Light Responsive Linchpins for Accelerated Discovery of Light-Responsive Mono-

and Bicyclic Peptide Ligands

by

Seyed Mohammadreza Jafari

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Chemistry

Department of Chemistry

University of Alberta

Abstract

In this thesis I describe the development and application of light-responsive (LR) linchpins for efficient discovery of LR-peptides from genetically encoded peptide libraries. LR-linchpins have been used for dynamic control of the structure of peptides and their biochemical properties such as binding to a protein receptor. Rational design of such LR binders is not trivial and it mandates previous knowledge of the structure of the target receptor and the binding mechanism.

In the first chapter of this thesis, I briefly review several linchpins that have been used to cyclize peptides. In Chapter 2 of this thesis, I describe integration of a well-known LR-linchpin 3,3'-bis(sulfonato)-4,4'-bis(chloroacetamido)azobenzene (BSBCA) in a phage displayed library of peptides for synthesis of a library of LR-macrocycles displayed on the coat protein of the phage. Selection from this library using streptavidin as a bait resulted in identification of three cyclized LR-peptides that bind to streptavidin. We show that binding affinity decreases reversibly in response to blue light (370 nm) as measured by quantitative mass spectroscopy. In the same chapter, we also describe a method for quantification of the efficiency of cyclization of the phage library with the BSBCA linchpin. Such quantification is vital for development and screening of chemically-modified phage displayed libraries, because it provides a measure for the quality of the library before selection. Appendix 1 of this thesis discusses the possible reasons of discrepancy observed in the reaction of phage display library and mono-clonal phage. In Appendix 2, I demonstrate screening of an LR-library against anti FLAG antibody which resulted in enrichment of peptide binders against a secondary target present in the panning, protein A.

To extend the repertoire of LR-linchpins, in Chapter 3, I describe the synthesis of 3,3'-bis(sulfonato)-4,4'-bis(buta-2,3-dienoylamido)azobenzene (BSBDA), a new LR-linchpin that can undergo fast reaction with cysteine containing peptides – with $k = 30$ M^{-1}s^{-1}. We designed this linchpin based on BSBCA which has been used by several groups to produce cyclic LR-peptide scaffolds. We used BSBDA to cyclize a streptavidin binding peptide precursor and confirmed the structure with Nuclear magnetic resonance (NMR) and mass spectroscopy (MS) analysis. We also showed that BSBDA can react with the thiols of cysteines displayed on coat protein of the phage, in the presence of reducing agent tris(2-carboxyethyl)phosphine (TCEP), with > 98% yield in only 20 min at pH 8. The fast reaction of linchpins with phage is essential because it reduces the exposure time of the phage proteins with toxic chemicals, such as TCEP.

Finally, I describe the first example of a C3$_\mathrm{V}$ symmetric LR-linchpin that can produce bicyclic LR-scaffolds from peptides consisted of unprotected natural amino acids. The reaction of the linchpin with the thiols and glyoxal produced by oxidation of N-terminal serine in the peptide sequence SX_nCRRRRC (where n = 1- 4) resulted in formation of bicyclic LR-peptides in 11-73% isolated yield. All four bicyclic peptides isomerized to *cis* conformer after irradiation with 365 nm light. The switching efficiency and the thermal relaxation half-life varied for different peptide sequences. I anticipate that bicyclic LR-peptides can be used in studies where a dynamic surface and a static motif are required on the same molecule where each motif has specific biological function. One of the possible future explorations is inhibition of dimerization of EGFR by tuning the static surface for cell penetration and the dynamic surface to inhibit the coiled-coil formation of the intracellular helix of EGFR which is necessary for its function.

