bead. This method enables a rapid generation of a high number of compounds. However, it requires also a deconvolution step or the use of an encoding method for hit identification.
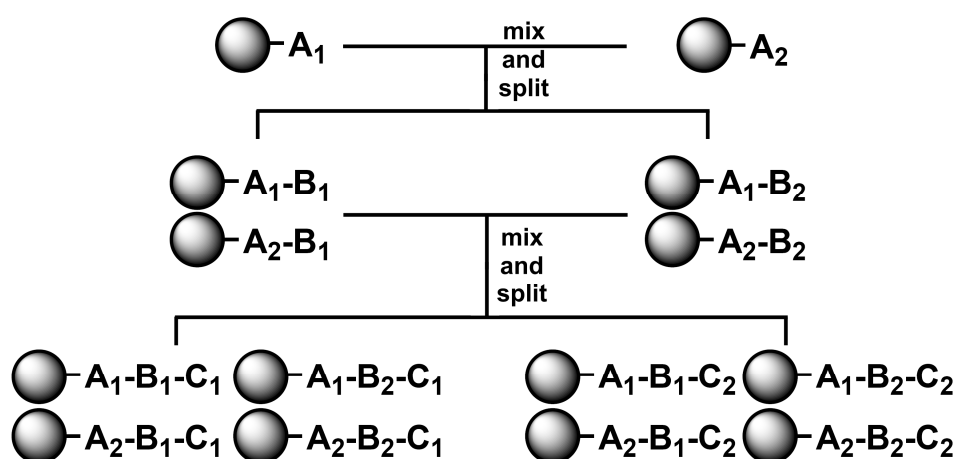
Figure 6: Furka's mix and split approach

II.1.4. Assays for lead discovery

II.1.4.1. *High-throughput screening*

In the last decades, high- throughput screening (HTS) became a common technique for lead discovery. However, whereas improved equipment enabled the screening of millions of randomly chosen compounds, the problem of relatively low hit rates pointed out the question of compounds selection and library design.

Compounds for HTS are traditionally stored as frozen DMSO solutions on MTP in 96-, 384-, 1536-, or 3456-wells format. With screening robots, they are automatically

transferred on assay plates to be submitted to previously defined assay protocols. The structure of the resulting "hits" is then tested to confirm their integrity.

Indeed, in spite of all precautions by storage and sampling, some degradation of compounds can occur and a repetition of assays on reordered or resynthesized hits should therefore always be conduced to confirm the first results. Finally, additional independent secondary assays are necessary to completely validate them.

II.1.4.2.   Functional assays

Enzymes are biomolecules that catalyze chemical reactions; their activity can be therefore quantified using functional assays in which the conversion rate of substrates into products is measured and analyzed using an appropriate model. The signal can be detected by different techniques such as spectrophotometry, radiometry, fluorescence, or mass spectrometry.

Spectrophotometric assays can only be performed when the absorbance properties of the substrate and the product are clearly distinguished. For radiometric assays, the incorporation of a radioactive label is necessary; the same applies to fluorescence-based assays. The advantage of both last techniques is nevertheless the high sensitivity that enables screening at very low concentration.

II.1.4.3.   Binding assays

Proteins with no enzymatic activity are more difficult to study; and other assays have to be used to evaluate their binding with a ligand or a protein. For HTS, HPLC-MS detection can be used, as well as fluorescence-based techniques can such as ELISA (enzyme-linked immunosorbent assay), FRET (fluorescence / Förster resonance energy transfer), or FP (fluorescence polarization). Other assays are more difficult to apply to the HTS format and are therefore mainly used for hit validation: e.g. NMR (nuclear magnetic resonance), ITC (isothermal titration calorimetry), CD (circular dichroism), or TSA (thermal shift assay).

II.1.4.4.   Cellular assays

Also assays on living cells can be realized in HTS format. Different properties such as cell toxicity, cell permeability, or distribution of compounds within the cell can be

studied. Cellular assays include furthermore functional investigations like reporter gene assays that can be used to study protein-protein interactions (e.g. two-hybrid screening), or translocation assays between different subcellular structures.

II.1.5. Design of small molecule libraries for screening

II.1.5.1. Large libraries vs. small fragment libraries

Although very large libraries – thousands or millions of compounds – are commonly used for HTS, the resulting hit rates are often very low. Molecules are indeed tested without any selection and cover, in spite of their number, only a limited proportion of the chemical space. Furthermore, difficulties are encountered to maintain the quality of such huge libraries, because many substances decompose over longer storage time and multiple freezing and defrosting steps.

An alternative to HTS is fragment-based screening; it is based on the idea that all drugs can be described by two or more fragments linked together [14]. As these building-blocks are smaller than traditional small molecules used for HTS. They are expected to bind only weakly, but once recombined in the right manner they should open the way for better ligands (Figure 7). This technique is often used by NMR or X-ray crystallography screenings.
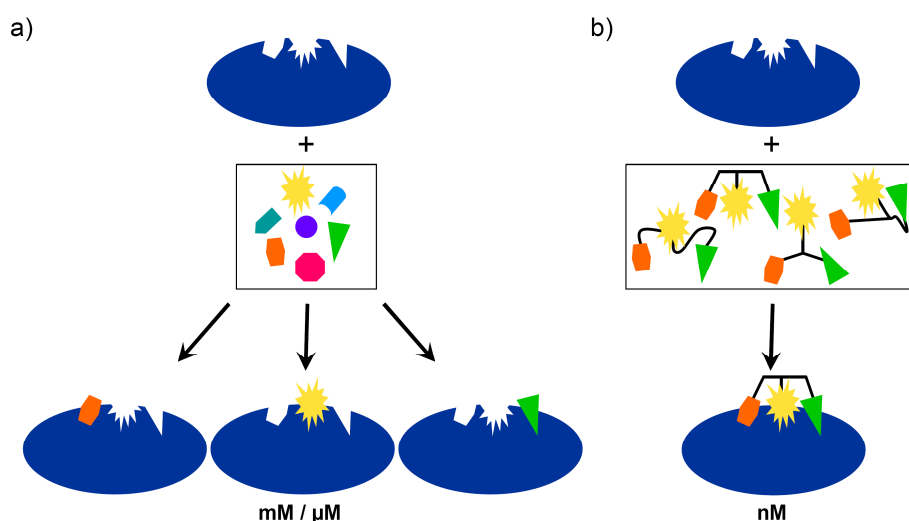
Figure 7: Concept of fragment-based screening [14]. After weak inhibitory fragments have been found for different adjacent binding pockets (a), they are bound to each other in different ways and tested again to find stronger inhibitors (b).

II.1.5.2. Design of fragment libraries

Fragment libraries have a limited number of compounds; these should therefore be accurately selected. Small, simple molecules are well-suited. The fragments should also be selected as various as possible (Figure 3). An analysis of known drugs can be very useful to determine the most important scaffolds present in drugs and facilitate the design of fragment libraries.

Such an analysis was done by M. Lisurek (group of R. Kühne, FMP, Berlin) using the MCS (Maximum Common Substructure) concept to design the ChemBioNet screening collection. The main tool of this strategy was an algorithm that simulates an experienced chemist's reasoning when classifying and considering the molecules. An example of substructures selected by this study and derived from the WDI (World Drug Index) is shown on Figure 8.

Former studies published by G. W. Bemis et al. [15] or Y. J. Xu were already addressing this topic by studying the CMC (comprehensive medicinal chemistry) database under the aspects framework, side chain, or linker [16].
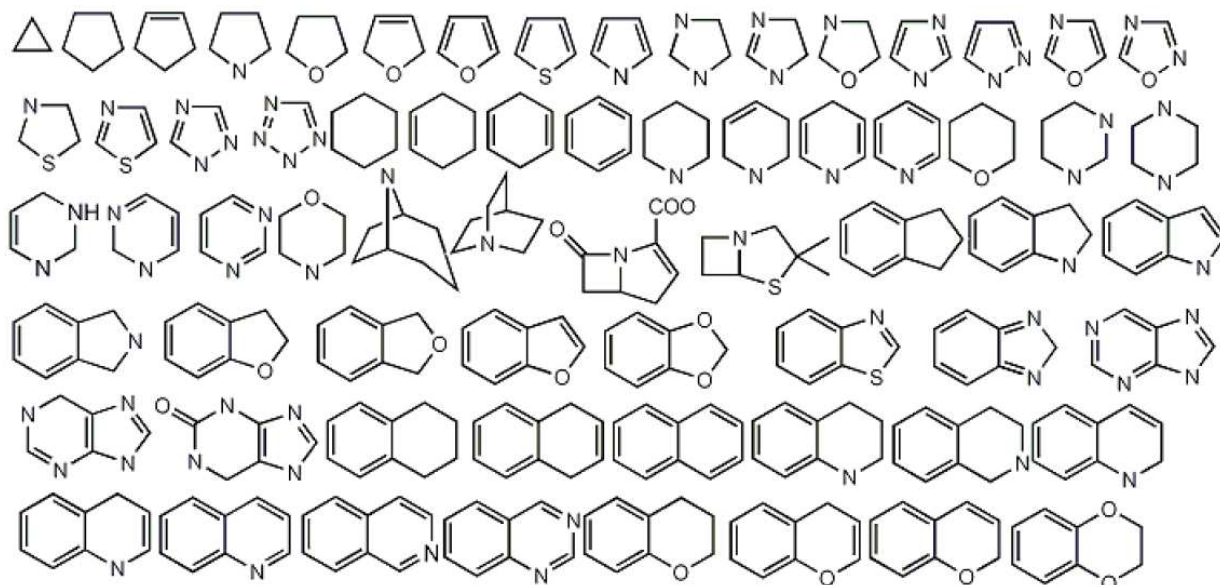
Figure 8: Overview of 66 of the 561 identified biologically active substructure classes, ordered by increasing complexity. Structures were derived by ClassPharmer from the WDI (M. Lisurek, group of R. Kühne, FMP, Berlin).

II.2. Solid phase synthesis

II.2.1. Emergence of solid phase synthesis

II.2.1.1. Merrifield's new concept to improve peptide synthesis

Merrifield introduced the concept of solid-phase synthesis in the early sixties by using a PS (polystyrene) polymer cross-linked with DVB (divinylbenzene) for the synthesis of a tetrapeptide[17]. The idea was to use an insoluble but porous polymer that should act as a carrier for chemical reactions; between two coupling steps, the polymer beads should only be washed without any other kind of purification or intermediary analysis.

This was a complete break with traditional organic solution synthesis where isolation, purification, and characterization of all intermediates were essential. But many doubts subsisted about the future of this technique[18]. Merrifield convinced a larger public only after publication of his next article: the efficient synthesis of a nonapeptide isolated in 68% overall yield, highly pure and active, a result that exceeded all solution phase attempts[19,20].

In the following years, the concept was acclaimed by many biochemists and pharmacologists; it was also extended to oligonucleotide synthesis. The nature of linkers and protecting groups was varied to find reagents compatible with milder reaction conditions and other strategies concurrent to Merrifield's method were developed.

II.2.1.2. Principle of solid phase synthesis

The basic concept of solid phase synthesis is the use of an insoluble polymer as support for chemical reactions. The product remaining attached to the support, all reagents in excess or side-products in solution can be washed off. After completion of the whole synthesis sequence, the resulting pure product is cleaved from the resin and isolated.

Two procedures can be applied: "batchwise" and "continuous-flow" (Figure 9). In the batchwise process, reagents are introduced through the top, the vessel is shaken during the reaction, and the waste is finally washed off from the bottom. In the

continuous-flow process, the reaction mixture is continuously pumped through the resin beads.

The system of choice depends on the resin/solvent combination; for resins having a large swelling volume, the batchwise procedure is preferentially used. The corresponding equipment consists of a fritted tube as reaction vessel, a shaker, and a washing line. If reactions are performed at room temperature, a disposable polypropylene syringe can be used after incorporation of a polyethylene frit. For reactions at higher temperature, fritted glass reactors are more appropriate. If a strongly inert atmosphere is necessary, special vessels that are closed with a septum are available.
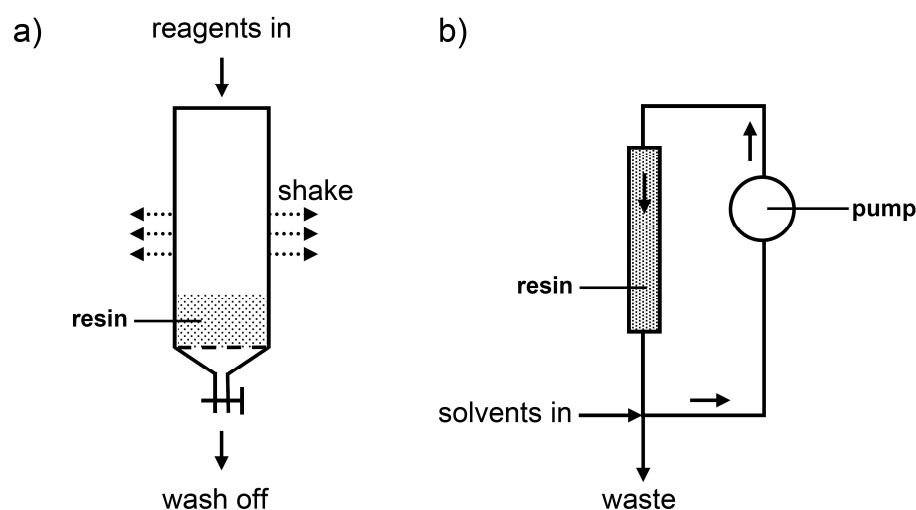
Figure 9: Batchwise and continuous-flow solid phase synthesis [21]**.** In the batchwise process (a), the reaction vessel containing loose resin beads and reagents in solution is shaken, whereas in the continuous-flow synthesis, the reaction mixture is continuously pumped through immobilized compact resin beads.

Solid phase being often used for multiple steps synthesis, the completion of each reactional step is crucial. If this one is only good and not excellent, the overall yield can be very low after only several steps (Table 1, Entry 3), resulting in mixtures of products. The necessity to purify the product after cleavage from the resin lowers the benefit of SPS.

To maximize the success of each step and thus the overall yield (Table 1, Entry 1), reagents are introduced in high excess and at high concentrations. When reaction conditions developed for solution synthesis are adapted for on the solid phase, longer

reaction times are often required. The temperature can also be enhanced and each step can be repeated if necessary.

Entry	Yield per step	Number of steps				
		2	4	6	8	10
1	99	98	96	94	92	90
2	90	81	66	53	43	35
3	80	64	41	26	17	11

Table 1: Overall yield of a multiple step synthesis as a function of the yield per step and the number of steps [3]**.** To become a correct overall yield, the yield of each step has to be excellent.

II.2.1.3. Analysis of reactions on solid support

Characterization of intermediates is quite difficult in SPS. Nevertheless, some transformations can be followed by studying the resin beads.

Colorimetric methods can be used for the detection of specific functional groups. The most common is the Kaiser test for detection of primary amines. The chloranil test can also be used and enables furthermore the detection of secondary amines.

Spectroscopic methods are also very helpful: IR (infra-red) spectra can be recorded directly from polymer beads with a special ATR-IR equipment (attenuated total reflection-IR); standard NMR spectra in gel-phase can be measured, but give interesting results only for nuclei with strong chemical-shift dispersion (^{13}C, ^{15}N, ^{19}F, ^{31}P) and are not applicable for ^{1}H. For this purpose, the magic-angle spinning (MAS) NMR can be used, but requires also special equipment and needs optimization before obtaining high quality ^{1}H-NMR spectra.

Another standard way to analyze reaction intermediates is to cleave a small amount of the resin and analyze it per HPLC, LC-MS or NMR.

II.2.2.   Materials for solid phase synthesis

II.2.2.1. Properties of polymer beads

The most common support for solid phase synthesis is a copolymer of styrene and divinylbenzene (1–2%) (Figure 10). Cross-linked polystyrene is prepared by radical

polymerization of suspensions of styrene and divinylbenzene. The size of the obtained beads (0.04–0.15 mm) is controlled by addition of surfactants, adjustment of the solvent and stirring conditions.
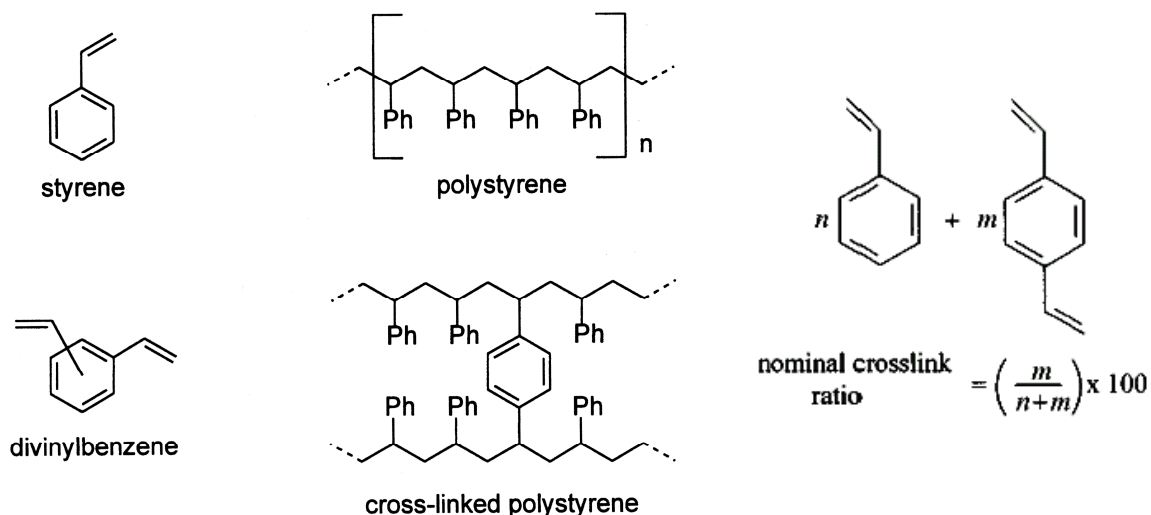
Figure 10: Structure of linear and cross-linked polystyrene [3]

Solvents used for reactions on solid support should enable a good swelling of the resin, so that the bead size increases and the reagents penetrate the pores more efficiently (Figure 11). For polystyrene, a hydrophobic and polarizable material, the swelling is strong in dipolar aprotic solvents like chloroform, dichloromethane, tetrahydrofurane or dimethylformamide, but very poor in methanol or water. Consequently, the diffusion of lipophilic reagents is generally fast, whereas ionic reagents are diffused slowly.
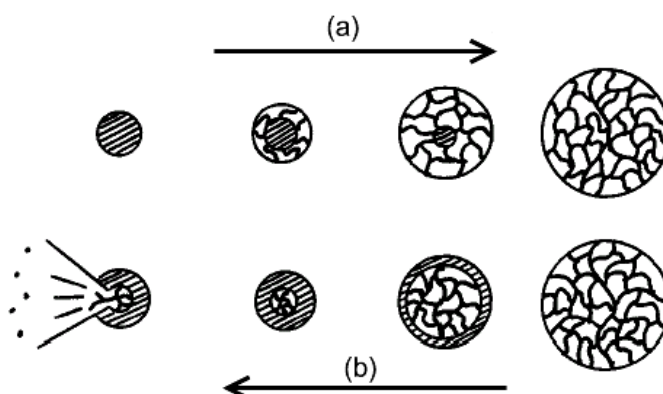
Figure 11: Solvent response of gel-type resins. In a good solvent, an expanded gel is formed (a), whereas upon addition of a bad solvent, contraction of swollen gel can result in osmotic shock (b).

Another important point by the choice of the solid support is the chemical stability of the polymer. Cross-linked polystyrene tolerates a broad range of reaction conditions including strong acids, bases, and weak oxidants. All reactions that affect benzylic positions are, however, inapplicable.

Although the first solid phase support, the polystyrene divinylbenzene copolymer, is still mainly used today, other polymers have been developed with diversified physicochemical properties. Polyethylene glycol-polystyrene graft polymers (e.g. Tentagel) are more mobile than cross-linked polystyrene, so that the created environment is more likely to solution phase synthesis, but they are less stable under strong acidic conditions or high temperature. Soluble or cross-linked insoluble polyethylene glycol, more hydrophilic polyacrylamides or polyacrylamide–PEG copolymers are also used for solid phase synthesis with variable chemical and mechanical stabilities. For the solid phase synthesis of oligonucleotides, "controlled pore glass" (a silica gel with large pore size) or the polysaccharides sephadex and sepharose have also proven to be efficient supports.

II.2.2.2.   *A broad range of linkers for various reactions*

To enable synthesis, a polymer has to be derivatized with suitable functional groups. The introduction of a so-called "linker" can occur either before the polymerization process by copolymerization of functionalized monomers, or after that, by chemical transformation of the unfunctionalized polymer.

Many various linkers are available, enabling an optimal choice for each application. An appropriate linker should enable an easy coupling of the first building block, it should be easily cleavable without affecting the rest of the molecule, and remain nevertheless stable during the whole synthesis process.

In SPPS (solid phase peptide synthesis) amino acids are attached to the solid support with their carboxylic acid function. Different linkers enable cleavage of this bond under various conditions (Table 2). The acid labile Wang resin is cleaved by 95% TFA delivering a fully deprotected peptide acid, whereas under the same conditions the Rink amide resin delivers a peptide amide. The hyperacid labile 2-chlorotrityl chloride resin is already cleaved by 1–5% TFA or even milder acids [22,23], what enables the recovery of a peptide with still protected side chains and a carboxylic acid at the *C-*

terminus. To obtain a protected peptide amide, the Sieber amide resin can be used. In a similar way, there are also base-cleavable as well as photocleavable linkers, which can be coupled to carboxylic acids or other functional groups.
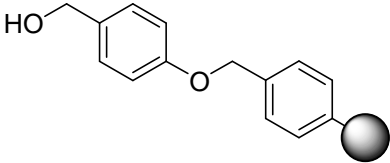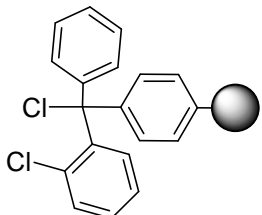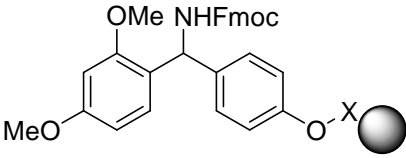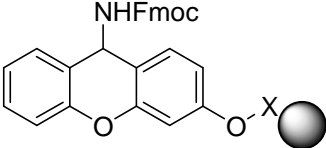
Entry	Linker name	Linker structure	Cleavage conditions
1	Wang resin		90–95% v/v TFA, 60–120 min
2	2-Chlorotrityl chloride resin		90% v/v TFA Ac-OH/TFE/DCM (1:8:8), 15–60 min 1–5% TFA in DCM, 1 min 20% HFIP in DCM, 3–5 min
3	Rink amide resin		50% TFA in DCM, 60 min 90–95% v/v TFA, 60 min
4	Sieber amide resin		2% TFA in DCE, 5–10 min 1–5% TFA in DCM, 5–15 min 90% v/v TFA

Table 2: Linkers commonly used for SPPS. X: CH_2-; CH_2(C=O)-

II.2.3.   Solid phase peptide synthesis (SPPS)

II.2.3.1.   Boc strategy

The synthesis strategy developed by Merrifield[17,24] was first performed on chloromethyl polystyrene resin (Figure 12a) with Cbz as protecting group for amino acids (Figure 12c). As this group requires very strong acidic conditions for the deprotection– hydrogen bromide in glacial acetic acid – which are also affecting the peptide-benzyl linkage, alternatives were investigated.

A more efficient protecting group was found: *tert*-butyl carbamate (Boc)[19,20] (Figure 12d): a few minutes treatment with 1 N HCl in acetic acid are sufficient for its deprotection and peptide loss by undesirable cleavage is thereby avoided[25]. Boc can

also be cleaved with TFA. Furthermore, this treatment reduces peptide aggregation by destroying secondary interactions. The polymer support currently used for Boc SPPS is 4-hydroxymethyl-phenylacetamidomethyl polystyrene (PAM resin, Figure 12b), which avoids side reactions and undesirable cleavage of the labile peptide-benzyl linkage.
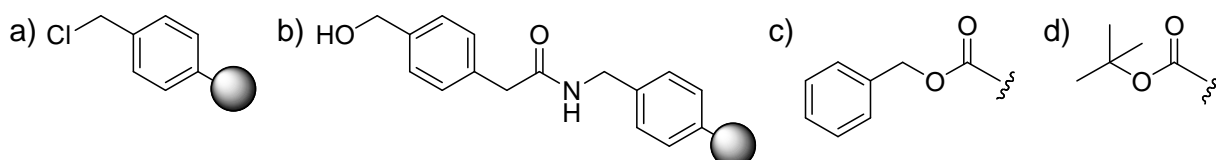
Figure 12: Resins and amine protecting groups for Boc SPPS. a) Chloromethyl resin; b) PAM resin; c) Cbz protecting group; d) Boc protecting group.

The principle of Boc strategy is briefly summarized on Scheme 2. After removal of the Boc protecting group under acidic conditions and deprotonation of the ammonium salt with DIPEA, an amino acid is coupled. The sequence is then repeated for each subsequent amino acid. The resulting peptide is cleaved from the resin by HF and the protecting groups of the side chains are also removed. Additional treatment with TFA after each coupling step enables a reduction of peptide aggregation. The main disadvantage of this synthesis way is the use of highly toxic HF. Therefore, the Fmoc strategy is often preferred.
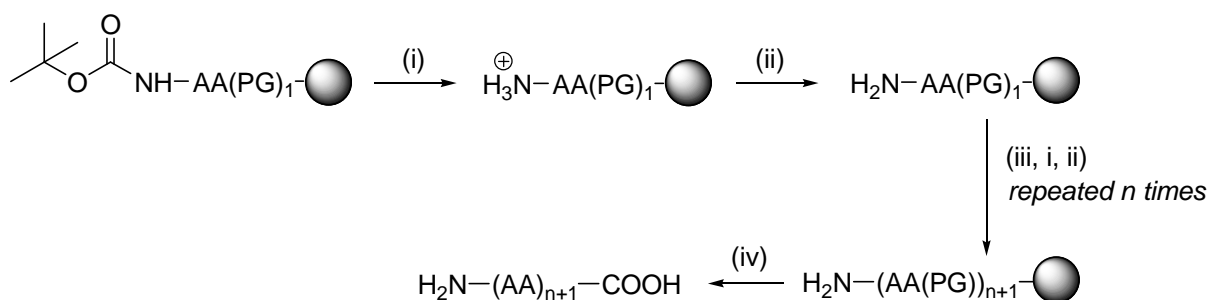
Scheme 2: Principle of Boc SPPS. Conditions: (i) TFA / DCM; (ii) DIPEA / DCM; (iii) (Boc-AA$_i$-O)$_2$O; (iv) HF / anisole.

II.2.3.2. Fmoc strategy

After the 9*H*-fluorenylmethoxycarbonyl (Fmoc) group has been presented by Carpino for the protection of amines [26] and the synthesis of dipeptides in solution [27], Atherton introduced this strategy for the synthesis of peptides on solid support [28]. The

superiority of this method over Merrifield's strategy consists in the mild cleavage conditions: the acid-labile side chain protecting groups can be easily removed, orthogonally to the base-labile Fmoc group [29]. Another advantage is the suppression of one step for each amino acid (neutralization of the ammonium salt following the Boc deprotection).

The principle of Fmoc strategy is briefly summarized on Scheme 3. After removal of the Fmoc protecting group under basic conditions, an amino acid is coupled. The sequence is then repeated until the whole peptide has been synthesized. The resulting peptide is cleaved from the resin by TFA with concomitant side chains deprotection. If, however, the side chains have to remain protected, milder conditions can also be used for the cleavage from the resin (Table 2).
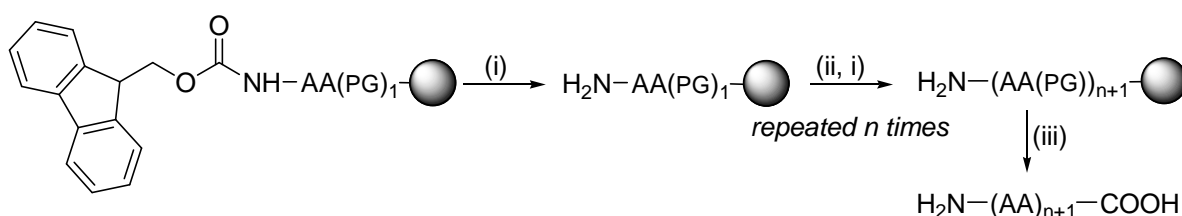
Scheme 3: Principle of Fmoc SPPS. Conditions: (i) Piperidine / DMF; (ii) FmocAA$_i$ / HBTU / DIPEA; (iii) TFA / H$_2$O.

Different resins with various degrees of lability have been developed for Fmoc SPPS. They enable either the preparation of peptide acids or of peptide amides (as already shown on Table 2). Many protecting groups have also been developed; the most common are shown on Figure 13.
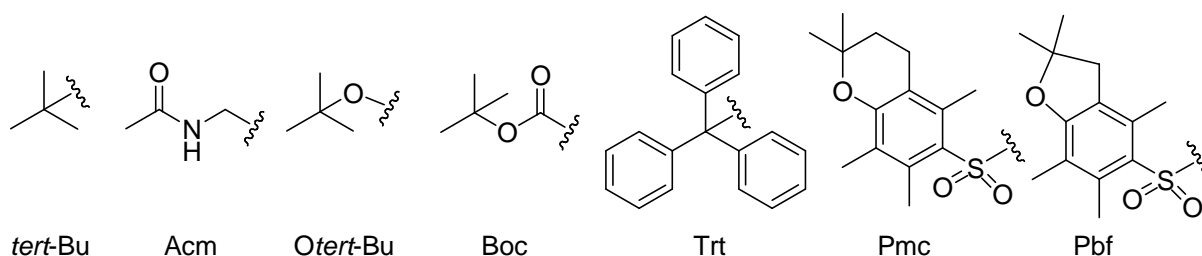
Figure 13: Side chain protecting groups for Fmoc SPPS

Fmoc strategy is the most popular approach for peptide synthesis. A lot of sequences, including so called "difficult peptides" could be successfully obtained. However, the major problem of this strategy remains peptide aggregation.

II.2.3.3. Coupling methods for the acylation of amines

Amide bond formation is one of the most investigated reactions on solid phase. Many protocols have been developed to obtain a quantitative, racemization-free reaction.

a. Use of isolated acylation agents

One possible way is the use of isolated acylating agents such as acyl chlorides [30], fluorides [31], symmetric anhydrides or *N*-hydroxysuccinimidyl esters (Scheme 4).
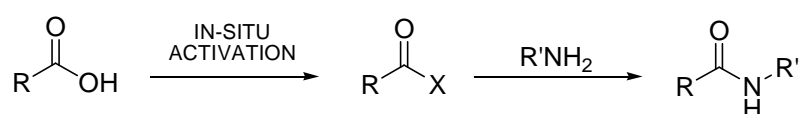
Scheme 4: Carboxylic acids acylation with isolated acylation agents. X: leaving group.

The most reactive are acyl halides, but they are therefore also not very selective. The addition of a base is furthermore required, what enhances racemization by oxazolinone formation, in particular in the case of acyl chlorides (Scheme 5). Therefore, the use of acyl fluorides should be limited to difficult couplings such as the incorporation of sterically hindered residues [31,32].
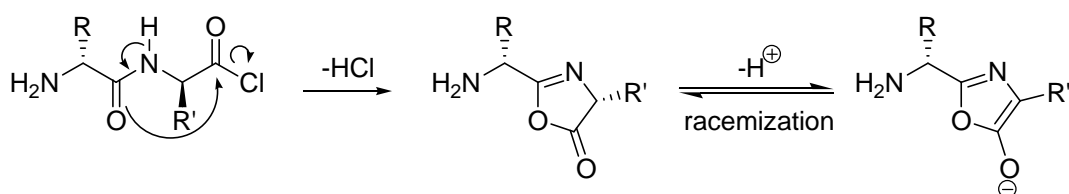
Scheme 5: Racemization of activated *N*-acyl amino acids under basic conditions

Symmetric or mixed anhydrides (e.g. with pivalic or 2,6-dichlorobenzoic acid) are an interesting alternative. They are less reactive, but still suitable for difficult couplings.

Another way for the acylation of amines is in-situ activation, mainly using carbodiimides, phosphonium salts, and uronium salts. Some other reagents like MSNT – generating mixed carboxylic sulfonic anhydrides – are also sporadically used.

b. Activation of acids with carbodiimides

Carbodiimide activation was introduced in 1955 by Sheehan and Hess [33]: DCC (*N,N'*-dicyclohexylcarbodiimide) has been the standard reagent for many years, before

being gradually replaced by the easy-to-handle liquid DIC (diisopropylcarbodiimide) or the water-soluble EDC (N-ethyl-N'-(3-dimethylaminopropyl)carbodiimide) that can be easily removed by extraction (Figure 14a).
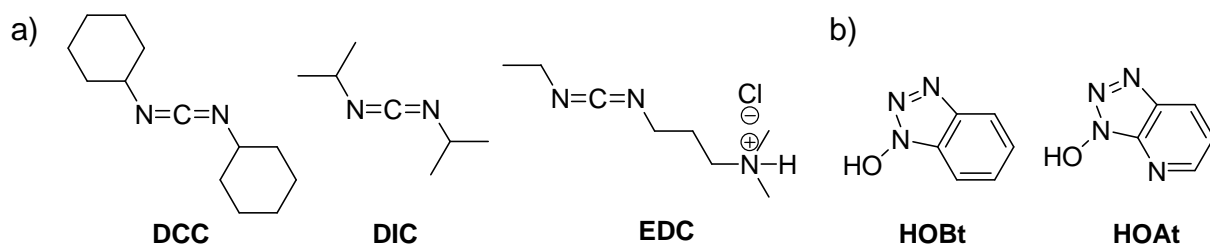
Figure 14: Carbodiimides for carboxylic acid activation (a) and adjuvants for active ester formation (b).

The mechanism of carbodiimide activation is illustrated on Scheme 6:
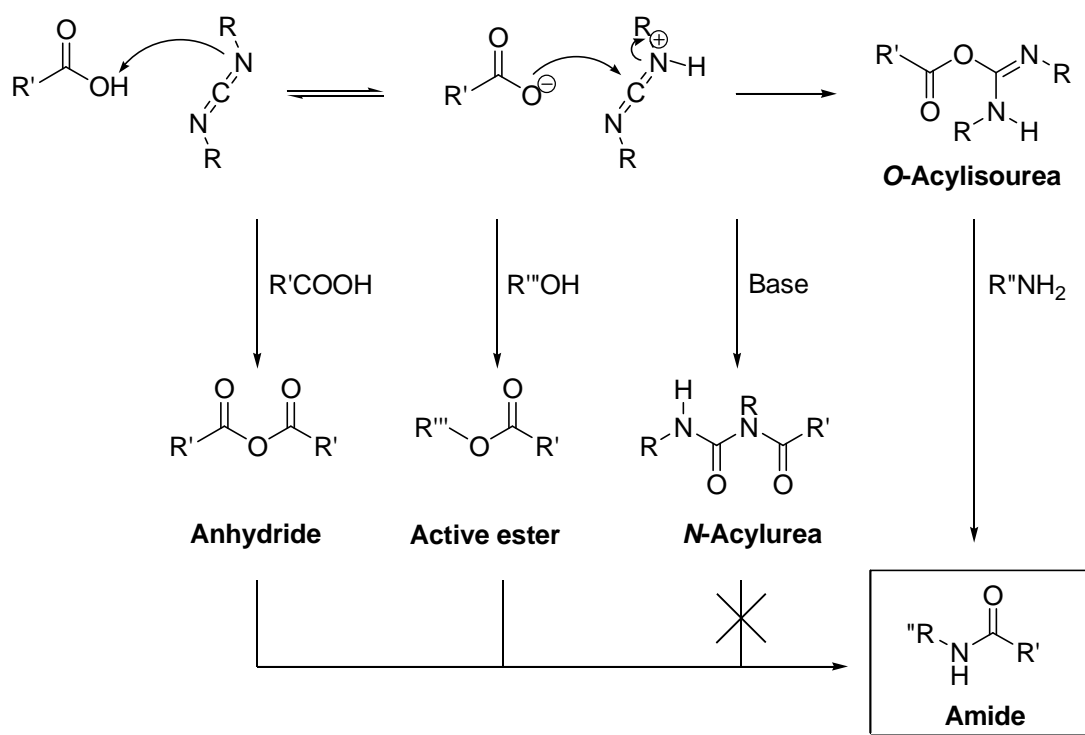
Scheme 6: Mechanism of carboxylic acid activation with carbodiimides

An O-acylisourea is first generated by attack of the carboxylate at the carbodiimide. In the absence of any other nucleophile, this intermediate can react with another carboxylic acid molecule to yield an anhydride. To prevent the rearrangement from active O-acylisourea to stable N-acylurea that can occur faster than anhydride

formation under basic conditions, an excess of carboxylic acid is used. In the presence of additives such as HOBt (1-hydroxybenzotriazole) or HOAt (4-aza-3-hydroxybenzotriazole) (Figure 14b), an active ester is preferably formed. Finally, after attack of the amine at the active ester, the expected amide is obtained. The auxiliary nucleophiles are not only beneficial as excellent leaving group providers, but also as proton acceptors, what enhances the reaction rate even more.

Therefore, in SPPS it makes a difference whether the amine or the acid is bound to the polymeric support. Indeed, only the reagent in solution can be added in excess. Thus, the carboxylic acid should be used in solution and the free amine should be attached to the resin. In the particular case of reactions involving acids or amines containing a leaving group, no HOBt should be used to prevent formation of substitution side products.

c. _Activation of acids with phosphonium salts_

The most common triaminophosphonium salts [34] are shown on Figure 15. BOP (benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate) was one of the first phosphonium salts, but leads to the formation of mutagenic HMPA (hexamethylphosphoramide). It was therefore replaced by its less hazardous analog pyBOP ((benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate).
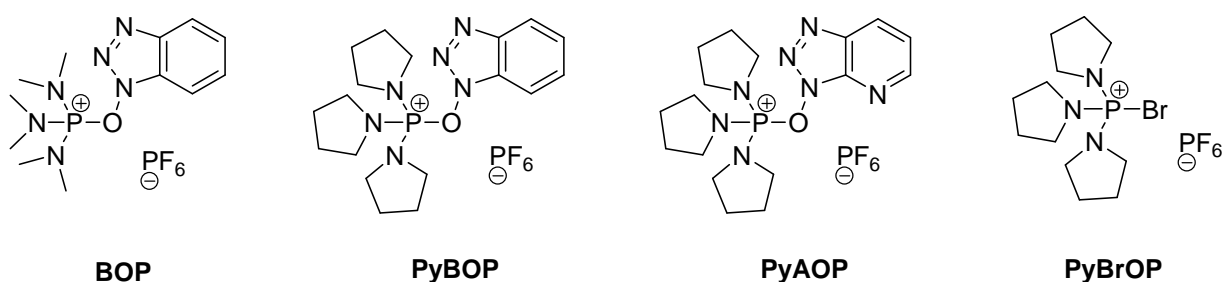
Figure 15: Triaminophosphonium salts for carboxylate activation

The mechanism is illustrated on Scheme 7. The acyloxyphosphonium salt formed by reaction between the carboxylic acid and the phosphonium salt leads either directly to an amide or, depending on the reaction conditions and the nature of the counterion, through formation of a symmetric anhydride or an active ester.
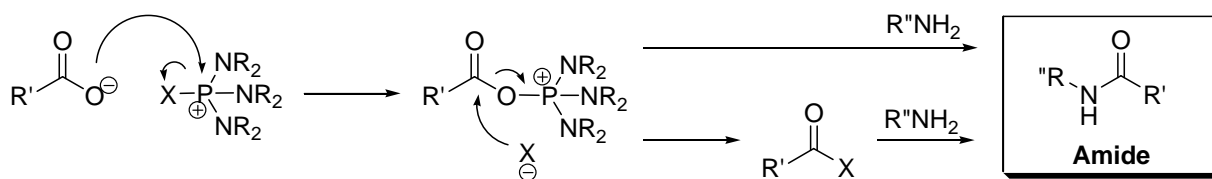
Scheme 7: Mechanism of the carboxylate activation with phosphonium salts. X: leaving group.

d. Activation of acids with uronium salts

By activation of a carboxylic acid with an uronium salt (Figure 16) an *O*-acylisouronium salt is first formed. This one is more reactive than the *O*-acylisourea formed by activation with a carbodiimide and cannot undergo intramolecular rearrangement. Then, it can be converted to a HOBt ester, a pentafluorophenyl ester, or an acylhalide.
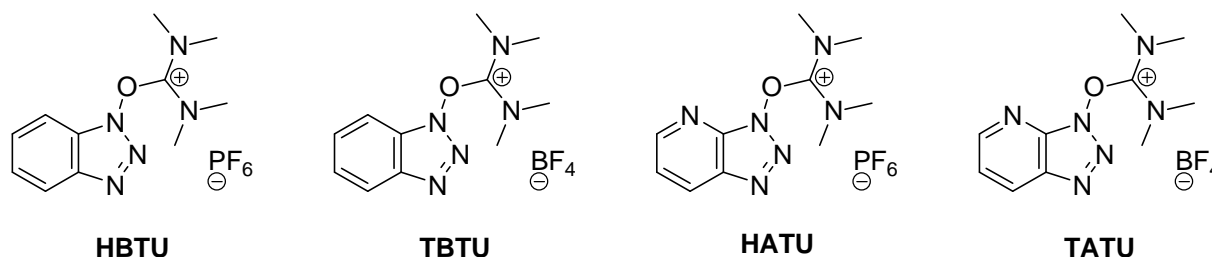
Figure 16: Uronium salts for carboxylate activation

Thus, uronium salts are powerful reagents, which are used when traditional activation methods remain unsuccessful. A possible side reaction is nevertheless the guanidine formation [35,36]. In particular in the case of hindered carboxylic acids or acids involved in a cyclization step, the activation is slow. The uronium salt can therefore react with the *N*-terminal amine, leading to a guanidine derivative. As a consequence, the peptide chain elongation is early terminated.

II.2.4. Solid phase organic synthesis (SPOS)

II.2.4.1. Application of the SPS concept to the synthesis of small organic molecules

Although solid phase synthesis rapidly became the method of choice for peptide synthesis, its use for the synthesis of non-peptidic compounds [37] remained marginal until the nineties.

The major difficulty of SPOS is the complexity of the transfer of reactions already known in solution to the solid support. For each reaction type, the polymer and the linker have to be accurately selected and the reaction conditions such as solvent, stoechiometry, reaction time and temperature have to be adapted. Reactions on solid phase are indeed slower than in solution, due to the heterogeneity of the system. The reagents in solution need longer diffusion times to penetrate the polymer beads. The introduction of reagents in excess or an elevation of the temperature can be helpful to accelerate the reaction [38].

II.2.4.2.   Advantages of reactions on solid support

Reactions conducted on solid support enable the isolation of pure products without purification. This is of particular interest for parallel synthesis of small molecules libraries. Reactions are mostly carried out with starting material on the polymer and reagents in solution, but can also be done with reagents on solid support and starting material in solution. The benefit of this method is illustrated by the Wittig reaction: triphenylphosphine oxide is a side product quite difficult to separate in solution from the Wittig product. Binding the reagent to a solid support enables an easy removal, once the reaction is completed.

Another technique is the use of polymers as scavengers or fishing tools. After carrying out the reaction in solution, polymer beads are introduced into the reaction mixture either to remove side-products or to fish out the desired product.

II.2.4.3.   Limitations of reactions on solid support

SPOS cannot be applied for all reactions. Some limitations are due to the intrinsic nature of the polymer beads. For example, the use of polar solvents like methanol or water is not possible with polystyrene resins because of bad swelling properties. The use of a different resin (e.g. Tentagel) can circumvent this problem if all other reaction steps are also compatible with this resin. Heterogeneous reactions are not possible since solid reagents cannot be separated from the resin beads and removed with simple washing steps. The economical aspect has also to be considered. Finally, reaction monitoring [39] is not as easy as in solution phase synthesis. All these

particularities limit an extended use of solid-supported synthesis, which is otherwise a very efficient tool.

II.3.   Fluorescence-based ligand discovery

II.3.1.   What is fluorescence?

II.3.1.1.   Introduction

Light emission consecutive to photoexcitation called "photoluminescence" is used in many domestic products, e.g. "glow-in-the-dark" toys, synthetic textile fibers, inks or plastics with daylight fluorescent pigments. It is also used in science, e.g. in biology or medicine, for the study of cellular structures or for diagnosis purposes.

II.3.1.2.   Jablonski energy diagram
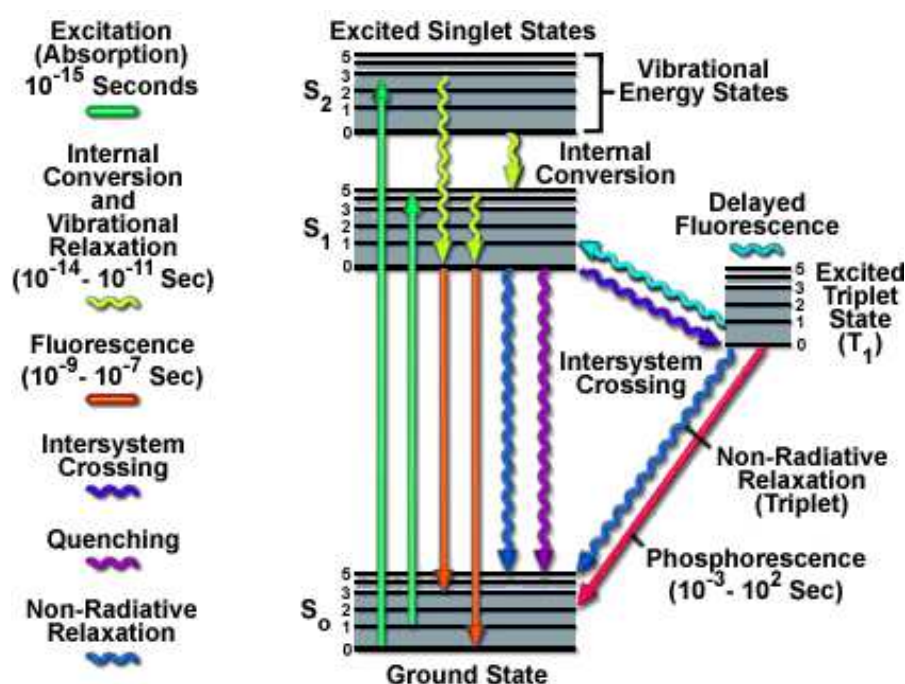
Figure 17: Jablonski energy diagram. Each excited singlet (S_i) or triplet (T_i) state consists of many vibrational energy levels. Transitions between different energy levels can occur by different radiative or non-radiative processes [41].

There are two kinds of photoluminescence: fluorescence and phosphorescence. They can be defined via the Jablonski diagram (Figure 17), named after the polish physicist Aleksander Jabłoński (1898–1980). This representation illustrates the process which occurs during light absorption and emission [40].

After absorption of a photon, a molecule is excited and transferred from its ground state S_0 to a higher energy electronic state. According to the Franck-Condon principle, this transition occurs "vertically", without changes in the nuclei spin orientation. This is due to the fact that electronic transitions are essentially instantaneous (10^{-15} s) compared with the time scale of nuclear motions. The reached electronic state is not stable and the molecule returns to its ground state either by radiative or non-radiative processes.

II.3.1.3. *Non-radiative relaxation processes*

Internal conversion, intersystem crossing, and vibrational relaxation are the non-radiative relaxation processes. By internal conversion, a molecule is transferred from an excited state to another excited state with lower energy and the same spin. When the transition occurs between two different spin states, it is called intersystem crossing. However, the most common way for a molecule to return to its ground state is vibrational relaxation. This occurs quickly ($< 10^{-12}$ s) and is enhanced by quenching – collision with other molecules associated to energy exchange.

II.3.1.4. *Radiative relaxation processes: fluorescence and phosphorescence*

In some cases, the relaxation is associated with light emission. If this transition occurs between energy levels with the same spin state, it is called fluorescence: usually, the transition happens between the ground vibrational level of the first excitation electronic state $S_{1,0}$ and the various vibrational levels of the electronic ground state $S_{0,i}$ (Figure 17, in red). If the molecule is excited to higher electronic and vibrational energy levels, the excess energy is quickly dissipated by non-radiative processes. As a consequence, fluorescence emission spectra are usually independent of the excitation wavelength; this is known as Kasha's rule [40,42].

Therefore, the energy of emitted light is lower than the energy necessary for excitation. Thus, according to the Planck's law linking energy and wavelength

(Equation 1), the wavelength of the fluorescence emission is higher than the excitation wavelength, and the resulting difference is called Stokes' shift.

When the molecule undergoes intersystem crossing prior to deactivation, the electron spins are parallel and the molecule is in a triplet state. The transition to the ground state (T_1 to S_0), taking place between two different spin states, is called phosphorescence (Figure 17, in pink).

$$E = h \cdot v = \frac{h \cdot c}{\lambda}$$

Equation 1: Planck's law. E: energy / J; h: Planck's constant / J·s; v: frequency / Hz; c: velocity of light / m·s^{-1}; λ: wavelength /m.

Fluorescence and phosphorescence are two competing processes. Phosphorescence is a very slow phenomenon (phosphorescence lifetimes are about 10^{-4} s to even some hours, whereas fluorescence lifetimes only 10^{-9} to 10^{-5} s). Furthermore, it is rarely observed in solution at room temperature because of the multiple other faster deactivation ways. Thus, statistically, fluorescence is much more likely to happen than phosphorescence.

II.3.2. Characteristics of fluorophores

II.3.2.1. Introduction

Beside the above mentioned absorption and emission spectral properties, fluorescence quantum yield and lifetime are the most important characteristics of a fluorophore. Regarding these criteria the adequate fluorophore can be selected for a particular application. It can be either a molecule naturally present like tryptophan to study proteins, or an extrinsic fluorophore that has to be incorporated to the system under investigation.

II.3.2.2. Fluorescence quantum yield

Upon excitation the photon population is disseminated on different energy levels. Then, each photon returns on its own way to the ground state. Per definition, the fluorescence quantum yield ϕ_{fl} is the ratio of the number of photons emitted by

fluorescence to the number of photons absorbed. Considering the rate constants k_i of all processes depopulating the excited state, ϕ_{fl} can also be expressed as a function of these constants (Equation 2). If the radiationless decay rate is minimal in comparison to the radiative decay rate, the quantum yield is close to 1.

$$\phi_{fl} = \frac{N_{photons,\,em}}{N_{photons,abs}} = \frac{k_{em}}{k_{em} + k_{n.r.}}$$

Equation 2: Definition of the quantum yield of fluorescence. $N_{photons,\,em}$: number of photons emitted by fluorescence; $N_{photons,\,em}$: number of photons absorbed; k_{em}: rate constant of deactivation through fluorescence; $k_{n.r.}$: rate constant of deactivation through all non-radiative processes.

II.3.2.3. Fluorescence lifetime

The fluorescence lifetime τ is defined as the average time spent by a molecule in the excited state prior to deactivation. The behavior of a population of molecules can be described by an exponential decay, and the fluorescence lifetime corresponds to the time for which the number of excited molecules decreases of 1/e (or 36.8 %), as illustrated with Equation 3.

$$N^*(t) = N_0 \cdot e^{-t/\tau}$$
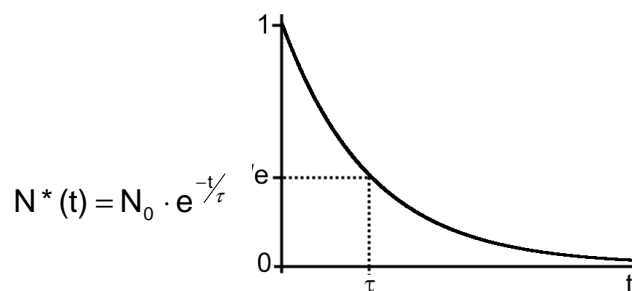
Equation 3: Fluorescence lifetime. $N^*(t)$: number of excited molecules at time t; τ: fluorescence lifetime.

II.3.2.4. Intrinsic and extrinsic fluorophores

a. Intrinsic fluorophores

Intrinsic fluorophores are those which occur naturally in biomolecules. For example, proteins are self-fluorescent through their aromatic amino-acids: tryptophan, tyrosine, and phenylalanine.

Entry	Species	λ_{ex} / nm	λ_{em} / nm	ϕ	τ / ns
1	Phenylalanine	260	282	0.02	6.8
2	Tyrosine	275	304	0.14	3.6
3	Tryptophan	295	353	0.13	3.1

Table 3: Fluorescence parameters of aromatic amino acids in water at neutral pH

Phenylalanine (Table 3, Entry 1) is the weaker fluorophore compared to the two other fluorescent amino-acids, with a quantum yield of only 0.02 [40]. Phenylalanine emission is therefore only studied in the absence of Tyrosine and Tryptophan, which is quite a rare case.

Tyrosine fluorescence is often quenched in native proteins due to its interaction with the peptidic chain. This property can be used to study protein stability: indeed, an increased tyrosine fluorescence emission is the sign of protein denaturation.

The major source of fluorescence emission in proteins is tryptophan (Table 3, Entry 3): spectral shifts can be observed upon binding of small ligands, proteins, or upon exposure to the aqueous phase. Furthermore, tryptophan emission is quenched by iodide, acrylamide, disulfide groups, or electron deficient residues. Due to its high sensitivity to the local environment, tryptophan is a very good reporter group for protein conformational changes.

Another widespread example of intrinsic fluorescence in proteins is the green fluorescent protein (GFP) [43] from *coelenterates.* The gene for GFP has been isolated and some mutants having an increased fluorescence and different absorption and emission maxima have been generated.
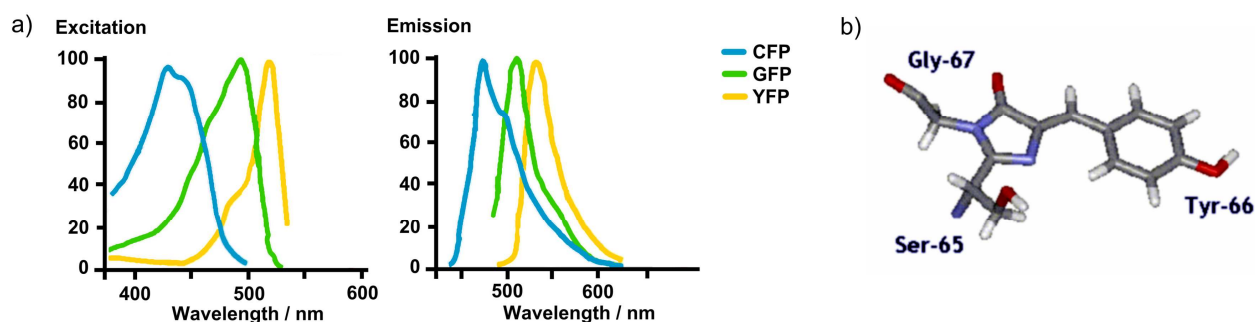
Figure 18: (a): Normalized excitation and emission spectra of the GFP and some of its variants; (b): The fluorophore center of GFP [44]. CFP: Cyan F P; YFP: Yellow F P.

The spectra of CyanFP, GreenFP and YellowFP are shown on Figure 18a. In this case, fluorescence results from an internal Ser-Tyr-Gly sequence post-translationally modified to a 4-(*p*-hydroxybenzylidene)- imidazolidin-5-one structure (Figure 18b). The advantage of GFP over other fluorescence labeling methods is that labeling occurs already on the genetic level, by introduction of fusion proteins, thus enabling applications in living cells.

b. Extrinsic fluorophores

Fluorophores that do not exist naturally and have first to be introduced in cells or organisms are called extrinsic fluorophores. Many different types of extrinsic fluorophores have been developed; some commonly used molecules are shown in Table 4 and Figure 19.
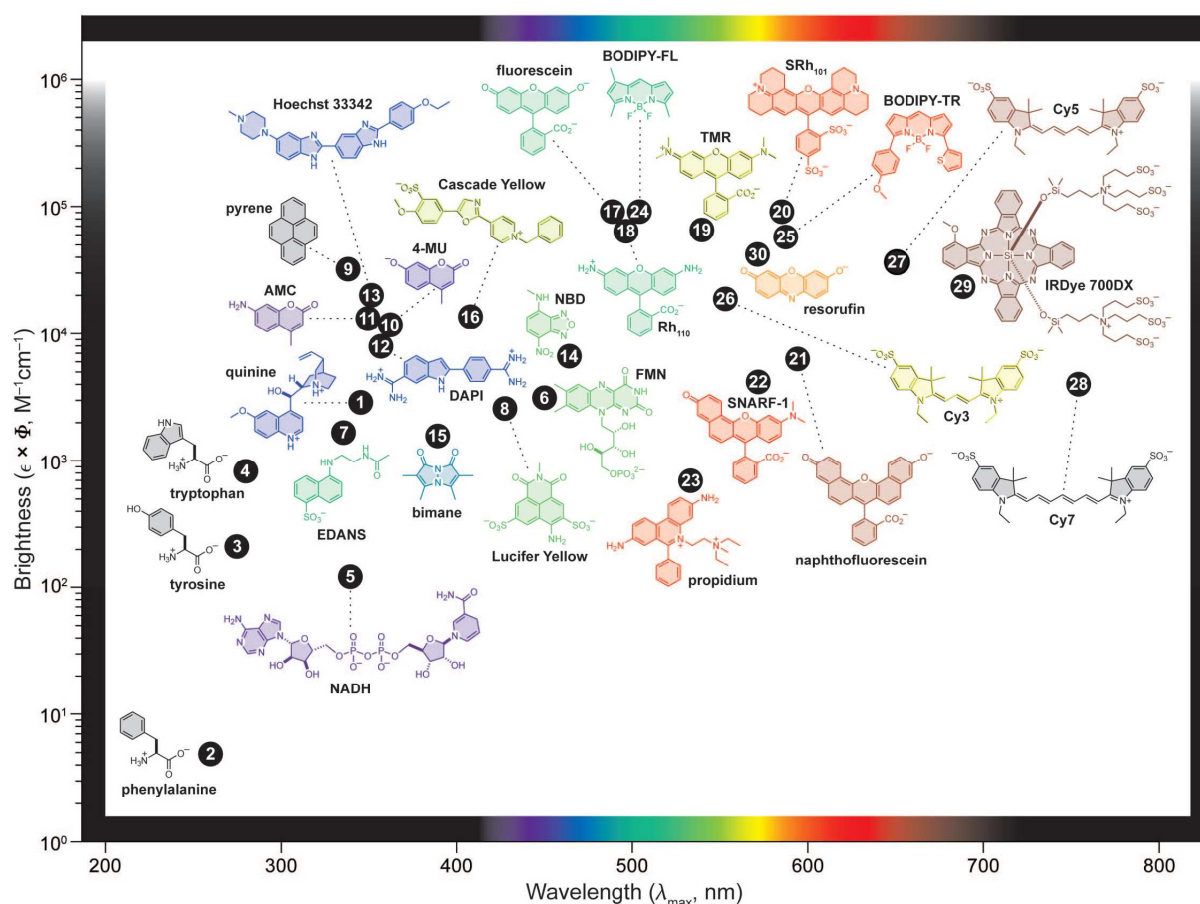
Figure 19: Extrinsic fluorophores. Many different fluorophores have been developed, covering the whole visible, ultra-violet and near infra-red spectral range [45].