Preface

Chapter 1 of this thesis contains an original brief review of molecular linchpins that are used for cyclization of peptides and has not been published. I partially wrote the manuscript. R. Derda wrote part of the manuscript and provided advice

Chapter 2 of this thesis has been published as M. R. Jafari, L. Deng, P. I. Kitov, S. Ng, W. L. Matochko, K. F. Tjhung, A. Zeberoff, A. Elias, J. S. Klassen, and R. Derda, "Discovery of Light-Responsive Ligands through Screening of a Light-Responsive Genetically Encoded Library", *ACS Chem Biol*, 2014, **9**: 443-450. I designed and performed the experiments for phage modification and quantification, library screening, ligand synthesis and modification as well as the light-responsiveness experiments. I partially contributed to writing the paper. S. Ng contributed with phage modification. W. L. Matochko, A. Zeberoff, and A. Elias contributed to emulsion amplification of phage. K. F. Tjhung performed the DNA sequencing of the phage. L. Deng performed binding measurement of the selected ligands to the protein by mass spectrometry. J. S. Klassen advised L. Deng for mass spectrometry methods. P. Kitov performed Molecular Dynamics and docking studies. R. Derda contributed in the design of experiments, calculation of binding measurements and writing the manuscript. F. Deiss critically reviewed the manuscript. J. Zheng, and M. Miskolzie assisted with mass spectroscopy and NMR analyses respectively. N. Mokhtarihaj assisted with synthesis of chemical intermediates.

Chapter 3 has not been submitted for publication yet. I designed and performed all the experiments and part of the analysis of data. J. Lakusta contributed to optimization of synthesis. R. Lundgren critically reviewed the manuscript and advised J. Lakusta on syn-

thesis optimization. R. Derda wrote part of the manuscript and advised me on the experiments and analysis.

Chapter 4 has not been submitted for publication yet. I contributed to the design of a light-responsive linker for generation of light-responsive bicyclic peptides. I designed and performed all the synthesis and experimental measurements. R. Derda advised me on experimental and analysis. Mark Miskolzie helped in acquisition and interpretation of NMR spectra.

Appendix 1 has not been submitted for publication yet. I performed all the phage experiments and measurement and part of the analysis. W. Matochko contributed in deep sequencing studies. R. Derda contributed in analysis of deep sequencing data.

Appendix 2 has not been submitted for publication yet. I perfumed the phage modification and screening, peptide array synthesis, protein labeling, ELISA assays and fluorescent polarization assays. W. Matochko contributed in deep sequencing studies. R. Derda contributed in analysis of deep sequencing data.

I dedicate this thesis to my wife, Andia, who supported me in this long journey to accomplish my PhD studies.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Abbreviations

Arg	arginine
Asn	asparagine
Asp	aspartate
Azb	azobenzene
BIA	biotin-peg20iodoacetamide
BSA	bovine serum albumin
BSBCA	3,3'-bis(sulfonato)-4,4'-bis(chloroacetamido)azobenzene
BSBDA	3,3'-bis(sulfonato)-4,4'-bis(buta-2,3-dienoylamido)azobenzene
BSH	Cystein-*N*-biotin
cfu	colony forming unit
Cys	cysteine
DCM	dichloromethane
DIPEA	*N,N*-diisopropylethylamine
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dsDNA	double stranded deoxyribonucleic acid
E. coli	*Escherichia coli*
ELISA	enzyme linked immunosorbant assay
ESI	electron spray ionization
EtOAc	ethyl acetate
EtOH	ethanol

FBS	fetal bovine serum
Gln	glutamine
Glu	glutamate
Gly	glycine
HATU	3-[Bis(dimethylamino)methyliumyl]-3H-benzotriazol-1-oxide hexafluorophosphate
HBTU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
His	histidine
HPLC	high-performance liquid chromatography
HRMS	high resolution mass spectrometry
HRP	horseradish peroxidase
Ile	isoleucine
IPTG	isopropyl b-d-1-thiogalactopyranoside
IR	infrared spectroscopy
LB	lysogeny broth
LCMS	liquid chromatography- mass spectrometer
Leu	leucine
LR	light-responsive
Lys	lysine
MALDI	matrix assisted laser desorption/ionization
MeCN	acetonitrile
MeOH	methanol

Met	methionine
MS	mass spectra
NMR	nuclear magnetic resonance
PCR	polymerase chain reaction
PEG	poly(ethylene glycol)
pfu	plaque forming units
Phe	phenylalanine
Pro	proline
RNA	ribonucleic acid
r.t.	room temperature
Ser	serine
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
THF	tetrahydrofuran
Thr	threonine
Trp	tryptophan
Tyr	tyrosine
Val	valine
WT	wild type
X-gal	5-bromo-4-chloro-3-indolyl β-D-galactopyranoside

Chapter 1 Molecular Linchpins: From Enforcing Productive Conformations to Accelerated *de Novo* Discovery and Dynamic Control of Bioactive Conformations in Peptides

1.1. Overview of Peptide Therapeutics

Peptides and peptide drugs constitute a significant fraction of the therapeutics for various types of diseases. The global market of peptide therapeutics was $18.9 billion in 2013 and is estimated to reach $23.7 billion by 2020.[1] The world pharmaceutical market was worth an estimated $865 billion in 2014.[2] A significant fraction of the peptide-derived drugs is composed of an amino acid sequence and additional chemical functionalities that improve biological and pharmacological properties of the drug. In this chapter, I focus on synthetic cross-linkers that I refer to as "linchpins", which are commonly used to constrain the secondary structure of peptides and equip them with added benefits such as resistance to proteolytic degradation and conformational stability. The latter properly leads to an increase in binding potency and increased bioavailability due to increased permeation through biological membranes. Some linchpins can even introduce properties not found in any natural peptides such as light-responsiveness.

Peptides cyclized by linchpins, can be viewed as a sub-class of a larger family of peptide-derived drugs with desired pharmacological performance *in vivo*. To understand how chemical modifications by linchpins improve drug discovery, I

briefly review several canonical example of chemical modification used in modern peptide therapeutics.

Peptide drugs are usually defined as molecules with less than 50 amino acids to discriminate them from larger biological products such as antibodies.[3,4] In 2014, there were more than 60 FDA approved peptide drugs on the market, 140 peptide drugs in clinical trials and more than 500 peptide drug candidates in pre-clinical trial.[5,6] The sales for the top seven selling peptide drugs (i.e. Copaxone®, Lupron®, Sandostatin®, Zoladex®, Victoza®, Forteo®, Byetta®) were more than $9.0 billion in 2011.[5]

Peptide drugs occupy a unique chemical space between small molecule therapeutics and antibodies. Antibody-based therapeutics is a rapidly growing class of therapeutics constituting 10% of the newly approved drugs in the past 10 years, however, limited tissue penetration is a fundamental drawback of antibodies. Peptide drugs can penetrate into tissues due to their smaller size.[7] In contrast to many small molecule drugs, peptides accumulate less in the tissues and often show higher specificity.[8]

Suboptimal pharmacokinetic properties of the peptides are a major challenge in the use of peptides as drugs. The susceptibility of the peptides to peptidases in the gastrointestinal tract reduces their bioavailability significantly through oral administration. Another barrier to oral administration is the hydrophilic structure of the peptides that dramatically reduces their delivery to the systemic circulation through passive diffusion. Even if peptides pass the digestive system and intestinal barrier, renal and hepatic clearance as well as degradation by peptidases in tissues

and blood shortens the half-life of the peptide drugs. The half-life of peptides administered directly in the circulation can be as short as a few minutes.[9,10]