Entry	Fluorophore	$\lambda_{absorption}$ / nm	ϵ / $M^{-1} \cdot cm^{-1}$	$\lambda_{emission}$ / nm	ϕ	τ / ns
1	DNS-Cl*	340–350	4 300	510–560	0.1–0.3	10–15
2	FITC*	492	72 000	516–525	0.3–0.85	4.5
3	TRITC*	535–545	107 000	570–580	–	2.0

Table 4: Fluorescence spectral properties of commonly used extrinsic probes [40]**.** *DNS-Cl: dansyl chloride; FITC: fluorescein isothiocyanate; TRITC: tetramethylrhodamine isothiocyanate; ** : not mentioned.

The advantage of extrinsic fluorophores upon intrinsic fluorophores is the possibility to modify some characteristics like fluorescence quantum yield, or absorption and emission wavelengths, by specific changes on the structure [46-48].

II.3.3.   Overview of fluorescence methods

II.3.3.1.   Fluorescence imaging

Microscopy is a fundamental tool in biological research and medical diagnosis. Unfortunately, living matter is not very suitable for direct optical examination. For decades specimens had to be chemically fixed before observation, until the emergence of novel imaging methods. Using convenient fluorophores, a specific labeling of distinct cell components can be realized (Figure 20). Thus, fluorescence microscopy is an ideal technique for the observation of fixed as well as living biological samples [49,50], allowing a selective detection with a good signal-to-background ratio [51].
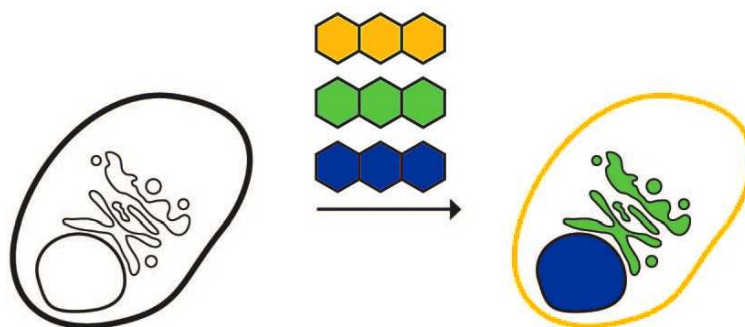
Figure 20: Specific labeling of cell components [45]

Traditional fluorescence microscopy is excellent for thin samples[52]. In thicker samples, excited photons from out-of-focus fluorophores interfere with the observation. To avoid this, confocal scanning microscopy[51] has been developed. This technique restricts photodetection to light emerging from the focal point. As a large part of the emitted photons are excluded, the excitation signal has to be increased to enable an efficient detection. The problem of this high intensity is, however, a rapid loss of the fluorescent properties by so-called photobleaching. Indeed, after a fluorophore undergoes some thousands of excitation-emission cycles, it loses the ability to fluoresce. This is due to photon-induced chemical damage and covalent modification by interaction with molecules of the environment and is probably occurring after transition from an excited singlet state to the relatively long-lived excited triplet state.

II.3.3.2. *Fluorescence / Förster resonance energy transfer*

Fluorescence resonance energy transfer (FRET) is a useful tool for quantifying molecular dynamics such as protein-protein interactions, protein-DNA interactions, and protein conformational changes[53,53]. This process, discovered by Theodor Förster in 1946, is based on a non-radiative energy transfer between a donor and an acceptor chromophore that occurs when both are less than some nanometers (Förster distance) away from each other[54].

To study the interaction between two molecules, one of them is labeled with a donor, the other one with an acceptor fluorophore. In case of binding, the energy absorbed by the donor upon excitation is transferred to the acceptor, which releases a signal at its own emission wavelength (Figure 21a). If there is no interaction between both molecules, no energy transfer can occur and the emitted signal is detected at the wavelength of the donor (Figure 21b).

To discriminate both signals, the emission ranges of donor and acceptor should not overlap. A common example of donor / acceptor pair used in FRET assays is the fluorescein / TAMRA system with a Förster distance of 4.9 to 5.5 nm, or the DNS / FITC system with 3.3 to 4.1 nm [55].
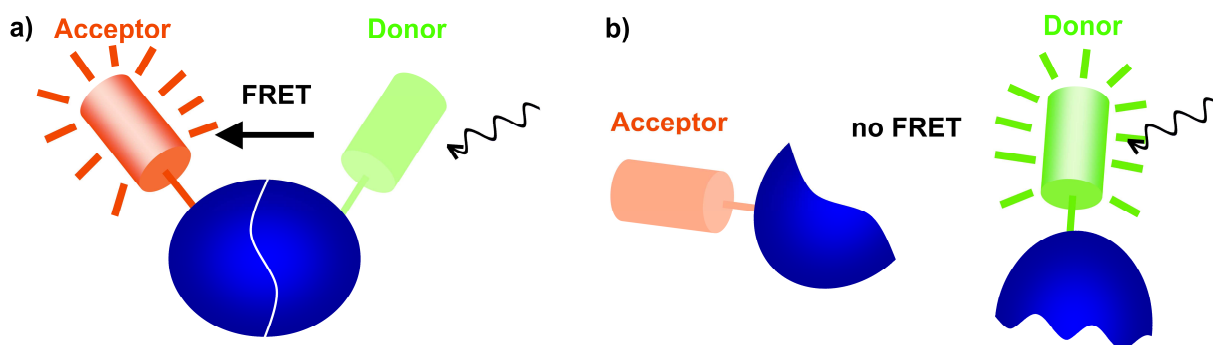
Figure 21: Principle of FRET. (a): If donor and acceptor are closer than the Förster distance, the energy absorbed by the donor is transferred to the acceptor, which emits radiations; (b): if both molecules are too far away, no energy transfer can occur and light is directly emitted by the donor.

II.3.3.3. *Fluorescence polarization*

The theory of fluorescence polarization (FP), first described by Perrin in 1926, is based on the following observation: fluorescent molecules in solution, which are excited with linear-polarized light, emit light back into a fixed direction if the molecules remain stationary during the fluorescence lifetime. This is described by the Perrin equation (Equation 4):

$$\frac{1}{p} - \frac{1}{3} = \left(\frac{1}{p_0} - \frac{1}{3} \right) \cdot \left(1 + \frac{RT}{\eta V} \tau_0 \right)$$

Equation 4: The Perrin equation. p: polarization; p_0: value of p when no depolarization takes place; R: ideal gas constant; T: absolute temperature; η: viscosity of the solvent; V: molecular volume of the fluorescent molecule; τ_0: fluorescence lifetime of the fluorophore

Considering a fluorophore with a fluorescence lifetime τ and a rotational correlation time θ_{fl} as well as the molecule to study with a rotational correlation time θ, the resulting polarization signal depends on the size of the molecule. If the molecule is large, its rotational correlation time θ is high and the molecule almost does not move during the time τ. The polarized light that is absorbed by the molecules with an adequate orientation is emitted immediately and remains polarized[56], because it is emitted from all molecules in solution in a common direction. In case of smaller molecules for which rotation and tumbling are faster (low θ), the fluorescence lifetime

- 33 -

is sufficient to obtain an almost random orientation of these molecules; the light is therefore emitted in all directions and is no more polarized.

In the study of protein-ligand interactions in an FP assay, fluorescence labeled small molecules that do not bind to the protein rotate freely in solution and emit light with almost no polarization, whereas small molecules that bind to the bigger protein deliver a polarized signal [57] (Figure 22).
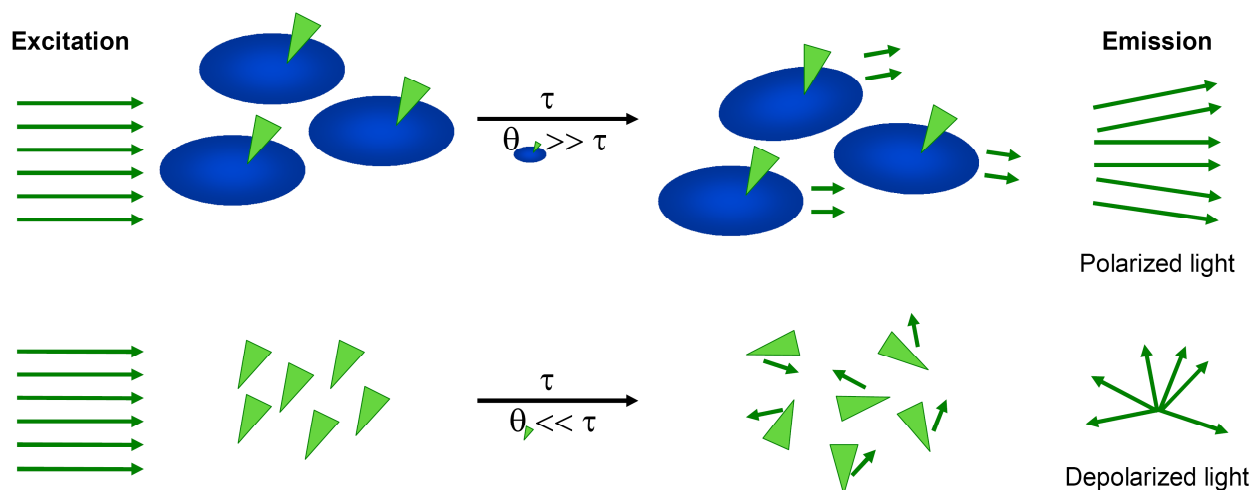
Figure 22: Principle of fluorescence polarization. τ: fluorescence lifetime; θ: rotational correlation time. All small fluorophores aligned with the polarization plane are excited at t=0. If they are bound to a large molecule (upper part), they almost do not rotate during the fluorescence lifetime ($\theta \gg \tau$), so that the emitted light remains polarized. If they are free in solution (lower part), they rotate rapidly ($\theta \ll \tau$), so that light is emitted in all directions and polarization is lost.

II.3.3.4. Fluorescence correlation spectroscopy

Fluorescence correlation spectroscopy (FCS) has been developed in the early seventies. This technique enables high-resolution spatial and temporal analysis of biomolecules at extremely low concentrations. In contrast to other fluorescence techniques, the parameter of primary interest is not the emission intensity itself, but intensity fluctuations due to phenomena like diffusion, physical and chemical reactions, or aggregation.

All physical parameters generating fluctuations in the fluorescence signal are accessible by FCS [58], for example local concentrations, mobility coefficients or characteristic rate constants of inter- or intramolecular reactions of fluorescently labeled biomolecules in nanomolar concentrations.

In the study of protein-ligand interactions, a small labeled ligand and a large non-fluorescent protein are mixed. The diffusion times of both free and bound ligand can be measured and their respective proportions can be determined, after application of the right model for data treatment (Figure 23).

However, for a significant change in the diffusion time, the mass ratio between free and bound entities should be at least eight, preferably more than an order of magnitude, due to the approximate cube root dependence of the diffusion coefficient on molecular mass.
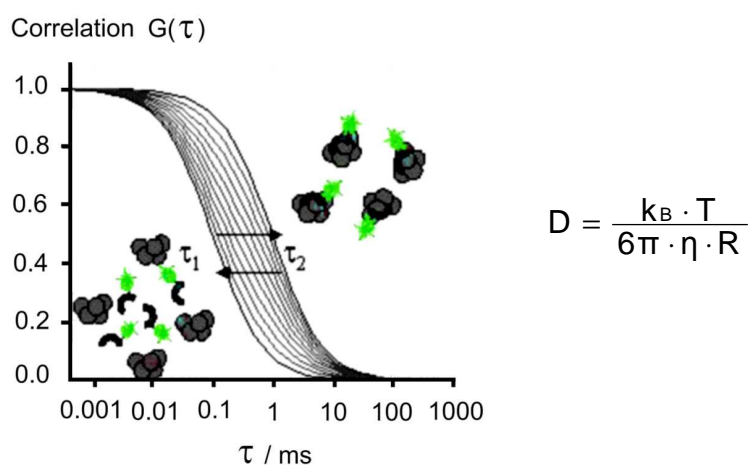
Figure 23: Changes in diffusion time of a small ligand upon binding to a heavy protein; Stokes-Einstein relationship. D: diffusion coefficient; k_B: Boltzmann's constant; T: absolute temperature; η: viscosity; R: radius of the globular molecule.

II.3.4. Fluorescence labeling methods

II.3.4.1. Labeling in solution

Fluorescein labeling in solution can be performed with different reagents such as 4(5)-fluorescein isothiocyanate (FITC), 4(5)-carboxyfluorescein-*N*-succinimidylester (CFSE), or 5-iodoacetamidofluorescein (5-IAF).

A selective labeling of the *N*-terminus of peptides is quite difficult in the presence of other amines or thiols, e.g. ε-amino groups of lysine or cysteine residues. This can be partially reached by keeping the pH near to neutral, but is not really efficient because of the low pKa difference between the *N*-terminus and the other interfering groups. Complex mixtures are often obtained.

Another labeling method used for proteins in solution is the modification of a thiol group, e.g. by coupling 5-iodoacetamide fluorescein[59]. This method is, however, limited to peptides containing at least one thiol in their sequence. It is possible to obtain almost exclusive labeling of the thiols, also in the presence of lysine side-chains, by keeping the pH near to neutral. Indeed, in this case the gap is 2 pH units. But it is still not selective to the *N*-terminus. Moreover, it cannot be really applied to synthetic scale due to the high price of this labeling reagent.
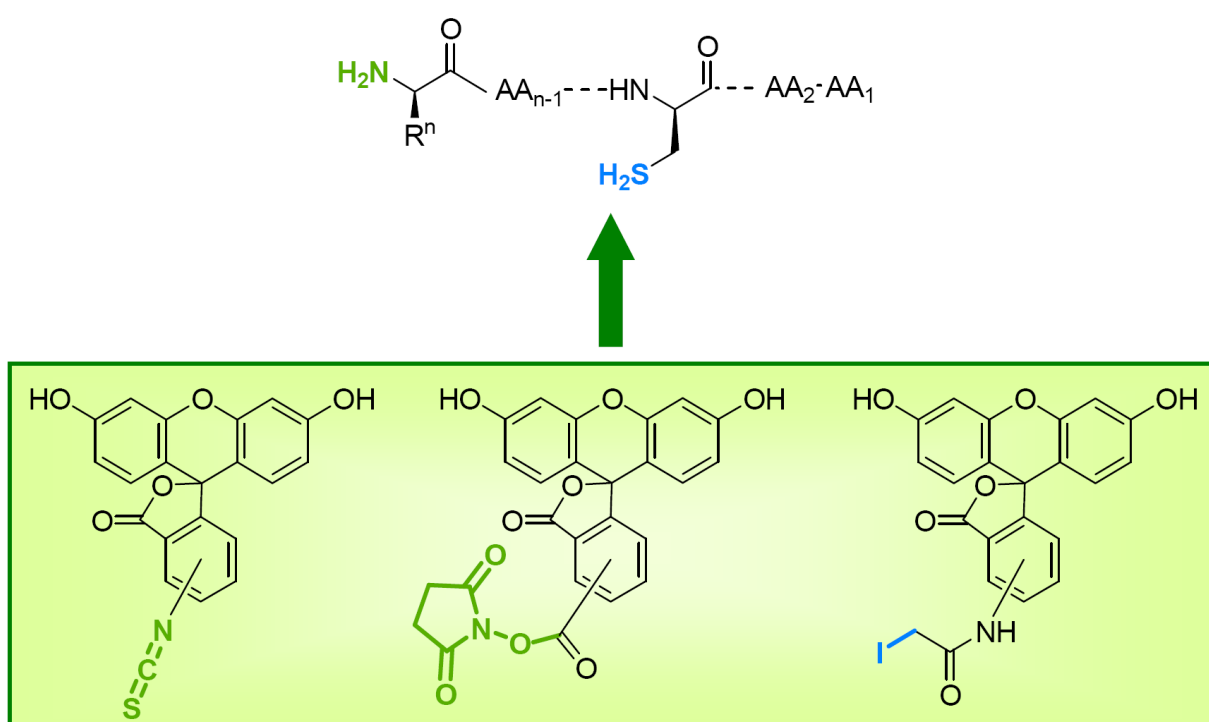
Figure 24: Peptide labeling in solution

For the labeling of a single small molecule, different strategies are compared[60] and the most convenient is selected. In the case of a whole library of small molecules, the labeling is, however, more difficult. In particular, if the molecules are diverse, with different reactive groups and solubilities. For such an application, another tool is required.

II.3.4.2.   Labeling on solid support

For molecules that are synthesized on solid support, labeling can follow the synthesis

and be thus directly done on the resin. The most popular example is peptide labeling.

Modification of a cysteine residue can e.g. be applied to labeling of resin-bound peptides [61], as described in solution.

Another very efficient technique for fully protected peptides is *N*-terminal labeling. This was first performed using the same reagents as for solution labeling, namely FITC and CFSE [62]. Then, it was shown that comparable results can be obtained by coupling the less expensive 4(5)-carboxyfluorescein (CF) [63] with DIC / HOBt activation.
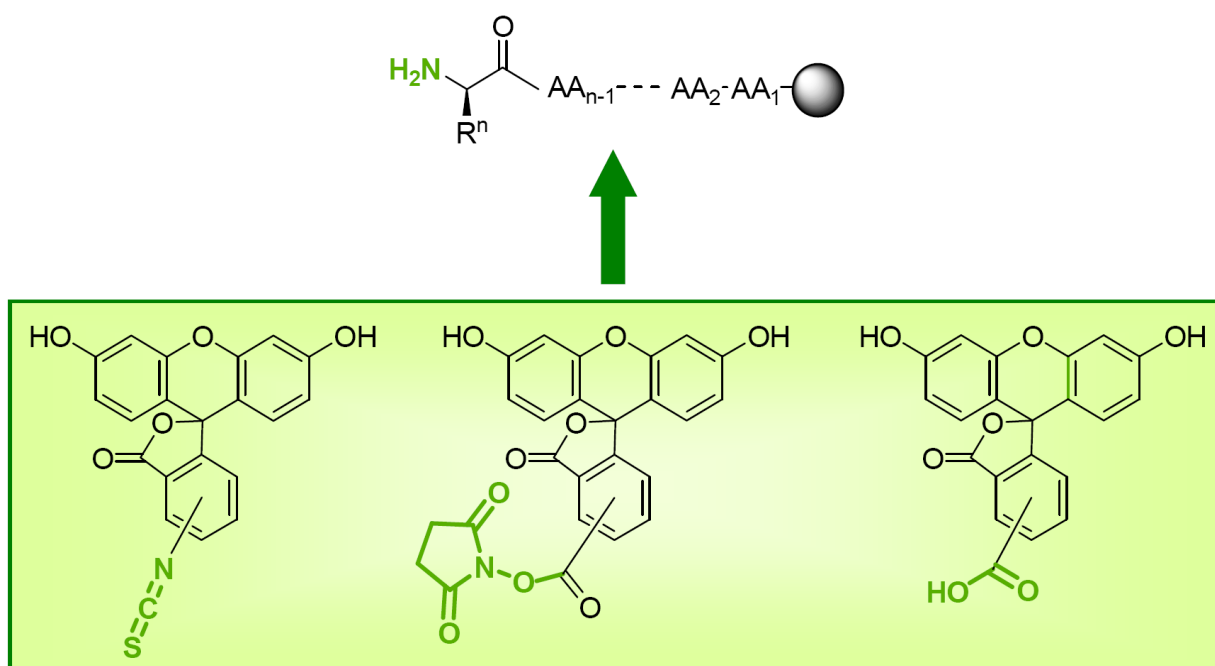
Figure 25: Peptide labeling on solid phase

For *C*-terminal labeling, a first approach has also been developed [64] (Figure 26): a lysine residue is first coupled to a resin, Fmoc is cleaved, and CF is then coupled to the amine. In a second step, the side-chain protecting group Dde is cleaved and the peptide can be built-up on the lysine side chain.

The resulting fluorescent label is, however, only loosely bound to the fluorophore. This can be interesting in the case of a small binding pocket, enabling the fluorophore to point out. For specific applications that require particular rigidity of the binding like FP, another solution has to be found.
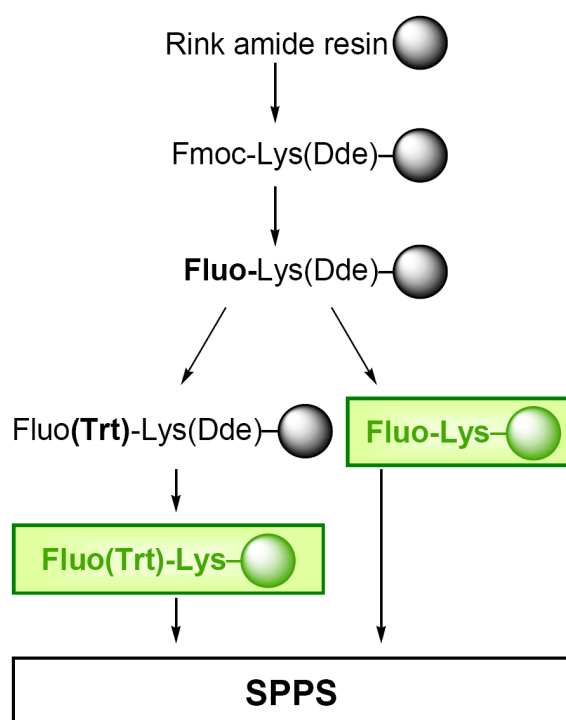
Figure 26: Preparation of N^α-carboxyfluorescein-labeled lysyl-Rink amide resins with a free ε-amino acid group for peptide assembly.

II.4. Protein-protein interactions

II.4.1. Proteins

II.4.1.1. Introduction

Proteins were recognized as a distinct class of biological molecules in the eighteenth century by Antoine Fourcroy and others [65]. The difficulty in purifying large quantities made them, however, very difficult to study for early protein biochemists.

Linus Pauling is credited with the successful prediction of regular protein secondary structures, based on hydrogen bonding – an idea first put forth by William Astbury in 1933 [66]. Later work of Walter Kauzmann [67] on denaturation contributed to understand protein folding mediated by hydrophobic interactions. In 1949 Fred Sanger determined the amino acids sequence of insulin [68], thus demonstrating that proteins are linear

polymers of amino acids. The first atomic-resolution structures of proteins were solved by X-ray crystallography in the 1960s and by NMR in the 1980s. As of 2009, the Protein Data Bank[69] has nearly 56 000 atomic-resolution structures of proteins.

II.4.1.2. Structure and function of proteins

The human proteome consists of a huge number of proteins, about 400 000. Proteins can be characterized by a primary structure (amino acid sequence), a secondary structure (local assemblies stabilized by hydrogen bonds, e.g. alpha helix or beta sheet), a tertiary structure (spatial relationship between the secondary structures, stabilized by nonlocal interactions such as the formation of a hydrophobic core, salt bridges, hydrogen or disulfide bonds) and for some of them a quaternary structure (interactions between different protein subunits).

Proteins can bind to other proteins as well as to small molecules very specifically and tightly and are thereby involved in many processes within the cell. The best known proteins are proteins that catalyze chemical reactions, called enzymes. Enzymes are in particular involved in metabolic pathways[70] or in DNA processes such as replication[71], transcription, or splicing. Signaling transduction[72][73] is also operated by proteins, as well as structural functions like organization of cell shape and motility. Many proteins remain, however, still not fully understood. Function elucidation is difficult due to the number of pathways in which each single protein is involved and the number of different proteins that are necessary for the obtention of a particular effect[74,75].

II.4.2. Analysis of protein-protein interactions

II.4.2.1. Protein-protein interactions vs. protein-small molecule interactions

The particularity of protein-protein interactions[76] in comparison with protein-small molecule interactions is the extended interface responsible for the binding. Whereas enzymes always have a well-defined binding pocket and known small molecule substrates that can be used as a starting point for drug design, proteins involved in protein-protein interactions have mostly no known small molecule ligand[77].

II.4.2.2. *Assays for studying protein-protein interactions*

One approach for studying protein-protein interactions is to map the epitope of one of the proteins on a small peptide or peptidomimetic. X-ray structures of protein-protein pairs [78] can also be helpful, and although they do not usually show deep cavities looking like small molecule binding sites, the few "hot-spots" – subsets of the interface that contribute to high affinity binding – are a sufficient starting point to study the interaction. A particularity of these areas is that proteins tend to use the same hot-spots in different interactions and thus adapt their conformation to reach the best binding. This high conformational flexibility makes traditional docking and virtual screening for protein-protein interaction inhibitors quite inapplicable. Another difficulty, once an inhibitor has been found, is to precisely define the mode of interaction, in particular which of both interacting proteins is bound to the inhibitor.

II.4.3. **Adaptor domains and PRS binding domains**

II.4.3.1. *Adaptor domains*

Adaptor domains [79] are small recognition domains; they have no enzymatic activity and bind to various classes of bioactive molecules such as nucleic acids, lipids, and proteins with a rather low affinity. Through their binding to main proteins, protein-binding adaptor domains are involved in many biologically relevant phenomena, in particular transduction pathways or formation of protein complexes.
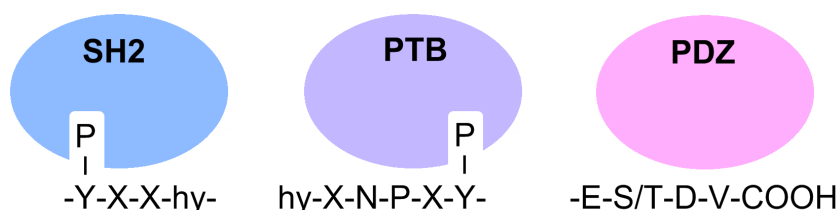
Figure 27: Three examples of protein adaptor domains: SH2, PTB, and PDZ. [84] The specific sequences recognized by the adaptor domains on their target acceptor proteins are indicated in single-letter code; hy: hydrophobic residue.

The Src homology 2 (SH2) domain 10 – a small protein domain binding to phosphotyrosines – was the first identified adaptor domain [80][81][82]. Then, many other adaptor domains have been found [83](Figure 27); they mainly recognize short linear

peptide sequences containing post-translationally modified amino acids (e.g. phosphotyrosines or acylated lysines), proline residues, or the *C*-terminal carboxy-group.

II.4.3.2.   *Proline-rich sequence binding domains*

Proline-rich sequences (PRS) are among the most common peptide motifs, but they are only recognized by a limited number of proline-rich sequence recognition domains (PRD). SH3 and WW domains are the most famous PRD families (Figure 28).
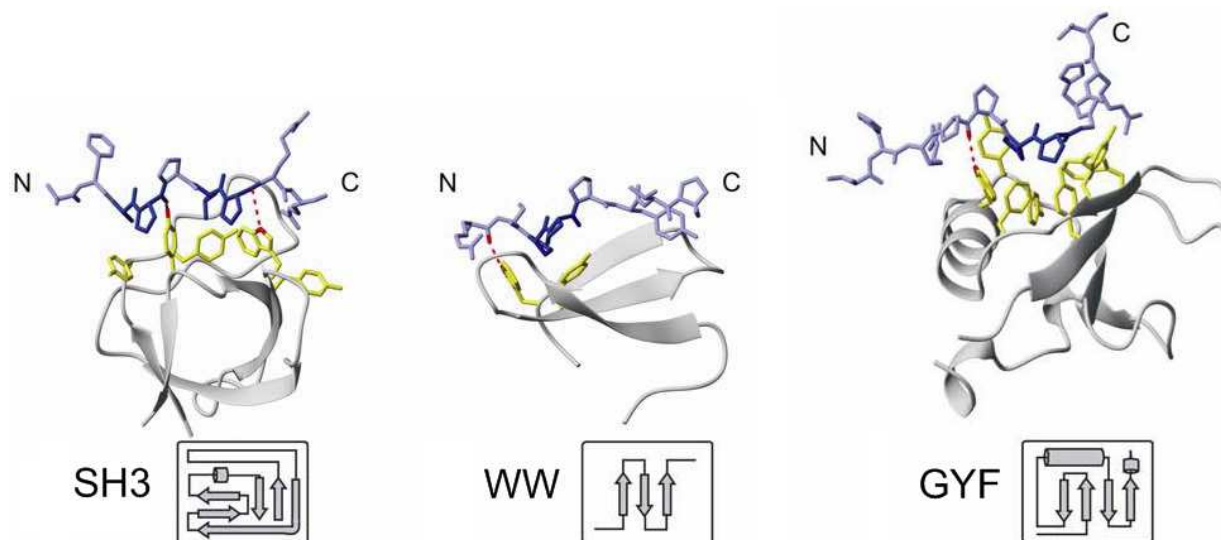
Figure 28: Structure of different prolin-recognition domains in complex with proline-rich ligands. For SH3 and WW, only one of the two opposite ligand binding orientations is exemplified. Aromatic residues which form part of the hydrophobic binding site are depicted in yellow. Residues of the ligands are colored in shades of blue with dark blue highlighting residues that interact with the hydrophobic pocket(s). The Hbonds between backbone carbonyl oxygens of the ligands and conserved aromatic residues (tryptophans and tyrosines) of the domains are represented as red, dotted lines. The topologies of the different PRD are shown below the structures. Arrows stand for β-strands, cylinders for helices. Protein Data Base (PDB) accession codes for presented structures are: 1PRM (SH3), 1EG4 (WW), 1L2Z (GYF).

SH3 (Scr Homology 3) domains are 50–70 amino acids long, fold as compact *β*-barrel of five to six anti-parallel *β* -strands connected by hydrophilic loops. The optimal ligand preference for each domain varies around the PxxP core. SH3 domains are involved in many signaling pathways, cell-cell communication, signal transduction from the cell surface to the nucleus.

WW domains are small globular modules composed of 38–40 amino acids. The name refers to two conserved tryptophan (W) residues that are spaced 20–22 amino acids

apart and play an important role in the structure and function of the domain. WW domains have an antiparallel three-stranded *β*-sheet structure that forms a shallow binding pocket for ligands containing PPxY or PPLP core motifs, usually flanked by additional prolines. Immediately after its discovery, the WW domain attracted attention because the signaling complexes it mediates have been implicated directly or indirectly in several diseases including Alzheimer's and Huntington's diseases [83a].

Beside these explicitly studied domains, other adaptor domains such as EVH1 (Enabled, VASP, Homology 1), UEV (Ubiquitin E2 Variant), prolifin, or GYF domains are still not so well known.

III. AIMS OF THE PROJECT

Various techniques are available for the labeling of compounds with fluorescent markers [85-88]. Most of these techniques have, however, been developed for biological and biochemical applications such as labeling of peptides, proteins, or cell components. The resulting molecules are used in assays based on fluorescence microscopy, ELISA [89], or FRET experiments using intensity detection. Therefore, existing reagents are all designed to enable a loose binding, avoiding an interference of the fluorophore in the behavior of the molecule of interest. This absence of rigidity is the reason why such labeling reagents cannot be used for fluorescence polarization experiments.

Fluorescence polarization detection is nevertheless already used for the screening of compound libraries for their binding to a macromolecule by displacement assays. Many known ligands are peptides; these can be labeled at the end of the solid phase peptide synthesis. As a result, *N*-terminally labeled peptides are obtained. However, such an important modification of the *N*-terminus is not always desirable. Indeed, the binding could be strongly affected, e.g. in the case of enzymes recognizing one or more residues *N*-terminally to the cleavage site [90]. In other situations, it could be interesting to label both the *N*- and the *C*-terminus of a peptide, e.g. to build a FRET system. Therefore, a new technique that enables the *C*-terminal labeling of peptides and preserves the rigidity of the system would be very usefull.

Small organic molecules are also used as fluorescent probes. However, the optimal labeling mode is often readjusted for each new set of molecules, resulting in a very specific procedure that is only used for the labeling of this focused library. Thus, there is a need for setting up a labeling method for obtaining FP probes, applicable to larger libraries and diverse organic compounds with a single universal procedure.

III.1. Development of a versatile fluorescence labeling method

In this dissertation project, a fluorescence labeling method should be developed for the labeling of a wide range of compounds, with the realization of fluorescence polarization assays as final objective.

The fluorophore should therefore be linked to the compounds through a short linker to avoid relaxation phenomena and enable accurate FP measurements. The synthesis strategy should be optimized to enable the obtention of high product amounts (10–20 g) without tedious purification steps.

III.2. Applications

III.2.1. Synthesis of *C*-terminally fluorescence labeled peptides for the screening of chemical libraries for their binding to the CD2BP2 and PERQ2 GYF domains via a competitive FP assay

As a model for protein-protein interactions, the binding of the CD2BP2 GYF domain to the CD2 protein – an interaction not so broadly studied so far – was selected. As every protein-protein interaction, this binding can be reduced to a protein-peptide interaction by a judicious choice of the amino acid sequence. In previous analyses [91], some peptides have already been shown to bind to CD2BP2 in the micromolar range. These peptides should be used as starting point for further investigations using FP based assays, in order to find non-peptidic ligands.

A small library of fluorescence labeled peptides should be synthesized first, using the developed labeling strategy. Dissociation constants of all peptides should be determined with FP measurements. After optimization, the best labeled peptidic ligand should be used to establish a displacement. Finally, the 16,896 compounds of the ChemBioNet library should be screened to find small molecule ligands of the CD2BP2 GYF domain.

An analog study should also be performed with the PERQ2 GYF domain.

III.2.2.   Synthesis of a diversity oriented library of fluorescence labeled small molecules to be used in a "direct" FP screening for their binding to different proteins

In the last decades, hudge compound libraries have been used for high-throughput screening. The disadvantage of such an approach is, however, the low hit rate. Therefore, even more strategies have been developed for low-throughput screenings, using smaller libraries consisting of a limited number of accurately selected compounds.

After an accurate selection of appropriate fragments according to diversity and drugability criteria, the labeling strategy developed here should enable the synthesis of a diversity set of fluorescence labeled small molecules. This library should be applied for different fluorescence experiments. In particular, the CD2BP2 GYF domain should be reviewed with this assay, and different other proteins should also be tested to evaluate the specificity of the hits.

IV. RESULTS AND DISCUSSION

IV.1. Fluorophore selection and synthesis strategy

IV.1.1. Prerequisites for a practical fluorophore for the synthesis of labeled libraries for FP assays

As presented in part II.3.2.4, many different fluorophores are available. They vary in their absorption and fluorescence emission wavelengths, their quantum yields, and their reactive sites. Different fluorophores are also appropriate for different mediums.

The choice of the right fluorophore depends essentially on the intended application. In this work, a fluorescence labeling method should be developed for the generation of fluorescent molecules to be used in FP assays.

IV.1.1.1. Absorption and emission wavelengths

To ensure accurate results, the fluorescence absorption and emission ranges of the fluorophore should not overlap with the intrinsic fluorescence of peptides, proteins and small molecules. Therefore, fluorophores like dansyl, quinine, or coumarin cannot be used. Dyes like fluorescein, rhodamine, cyanine or BODIPY are more suitable (see Part II.3.2.4, Table 4 and Figure 19).

IV.1.1.2. Fluorescence lifetime

Another parameter that influences the polarization value is the fluorescence lifetime τ, which is specific to each fluorophore. An example of the impact of fluorescence lifetime changes on the relationship between polarization and molecular weight is illustrated on Figure 29.

A high fluorescence lifetime is associated with a low polarization value: indeed, if the period between light absorption and emission is long enough, even large molecules can partially relax. Dansyl dyes can therefore be excluded due to their high fluorescence lifetime; they would only be applicable to large proteins (50–75 kDa).

A lower fluorescence lifetime is more convenient because of its very high sensitivity, but if this is too low (<1 ns), it could be also problematic if used with large peptidic probes: these would deliver a polarization signal even at the free state, what would induce a low signal to baseline ratio. Cyanine dyes that can have, depending on the spacer, a very low fluorescence lifetime (0.3 ns in case of Cy3; 1.0 ns in case of Cy5) should therefore be used carefully.

With a fluorescence lifetime of 2–8 ns, fluorescein, rhodamine and BODIPY are well-suited for protein-ligand interaction studies.
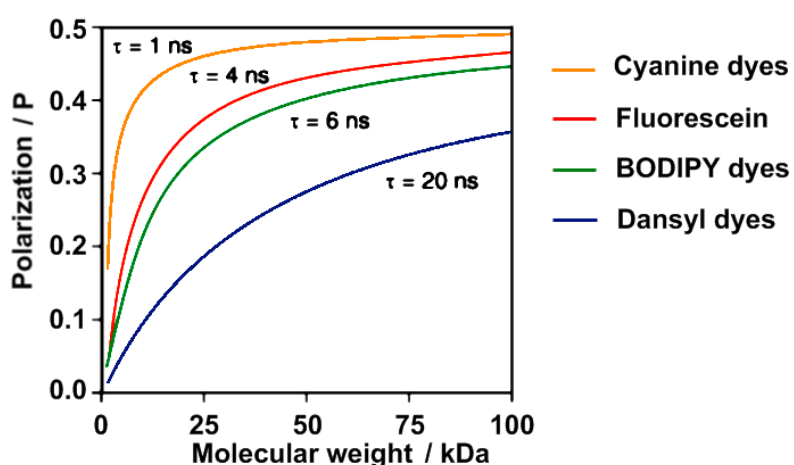
Figure 29: Simulation of the relationship between molecular weight and fluorescence polarization[92]**.** These simulations assume $P_0=0.5$, and a rigid attachment of dyes to spherical carriers. For proteins in aqueous solutions, the curves would vary.

IV.1.1.3. Functionalization potential

To enable the synthesis of libraries of labeled molecules, the fluorophore should exist under a form enabling its functionalization. It should also be commercially available in synthesis-scale amount. The BODIPY dye that is similar to fluorescein from the fluorescence properties point of view can, however, not be purchased in an amount enabling large scale synthesis.

IV.1.1.4. Selection of the fluorophore

Considering the different criteria mentioned above, fluorescein and rhodamine have been selected for this project. These are two of the most popular fluorophores for assays and imaging in biochemical, biological and medicinal research. Both are

commercially available – what was a prerequisite for the development of a synthesis strategy that should be used for the labeling of hundreds of compounds. They can be found under different reactive species: as isothiocyanates, *N*-hydroxysuccinimide esters or carboxylic acids; fluorescein can also be purchased as an amine.

IV.1.2.   Reactivity of fluorescein and rhodamine

Fluorescein can be present under seven prototropic forms (Scheme 8) [93-96]: one dianion, two monoanions, three neutral species, and one cation. As monoanion, the carboxylate form is highly predominant. The neutral fluorescein form, however, results in an equilibrium between a lactone (70 %), a carboxylic acid (15 %) and a zwitterion (15 %). These multiple species and reactive sites make the study of reactions involving fluorescein an interesting and challenging purpose.
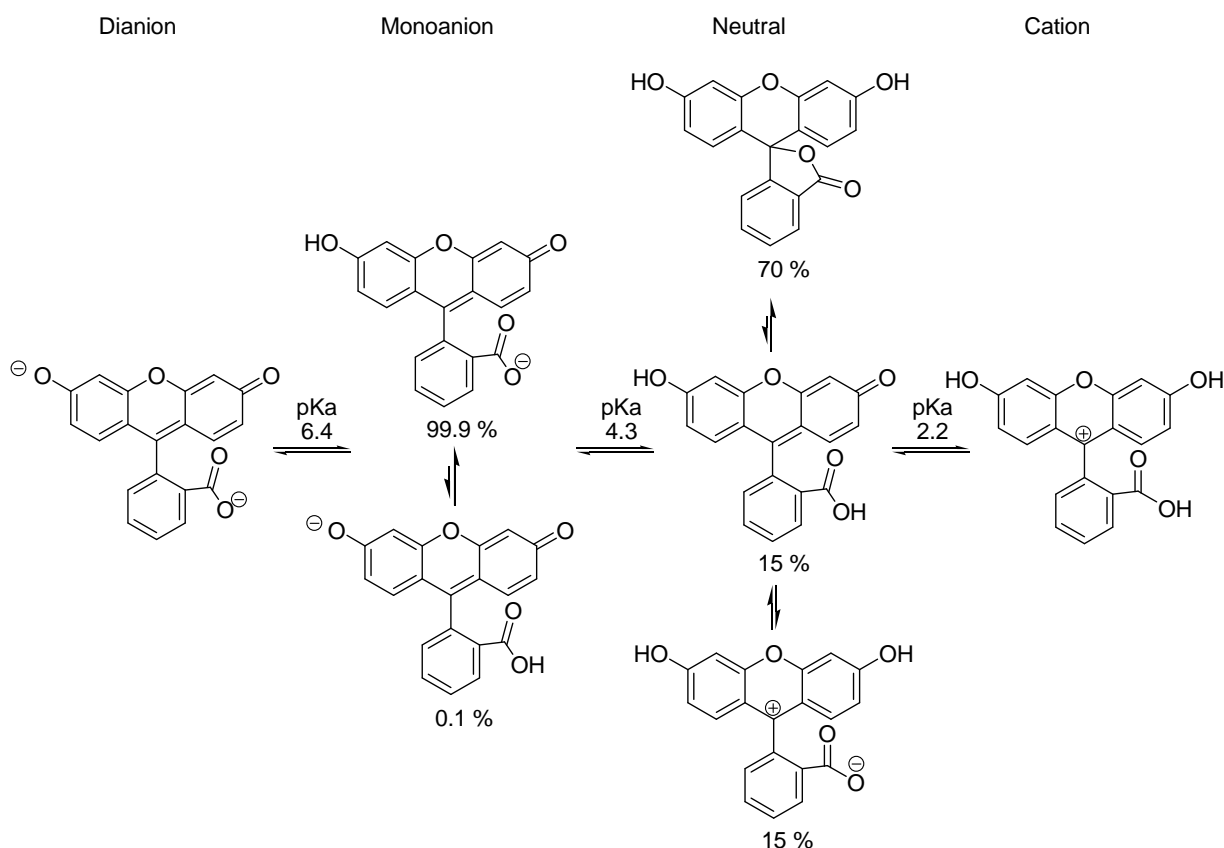
Scheme 8: Seven prototropic forms of fluorescein

Unlike fluorescein, rhodamine is commercially available as an ammonium chloride. The carboxylic acid can be closed to form a lactone ring under basic conditions

(Scheme 9); the obtained rhodamine specie should therefore be easier to control in solution synthesis compared to fluorescein.
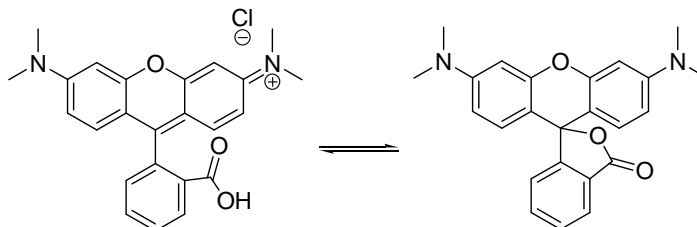
Scheme 9: Rhodamine forms

IV.1.3.   Synthesis strategy

A fluorescence reagent for the labeling of small molecules to be used in FP assays should be a system composed of a fluorophore and a short linker. Indeed, the fluorophore should not be directly bound to the small molecule so that it does not affect significantly the interaction with the protein, but should be close enough to preserve the system rigidity for polarization measurements (Figure 30). Finally, the linker should have a reactive site to react with the small molecule.

Such a system can be synthesized in different ways. The strategy of labeling in solution was examined first.
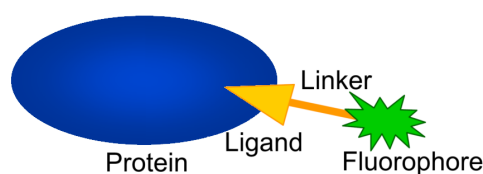
Figure 30: Model of a fluorescent labeled compound interacting with a protein

2-(2-Aminoethoxy)ethanol was chosen as a linker: the amino group should be attached to the fluorophore, the alcohol converted into a leaving group to enable nucleophilic substitution with various nucleophiles, and the oxygen within the chain would act as solubilizing group in aqueous medium

In order to couple selectively the linker to the fluorophore on the amine side, the alcohol was protected [97] by silyl groups that can be cleaved under acidic conditions. Furthermore, the cleavage product can be easily removed from the reaction mixture.

After testing different solvents (DCM, Et$_2$O, DMF) and bases (NEt$_3$, DIPEA, imidazole) [98], the protected 2-(2-aminoethoxy)ethanol was successfully obtained with both TES [99-102] and TBDMS [103,104] protecting groups (Scheme 10).
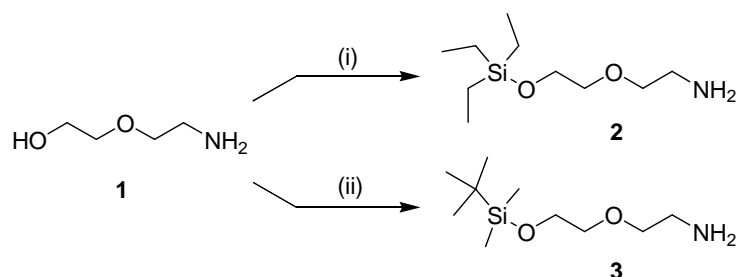
Scheme 10: Protection of 2-(2-aminoethoxy)ethanol with triethylsilyl chloride and tert-butyldimethylsilyl chloride. (i) TES-Cl (1.1 equiv.), imidazole (5.0 equiv.), DMF, 0°C to RT, yield: 94%; (ii) TBDMS-Cl (1.1 equiv.), imidazole (5.0 equiv.), DCM, 0°C to RT, yield: 92%.

The protected linker was then coupled to rhodamine and fluorescein moieties. After testing different carbodiimide activating agents like EDC or DIC, with and without addition of base or DMAP, which were all unsuccessful, coupling to TAMRA was reached using pyBOP, the less mutagenic analog of BOP (Scheme 11) [105].
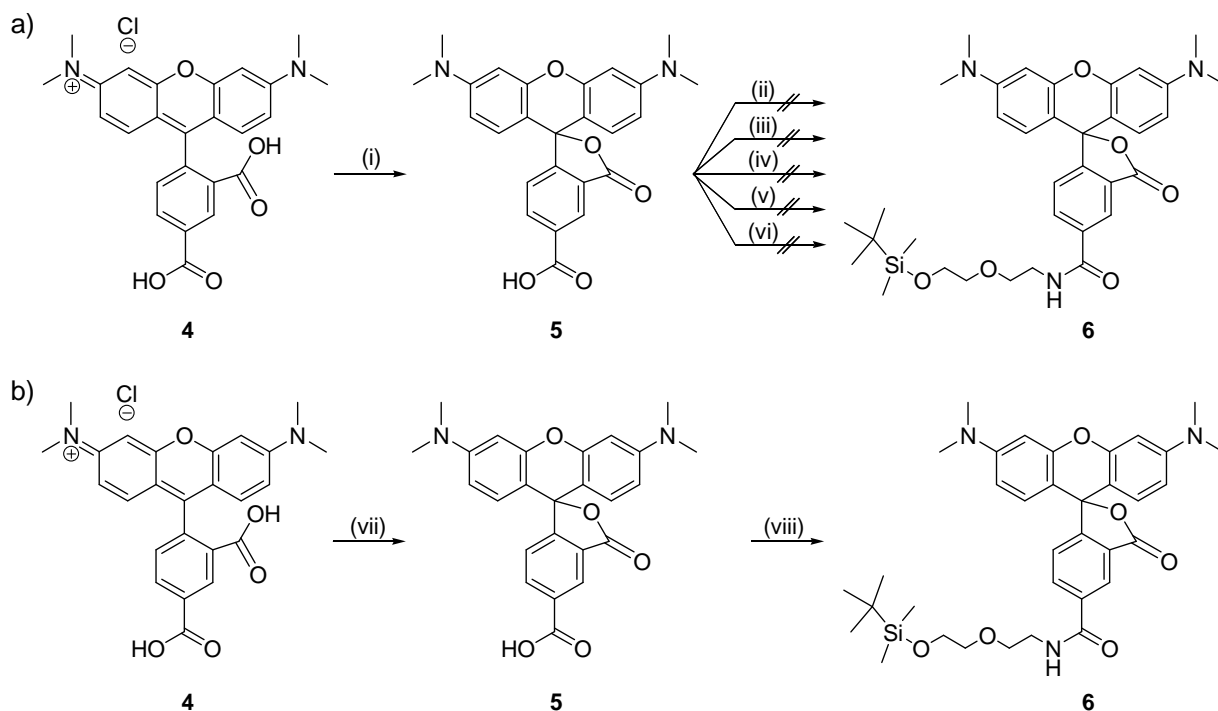
Scheme 11: Coupling of 2-[2-(*tert*-butyldimethyl-silanyloxy)-ethoxy]-ethylamine to TAMRA. a) (i): DIPEA, DMF; (ii): 2-[2-(*tert*-butyldimethyl-silanyloxy)-ethoxy]-ethylamine, EDC, DIPEA, DMF, conversion < 20%; (iii): 2-[2-(*O*-TBDMS)-ethoxy]-ethylamine, EDC, DMAP, DMF, conversion < 20%;

(iv): 2-[2-(*O*-TBDMS)-ethoxy]-ethylamine, DIC, DIPEA, DMF, conversion < 20%; (v): 2-[2-(*O*-TBDMS)-ethoxy]-ethylamine, DIC, DMAP, DMF, conversion < 20%; (vi): 2-[2-(*O*-TBDMS)-ethoxy]-ethylamine, pyBOP, DIPEA, DMF, conversion : 40–50%; b) (vii): NMM; (viii): 2-[2-(*O*-TBDMS)-ethoxy]-ethylamine (2.0 equiv.), pyBOP (1.1 equiv.), NMM (2.0 equiv.), DMF, conversion: 100%, yield: 40–50%.

The TBDMS protected compound, more stable than the TES protected, was further used. The primary alcohol should after deprotection be converted into a leaving group, but neither tosyl, nor mesyl or triflate esters could be obtained. Attempt to oxidize the alcohol to an aldehyde or a carboxylic acid was also unsuccessful because of difficulties with the work-up of the water-soluble rhodamine derivatives.
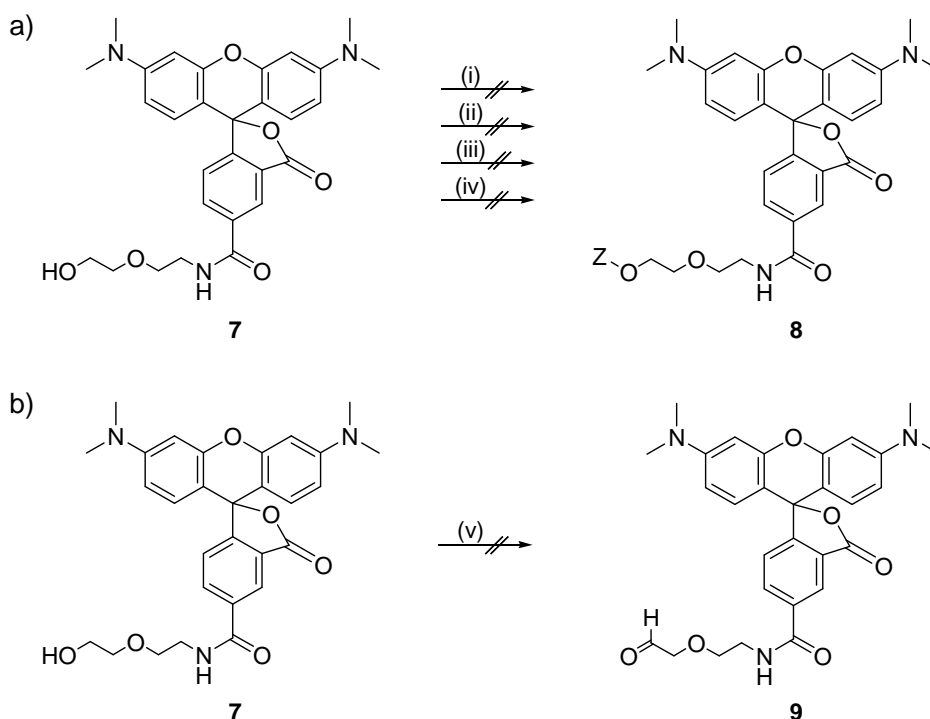
Scheme 12: Introduction of a leaving group or oxidation of the primary alcohol to obtain a rhodamine labeling reagent for nucleophiles. Z: Mesyl, tosyl, or triflate. a) (i): Mes-Cl or Ts-Cl or Tf₂O, DIPEA, DCM, 0℃; (ii): Mes-Cl or Ts-Cl or Tf₂O, DIPEA, DCM, RT; (iii): Mes-Cl or Ts-Cl or Tf₂O, NMM, DCM, 0℃; (iv): Mes-Cl or Ts-Cl or Tf₂O, NMM, DCM, RT; b) (v): Dess-Martin periodinane; DMF, RT.

To control the reaction on the fluorescein moiety, the phenolic hydroxyls were first protected with groups like acetyl[60,103,106,107], silyl[108-110] or pivaloyl[111]. Success of these reactions varied and depended on the functionalization of fluorescein (Scheme 13). The coupling of a short linker on protected fluorescein molecules in a subsequent step could, however, never be obtained (Scheme 14).

Subsequently, the inverse approach was studied. Coupling of the linker was done first, followed by the protection of both phenols (Scheme 15), deprotection of the alcohol and its conversion into an electrophile, but the same problem as by syntheses with rhodamine appeared: the difficulty to convert the primary alcohol to a leaving group (Scheme 16).
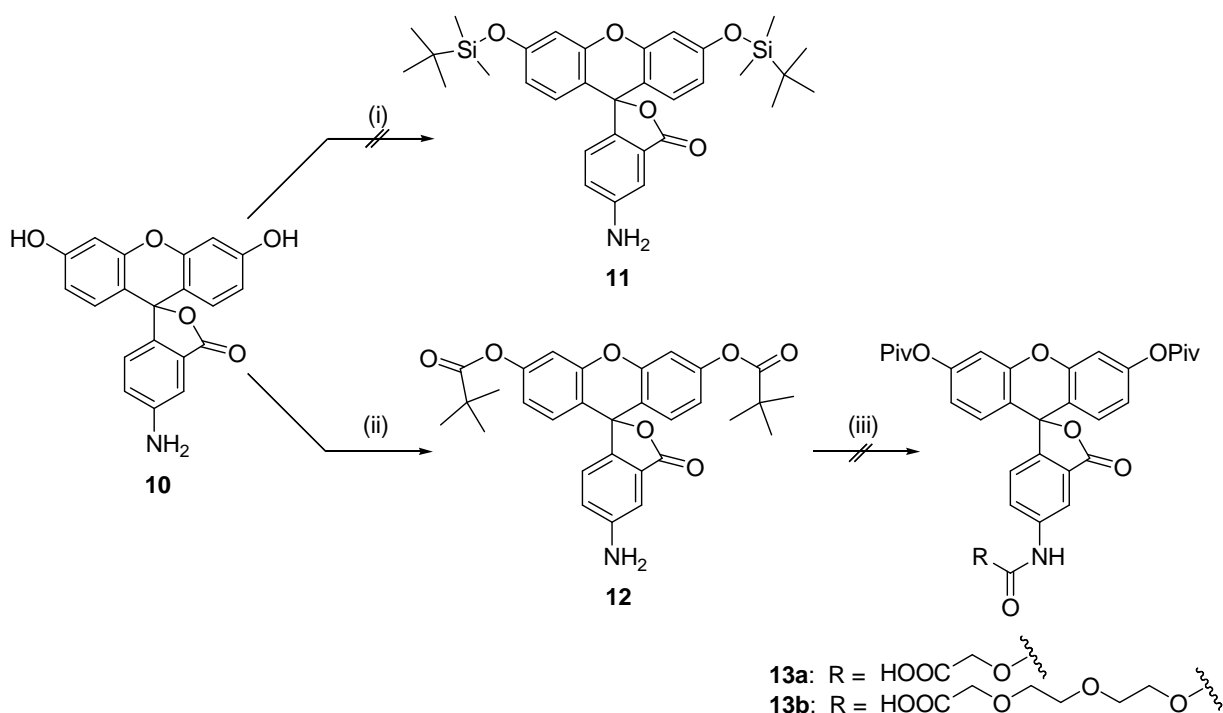
Scheme 13: Protection of fluoresceinamine and coupling of short PEG linkers. (i): TBDMS-Cl, imidazole, DMF, RT; (ii): Piv$_2$O (2.1 equiv.), Cs$_2$CO$_3$ (1.1 equiv.), DMF, RT; 3h; (iii) RCOOH, DIPEA or NMM, pyBOP, DMF, RT.
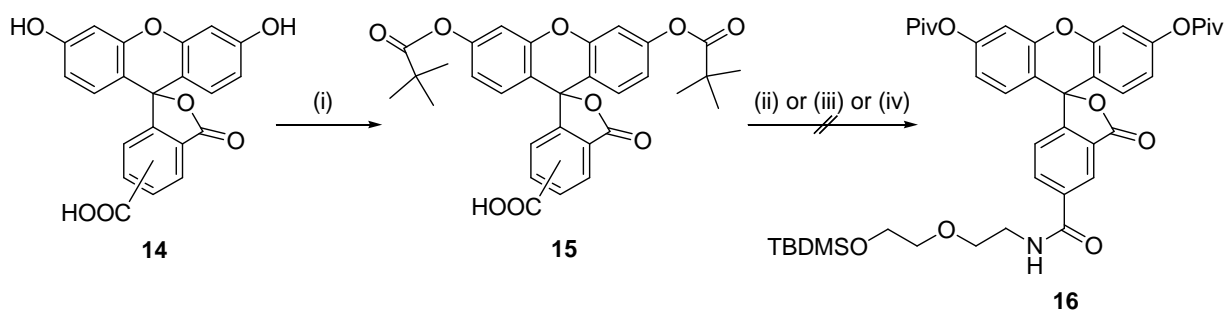
Scheme 14: Protection of 5(6)-carboxyfluorescein and coupling of 2-[2-(*tert*-butyldimethyl-silanyloxy)-ethoxy]-ethylamine. (i): Piv$_2$O (2.1 equiv.), Cs$_2$CO$_3$ (1.1 equiv.), DMF, RT; 3h; (ii): 2-[2-(*O*-TBDMS)-ethoxy]-ethylamine, DIPEA or NMM, pyBOP, DMF, RT; (iii): 2-[2-(*O*-TBDMS)-ethoxy]-ethylamine, DIPEA, EDC, DMF, RT; (iv): 2-[2-(*O*-TBDMS)-ethoxy]-ethylamine, DMAP, EDC, DMF, RT.
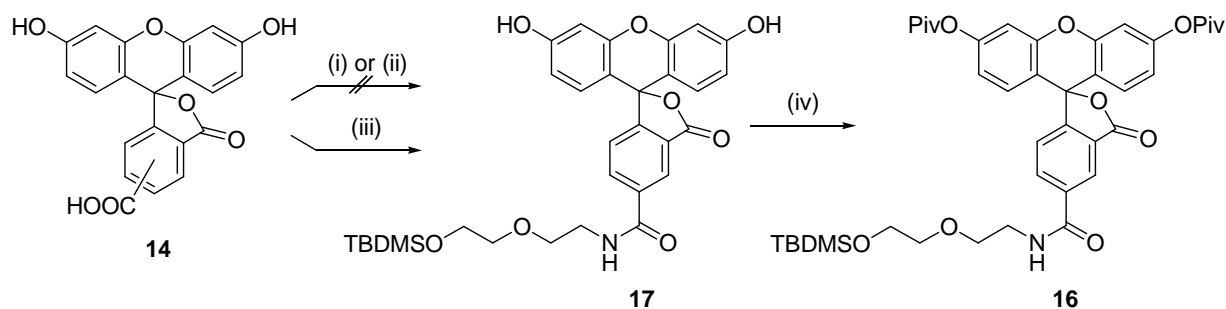
Scheme 15: Coupling of 2-[2-(*tert*-butyldimethyl-silanyloxy)-ethoxy]-ethylamine to 5(6)-carboxy-fluorescein and protection of the phenols. (i): 2-[2-(*O*-TBDMS)-ethoxy]-ethylamine, DIPEA, EDC, DMF, RT; (ii): 2-[2-(*O*-TBDMS)-ethoxy]-ethylamine, DMAP, EDC, DMF, RT; (iii): 2-[2-(*O*-TBDMS)-ethoxy]-ethylamine (1.0 equiv.), NMM (1.0 equiv.), pyBOP (1.1 equiv.), DMF, RT; (iv): Piv$_2$O (2.1 equiv.), Cs$_2$CO$_3$ (1.1 equiv.), DMF, RT; 3h.
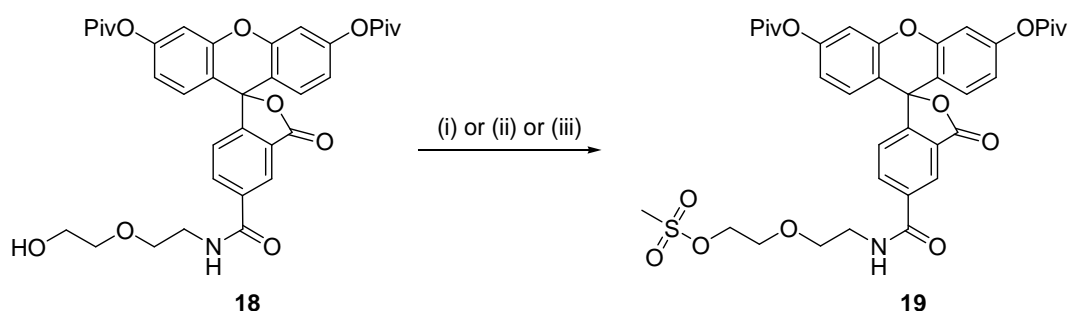
Scheme 16: Introduction of a leaving group to obtain a fluorescein labeling reagent for nucleophiles. (i): Mes-Cl, DIPEA, DCM, 0°C or RT, conversion: 14%; (ii): Mes-Cl, NMM, DCM, 0°C, conversion: 18%; (iv): Mes-Cl, NMM, DCM, RT, conversion: 30%. The equivalent reactions with tosyl chloride or triflate anhydride delivered almost no product (conversion < 5%).