Such a short half-life has evolutionary origins: many peptides act as messenger hormones, cytokines, neural transmitters and metabolic information. Rapid degradation of the messenger peptides ensures effective turn-off mechanism and high temporal fidelity of the message. However, when peptides are adapted as drugs, in many cases the life-time of the peptide has to be extended to avoid frequent or continuous administration.
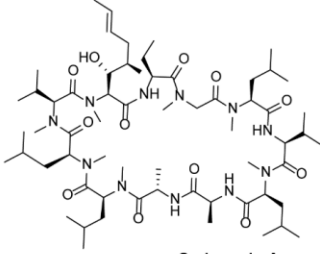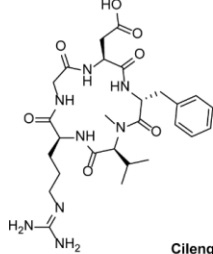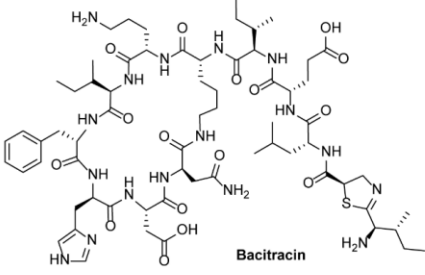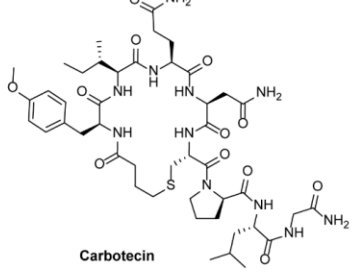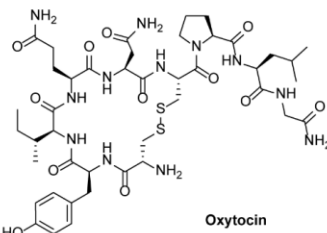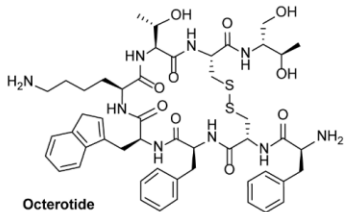
Both physical and chemical solutions have been devised to increase the bioavailability and resistance of the peptides. Examples of physical modifications are coating of the peptide drug with polymers to protect the peptides' gastric degradation and formulation of the peptide with protease inhibitors.[11] Physical protection and peptide formulation is beyond the scope of this review and is reviewed elsewhere.[11,12] Examples of chemical modification of peptides that improves their pharmacokinetics and pharmacodynamics are conjugation of peptides with fatty acids to increase circulation half-life *via* interaction with serum albumin and pegylation that increases the hydrodynamic radius. Both modifications delay renal elimination.[8,13] An example of the first type of modification can be found in Levemir®, a chemically modified insulin developed by NovoNordisk, which is an FDA approved form of insulin myristoylated at a lysine residue. A seemingly minor modification – 17 carbons added to a polypeptide with 250 carbons – yields a dramatic increase in half-life from 2.8 h to 8.8 h.[14] This increase is even more dramatic in Victoza®, a 37 amino acid glucagon-like peptide-1 receptor (GLP-1) agonist, also

developed by NovoNordisk, which is used for management of type 2 diabetes. Palmitoylation of a lysine residue in this peptide increased the half-life of the subcutaneously injected drug from 1 h to 11-15 h.[15] Examples of second type of modification are two FDA approved drugs Pegasys®, pegylated interferon developed by Genentech, and Pegintron®, another variant of pegylated interferon distributed by Schering. These technologies highlight the benefits of introducing non-peptide foreign structures, which equip the peptides with beneficial pharmacokinetic properties, while providing only a minor interference with their biological function.
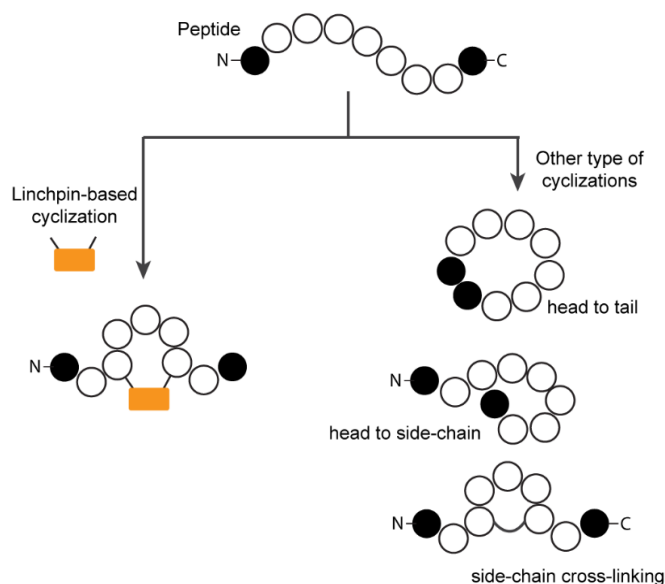
Figure 1.1. Linchpin based cyclization *vs.* other types of cyclization.

Linchpins are an emerging class of chemical modifiers that can be defined as molecules that convert a linear peptide to a cyclic peptide. This chemical modification can potentially provide several advantages: (i) proteolytic stability (ii)

4

Cyclization	Natural peptide	Synthetic peptide
Head to tail	Cyclosporin A	Cilengitide
Head/tail to side chain	Bacitracin	Carbotecin
Disulfide	Oxytocin	Octerotide

Figure 1.2. Chemical structure of selected FDA approved cyclic peptide drugs.

restricted conformation mobility (iii) increased propensity to permeate through bio-membranes.

Development of the linchpins originated from the observation that many successful natural or synthetic peptide-based therapeutics have cyclic structure. These derivatives were made using three general approaches to cyclization (Figure 1.2): (i) head to tail cyclization (or macrolactamization) (ii) head-to-side chain or tail-to-

side chain cyclization, and (iii) cross-linking of functional groups in side chains, such as conversion of thiols of cysteines to disulfide bond. Reviews of these cyclizations can be found elsewhere.[16] Figure 1.2 depicts the structure of three FDA approved synthetic drugs and three natural peptides highlighting these three types of cyclization. To delineate cyclization assisted by linchpins from the canonical approaches to macrocyclization, we briefly review general cyclization approaches and introduce advantages offered by linchpin based cyclization over the other routes.

1.1.1. Improving the Pharmacodynamics of Peptide by Macrocyclization

Preorganization of structure of the peptide by macrocyclization can increase the binding affinity by decreasing the entropic penalty resulting from binding.[17,18] A linear flexible molecule exists in multiple conformational states and the contribution of the conformer necessary to achieve the most proper fit for a potent interaction is often small (Figure 1.3A). Constraining the skeleton of the ligand can preorganize its structure by maximizing the fraction of the conformers that exist in the structure optimal for the interaction with the receptor (Figure 1.3B). Therefore, macrocyclization can result in a higher binding association constant (K_a) often by increasing the on-rate of binding. Furthermore, dynamic linchpins can change the conformation in response to external stimuli. Such conformational change in a small motif of a macrocycle can dramatically change its structure, favoring a defined set of conformers over the others. When these changes are designed or selected for, they can result in dynamic a increase or decrease in K_a. (Figure 1.3C).