IV.2. Synthesis of a solid phase reagent for fluorescence labeling

IV.2.1. General strategy: selection of the most convenient fluorophore, resin, and protecting group

Due to the multiple difficulties encountered with fluorescein and rhodamine derivatives in solution such as multiple reactivity, absence of reactivity, solubility and work-up problems, the synthesis strategy was modified.

Attaching the fluorophore on solid support should simplify the synthesis procedure, and deliver pure products without tedious work-up.

- 53 -

IV.2.1.1. Selection of the fluorophore

A suitable fluorophore should have two different reactive sites: an attachment site for on-resin immobilization and a reactive site for the coupling of small molecules or peptides. Dyes of both rhodamine and fluorescein families came into consideration (Scheme 17).
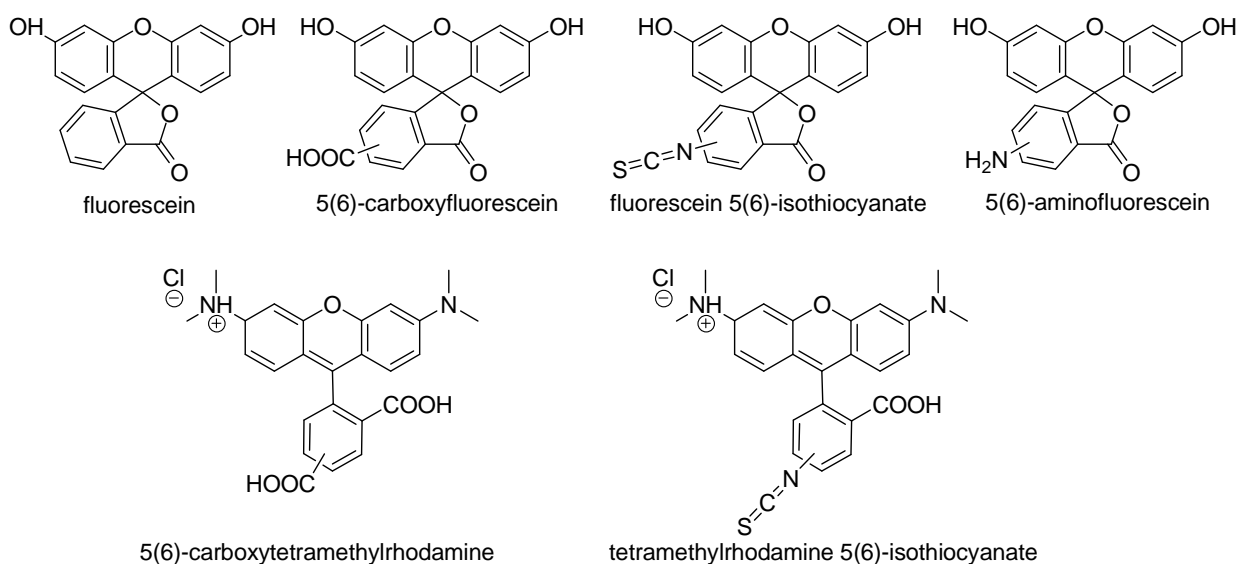
Scheme 17: Examples of fluorescein and rhodamine derivatives

The simple fluorescein molecule has no second reactive site for the coupling of small molecules. 5(6)-Carboxyfluorescein and 5(6)-carboxytetramethylrhodamine have a second carboxylic acid which could cause selectivity problems. Isothiocyanate derivatives of both fluorophores could be used but are much more expensive than aminofluorescein, which can moreover be purchased as a pure isomer. Aminofluorescein has three different reactive sites – phenol, carboxylic acid and aniline – which can be differentiated. It was selected for the synthesis as it presents all the prerequisites.

IV.2.1.2. Selection of the linker

As mentioned in part II.2.2.2, linkers can be discriminated through their cleavage mode. They can be cleaved by acidic treatment, by basic treatment, or by UV radiations. Photocleavable linkers, however, should be avoided when working with a fluorophore. Such cleavages often require long irradiation times (> 10 h) [3] which could

affect the fluorescence properties, even if the applied light does not exactly correspond to the fluorophore excitation wavelength.

The possibilities to link a phenol to a resin are shown on Scheme 18. Phenols can be linked as ethers (e.g. benzyl ethers) or acetals and can be cleaved under acidic conditions. They can also be linked as silyl ethers and cleaved with HF or TBAF, or linked as esters and cleaved under basic conditions. Resins of the trityl family were chosen since they are widely available and have been extensively used in SPPS.
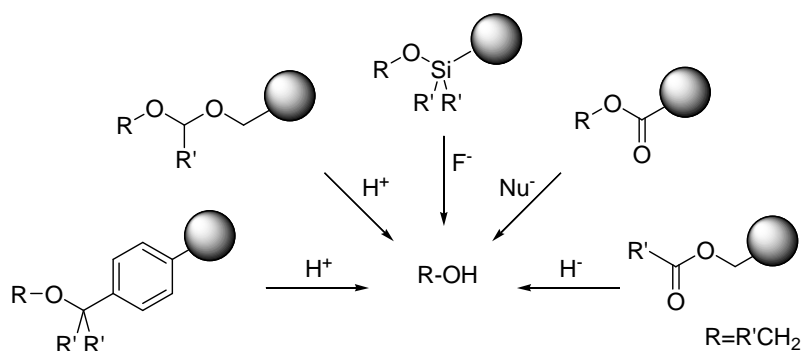
Scheme 18: Linkers for alcohols and phenols

Varying the substituents on the phenyl ring affects the stability of the trityl resin [112,113]: the 4-methoxytrityl linker is more labile than non-substituted trityl [22,114] because of the mesomeric donor effect of the methoxy group that is stabilizing the carbocation (Scheme 19a), thereby promoting the cleavage. In contrast, the 2-chlorotrityl linker is more stable than non-substituted trityl because of the inductive attractor effect (-I) of the chlorine that dominates in an aromatic system its mesomeric donor effect (+M) and destabilizes the trityl cation so that cleavage is hindered (Scheme 19b).
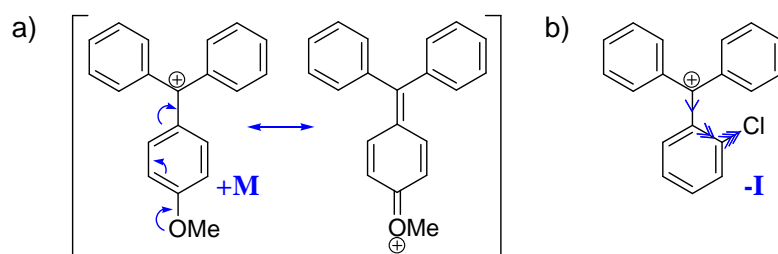
Scheme 19: Compared stability of 4-OMe (a) and 2-Cl (b) substituted trityl carbocations. The 4-methoxytrityl cation is stabilized by the mesomeric donor effect of the methoxy group (a), whereas the 2-chlorotrityl cation is destabilized by the inductive attractor effect of the chlorine (b).

The stable 2-chlorotrityl chloride resin was chosen. It can be cleaved by different methods. Either by 95% TFA / H$_2$O delivering fully deprotected peptides, or at milder conditions like AcOH / TFE / DCM (1:1:8) [115] or 20% HFIP in DCM [23], with the peptidic side chains remaining protected. The latter conditions are more convenient for the synthesis of sensitive molecules.

IV.2.1.3. A necessary protection

To test if non-protected aminofluorescein **10** could be directly coupled to 2-chlorotrityl chloride resin, the analogous reaction was performed with trityl chloride in solution. The resulting product was analyzed by NMR (Figure 31), its spectrum was compared to the spectrum of aminofluorescein **10** and showed the exclusive formation of *N*-tritylated aminofluorescein **20**.

Indeed, on the product spectrum, the integration of the aromatic signals (e.g. three protons **4** to one proton **3**) corresponds to the presence of exactly one trityl group on the molecule. If this group was bound to a phenol, the molecule would no longer be symmetric and the signals of both phenyls of the xanthene moiety (in orange) would be different, which is not the case here: protons **1** and **8** are identical, as well as protons **2** and **7**, and protons **4** and **5.**

Furthermore, only the chemical shifts of protons of the fluorescein phenyl ring (in green) are affected by trityl binding, if compared with the NMR of unsubstituted aminofluorescein **10** (Figure 32). Protons **3**, **5**, and **6** are shifted to the right (higher field), whereas the protons of the xanthene ring remain almost in the same area. It should be noted that both spectra could not be directly superimposed, because they were measured under different conditions.

Subsequently, assuming that the chemoselectivity of the solid phase reaction would be the same as in solution, the attachment to the resin would occur on the aniline, so that no further coupling could be done. Therefore, the aniline should be capped first, before attaching aminofluorescein **10** to trityl resin on one of the remaining positions. The Fmoc protecting group, traditionally used by SPPS, was selected for this purpose. It enables resin loading determination by UV spectroscopy, as usually performed when coupling the first amino acid of a peptide chain to the solid support.
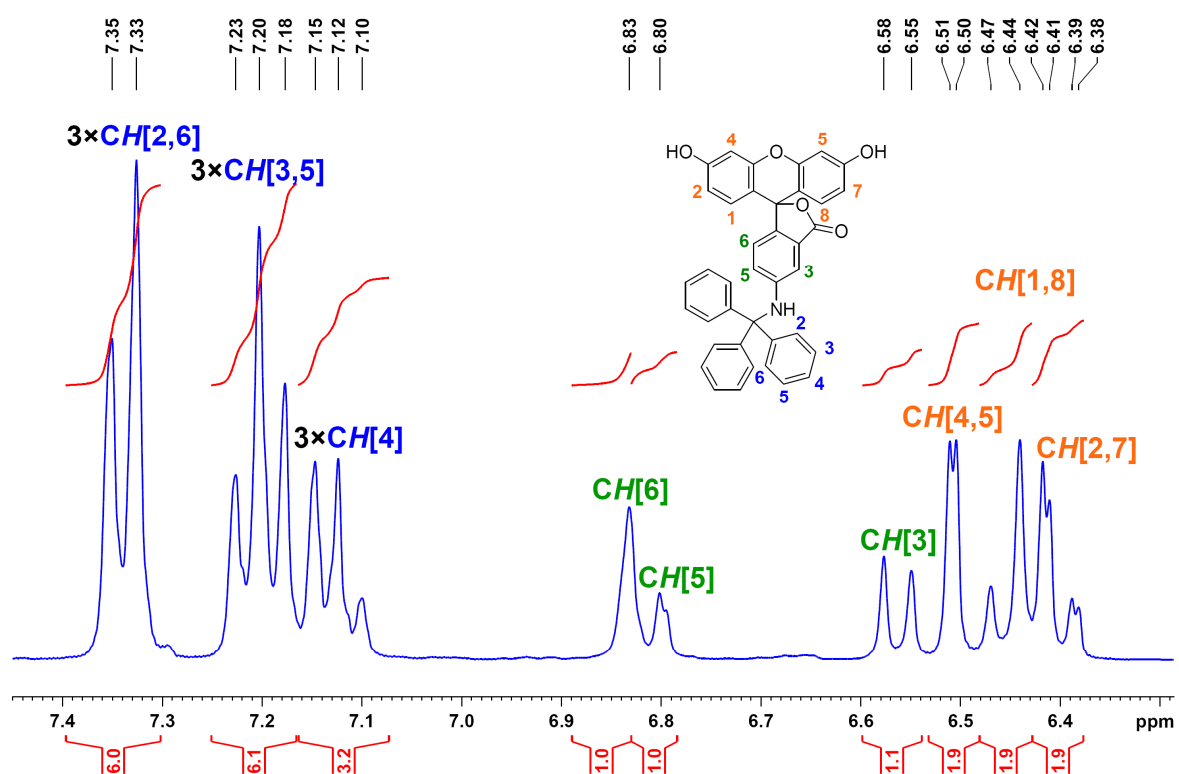
Figure 31: NMR spectrum of tritylated aminofluorescein 20
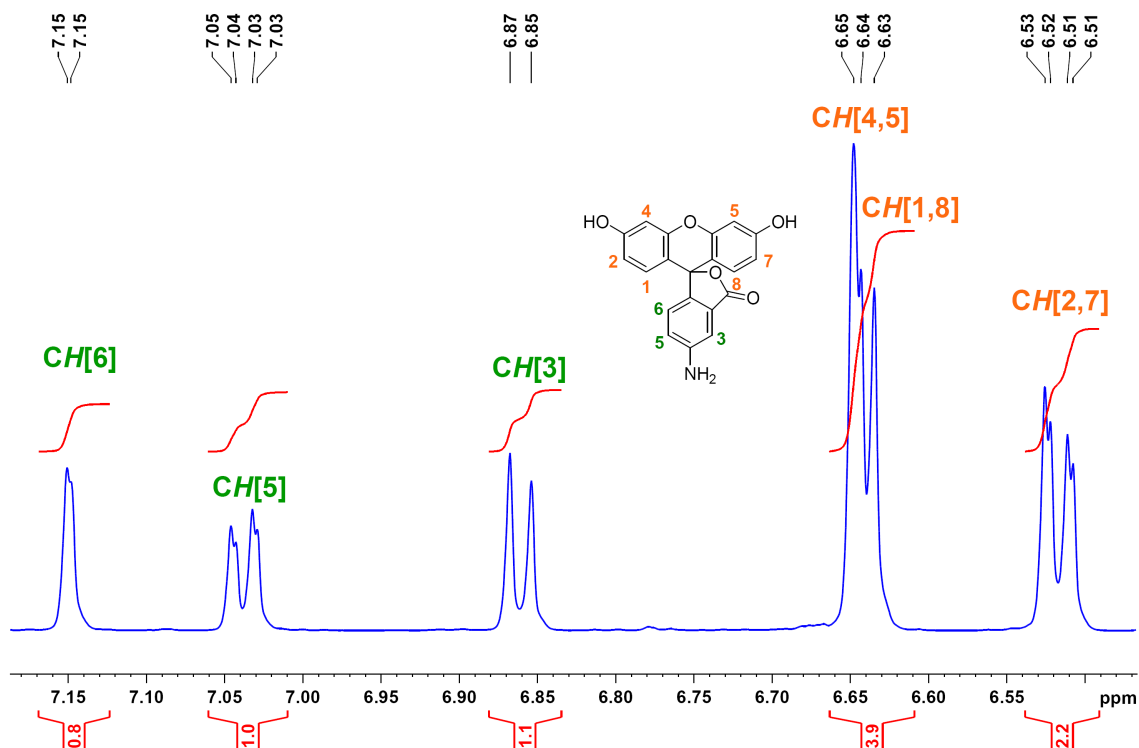
Figure 32: NMR spectrum of aminofluorescein 10

IV.2.2. Synthesis in solution of *N*-Fmoc aminofluorescein 21

IV.2.2.1. Direct aminofluorescein protection

a. *Protection with Fmoc chloride*

The synthesis of *N*-Fmoc protected aminofluorescein **21** was performed first by direct coupling of Fmoc chloride. The reaction was conducted under the Schotten-Baumann conditions in presence of aqueous NaOH (Scheme 20).
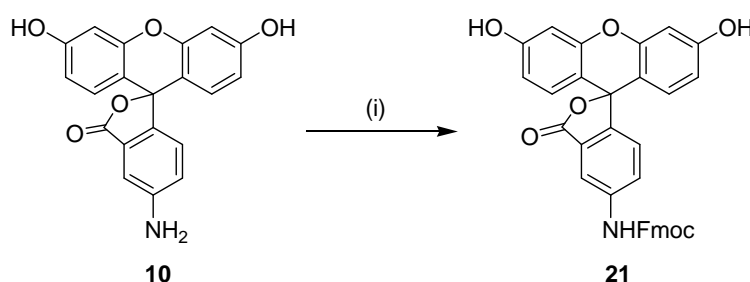
Scheme 20: Aminofluorescein direct protection with Fmoc chloride. (i) Fmoc-Cl (1.2 equiv), 1 M aq. NaOH (1.0 equiv), THF, 0°C to RT, 180 min.
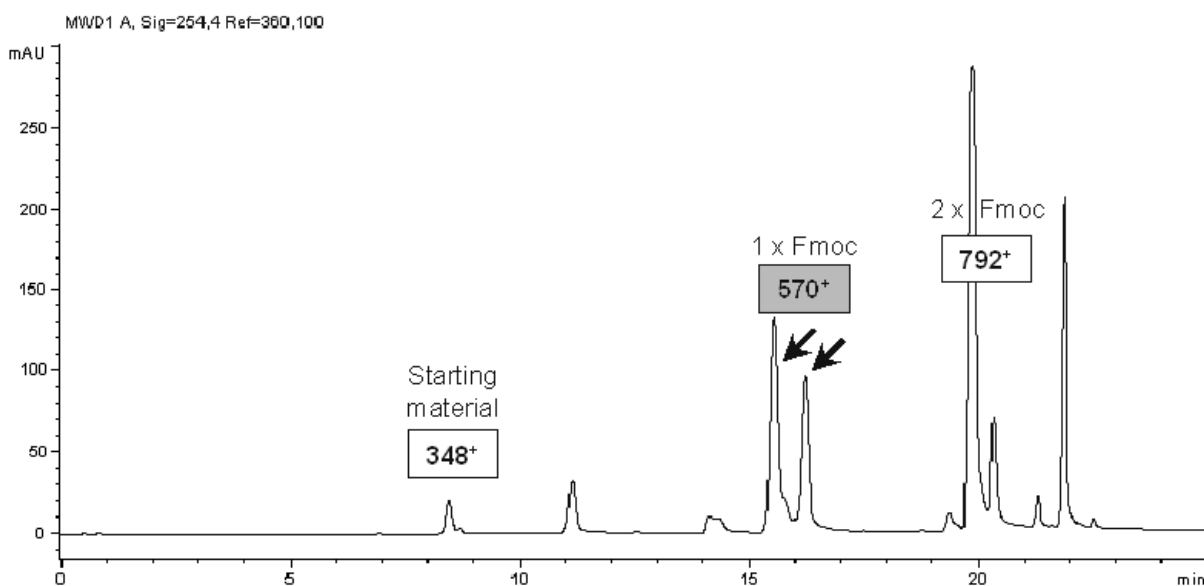
Figure 33: LC-MS analysis of the crude product resulting from the direct Fmoc protection of aminofluorescein in solution, before optimization of the reaction conditions. The reaction delivered a mixture of starting material ([M+H]$^+$ = 348), mono-protected ([M+H]$^+$ = 570) and bis-protected ([M+H]$^+$ = 792) aminofluorescein.

Traditionally, the name "Schotten-Baumann conditions" is used for a system consisting of water and a non-water-miscible organic solvent. The base in the aqueous phase neutralizes the acid generated in the reaction, while the starting materials and product remain in the organic phase. In this particular case, the reaction was adapted from the general procedure[116,117] to the particular case of aminofluorescein **10**. Fmoc chloride was dissolved in THF because aminofluorescein was not soluble in other organic solvents that are completely non-miscible with water such as DCM or Et$_2$O.

Unfortunately, a mixture of compounds was obtained (Figure 33), the main peaks representing starting material **10**, mono *N-* and *O*-protected aminofluorescein molecules, as well as bis-protected aminofluorescein.

b. Protection with Fmoc succinimide

To prevent the formation of bis-protected species, coupling with the less reactive Fmoc succinimide was tested (Scheme 21) but almost no reaction occurred in this case. Only traces of product were found in the LC-MS spectrum, even after longer reaction times and it was decided to return to the strategy with Fmoc chloride and optimize the reaction conditions.
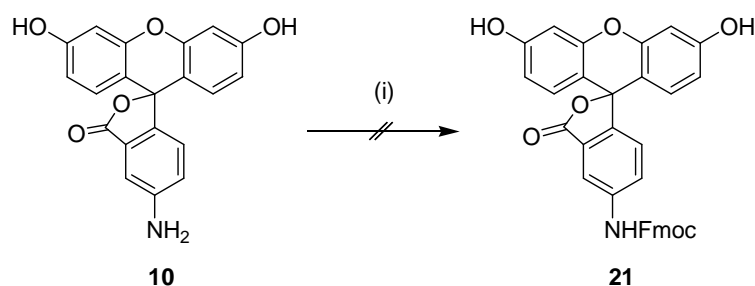
Scheme 21: Aminofluorescein direct protection with Fmoc-hydroxysuccinimide ester. (i) Fmoc-OSu (1.1 equiv), NaOH (1.0 equiv.), DMF, RT, 1 h to 72 h.

c. Optimization of the protection with Fmoc chloride

The decisive improvement was the operation mode: a very slow addition of very small amounts of Fmoc chloride and aqueous NaOH. The reaction mixture was kept at 0 °C during the addition of the reagents and was then allowed to return to room temperature.

After optimization, only three products were obtained: starting material, expected product, as well as bis-protected aminofluorescein – was obtained, as shown on Figure 34a. After work-up and flash chromatography purification on a silica gel column with Hex / EtOAc mixture (2:1) and (1:1), a very pure product was isolated (97%) (Figure 34b) but only with 40% yield. The side-products could be recycled: bis-protected aminofluorescein was completely deprotected by basic treatment and used again for the reaction. However, this route was not optimal for larger scale synthesis.
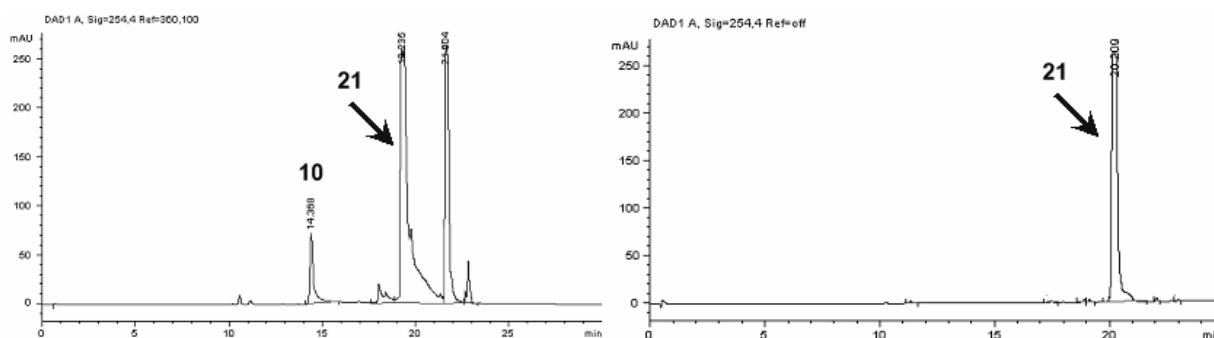
Figure 34: LC-MS analysis of the product resulting from the direct aminofluorescein protection with Fmoc chloride, after optimization of the reaction conditions. The crude product (a) is a mixture of starting material, mono- and bis-protected aminofluorescein, the expected product being predominant. The pure product (b) is isolated after silica gel chromatographic purification.

IV.2.2.2. Three-step synthesis through a protection-deprotection sequence

To avoid Fmoc addition on the phenolic -OH, another strategy was developed: the phenols should be first selectively protected with an acid labile group, and after Fmoc addition on the aniline, removed leaving pure *N*-Fmoc aminofluorescein.

a. Protection of the aminofluorescein phenolic hydroxyls

During solution phase experiments, different groups have been tested for aromatic hydroxyl protection on carboxyfluorescein or fluorescein isothiocyanate with more or less success: triethyl silyl, *tert*-butyldimethyl silyl and pivaloyl [111,118]. They were reconsidered here for use on aminofluorescein.

Aminofluorescein protection with silyl chlorides was tested first. The reaction was performed in DMF with an excess of imidazole. As a result a mixture of mono-protected product, starting material, and some bis-protected aminofluorescein was

obtained. This reaction has already been published [119]. However, no details regarding the yield, purity, or synthesized amounts have been given. Although product was obtained, what was probably sufficient for the work mentioned above, it was not pure enough to be adopted in the present strategy, where a very good yield and purity were crucial. Some variations of the reaction conditions were explored, such as the use of pyridine as solvent and base, and a range of temperatures from room temperature up to 80°C, without any improvement.

Pivaloyl protection was then tested. First, a general procedure with pivaloyl chloride [120,121] was used, but a mixture of different species was obtained. Protection with pivaloyl anhydride was tested next. This reagent had already been used for aminofluorescein protection by T. Hirano et al. [118] (Scheme 22). Cesium carbonate acted both as a base and a catalyst for selective protection of the phenolic groups with pivalic anhydride.
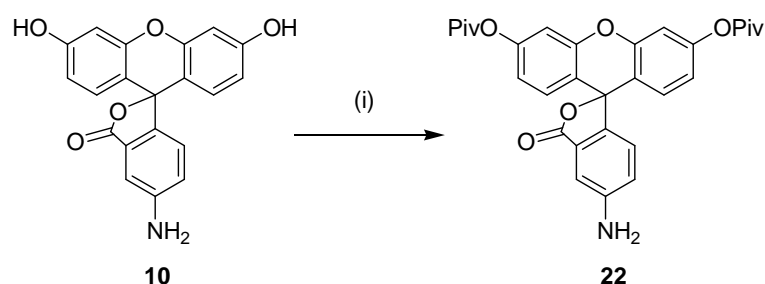
Scheme 22: Protection the aminofluorescein phenols with pivalic anhydride. (i) Cs_2CO_3, Piv_2O, DMF, t, T.
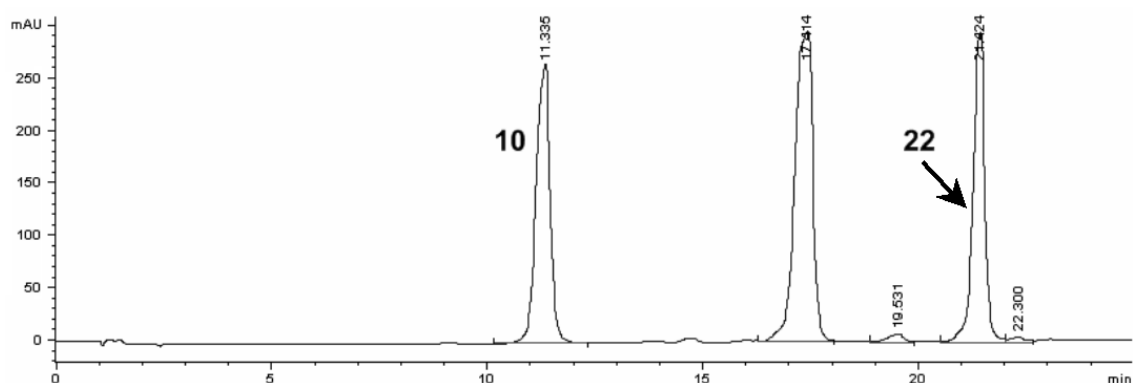
Figure 35: LC-MS analysis of the crude product resulting from the reaction of aminofluorescein with pivalic anhydride, before optimization of the reactional conditions. A mixture of unprotected (11.3 min), mono-protected (17.4 min) and bis-protected (21.4 min) aminofluorescein was obtained.

The results were not optimal, resulting in the formation of a mixture of unprotected, mono-protected and the expected bis-protected aminofluorescein **22** (Figure 35), but the conditions could be improved after a few attempts.

Optimizations were performed first on 5 mg scales, by varying the cesium carbonate and pivalic anhydride ratios, as well as the reaction time (Table 5). Increasing the cesium carbonate amount to 1.2 equiv and prolongation of the reaction time up to 60 min lead to formation of the desired product in up to 95% yield (Table 5, entries 5 and 6).

| Entry	Cs_2CO_3 / equiv	Piv_2O / equiv	t / min	T / °C	Ratio (0 Piv : 1 Piv : 2 Piv : 3 Piv)*
1	0.8	2.0	30	25°C	47 : 40 : **13** : 0
2	0.8	2.5	30	25°C	6 : 36 : **56** : 2
3	1.2	2.0	30	25°C	0 : 20 : **79** : 1
4	1.2	2.5	30	25°C	1 : 23 : **75** : 1
5	1.2	2.0	60	25°C	0 : 7 : **92** : 1
6	1.2	2.5	60	25°C	0 : 2 : **95** : 3

Table 5: Optimization of the aminofluorescein protection with pivalic anhydride at a low scale (5 mg aminofluorescein). * The ratio between the different species was determined by analyzing HPLC spectra at 254 nm.

A direct up-scale of these conditions to protect 100 mg aminofluorescein **10** resulted in a slight decrease of yield and purity (Table 6, entry 1). Small variations of the proportions did not improve it considerably, but once more, the operation mode was decisive: the slow addition of pivaloyl anhydride at lower temperature (0 °C) enabled to obtain a very good result (Table 6, entry 5) with a yield of 98% and a purity of 96% (**22**, Figure 36).

Entry	Cs_2CO_3 / equiv	Piv_2O / equiv	t / min	T / °C	Ratio (0 Piv : 1 Piv : 2 Piv : 3 Piv)*
1	1.2	2.5	60	25°C	0 : 9 : **82** : 9
2	1.0	2.1	60	25°C	1 : 13 : **84** : 2
3	1.0	2.2	60	25°C	0 : 5 : **86** : 9
4	1.1	2.5	60	25°C	0 : 5 : **91** : 4
5	1.1	2.5	60	0 - 25°C	0 : 2 : **96** : 2

Table 6: Optimization of the aminofluorescein protection with pivalic anhydride at a larger scale (100 mg aminofluorescein). * The ratio between the different species was determined by analyzing HPLC spectra at 254 nm.
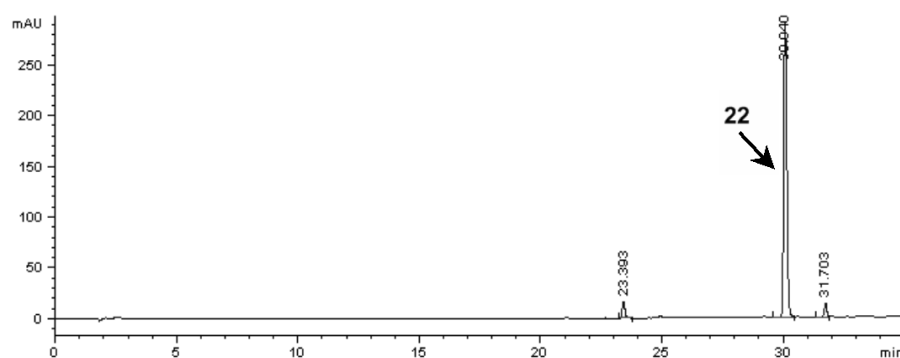
Figure 36: LC-MS analysis of the crude product after the protection of aminofluorescein using pivaloyl anhydride under reaction conditions described in Table 6, Entry 5. The 96% pure *N*-Fmoc aminofluorescein was used without further purification in the next step.

b. *Protection of the remaining aniline*

Fmoc protection of the aromatic amine of product **22** worked smoothly. Reaction conditions analogous to those already described in Part IV.2.2.1. were used, slightly increasing the excess of Fmoc chloride. The fully protected fluorophore was subsequently crystallized as a white powder **23** with a final yield of 85% over the two steps (Figure 37).
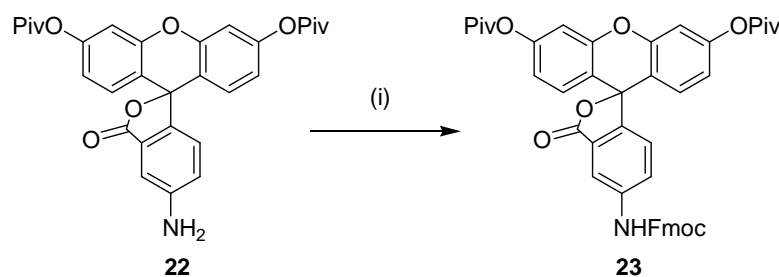
Scheme 23: Protection of the aniline of dipivaloylaminofluorescein 22 with Fmoc chloride. (i) Fmoc-Cl (1.5 equiv), NaOH (1.0 equiv), THF, 0 to 25°C, 180 min.
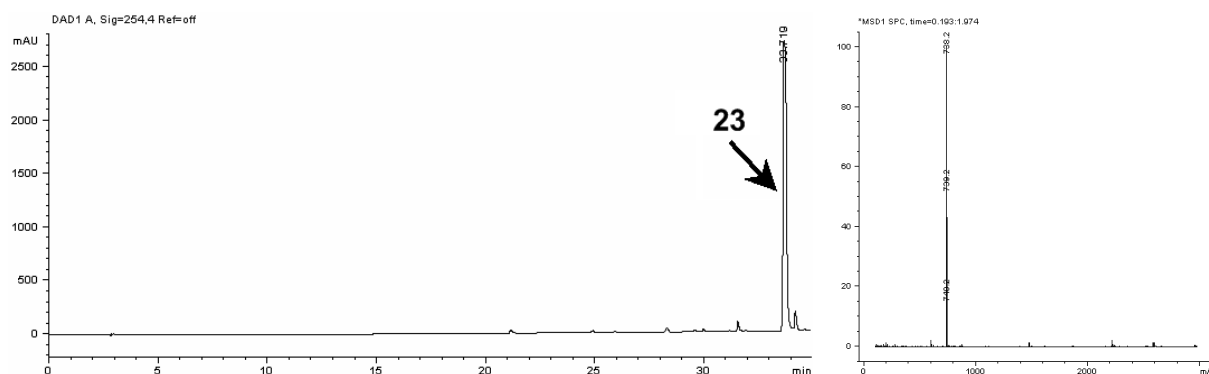
Figure 37: LC-MS analysis of the fully protected di-*O*-pivaloyl-*N*-Fmoc-aminofluorescein 23 after crystallization.

c. _Final cleavage of the pivaloyl groups_

Although pivaloyl groups are cleaved commonly under basic conditions – for example by treatment with sodium methoxide [122] – the deprotection can also be performed by treatment with a strong acid [123]. This way was selected here because of the presence of the base-labile Fmoc group on compound **23**.

Treatment with 95% TFA at room temperature for 3–6 hours led to the cleavage of the first pivaloyl group, whereas almost one day was necessary to cleave the second group. By increasing the temperature to 60 °C, complete pivaloyl deprotection could be achieved in only 3 h (Scheme 24). For the synthesis of large amounts (10–20 g), the reaction was conducted over night to ensure total deprotection. No further side-product formation was observed in that time (Figure 38) and after work-up, compound **21** was obtained in 92% yield.
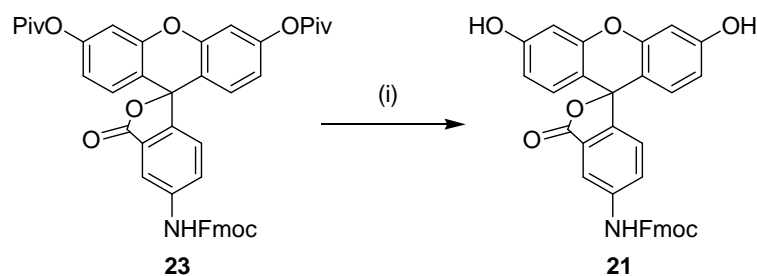
Scheme 24: Deprotection of di-_O_-pivaloyl-_N_-Fmoc-aminofluorescein 23 under highly acidic conditions to afford _N_-Fmoc aminofluorescein 21. (i) 95% aq. TFA, 60 °C, 180 min.
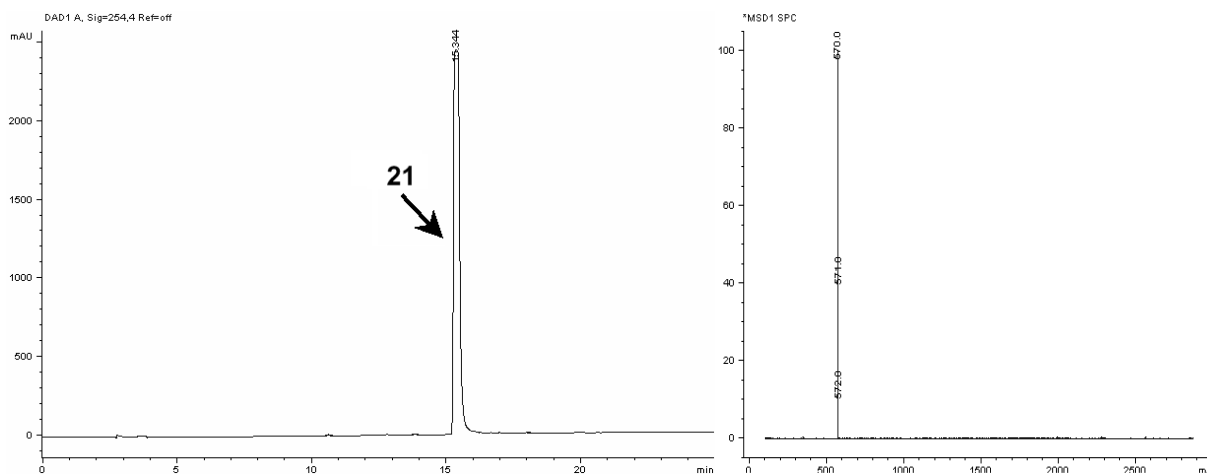
Figure 38: LC-MS analysis of _N_-Fmoc-aminofluorescein 21 resulting from the deprotection of di-_O_-pivaloyl-_N_-Fmoc-aminofluorescein 23.

IV.2.2.3. Comparison of the two synthesis routes for N-Fmoc-aminofluorescein
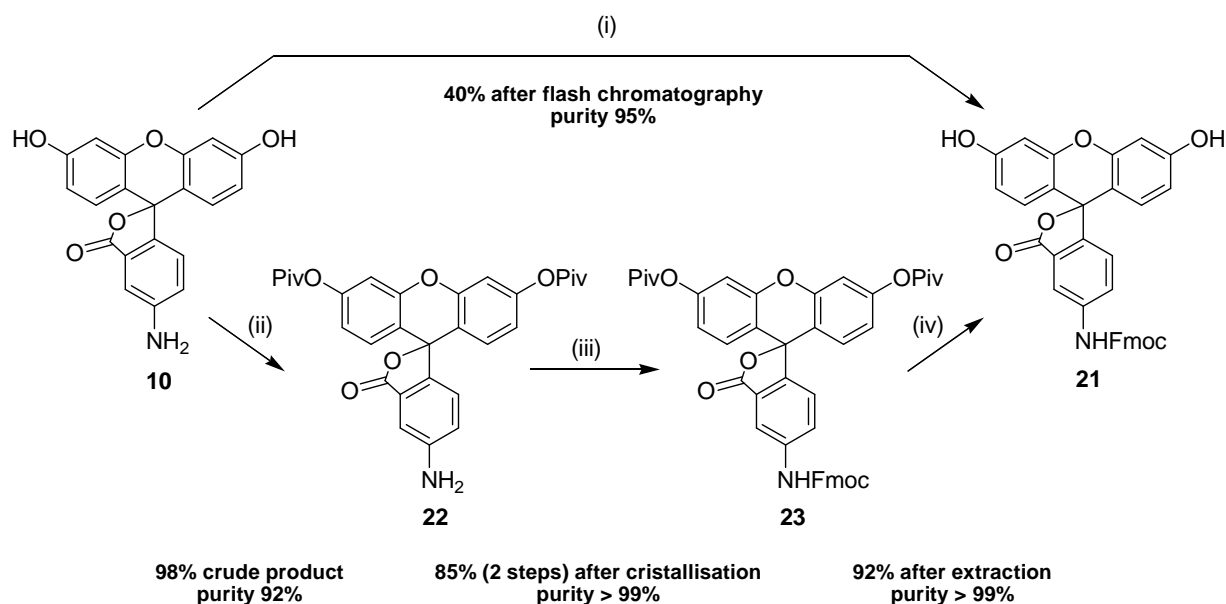
Scheme 25: Overview of the synthesis strategy for *N*-Fmoc aminofluorescein 21. A comparison of the two developed synthesis ways shows the superiority of a three-step approach, in particular for large-scale synthesis. (i) Fmoc-Cl (1.2 equiv), 1 M aq. NaOH (1.0 equiv), THF, 0°C to RT, 180 min; (ii) Cs$_2$CO$_3$ (1.1 equiv), Piv$_2$O (2.5 equiv), DMF, 0°C to RT, 60 min (iii) Fmoc-Cl (1.5 equiv), NaOH (1.0 equiv), THF, 0 to 25°C, 180 min; (iv) 95% aq. TFA, 60 °C, 180 min.

Direct protection of aminofluorescein **10** by addition of Fmoc chloride is rapid, but product **21** is only obtained when keeping conditions accurately, with a yield under 50% after a tedious chromatographic purification. The major side products – free and bis-protected aminofluoresceins – could be recycled to limit the material loss. The use of an MPLC in place of a hand-made column could also enable an automated and more efficient purification. However, high solvent volumes are necessary and a part from the material is nevertheless lost, so that this method is not suitable for large scale synthesis.

The second synthesis route consists of three steps and requires additional time, for example for crystallization, but delivers a very pure product without tedious purification. It was therefore selected for the synthesis of a larger amount of resin-bound aminofluorescein.

IV.2.3. Coupling of *N*-Fmoc-aminofluorescein 21 to 2-chlorotrityl chloride resin

Coupling of amino acids to 2-chlorotrityl chloride resin [124-126] is usually performed in DCM, but *N*-Fmoc aminofluorescein **21** is almost non-soluble in this solvent. Therefore, the coupling of **21** was first performed in DMF, with addition of an excess of the non-nucleophilic base DIPEA. The best coupling was obtained after 3 h (Table 7, entry 2) and could be increased neither by heating, by longer reaction times (Table 7, entry 3) nor by microwave-assisted synthesis (Table 7, entries 4–7).

A further improvement could be reached by reducing the DMF volume to the minimum necessary for *N*-Fmoc aminofluorescein dissolution and adding a larger volume of DCM to obtain a ratio of about 1:6. This enabled an increase of the yield from 35% to almost 50% (Table 7, entry 8), corresponding to a maximal resin loading of 0.38 mmol·g^{-1}.

This value could not be exceeded, probably because of partial cross-linking on the resin or steric hindrance of the quite large *N*-Fmoc aminofluorescein (M = 569 g·mol^{-1}). To validate this hypothesis, a test coupling could be done on a resin with lower loading, but no commercially available trityl chloride resin with lower loading was found. However, it was possible to reuse the *N*-Fmoc aminofluorescein solution (recovered after a first attachment reaction) on a new portion of trityl resin and hence limit the lost of *N*-Fmoc aminofluorescein **21**. Finally, the remaining reactive sites were capped by methanol.

Entry	Base	Solvent	Temperature / °C	Reaction time / min	Yield / %
1	DIPEA, 2.0 equiv	DMF	RT	60	13
2	DIPEA, 2.0 equiv	DMF	RT	180	35
3	DIPEA, 2.0 equiv	DMF	RT	1440	31
4	DIPEA, 2.0 equiv	DMF	MW, 60	3	2
5	DIPEA, 2.0 equiv	DMF	MW, 60	10	8
6	DIPEA, 2.0 equiv	DMF	MW, 80	3	2
7	DIPEA, 2.0 equiv	DMF	MW, 80	10	2
8	DIPEA, 2.0 equiv	DMF/DCM	RT	180	47

Table 7: Optimization of the coupling conditions for the attachment of *N*-Fmoc aminofluorescein 21 to 2-chlorotrityl chloride resin.

IV.3. Properties of resin-bound *N*-Fmoc-aminofluorescein

IV.3.1. Loading determination

Quantification of the coupling yield could be achieved by evaluation of the resin loading. Two methods were used: weighing and spectroscopic measurements.

IV.3.1.1. Loading determination by weighing

The entire resin was precisely weighed before the reaction and after the attachment of the first compound. The theoretical mass difference resulting from the various molecular masses of the leaving group (chloride) and the attached group (*N*-Fmoc aminofluorescein minus a proton) was calculated and compared to the experimentally measured mass difference (see Part VI.1.16.1)

Weighing is a general method that can be used for the attachment of all fragments, without having particular groups with a specific absorption. To get reliable values, this method should, however, be restrained to the coupling of fragments with a relatively high molecular weight – what was the case here with *N*-Fmoc aminofluorescein – so that the mass difference can be measured with enough precision. For the same reason relatively high resin amounts should be used. The resin should also be properly washed and dried to remove completely any trace of reagent or solvent.

IV.3.1.2. Loading determination by spectroscopic measurements

A small resin sample was taken from the reaction vessel, accurately weighed, and submitted to Fmoc cleavage in a defined volume of 20% piperidine / DMF (Scheme 26). The concentration of the fluorene was determined by absorption measurements using the Beer-Lambert law, and used to calculate the resin loading (see Part VI.1.16.2).

This method can be applied not only for all molecules having a specific absorption with known extinction coefficients, but also for unknown systems. Indeed, calibration curves can be registered with previously prepared solutions of defined concentrations. Thus, this method is broadly applicable.
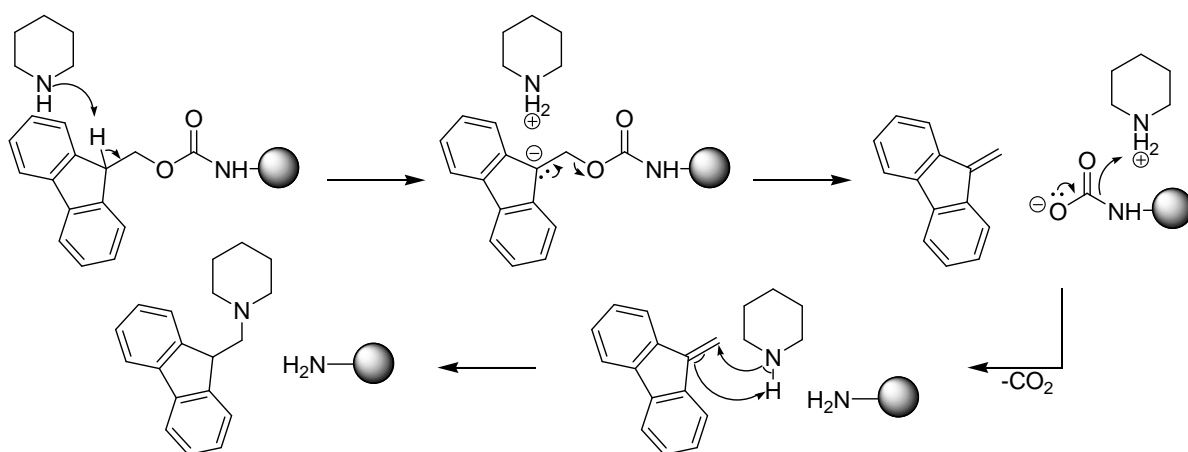
Scheme 26: Fmoc cleavage. The cleavage occurs by an E1cb (cb: conjugate base) mechanism with elimination of a halide ion with piperidine followed by release of carbon dioxide.

IV.3.1.3. Comparison of both methods

Both methods were compared on an example (coupling Table 7, entry 2). A yield of 33.3% was determined by weighing and 34.6% by Fmoc absorption, corresponding to a resin loading of 0.28 mmol·g^{-1}. Both results showed no significant difference, remaining within the error range. For further experiments, the UV absorption method was used due to its applicability on even very low resin amounts.

IV.3.2. Attachment mode of *N*-Fmoc aminofluorescein to 2-chlorotrityl chloride resin.

The aniline being protected with Fmoc, two possibilities remained for the attachment to the solid support: the free phenolate or the carboxylate resulting from an opening of the lactone could both substitute the trityl chloride (Figure 39).
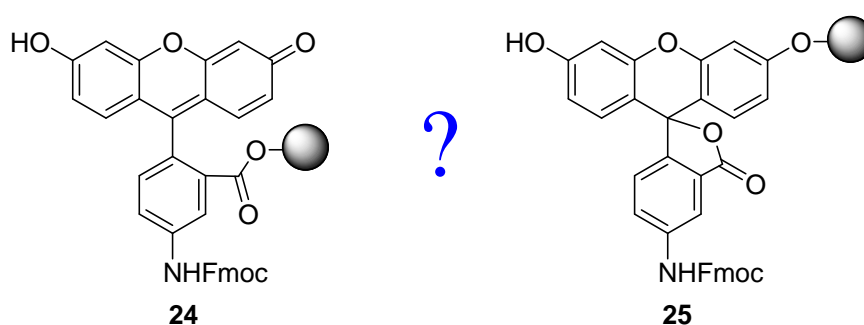
Figure 39: Two possibilities for *N*-Fmoc aminofluorescein attachment to the resin

IV.3.2.1. Expected attachment mode

Depending on the pH value, fluorescein can be present under four prototropic species, as already described in Part IV.1.2, Scheme 8. The most reactive site of the dianion is the phenolate: indeed, the transition from the dianion to the carboxylate monoanion occurs through protonation of the phenol.
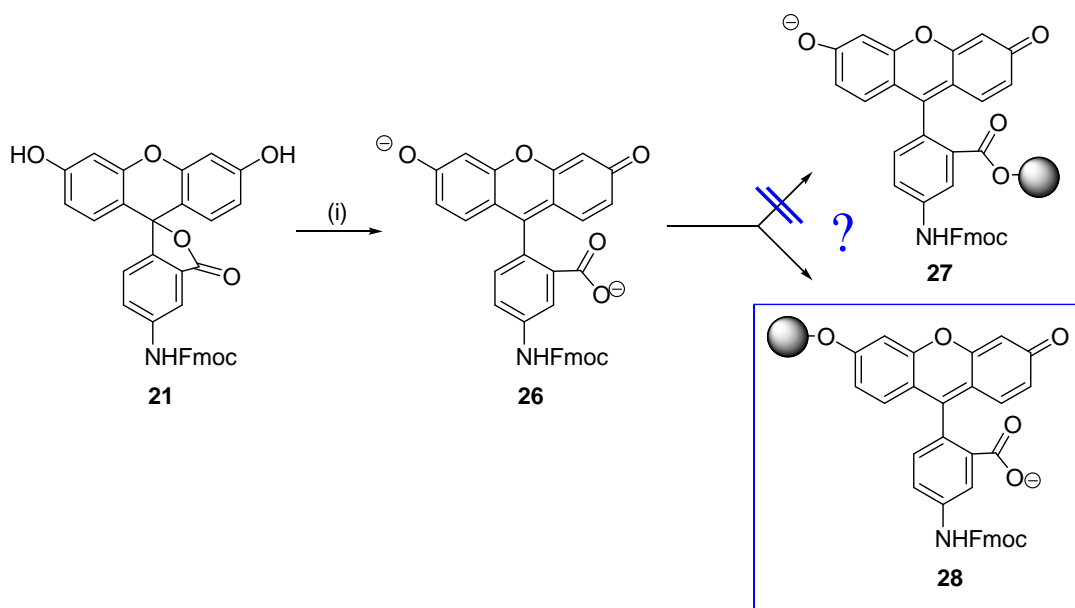
Figure 40: Expected attachment mode of aminofluorescein to the solid support. If the *N*-Fmoc aminofluorescein reacts with the trityl cation in the same way as fluorescein with a proton, product 14 should be formed. (i) Base.

Assuming that the coupling to the solid support – carried out under basic conditions resulting in a complete deprotonation of *N*-Fmoc aminofluorescein to the dianionic form 12 – occurs with the same reactivity, trityl ether 14 should be formed (Figure 40).

To verify this theory, the structure of the synthesized product had to be confirmed experimentally. For this purpose, two experiments were carried out: the NMR analysis of an analogous tritylation in solution, and the ATR-IR analysis of *N*-Fmoc aminofluorescein resin.

IV.3.2.2. NMR analysis of N-Fmoc aminofluorescein tritylation in solution

The reaction of *N*-Fmoc aminofluorescein **21** with trityl chloride was performed in solution [127] in the presence of DIPEA. Deuterated DMF was used as a solvent to enable direct monitoring by ^{1}H-NMR analysis. The spectrum of *N*-Fmoc-

aminofluorescein in DMF was recorded (Figure 41, in blue). After addition of DIPEA (1.0 equiv), the spectrum was measured again and no changes could be observed. One equivalent of trityl chloride was then added, and a new spectrum was recorded after 1 h (Figure 41, in red). Finally, after addition of a further equivalent DIPEA and trityl chloride, and one further hour reaction, a last spectrum was recorded (Figure 41, in green).
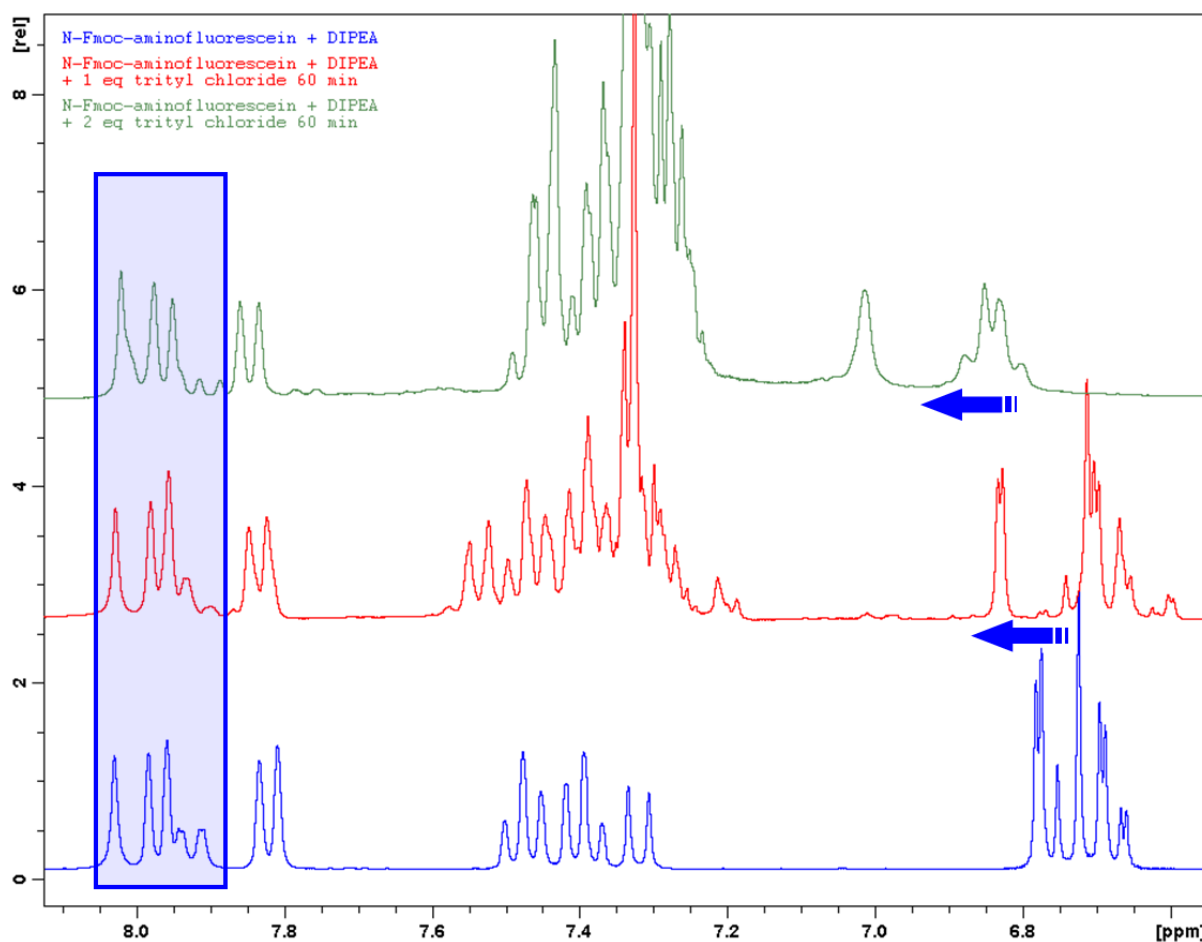
Figure 41: ^{1}H-NMR analysis of *N*-Fmoc-aminofluorescein tritylation in solution. After addition of 1.0 equiv Trt-Cl (in red), a change in the xanthene signals can be observed in comparison to *N*-Fmoc aminofluorescein (in blue). This shows that both parts of the xanthene moiety are no more equivalent and formation of a mono-tritylated phenol occurs. After further addition 1.0 equiv Trt-Cl (in green), the original form of the xanthene signals is recovered and a shift is observed. This shows the attachment of trityl on both phenols, the xanthene moiety recovering its symmetry. In all cases, the aniline signals are not affected, so that the carboxylate should not have been tritylated.

On the left side, the range corresponding to the three protons of aniline (see Part VI.2.1) remained unchanged during the whole experiment, whereas the right side corresponding to the six protons of the xanthene moiety has been varying. After

addition of 1.0 equiv trityl chloride, the xanthene signals were shifted and additional signals appeared, showing that the molecule was no more symmetric. When an excess of trityl was added, the symmetry of the molecule was recovered but the signals were shifted further more.

From this analysis, it can be concluded that a phenol group – and not the lactone – reacted first with the trityl, when this was stoichiometric added. With an excess of trityl, both phenols reacted.

So, for the attachment of *N*-Fmoc aminofluorescein to 2-chlorotrityl chloride resin, the reaction should also occur on the same way resulting in a phenol ether bond and so delivering the product expected from the theoretical analysis.

IV.3.2.3. ATR-IR analyses

In IR measurements, vibrations of an ester carbonyl can be differentiated from vibrations of a lactone carbonyl [128]. Benzoate esters are found in the region of 1730–1715 cm^{-1}. Saturated γ-lactones (five-membered ring) absorb in the 1795–1760 cm^{-1} region; an unsaturation α to the C=O reduces the absorption frequency whereas an unsaturation α to the –O– group increases it.
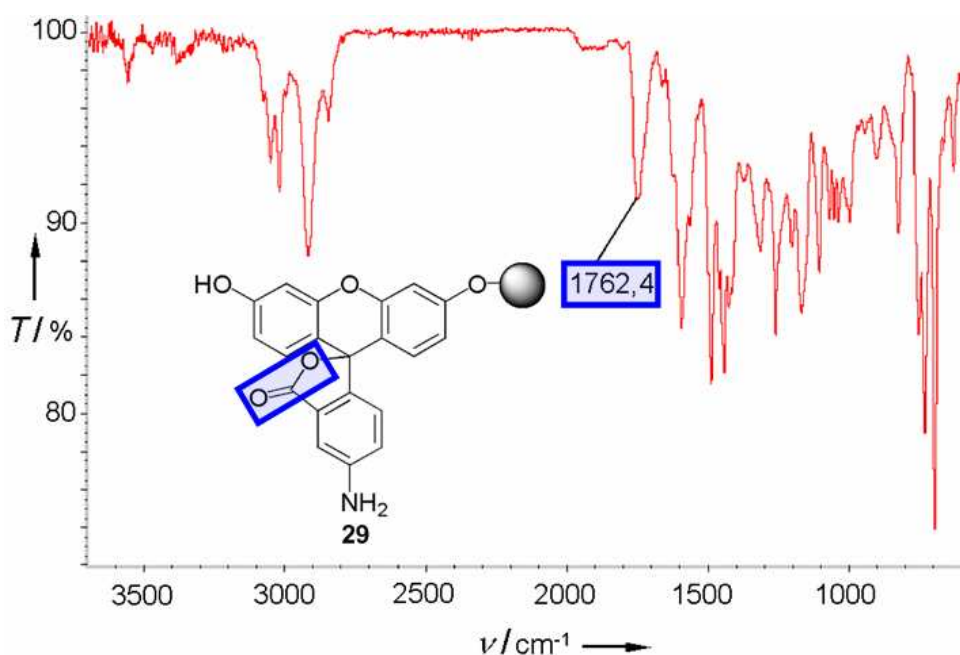
Figure 42: ATR-IR spectrum of resin-bound aminofluorescein. The signal by 1762 cm^{-1} reveals the presence of a γ-lactone with an unsaturation α to the C=O, confirming the attachment of *N*-Fmoc aminofluorescein to the solid support by the phenol group.

This can be used for determination of the attachment point of fluorescein to 2-chlorotrityl chloride resin. Indeed, if *N*-Fmoc aminofluorescein is attached to the resin via the carboxylate, an ester carbonyl should be found in the spectrum, whereas if it is bound via the phenol ether, the lactone ring remains closed, and a lactone carbonyl signal should be visible.

N-Fmoc-aminofluorescein was analyzed with this method, but the signal of the carbonyl of interest was covered by the Fmoc carbonyl. For a better read-out, Fmoc was cleaved and the spectrum recorded once more. Only one peak remained in the carbonyl range, with a value of 1762 cm^{-1} (Figure 42) corresponding to a lactone. This analysis confirmed the exclusive formation of the phenol ether bound product, confirming the results obtained from NMR analysis of the solution phase reaction.

IV.3.2.4. Conclusion: attachment to the resin

Both ATR-IR analysis of resin-bound aminofluorescein and NMR analysis of the analogous *N*-Fmoc-aminofluorescein tritylation in solution confirmed the theoretically predicted attachment mode through the phenolate.

IV.4. Reactivity of resin-bound aminofluorescein

IV.4.1.   Reactivity towards amino-acids: amide coupling

IV.4.1.1. Anilines: poor nucleophiles

In comparison with aliphatic amines, and in particular with secondary or primary amines, anilines are rather poor nucleophiles. Indeed, the doublet of the nitrogen atom is delocalized over the aromatic benzene ring, so that the electrons are less available for nucleophile attack.

In the case of aminofluorescein, this phenomenon is even stronger due to the presence of the additional xantene ring system that further extends the aromatic system (Scheme 27). Therefore, for reactions involving aminofluorescein as nucleophile, specific conditions have to be used to reach a complete conversion.
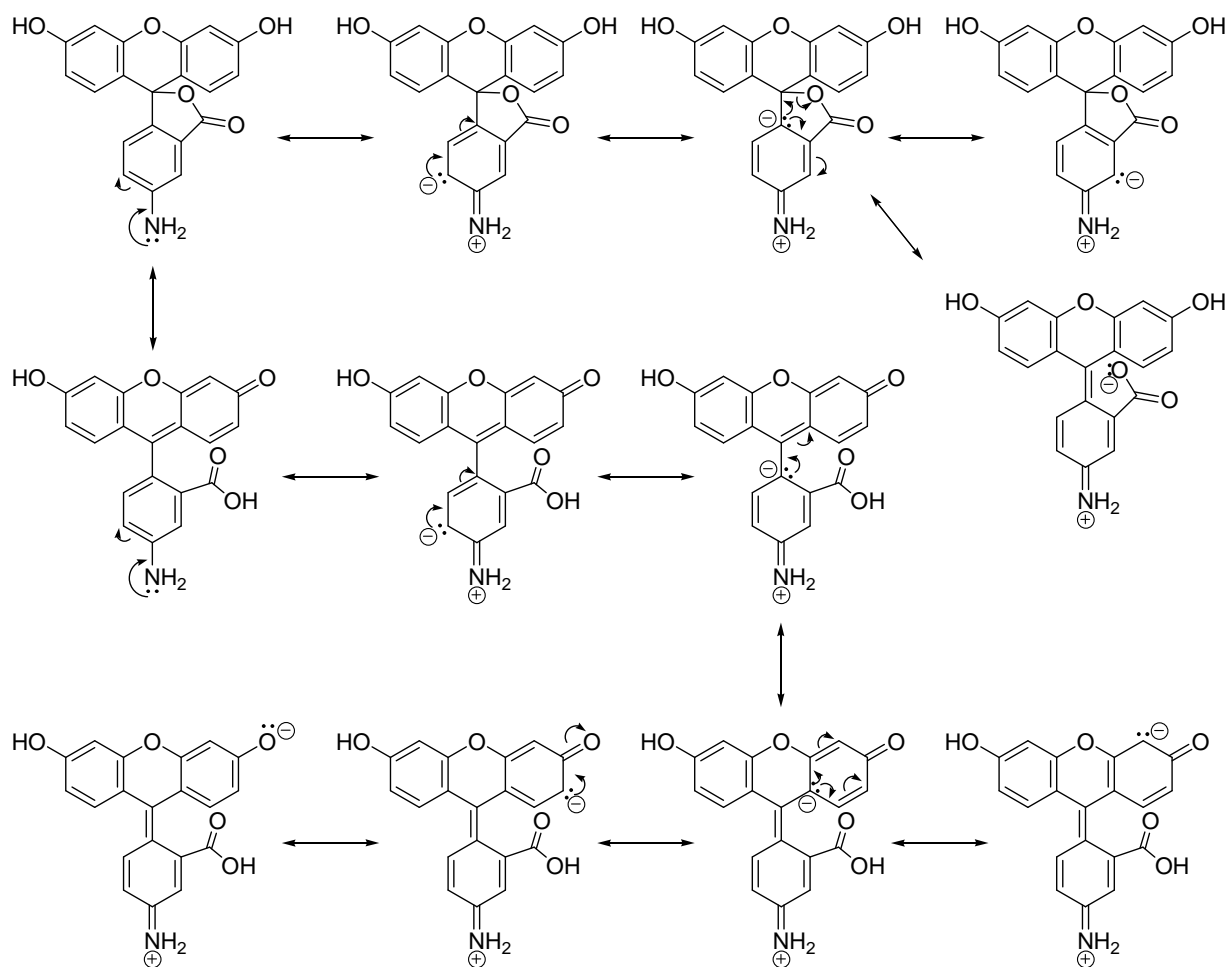
Scheme 27: Aminofluorescein, a bad nucleophile. Due to the delocalization of the nitrogen electron pair over the whole conjugated system, the nucleophilic character of the aminofluorescein aniline is rather weak.

IV.4.1.2. Specific coupling conditions

Different coupling conditions were tested. With the method traditionally used by SPPS – DIC / HOBt as activating reagent (Table 8, Entry 1) – almost no coupling occurred. Better results were obtained by changing the solvent to DCM (Table 8, Entry 5). By removing HOBt, a further improvement could be reached. Indeed, HOBt is used to avoid racemization in peptide synthesis, but its use also decreases the reaction rate, so that it has been expected that its absence would be beneficial for a difficult coupling. The best results were finally obtained using DIC in a mixture of THF / DCM (1:1) with some DMF (Table 8, Entry 6) – as little as necessary for complete dissolution of the amino acid. This procedure had already been used for the coupling of aminocoumarin [129]. As an alternative, the more expensive HATU could also be used with DIPEA in DCM or DMF (Table 8, Entries 7 and 8).
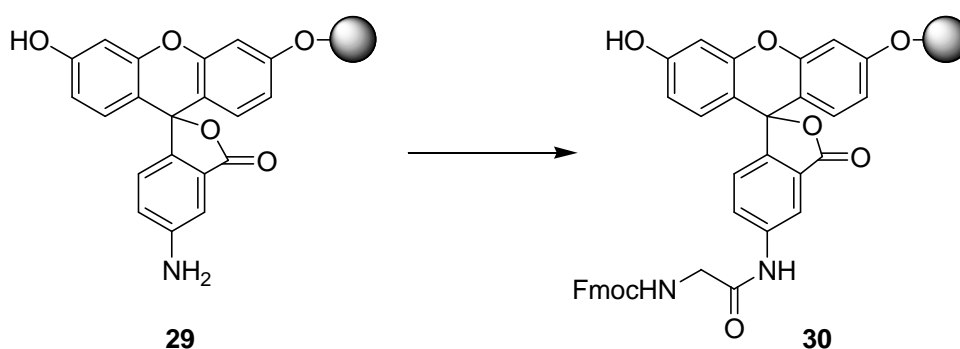
Entry	Amino acid (5 equiv)	Activating agent (5 equiv)	Base (5 equiv)	Solvent	Conversion / %
1	Fmoc-Gly-OH	DIC / HOBt	-	DMF	0
2	Fmoc-Gly-OH	DIC	-	DMF	0
3	Fmoc-Gly-OH	DIC	-	THF	9
4	Fmoc-Gly-OH	DIC	-	DCM	77
5	Fmoc-Gly-OH	DIC / HOBt	-	DCM	68
6	Fmoc-Gly-OH	DIC	-	THF/DCM (1:1) *	92
7	Fmoc-Gly-OH	HATU	DIPEA	DMF	92
8	Fmoc-Gly-OH	HATU	DIPEA	DCM	91
9	Fmoc-Gly-OH	HATU	collidine	DMF	77
10	Fmoc-Gly-OH	HATU	collidine	DCM	73
11	Fmoc-Gly-OH	HATU	2,6-lutidine	DMF	60
12	Fmoc-Gly-OH	HATU	2,6-lutidine	DCM	50
13	FmocNH-CH$_2$-COF	-	DIPEA	DCM	78

Table 8: Coupling of Fmoc-amino acids to of resin-bound *N*-Fmoc-aminofluorescein. (i) Amino acid, activating agent, base, solvent, 90 min, RT; (ii) 20 % piperidine in DMF, RT, 2 × 15 min. * with the addition of some drops DMF for complete dissolution

IV.4.2. Protection of the phenols of resin-bound *N*-Fmoc-aminofluorescein

The phenols of resin-bound *N*-Fmoc-aminofluorescein **25** are good nucleophiles and react also in the presence of an electrophile. In the case of repeated amino acid couplings, the formed ethers are cleaved during the piperidine treatment subsequent to the coupling. However, in some cases, it can be usefull to protect these phenols, e.g. to avoid side-products and save starting material, or when no piperidine treatment is done.

The right protecting group should be cleavable under acidic conditions, removed together with the final cleavage from the resin, and should be easy to eliminate from the solution, e.g. by evaporation.

Three different groups were investigated for this purpose: 2-methoxyethoxymethyl (MEM), *tert*-butyldimethyl silyl (TBDMS), and triethylsilyl (TES) ethers, yielding resins **31a**, **31b** and **31c** respectively (Scheme 28).

The efficient applicability of these resins was evaluated by LC-MS analyses: after the phenols protection step, the Fmoc group was removed, an Fmoc-glycine was coupled, and the resulting product was cleaved from the resin and analyzed in LC-MS.

All three groups provided product **34** with high purity in excellent yields. The silyl ethers, however, were slightly labile on base treatment, whereas the MEM group was totally stable under these conditions and was therefore selected for the subsequent synthesis.

25

31a: R=MEM
31b: R=TBDMS
31c: R=TES

32a: R=MEM
32b: R=TBDMS
32c: R=TES

34

33a: R=MEM
33b: R=TBDMS
33c: R=TES
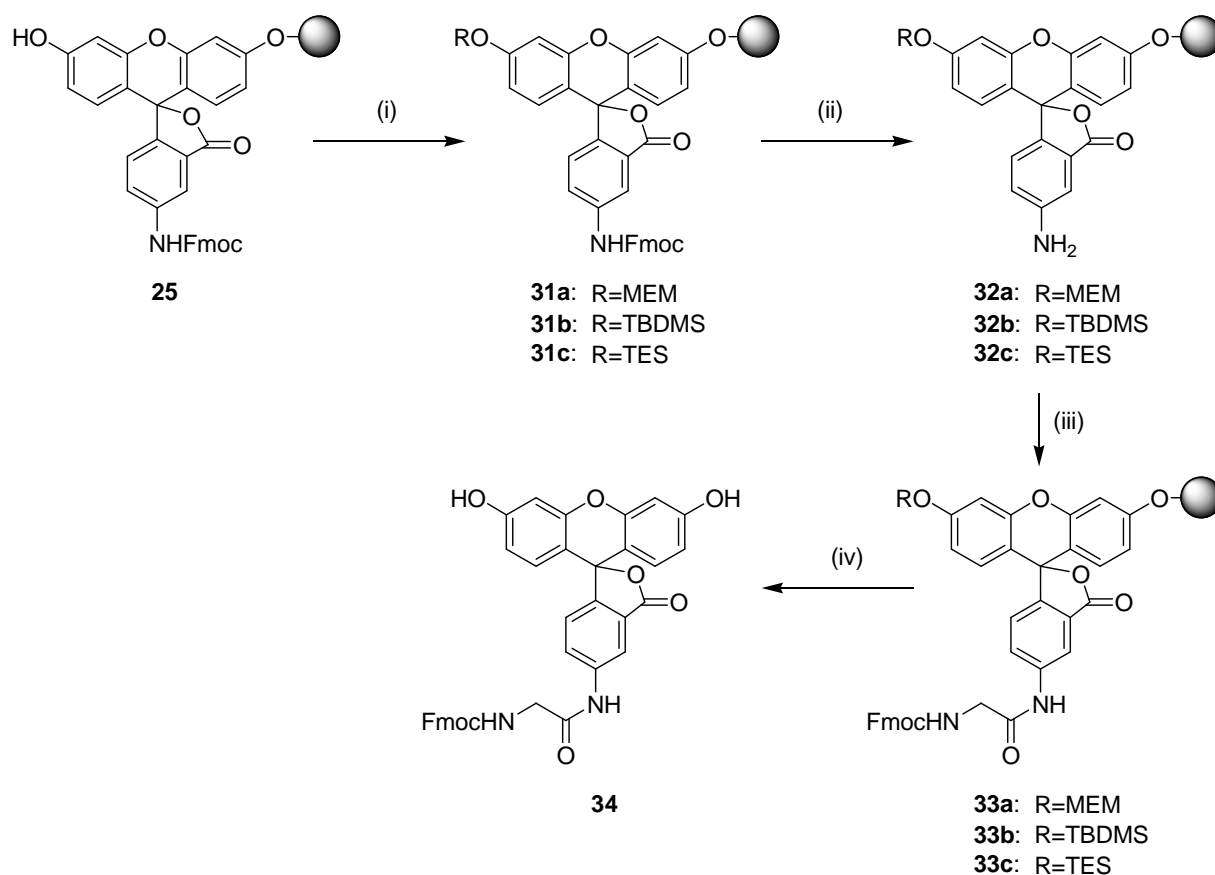
Scheme 28: Protection of the phenols of resin-bound *N*-Fmoc-aminofluorescein. (i) MEM-chloride (xa) or TBDMS-Cl (xb) or TES-Cl (xc), DIPEA, DMF, RT, 2 × 120 min; (ii) 20 % piperidine in DMF, RT, 2 × 15 min ; (iii) Fmoc-glycine, DIC, THF / DCM, RT, 60 min; (iv) 95 % TFA / H_2O.

IV.4.3. Conversion into an electrophile: α-bromoacetamide formation

IV.4.3.1. Reaction of resin-bound aminofluorescein 25 with bromoacetyl chloride

The coupling of bromoacetyl chloride to an amine or an aniline is known to form selectively the bromoacetylamide [130,131]. The reaction was tested here without real success and due to the corrosive character of bromoacetyl chloride and its difficult handling and storage, the equivalent acid [132] was preferred.

IV.4.3.2. Reaction of resin-bound aminofluorescein 25 with bromoacetic acid
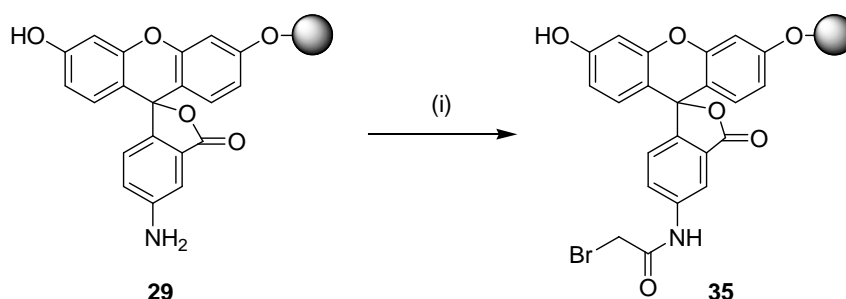
Scheme 29: Coupling of bromoacetic acid to resin-bound aminofluorescein. (i) Bromoacetic acid (5 equiv), DIC (5 equiv) in THF / DCM (1:1), 15 min, RT.

Bromoacetic acid coupling was performed using the previously determined conditions (Scheme 29). Reaction completion was checked by the cleavage of a small aliquot of the resin with 20% HFIP in DCM and its analysis in LC-MS.
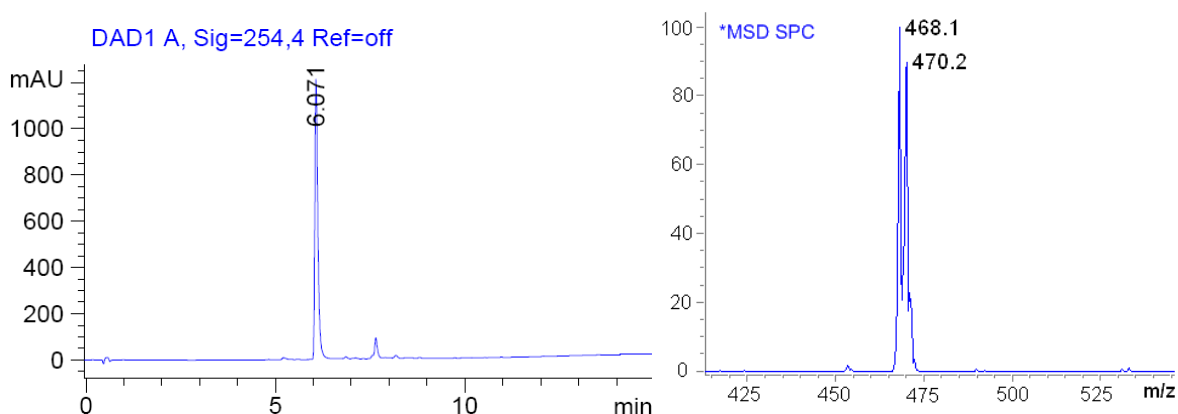
Figure 43: LC-MS analysis of bromoacetylamidyl fluorescein 35

The expected product **35** was obtained already after few minutes with 95% purity, the remaining 5% being the di-acetylated compound (Figure 43). This can either be cleaved of using piperidine after the nucleophilic substitution, or avoided using MEM-protected aminofluorescein resin as starting material. The resulting reagent was used for labeling small molecules by nucleophilic substitution.

IV.4.3.3. Nucleophilic substitution on bromoacetic acid

The obtained bromoacetylamidyl fluorescein resin **35** was used for the labeling of small molecules by nucleophilic substitution. An example is shown on Scheme 30:
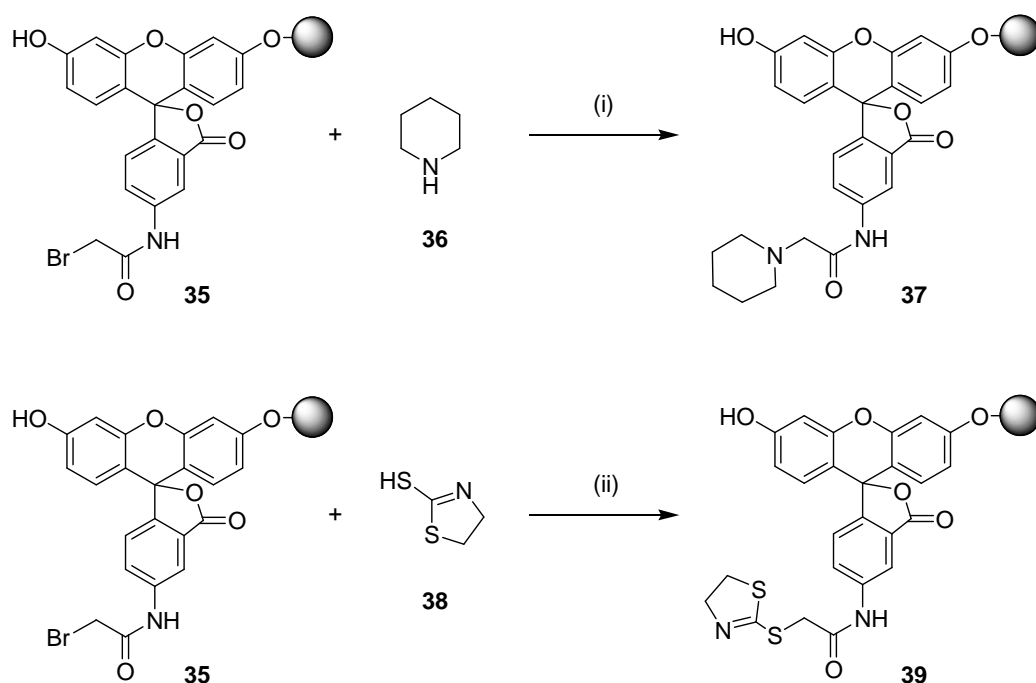
Scheme 30: Nucleophilic substitution on bromoacetylamidyl fluorescein resin. (i) RT, 60 min, conversion: 100 %, purity: 98 %; (ii) DCM, RT, 180 min, conversion: 100 %, purity: 92 %.

The reaction was further done with some more thiols, amines, and phenols with variable success. The conversion depended on the nucleophilicity of the small molecule. However, the reaction conditions should be further studied – influence of the solvent, reaction time and temperature – and improved to really cover a broad range of compounds with an optimal labeling yield.

IV.5. Application 1: Synthesis of a library of *C*- and *N*-terminally fluorescein labeled peptides to be used as fluorescent probes for CD2BP2 and PERQ2 GYF domains

Peptides are very useful tools for the investigation of the role of proteins [133]. Indeed, protein-protein interactions occurring in tissues can be reduced to protein-peptide interactions by a judicious choice of the protein fragment. The resulting peptide can be used in assays for functional elucidation.

Whereas peptides are traditionally labeled at their *N*-terminus using carboxyfluorescein (see Part II.3.4), the synthesized *N*-Fmoc aminofluorescein resin enables peptide labeling at the *C*-terminus. Libraries of *C*- and *N*-terminally fluorescein labeled peptidic ligands for two proteins of the GYF family (CD2BP2 and PERQ2) have been synthesized and tested in FP assays. They were used to develop a fluorescence probe for a competitive FP screening in order to find non-peptidic ligands for CD2BP2 and PERQ2.

IV.5.1.   GYF domains

IV.5.1.1. Proline-Rich Sequences

Protein-protein interactions involving small adaptor domains occur over short regions, often less than ten amino acids. Proline is critical for many of these interactions. This is the only amino acid having its side-chain cyclized on the backbone nitrogen; polyproline sequences tend therefore to adopt the PP II helix, a particular structure where a continuous hydrophobic surface is formed at the external part of the helix. Such helixes are found in known globular proteins, but are probably even more common in proteins that have not yet been structurally characterized because of the difficulty to analyse such regions with NMR or X-ray diffraction [134].

IV.5.1.2. The CD2BP2 and PERQ2 GYF domains

The GYF domains under study were derived from the human proteins CD2BP2 and

PERQ2 (Swiss-Prot entry name O75137). CD2BP2 was identified in a yeast two-hybrid screen as a binding partner of the T cell adhesion protein, CD2. NMR studies showed that the last 62 amino acids of CD2BP2 fold into a compact domain, the CD2BP2 GYF domain [108,135]. Additional characteristics of CD2BP2-type GYF domains are their strict localization to the very *C*-terminus and their absence in plant proteins.

The majority of GYF domains are, however, localized in the center of the respective proteins and share a shorter loop between strands β1 and β2. Furthermore, they predominantly contain aspartate at position 8 instead of tryptophan and lack phenylalanine at position 58. Therefore, these GYF domains constitute a second subgroup termed SMY2 subfamily of GYF domains, referring to its most prominent member, suppressor of myo2-66. The PERQ2 GYF domain is a member of this GYF family.

IV.5.2.   Design of libraries of fluorescent probes for FP assays

IV.5.2.1. Libraries of peptides for the PERQ2-GYF domain

As a starting point for the selection of an appropriate peptide sequence, the investigations previously conducted by M. Kofler (group of C. Freund, FMP, Berlin) [136] were used. Substitution analyses of individual peptides selected by phage display had been performed to identify residues within the wild-type binding peptide that were important for binding.

As shown on Figure 44, the central amino acids P, P, G, and L could not be exchanged without complete loss of peptide affinity. Thus, this central motif was conserved, whereas neighboring amino acids were modified according to the results of substitution analyses. The length of the peptide and the position of the fluorescent label were varied. Furthermore, both *C*- and *N*-terminally labeled peptides were compared. A second peptide had been also found to have a great affinity to PERQ2 (M. Kofler, unpublished results) and was used as a basis for other fluorescent peptides. All peptides to be synthesized are listed in Table 9.
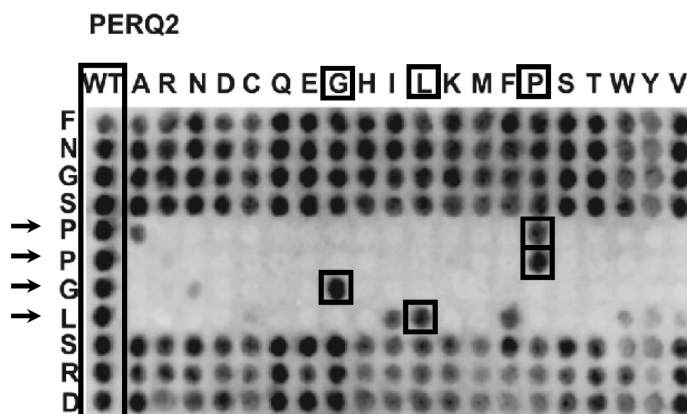
Figure 44: Substitution analysis of peptides selected by phage display (data from M. Kofler, group of C. Freund, FMP, Berlin [136]). All possible single substitution analogs of the peptide were synthesized on a membrane. The single letter code above each column indicates the amino acid that replaces the corresponding wild-type residue; rows define the position of the substitution within the peptide. Spots in the far left column (WT) represent the wild-type peptide.

Peptide sequence (*C*-terminal labeling)	Peptide sequence (*N*-terminal labeling)
Ac-F N G S **P P G L** G G G Fluo	Fluo F N G S **P P G L** G G
Ac-F N G S **P P G L** E R G Fluo	Fluo **P P G L** S R D
Ac-F N G S **P P G L** G G Fluo	Fluo S **P P G L** S R D
Ac-F N G S **P P G L** G Fluo	Fluo N G S **P P G L** S R D
Ac-F N G S **P P G L** Fluo	Fluo W R P G P P P P P **P P G L** V
Ac-W R P G P P P P P **P P G L** V Fluo	Fluo R P G P P P P P **P P G L** V
Ac-W R P G P P P P P **P P G L** Fluo	Fluo P G P P P P P **P P G L** V
	Fluo W R P G P P P P P **P P G L**

Table 9: Fluorescein labeled peptidic ligands for the PERQ2 GYF domain. *C*- and *N*-terminally labeled peptides were selected conserving the central PPGL motif and varying the neighbor amino acids as well as the distance to fluorescein.

IV.5.2.2. Library of peptides for the CD2BP2-GYF domain

The same strategy was adopted for CD2BP2 binding peptides. The central PPG motif was conserved and the neighboring amino-acids were varied taking into account the substitution analysis (M. Kofler, Figure 45) as well as the sequence of another good binding peptide previously evaluated in NMR experiments. Peptides chosen for synthesis are listed in Table 10.
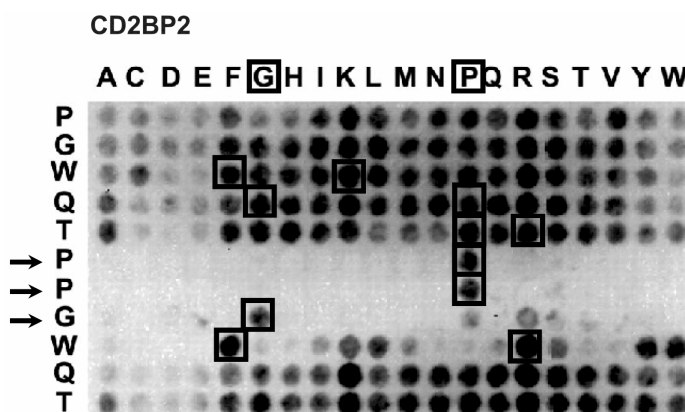
Figure 45: Substitution analysis of peptides selected by phage display (data from M. Kofler, group of C. Freund, FMP, Berlin)[91]. All possible single substitution analogs of the peptide were synthesized on a membrane. The single letter code above each column indicates the amino acid that replaces the corresponding wild-type residue; rows define the position of the substitution within the peptide.

Peptide sequence (*C*-terminal labeling)	Peptide sequence (*N*-terminal labeling)
Ac-E F G P **P P G** W L G R Fluo	Fluo E F G P **P P G** W L G R
Ac-E F G P **P P G** W K G Fluo	Fluo E F G P **P P G** W L G
Ac-E F G P **P P G** W K Fluo	Fluo E F G P **P P G** W L
Ac-E F G P **P P G** R K P Fluo	Fluo E F G P **P P G** W
Ac-E F G P **P P G** R K Fluo	Fluo E G P **P P G** W L G R
Ac-E F G P **P P G** F Fluo	Fluo G P **P P G** W L G R
	Fluo P **P P G** W L G R
	Fluo **P P G** W L G R
	Fluo E F G P **P P G** W K G

Table 10: Fluorescent labeled peptidic ligands for the CD2BP2 GYF domain. *C*- and *N*-terminally labeled peptides were selected conserving the central PPG motif and varying the neighbor amino acids as well as the distance to fluorescein.

IV.5.3. Synthesis of fluorescent probes for PERQ2 and CD2BP2

*IV.5.3.1. Synthesis of C-terminally fluorescein labeled peptides **43a – 43m***

All *C*-terminally labeled peptides were synthesized on the aminofluorescein resin **32a** as shown on Scheme 31. The first peptide coupling was done following the conditions described in Part IV.4.1.2 and product **40** was obtained. After reaction completion, the

Fmoc protecting group was removed and the remaining amino acids were coupled (**41**) using carbodiimide activation (Fmoc-AA, DIC, HOBt, DMF, 30 min). After each coupling step, the absence of free amine was checked with the Kaiser test (VI.1.17), or with the chloranil test in the case of a proline (VI.1.18).

The last amino acid was acetylated (**42**), and the final cleavage from the resin was performed with 95% TFA / H_2O + 2% triisopropylsilane with addition of a phenol scavenger for peptides containing a Pmc protecting group. Peptides **43** were finally precipitated by addition of the cleavage solution to cold diethyl ether (4 °C), isolated by centrifugation and lyophilized from *tert*-BuOH / H_2O 4:1 (v/v).
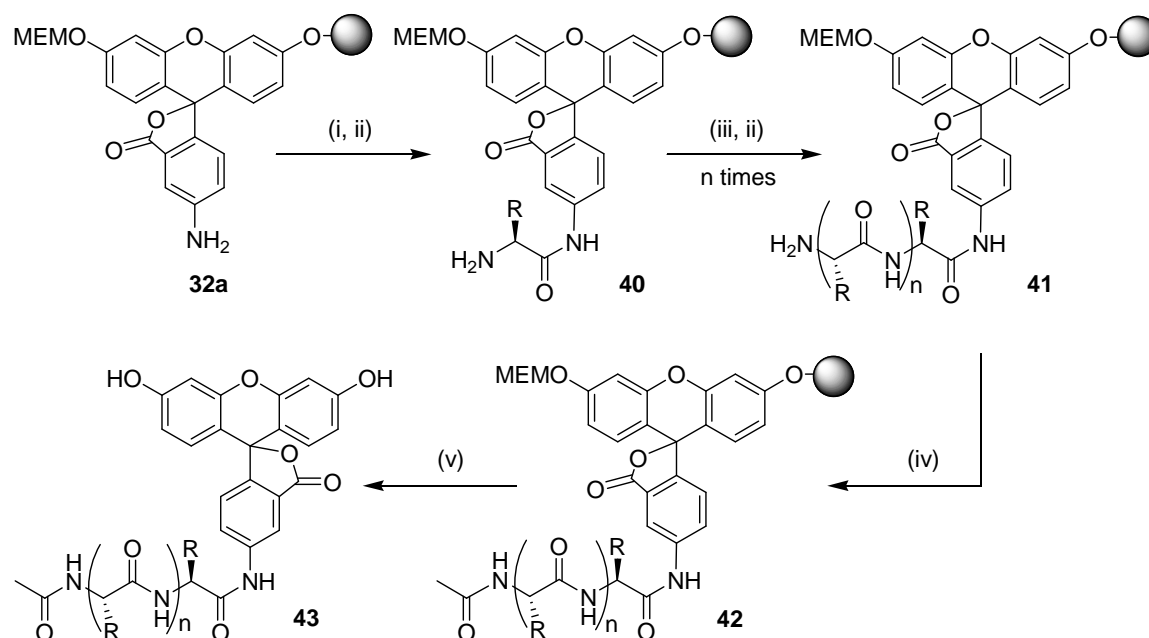
Scheme 31: Synthesis of *C*-terminally fluorescein labeled peptides. (i) Fmoc-AA, DIC, THF / DCM, RT, 60 min; (ii) 20% piperidine in DMF, RT, 2 × 15 min; (iii) Fmoc-AA, DIC, HOBt, DMF, 30 min; (iv) Ac_2O, DIPEA, DMF, RT, 15 min; (v) 95% TFA in H_2O, PhOH (if Pmc protecting group), TMS (if Trt protecting group).

Crude peptides were obtained in a similar purity as peptides resulting from traditional Fmoc SPPS (80 to 90%). In particular peptides with long polyproline sequences[137] **43m** and **43l** were obtained with a very high purity. Finally, the peptides were purified on a preparative HPLC column to reach purities above 90% (cf. Table 13, peptides **43a – 44m**).

*IV.5.3.2. Synthesis of N-terminally fluorescein labeled peptides **44a – 44s***

All *N*-terminally fluorescein labeled peptides were synthesized with traditional methods of SPPS and labeled with 5-(6)carboxyfluorescein before resin cleavage (see II.2.3.2 and VI.2.3). All crude peptides obtained from the synthesizer were purified on preparative HPLC columns to reach purities above 90% (cf. Table 14, peptides **44a – 44s**).

IV.5.4. Fluorescence properties of *C*-terminally fluorescein labeled peptides

Fluorescence properties of the synthesized *C*-terminally labeled peptides were investigated and compared with the fluorescence properties of fluorescein, aminofluorescein, and *N*-terminally labeled peptides. The fluorescence absorption and emission spectra were recorded and the quantum yields of the different compounds calculated by comparison of the areas of unknown compounds with the area of a fluorescein standard solution with known quantum yield of fluorescence.
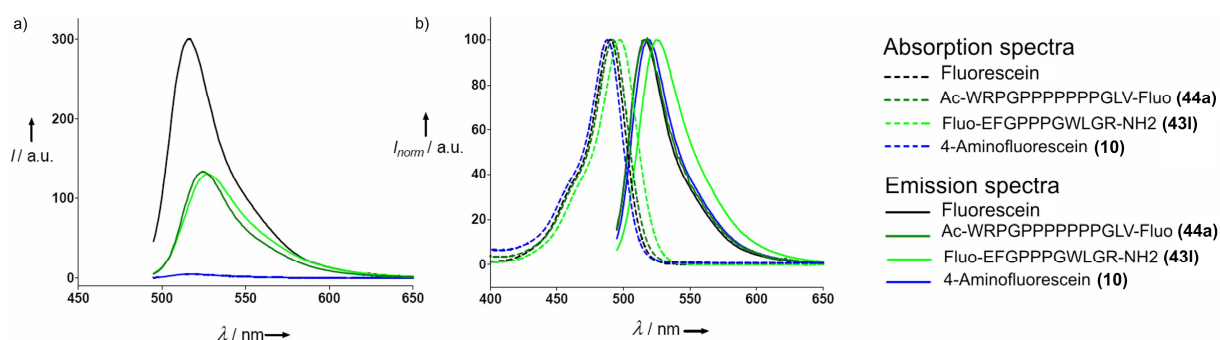
Figure 46: Compared fluorescence spectra of *N*- and *C*-terminal fluorescein labeled peptides

Whereas fluorescein is highly fluorescent (Figure 46 in black)[138], 4-aminofluorescein **10** is practically non-fluorescent under its amine form[139,140] (Figure 46 in blue). Only after conversion into an amide by coupling of an amino acid, the molecule recovers its fluorescence (Figure 46, peptide **44a** in dark green).

The fluorescence absorption and emission maxima of *C*-terminally labeled peptide **43l** are the same as these of fluorescein and aminofluorescein **10**, whereas these of *N*-terminally labeled peptide **44a** are slightly shifted – remaining nevertheless in the

spectral domain delimited by fluorescein filters. Fluorescence quantum yields of both labeled peptides are in the same range, namely almost 0.5.

As a result, it was shown that the novel *C*-terminally labeled and the classical *N*-terminally labeled peptides have similar excitation and emission maxima as well as fluorescence quantum yields, so that they are also convenient for fluorescence assays.

Compound	ϕ_{fl}	λ_{abs} / nm	λ_{em} / nm
Peptide **43l**	0.49*	497	525
Peptide **44a**	0.44*	492	519
4-Aminofluorescein	0.02*	488	518
Fluorescein	0.95**	490	516

Table 11: Fluorescence quantum yields of *N*- and *C*-terminal fluorescein labeled peptides. *calculated; ** standard (see Part VI.1.11)

IV.5.5.   Fluorescence polarization assays

IV.5.5.1. Set-up of fluorescence polarization assays

Before running a successful assay, different parameters had to be adjusted and optimized, such as the type of microtiter plates (MTP), the assay volume, the buffer system, and the concentrations of the different assay components.

a. MTP selection

First measurements were performed on standard MTP that were available in the lab (Corning Nr. 3710). After some test measurements with fluorescein labeled peptides, it appeared that these plates were not suitable for FP analyses because of a very strong inhomogeneity probably due to minimal deformations of the material by production. The measurements were repeated with free fluorescein in solution to avoid peptide aggregation effects, but the measured FP values were still inhomogeneous and furthermore too high (Figure 47) – indeed, free fluorescein in solution should have an FP value of about 20 mP.

<>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	183	212	173	172	148	130	111	127	139	115	129	165	155	125	112	131	127	105	128	150	166	278	204	172
B	102	179	141	141	119	73	89	123	161	157	163	177	175	156	151	158	122	92	81	124	137	144	179	117
C	135	186	157	133	85	90	78	84	98	107	122	163	160	121	99	89	81	76	85	88	123	148	174	130
D	137	173	154	135	104	88	73	73	75	96	127	171	163	118	92	73	68	76	92	105	133	144	170	134
E	132	187	159	161	137	111	99	86	81	82	98	167	159	95	81	78	83	97	114	134	155	143	176	133
F	128	185	159	185	160	146	131	108	87	74	75	151	132	80	74	87	102	124	148	156	177	158	173	126
G	125	199	193	206	185	177	162	152	137	105	79	171	172	78	98	129	144	156	176	183	198	186	194	108
H	201	216	218	214	203	198	186	186	177	171	129	104	93	127	170	171	174	178	194	203	206	206	207	177
I	182	219	212	216	207	204	194	189	180	172	157	106	95	160	174	178	187	193	204	199	209	197	205	143
J	132	193	178	198	187	182	170	157	134	94	74	156	152	71	93	123	151	174	182	184	182	170	186	107
K	133	186	158	181	158	153	132	116	89	80	92	173	175	87	70	92	116	129	154	157	172	150	173	140
L	124	170	133	164	122	103	96	84	104	96	144	176	175	127	81	88	85	94	106	120	148	134	166	137
M	151	152	145	135	106	99	74	74	92	133	155	173	176	157	136	94	71	79	92	103	127	134	150	134
N	141	163	135	120	66	82	83	86	104	133	110	182	178	96	118	111	91	73	81	72	119	125	159	115
O	125	189	145	127	119	85	112	136	154	152	127	195	195	113	144	172	132	97	83	117	119	138	188	115
P	126	206	168	144	141	121	136	129	130	124	118	146	166	114	120	128	...	...	...	...	...	...	...	...

Figure 47: FP values for a 5 nM fluorescein solution on standard non-treated MTP. The recorded FP values are too high, probably because of binding of the molecules to the plate surface due to and fully non-homogenate; the cross-pattern appears.

After some investigations, plates with a special non-binding surface (NBS™) (Figure 48a) were chosen to prevent binding of the compounds to the well surface and recording of inaccurate FP values. A MTP model with round-bottom for very low volume was selected, enabling to work with very low volumes: 2–20 µL (Figure 48c). As a standard, a volume of 6 µL was used for the assays with very good reproducibility. The results obtained with these new plates were much more accurate, in both low and high FP ranges (Figure 54).

PS   **Hydrophilic PEO-like linker**   **Non-ionic hydrophilic tail**
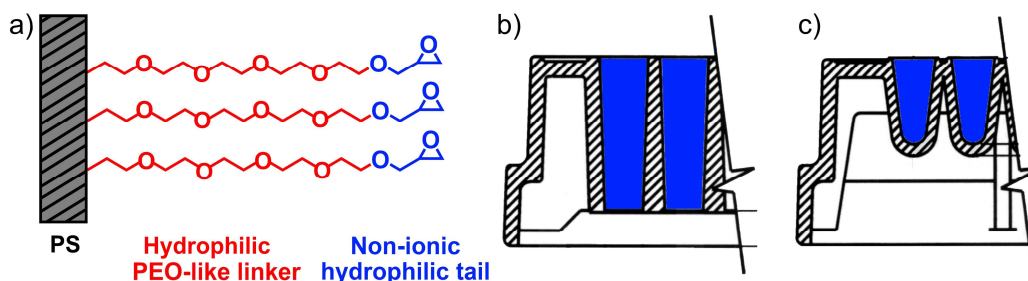
Figure 48: Special MTP for FP assays: NBS™ treatment and very-low volume design. Plates with a non-binding surface treatment (a) avoid protein or peptide binding to the well surface and enhanced FP values. Whereas the standard flat bottom design (b) requires minimal volumes in the 20–100 µL range, the small volume round bottom design (c) can be used with low amounts (2–20 µL).

b. *Detergent selection*

The influence of detergents was also studied[135,141]. Analysis of the FP values for a 5 nM fluorescein solution in H_2O with different detergents is shown on Table 12 .

Detergent	Amount / %	Polarization / mP
–	–	33
CHAPS	0.50	54
CHAPS	1.00	94
Tween 20	0.05	34
Tween 20	0.50	42
Tween 20	1.00	67
Triton-X	0.05	33
Triton-X	0.50	44
Triton-X	1.00	54

Table 12: Influence of detergents on FP values (5 nM fluorescein in H$_2$0)

With diluted Tween 20 or Triton-X, the obtained values were similar to a solution without detergent. Only under high detergent concentrations (1.0%), an increase of the FP values could be observed. For further experiments, even if no aggregation was observed with fluorescein labeled peptides, 0.1% Tween 20 was chosen as a detergent to prevent aggregation.

c. Assay set-up

Different assay volumes were tested in the plate: 6, 10 and 20 µL per well. Results were similar in all three cases, so that the lower volume of 6 µL was chosen. It enabled very low protein consumption and still comfortable pipetting.

The fluorescent probe was tested at different concentrations: 5 nm, 10 nm and 20 nm. The second value (10 nm) was selected for assays, enabling a total intensity signal (RFU) of about 30 000–70 000 with a gain of 80–90, manually adjusted for each measurement.

IV.5.5.2. FP measurements: influence of the label position on the protein-peptide interaction

For fluorescence polarization measurements peptides were dissolved in DMSO at 10 mM and diluted with PBS buffer (pH 7.3) to reach a concentration of 10 nm in the final assay mixture [142]. Samples with increasing amounts of protein over a concentration range of 0.01 mM to 350 mM were prepared. After addition of both protein and peptide, the plate was shaken briefly and polarization of the emitted light was recorded.

Entry	Compound	Peptide sequence	Purity[c] / %	K_D / µM
1	43a [a]	Ac-E F G P P P G W L G R Fluo	96	111 ± 43
2	43b [a]	Ac-E F G P P P G W K G Fluo	95	49 ± 15
3	43c [a]	Ac-E F G P P P G W K Fluo	95	312 ± 106
4	43d [a]	Ac-E F G P P P G R K P Fluo	98	385 ± 102
5	43e [a]	Ac-E F G P P P G R K Fluo	95	460 ± 320
6	43f [a]	Ac-E F G P P P G F Fluo	96	135 ± 31
7	43g [b]	Ac-F N G S P P G L G G G Fluo	95	92 ± 32
8	43h [b]	Ac-F N G S P P G L E R G Fluo	99	168 ± 37
9	43i [b]	Ac-F N G S P P G L G G Fluo	95	67 ± 21
10	43j [b]	Ac-F N G S P P G L G Fluo	93	149 ± 26
11	43k [b]	Ac-F N G S P P G L Fluo	92	188 ± 42
12	43l [b]	Ac-W R P G P P P P P P P G L V Fluo	97	441 ± 339
13	43m [b]	Ac-W R P G P P P P P P P G L Fluo	96	80 ± 20

Table 13: Affinities of *C*-terminally labeled peptides for two GYF proteins, determined by FP. (a) for the CD2BP2-GYF domain; (b) for the PERQ2-GYF domain; (c) HPLC purity (220 nm) after HPLC purification.

Entry	Compound	Peptide sequence	Purity[c] / %	K_D / µM
1	44a [a]	Fluo E F G P P P G W L G R	99	12 ± 1.9
2	44b [a]	Fluo E F G P P P G W L G	93	27 ± 7.5
3	44c [a]	Fluo E F G P P P G W L	98	82 ± 19
4	44d [a]	Fluo E F G P P P G W	99	183 ± 96
5	44e [a]	Fluo E G P P P G W L G R	99	108 ± 24
6	44f [a]	Fluo G P P P G W L G R	92	148 ± 33
7	44g [a]	Fluo P P P G W L G R	92	132 ± 31
8	44h [a]	Fluo P P G W L G R	95	295 ± 33
9	44i [a]	Fluo E F G P P P G W K G	99	16 ± 2.7
10	44j [a]	Fluo Bpa E F G P P P G W L G R	99	2.3 ± 0.4
11	44k [a]	Fluo E F G P P P G W L G R Bpa	99	10 ± 7.0
12	44l [b]	Fluo F N G S P P G L G G	95	285 ± 185
13	44m [b]	Fluo P P G L S R D	96	177 ± 24
14	44n [b]	Fluo S P P G L S R D	98	690 ± 218
15	44o [b]	Fluo N G S P P G L S R D	95	994 ± 855
16	44p [b]	Fluo W R P G P P P P P P P G L V	95	75 ± 16
17	44q [b]	Fluo R P G P P P P P P P G L V	99	202 ± 38
18	44r [b]	Fluo P G P P P P P P P G L V	98	338 ± 143
19	44s [b]	Fluo W R P G P P P P P P P G L	95	46 ± 21

Table 14: Affinities of *N*-terminally labeled peptides for two GYF proteins, determined by FP. (a) for the CD2BP2-GYF domain; (b) for the PERQ2-GYF domain; (c) HPLC purity (220 nm) after HPLC purification

FP values were plotted as a function of the logarithm of the protein concentration. Binding curves were fitted with the software Prism 5.0 (Graph Pad Software, Inc., San Diego, CA) using a sigmoidal dose–response model (see VI.3.3) and K_D values [143-145] in the micromolar range were obtained (Table 13, Table 14).

For the CD2BP2-GYF domain, *C*-terminally labeled peptides with binding affinities around or below 100 µM were identified (**43a**, **43b**, **43f**), **43b** as having the highest affinity. To investigate the effect of the label position, the *N*-terminally labeled derivatives were also prepared and showed consistently higher affinity (namely **44a** and **44i** with 12 µM and 16 µM). The *C*-terminal (**44b**–**44d**) and *N*-terminal (**44e**–**44h**) shortening of the sequence led to drastic decrease in binding (Table 14), and so **44a** (Fluo- EFGPPPGWLGR-NH$_2$) was identified being the best fluorescence probe (12 µM) for the CD2BP2-GYF domain.

For the PERQ2-GYF domain, the *C*-terminally labeled peptide **43i** (67 mM) was identified as the highest-affinity probe (Table 13). Again, to test the effect of the labeling position, the *N*-terminally labeled analog **44l** was prepared and had again a significantly reduced affinity (285 µM, Table 14).

For another test, a sequence containing seven consecutive prolines was investigated, and in this case, the *N*-terminally labeled peptide **44s** (46 µM) displayed a higher affinity than the *C*-terminally labeled analog **43m**.

It can be concluded that both *C*- and *N*-terminal labeling modes can be useful for FP measurements. The effect of the label position cannot be generalized, and seems to depend on the peptide and/or protein structure.

IV.5.5.3. *Evaluation of the binding by NMR experiments*

The binding of selected *N*-terminally fluorescein labeled peptides to GYF domain was confirmed by NMR spectroscopy (D. Kosslick, group of C. Freund, FMP, Berlin). Increasing amounts of a fluorophore-modified peptide were added to [15]*N*-labeled GYF domain. HSQC-NMR spectra were recorded and superposed to identify *NH* resonances of amino acid residues significantly shifted upon ligand addition (Figure 49a). Shifted amino acids were consistently located at the reported ligand binding site of this GYF domain [146-148]. For five fluorescein-labeled peptides the reversible binding affinities were determined by NMR titration. Apparent K_D values were calculated by
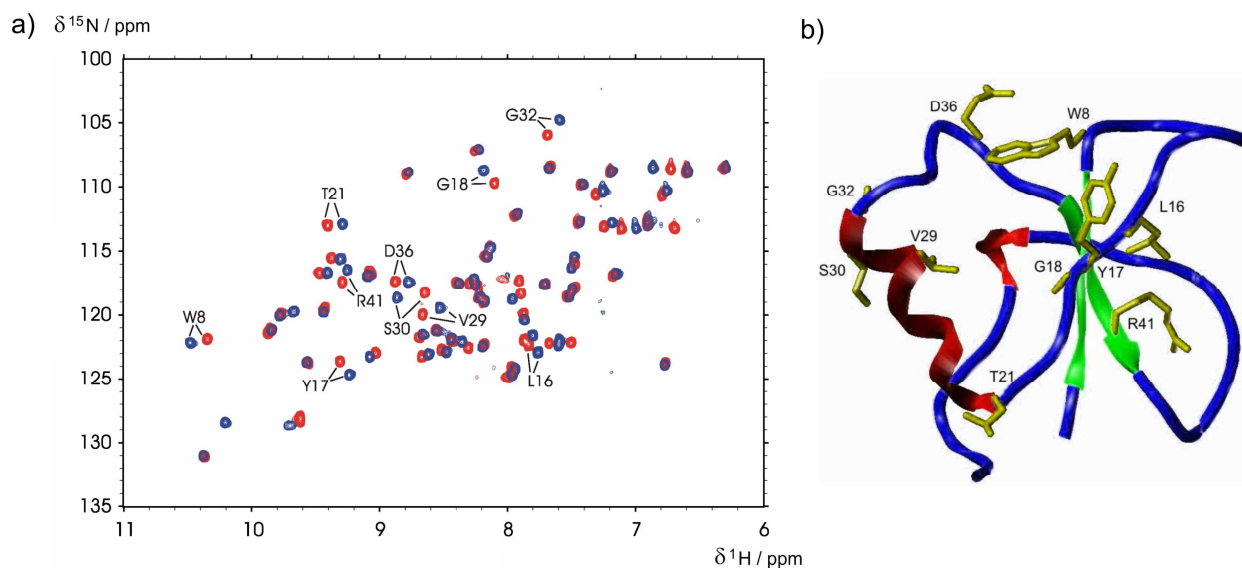
Figure 49: Evaluation of the binding to the CD2BP2-GYF domain by NMR titration experiments (measurements done by D. Kosslick, group of C. Freund, FMP, Berlin). a) Overlays of the ^{15}N-HSQC spectra for the CD2BP2-GYF domain without ligand (red) and upon addition of Fluo-EFGPPPGWLGR-NH2 **44a** (2 mM, blue). Ten resonances showing the largest chemical shifts between the bound and unbound state used for K_D determination are labeled according to residue type and number. b) Ribbon structure of the CD2BP2-GYF domain (PDB ID: 1GYF), highlighting the secondary structure of the protein. Side chains of residues that were used as probes in the NMR titration experiments are shown in yellow and are labeled by amino acid type and sequence number.

Entry	Compound	Peptide sequence	Purity [c] / %	K_D (µM)
1	44a [a]	Fluo E F G P P P G W L G R	99	72 ± 8
2	44h [a]	Fluo P P G W L G R	95	728 ± 118
3	44m [b]	Fluo P P G L S R D	96	311 ± 52
4	44n [b]	Fluo S P P G L S R D	98	487 ± 115
5	44p [b]	Fluo W R P G P P P P P P P G L V	95	119 ± 11

Table 15: Affinities of *N*-terminally labeled peptides for two GYF proteins, determined by NMR spectroscopy. (a) for the CD2BP2-GYF domain; (b) for the PERQ2-GYF domain; (c) HPLC purity (220 nm) after HPLC purification.

recording the chemical shifts of ten characteristic *NH* signals with increasing peptide concentration. All ten considered *NH* signals were located in proximity to the protein binding site as indicated in the three-dimensional protein structure (Figure 49b). All peptides displayed slightly reduced binding affinities in the NMR titration experiments relative to fluorescence polarization. It should be noted that the NMR spectroscopy experiments record the binding to single *NH* positions, which can be reduced by the

flexibility of the peptide ligand in the binding pocket. Moreover, the significantly higher DMSO concentrations in the NMR spectroscopy experiments can be the reason for overestimated K_D values, since it was found that DMSO can bind to the hydrophobic binding sites of GYF domains (M. Kofler, group of C.Freund, FMP Berlin) Relative affinities of different peptides, however, were identical for both methods (Table 15).

IV.5.5.4. Peptide selection for competitive FP screening

As mentioned above, the best peptide of the previously synthesized library was Fluo-EFGPPPGWLGR **44a**, with a K_D of 12 µM. An option to increase the FP maximum was the use of a photo cross-linking agent. UV irradiation of a solution containing a protein and a ligand labeled with a photo cross-linking agent should induce covalent binding of the ligand to the protein by formation of radicals in the binding pocket, leading to an increase of the recorded FP values (Figure 50b). Different photo cross-linking agents can be used like aryl azides [86] or benzophenone [149]. This latter was first investigated.

To validate the concept, two peptides were synthesized, containing Bpa either at the *C*-terminus next to the arginine or at the *N*-terminus between the fluorescein and the glutamic acid. Benzophenone was introduced on the peptide by coupling Fmoc protected 4-benzoyl-phenylalanine (Fmoc-Bpa-OH, Figure 50a) with the standard methods of SPPS.
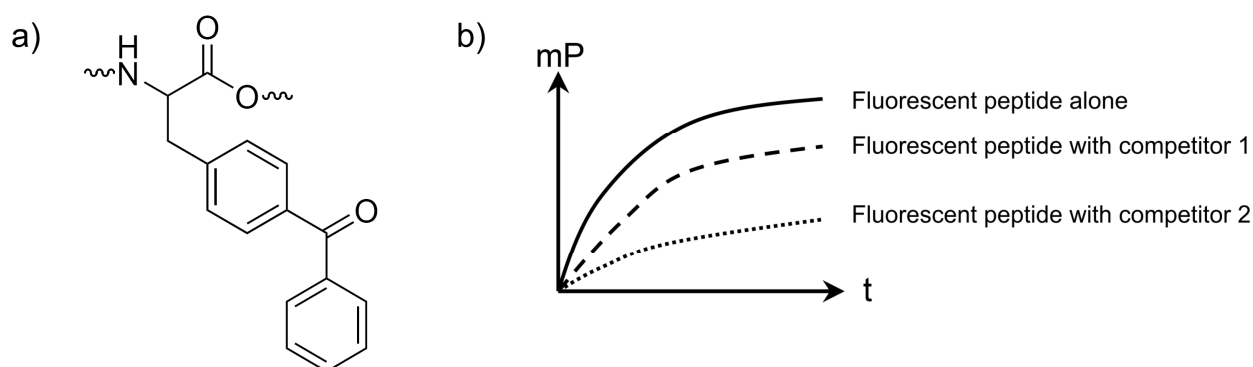
Figure 50: Influence of cross-linking on FP assays. The Bpa group (a) should induce cross-linking between the fluorescent peptide and the protein, increasing the FP signal (b).

FP binding curves were first recorded without any irradiation. They are shown for the two peptides containing Bpa on Figure 51 (**44j** in blue and **44k** violet) and compared to

the non-Bpa labeled peptide (in black) and the equivalent with *N*-terminal fluorescein (in green). An important improvement of the binding could already be observed, especially for one of the new peptides: Fluo-(Bpa)EFGPPPGWLGR **44j**, reaching a K_D value of 2.3 µM.
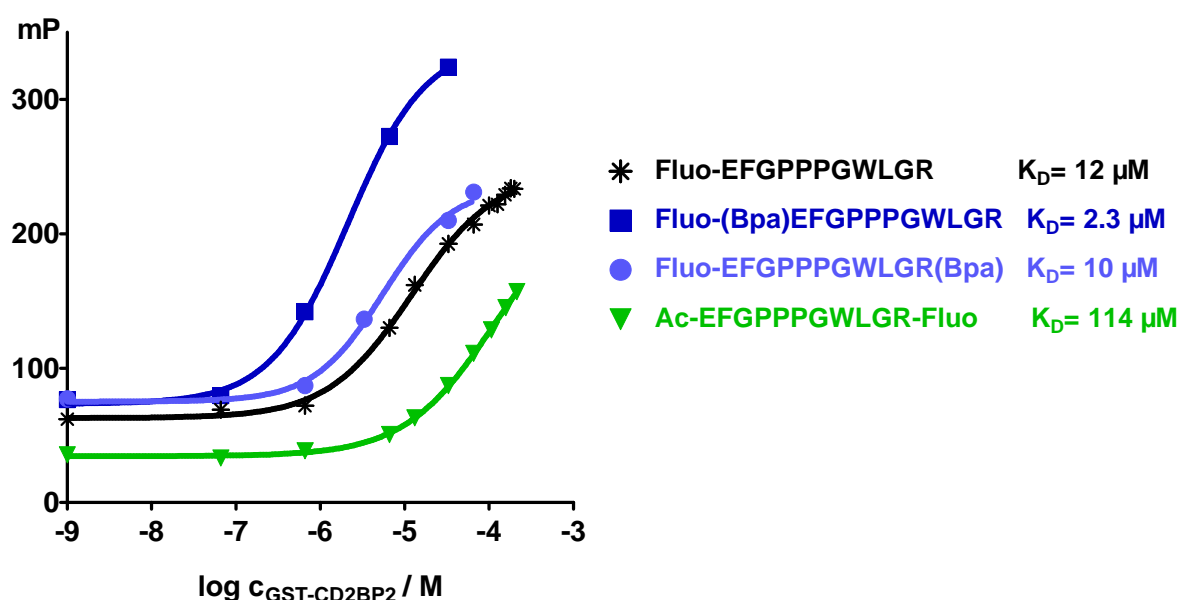
Figure 51: Determination of the K_D of some fluorescence labeled peptides by binding to the GST-CD2BP2 GYF domain. A binding probe with a very high affinity was found: Fluo-(Bpa)EFGPPPGWLGR (**44j**), with a K_D value of 2.3 µM.