6
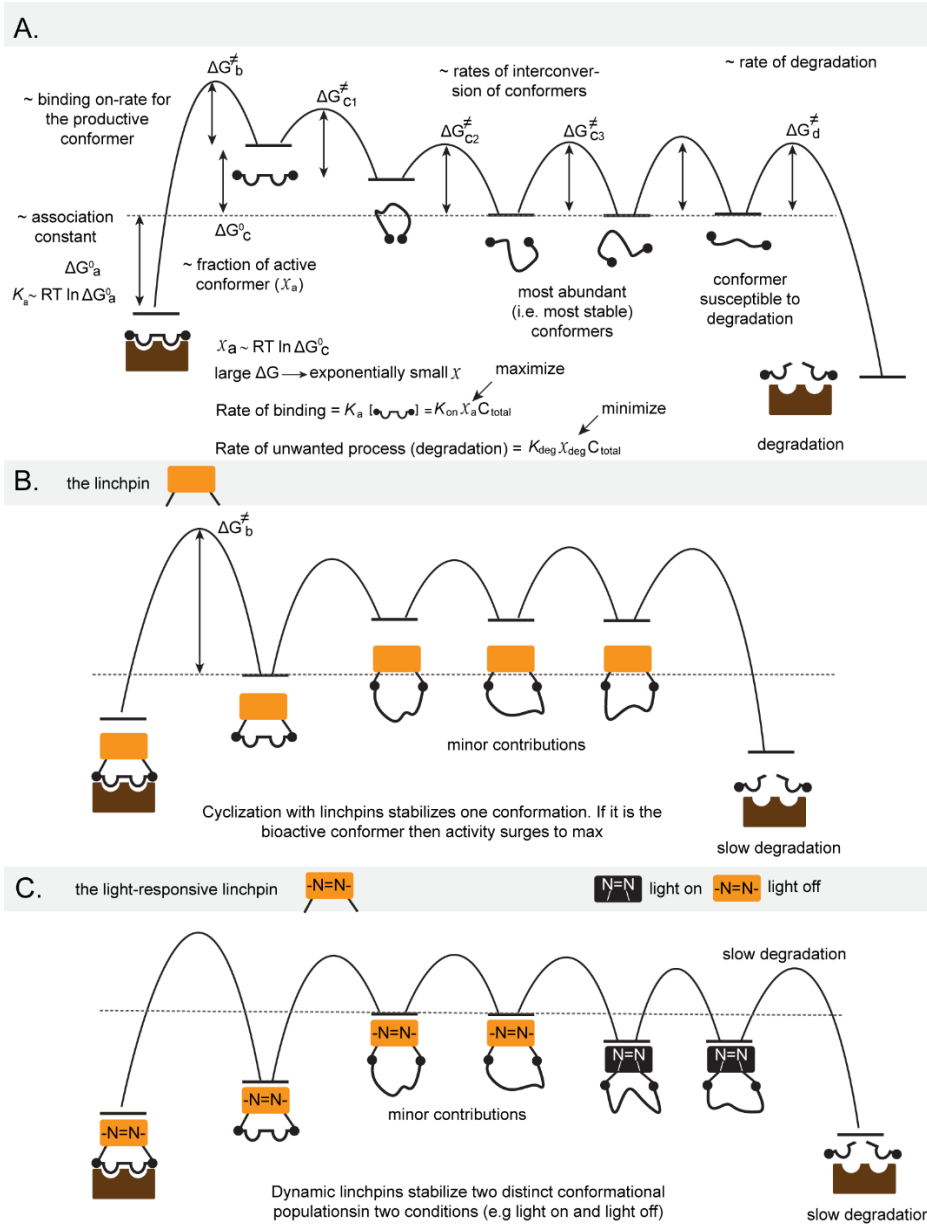
Figure 1.3. The effect of cyclization by linchpins on peptide conformation. A) A linear peptide exists in multiple conformation, most of which are not optimal for interaction with the receptor. B) Macrocyclization can increase the population of the conformers that exhibit the best fit for the receptor. C) Macrocyclization with linchpins that respond to external stimuli, can enrich the fraction of a set of productive conformer in "light-on" condition, and deplete it in "light-off" condition.

7

1.1.2. Improving the Pharmacokinetics of Peptides by Macrocyclization

Peptide cyclization can overcome difficulties in active or passive intestinal absorption and/or cell penetration, because macrocyclization of the backbone can facilitate formation of internal hydrogen bonds which in turn are known to facilitate masking of hydrophilic areas and exposure of hydrophobic side chains when passing thorough hydrophobic media such as the interior of a lipid bilayer.[19] In contrast, a large exposed polar surface area is detrimental to the peptide diffusion into the cell.[20] Cyclosporine A is a well-known example of an orally bioavailable peptide that permeates the cell membrane by this mechanism. Cyclosporine shows the same conformation in aqueous solvent as in co-crystal structure in complex with its receptor cyclophiline. However, upon entering the hydrophobic environment (cell membrane) cyclosporine changes its conformation to expose its *N*-methylated hydrophobic backbone to the solvent by making internal hydrogen bonds.[18] Lockey and co-workers generalized this observation and made similar *N*-methylated peptide scaffolds that exhibit bi-stability upon transfer from polar to non-polar environments. [21] The structures are currently evaluated by Circle Pharma for their potential to serve as orally bioavailable pharmaceuticals.[22]

Another potentially general solution to the cell permeability problem that uses macrocyclization of peptides is cross-linking of unnatural hydrocarbon side chains which in several reports has shown to yield "stapled" peptides. About 15-20 years since the original report by Blackwell and Grubbs,[23] and later by Verdine and co-

workers, [24] the stapling concept has been adapted by several pharmaceutical companies. Aileron, for example is conducting a phase I clinical trial of a stapled peptide that binds to p53 protein.

Macrocyclization has been also applied to constrain canonical cell-penetrating sequences such as poly-arginine (R_x, where $x = 4 - 9$). Such peptides showed higher cell penetration than their linear counterparts.[25,26] Cyclization with linchpins can introduce extra features in the structure of a potential peptide drug such as dynamic control, hydrophobicity and albumin binding, where devising such features with conventional macrolactamization is challenging.