The assay plate was then irradiated with UV for 30 min, following by direct FP measurement. During the excitation process, the plate was cooled on ice to avoid heating and protein deterioration. After that, the plate was again irradiated for further 60 min and FP recorded.

As shown on Figure 52, the irradiation had no effect on the polarization values. Thus, the presence of a covalent binding was tested: after mixing the protein with the peptide and UV irradiation, a non-labeled binding peptide was introduced as a competitor. A displacement of the ligand was observed (decrease of the FP value). This confirmed that no cross-linking had occurred. This can be explained by an inappropriate orientation of the benzophenone moiety in the binding pocket. Indeed, the geometry has a great influence on the success of Bpa covalent binding [149]. For a successful cross-linking, an azide-based agent could be tested.
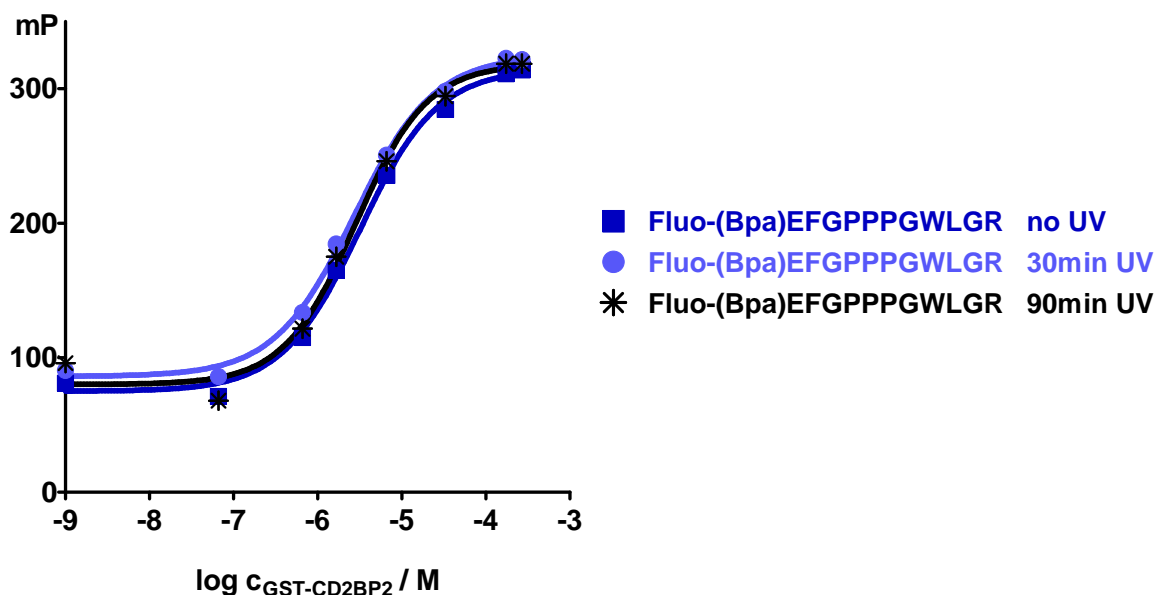
Figure 52: Determination of the binding of Fluo-(Bpa)EFGPPPGWLGR 44j to the GST-CD2BP2 GYF domain before and after UV irradiation. No cross-linking effect can be observed, even after long irradiation times.

However, the synthesized Fluo-(Bpa)EFGPPPGWLGR (**44j**) peptide, with a ten times stronger binding (2.3 µM) than the non-Bpa-labeled equivalent, was sufficient for screening and no further K_D improvement was required.

IV.5.5.5. Analysis of the Fluo-(Bpa)EFGPPPGWLGR peptide: reversibility and Z' factor determination

To be suitable for screening, the labeled peptide cannot bind covalently to the protein. Indeed, in a competitive screening, small molecules are added to the protein-peptide complex and should be able – if they have enough affinity – to displace the peptide from the binding pocket.

As a control [150], a displacement assay was performed using as competitive ligand the similar non-labeled peptide Ac-EFGPPPGWLGR (**45**) (Figure 53). The protein concentration was set-up to 3 µM – in the range of the K_D value – to become significant changes in FP upon ligand addition.

The Z' factor [133,151,152] was furthermore calculated for the fluorescein labeled peptide at the above mentioned protein concentration. The maximum value was obtained with a mixture of protein with the fluorescein labeled peptide; the minimum, after full

peptide displacement upon further addition of an excess of the non labeled Ac-EFGPPPGWLGR (Figure 54).
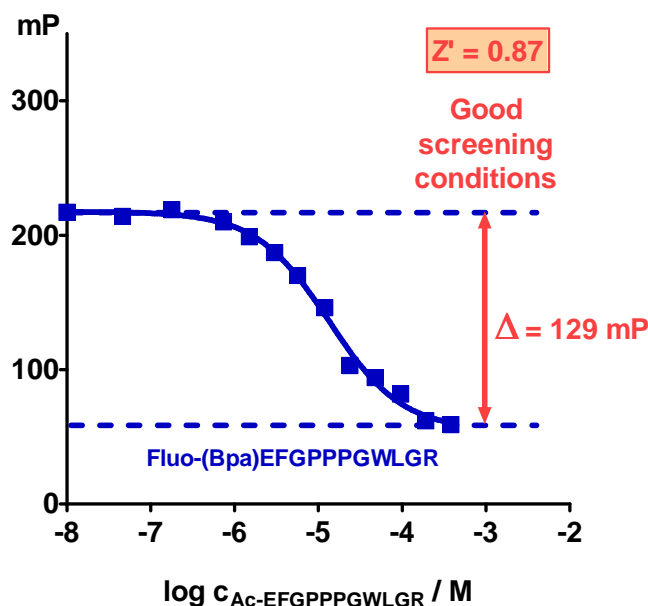
Figure 53: Displacement assay: Ac-EFGPPPGWLGR (45) against Fluo-(Bpa)EFGPPPGWLGR (44j), at a CD2BP2 concentration of 3 µM. The assay shows the non-covalent character of the protein-peptide binding. Furthermore, the high maximum to minimum ratio (129 mP) contributes to a good Z' factor.

a)

<>	1	2	3	4	5	6	7	8	9	10	11	12
A	201	202	202	203	201	202	199	204	203	203	200	202
B	202	207	205	206	205	201	204	204	203	204	199	202
C	203	205	201	202	204	202	203	202	203	201	203	203
D	202	204	201	203	200	200	202	204	200	204	197	204
E	205	204	205	204	204	201	202	203	197	201	199	202
F	201	204	204	203	203	200	200	202	202	202	199	201
G	203	200	201	205	200	202	198	204	199	199	199	199
H	199	204	202	202	202	199	200	200	200	201	204	201

b)

<>	13	14	15	16	17	18	19	20	21	22	23	24
C	70	69	71	76	76	73	72	77	82	71	73	68
D	71	71	72	75	72	73	70	71	75	78	69	69
E	69	69	71	69	73	72	72	73	77	70	71	73
F	68	70	70	70	70	66	71	79	75	77	74	75
G	69	72	73	68	71	71	74	67	71	72	74	75
H	71	75	68	71	72	70	73	74	77	72	73	73
I	73	72	75	72	71	71	76	79	77	77	76	78

Figure 54: FP maxima (a) and minima (b) for Z' factor determination. On the very-low volume MTP with non-binding surface treatment, FP values are very homogeneous. This contributes to a high Z' factor.

With standard deviations of 2.0 for the maximum and 3.1 for the minimum as well as a signal to baseline ratio of 129 mP, the resulting calculated Z' factor was 0.87 (see

VI.3.6). As a Z' value higher than 0.5 is considered to be suitable for an assay, the synthesized peptide was therefore a very good probe for FP screening of the ChemBioNet library for CD2BP2 ligands.

IV.6. A fluorescence polarization assay for the screening of chemical libraries for CD2BP2 GYF domain ligands

IV.6.1. Screening preparation and realization

The ChemBioNet library was screened with an FP displacement assay as described in Part IV.5.5.5, using Corning Nr. 3676 small volume round bottom black non binding surface 384-well microtiter plates. The 16 896 compounds were dissolved at a concentration of 10 mM in DMSO on 48 MTP from column 1 to 22. The two last columns (23 and 24) remained free of compound, only filled-up with DMSO for negative and positive controls [153] (Figure 55).

Figure 55: Layout of an assay plate. Compounds are present in wells A1 to P22; positions A23 to P23 and A24 to P24 are reserved for negative and positive controls.

The assay plates were first pre-filled with buffer with a dispenser. In a second step, compounds of the stock solution plate were transferred to the assay plate by the robot, and the protein-peptide solution was added. Finally, a solution containing a known competitor was manually added in the negative control wells, as described on Figure

56. The plates were then centrifuged to eliminate air bubbles due to pipetting, and FP values were directly recorded.

Per well:
- 4.9 µL 1x PBS buffer pH 7.3 (dispenser)
- 0.1 µL compound from the 10mM stock solution in DMSO (robot)
- 5.0 µL solution (A) (robot)

Positive control:
- 4.9 µL 1x PBS buffer pH 7.3 (dispenser)
- 0.1 µL DMSO (robot)
- 5.0 µL solution (A) (robot)

Negative control:
- 4.9 µL 1x PBS buffer pH 7.3 (dispenser)
- 0.1 µL DMSO (robot)
- 5.0 µL solution (A) (robot)
- 0.5 µL Ac-EFGPPPGWLGR-NH$_2$, 5mM (manually)

Solution (A):
GST-CD2BP2, 6µM; Fluo-(Bpa)EFGPPPGWLGR-NH$_2$, 20nM; Tween 0.2%, in 1x PBS buffer pH 7.3

Preparation of 100mL solution (A): 40 mL GST-CD2BP2, 15µM + 40 mL Fluo-(Bpa)EFGPPPGWLGR-NH$_2$, 50nM + 20mL Tween 1.0%, in 1x PBS buffer pH 7.3

Figure 56: Screening protocol

IV.6.2.   Data analysis and hits selection

IV.6.2.1. Data formatting

For an efficient results analysis, a ChemFinder database was created. Each Excel file generated by the reader (one table of results per plate) was converted into an sd-file. All resulting sd-files were incorporated into one ChemFinder database (Figure 57).

This format enabled a rapid access to hits: polarization and intensity criteria could be very easily defined and modified. The time invested in data reformatting was recovered at the later stage by the high flexibility and adaptability of results analysis.

IV.6.2.2. Data analysis

Compounds were first sorted through their polarization value: all compounds inducing more than 75% decrease in fluorescence polarization were retained (Figure 58). Then, compounds with autofluorescence were eliminated[153,154]. Indeed, as shown by the polarization equation (Equation 5), if the total intensity is increased by a third intensity

Figure 57: Screening data conversion and analysis. Data can be easily analyzed through the multicriteria search functionality offered by the ChemFinder interface.

term coming from the added small molecule, the resulting polarization value is artificially decreased, resulting in false positives [153].

$$P = \frac{I_{/\!/} - I_{\perp}}{I_{/\!/} + I_{\perp}}$$

$$P = \frac{I_{/\!/} - I_{\perp}}{I_{/\!/} + I_{\perp} + I_{s.m.}}$$
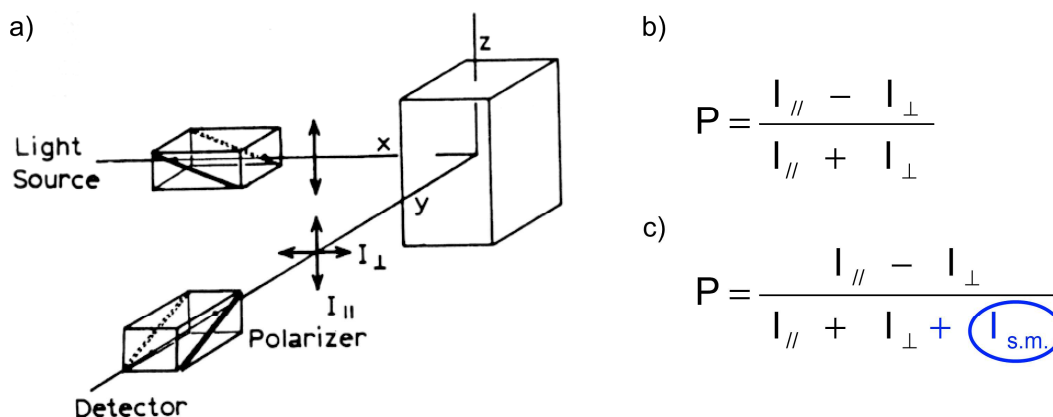
Equation 5: Calculation of fluorescence polarization. Both parallel and perpendicular intensities are recorded (a) and used to calculate the fluorescence polarization value (b). If the added small molecule is also emitting light in the same wavelength range that fluorescein, its intensity has to be added to the denominator (c) resulting in a decrease of the FP value.

The twenty-four remaining compounds were tested in an FP displacement assay, confirming their activity. Molecules with a K_D value under 80 μM were selected and analyzed in LC-MS to confirm their intactness and purity: ten compounds were finally available for further analysis.

→ Polarization value < 100 mP (75 % displacement)   107
→ Elimination of molecules with autofluorescence   24
→ K_i determination with FP displacement curves, K_D < 80 μM   13
→ LC/MS analysis, purity >90%   10
→ complementary measurements   ...
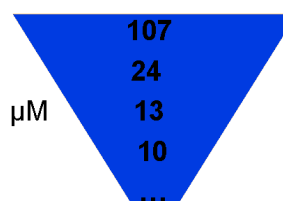
Figure 58: Analysis of screening results and hits validation process

IV.6.2.3. Hits selection

To enable further investigations, compounds had to be reordered. From the ten selected compounds, six could be purchased. The FP displacement assay was done once more and confirmed the precedent results (Figure 59).
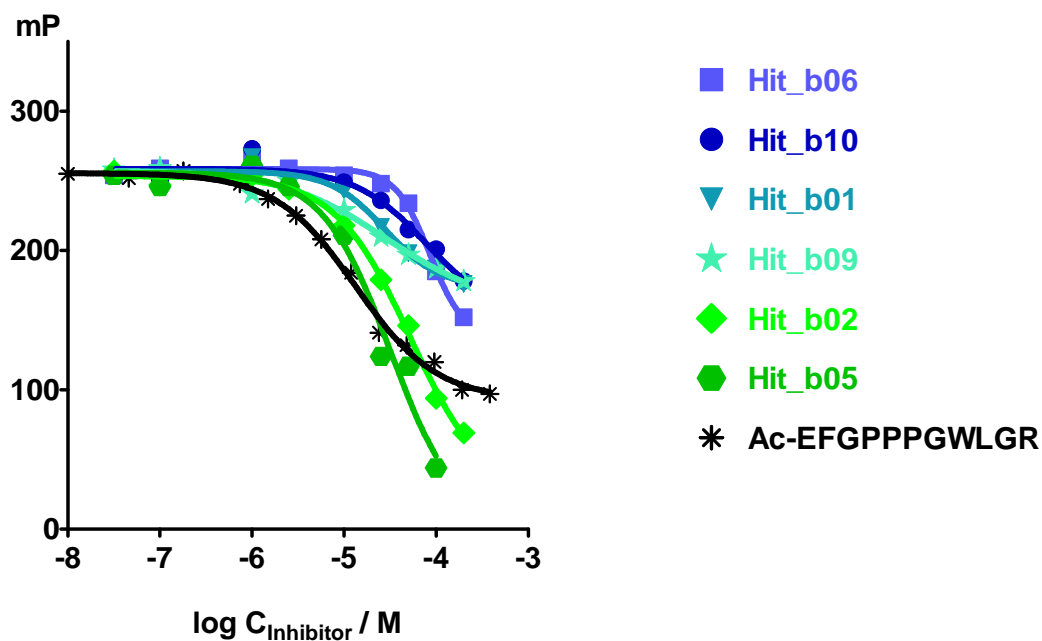
Figure 59: FP displacement curves for the six reordered screening hits
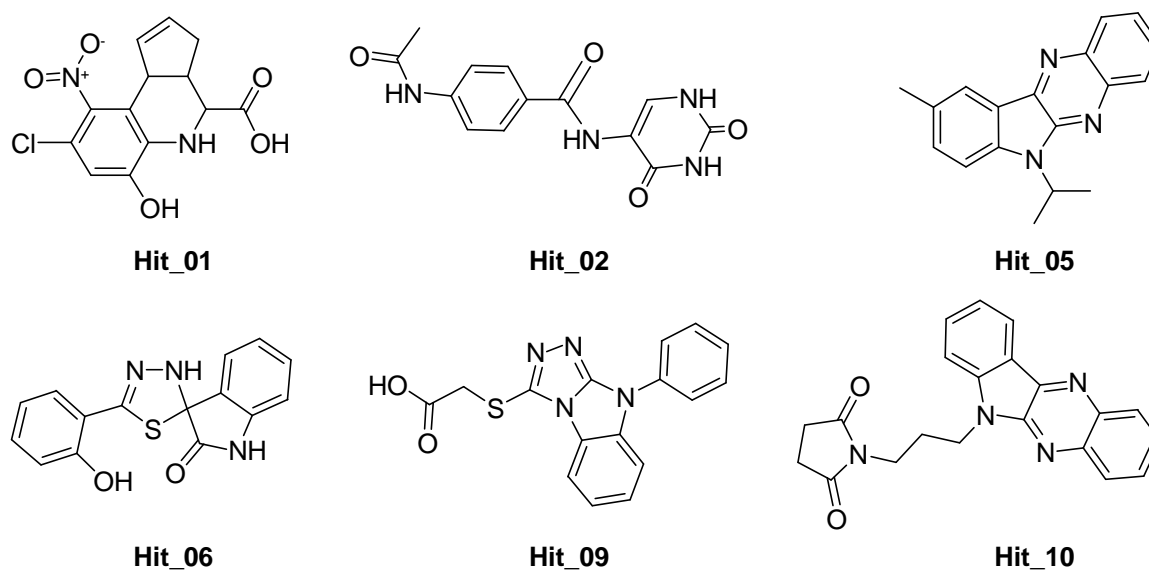
Figure 60: Structure of the six reordered screening hits

IV.6.3. Hits validation

a. *NMR validation*

To confirm the hits activity, binding assays had to be repeated with an independent method. NMR was first selected because of the precision of the delivered information.

It enables not only the determination of the binding constant by titration, but also a precise visualization of the binding pocket and the interacting residues.

Binding assays were done by D. Kosslick (group of C. Freund, FMP, Berlin). First, a direct binding assay was carried out, but no interaction could be proved. As it is known that DMSO has a certain affinity with the CD2BP2-GYF protein and induces small shifts in the NMR spectrum, and as the compounds could not be dissolved in aqueous buffer without addition of DMSO, they were tested again in a second assay to make sure that the interaction has not been hidden by the DMSO effect. The second assay was a competitive experiment, where the small molecules were added to a protein-ligand complex. No spectral changes indicating displacement of the ligand could be observed.

b. ITC validation

As no binding could be shown in NMR experiments, ITC spectra were measured with the help of A. Sieber (group of S. Keller, FMP, Berlin). Similar problems occurred first, namely signal perturbations due to DMSO (Figure 61).
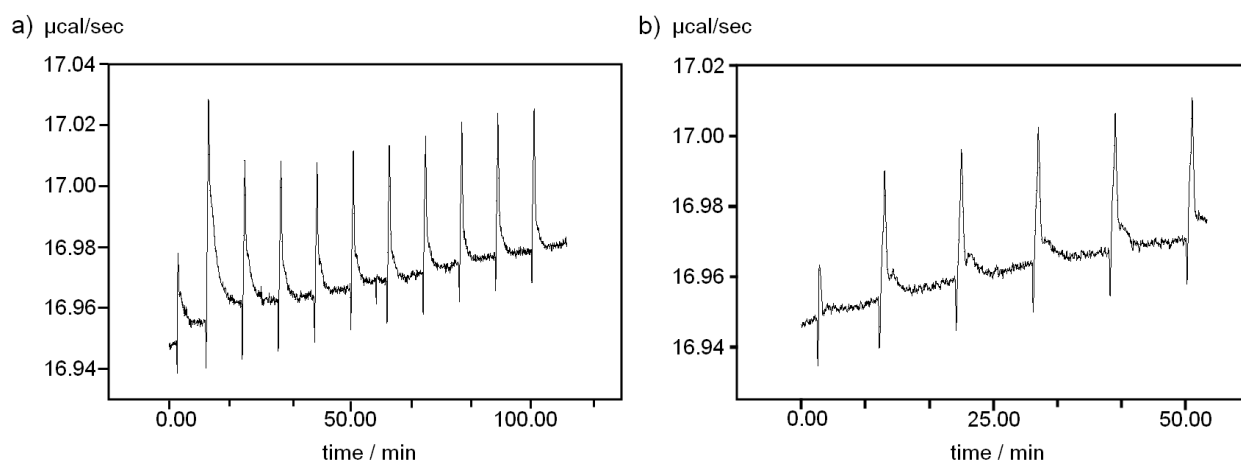
Figure 61: ITC titration of Hit_b02 with the CD2BP2-GYF protein; incidence of DMSO. Aliquots of 1 mM Hit_b02 in PBS buffer with 2% DMSO were added to a 50 µM CD2BP2 solution at 4℃ (a). Addition of a 2% DMSO solution to the protein solution in the same conditions (b) resulted in a very similar signal. So, although the DMSO concentration was very low, it disturbed completely the measurement.

A repetition of the experiment with additional presence of 2% DMSO in the protein solution delivered also a similar graph with low, wide peaks and much noise.

Compound Hit_b09 could be directly dissolved in the PBS buffer without DMSO addition. The observed peaks were more proper – high and fine – (Figure 62) but still no binding could be detected. The experiment was also repeated at higher temperature, leading to equivalent results.
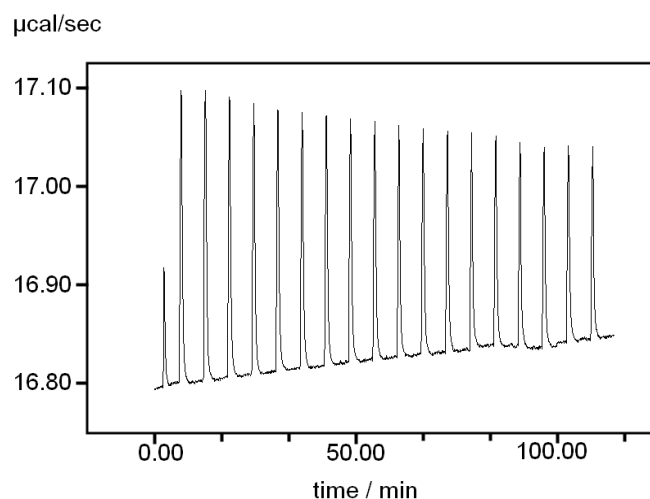
Figure 62: ITC titration of Hit_b09 with the CD2BP2-GYF protein. Aliquots of 1 mM Hit_b09 in PBS buffer were added to a 50 µM CD2BP2 solution at 4°C . No binding could be observed. The light linear decrease in peaks area is due to solvation effects. A similar graph was obtained by addition of the compound to a buffer solution without protein.

c. *CD spectroscopic validation*
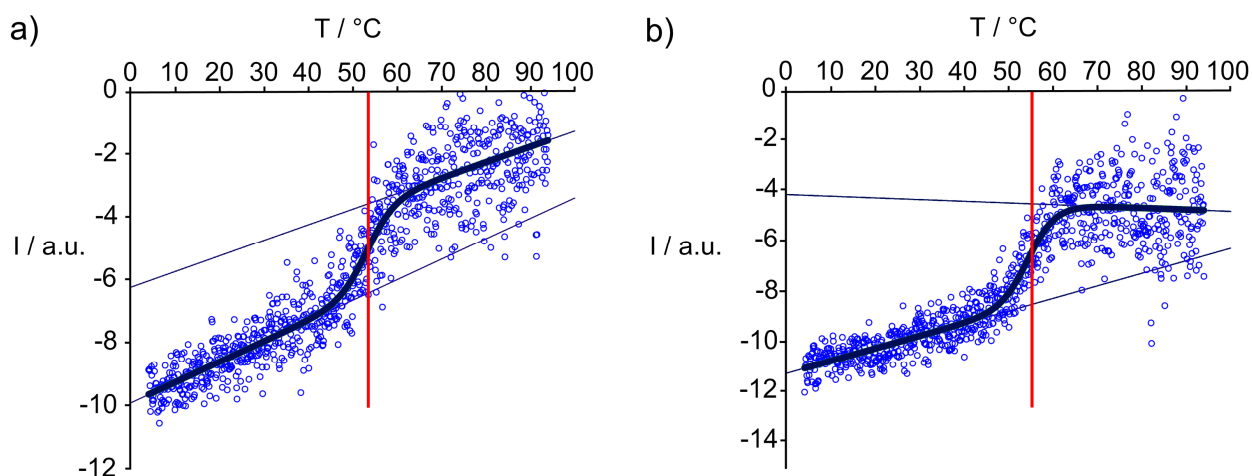
Figure 63: CD spectroscopic evaluation of a ligand binding to the CD2BP2-GYF domain upon melting temperature (T_m) measurement. a) 20 µM CD2BP2-GYF protein; $T_m = 53.5$°C; b) 20 µM CD2BP2-GYF protein with 300 µM ligand $T_m = 55.0$°C. The minor T_m variation is not significant, considering the high dispersion of the recorded values, so that no binding could be confirmed.

In parallel to ITC experiments, circular dichroism [155] (CD) measurements were carried out with the help of M. Georgi (group of S. Keller, FMP, Berlin). The curves obtained for the protein alone and the protein with ligand were quite similar (Figure 63). After fitting, the melting temperatures found for the protein alone (53.5 °C) and in the presence of the ligand (55.0 °C) only separated by 1.5 °C would correspond to a binding constant of about 500 µM.

IV.6.4. Conclusion

a. *A problematic interference of autofluorescence of the tested small molecules with FP measurements*

The limited success of hit validation showed that there was a high proportion of false positives within the hits, probably due to autofluorescence: indeed, a look at the intensity emitted by the molecules during the FP assay brought out that there was a high rate of compounds with a very high autofluorescence (Figure 64). Only for 50% of the molecules, the fluorescence intensity deviation was lower than 20%.
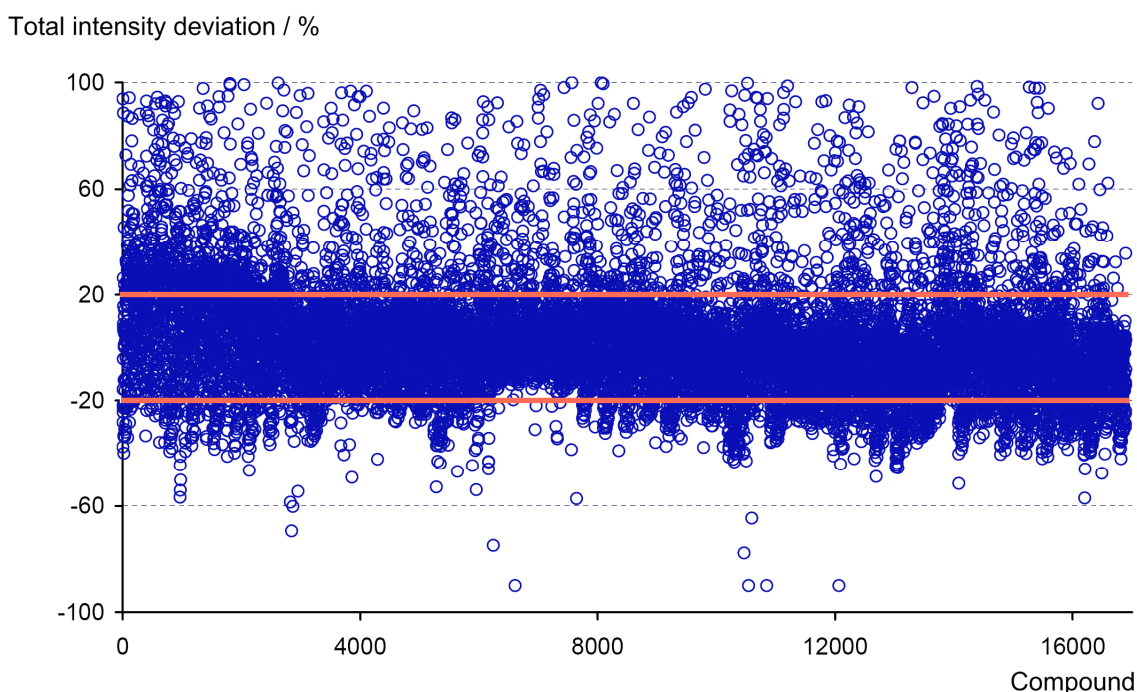
Figure 64: Total intensity of all screening wells. Only for 50% of all compounds, the fluorescence intensity deviation is lower than 20%. Furthermore, 6% of the compounds have a much higher intensity and are out of the graph range.

A further attempt to validate other hits from the FP screening was done with eight molecules having apparently no strong autofluorescence (deviation within the 20% range).

Unfortunately, the previous observations were confirmed: at high concentrations, the fluorescence intensity increased, and parallel to this, the FP value decreased. So, there is in fact a parallel between low FP and high fluorescence intensity (Figure 65) and it is therefore difficult to discriminate between real hits and false positives.
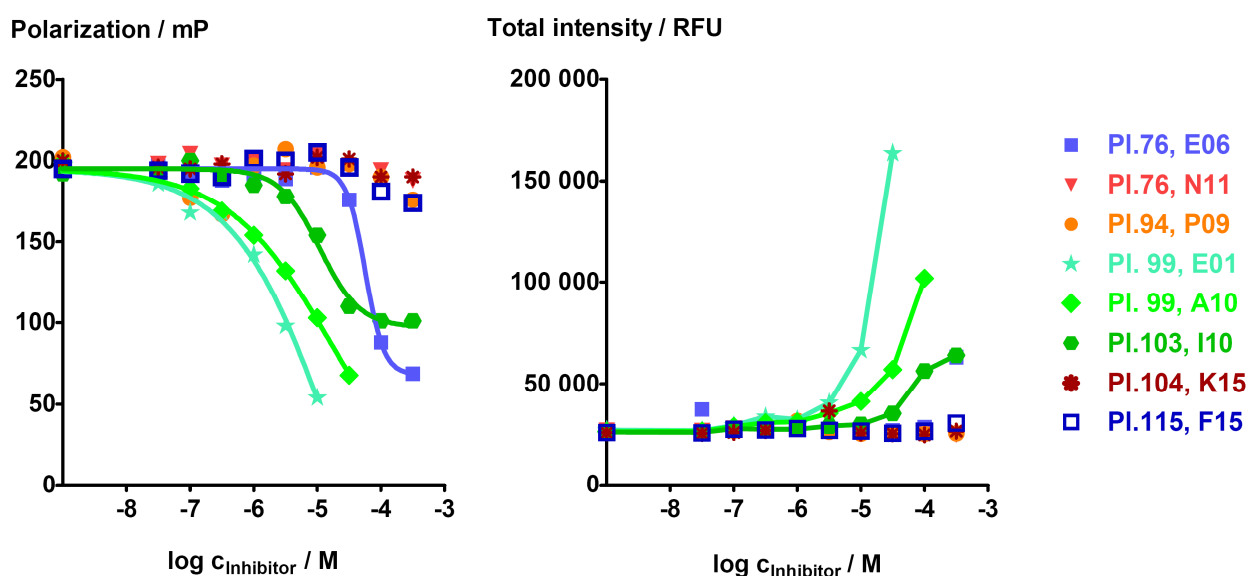
Figure 65: Correlation between the emission of tested small molecules and the decrease in FP values. The FP decrease (left) is for most of the compounds due to autofluorescence of the small molecules (right), inducing false positive hits.

b. *Alternative solutions to improve FP assays.*

As FP assays are nevertheless very powerfull tools, enabling e.g. the use of very small volumes for the determination of interactions and an automated screening of large libraries of compounds, it would be interesting to develop a method to improve them and circumvent the problem of numerous false positives.

A first solution to eliminate false positives would be to combine the established FP screening to size-exclusion chromatography (SEC). The SEC principle had already been adapted to 96-well MTP format by Muckenschnabel et al. (Novartis Pharma, Basel) [156] and used for screening in combination with LC-MS detection. The used plate system is shown on Figure 66. It enables the simultaneous filtration of 96

compounds. Plate B is packed with size-exclusion material; compounds are then filled on plate A; finally the plate system is centrifuged and the products are recovered on plate C.

Figure 66: SpeedScreen "sandwich" assembly. Pinhole plate (A), size-exclusion chromatography plate (B) and collection plate (C), for processing all samples simultaneously in the 96-well format [156].

This technique could be employed to eliminate false positives in FP screenings. Indeed, false positives are mostly due to autofluorescence of the tested small molecules. These compounds have an emission in the same range that fluorescein, and their high excess (100 µM) in comparison to the fluorophore concentration (10 nm) induces a higher total intensity. This modifies artificially the denominator of the equation for FP calculation (Equation 5), which should otherwise remain constant, and the resulting calculated polarization value is consequently decreased, even if no real polarization of the light emitted by fluorescein occurs.

To prevent this phenomenon, the disturbing non-bound compounds that artificially reduce the polarization value should be eliminated. SEC can be a very useful method. As shown on Figure 67, all small molecules that are not bound could be retained in the SEC material, so that the resulting FP value would still be at the maximum, whereas all bound compounds would remain bound to the protein and would be recovered after filtration, so that the FP value would be reduced. The total intensity could either be

reduced, if the compound had no autofluorescence or remain high, without any perturbation of the FP detection.
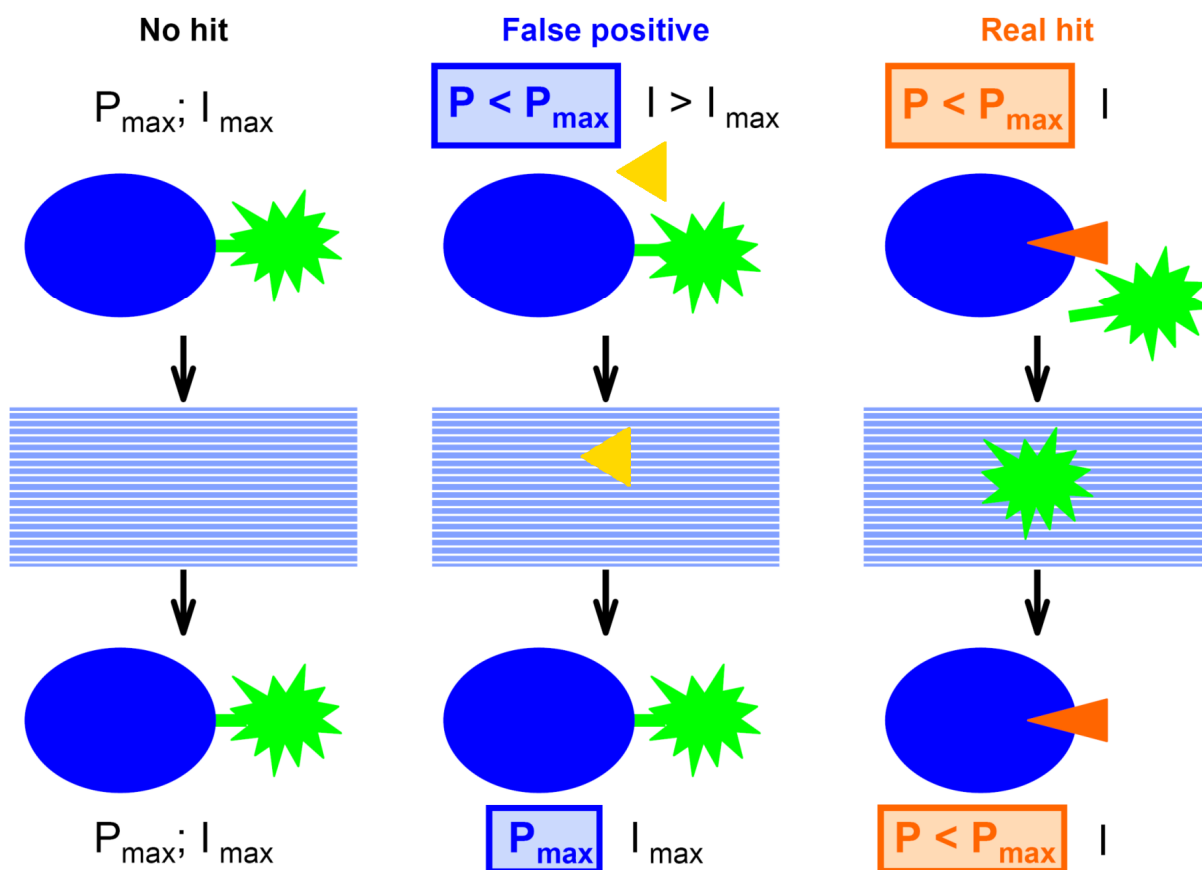
Figure 67: Principle of a combined SEC-FP screening. False positives can be eliminated by filtration through size-exclusion material for the FP detection.

A second solution would be the use of labeled libraries. This assay type would have an inverse read-out: after analogous FP measurements, hits would be detected as molecules having an increased FP value, so that autofluorescence would no more be a disturbing factor.

Preliminary tests were done with the SEC-FP assay. In some cases, results were corresponding to the predictions but they were not always repeatable. The reproducibility problems were due to difficulties to obtain a good compact filling of the plate with the wet filtration material without formation of cracks.

The second way was thus chosen, and a library of fluorescein labeled fragments was synthesized.

IV.7. Application 2: Synthesis of a library of fluorescein labeled small molecules

IV.7.1. Selection of relevant scaffolds

As described in Part IV.3, the synthesized resin enabled the labeling of diverse small molecules either by coupling the aniline form to various acetic acids or by nucleophilic substitution on the bromoacetamide resin.

Appropriate fragments were selected using studies of M. Lisurek (group of R. Kühne, FMP, Berlin) as well as the work of Bernis and Murcko [15,16] to become a diversity set of drug-like building blocks. Analyses of the frameworks and side chains of drugs listed in the CMC (Comprehensive Medicinal Chemistry) catalogue revealed indeed that the whole CMC set (~ 5000 compounds) can be resumed to a very limited number of frameworks. Compounds that were selected for the library should be also commercially available at synthesis scale and not only as HTS samples.

IV.7.2. Synthesis of the fluorescein labeled small molecules library (FSML)

IV.7.2.1. General synthesis strategy
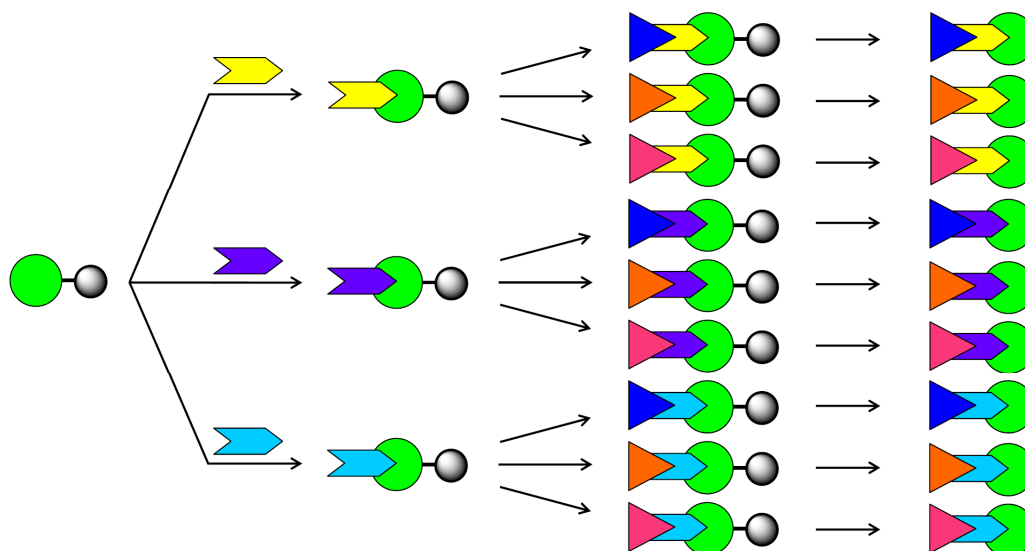
Figure 68: Synthesis strategy for the fluorescein labeled small molecules library. First, a bifunctional residue is coupled to the resin, then a second fragment, finally the resulting product is cleaved from the resin.

IV.7.2.2. Coupling of the first bifunctional fragment on the aminofluorescein resin

The coupling was done as described in Part IV.4.1.2 with the fragments listed on Scheme 32. A complete conversion of aminofluorescein **10** into the products **46ⁱ** was obtained, as seen on control LC-MS.
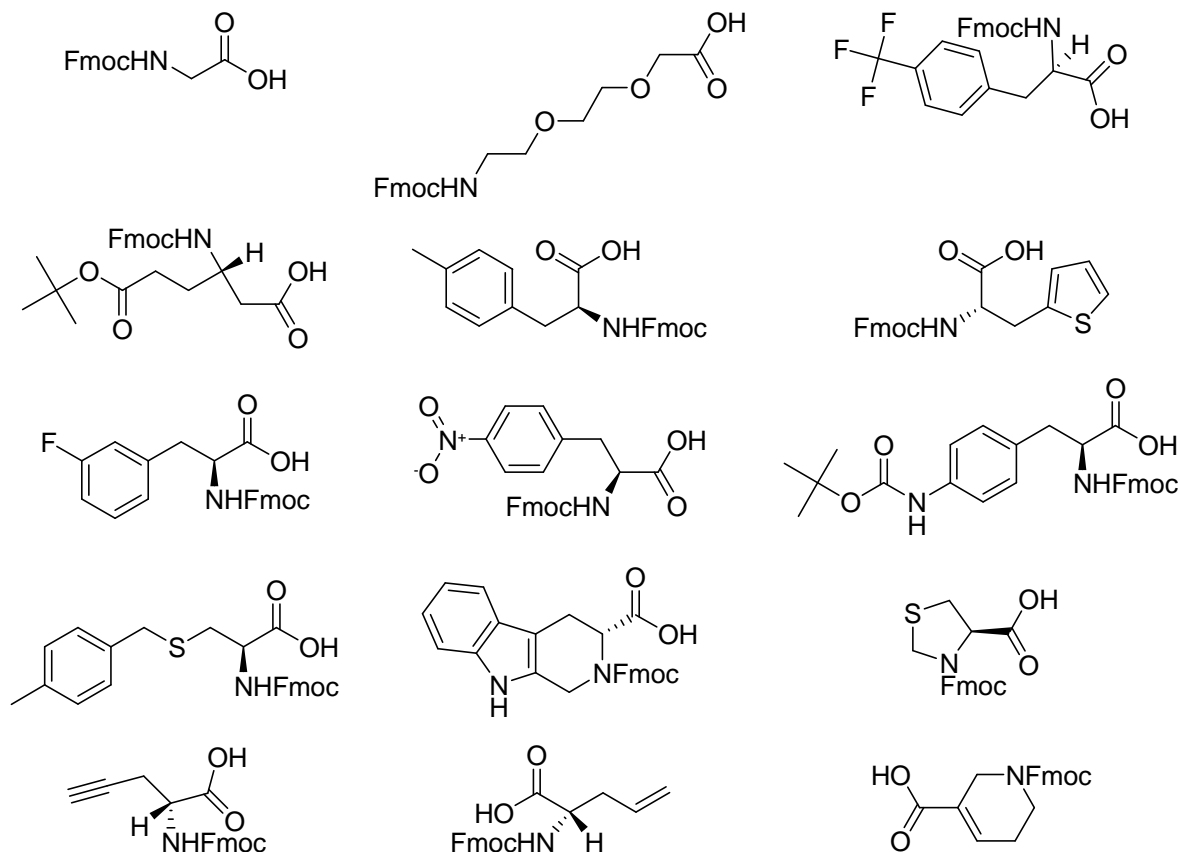
Scheme 32: Bifunctional fragments coupled to aminofluorescein resin

IV.7.2.3. Coupling of the second fragment on the aminofluorescein resin

After Fmoc cleavage of the resins **46ⁱ**, the carboxylic acids listed on Scheme 33 were coupled and resins **47ⁱ** were obtained. Finally, after resin cleavage, the fluorescein labeled small molecules library (FSML), composed of 265 labeled compounds – **FSML-001** to **FSML-265** – was obtained.

This coupling was successful for most of the fragments, delivering products with purities above 90% in quantitative yields. Purities were determined for all compounds by LC-MS analysis; yields were determined by weighing for the first forty compounds, and it was assumed the reaction is also quantitative for all remaining compounds.
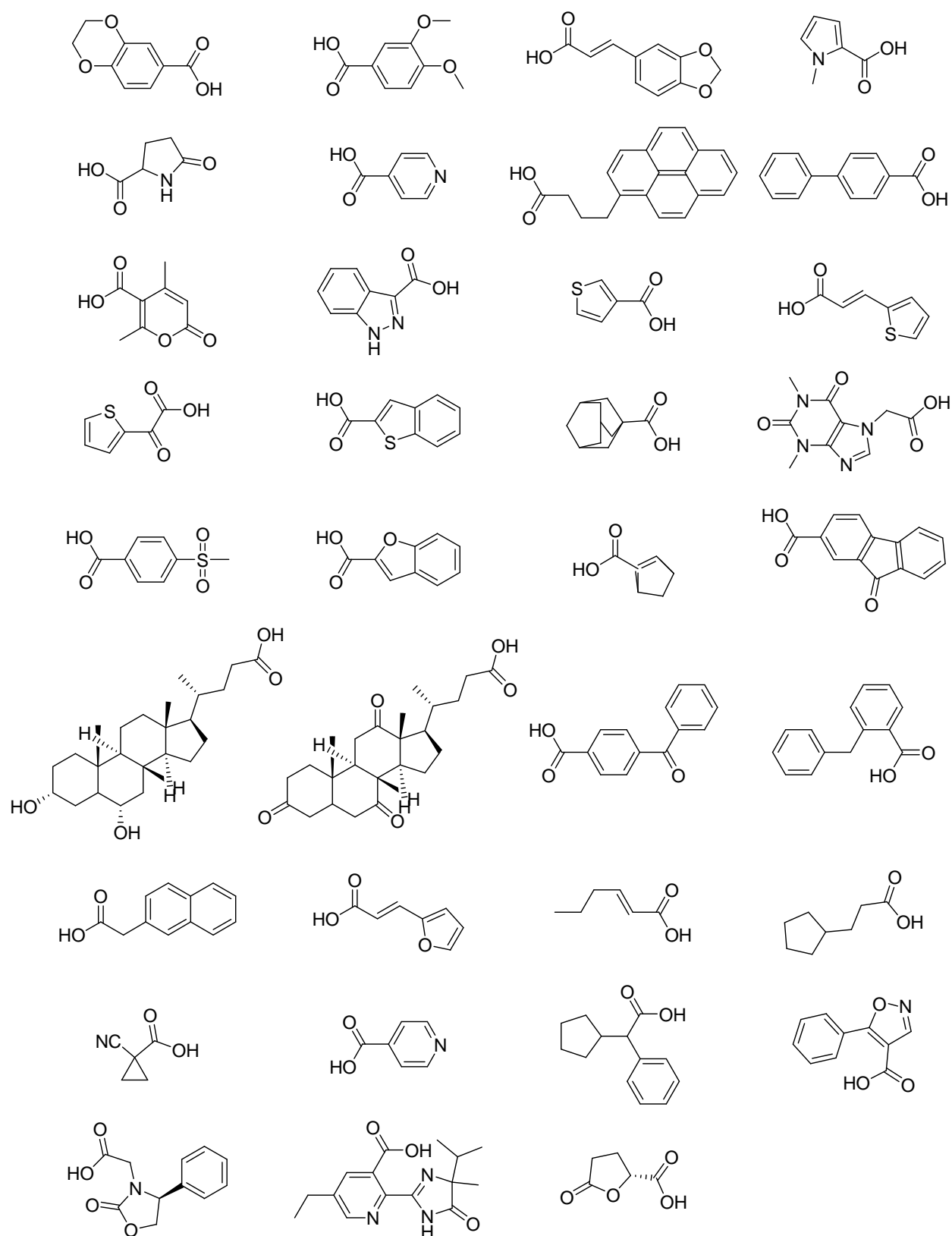
Scheme 33: Carboxylic acids coupled as second fragment to monofunctionalized aminofluorescein resin.

IV.7.3.   Preparation of the FSML stock solutions

A small sample of each compound (~ 0.5–2 mg) was precisely weighed and dissolved in an appropriate volume of DMSO to reach a concentration of 10 mM. All samples were then filled on a MTP (Figure 69), covered with an aluminum sheet and stored at - 18 °C.

All data concerning the library – e.g. structure of the molecule, molecular formula, molecular weight, localization on the plate, building blocks used for the synthesis – were collected in a ChemFinder database. This format offers the possibility to convert the data in other formats usefull for further calculations, as well as to add easily new compounds and fields, in particular the results of subsequent assays.
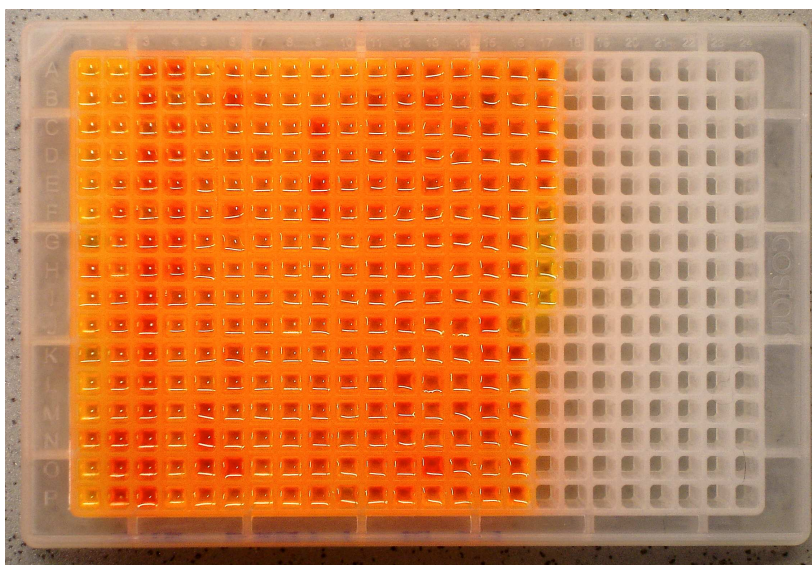
Figure 69: Microtitterplate containing the FSML stock solution, 10 mM in DMSO

IV.7.4.   Diversity analysis of the obtained library (calculations done by M. Lisurek, group of R. Kühne, FMP, Berlin)

Statistics were determined for different parameters to evaluate the diversity of the synthesized library. As shown on Figure 70, a limited number of building blocks used as starting material (a) was sufficient to create a library of relatively diverse labeled molecules (b).

The molecular mass M and the log $P_{o/w}$ values (Figure 71) are also well distributed over a range from about 500 – 800 g·mol^{-1} and 2 – 7, with several compounds above

these limits. The number of hydrogen donors is from 6 to 11, with a value of 6 – 8 for most of the molecules; the number of hydrogen acceptors is about 4 – 5 (Figure 72).

The obtained labeled compounds are not always respecting the Lipinski rule of five, but this is mostly due to the fact that a fluorescein moiety is bound to each molecule (thereby increasing e.g. all molecular masses or the number of hydrogen donors) and the compounds of interest (without the fluorescein moiety) are nevertheless well-suited
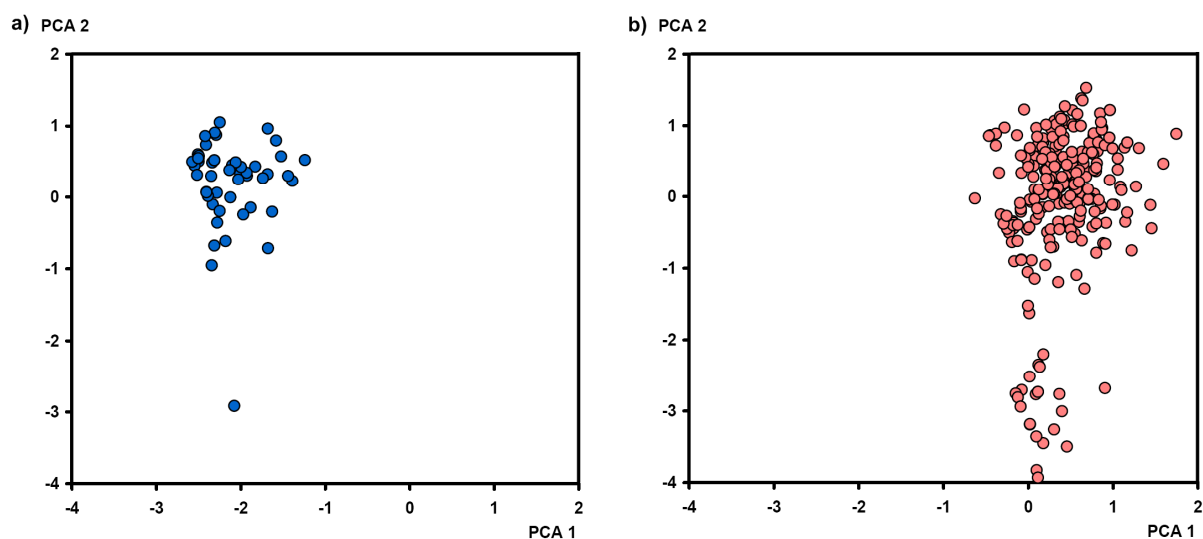
Figure 70: Repartition on the chemical space of the building blocks used for the synthesis of the library (a) and of the resulting compounds FMSL-001−FSML-265.
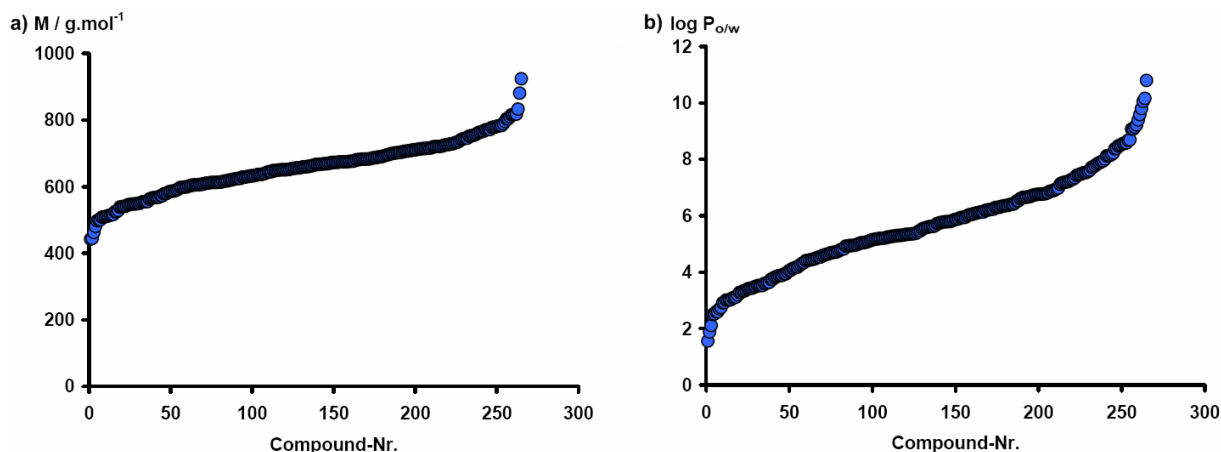
Figure 71: Molecular mass (a) and log $P_{o/w}$ (b) of the compounds FSML-001−FSML-265.
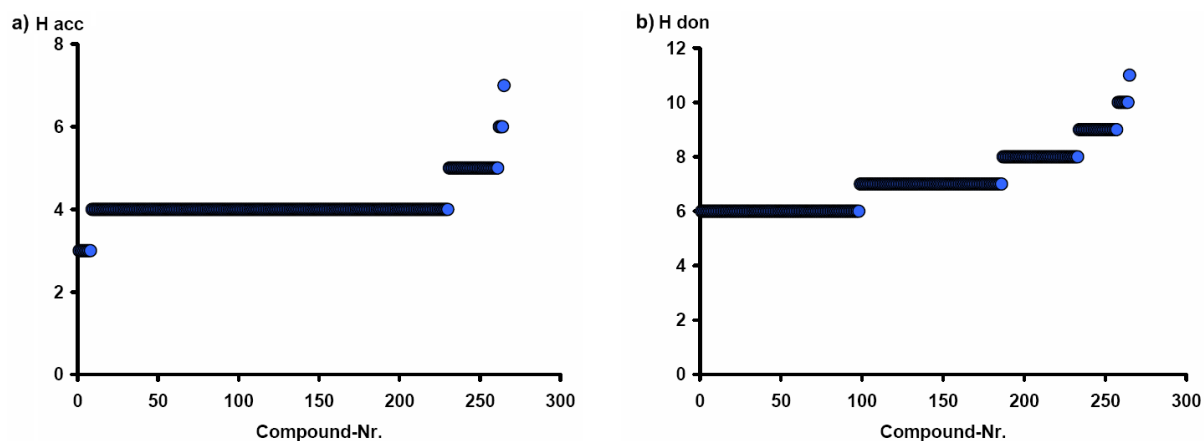
Figure 72: Number of hydrogen acceptors (a) and hydrogen donors (b) for the compounds FSML-001–FSML-265.

IV.8. Fluorescence polarization "direct" screening of the FSML for low-affinity binding to different proteins

IV.8.1. Preparation of diluted assay plates

With the help of Jens von Kries (Screening Unit, FMP, Berlin) for the programmation and control of the automated pipetting roboter, the substances were transferred from the stock solution plate into the assay plates: a black plate for fluorescence emission measurement (Greiner Nr. 781000) and a clear plate for absorption measurement (Corning Nr. 3675), previously filled-up with H_2O + 0.03% Tween 20.
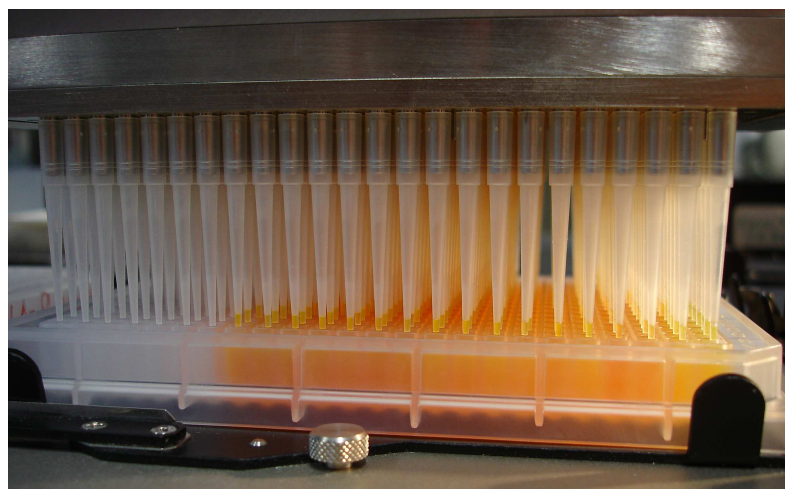
Figure 73: Preparation of assay plates with the pipetting roboter

IV.8.2.   Analysis of the FSML: absorption and fluorescence emission

The absorption of all compounds was recorded. Assuming that all small molecules do not absorb at the fluorescein absorption wavelength (492 nm), the value should remain constant. As shown on Figure 74, they were variations of the absorption, showing an influence of the small molecules on the recorded value.
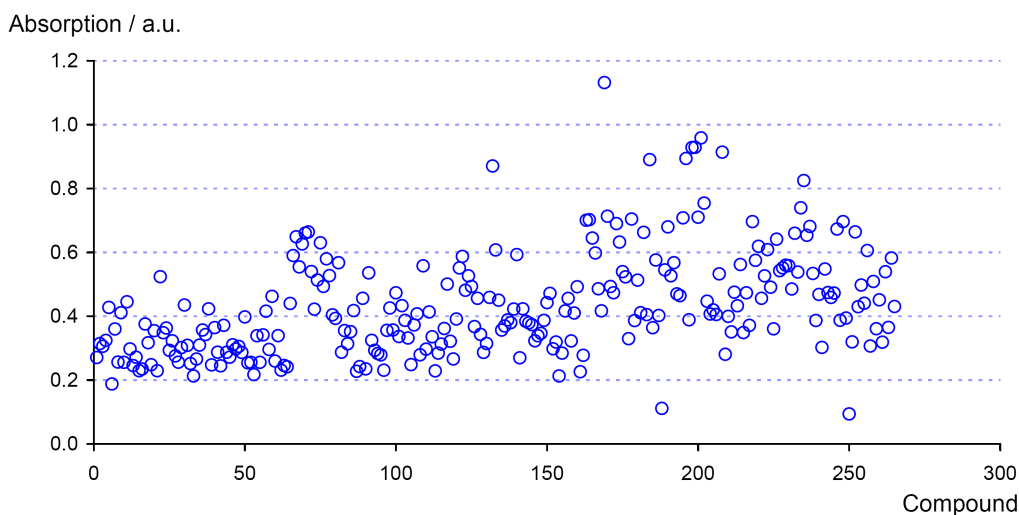
Figure 74: Absorption of the FSML at 492 nm

The fluorescence intensity was also recorded (Figure 75) and large variations could be observed. This is representative of the variety of fragments and confirms that such a library could not be used for a displacement assay with FP detection for low-affinity ligands. For the use in a direct FP assay it is, however, no more problematic.
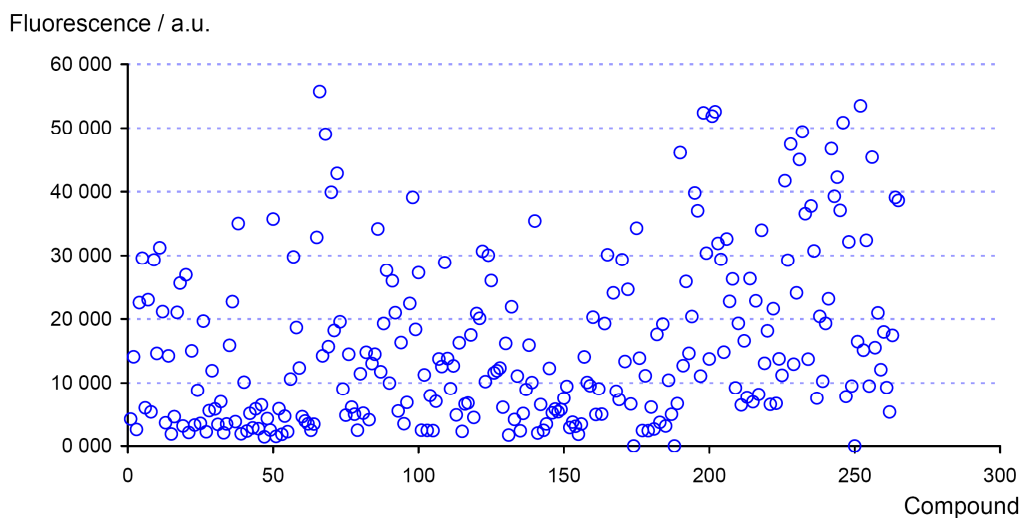
Figure 75: Fluorescence intensity of the FSML at 519 nm

IV.8.3.   Preparation and realization of the screening

The fluorescent small molecules library (FSML) was tested for its binding to a set of 12 proteins by a direct measurement of the polarization value of a protein-fluorescent small molecule mixture. The FSML plate was diluted several times with H_2O + 0.05 % Tween with a roboter to reach a concentration of 30 nM. A small volume (2 µL) of this solution was filled on the assay plate and 4 µL of the protein solution were added to reach a final assay volume of 6 µL and a final small molecule concentration of 10 nM. The final protein concentration was depending on the protein. The plate was directly centrifuged and FP was measured. A second plate containing only the small molecules in protein buffer was also measured to evaluate the ground signal of each compound.

IV.8.4.   Data analysis and hits selection

	GST-CD2BP2	GST-AF6	GST-tudor-Y65N	GST-MTB	DAGK	proteinase_K
VUP-FSML1-003						
VUP-FSML1-007						
VUP-FSML1-013						
VUP-FSML1-015						
VUP-FSML1-017	x	x	x			
VUP-FSML1-019						
VUP-FSML1-024						
VUP-FSML1-031	x					
VUP-FSML1-033		x	x			
VUP-FSML1-034			x			
VUP-FSML1-043						
VUP-FSML1-049	x		x		x	
VUP-FSML1-063						x
VUP-FSML1-075						x
VUP-FSML1-076		x				x
VUP-FSML1-100						
VUP-FSML1-113						
VUP-FSML1-114						
VUP-FSML1-121			x			
VUP-FSML1-122						x
VUP-FSML1-124						
VUP-FSML1-129					x	
VUP-FSML1-161					x	
VUP-FSML1-170	x					
VUP-FSML1-171						x
VUP-FSML1-174						x
VUP-FSML1-193		x				
VUP-FSML1-201		x				
VUP-FSML1-210						
VUP-FSML1-218						x
VUP-FSML1-230					x	
VUP-FSML1-240						x
VUP-FSML1-255						

	Lysozym	BSA	IK-domain	RKIP	VBC	Shank3
VUP-FSML1-003				x		
VUP-FSML1-007	x	x	x		x	x
VUP-FSML1-013				x		
VUP-FSML1-015				x		x
VUP-FSML1-017						
VUP-FSML1-019	x	x		x	x	
VUP-FSML1-024				x		
VUP-FSML1-031						
VUP-FSML1-033						
VUP-FSML1-034						
VUP-FSML1-043	x	x				
VUP-FSML1-049	x					
VUP-FSML1-063						
VUP-FSML1-075						
VUP-FSML1-076						
VUP-FSML1-100					x	
VUP-FSML1-113		x		x		
VUP-FSML1-114		x		x		
VUP-FSML1-121						
VUP-FSML1-122		x				
VUP-FSML1-124		x				
VUP-FSML1-129		x	x		x	
VUP-FSML1-161			x		x	
VUP-FSML1-170						
VUP-FSML1-171						
VUP-FSML1-174						
VUP-FSML1-193						
VUP-FSML1-201						
VUP-FSML1-210	x					
VUP-FSML1-218						
VUP-FSML1-230						
VUP-FSML1-240						
VUP-FSML1-255					x	

Table 16: Results of the screening of FSML for binding to different proteins with FP detection. The increase in FP upon protein addition was considered. For proteins with a GST tag, compounds binding to GST were eliminated. The better hits for each protein are shown in this table.

FP values of compounds with and without protein were compared and molecules having high increases of the FP signal upon binding to the protein were selected as hits. Four proteins had a GST tag. The binding to GST alone was also measured on a separate plate and hits that were also binding to GST were eliminated. Results are summarized on Table 16.

For almost each protein, hits could be found. The only one without hits was also the lowest concentrated (18 µM). A repetition of the assay at a higher concentration would maybe also have shown some positive results.

Two kinds of hits could be identified. Some molecules were binding to several proteins whereas some others were specific.

IV.8.5.   Hits validation

Further investigations were first done on the CD2BP2 GYF domain. Sixteen molecules were selected and FP titration curves were measured, as already described above in the case of peptides. The results are shown on Figure 76:
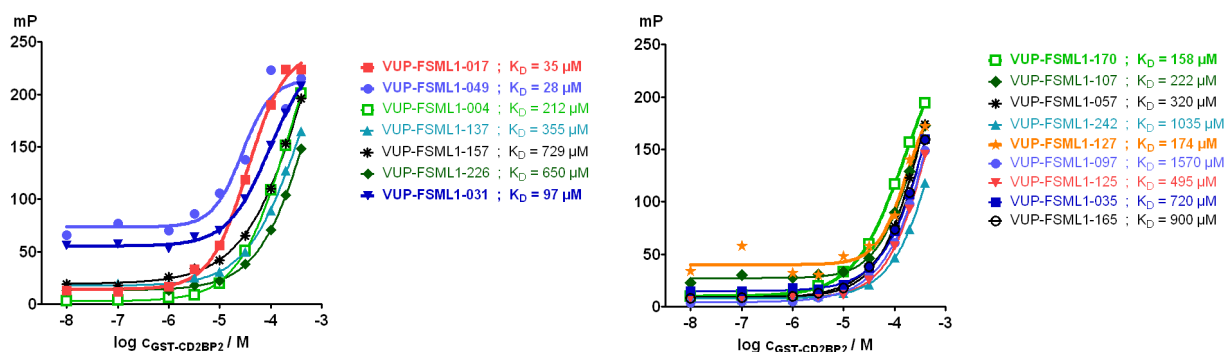
Figure 76: K_D determination for 16 best hits binding to the CD2BP2 GYF domain with FP.

All compounds had a titration curve fitting to the model very well and three of them had even a K_D value under 100 µM. These three molecules are shown on Figure 77. Two of them are multiple hits (VUP-FSML1-017 and VUP-FSML1-049) whereas the third (VUP-FSML1-031) binds only to the CD2BP2 GYF domain.
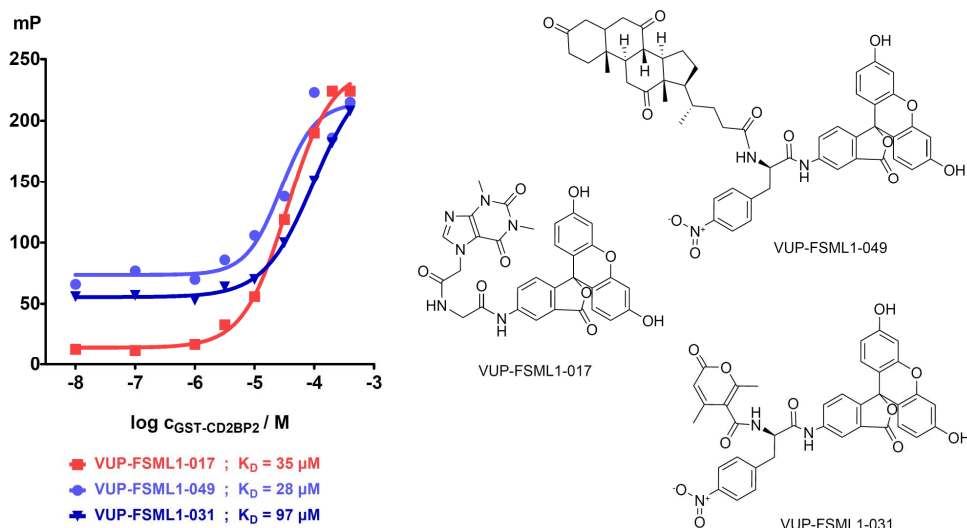
Figure 77: Three best hits binding to the CD2BP2 GYF domain with an affinity better than 100 µM.

These compounds will further be submitted to NMR analyses by D. Kosslick (group of C. Freund, FMP, Berlin) to confirm the binding and identify precisely the interacting sites.

V. CONCLUSION AND OUTLOOK

A novel strategy has been developed for the *C*-terminal labeling of peptides and the labeling of small molecules: a labeling resin carrying aminofluorescein has been synthesized by linking *N*-Fmoc-aminofluorescein to 2-chloro-trityl chloride resin after the synthesis way of *N*-Fmoc-aminofluorescein has been established through a protection-deprotection sequence.

This resin was used to build libraries of fluorescein labeled peptides binding to two protein domains of the GYF family: the CD2BP2 and PERQ2 GYF domains. An FP probe Fluo-(Bpa)EFGPPPGWLGR having a K_D value of 2.3 µM could be identified and enabled the realization of an FP screening of the ChemBioNet library to find small molecule ligands of CD2BP2.

The resin was further used for the synthesis of a library of fluorescein labeled small molecules, starting from fragments that were selected to have a high diversity. The binding of the library compounds to a set of 12 proteins was evaluated with direct FP measurements. More precise investigations were continued on the CD2BP2 GYF domain and three hits under 100 µM could be successfully identified.

The resin practicability could be extended to the synthesis of larger libraries using equipment for automated synthesis, and to different reactions. The labeled library could be tested in further assays, e.g. cellular assays to evaluate characteristics of the molecules in living cells [157-159].

VI. EXPERIMENTAL PART

VI.1. Materials and methods

VI.1.1. Chemicals and solvents

Solvents were purchased from Fisher Scientific, J.T. Baker, Acros Organics, VWR, Fluka in p.a. quality. Anhydrous solvents, puriss. Absolute grade, over molecular sieve, $H_2O \leq 0.005\%$, were purchased by Fluka.

HPLC solvents – acetonitrile, methanol – were HPLC gradient grade, purchased by Fisher Scientific. HPLC-water was obtained from a Millipore Milli-Q$^®$ Biocel A10 purification system. Formic acid for LC-MS was p.a. 98-100 % grade from Merck. TFA for preparative HPLC was 99 % extra pure grade from Acros Organics. Solvents for preparative HPLC were degassed prior to use under helium gas stream.

Deuterated solvents for NMR were purchased by Deutero GmbH (Kastellaun).

Screened small molecules were part of the FMP library, purchased by ChemDiv as well as reordered compounds.

All other chemicals, reagents and resins were purchased by Sigma-Aldrich, Acros Organics, Merck, Novabiochem, EMP Biotech GmbH and used without further purification.

VI.1.2. Thin-layer chromatography (TLC)

TLC were done on silica gel 60 F_{254}, 5×10 cm, coated aluminium sheets from Merck. Products were observed under UV-light ($\lambda = 254$ nm, $\lambda = 366$ nm).

VI.1.3. Analytical liquid chromatography – mass spectroscopy (LC-MS)

LC-MS analysis were carried out on an Agilent 1100 Series Liquid Chromatography Station equipped with a diode-array detector (DAD) and a Single Quadrupole Mass

Spectrometer with electrospray ionization (ESI). The following eluent solutions were used. A: water + 0.1 % formic acid and B: acetonitrile + 0.1 % formic acid. A switching valve allowed the use of 6 different columns.

A C_{18} endcapped, 100 Å, 5 μM, 4 × 250 mm column (Macherey-Nagel, Nucleodur®) was used with the following gradient method: 5 to 100 % B in 30 min, then 5 min isocratic 100 % B, flow 0.8 mL·min^{-1}

A Zorbax® SB-C18, 3.5 μM, 4.6 × 30 mm rapid resolution column (Agilent Technologies) was used with the following gradient method: 5 to 100 % B in 6 min, then 2 min isocratic 100 % B, flow 1.2 mL·min^{-1} or 5 to 100 % B in 12 min, then 3 min isocratic 100 % B, flow 0.8 mL·min^{-1}.

The *m/z* range 100-3000 was scanned in positive mode with a voltage of 3000 V and a fragmentation of 70 V. Selected ion monitoring (SIM) was used on a second channel for selective mass detection. For a good ionization in positive mode, formic acid was added to the eluents.

Data were analyzed with the Agilent LC/MSD ChemStation B.01.03.

HPLC purity was determined at 254 ± 4 nm or 220 ± 4 nm.

VI.1.4. Preparative high performance liquid chromatography (HPLC)

Purifications were carried out on an Agilent 1100 Series preparative HPLC with the following eluent solutions. A: water + 0.1 % TFA and B: acetonitrile + 0.1 % TFA. For good peak sharpness and separation by peptide purification, trifluoroacetic acid was used. The gradients used for purifications were determined on the analytical LC-MS and scaled-up to the preparative instrument. A volume of 2 to 8 mL was injected and flows of 10 to 15 mL·min^{-1} were used on a C_{18} endcapped, 100 Å, 7 μM, 21 × 250 mm column (Macherey-Nagel, Nucleodur®).

VI.1.5. HR-MS

Samples were analyzed by Andreas Springer in the mass group of the organic chemistry department, Freie Universität, Berlin.

VI.1.6.   NMR spectroscopy

All one-dimensional (^{1}H, ^{13}C and DEPT) and two-dimensional (HMQC, DQF-COSY, HMBC) NMR spectra were recorded on a Bruker AVANCE™ 300 MHz spectrometer and analyzed with Topspin 2.0.a

VI.1.7.   ATR-IR spectroscopy

IR-spectra of resins were recorded on a Nicolet Impact 400 series FT-IR spectrometer equipped with a PIKE MIRacle™ ATR-IR (KBr) detection system with a ZnSe crystal for pressing the sample. The measurements were performed on the 400–4000 cm^{-1} range, with a number of scans of 32, a resolution of 16, a Happ-Genzel apodization and manual background substraction.

VI.1.8.   UV spectroscopy

UV/Vis absorption spectra were measured on a JASCO V-550 UV/Vis Spectrophotometer and analyzed with the Jasco Spectra Manager 1.54.03.

VI.1.9.   Fluorescence spectroscopy

Fluorescence emission spectra were measured on a Cary Eclipse Fluorescence Spectrophotometer from Varian.

VI.1.10.   Fluorescence excitation and emission spectra

For a representation of the fluorescence excitation and emission spectra, the two absorption and fluorescence emission datasets were imported from the two different spectrometers. The excitation data were exported from the JASCO UV/Vis spectrophotometer under ".txt" format, the emission data were exported from the Cary Eclipse Fluorescence Spectrophotometer under ".csv" format. Both were imported in Microsoft Excel, normalized and finally imported and plotted with Origin 7.

VI.1.11. Fluorescence quantum yield determination

For quantum yield determination, the samples were compared to a reference solution of 10 µM fluorescein in 0.1 M aq. NaOH with a quantum yield of fluorescence of 0.97. The unknown products were dissolved at the same concentration in the same solvent as the reference solution. The "intersection point method" was applied.

First, absorption spectra of unknown sample and reference solution were measured to determine the wavelength at the intersection of both curves, where the two products have the same extinction coefficient. This was chosen as excitation for the fluorescence emission measurements. Finally, the fluorescence data were analyzed in Origin 7.0 and the surfaces under both curves determined.

Under these conditions, the fluorescence quantum yield Φ of the unknown product is given by Equation 6:

$$\Phi = \Phi_{ref} \left(I / I_{ref} \right)$$

Equation 6: Determination of the fluorescence quantum yield. Φ: quantum yield of the sample; Φ_{ref}: quantum yield of the standard solution; I: fluorescence intensity of the sample; I_{ref}: fluorescence intensity of the standard solution.

VI.1.12. Fluorescence polarization

Fluorescence polarization measurements were carried out on low-volume round-bottom black non binding surface 384-multiplates from Corning (Nr. 3676) with a GenioPro Reader from Tecan. For molecules labeled with fluorescein, the following filters were incorporated (installed): excitation at 485 $\pm$ 20 nm and emission at 535 $\pm$ 25 nm and the following polarization settings were used: a gain of 90, 10 reads per well and an integration time of 40 µs. Previous to measurement, the plates were mixed on an Eppendorf® MixMate™ shaker at 1950 tr·min^{-1} for 15 s, and centrifuged on an Eppendorf® Centrifuge 5810 R at 4000 tr·min^{-1} for 2 min to eliminate bubbles.

The results were analyzed by plotting and fitting the curves with the software Prism 5.0 (Graph Pad Software, Inc., San Diego, CA).

VI.1.13. Isothermal titration calorimetry (ITC)

ITC measurements were performed on a VP-ITC microcalorimeter from MicroCal and data analyzed with the MicroCal AddOn of Origin 7.

VI.1.14. Circular Dichroism (CD)

CD measurements were performed on a JASCO J-720 spectropolarimeter.

VI.1.15. Solid phase synthesis

Reactions on solid phase were performed in 2, 5, 10, or 20 mL polypropylene syringes with adapted polyethylene frits and caps, all purchased by Roland Vetter Laborbedarf OHG, Ammerbuch. For small amounts, a VXR basic IKA-Vibrax® shaker was used, and a KS-15 B control shaker from Edmund Bühler GmbH for larger scale syntheses.

VI.1.16. Determination of the resin loading / coupling yield

VI.1.16.1. By weighing of the resin.

The resin was precisely weighed before the reaction, and once the coupling accomplished after having been washed and dried. The resulting experimental mass difference is given by Equation 7:

$$\Delta m_{exp} = m_{exp} - m_0$$

Equation 7: Determination of the experimental mass difference. m_{exp}: resin weight after the reaction; m_0: resin weight for the reaction.

The theoretical mass difference was calculated using Equation 8:

$$\Delta m_{th} = m_0 \cdot x_0 \cdot \Delta M_{th} = m_0 \cdot x_0 (M - M_X)$$

Equation 8: Determination of the theoretical mass difference. m_0: resin weight for the reaction / g; x_0: resin loading for the reaction / mol·g^{-1}; M: molecular mass of the molecule added in the reaction / g·mol^{-1}; M_X: molecular mass of all substracted atoms / g·mol^{-1}.

The coupling yield is given by the quotient of the two above mentioned values (Equation 9):

$$\text{coupling yield}: \frac{\Delta m_{exp}}{\Delta m_{th}}$$

Equation 9: Coupling yield, as calculated by weighing of the resin

VI.1.16.2. By spectroscopic measurement after Fmoc cleavage.

After washing and drying, about 2 mg resin were weighed in a 2 mL volumetric flask, which was filled up with a 20% piperidine in DMF solution and shaken 30 min. The concentration of the cleaved Fmoc residue was determined by UV absorption measurements, using a 20% piperidine in DMF solution as reference.

Spectra were measured in the 320–250 nm range.

For quantification, the Beer-Lambert law was used (Equation 10):

$$A_\lambda = \varepsilon_\lambda \cdot \ell \cdot c$$

Equation 10: Beer-Lambert law. A_λ: absorbance at the wavelength λ; ε_λ: molar extinction coefficient at the wavelength λ / L·mol^{-1}·cm^{-1}); ℓ: cuvette length / cm; c: concentration / mol·L^{-1}.

Absorptions on three characteristic peaks of the Fmoc group were noted: at λ_1 = 267 nm ($\varepsilon_{\lambda 1}$ = 17 500 L·mol^{-1}·cm^{-1}); at λ_2 = 289 nm ($\varepsilon_{\lambda 2}$ = 5 800 L·mol^{-1}·cm^{-1}); and at λ_3 = 301 nm ($\varepsilon_{\lambda 3}$ = 7 800 L·mol^{-1}·cm^{-1}).

The concentration of the Fmoc residue was given by Equation 11:

$$c = \frac{1}{3} \sum_{i=1}^{3} \frac{A_{\lambda i}}{\varepsilon_{\lambda i} \cdot \ell}$$

Equation 11: Concentration of the Fmoc residue. c: concentration / mol·L^{-1}; $A_{\lambda i}$: absorbance at the wavelength λ_i; $\varepsilon_{\lambda i}$: molar extinction coefficient at the wavelength λ_i / L·mol^{-1}·cm^{-1}; ℓ: cuvette length / cm.

The experimental resin loading was given by Equation 12:

$$x_{exp} = \frac{c \cdot V}{m}$$

Equation 12: Determination of the resin loading in SI units. x_{exp}: resin loading / mol·g^{-1}; m: resin weight / g; V: solution volume / L; c: concentration / mol·L^{-1}.

To become a loading in the unit commonly used in SPS: mmol·g^{-1}, a correction factor of 10^3 had to be used (Equation 13):

$$x_{exp}(mmol \cdot g^{-1}) = \frac{1000 \cdot V(mL)}{3 \cdot m(mg)} \sum_{i=1}^{3} \frac{A_i}{\varepsilon_{\lambda i}(L \cdot mol^{-1} \cdot cm^{-1}) \cdot \ell(cm)}$$

Equation 13: Determination of the resin loading in commonly used units. x_{exp}: resin loading / mmol·g^{-1}; V: solution volume / mL; m: resin weight / mg; $A_{\lambda i}$: absorbance at the wavelength λ_i; $\varepsilon_{\lambda i}$: molar extinction coefficient at the wavelength λ_i / L·mol^{-1}·cm^{-1}; ℓ : cuvette length / cm.

The theoretical maximal loading, in case of 100 % coupling efficiency, could be determined by considering the mass variation after introduction of the *N*-Fmoc-aminofluorescein.

The loading of a resin is the inverse of the apparent molecular mass (Equation 14):

$$x_0 = \frac{1}{M_{res} + M_0} \quad \text{and} \quad x_{th} = \frac{1}{M_{res} + M'}$$

Equation 14: Relation between the resin loading and the molecular mass of the bound moieties. x_0: loading of the initial resin / mol·g^{-1}; x_{th}: theoretical maximal loading of the resin after reaction / mol·g^{-1}; M_0: molecular mass of the molecule initially bound to the resin / g·mol^{-1}; M_{res}: molecular mass corresponding to the unvariable part of the resin / g·mol^{-1}; M': molecular mass of the molecule finally bound to the resin / g·mol^{-1}.

The resulting molecular mass difference between product and starting material is shown on Equation 15:

$$M' - M_0 = M - M_X$$

Equation 15: Molecular mass difference between product and starting material. M': molecular mass of the molecule finally bound to the resin / g·mol^{-1}; M_0: molecular mass of the molecule initially bound to the resin / g·mol^{-1}; M: molecular mass of the molecule added in the reaction / g·mol^{-1}; M_X: molecular mass of all substracted atoms / g·mol^{-1}.

Resulting from the precedent equations:

$$M_{res} + M' = M_{res} + M_0 + M - M_X = M_0\left(1 + \frac{M - M_X}{M_0}\right) = \frac{1 + x_0(M - M_X)}{x_0}$$

And the theoretical maximal loading x_{th} is:

$$x_{th} = \frac{x_0}{1 + x_0 (M - M_x)}$$

Equation 16: Determination of the theoretical maximal resin loading in SI units. x_{th}: theoretical maximal loading of the resin after reaction / mol·g^{-1}; x_0: loading of the initial resin / mol·g^{-1}; M: molecular mass of the molecule added in the reaction / g·mol^{-1}; M_x: molecular mass of all substracted atoms / g·mol^{-1}.

To become a loading in the commonly used unit: mmol·g^{-1}, a correction factor of 10^3 had to be used:

$$x_{th} \,(mmol \cdot g^{-1}) = \frac{x_0 \,(mmol \cdot g^{-1})}{1 + \dfrac{x_0 \,(mmol \cdot g^{-1})}{1000} \left[M\,(g \cdot mol^{-1}) - M_x\,(g \cdot mol^{-1}) \right]}$$

Equation 17: Determination of the theoretical maximal resin loading in commonly used units. x_{th}: theoretical maximal loading of the resin after reaction / mmol·g^{-1}; x_0: loading of the initial resin / mmol·g^{-1}; M: molecular mass of the molecule added in the reaction / g·mol^{-1}; M_x: molecular mass of all substracted atoms / g·mol^{-1}.

The yield of the reaction was determined by the quotient of the two above mentioned values (Equation 18):

$$coupling\ yield : \frac{x_{exp}}{x_{th}}$$

Equation 18: Coupling yield, as calculated after spectroscopic measurements. x_{exp}: real resin loading after reaction, resulting from spectroscopic measurements / mmol·g^{-1}; x_{th}: theoretical maximal loading of the resin after reaction / mol·g^{-1}.

VI.1.17. Kaiser test [160,161]

The Kaiser test was used to find traces of primary amines, in particular for the control of peptide coupling completion. To a few resin beads (1-3 mg) preleved in a 500 µL Eppendorf-Cup, 50 µL of the three following solutions were added:

Solution I :   1 mL 1mM aq. KCN + 49 mL pyridine
Solution II :   2 g ninhydrin in 40 mL ethanol
Solution III : 16 g phenol in 40 mL ethanol

The mixture was heated for 5 min at 110℃ and the r esulting color was observed. In absence of free primary amines, the color remains yellow, whereas in presence of a primary amine (all amino-acids excepted proline), the reaction described below happens and a blue coloration of the resin beads or of the solution can be observed.
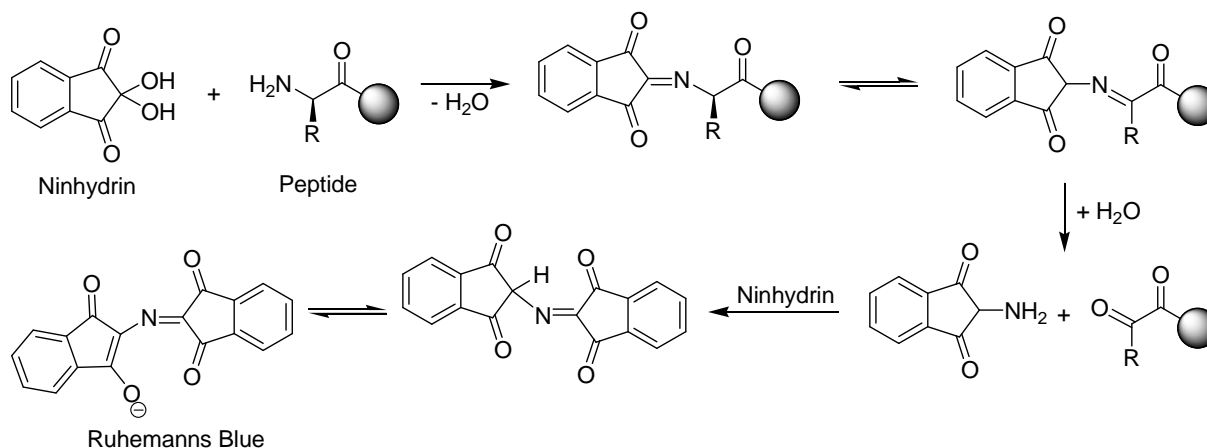
Scheme 34: Mechanism of the Kaiser test

VI.1.18.   Chloranil test

The chloranil test was applied for identification of secondary amines. To a small sample of the resin (1-3 mg) in the presence of 200 µL acetone, 50 µL of a saturated solution of chloranil (2,3,5,6-tetrachloro-1,4-benzoquinone) in toluene were added, and the sample was allowed to react for 5 min at RT. In the presence of free amines, a charge-transfer complex is formed between the amine and the chloranil, resulting in a blue or green coloration of the resin beads, whereas in the absence of any amine, the beads remain yellow.
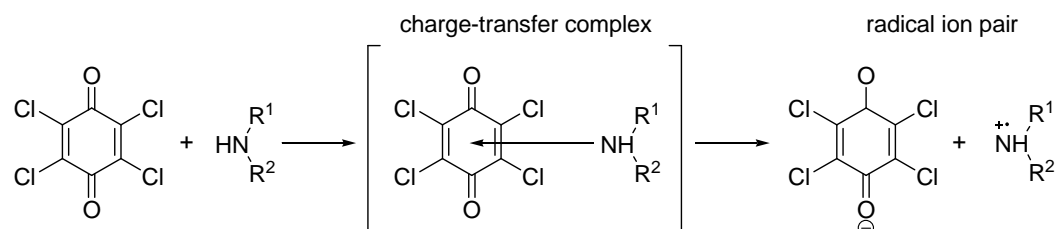
Scheme 35: Mechanism of the chloranil test

VI.1.19. Lyophilization

Lyophilization of a product consists in sublimation of the water contained in it, that means direct transformation from solid into gas, as shown on the phase diagram of water (Figure 78), in order to obtain a more voluminous powder, that is easier to handle, for example by weighing .
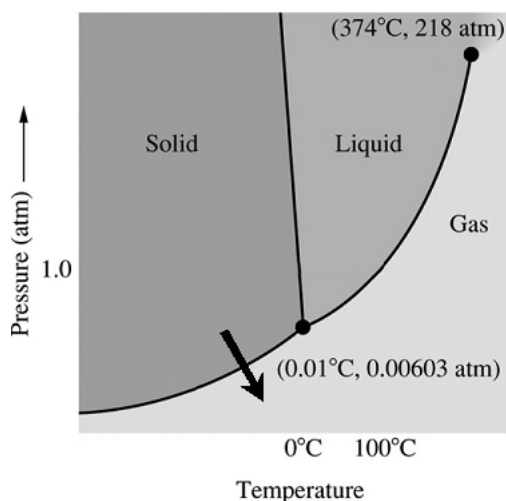
Figure 78: Phase diagram of water

Substances were dissolved in a mixture of H_2O / *tert*-BuOH, frozen in liquid nitrogen and lyophilized on a Christ Alpha 2-4 LSC instrument at -90 °C under a vacuum of 0.01–0.05 mbar.

VI.1.20. Molecular modeling and database analysis

Three-dimensional structures of proteins and protein-ligand interactions were viewed and modeled with the software MOE 2006.08 from the Chemical Computing Group, Inc.

Chemicals databases were created and studied with ChemFinder Ultra 8.0 and MOE 2006.08.

VI.2. Synthesis

VI.2.1. Synthesis of the resin-bound fluorescein labeling reagent

Protection of the phenols of 4-aminofluorescein (10) with pivalic anhydride to di-*O*-pivaloyl-4-aminofluorescein (22).

Cs_2CO_3 (9.38 g, 28.8 mmol) was added to a solution of 4-aminofluorescein **10** (10.03 g, 28.8 mmol) in DMF (250 mL). The reaction mixture was cooled to 0°C with an ice bath and pivalic anhydride (13.37 mL, 72.0 mmol) was added dropwise. The solution was stirred for 2 h at room temperature. Cs_2CO_3 was filtered off, DMF was evaporated *in vacuo*. The residue was diluted in ethyl acetate, washed with water and brine, dried with Na_2SO_4, filtered, and evaporated to yield a yellow powder **22** (14.61 g, 98%) which was used as crude material in the next step.

Di-*O*-pivaloyl-4-aminofluorescein (22)
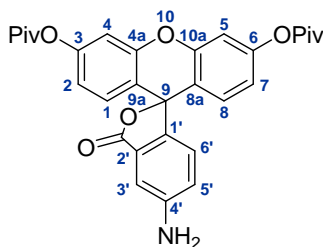
LC-MS: HPLC r_t = 30.0 min, purity 92%; ESI-MS calcd 516.2 *m/z*, found 516.2 [M + H]+.

^{1}H-NMR (300 MHz; CD$_2$Cl$_2$): δ 1.36 (s, 18 H, C*H$_3$ tert*-Bu), 6.80 (dd, *J* = 8.6 and 2.2 Hz, 2 H, xanthene C*H*-2, 7), 6.90 (d, *J* = 8.2 Hz, 1 H, phenyl C*H*-6'), 6.91 (d, *J* = 8.5 Hz, 2 H, xanthene C*H*-1, 8), 6.98 (dd, *J* = 8.2 and 2.0 Hz, 1 H, phenyl C*H*-5'), 7.06 (d, *J* = 2.1 Hz, 2 H, xanthene C*H*-4, 7), 7.18 (d, *J* = 1.9 Hz, 1 H, phenyl C*H*-3').

^{13}C-NMR (300 MHz; CD$_2$Cl$_2$): δ 27.37 (CH$_3$, *tert*-Bu), 39.63 (C quat., *tert*-Bu), 82.03 (C quat., xanthene C-9), 108.71 (CH, phenyl C-3'), 110.70 (CH, xanthene C-4, 5), 117.73 (C quat., xanthene C-8a, 9a), 118.22 (CH, xanthene C-2, 7), 123.00 (CH, phenyl C-5'), 125.01 (CH, phenyl C-6'), 128.26 (C quat., phenyl C-2'), 129.51 (CH, xanthene C-1, 8), 142.64 (CH, phenyl C-1'), 149.71 (CH, phenyl C-4'), 152.28 (C quat., xanthene C-4a, 10a), 153.16 (C quat., xanthene C-3, 6), 169.98 (C quat., lactone C=O), 177.10 (C quat., pivaloyl C=O).

HR-MS: calcd 516.2017, found 516.2019 (Δm = 0.39 ppm).

Protection of the aniline of di-*O*-pivaloyl-4-aminofluorescein (2) to di-*O*-pivaloyl-4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluorescein (23).

1 N NaOH (26.5 mL, 26.5 mmol) and Fmoc-chloride (10.28 g, 39.7 mmol) were added stepwise over 1 h to a solution of **22** (14.61 g, 26.5 mmol) in THF (250 mL) at 0°C. The solution was stirred overnight at room temperature. THF was evaporated, water was added (250 mL) and the pH was adjusted at 2 with addition of KHSO$_4$. The product was extracted with ethyl acetate. The combined organic layers were washed with water and brine, dried with Na$_2$SO$_4$, filtered, and evaporated. The crude product was recristallized from CH$_2$Cl$_2$ / MeOH to yield white crystals of **23** (16.25 g, 85%).

Di-*O*-pivaloyl-4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluorescein (23)
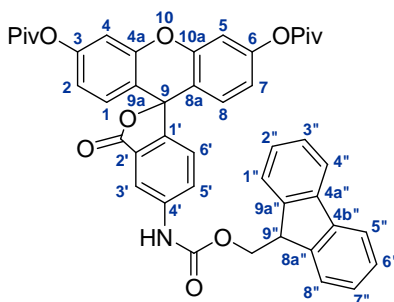
LC-MS: HPLC r$_t$ = 33.7 min, purity > 99%; ESI-MS calcd 738.3 *m/z*, found 738.2 [M + H]$^+$.

^{1}H-NMR (300 MHz; CD$_2$Cl$_2$): δ 1.36 (s, 18 H, C*H$_3$ tert*-Bu), 4.30 (t, *J* = 6.3 Hz, 1H, Fmoc C*H*-9"), 4.59 (d, *J* = 6.3 Hz, 2H, Fmoc C*H$_2$*), 6.80 (d, *J* = 8.6 Hz, 2H, xanthene C*H*-2, 7), 6.86 (d, *J* = 8.6 Hz, 2H, xanthene C*H*-1, 8), 7.07 (s, 2H, xanthene C*H*-4, 5), 7.11 (d, *J* = 8.3 Hz, 1H, phenyl C*H*-6'), 7.33 (t, *J* = 8.3 Hz, 2H, fluorenyl C*H*-2", 7"), 7.42 (t, *J* = 8.3 Hz, 2H, fluorenyl C*H*-3", 6"), 7.66 (d, *J* = 7.3 Hz, 2H, fluorenyl C*H*-1", 8"), 7.68 (d, 1H, phenyl C*H*-5'), 7.80 (d, *J* = 7.4 Hz, 2H, fluorenyl C*H*-4",5"), 8.10 (s, 1H, phenyl C*H*-3').

^{13}C-NMR (300 MHz; CD$_2$Cl$_2$): δ 27.38 (CH$_3$, *tert*-Bu), 39.66 (C quat., *tert*-Bu), 47.66 (CH, fluorenyl C-9"), 67.65 (CH$_2$, Fmoc), 82.24 (C quat., xanthene C-9), 110.86 (CH, xanthene C-4, 5), 114.37 (CH, phenyl C-3'), 116.95 (C quat., xanthene C-8a, 9a), 118.37 (CH, xanthene C-2, 7), 120.60 (CH, fluorenyl C-4", 5"), 125.10 (CH, phenyl C-6'), 125.52 (CH, fluorenyl C-1", 8"), 126.45 (CH, phenyl C-5'), 127.71 (CH, fluorenyl C-2", 7"), 127.92 (C quat., phenyl C-2'), 128.38 (CH, fluorenyl C-3", 6"), 129.43 (CH, xanthene C-1, 8), 140.84 (C quat., phenyl C-1'), 141.94 (C quat., fluorenyl C-4a", 4b"), 144.30 (C quat., fluorenyl C-8a", 9a"), 147.71 (C quat., phenyl C-4'), 152.25 (C quat., xanthene C-4a, 10a), 153.35 (C quat., xanthene C-3, 6), 153.78 (C quat., Fmoc C=O), 169.24 (C quat., lactone C=O), 177.10 (C quat., pivaloyl C=O).

^{1}H-NMR (300 MHz; DMF-d7): δ 1.36 (s, 18 H, CH_3 tert-Bu), 4.41 (t, J = 6.6 Hz, 1H, Fmoc CH-9"), 4.60 (d, J = 6.6 Hz, 2H, Fmoc CH_2), 7.04 (dd, J = 8.6 and 2.1 Hz, 2H, xanthene CH-2, 7), 7.10 (d, J = 8.6 Hz, 2H, xanthene CH-1, 8), 7.31 (d, J = 2.1 Hz, 2H, xanthene CH-4, 5), 7.40 (t, J = 7.3 Hz, 2H, fluorenyl CH-2", 7"), 7.44 (d, 1H, phenyl CH-6'), 7.48 (t, J = 7.3 Hz, 2H, fluorenyl CH-3", 6"), 7.83 (d, J = 7.8 Hz, 2H, fluorenyl CH-1", 8"), 7.95 (d, 1H, phenyl CH-5'), 7.98 (d, J = 7.8 Hz, 2H, fluorenyl CH-4",5"), 8.39 (s, 1H, phenyl CH-3').

^{13}C-NMR (300 MHz; DMF-d7): δ 26.73 (CH$_3$, tert-Bu), 39.16 (C quat., tert-Bu), 47.28 (CH, fluorenyl C-9"), 66.81 (CH$_2$, Fmoc), 81.60 (C quat., xanthene C-9), 110.66 (CH, xanthene C-4, 5), 112.88 (CH, phenyl C-3'), 117.15 (C quat., xanthene C-8a, 9a), 118.72 (CH, xanthene C-2, 7), 120.47 (CH, fluorenyl C-4", 5"), 125.08 (CH, phenyl C-6'), 125.53 (CH, fluorenyl C-1", 8"), 126.07 (CH, phenyl C-5'), 127.40 (C quat., phenyl C-2'), 127.48 (CH, fluorenyl C-2", 7"), 128.10 (CH, fluorenyl C-3", 6"), 129.62 (CH, xanthene C-1, 8), 141.56 (C quat., fluorenyl C-4a", 4b"), 142.12 (C quat., phenyl C-1'), 144.40 (C quat., fluorenyl C-8a", 9a"), 146.68 (C quat., phenyl C-4'), 151.75 (C quat., xanthene C-4a, 10a), 153.17 (C quat., xanthene C-3, 6), 154.18 (C quat., Fmoc C=O), 168.91 (C quat., lactone C=O), 176.50 (C quat., pivaloyl C=O).

HR-MS: calcd 738.2698, found 738.2737 (Δm = 5.28 ppm).

Synthesis of 4-(9-Fluorenylmethoxycarbonyl)-aminofluorescein (21) through direct protection of 4-aminofluorescein (10)

1 N NaOH (2.88 mL, 2.88 mmol) and Fmoc-chloride (1.12 g, 4.32 mmol) were added stepwise over 30 min to a solution of **10** (1.00 g, 2.88 mmol) in THF (10 mL) at 0°C. The solution was stirred 3 h at room temperature. THF was evaporated, water was added (25 mL) and the pH was adjusted at 2 with addition of KHSO$_4$. The product was extracted with ethyl acetate. The combined organic layers were washed with water and brine, dried with Na$_2$SO$_4$, filtered, and evaporated. The crude product was purified with flash chromatography using first hexane / ethyl acetate to elute the twice protected by-product; then, the column was washed with dichloromethane / methanol to recover **21** (0.75 g, 46%, HPLC-purity 96%).

Synthesis of 4-(9-Fluorenylmethoxycarbonyl)-aminofluorescein (21) through deprotection of di-*O*-pivaloyl-4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluorescein (23).

A solution of **23** (16.25 g, 22.1 mmol) in 95% TFA (250 mL) was stirred overnight at 60°C. TFA was evaporated, the residue was dissolved in ethyl acetate and washed with water. The organic layer was extracted with a Na_2CO_3 aqueous solution at pH 10 to transfer the product **21** to the aqueous layer, leaving unreacted traces of starting material **23** in the organic layer. The aqueous layer was washed with ethyl acetate, the pH was adjusted to pH 2-4 with addition of $KHSO_4$ salt, and extracted with ethyl acetate to transfer the product **21** back in the organic layer. The organic layer was washed with water and brine, dried with Na_2SO_4 and evaporated to yield an orange powder **21** (11.54 g, 92%, HPLC-purity > 99%).

4-(9-Fluorenylmethoxycarbonyl)-aminofluorescein (21)
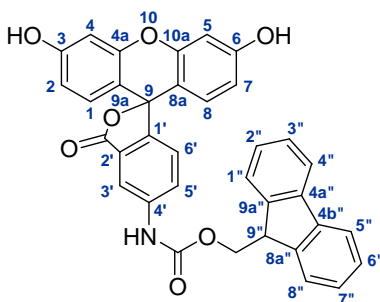
LC-MS: HPLC r_t = 24.95 min, purity 96% (*from compound* **10**), purity > 99% (*from compound* **23**); ESI-MS calcd 570.1 *m/z*, found 570.0 [M + H]$^+$.

^{1}H-NMR (300 MHz; DMF-d7): δ 4.40 (t, *J* = 6.9 Hz, 1H, Fmoc C*H*-9"), 4.59 (d, *J* = 7.0 Hz, 2H, Fmoc C*H_2*), 6.68 (dd, *J* = 8.5 and 2.0 Hz, 2H, xanthene C*H*-2, 7), 6.75 (d, *J* = 8.5 Hz, 2H, xanthene C*H*-1, 8), 6.78 (d, *J* = 1.9 Hz, 2H, xanthene C*H*-4, 5), 7.33 (d, *J* = 8.4 Hz, 1H, phenyl C*H*-6'), 7.40 (t, *J* = 7.3 Hz, 2H, fluorenyl C*H*-2", 7"), 7.48 (t, *J* = 7.3 Hz, 2H, fluorenyl C*H*-3", 6"), 7.83 (d, *J* = 7.1 Hz, 2H, fluorenyl C*H*-1", 8"), 7.93 (dd, 1H, *J* = 8.4 and 1.6 Hz, phenyl C*H*-5'), 7.98 (d, *J* = 7.8 Hz, 2H, fluorenyl C*H*-4",5"), 8.32 (d, *J* = 1.4 Hz, 1H, phenyl C*H*-3').

^{13}C-NMR (300 MHz; DMSO-d6): δ 46.67 (CH, fluorenyl C-9"), 65.75 (CH$_2$, Fmoc), 79.16 (C quat., xanthene C-9), 102.25 (CH, xanthene C-4, 5), 109.82 (C quat., xanthene C-8a, 9a), 112.51 (CH, phenyl C-3'), 112.9 (CH, xanthene C-2, 7), 120.20 (CH, fluorenyl C-4", 5"), 124.46 (CH, phenyl C-6'), 125.08 (CH, fluorenyl C-1", 8"), 125.67 (CH, phenyl C-5'), 127.15 (C quat., phenyl C-2'), 127.15 (CH, fluorenyl C-2", 7"), 127.72 (CH, fluorenyl C-3", 6"), 129.10 (CH, xanthene C-1, 8), 140.84 (C quat., phenyl C-1'), 140.86 (C quat., fluorenyl C-4a", 4b"), 143.68 (C quat., fluorenyl C-8a", 9a"), 146.14 (C quat., phenyl C-

4'), 151.97 (C quat., xanthene C-4a, 10a), 153.54 (C quat., xanthene C-3, 6), 159.51 (C quat., Fmoc C=O), 168.64 (C quat., lactone C=O).

HR-MS: calcd 570.1547, found 570.1571 (Δm = 4.21 ppm).

Coupling of 4-(9-Fluorenylmethoxycarbonyl)-aminofluorescein (4) to trityl chloride resin yielding 4-(9-Fluorenylmethoxycarbonyl)-aminofluorescein trityl resin (25).

25
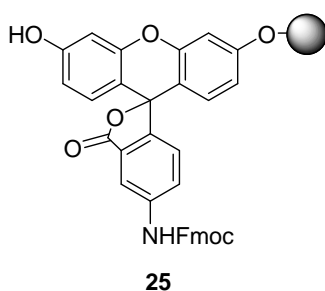
DIPEA (7.1 mL, 40.6 mmol) and 2-chlorotrityl chloride resin (11.2 g, 15.5 mmol, 1.4 mmol/g, 100-200 mesh, 1% DVB, Novabiochem®) were added to a solution of **21** (11.54 g, 20.3 mmol) in DMF (200 mL) and the reaction mixture was shaken for 3 h at room temperature. The solution was filtered off and the resin was washed with DMF. The remaining free reactive sites were blocked shaking the resin 2 times for 15 min in a mixture MeOH / DMF (1:9, v/v) with DIPEA (2.7 mL, 154.5 mmol). The resin was washed with DMF, THF and DCM to furnish an orange resin **25** (0.38 mmol.g^{-1}, 47%). The resin loading was determined by UV quantification of the Fmoc residue after Fmoc cleavage on a small sample.

Protection of the free phenol and carboxylic acid of 4-(9-Fluorenylmethoxycarbonyl)-aminofluorescein trityl resin (25) with the methoxyethoxymethyl, *tert*-butyldimethylsilyl and triethylsilyl groups.
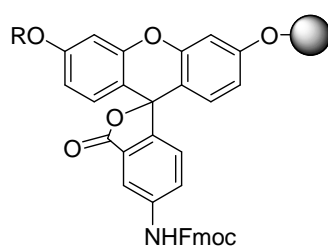
31a: R = MEM
31b: R = TBDMS
31c: R = TES

Di-*O*-(methoxyethoxymethyl)-4-(9-fluorenylmethoxycarbonyl)-aminofluorescein trityl resin (31a).

DIPEA (11.1 mL, 63.4 mmol) and MEM-Cl (6,24 mL, 52.8 mmol) were added to a suspension of **25** in DMF (200 mL). The reaction mixture was shaken for 90 min at room temperature. The resin was washed with DMF, THF and DCM to yield **31a** (0.25 mmol. g^{-1}, 100%) as a pale yellow resin.

Di-*O*-(*tert*-butyldimethylsilyl)-4-(9-fluorenylmethoxycarbonyl)-aminofluorescein trityl resin (31b).

Same procedure as for **31a**, using TBDMS-Cl.

Di-*O*-(triethylsilyl)-4-(9-fluorenylmethoxycarbonyl)-aminofluorescein trityl resin (31c).

Same procedure as for **31a**, using TES-Cl.

Deprotection of the aniline of Di-*O*-(methoxyethoxymethyl)-4-(9-fluorenyl-methoxycarbonyl)-aminofluorescein trityl resin (31a) to di-*O*-(methoxy-ethoxymethyl)-4-aminofluorescein trityl resin (32a).

32a
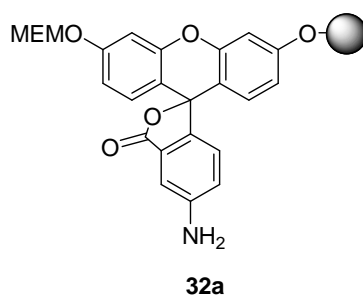
A suspension of **31a** in 20% piperidine in DMF (200 mL) was shaken for 2 min at room temperature, washed with DMF and shaken for another 10 min. The solution was filtered off and the resin was washed with DMF, THF, and DCM to furnish **32a** as an orange resin.

VI.2.2. Synthesis of *C*-terminally fluorescein labeled peptides with Fmoc Solid Phase Peptide Synthesis (SPPS) on the resin-bound fluorescein.

Coupling of the first amino-acid to di-*O*-(methoxyethoxymethyl)-4-aminofluorescein trityl resin (32a).

40
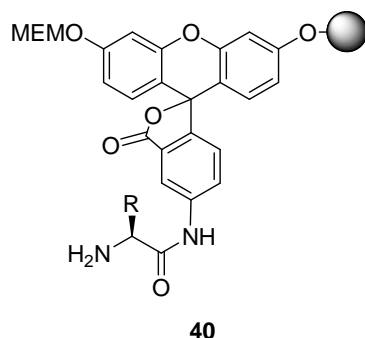
Fmoc-protected amino-acid (0.5 mmol) and DIC (77.6 µL, 0.5 mmol) were added to a suspension of **32a** (100 mg, 0.05 mmol) in DCM / THF 1:1 (v/v). The reaction mixture was shaken for 180 min at room temperature. The resin was washed with DMF, THF and DCM to yield **40**. A test cleavage was done with 95% TFA / H_2O and the resulting solution analyzed with LC-MS to control the reaction completion. If necessary, the coupling was repeated once more.

Coupling of further amino-acids with SPPS.

42
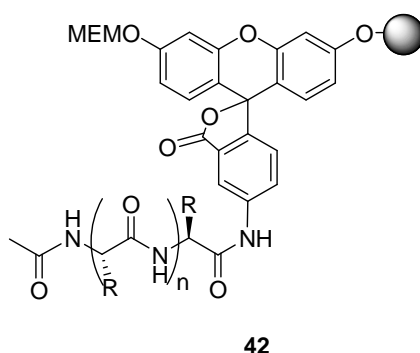
The Fmoc strategy was used to synthesize the peptides. Amino-acids with a Fmoc-protected amine function, a free carboxylic acid and side chain protecting groups resistant to the basic Fmoc removal procedure and acidic labile were used:

A: Fmoc-Ala; R: Fmoc-Arg(Pmc); N: Fmoc-Asn(Trt); D: Fmoc-Asp(O^{t}Bu); E: Fmoc-Glu(O^{t}Bu); G: Fmoc-Gly; L: Fmoc-Leu; K: Fmoc-Lys(Boc); F: Fmoc-Phe; P: Fmoc-Pro; S: Fmoc-Ser(^{t}Bu); W: Fmoc-Trp(Boc); V: Fmoc-Val.

From the previously obtained resin **40**, the Fmoc protecting group was removed with 20% piperidine in DMF. The next amino-acid (0.25 mmol) was coupled adding DIC (38.8 μL, 0.25 mmol) and HOBt (38.3 mg, 0.25 mmol) in DMF for 30 min. Completion of the reaction was controlled with the Kaiser test and the coupling repeated if necessary. The Fmoc group was cleaved, and the procedure repeated until the required sequence was obtained (**41**). The last amino-acid was acetylated with acetic anhydride (0.2 mL), DMF (1.4 mL) for 15 min to yield **42**.

Cleavage from the resin with side-chain deprotection and peptide recovery.

43
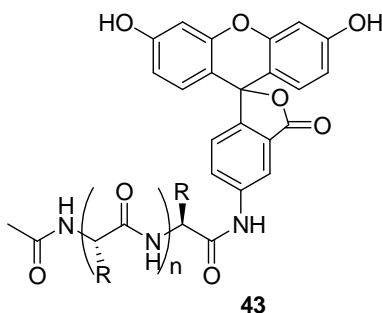
The peptide was cleaved off the resin and side-chains as well as fluorescein were deprotected by treatment with 95% TFA in H_2O for 60 min, with addition of triisopropylsilane (2 %) and addition of phenol (5 %) for peptides containing a Pmc protecting group. The product **43** was precipitated by addition of the cleavage solution to cold diethyl ether (4℃). After centrifugation (4000 tr·min^{-1}, 5 min), the upper solution was removed from the vial, diethyl ether was added, the solid was washed 5 min in ultrasonic bath and the vial was centrifuged. The procedure was repeated once more and the remaining solid was dissolved in *tert*-BuOH / H_2O (1:1) (v/v) and lyophilized.

Purification of the peptides.

All peptides with a purity below 90% were purified on a reversed phase preparative HPLC (see Part VI.1.4).

VI.2.3. Synthesis of *N*-terminally fluorescein labeled peptides 44 with Fmoc SPPS.
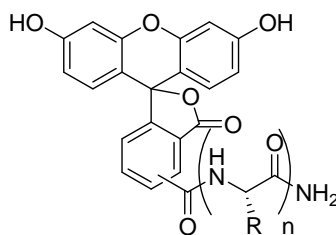
44

All *N*-terminally labeled peptides were synthesized by B. Schmikale and A. Klose (group of M. Beyermann, FMP, Berlin) on a Multisyntech Syro multiple peptide synthesizer and finally purified on a reversed phase preparative HPLC by D. Krause (group of M. Beyermann, FMP, Berlin).

VI.2.4. Synthesis of the library of *C*-terminally fluorescein labeled small molecules: FSML (Fluorescent Small Molecules Library)

Coupling of the first fragment (46).

The same procedure (DIC in THF / DCM) as for the coupling of amino-acids to the fluorescein resin was used (V.2.2), with non-natural amino acids, and resins 46 werw obtained. For all fragments, a single coupling step was sufficient to obtain a complete conversion.

Coupling of the second fragment (47).

Each resin **46** was divided in 50 mg (resp. 20 mg and 10 mg) amounts and coupled once more after Fmoc deprotection to various carboxylic acids with the previously described DIC / HOBt method to yield resins **47**.

Cleavage from the resin and product recovery.

The products were cleaved from the resin with 20% HFIP in DCM + 2% TIPS, the solvents were evaporated in vacuo, and the products lyophilized from H$_2$O / *tert-*

BuOH. The reactions were quantitative (as determined by weighing for a representative panel of molecules) and delivered about 10 mg (resp. 4 mg and 2 mg) product in high purity (> 90 %) that was checked with LC-MS analysis.

As a result, a library of 265 compounds labeled with fluorescein was obtained as pale to dark orange powders: VUP-FSML-001 to VUP-FSML-265. The HR-MS data of the library compounds are shown in the following table.

Storage of the library.

Small samples of each compound were precisely weighed and dissolved in DMSO to provide 10 mM stock solutions. 80 µL of each solution were pipetted into a 384-wells MTP with the following order: A1, B1, C1, …, P1, A2, B2, … . This plate was splitted into two parts with the pipetting roboter to deliver two duplicates VUP-FSML1-Plate_01 and VUP-FSML1-Plate_02 that were covered with an aluminium sheet and stored at -20℃.

All remaining solids were furthermore stored in glass vials at +4 ℃.

VI.2.5. Synthesis of the bromoacetic acid derivative of resin-bound aminofluorescein (35)

35
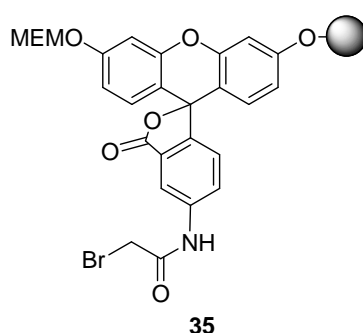
Coupling of bromoacetyl chloride.