1.2. Applications of Molecular Linchpins for Macrocyclization

Presence of multiple reactive side chains (i.e. Lys, Cys, Asp, Glu, *N*-terminal amine, *C*-terminal carboxylic acid) in amino acid oligomers often demands incorporation of unnatural amino acids, such as those used in stapling, or orthogonal protection of reactive side chains in synthetic peptides, such as those used in classical macrolactamization approaches. Many of the biologically relevant polypeptides (e.g, insulin, epidermal growth factor, etc.), however, are ribosomally synthesized. While incorporation of orthogonal handles in ribosomally synthesized peptides and proteins is possible, it often faces problems with translational fidelity and low efficiency of amber codon suppression.[27] The use of linchpins allows building the orthogonality of the reaction into the design of the small molecule linker to permit using unprotected natural amino acid side chains as handles for cyclization.

The linchpin links two or more reactive groups present in amino acid side chains. Connection of *n+1* anchoring amino acid will result in production of *n* cycles. Mono-,[28,29] bi-[30,31] and tricyclic[32-34] peptides have been produced using linchpins with 2, 3 and 4 anchors, respectively. The advantages of using linchpins for cyclization are: (i) Orthogonal and biocompatible functionalities can be programmed in the unnatural structure of the linchpins, bypassing the use of unnatural structures in the peptide chain. (ii) Linchpins can introduce extra features into peptide structure such as hydrophobicity,[35,36] light-responsiveness,[37,38] conformational stability,[34,39] etc. (iii) A single linchpin linker can be used for cyclization of various peptides with different binding targets.

There are three main routes for cyclization of peptides by linchpins: (i) Chemical modification using native side chain functionalities of the natural amino acids. Nucleophiles in cysteine, lysine and tyrosine[40] have been widely used for such bio-conjugation. (ii) Pre-activation of the amino acid side chains to provide new orthogonal functionalities (e.g. glyoxal,[41] dehydroalanine,[42,43] etc.). Examples are bio-orthogonal oxime bond ligation[44] or Michael addition.[45] (iii) Incorporation of unnatural amino acids with orthogonally reactive functionality.[46,47]

The use of linchpins is even more attractive in genetically encoded libraries of peptides, such as those displayed on phage or mRNA. Introduction of an unnatural linker that is reactive towards natural amino acids is a convenient strategy to constrain $10^9 - 10^{12}$ peptides at once. We and others have already shown the cyclization of peptides displayed on bacteriophage coat protein.[28,30] In the following sections we will focus on selected molecular linchpins that have been used to cyclize linear

peptides through covalent bonds. Other methods for stabilization of the peptide structure (i.e. metal coordination,[48,49] hydrogen bonding[50] and other non-covalent strategies) have been reported, but they will not be discussed in this chapter.

1.2.1. Linchpins that Form Amide Bonds

Macrolactamization is commonly used for peptide cyclization, since the reactive precursors – amine and carboxylic acid – are readily available in the backbone of the peptide. Amine-reactive linchpins made of bis-succinimidyl esters with different lengths are commercially available or can be readily generated. The availability of reactive groups brings about both convenience and challenge. Amine side chains are common reactive residues and are often present in the "active" form of the peptide. Unwanted cross-linking of lysines by the linchpin can destroy the active structure. Use of unprotected natural amino acids during cyclization is possible, but in only a narrow class of peptide sequences. In all other peptides, such cyclization mandates the use of orthogonal protecting groups. Despite the need for extra synthetic steps, the amide-forming linchpins are popular due to the wide availability of orthogonally protected Lys building blocks and robustness of amide bond forming reactions.