Bromoacetyl chloride was added to a suspension of **32a** (100 mg, 0.05 mmol) in DCM / THF 1:1 (v/v) with DIPEA; the reaction mixture was shaken for at room temperature; the resin was washed with DMF, THF and DCM. The reaction progress

Compound	Formula	Mass (calcd)	Mass (exp)	Δ / ppm
VUP-FSML1-001	C31 H22 N2 O9	566.1325	566.1331	1.09
VUP-FSML1-002	C31 H24 N2 O9	568.1482	568.1488	1.07
VUP-FSML1-003	C32 H22 N2 O9	578.1325	578.1327	0.27
VUP-FSML1-004	C28 H21 N3 O7	511.1380	511.1386	1.23
VUP-FSML1-005	C27 H21 N3 O8	515.1329	515.1328	-0.22
VUP-FSML1-006	C28 H19 N3 O7	509.1223	509.1224	0.28
VUP-FSML1-007	C30 H21 F3 N2 O6	562.1352	562.1353	0.31
VUP-FSML1-008	C35 H24 N2 O7	584.1584	584.1582	-0.25
VUP-FSML1-009	C30 H22 N2 O9	554.1325	554.1329	0.75
VUP-FSML1-010	C30 H20 N4 O7	548.1332	548.1334	0.33
VUP-FSML1-011	C30 H30 N2 O8	546.2002	546.2005	0.57
VUP-FSML1-012	C27 H18 N2 O7 S	514.0835	514.0838	0.71
VUP-FSML1-013	C29 H20 N2 O7 S	540.0991	540.0992	0.22
VUP-FSML1-014	C28 H18 N2 O8 S	542.0784	542.0785	0.24
VUP-FSML1-015	C31 H20 N2 O7 S	564.0991	564.0995	0.75
VUP-FSML1-016	C33 H30 N2 O7	566.2053	566.2061	1.39
VUP-FSML1-017	C31 H24 N6 O9	624.1605	624.1610	0.80
VUP-FSML1-018	C30 H22 N2 O9 S	586.1046	586.1049	0.50
VUP-FSML1-019	C31 H20 N2 O8	548.1220	548.1227	1.32
VUP-FSML1-020	C30 H24 N2 O7	524.1584	524.1588	0.76
VUP-FSML1-021	C42 H30 N2 O7	674.2053	674.2063	1.46
VUP-FSML1-022	C36 H22 N2 O8	610.1376	610.1372	-0.65
VUP-FSML1-023	C46 H54 N2 O9	778.3829	778.3838	1.11
VUP-FSML1-024	C46 H48 N2 O10	788.3309	788.3317	1.01
VUP-FSML1-025	C36 H26 N2 O7	598.1740	598.1741	0.09
VUP-FSML1-026	C29 H20 N2 O8	524.1220	524.1222	0.38
VUP-FSML1-027	C36 H24 N2 O8	612.1533	612.1541	1.37
VUP-FSML1-028	C34 H24 N2 O7	572.1584	572.1591	1.23
VUP-FSML1-029	C28 H24 N2 O7	500.1584	500.1589	1.15
VUP-FSML1-030	C42 H29 N3 O9	719.1904	719.1918	2.04
VUP-FSML1-031	C37 H27 N3 O11	689.1646	689.1653	1.08
VUP-FSML1-032	C36 H24 N6 O9	684.1605	684.1607	0.32
VUP-FSML1-033	C34 H23 N3 O9 S	649.1155	649.1164	1.38
VUP-FSML1-034	C39 H27 N3 O11	713.1646	713.1649	0.42
VUP-FSML1-035	C28 H20 N2 O6	480.1321	480.1322	0.19
VUP-FSML1-036	C34 H26 N4 O10	650.1649	650.1647	-0.34
VUP-FSML1-037	C35 H24 N4 O9	644.1543	644.1548	0.67
VUP-FSML1-038	C30 H24 N2 O6	508.1634	508.1638	0.78
VUP-FSML1-039	C36 H25 N3 O9 S	675.1312	675.1311	-0.10
VUP-FSML1-040	C35 H23 N3 O10 S	677.1104	677.1109	0.66
VUP-FSML1-041	C38 H25 N3 O9 S	699.1312	699.1312	0.11
VUP-FSML1-042	C40 H35 N3 O9	701.2373	701.2378	0.64

Compound	Formula	Mass (calcd)	Mass (exp)	Δ / ppm
VUP-FSML1-043	C38 H29 N7 O11	759.1925	759.1926	0.19
VUP-FSML1-044	C37 H27 N3 O11 S	721.1366	721.1369	0.31
VUP-FSML1-045	C38 H25 N3 O10	683.1540	683.1543	0.43
VUP-FSML1-046	C37 H29 N3 O9	659.1904	659.1907	0.53
VUP-FSML1-047	C49 H35 N3 O9	809.2373	809.2359	-1.80
VUP-FSML1-048	C43 H27 N3 O10	745.1696	745.1693	-0.51
VUP-FSML1-049	C53 H53 N3 O12	923.3629	923.3634	0.49
VUP-FSML1-050	C29 H21 F N2 O6	512.1384	512.1385	0.33
VUP-FSML1-051	C41 H29 N3 O9	707.1904	707.1907	0.46
VUP-FSML1-052	C36 H25 N3 O10	659.1540	659.1541	0.19
VUP-FSML1-053	C35 H29 N3 O9	635.1904	635.1910	0.91
VUP-FSML1-054	C38 H29 N3 O11	703.1802	703.1808	0.89
VUP-FSML1-055	C43 H31 N3 O9	733.2060	733.2061	0.05
VUP-FSML1-056	C38 H27 N3 O11	701.1646	701.1652	0.97
VUP-FSML1-057	C27 H20 N2 O6 S	500.1042	500.1044	0.37
VUP-FSML1-058	C30 H28 N2 O7	528.1897	528.1905	1.57
VUP-FSML1-059	C34 H36 N2 O9	616.2421	616.2423	0.40
VUP-FSML1-060	C38 H33 F3 N2 O7	686.2240	686.2251	1.56
VUP-FSML1-061	C38 H42 N2 O9	670.2890	670.2895	0.67
VUP-FSML1-062	C38 H36 N2 O7	632.2523	632.2522	-0.04
VUP-FSML1-063	C35 H32 N2 O7 S	624.1930	624.1938	1.17
VUP-FSML1-064	C37 H33 F N2 O7	636.2272	636.2280	1.28
VUP-FSML1-065	C27 H19 N3 O7	497.1223	497.1228	1.01
VUP-FSML1-066	C31 H27 N3 O9	585.1747	585.1748	0.12
VUP-FSML1-067	C35 H24 F3 N3 O7	655.1566	655.1575	1.30
VUP-FSML1-068	C35 H33 N3 O9	639.2217	639.2219	0.42
VUP-FSML1-069	C35 H27 N3 O7	601.1849	601.1858	1.53
VUP-FSML1-070	C32 H23 N3 O7 S	593.1257	593.1268	1.96
VUP-FSML1-071	C34 H24 F N3 O7	605.1598	605.1611	2.12
VUP-FSML1-072	C33 H25 N3 O9	607.1591	607.1600	1.60
VUP-FSML1-073	C37 H33 N3 O11	695.2115	695.2122	1.03
VUP-FSML1-074	C41 H30 F3 N3 O9	765.1934	765.1949	1.96
VUP-FSML1-075	C41 H33 N3 O9	711.2217	711.2226	1.26
VUP-FSML1-076	C38 H29 N3 O9 S	703.1625	703.1631	0.99
VUP-FSML1-077	C40 H30 F N3 O9	715.1966	715.1981	2.05
VUP-FSML1-078	C42 H35 N3 O9	725.2373	725.2384	1.45
VUP-FSML1-079	C47 H45 N3 O9	795.3156	795.3174	2.27
VUP-FSML1-080	C38 H32 N2 O7	628.2210	628.2220	1.65
VUP-FSML1-081	C38 H34 N2 O7	630.2366	630.2375	1.47
VUP-FSML1-082	C39 H26 N4 O10	710.1649	710.1649	-0.04
VUP-FSML1-083	C44 H36 N4 O10	780.2431	780.2444	1.55
VUP-FSML1-084	C42 H28 N4 O8	716.1907	716.1895	-1.64

Compound	Formula	Mass (calcd)	Mass (exp)	Δ / ppm
VUP-FSML1-085	C34 H23 N3 O8 S	633.1206	633.1204	-0.31
VUP-FSML1-086	C35 H23 N3 O8	613.1485	613.1483	-0.37
VUP-FSML1-087	C35 H25 N3 O8	615.1642	615.1639	-0.47
VUP-FSML1-088	C28 H19 N3 O7	509.1223	509.1227	0.85
VUP-FSML1-089	C32 H27 N3 O9	597.1747	597.1753	1.01
VUP-FSML1-090	C36 H24 F3 N3 O7	667.1566	667.1574	1.10
VUP-FSML1-091	C29 H21 N3 O8	539.1329	539.1332	0.69
VUP-FSML1-092	C36 H27 N3 O7	613.1849	613.1854	0.75
VUP-FSML1-093	C33 H23 N3 O7 S	605.1257	605.1262	0.92
VUP-FSML1-094	C35 H24 F N3 O7	617.1598	617.1605	1.14
VUP-FSML1-095	C35 H24 N4 O9	644.1543	644.1550	1.05
VUP-FSML1-096	C40 H34 N4 O9	714.2326	714.2335	1.32
VUP-FSML1-097	C30 H21 N3 O7 S	567.1100	567.1106	1.10
VUP-FSML1-098	C31 H21 N3 O7	547.1380	547.1387	1.36
VUP-FSML1-100	C34 H31 N3 O8	609.2111	609.2123	1.86
VUP-FSML1-101	C39 H25 F3 N2 O7 S	722.1335	722.1352	2.48
VUP-FSML1-102	C39 H34 N2 O9 S	706.1985	706.2003	2.61
VUP-FSML1-103	C39 H28 N2 O7 S	668.1617	668.1633	2.37
VUP-FSML1-104	C36 H24 N2 O7 S2	660.1025	660.1031	0.99
VUP-FSML1-105	C38 H25 F N2 O7 S	672.1367	672.1377	1.54
VUP-FSML1-106	C45 H45 N3 O9	771.3156	771.3174	2.40
VUP-FSML1-107	C36 H32 N2 O7	604.2210	604.2220	1.76
VUP-FSML1-108	C36 H34 N2 O7	606.2366	606.2379	2.10
VUP-FSML1-109	C34 H30 N2 O11 S	674.1570	674.1582	1.75
VUP-FSML1-110	C38 H27 F3 N2 O9 S	744.1389	744.1404	1.94
VUP-FSML1-111	C38 H36 N2 O11 S	728.2040	728.2058	2.56
VUP-FSML1-112	C38 H30 N2 O9 S	690.1672	690.1690	2.54
VUP-FSML1-113	C35 H26 N2 O9 S2	682.1080	682.1092	1.85
VUP-FSML1-114	C37 H27 F N2 O9 S	694.1421	694.1439	2.49
VUP-FSML1-115	C43 H35 N3 O10	753.2322	753.2342	2.58
VUP-FSML1-116	C41 H27 N3 O8	689.1798	689.1801	0.47
VUP-FSML1-117	C33 H22 N2 O8 S	606.1097	606.1102	0.79
VUP-FSML1-118	C34 H22 N2 O8	586.1376	586.1393	2.85
VUP-FSML1-119	C34 H24 N2 O8	588.1533	588.1544	1.93
VUP-FSML1-120	C39 H29 F3 N6 O9	782.1948	782.1962	1.81
VUP-FSML1-121	C39 H38 N6 O11	766.2599	766.2608	1.26
VUP-FSML1-122	C39 H32 N6 O9	728.2231	728.2242	1.49
VUP-FSML1-123	C36 H28 N6 O9 S	720.1638	720.1640	0.18
VUP-FSML1-124	C38 H29 F N6 O9	732.1980	732.1993	1.84
VUP-FSML1-125	C34 H32 N2 O9	612.2108	612.2111	0.59
VUP-FSML1-126	C38 H29 F3 N2 O7	682.1927	682.1941	2.13
VUP-FSML1-127	C38 H38 N2 O9	666.2577	666.2585	1.19

Compound	Formula	Mass (calcd)	Mass (exp)	Δ / ppm
VUP-FSML1-128	C38 H32 N2 O7	628.2210	628.2221	1.82
VUP-FSML1-129	C35 H28 N2 O7 S	620.1617	620.1630	2.13
VUP-FSML1-130	C37 H29 F N2 O7	632.1959	632.1964	0.85
VUP-FSML1-131	C48 H37 N3 O10	815.2479	815.2484	0.58
VUP-FSML1-132	C46 H29 N3 O8	751.1955	751.1963	1.11
VUP-FSML1-133	C39 H24 N2 O8	648.1533	648.1539	1.01
VUP-FSML1-134	C39 H26 N2 O8	650.1689	650.1702	1.95
VUP-FSML1-140	C40 H32 N2 O10	700.2057	700.2057	-0.06
VUP-FSML1-141	C44 H29 F3 N2 O8	770.1876	770.1891	1.89
VUP-FSML1-142	C44 H38 N2 O10	754.2526	754.2516	-1.38
VUP-FSML1-143	C41 H28 N2 O8 S	708.1566	708.1573	0.95
VUP-FSML1-144	C43 H29 F N2 O8	720.1908	720.1922	1.89
VUP-FSML1-145	C40 H34 N2 O9	686.2264	686.2278	2.03
VUP-FSML1-146	C44 H31 F3 N2 O7	756.2083	756.2099	2.08
VUP-FSML1-147	C44 H40 N2 O9	740.2734	740.2726	-1.04
VUP-FSML1-148	C44 H34 N2 O7	702.2366	702.2380	2.02
VUP-FSML1-149	C48 H41 N3 O9	803.2843	803.2842	-0.11
VUP-FSML1-150	C43 H31 F N2 O7	706.2115	706.2129	1.99
VUP-FSML1-151	C38 H32 N2 O9	660.2108	660.2113	0.72
VUP-FSML1-152	C42 H29 F3 N2 O7	730.1927	730.1950	3.18
VUP-FSML1-153	C42 H38 N2 O9	714.2577	714.2571	-0.86
VUP-FSML1-154	C42 H32 N2 O7	676.2210	676.2213	0.46
VUP-FSML1-155	C39 H28 N2 O7 S	668.1617	668.1609	-1.21
VUP-FSML1-156	C41 H29 F N2 O7	680.1959	680.1957	-0.22
VUP-FSML1-157	C33 H28 N2 O10	612.1744	612.1708	-5.90
VUP-FSML1-158	C37 H25 F3 N2 O8	682.1563	682.1561	-0.31
VUP-FSML1-159	C37 H34 N2 O10	666.2213	666.2212	-0.27
VUP-FSML1-161	C34 H24 N2 O8 S	620.1253	620.1257	0.54
VUP-FSML1-162	C36 H25 F N2 O8	632.1595	632.1599	0.60
VUP-FSML1-163	C39 H33 N3 O9 S	719.1938	719.1945	1.06
VUP-FSML1-164	C37 H25 N3 O7 S	655.1413	655.1416	0.35
VUP-FSML1-165	C29 H20 N2 O7 S2	572.0712	572.0707	-0.85
VUP-FSML1-166	C30 H20 N2 O7 S	552.0991	552.0999	1.37
VUP-FSML1-167	C30 H22 N2 O7 S	554.1148	554.1156	1.53
VUP-FSML1-168	C41 H35 N3 O9 S	745.2094	745.2102	1.02
VUP-FSML1-169	C39 H27 N3 O7 S	681.1570	681.1563	-0.96
VUP-FSML1-170	C31 H22 N2 O7 S2	598.0868	598.0876	1.33
VUP-FSML1-171	C32 H22 N2 O7 S	578.1148	578.1150	0.41
VUP-FSML1-172	C32 H24 N2 O7 S	580.1304	580.1323	3.20
VUP-FSML1-173	C40 H39 N3 O9	705.2686	705.2691	0.70
VUP-FSML1-174	C31 H26 N2 O7	538.1740	538.1757	3.22
VUP-FSML1-175	C31 H28 N2 O7	540.1897	540.1905	1.48

Compound	Formula	Mass (calcd)	Mass (exp)	Δ / ppm
VUP-FSML1-176	C46 H38 N2 O9	762.2577	762.2591	1.82
VUP-FSML1-177	C50 H35 F3 N2 O7	832.2396	832.2414	2.13
VUP-FSML1-178	C50 H44 N2 O9	816.3047	816.3052	0.62
VUP-FSML1-179	C50 H38 N2 O7	778.2679	778.2677	-0.27
VUP-FSML1-180	C47 H34 N2 O7 S	770.2087	770.2097	1.28
VUP-FSML1-181	C49 H35 F N2 O7	782.2428	782.2424	-0.50
VUP-FSML1-182	C39 H32 N2 O9	672.2108	672.2118	1.58
VUP-FSML1-183	C43 H29 F3 N2 O7	742.1927	742.1945	2.43
VUP-FSML1-184	C43 H38 N2 O9	726.2577	726.2595	2.42
VUP-FSML1-185	C43 H32 N2 O7	688.2210	688.2221	1.68
VUP-FSML1-186	C40 H28 N2 O7 S	680.1617	680.1613	-0.67
VUP-FSML1-187	C42 H29 F N2 O7	692.1959	692.1957	-0.21
VUP-FSML1-188	C34 H30 N2 O11	642.1850	642.1865	2.35
VUP-FSML1-189	C38 H27 F3 N2 O9	712.1669	712.1680	1.63
VUP-FSML1-190	C38 H36 N2 O11	696.2319	696.2315	-0.60
VUP-FSML1-191	C38 H30 N2 O9	658.1951	658.1962	1.69
VUP-FSML1-192	C35 H26 N2 O9 S	650.1359	650.1367	1.20
VUP-FSML1-193	C37 H27 F N2 O9	662.1701	662.1704	0.45
VUP-FSML1-194	C37 H33 N5 O8	675.2329	675.2315	-2.02
VUP-FSML1-195	C41 H41 N5 O10	763.2853	763.2849	-0.53
VUP-FSML1-196	C45 H47 N5 O10	817.3323	817.3340	2.10
VUP-FSML1-197	C32 H23 N3 O6	545.1587	545.1591	0.81
VUP-FSML1-198	C42 H37 N5 O8 S	771.2363	771.2381	2.41
VUP-FSML1-199	C44 H38 F N5 O8	783.2704	783.2718	1.77
VUP-FSML1-200	C49 H48 N6 O10	880.3432	880.3442	1.14
VUP-FSML1-201	C40 H35 N5 O8	713.2486	713.2499	1.85
VUP-FSML1-202	C40 H37 N5 O8	715.2642	715.2659	2.38
VUP-FSML1-205	C35 H25 F3 N2 O9	674.1512	674.1525	1.85
VUP-FSML1-206	C34 H25 F N2 O9	624.1544	624.1526	-2.85
VUP-FSML1-208	C35 H30 N2 O11	654.1850	654.1852	0.37
VUP-FSML1-209	C39 H27 F3 N2 O9	724.1669	724.1675	0.89
VUP-FSML1-210	C39 H36 N2 O11	708.2319	708.2329	1.37
VUP-FSML1-211	C39 H30 N2 O9	670.1951	670.1963	1.74
VUP-FSML1-212	C36 H26 N2 O9 S	662.1359	662.1375	2.46
VUP-FSML1-213	C38 H27 F N2 O9	674.1701	674.1700	-0.03
VUP-FSML1-214	C35 H32 N2 O11	656.2006	656.2011	0.75
VUP-FSML1-215	C39 H29 F3 N2 O9	726.1825	726.1838	1.80
VUP-FSML1-216	C39 H38 N2 O11	710.2476	710.2496	2.83
VUP-FSML1-217	C39 H32 N2 O9	672.2108	672.2115	1.10
VUP-FSML1-218	C36 H28 N2 O9 S	664.1516	664.1534	2.73
VUP-FSML1-219	C38 H29 F N2 O9	676.1857	676.1861	0.51
VUP-FSML1-220	C36 H30 N2 O11	666.1850	666.1854	0.60

Compound	Formula	Mass (calcd)	Mass (exp)	Δ / ppm
VUP-FSML1-221	C40 H27 F3 N2 O9	736.1669	736.1681	1.65
VUP-FSML1-222	C40 H36 N2 O11	720.2319	720.2329	1.33
VUP-FSML1-223	C40 H30 N2 O9	682.1951	682.1956	0.70
VUP-FSML1-224	C37 H26 N2 O9 S	674.1359	674.1370	1.68
VUP-FSML1-225	C39 H27 F N2 O9	686.1701	686.1707	0.91
VUP-FSML1-226	C32 H29 N3 O9	599.1904	599.1905	0.13
VUP-FSML1-228	C30 H25 N3 O8	555.1642	555.1645	0.55
VUP-FSML1-229	C40 H34 N4 O9	714.2326	714.2333	0.97
VUP-FSML1-230	C38 H26 N4 O7	650.1801	650.1805	0.58
VUP-FSML1-231	C30 H21 N3 O7 S	567.1100	567.1103	0.41
VUP-FSML1-232	C31 H21 N3 O7	547.1380	547.1382	0.43
VUP-FSML1-233	C31 H23 N3 O7	549.1536	549.1537	0.27
VUP-FSML1-234	C48 H41 N3 O9	803.2843	803.2850	0.93
VUP-FSML1-235	C39 H28 N2 O7	636.1897	636.1901	0.73
VUP-FSML1-236	C39 H30 N2 O7	638.2053	638.2056	0.53
VUP-FSML1-237	C46 H39 N3 O9	777.2686	777.2691	0.57
VUP-FSML1-238	C36 H26 N2 O7 S	630.1461	630.1465	0.60
VUP-FSML1-239	C37 H26 N2 O7	610.1740	610.1749	1.53
VUP-FSML1-240	C37 H28 N2 O7	612.1897	612.1899	0.41
VUP-FSML1-241	C35 H26 F3 N3 O8	673.1672	673.1683	1.58
VUP-FSML1-242	C35 H35 N3 O10	657.2322	657.2326	0.54
VUP-FSML1-243	C35 H29 N3 O8	619.1955	619.1961	1.04
VUP-FSML1-244	C32 H25 N3 O8 S	611.1362	611.1367	0.71
VUP-FSML1-245	C34 H26 F N3 O8	623.1704	623.1714	1.55
VUP-FSML1-246	C32 H27 N3 O9	597.1747	597.1750	0.47
VUP-FSML1-247	C36 H24 F3 N3 O7	667.1566	667.1569	0.43
VUP-FSML1-248	C36 H33 N3 O9	651.2217	651.2225	1.21
VUP-FSML1-249	C36 H27 N3 O7	613.1849	613.1855	0.96
VUP-FSML1-250	C33 H23 N3 O7 S	605.1257	605.1261	0.65
VUP-FSML1-251	C35 H24 F N3 O7	617.1598	617.1601	0.45
VUP-FSML1-252	C31 H26 N2 O9 S	602.1359	602.1363	0.69
VUP-FSML1-253	C35 H23 F3 N2 O7 S	672.1178	672.1184	0.93
VUP-FSML1-254	C35 H32 N2 O9 S	656.1829	656.1829	0.13
VUP-FSML1-256	C33 H28 N2 O9 S	628.1516	628.1521	0.92
VUP-FSML1-257	C37 H25 F3 N2 O7 S	698.1335	698.1341	0.86
VUP-FSML1-258	C37 H34 N2 O9 S	682.1985	682.1987	0.28
VUP-FSML1-259	C37 H28 N2 O7 S	644.1617	644.1623	0.88
VUP-FSML1-260	C34 H24 N2 O7 S2	636.1025	636.1030	0.80
VUP-FSML1-261	C36 H25 F N2 O7 S	648.1367	648.1377	1.56
VUP-FSML1-262	C35 H30 N2 O7	590.2053	590.2061	1.44
VUP-FSML1-264	C25 H18 N2 O6	442.1165	442.1167	0.41
VUP-FSML1-265	C25 H20 N2 O6	444.1321	444.1324	0.57

was analyzed by an LC-MS of a sample obtained after cleavage of a small amount resin with 95% TFA / H$_2$O. The product was not obtained in a good purity; furthermore the conservation and handling of bromoacetyl chloride were quite difficult, so that the next method using bromoacetic acid was preferred.

Coupling of bromoacetic acid.

The same procedure as for the coupling of amino-acids to the fluorescein resin was used (V.2.2): bromoacetic acid (69.5 mg, 0.5 mmol) and DIC (77.6 µL, 0.5 mmol) were added to a suspension of **32a** (100 mg, 0.05 mmol) in DCM / THF 1:1 (v/v); the reaction mixture was shaken for 15 min at room temperature; the resin was washed with DMF, THF and DCM to yield **35**. The total conversion of the starting material into the product **35** was confirmed by an LC-MS analysis of a sample obtained after cleavage of a small amount resin with 95% TFA / H$_2$O.

VI.2.6. Nucleophilic substitution on the bromoacetic acid fluorescein derivative
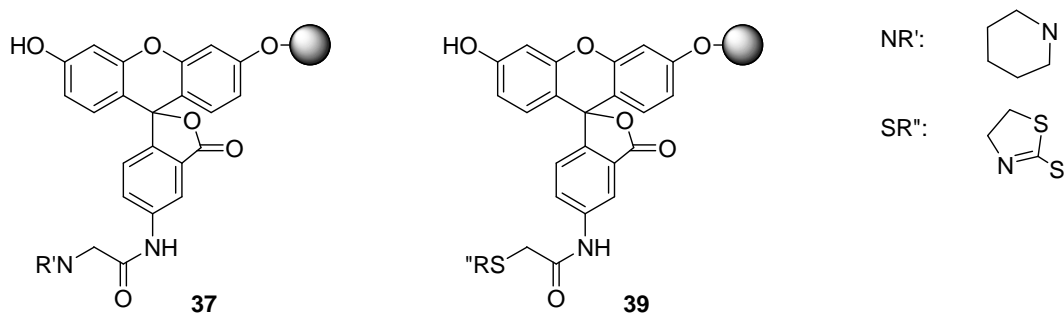
Substitution with thiols and amines.

Each thiol (respectively amine) (0.05 mmol) was dissolved in DCM / THF 1:1 (v/v) with addition of a little DMF if necessary and added to a suspension of resin **32a** (10 mg, 0.005 mmol) in DCM / THF and DIPEA (0.05 mmol). The reaction mixture was shaken for 180 min at room temperature. The resin was washed with DMF, THF and DCM. The resin cleavage was done with 20% HFIP in DCM + 2% TIPS, the solvents were

evaporated in vacuo, and the products lyophilized from H_2O / *tert*-BuOH. The resulting products were analyzed with LC-MS.

VI.3. Fluorescence polarization assays

VI.3.1. Proteins and peptides preparation

The proteins were provided by Michael Kofler and Daniela Kosslick from the group of Christian Freund, FMP, Berlin. The PERQ2 GYF and CD2BP2 GYF domains were expressed as GST fusion proteins *E coli* BL21 (DE3) cells harboring the plasmid pGEX 4T-1 with fragments of either PERQ2 (amino acids 531-596 of full-length PERQ2) or CD2BP2 (amino acids 280–341 of full-length CD2BP2). *E. coli* cells were grown in LB medium at 37℃ till A_{600} of 0.5, gene expression was induced with 1 mM IPTG, and the culture was harvested after 4 h of growth. GST fusion proteins were purified from the soluble fraction using affinity chromatography on a prepacked GSTrap™ HP glutathione-Sepharose column (GE Healthcare) and subsequent gel filtration (Superdex® 75, GE Healthcare) in phosphate buffered saline (pH 7.3). Final protein concentrations were in the range of 0.2–0.4 mM in PBS.

The synthesized peptides were used after HPLC purification, dissolved in DMSO at a concentration of 10 mM, then diluted with the assay buffer to the required concentration.

VI.3.2. Assay procedure for K_D determination

A solution was prepared for each fluorescein labeled peptide containing the peptide at a concentration of 30 nM and 0.3% Tween 20 in $1 \times$ PBS buffer pH 7.3. The assay plate was filled with 2 µL amounts of these solutions. Different solutions of the protein in $1 \times$ PBS buffer pH 7.3 were prepared to cover a concentration range from 0.001 µM to 200–400 µM. 4 µL were added to the plate previously prepared in the wells containing the peptide solution, so that the final assay composition was 10 nM fluorescein labeled peptide, 0.1 % Tween 20, in presence of growing protein

concentrations, in a total assay volume of 6 µL. For each protein concentration, the determination was done twice. After pipetting the solutions, the plate was briefly mixed, centrifuged and directly measured.

VI.3.3. Curves fitting and K_D determination

Data were plotted with the logarithm of the concentration on the x axis and the measured mP polarization value on the y axis. They were analyzed with the software Prism and fitted to a dose-response model with variable slope to determine the K_D values. These values are reported in Table 13 and Table 14, p. 87-87.

VI.3.4. Assay procedure for the displacement experiment

A solution of 4 µM CD2BP2-GYF protein (depending on the analyzed peptide, the concentration was chosen in the range of the peptide K_D value) with 12 nM fluorescein labeled peptide and 0.12% Tween 20 was prepared. On a 384-MTP, 1 µL of non-labeled peptide with the same sequence at different concentrations was pipetted. Then, 5 µL of the protein / fluorescent peptide solution was added to each well. The plate was briefly shaken, centrifuged and FP directly measured.

VI.3.5. Analysis of the displacement experiment

The data were plotted with the logarithm of the concentration on the x axis and the measured mP polarization values on the y axis. They were analyzed with the software Prism and fitted to a dose-response model with variable slope to determine the IC_{50} values.

These values enabled the determination of the K_D values using the Cheng-Prussoff equation [162] (Equation 19). Indeed, in this particular case, $[L_0] << K_D{}^* \Rightarrow [L^*] << K_D{}^* \Rightarrow [L^*] / K_D{}^* << 1 \Rightarrow IC_{50} \approx K_i$. The inhibition constant K_i obtained in this experiment is furthermore equal to the K_D value that would be determined in a direct binding assay [21], so that finally the measured IC_{50} value is equal to the dissociation constant K_D for the unlabeled peptide.

$$K_i = \frac{IC_{50}}{1 + \dfrac{[L^*]}{K_D^*}}$$

Equation 19: Cheng-Prussoff equation. Ki: inhibition constant of the unlabeled peptide; K_D^*: dissociation constant of the labeled peptide, [L*]: concentration of the ligand free in solution; IC_{50}: concentration at 50% inhibition.

VI.3.6.   Calculation of the Z' factor of a fluorescent probe

The quality of the fluorescent probe for HTS was determined by calculating the Z' factor (Equation 20):

$$Z' = 1 - \frac{3 \times (\sigma_{bound} + \sigma_{free})}{mP_{bound} - mP_{free}}$$

Equation 20: Determination of the Z' factor. σ_{bound}: standard deviation of polarization values for the peptide completely bound to the protein; σ_{free}: standard deviation of polarization values for the peptide completely free in solution; mP_{bound}: average fluorescence polarization value for the peptide completely bound to the protein; mP_{free}: average fluorescence polarization value for the peptide completely free in solution.

For the bound state, FP was measured for a solution of 10 nM fluorescein labeled peptide Fluo-EFGPPPGWLGR-NH$_2$ **44a** with 12 µM GYF-CD2BP2 in PBS buffer pH 7.3 and 0.1% Tween. To evaluate the free state, the mixture composition was the same with addition of the unlabeled peptide: Ac-EFGPPPGWLGR-NH$_2$ **45** at a final concentration of 100 µM. A Z' value of 0.74 was obtained.

VI.4. Protein NMR spectroscopy

The complete protein NMR measurements were done by D. Kosslick (group of C. Freund, FMP, Berlin). The untagged ^{15}N-labeled GYF domain of human CD2BP2 was cloned and expressed as described elsewhere [163]. The His$_6$-tagged GYF domain of PERQ2 was cloned as described by Kofler et al. [136]. The ^{15}N-labeled PERQ2 GYF domain was isolated from *Escherichia coli* BL21 (DE3) culture grown at 37 °C in M9

medium supplemented with $^{15}NH_4Cl$ until an A_{600} of 0.5 was reached. Gene expression was induced with IPTG (1 mM) for 4 h. Cells were disrupted by sonication, and the protein was purified from the soluble fraction by affinity chromatography on a prepacked HisTrap nickel-agarose column (GE Healthcare) and subsequent gel filtration. The NMR experiments were performed at 300 K on a Bruker DRX 600 instrument fitted with standard triple resonance probes. Data processing and analysis were carried out with the XWINNMR (Bruker) and SPARKY (SPARKY 3, version 3.1; T. D. Goddard, D. G. Kneller, University of California at San Francisco) software packages.

In the NMR titration experiments, increasing amounts of synthesized peptides were added to samples (0.1 mM) of the ^{15}N-labeled CD2BP2 or PERQ2 GYF domain. Sample temperature was 300 K, and PBS with DMSO (5 %) and D_2O (10 %) was used as buffer. The gradual changes in chemical shifts of several isolated peaks in the heteronuclear single quantum coherence spectra were used for K_D calculation (Figure 49, p.89). The chemical shift changes for ^{15}N and ^1H atoms in a sample with peptide were determined as $((\Delta^1H)^2+(\Delta^{15}N)^2)^{1/2}$, where ^1H is in units of 0.1 ppm and ^{15}N is in units of 0.5 ppm. For curve fitting and K_D calculation with Microcal^{TM} Origin^{TM} software, chemical shift changes were plotted versus the peptide concentration, and a simple two-state binding mode was assumed. The resulting values are summarized in Table 15.

VI.5. Fluorescence polarization competitive screening for CD2BP2-GYF domain ligands

VI.5.1. Screening procedure

The 16896 compounds of the FMP ChemBioNet small molecules library were stocked as 10 mM solutions in DMSO on 48 384-well microtitterplates. For an easy compound transfer and the direct use of the stock-solutions without previous dilution, the assay conditions had to be slightly adapted from the displacement experiment and the final

assay volume scaled-up from 6 to 10 µL. The plate preparation procedure is described on Figure 55 p. 94 and Figure 56 p. 95. The plates were subsequently briefly mixed, centrifuged and FP was directly measured.

VI.5.2. Analysis of the screening results

The ChemFinder database containing the screened compounds was converted into an sd-file, imported in Excel, where two columns were added: the measured FP and total intensity values were inserted. Finally, the obtained Excel document was exported as sd-file, imported in ChemFinder to create a new database containing the two measured values. This database enabled a much easier results analysis: the hits could be automatically sorted by choosing an interval for both the FP and the total intensity values, so that different parameter could be tried without tedious manual hits selection. Hits were chosen by looking at low mP values and total intensity values close to these of the control.

A displacement experiment was realized (see Parts V.3.4. and V.3.5.) and measured with FP. The molecules were also analyzed in LC-MS to check their inalterability. The commercial availability of the remaining molecules (10) was checked and 6 compounds could be reordered. The binding of these molecules to the CD2BP2-GYF domain was then checked in different experiments.

VI.5.3. Hits validation

VI.5.3.1. NMR spectroscopy.

To check the binding of the 6 reordered hits, a similar procedure to the peptides K_D determination was used, as described in Part VI.4.

VI.5.3.2. ITC titration.

For an ITC titration, three solutions were prepared. The reference cell was filled-up with degassed water. The reaction cell was filled-up with the protein solution: e.g. 400 µL of 50 µM CD2BP2-GYF (without GST tag) in $1 \times$ PBS buffer. The needle was filled-up with the ligand solution: e.g. 350 µL of 1 mM ligand in $1 \times$ PBS buffer.

As a control, to see the role of the dilution, a similar experiment was done, where the protein solution was replaced by a buffer solution. In case the ligand had to be first dissolved in DMSO, the same DMSO concentration was added in the protein solution. A second control was also done: injection of diluted DMSO into the protein solution.

Prior to starting the measurement, the internal temperature was stabilized. Then, every 6 to 8 min, 10 to 20 μL were injected, depending on the sample.

VI.5.3.3. CD spectroscopy.

First, the protein solution was analyzed: 200 μL of 20 μM CD2BP2-GYF (without GST tag). At 4℃, a complete absorption spectrum was me asured. The best wavelength was selected and the absorption at this wavelength was measured increasing the temperature from 4℃ to 95℃. Then, the temperature was decreased back to 4℃ and the absorption spectrum measured at the end of the experiment, to check the reversibility of the denaturation process.

A second experiment was done, adding ligand to the protein solution. Finally, both obtained unfolding temperatures were compared to check the ligand effect.

VI.6. Fluorescence polarization « direct » screening of the FSML

VI.6.1. Screening procedure

The stock-solutions of VUP-FSML1-Plate_01 were diluted using the pipetting robot to get nanomolar concentrations suitable for the assay, following the sequence described on Figure 79. Two assay plates were finally obtained: VUP-FSML1-Plate_01-C to measure the FP signal emitted by the labeled small molecules alone in solution, and VUP-FSML1-Plate_01-D to measure the eventual FP increase to due the binding of small molecules to the protein. FP was measured directly after pipetting and shaking each plate.

The following proteins were analyzed with the described procedure at a final assay concentration as listed in parentheses : GST-CD2BP2 GYF (100 μM, 38 μM and 18 μM) and GST (67 μM) provided by the group of C. Freund (FMP, Berlin); GST-AF

VUP-FSML1-Plate_01-A	50 µL H$_2$O + 0.03% Tween + 0.2 µL automatically transferred from VUP-FSML1-Plate_01
VUP-FSML1-Plate_01-B	50 µL H$_2$O + 0.03% Tween + 0.3 µL automatically transferred from VUP-FSML1-Plate_01-A
VUP-FSML1-Plate_01-C	5.8 µL PBS buffer + 0.1% Tween + 0.3 µL automatically transferred from VUP-FSML1-Plate_01-B
VUP-FSML1-Plate_01-D	1.8 µL PBS buffer + 0.3% Tween + 0.3 µL automatically transferred from VUP-FSML1-Plate_01-C + 4.0 µL protein solution, added manually

Figure 79: Preparation of assay plates by successive dilution steps

(33 µM), GST-Tudor_Y65N (67 µM), GST-MBT (12 µM), DAGK (67 µM), Proteinase K (67 µM), Lysozym (67 µM), BSA (67 µM), IK-domain (67 µM), RKIP (67 µM), VBC(53 µM) and Shank3 (67 µM), provided by the group of A. Diehl (FMP, Berlin).

VI.6.2. Analysis of the screening results

Analog to the analysis of the competitive screening (VI.5.2), the data were sorted with the Excel and Prism softwares, looking at the FP increase between small molecules free in solution and in the presence of protein.

### VI.6.3. Hits validation (for the CD2BP2 GYF domain)

Determination of FP binding curves.

Sixteen selected molecules were analyzed once more with FP: binding curves were determined, with the method already described in Part VI.3.2 and VI.3.3 for the labeled peptides. K$_D$ values in the micromolar range were found, with three values under 100 µM, as reported on Figure 77.
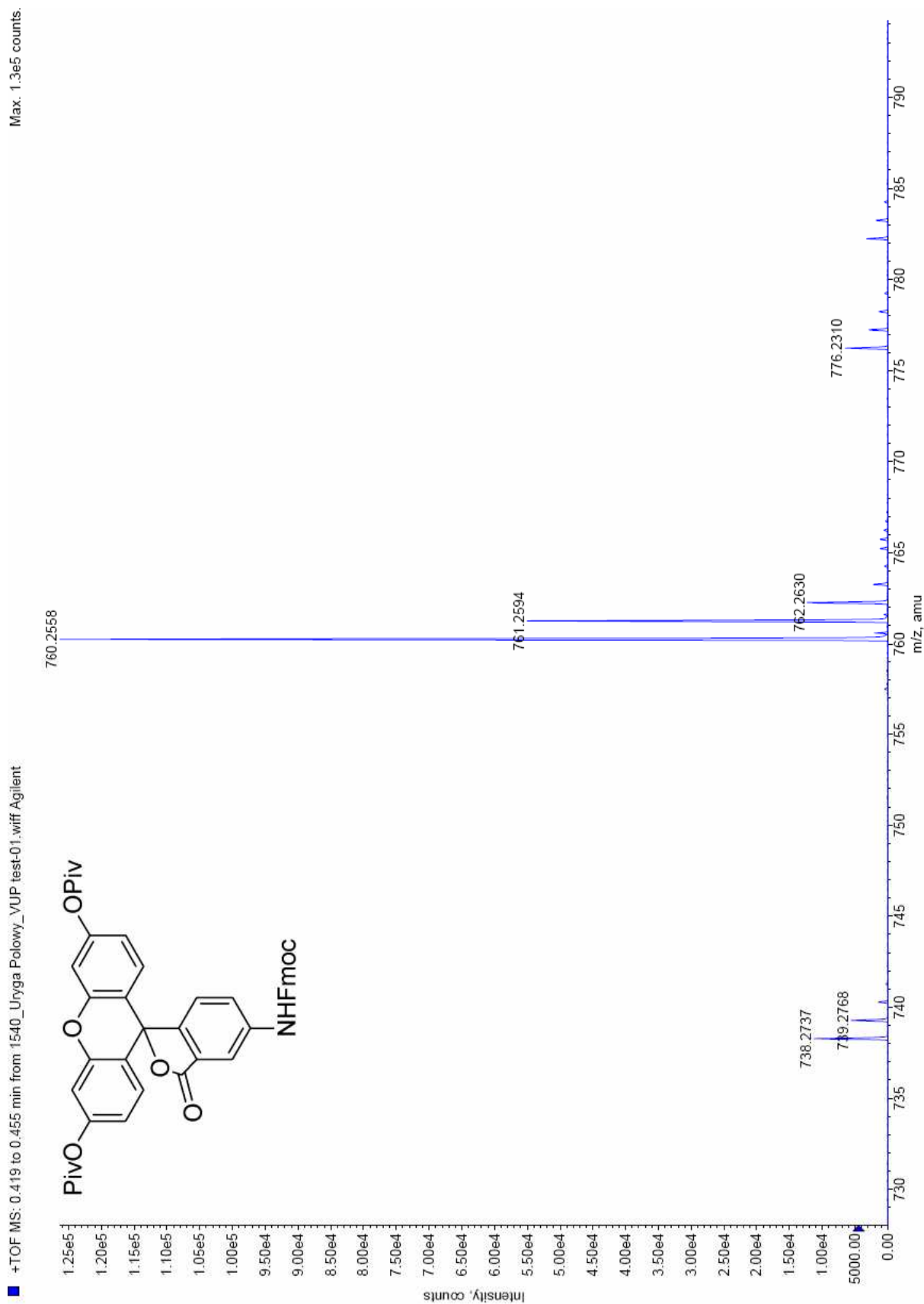
Confirmation of the binding by NMR.

The three best hits were further submitted to NMR binding experiments, as already described in Part VI.4.
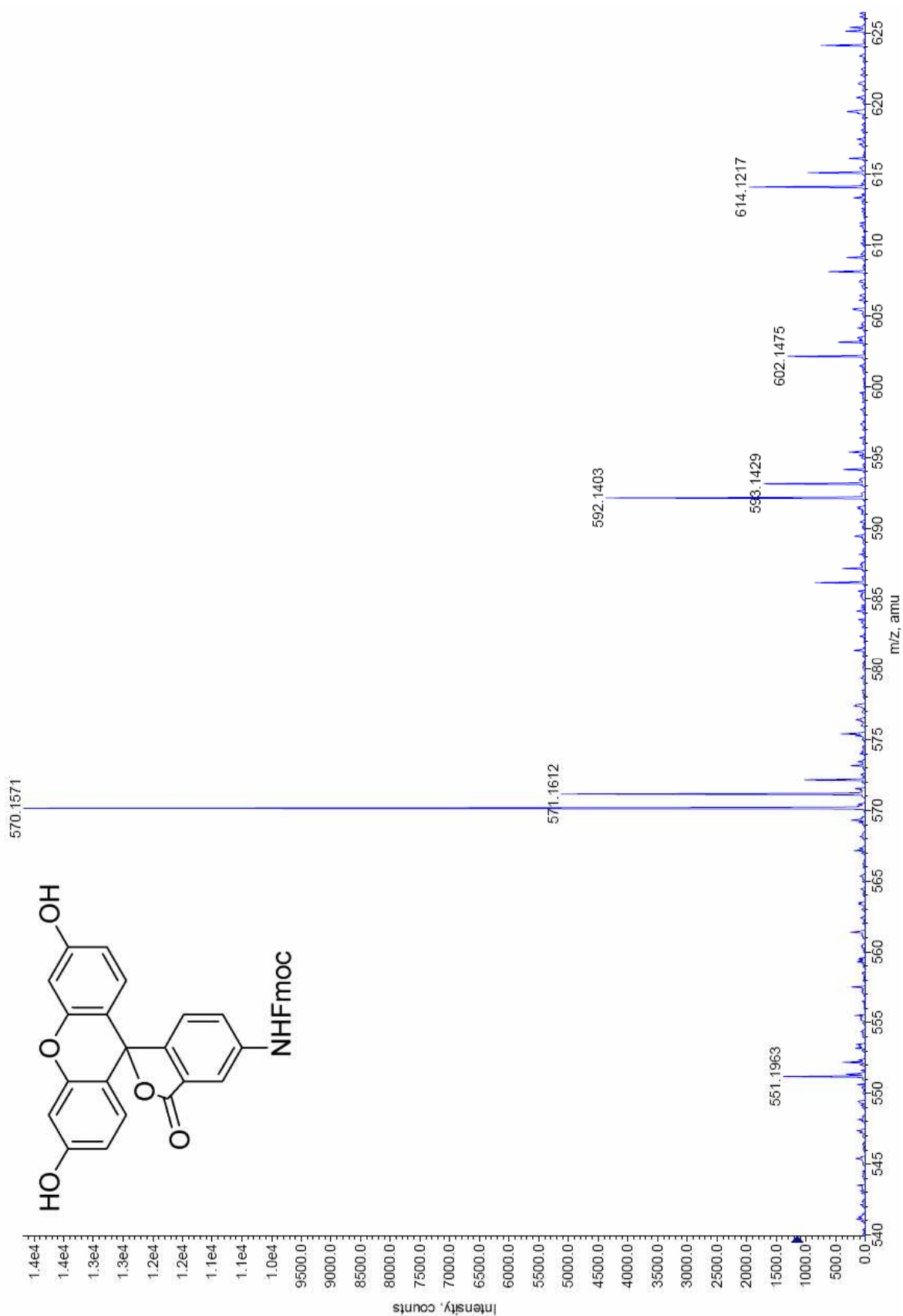
VII. ANNEX
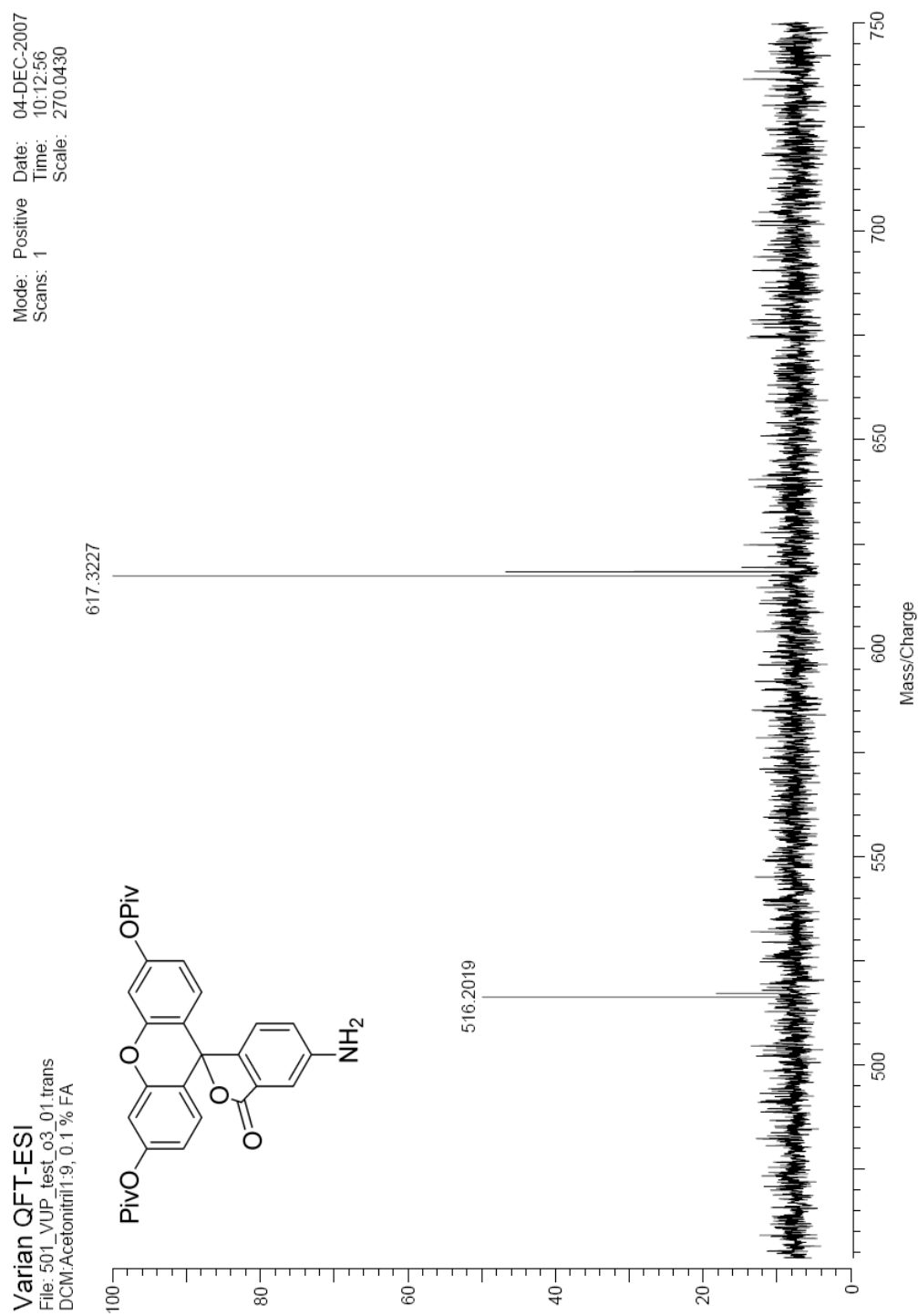
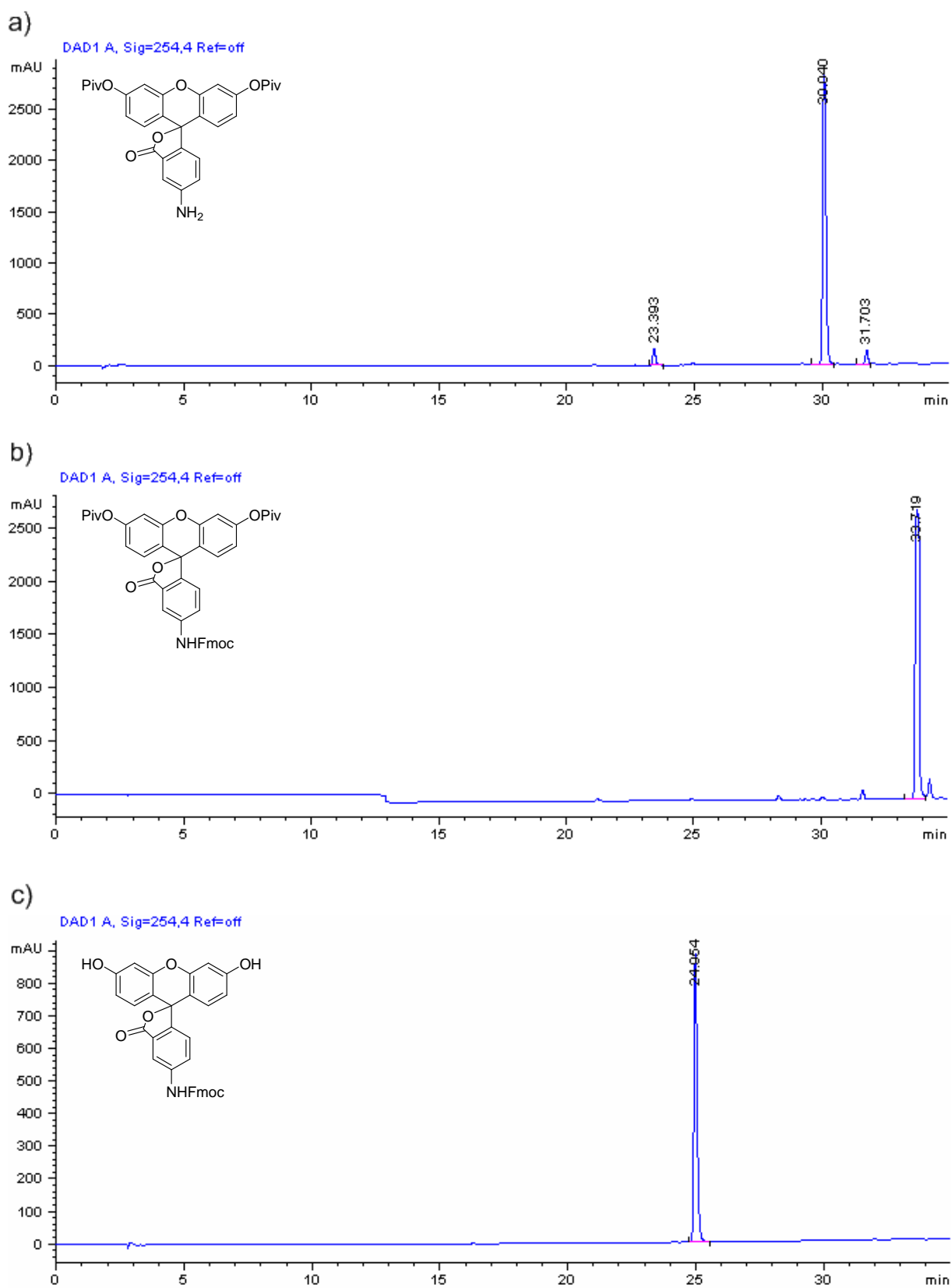**Annex - Fig. 1: HR-MS spectrum of di-*O*-pivaloyl-4-aminofluorescein (22)**

Annex - Fig. 2: HR-MS spectrum of di-*O*-pivaloyl-4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluorescein (23)

Annex - Fig. 3: HR-MS spectrum of 4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluo-rescein (21)

Annex - Fig. 4: HPLC spectra of (a): di-*O*-pivaloyl-4-aminofluorescein (22); (b): di-*O*-pivaloyl-4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluorescein (23); (c): 4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluorescein (21)
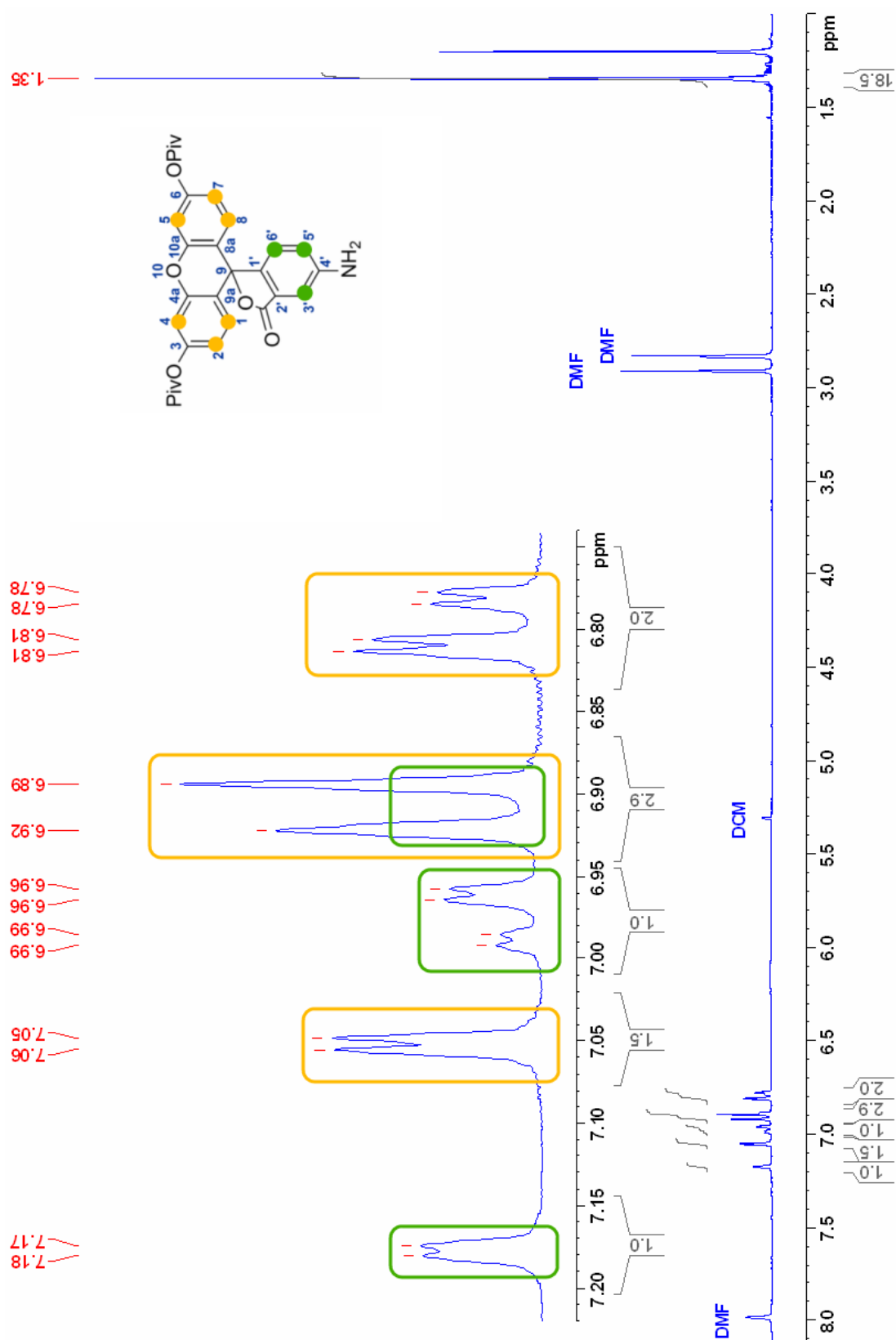
Annex - Fig. 5: VII.1.1.1. ^{1}H-NMR (300 MHz; CD2Cl2) of di-*O*-pivaloyl-4-aminofluorescein (22)
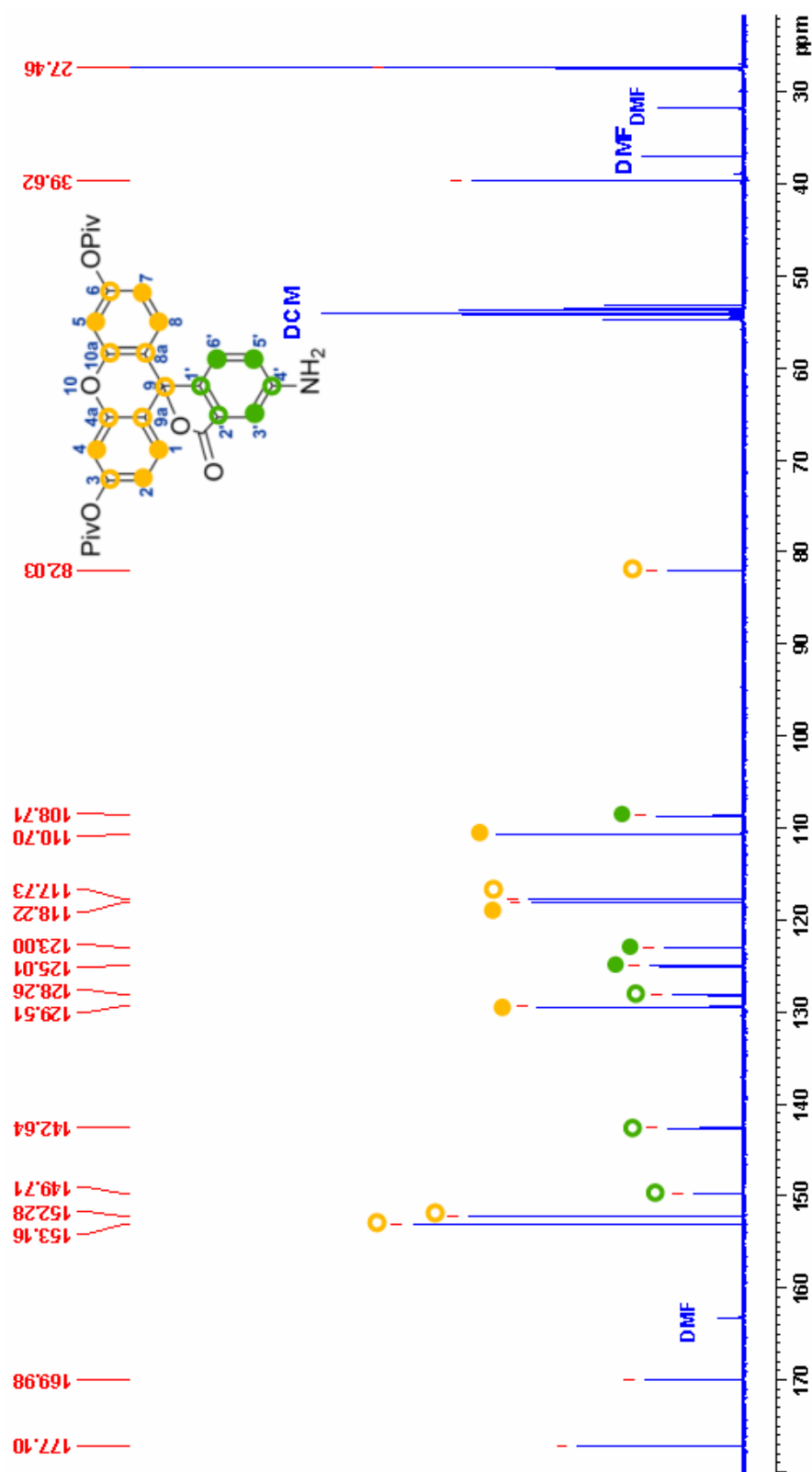
Annex - Fig. 6: ^{13}C-NMR (300 MHz; CD$_2$Cl$_2$) of di-*O*-pivaloyl-4-aminofluorescein (22)

Annex - Fig. 7: HMQC-NMR (300 MHz; CD$_2$Cl$_2$) of di-*O*-pivaloyl-4-aminofluorescein (22)
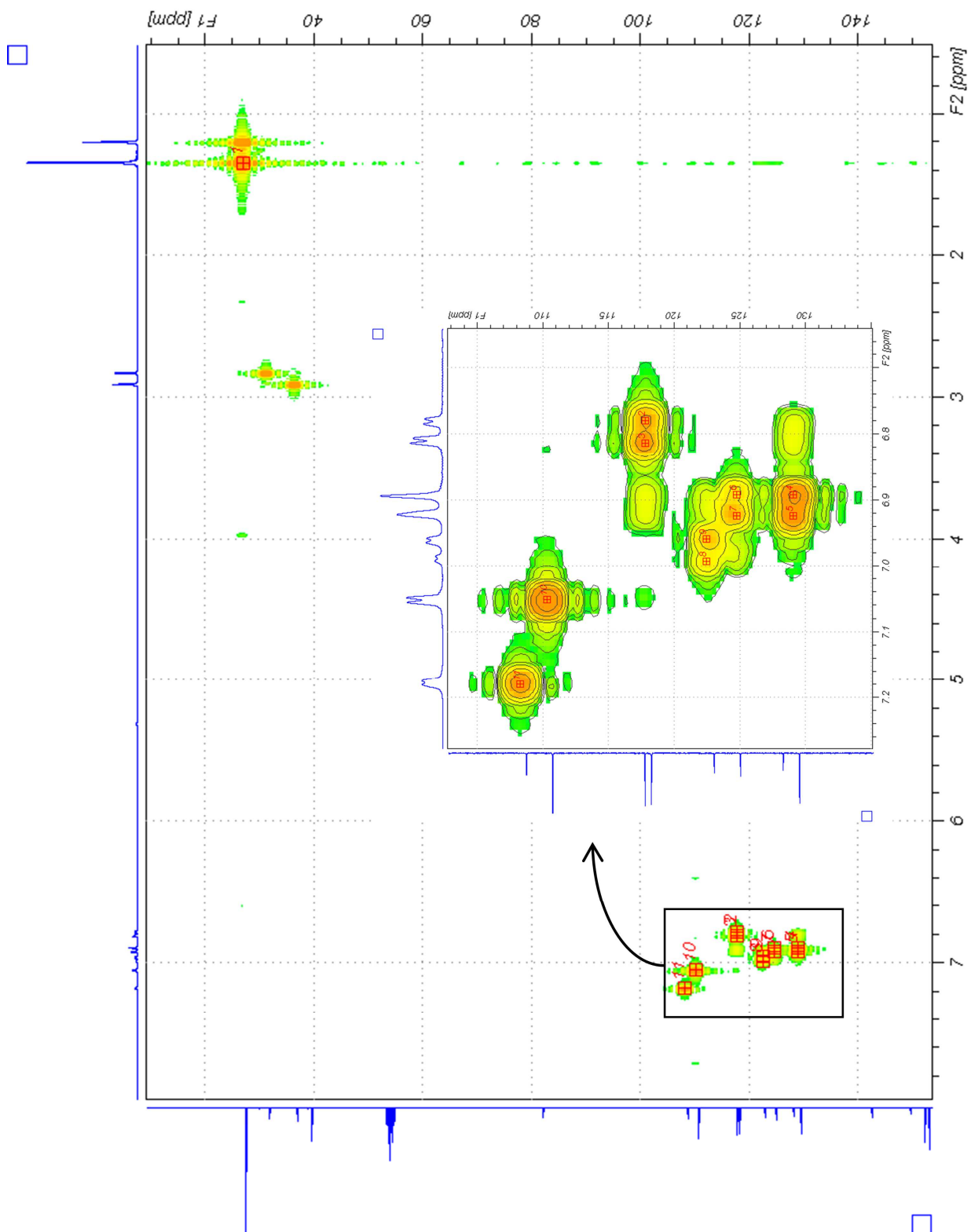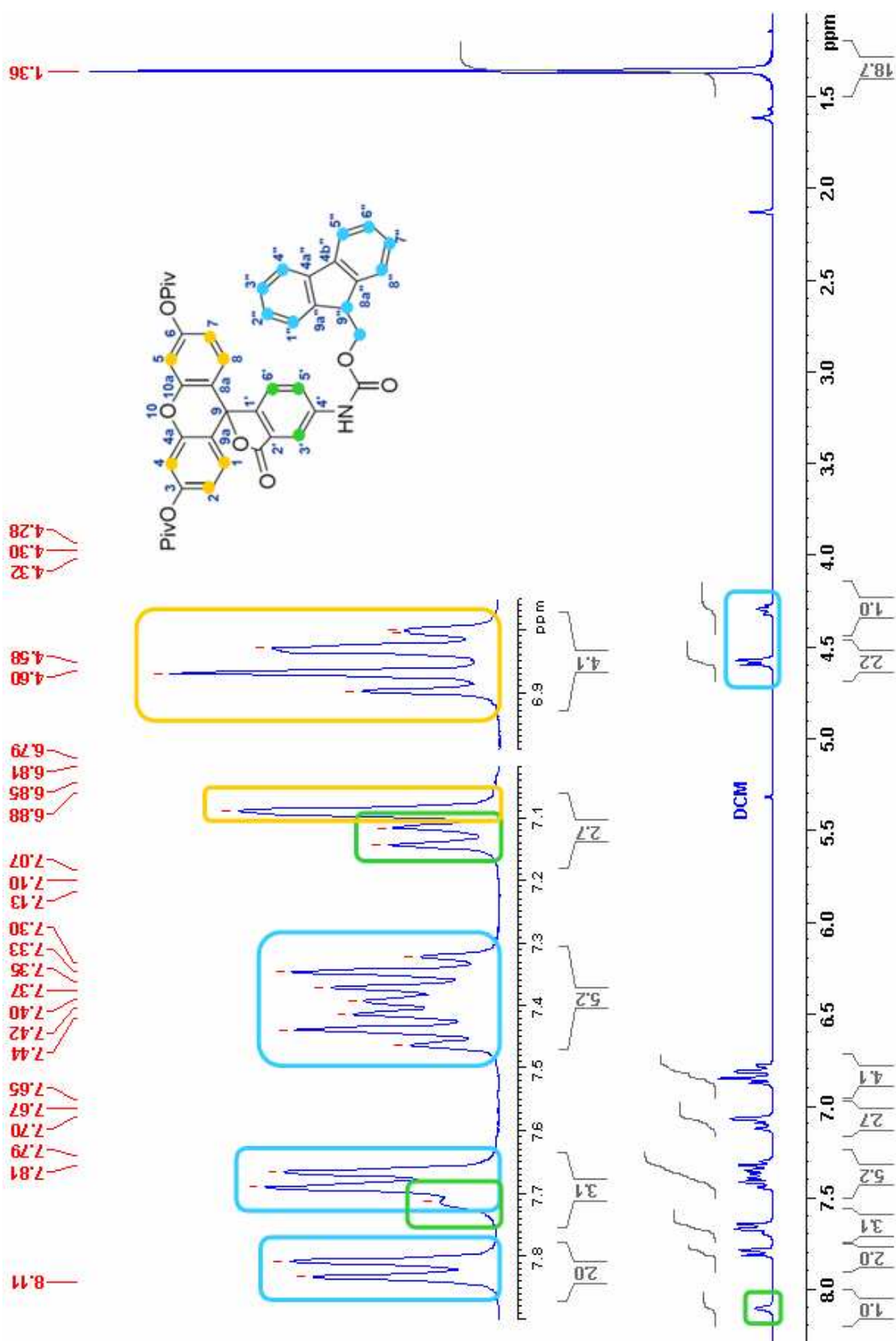
Annex - Fig. 8: ^{1}H-NMR (300 MHz; CD$_2$Cl$_2$) of di-*O*-pivaloyl-4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluorescein (23)
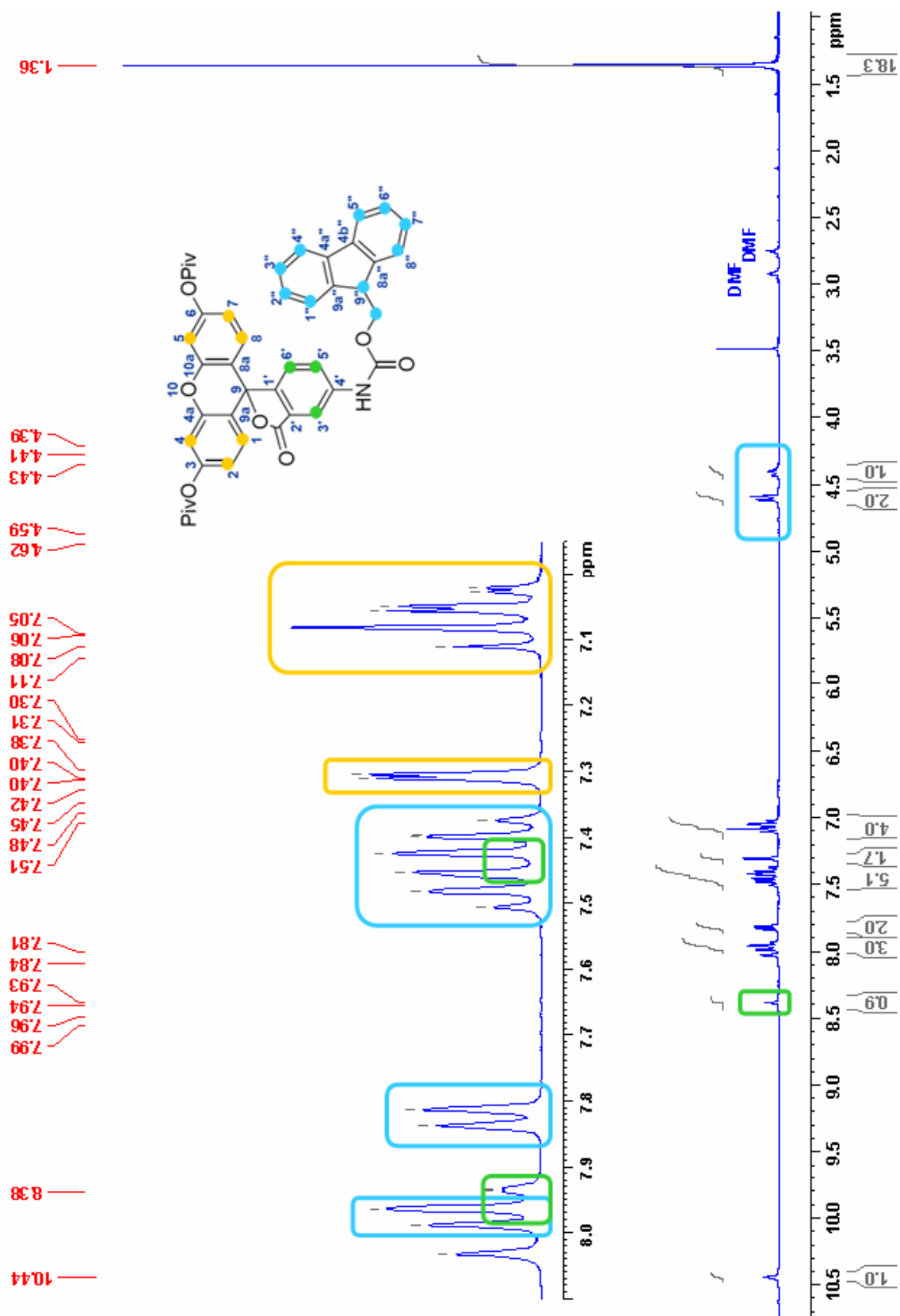
Annex - Fig. 9: ^{1}H-NMR (300 MHz; DMF-d7) of di-*O*-pivaloyl-4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluorescein (23)
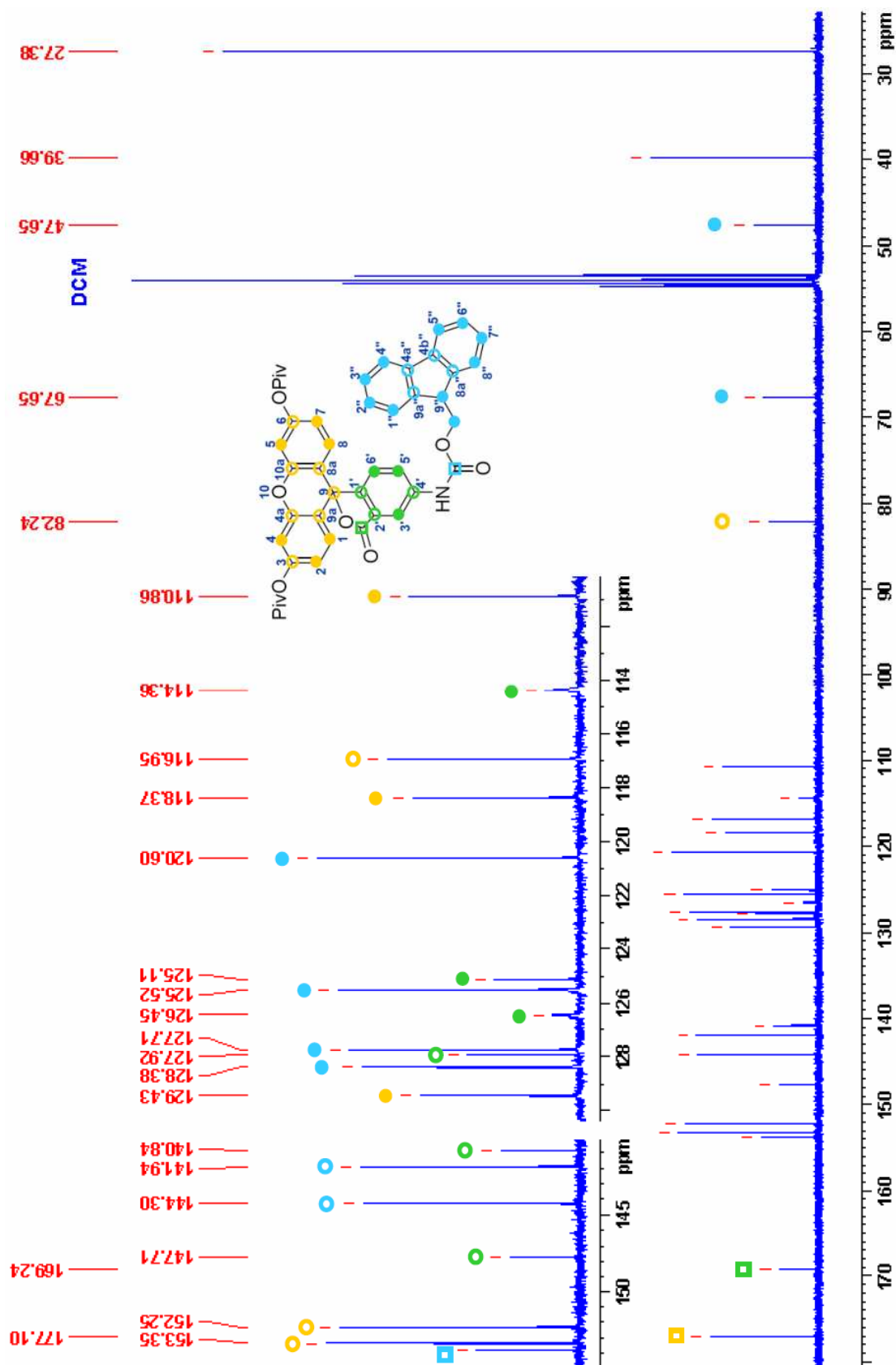
Annex - Fig. 10: ^{13}C-NMR (300 MHz; CD$_2$Cl$_2$) of di-*O*-pivaloyl-4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluorescein (23)
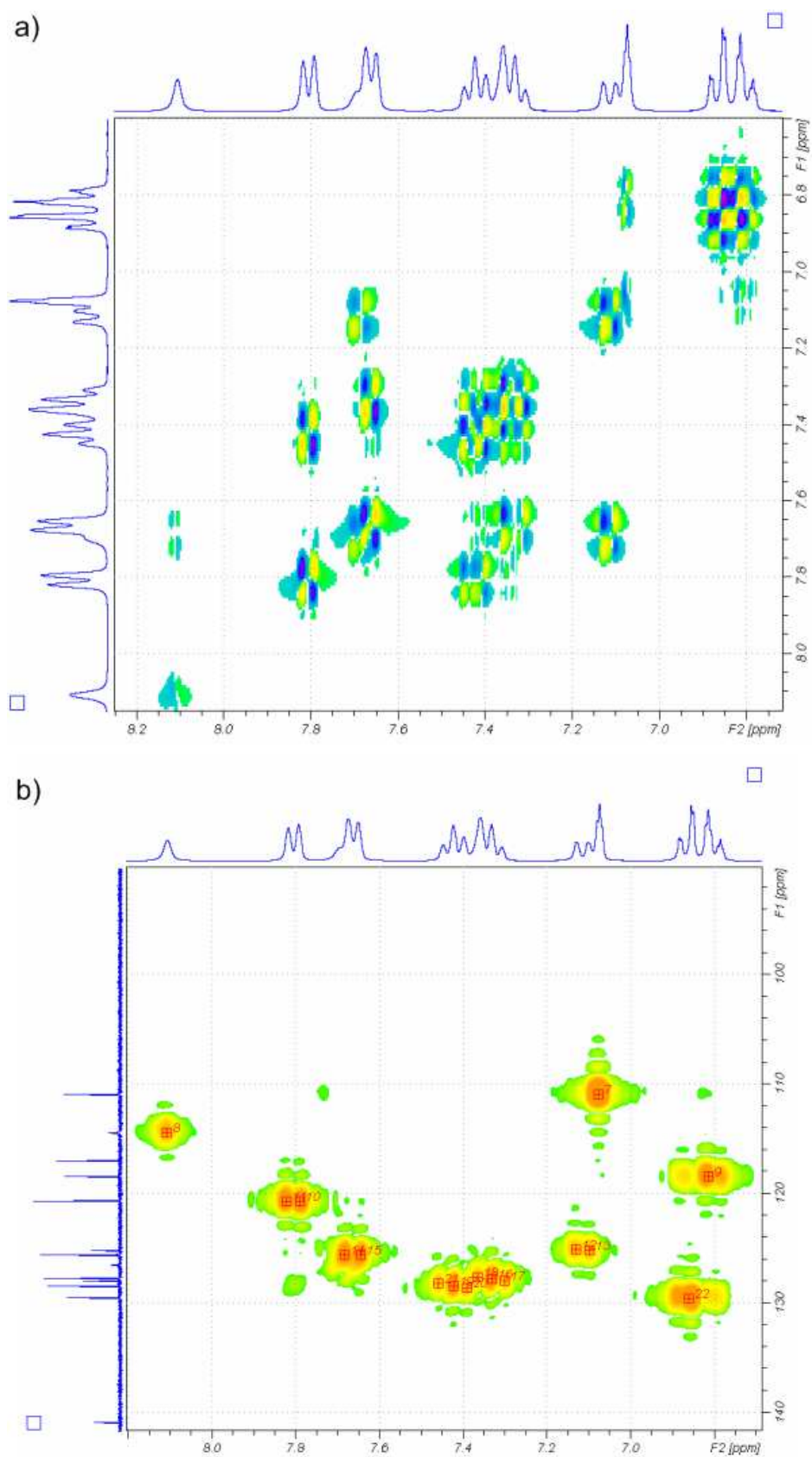
Annex - Fig. 11: DQF-COSY (a) and HMQC (b) NMR (300 MHz; CD$_2$Cl$_2$) of of di-*O*-pivaloyl-4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluorescein (23)
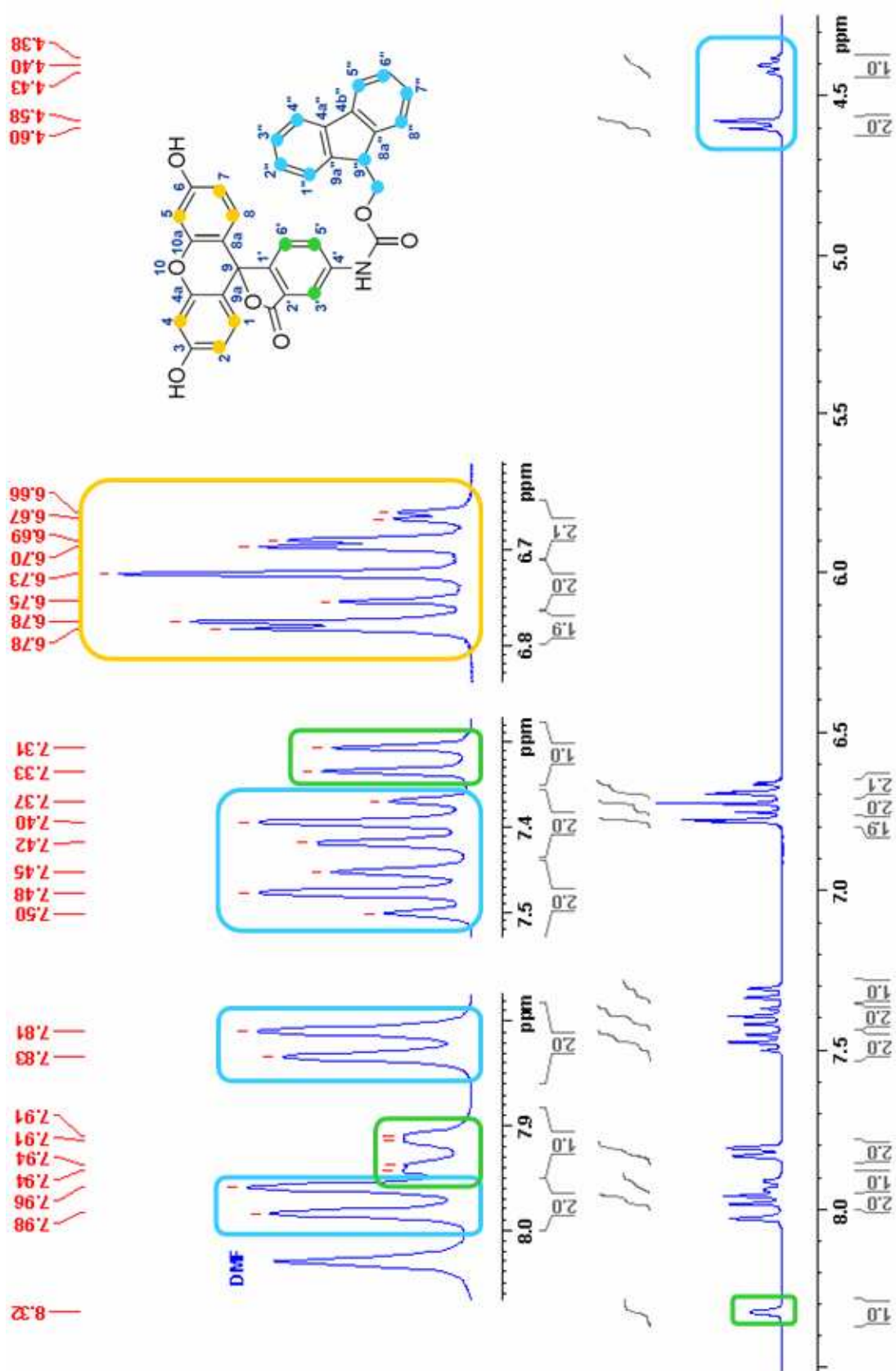
Annex - Fig. 12: ^{1}H-NMR (300 MHz; DMF-d7) of 4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluore-scein (21)

Annex - Fig. 13: ^{1}H-NMR (300 MHz; DMSO-d6) of 4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluore-scein (21)
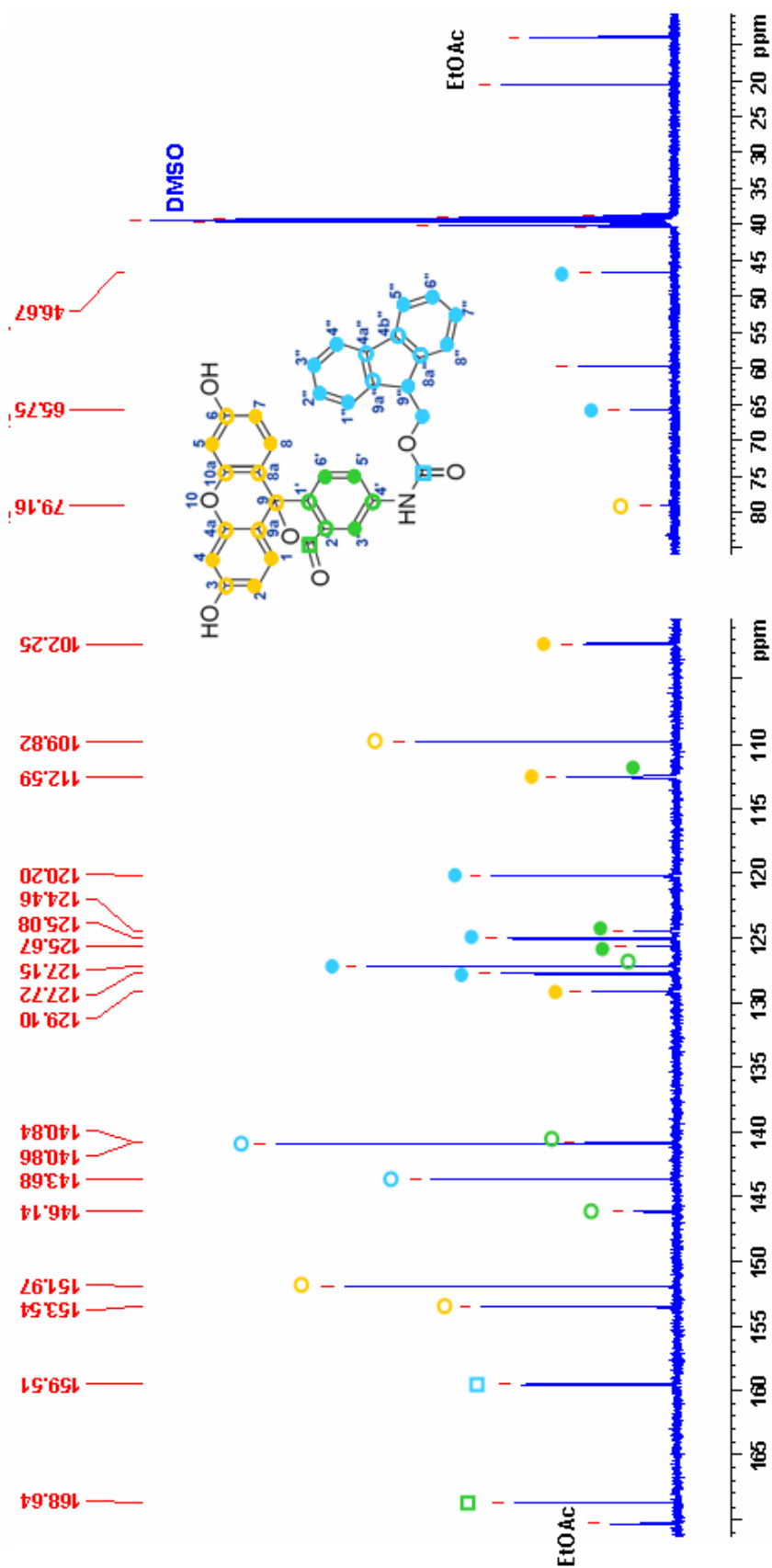
Annex - Fig. 14: ^{13}C-NMR (300 MHz; DMSO-d6) of 4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluore-scein (21)

Annex - Fig. 15: HMQC (300 MHz; DMSO-d6) of 4-(*N*-9-fluorenylmethoxycarbonyl)-aminofluorescein (21)
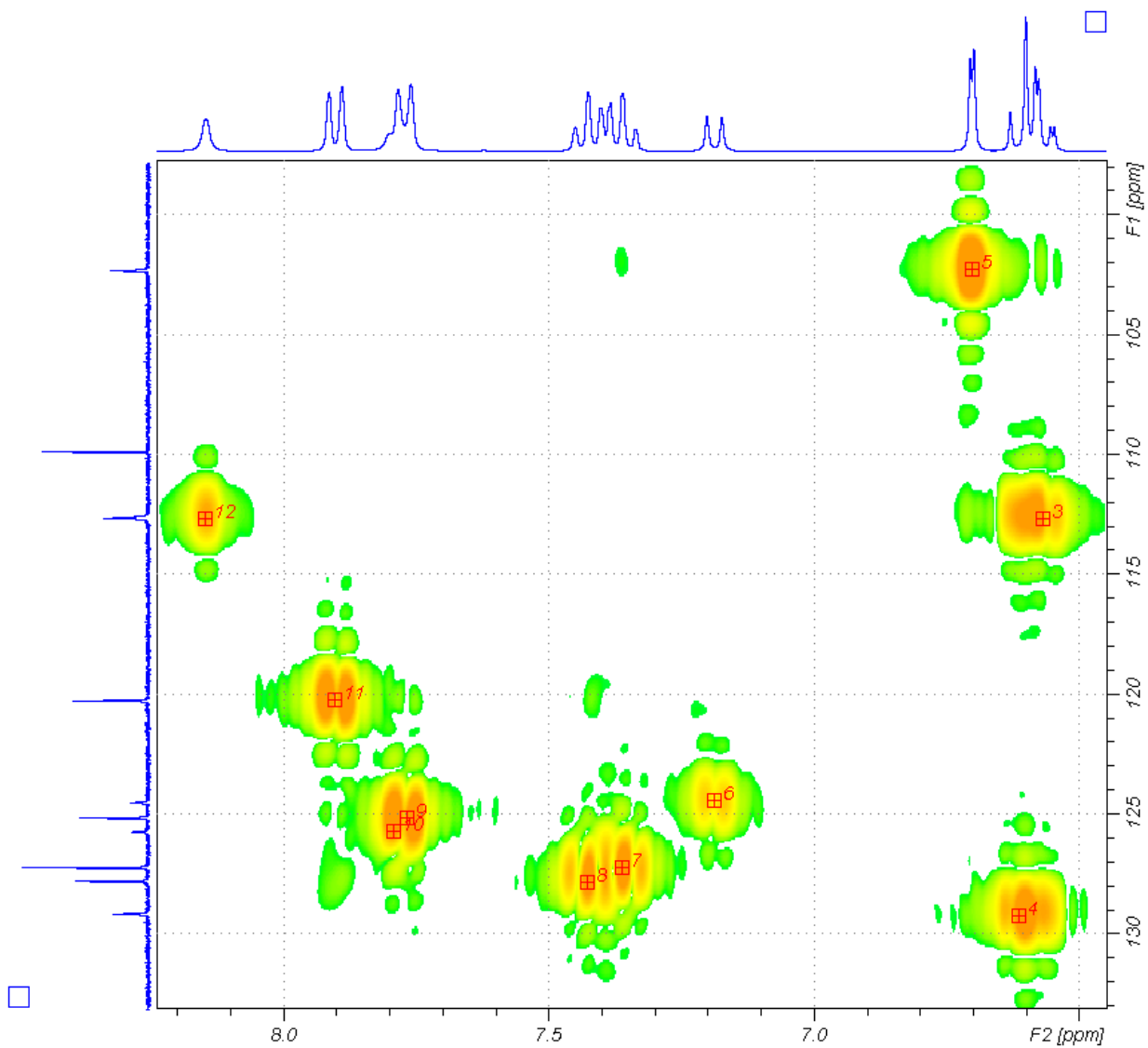
VIII. References

[1]   S. Riethmiller, *Chemotherapy* **2005**, *51*, 234–242.

[2]   F. Bosch, L. Rosich, *Pharmacology* **2008**, *82*, 171–179.

[3]   F. Zaragoza Dörwald, "General Techniques and Analytical Tools for Solid-Phase Organic Synthesis", in *Organic Synthesis on Solid Phase* **2000**, Wiley-VCH, p. 1–11.

[4]   C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25.

[5]   W. P. Walters, A. A. Murcko, M. A. Murcko, *Curr. Opin. Chem. Biol.* **1999**, *3*, 384–387.

[6]   C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26.

[7]   F. Cramer, "Emil Fischer's Lock and Key Hypothesis After 100 Years Towards a Supracellular Chemistry" , in *The lock and key principle, the state of the art 100 years on* **1994**, Ed. J-P Behr, p. 1–24.

[8]   H. R. Fana, W. G. Zhanga, *Asian J. Drug Metab. Pharmacokinet.* **2005**, *5(3)*, 163–172.

[9]   R. Frank, *Tetrahedron* **1992**, *48*, 9217–9232.

[10]   A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, *Int. J. Pept. Protein Res.* **1991**, *37*, 487–493.

[11]   K. S. Lam, M. Lebl, V. Krchnak, *Chem. Rev.* **1997**, *97*, 411–448.

[12]   A. Furka, F. Sebestyen, M. Asgedom, G. Dibo **1988**.

[13]   K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp, *Nature* **1991**, *354*, 82–84.

[14]   E. Jacobyl, J. Davies, M. J. J. Blommers, *Curr. Top. Med. Chem.* **2003**, *3*, 11–23.

[15]   G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1996**, *39*, 2887–2893.

[16]   G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1999**, *42*, 5095–5099.

[17]   R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.

[18]   G. R. Marshall, *J. Pept. Sci.* **2003**, *9*, 534–544.

[19]   R. B. Merrifield, *J. Am. Chem. Soc.* **1964**, *86*, 304–305.

[20]   R. B. Merrifield, *Biochemistry* **1964**, *3*, 1385–1390.

[21]   GlaxoWellcome **9999**.

[22]   K. Barlos, D. Gatos, J. Kallitsis, G. Papaphotiu, P. Sotiriu, Y. Wenqing, W. Schafer, *Tetrahedron Lett.* **1989**, *30*, 3943–3946.

[23]   R. Bollhagen, M. Schmiedberger, K. Barlos, E. Grell, *J. Chem. Soc. Chem. Commun.* **1994**, 2559–2560.

[24]   R. B. Merrifield, *Science* **1965**, *150*, 178–185.

[25]   G. R. Marshall, R. B. Merrifield, *Biochemistry* **1965**, *4*, 2394–2401.

[26]   L. A. Carpino, G. Y. Han, *J. Am. Chem. Soc.* **1970**, *92*, 5748–5749.

[27]   L. A. Carpino, G. Y. Han, *J. Org. Chem.* **1972**, *37*, 3404–3409.

[28]   E. Atherton, H. Fox, D. Harkiss, C. J. Logan, R. C. Sheppard, B. J. Williams, *Chem. Commun.* **1978**, 537–539.

[29]   W. Chan, *Fmoc Solid Phase Peptide Synthesis. A Practical Approach*, ed. W. C. Chan and P. D. White. **2000** Oxford University Press.

[30]   M. Beyermann, M. Bienert, H. Niedrich, L. A. Carpino, D. Sadat-Aalaee, *J. Org. Chem.* **1990**, *55*, 721–728.

[31]   L. A. Carpino, D. Sadat-Aalaee, H. G. Chao, R. H. DeSelms, *J. Am. Chem. Soc.* **1990**, *112*, 9651–9652.

[32]   H. Wenschuh, M. Beyermann, E. Krause, M. Brudel, R. Winter, M. Schuemann, L. A. Carpino, M. Bienert, *J. Org. Chem.* **1994**, *59*, 3275–3280.

[33]   J. C. Sheehan, G. P. Hess, *J. Am. Chem. Soc.* **1955**, *77*, 1067–1068.

[34]   F. Albericio, J. M. Bofill, A. El-Faham, S. A. Kates, *J. Org. Chem.* **1998**, *63*, 9678–9683.

[35]   F. Albericio, J. M. Bofill, A. El-Faham, S. A. Kates, *J. Org. Chem.* **1998**, *63*, 9678–9683.

[36]   L. A. Carpino, H. Imazumi, A. El-Faham, F. J. Ferrer, C. Zhang, C. Lee, B. M. Foxman, P. Henklein, C. Hanay, C. Mügge, H. Wenschuh, J. Klose, M. Beyermann, M. Bienert, *Angew. Chem.* **2002**, *114*, 457–461.

[37]   C. C. Leznoff, *Chem. Soc. Rev.* **1974**, *3*, 65–85.

[38]   H. J. Egelhaaf, J. Rademann, *J. Comb. Chem* **2005**, *7*, 929–941.

[39]   *J.Org.Chem.* **2001**.

[40]   J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*. Second Edition ed, K.A.P. Publishers. **1999**, New York 69.

[41]   http://www.olympusmicro.com/primer/java/jablonski/jabintro/index.html

[42]   M. Kasha, *Acta Physica Polonica A* **1999**, *95*, 15–35.

[43]   J. Lippincott-Schwartz, G. H. Patterson, *Science* **2003**, *300*, 87–91.

[44]   http://dwb.unl.edu/Teacher/NSF/C08/C08Links/pps99.cryst.bbk.ac.uk/projects/gmocz/gfp.htm

[45]   L. D. Lavis, R. T. Raines, *ACS Chem.Biol.* **2008**, *3*, 142–155.

[46]   T. Ueno, Y. Urano, K. Setsukinai, H. Takakusa, H. Kojima, K. Kikuchi, K. Ohkubo, S. Fukuzumi, T. Nagano, *J. Am. Chem. Soc.* **2004**, *126*, 14079–14085.

[47]   T. Miura, Y. Urano, K. Tanaka, T. Nagano, K. Ohkubo, S. Fukuzumi, *J. Am. Chem. Soc.* **2003**, *125*, 8666–8671.

[48]   Y. Urano, M. Kamiya, K. Kanda, T. Ueno, K. Hirose, T. Nagano, *J. Am. Chem. Soc.* **2005**, *127*, 4888–4894.

[49]   R. Pepperkok, J. Ellenberg, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 690–696.

[50]   P. Lang, K. Yeow, A. Nichols, A. Scheer, *Nat. Rev. Drug Discov.* **2006**, *5*, 343–356.

[51]   R. Yuste, *Nat. Methods* **2005**, 902–904.

[52]   J. W. Lichtman, J.-A. Conchello, *Nat. Methods* **2005**, 910–919.

[53]   P. R. Selvin, *Nat. Struct. Mol. Biol.* **2000**, *7*, 730–734.

[54]   T. Komatsu, K. Kikuchi, H. Takakusa, K. Hanaoka, T. Ueno, M. Kamiya, Y. Urano, T. Nagano, *J. Am. Chem. Soc.* **2006**, *128*, 15946–15947.

[55]   http://www.olympusfluoview.com/applications/fretintro.html

[56]   G. Weber, *Biochemistry* **1952**, *51*, 145–155.

[57]   G. J. Parker, T. L. Law, F. J. Lenoch, R. E. Bolger, *J. Biomol. Screen.* **2000**, *5*, 77–88.

[58]   S. T. Hess, S. Huang, A. A. Heikal, W. W. Webb, *Biochemistry* **2002**, *41*, 697–705.

[59]   M. Nakagawa, T. Yamagaki, H. Nakanishi, *J. Mass. Spectrom. Soc. Jpn.* **2003**, *51*, 554–558.

[60]   L. H. Thoresen, G.-S. Jiao, W. C.Haaland, M. L. Metzker, K. Burgess, *Chemistry - A European Journal* **2003**, *9*, 4603–4610.

[61]   M. Nakagawa, T. Yamagaki, H. Nakanishi, *J. Mass. Spectrom. Soc. Jpn.* **2003**, *51*, 554–558.

[62]   T. S. Seo, Z. Li, H. Ruparel, J. Ju, *J. Org. Chem.* **2003**, *68*, 609–612.

[63]   P. J. A. Weber, J. E. Bader, G. Folkers, A. G. Beck-Sickinger, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 597–600.

[64]   R. Fischer, O. Mader, G. Jung, R. Brock, *Bioconj. Chem.* **2003**, *14*, 653-660.

[65]   H. Spencer, *Popular Science* **1872**, p.583.

[66]   A. Rich, *Does RNA form a double helix?* in Inspiring Science: Jim Watson and the Age of DANN. CSHL Press ed. **2003**, p. 145–158.

[67]   J. L. Oncley, *Biophysical Science-A Study Program.* Wiley New York ed. **1959**.

[68]   A. O. W. Stretton, *Genetics* **2002**, *162*, 527–532.

[69]   http://www.pdb.org/pdb/home/home.do

[70]   A. J. Cozzon, *Ann. Rev. Microbiol.* **1988**, *42*, 97–125; D. Barford, A. K. Das, and M. P. Egloff, *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 133–164.

[71]   J. M. Berg, J. L. Tymoczko, L. Stryer, N. D. Clarke, *DNA Polymerases Require a Template and a Primer*, W.H. Freeman and Company ed. **2002**, Chapter 27, Section 2.

[72]   W. Yang, S. Xia, *World J. Gastroenterol.* **2006**, *12*, 7753–7757.

[73]   T. Pawson, P. Nash, *Science* **2003**, *300*, 445–452.

[74]   A. Abbott, *Nature* **2002**, *417*, 894–896.

[75]   T. Chouard, J. Finkelstein, *Nature* **2007**, *450*, 963–963.

[76]   E. M. Phizicky, S. Fields, *Microbiol. Rev.* **1995**, *59*, 94–123.

[77]   M. R. Arkin, J. A. Wells, *Nat. Rev. Drug Discov.* **2004**, *3*, 301–317.

[78]   D. D. Boehr, P. E. Wright, *Science* **2008**, *320*, 1429–1430.

[79]   D. C. Flynn, *Oncogene* **2001**, *20*, 6270–6272.

[80]   D. Anderson, C. A. Koch, L. Grey, C. Ellis, M. F. Moran, and T. Pawson, *Science* **1990**, *250*, 979–982.

[81]   M. F. Moran, C. A. Koch, D. Anderson, C. Ellis, L. England, G. S. Martin, and T. Pawson *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8622–8626.

[82]   J. Kuriyan, D. Cowburn, *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 259–288.

[83]   V. Neduva, R. Linding, I. Su-Angrand, A. Stark, F. de Masi, T. J. Gibson, J. Lewis, L. Serrano, R. B. Russell, PLoS. *Biol.* **2005**, *3(12)*, 2090–2099.

[83a] P. W. Faber, G. T. Barnes, J. Srinidhi, J. Chen, J. F. Gusella, M. E. Mac Donald, *Hum. Mol. Genet.* **1998**, *7*, 1463–1474.

[84] T. Pawson, J. D. Scott, *Science* **1997**, *278(5346)*, 2075–2080.

[85] E. P. Diamandis, T. K. Christopoulos, *Clin. Chem.* **1991**, *37*, 625–636.

[86] J. J. Tate, J. Persinger, B. Bartholomew, *Nucl. Acids Res.* **1998**, *26*, 1421–1426.

[87] I. Chen, M. Howarth, W. Lin, A. Y. Ting, *Nat. Methods* **2005**, *2*, 99–104.

[88] K. M. Marks, G. P. Nolan, *Nature Methods* **2006**, *3*, 591–596.

[89] A. J. Ozinskas, *Principles of Fluorescence Immunoassay*, in *Principles of Fluorescence Spectroscopy.* **1999**, Kluwer Academic / Plenum Publishers. p. 449–491.

[90] T. Hunter, *Cell* **2000**, *100*, 113-127.

[91] M. Kofler, K. Motzny, M. Beyermann, C. Freund, *J. Biol. Chem.* **2005**, *280*, 33397–3402.

[92] http://probes.invitrogen.com/handbook/boxes/1572.html

[93] K. Nectarios, H. S. William, *J. Fluoresc.* **1996**, *6*, 147–157.

[94] V. Zanker, W. Peter, *Chem. Ber.* **1958**, *91*, 572–580.

[95] J. Lohse, P. E. Nielsen, N. Harrit, O. Dahl, *Bioconjug. Chem.* **1997**, *8*, 503–509.

[96] M. Krol, M. Wrona, C. S. Page, P. A. Bates, *J. Chem. Theory Comput.* **2006**, *2*, 1520–1529.

[97] A. C. Spivey, A. Maddaford, *Annu. Rep. Prog. Chem., Sect. B*, **1999**, *95*, 83-95.

[98] M. Salunkhe, T. Wu, R. L. Letsinger, *J. Am. Chem. Soc.* **1992**, *114*, 8768-8772.

[99] T. Hanazawa, A. Koyama, K. Nakata, S. Okamoto, F. Sato, *J. Org. Chem.* **2003**, *68*, 9767–9772.

[100] M. J. Plunkett, J. A. Ellman, *J. Org. Chem.* **1997**, *62*, 2885–2893.

[101] M. D. Chappell, C. R. Harris, S. D. Kuduk, A. Balog, Z. Wu, F. Zhang, C. B. Lee, S. J. Stachel, S. J. Danishefsky, T. C. Chou, Y. Guan, *J. Org. Chem.* **2002**, *67*, 7730–7736.

[102] W. R. Roush, S. Russo-Rodriguez, *J. Org. Chem.* **1987**, *52*, 598–603.

[103] R. K. Kallury, U. J. Krull, M. Thompson, *J. Org. Chem.* **1987**, *52*, 5478–5480.

[104] K. B. Lindsay, S. G. Pyne, *J. Org. Chem.* **2002**, *67*, 7774–7780.

[105]   M. H. Lyttle, T. G. Carter, D. J. Dick, R. M. Cook, *J. Org. Chem.* **2000**, *65*, 9033–9038.

[106]   X. Li, J. S. Taylor, *Bioorg. Med. Chem.* **2004**, *12*, 545–552.

[107]   D. Yang, M. K. Wong, Z. Yan, *J. Org. Chem.* **2000**, *65*, 4179–4184.

[108]   T. R. Hoye, M. K. Renner, T. J. Vos-DiNardo, *J. Org. Chem.* **1997**, *62*, 4168–4170.

[109]   A. B. Smith, N. J. Liverton, N. J. Hrib, H. Sivaramakrishnan, K. Winzenberg, *J. Am. Chem. Soc.* **1986**, *108*, 3040–3048.

[110]   A. B. Smith, K. J. Hale, H. A. Vaccaro, R. A. Rivero, *J. Am. Chem. Soc.* **1991**, *113*, 2112–2122.

[111]   T. Hirano, K. Kikuchi, Y. Urano, T. Higuchi, T. Nagano, *J. Am. Chem. Soc.* **2000**, *122*, 12399–12400.

[112]   C. Zikos, E. Livaniou, L. Leondiadis, N. Ferderigos, D. S. Ithakissios, G. P. Evangelatos, *J. Pept. Sci.* **2003**, *9*, 419–429.

[113]   S. Mourtas, D. Gatos, K. Barlos, *Tetrahedron Lett.* **2003**, *42*, 2201–2204.

[114]   K. Barlos, D. Gatos, S. Kapolos, G. Papaphotiu, W. Schafer, Y. Wenqing, *Tetrahedron Lett.* **1989**, *30*, 3947–3950.

[115]   K. Barlos, O. Chatzi, D. Gatos, G. Stravropoulos, *Int. J. Pept. Protein Res.* **1991**, *37*, 513–520.

[116]   W. Q. Liu, M. Vidal, C. Olszowy, E. Million, C. Lenoir, H. Dhotel, C. Garbay, *J. Med. Chem.* **2004**, *47*, 1223–1233.

[117]   A. J. Moreno-Vargas, J. Jimenez-Barbero, I. Robina, *J. Org. Chem.* **2003**, *68*, 4138–4150.

[118]   D. Cabaret, J. Liu, M. Wakselman, R. F. Pratt, Y. Xub, *Bioorg. Med. Chem.* **1994**, *2*, 757–771.

[119]   D. Margulies, G. Melman, C. E. Felder, R. rad-Yellin, A. Shanzer, *J. Am. Chem. Soc.* **2004**, *126*, 15400–15401.

[120]   P. Nussbaumer, P. Lehr, A. Billich, *J. Med. Chem.* **2002**, *45*, 4310–4320.

[121]   W. A. Reckhow, D. S. Tarbell, *New compounds - Aryl Esters of Pivalic Acid* **1952**, *74*, 4968–4970.

[122]   M. Buback, T. Perkovi&cacute, S. Redlich, A. de Meijere, *European J. Org. Chem.* **2003**, 2375–2382.

[123]   C. Benard, F. Zouhiri, M. Normand-Bayle, M. Danet, D. Desmaele, H. Leh, J. F. Mouscadet, G. Mbemba, C. M. Thomas, S. Bonnenfant, M. Le Bret, J. d'Angelo, *Bioorg. Med. Chem Lett.* **2004**, *14*, 2473–2476.

[124] Z. Zhu, B. Mckittrick, *Tetrahedron Lett.* **1998**, *39*, 7479–7482.

[125] B. B. Shankar, D. Y. Yang, S. Girton, A. K. Ganguly, *Tetrahedron Lett.* **1998**, *39*, 2447–2448.

[126] C. A. Olsen, M. R. Jorgensen, S. H. Hansen, M. Witt, J. W. Jaroszewski, H. Franzyk, *Org. Lett.* **2005**, *7*, 1703–1706.

[127] M. Yamashita, K. Yamada, K. Tomioka, *J. Am. Chem. Soc.* **2004**, *126*, 1954–1955.

[128] R. M. Silverstein, F. X. Webster, **1998**, *6*.

[129] J. L. Harris, B. J. Backes, F. Leonetti, S. Mahrus, J. A. Ellman, C. S. Craik, *PNAS* **2000**, *97*, 7754–7759.

[130] R. R. Kale, C. M. Clancy, R. M. Vermillion, E. A. Johnson, S. S. Iyer, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2459–2464.

[131] K. Krowicki, J. W. Lown, *J. Org. Chem.* **1987**, *52*, 3493–3501.

[132] H. J. Olivos, P. G. Alluri, M. M. Reddy, D. Salony, T. Kodadek, *Org. Lett.* **2002**, *4*, 4057–4059.

[133] J. Schust, T. Berg, *Anal. Biochem.* **2004**, *330*, 114–118.

[134] B. K. Kay, M. P. Williamson, M. A. R. I. Sudol, *FASEB J.* **2000**, *14*, 231–241.

[135] B. Y. Feng, B. K. Shoichet, *Nat. Protocols* **2006**, *1*, 550–553.

[136] M. Kofler, K. Motzny, C. Freund, *Mol. Cell Proteomics* **2005**, *4*, 1797–1811.

[137] T. Wohr, F. Wahl, A. Nefzi, B. Rohwedder, T. Sato, X. Sun, M. Mutter, *J. Am. Chem. Soc.* **1996**, *118*, 9218–9227.

[138] T. Mineno, T. Ueno, Y. Urano, H. Kojima, T. Nagano, *Org. Lett.* **2006**, *8*, 5963–5966.

[139] K. Tanaka, T. Miura, N. Umezawa, Y. Urano, K. Kikuchi, T. Higuchi, T. Nagano, *J. Am. Chem. Soc.* **2001**, *123*, 2530–2536.

[140] C. Munkholm, D. R. Parkinson, D. R. Walt, *J. Am. Chem. Soc.* **1990**, *112*, 2608–2612.

[141] R. Buchli, R. S. VanGundy, H. D. Hickman-Miller, C. F. Giberson, W. Bardet, W. H. Hildebrand, *Biochemistry* **2004**, *43*, 14852–14863.

[142] P. Blommel, G. T. Hanson, and K. W. Voge, *J. Biomol. Screen.* **2004**, *9*, 294–302.

[143] L. Prystay, M. Gosselin, P. Banks, *J. Biomol. Screen.* **2001**, *6*, 141–150.

[144] R. Buchli, R. S. VanGundy, H. D. Hickman-Miller, C. F. Giberson, W. Bardet, and W. H. Hildebrand, *Biochemistry* **2005**, *44*, 12491–12507.

[145] M. H. A. Roehrl, J. Y. Wang, G. Wagner, *Biochemistry* **2004**, *43*, 16067–16075.

[146] M. Kofler, C. Freund, *FEBS J.* **2006**, *273*, 245–256.

[147] C. Freund, V. Dotsch, K. Nishizawa, E. L. Reinherz, G. Wagner, *Nat. Struct. Mol. Biol.* **1999**, *6*, 656–660.

[148] C. Freund, R. Kühne, H. Yang, S. Park, E. L. Reinherz, G. Wagner, *EMBO J.* **2002**, 5958–5995.

[149] G. Dorman, G. D. Prestwich, *Biochemistry* **1994**, *33*, 5661–5673.

[150] P. A. Wright, H. F.Boyd, R. C. Bethell, M. Busch, P. Gribbon, J. Kraemer, E. Lopez-Calle, T. H. Mander, D. Winkler, N. Benson, *J. Biomol. Screen.* **2002**, *7*, 419–428.

[151] W. P. Walters, M. Namchuk, *Nat. Rev. Drug Discov.* **2003**, *2*, 259–266.

[152] J. C. Owicki, *J. Biomol. Screen.* **2000**, *5*, 297–306.

[153] S. Turconi, K. Shea, S. Ashman, K. Fantom, D. L. Earnshaw, R. P. Bingham, U. M. Haupts, M. J. B. Brown, A. J. Pope, *J. Biomol. Screen.* **2001**, *6*, 275–290.

[154] **2002**, *poster pharmacopeia*.

[155] *Nat. Protoc.* **2006**.

[156] I. Muckenschnabel, R. Falchetto, L. M. Mayr, I. Filipuzzi, *Anal. Biochemistry* **2004**, *324*, 241–249.

[157] A. E. Carpenter, *Nat Chem Biol* **2007**, *3*, 461–465.

[158] K. A. Kelly, S. Y. Shaw, M. Nahrendorf, K. Kristoff, E. Aikawa, S. L. Schreiber, P. A. Clemons, R. Weissleder, *Integrative Biology* **2009**.

[159] S. Santra, H. Yang, D. Dutta, J. T. Stanley, P. H. Holloway, W. Tan, B. M. Moudgil, R. A. Mericle, *Chem. Commun.* **2004**, 2810–2811.

[160] W. D. Brown, S. Green, *Anal. Biochem.* **1970**, *34*, 593–595.

[161] V. K. Sarin, S. B. H. Kent, J. P. Tam, R. B. Merrifield, *Anal. Biochem.* **1981**, *117*, 147–157.

[162] Y.-C. Cheng, and W. H. Prusoff, *Biochem. Pharmacol.* **1973**, *22*, 3099–3108.

[163] M. Kofler, K. Heuer, T. Zech, C. Freund, *J. Biol. Chem.* **2004**, *279*, 28292–28297.

Abstract

Protein-protein interactions are involved in many metabolic pathways. However, their study and inhibition is difficult. In this dissertation project, a novel tool targeting protein-protein interactions has been designed. A labeling resin was synthesized by linking *N*-Fmoc-aminofluorescein to trityl resin and applied either for *C*-terminal labeling of peptides, or for small molecules labeling.

In the first part of the project, libraries of *C*- and *N*-terminally fluorescein labeled peptides for the CD2BP2 and PERQ2 GYF domains were synthesized and evaluated in FP assays. A probe suitable for screening was found for the CD2BP2 GYF domain: the peptide Fluo-Bpa-EFGPPPGWLGR-NH$_2$, with an affinity of 2.3 µM and a Z' factor of 0.87. A screening of the 16,896 compounds of the ChemBioNet library was realized using a displacement assay with FP detection, and lead to the discovery of primary hits that were further evaluated by independent methods: NMR, ITC and CD but could unfortunately not be validated. Indeed, autofluorescence of the small molecules interfered in the assay and lead to many false positives.

In a second approach, a diversity set of 265 fluorescein labeled small molecules was synthesized. The fragments were linked to the fluorescein resin either by amide bond formation, or by nucleophilic substitution after conversion of the aniline into an electrophile by addition of bromoacetic acid. This library was tested for affinity on twelve different proteins using a "direct" FP assay with inverse read-out, where an increased FP signal corresponds to binding of the small molecule to the protein. As a result, primary hits were found for almost all proteins, and validated with FP titration for the CD2BP2 GYF domain.

Zusammenfassung

Protein-Protein Wechselwirkungen sind komplexe Prozesse, die an vielen metabolischen Signalwegen beteiligt sind. Ihre Untersuchung ist schwierig. Im Rahmen dieser Doktorarbeit wurde ein neues Tool zur Aufklärung von Protein-Protein Wechselwirkungen mittels Fluoreszenzpolarisation (FP) entwickelt. Ein Festphasenlabelingreagenz für die C-terminale Markierung von Peptiden und kleinen Molekülen wurde durch Immobilisierung von *N*-Fmoc-Aminofluorescein an Tritylharz hergestellt und für das C-terminale Markieren von Peptiden oder von kleinen Molekülen verwendet.

Im ersten Teil des Projektes wurden Bibliotheken von *C*- und *N*-terminal gelabelten Peptiden für die CD2BP2- und die PERQ2-GYF Domänen hergestellt und in FP-Assays evaluiert. Eine geeignete Sonde für ein Screening mit der CD2BP2-GYF Domäne wurde gefunden: Das Peptid Fluo-Bpa-EFGPPPGWLGR-NH$_2$, mit einer Affinität von 2.3 µM und einem Z' Faktor von 0.87. Das Screening der 16 896 Verbindungen der ChemBioNet Substanzbibliothek wurde mit einem FP-Verdrängungsassay durchgeführt. Die primären Hits konnten jedoch durch Untersuchung mit weiteren unabhängigen Methoden (NMR, ITC, CD) nicht validiert werden. Die Eigenfluoreszenz vieler Verbindungen führte zu falsch positiven Hits.

Im zweiten Teil des Projekts wurde mit dem Festphasenlabelingreagenz eine Bibliothek hoher Diversität von 265 mit Fluorescein markierten Molekülen hergestellt. Die Fragmente wurden entweder durch Amidkupplung oder nach vorheriger Umwandlung des Anilins in ein Elektrophil mittels Kupplung von Bromessigsaüre durch nukleophilen Angriff an das Harz gekuppelt. Die Affinität dieser Moleküle zu zwölf verschiedenen Proteinen wurde in einem „direkten" FP Assay untersucht, bei dem eine Erhöhung des FP Signals einer starken Bindung des Moleküls zum Protein entspricht. Für die meisten Proteine wurden Hits gefunden und für die CD2BP2-GYF Domäne durch Titration validiert.

For reasons of data protection,

the curriculum vitae is not included in the online version