A typical example of a linchpin that employs amide bond formation is trimesic acid (**TA**). Pei and co-workers[31] demonstrated the use of **TA** to cross-link
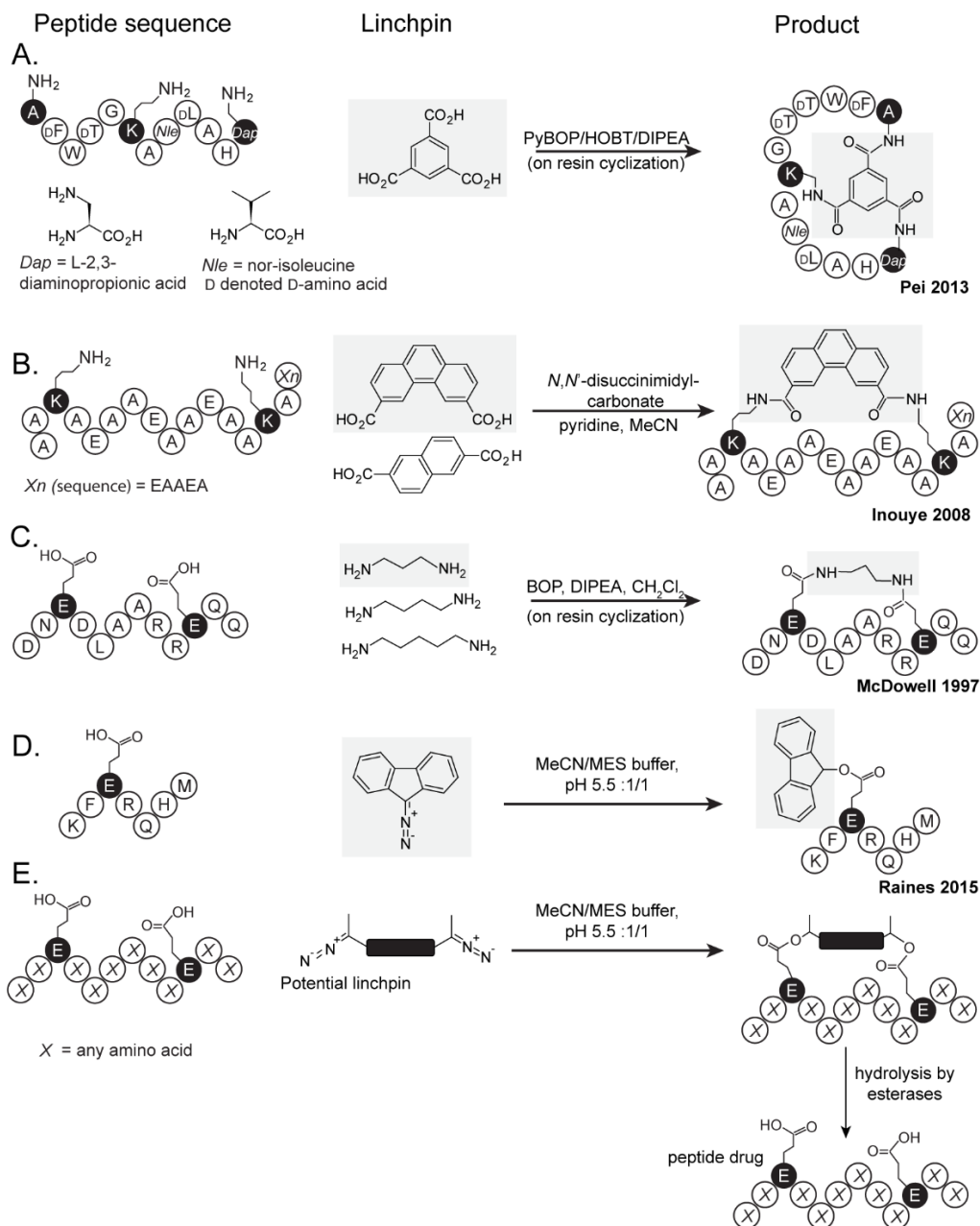
Figure 1.4. Cyclization of peptides with amide bond forming linchpins.

residues of a one-bead-two-compound peptide library that harbored three amines at fixed position in order to create a repertoire of bicyclic peptides (Figure 1.4A). A

potential limitation of this approach is the need for multiple rounds of protection/deprotection for stepwise unmasking of the reactive amines and cyclization. First, deprotection of allyloxycarbony (alloc)-protected 2,3-diaminopropionic acids (**Dap**) revealed the free amine used to attach the **TA** through a single amide bond. Then, the FMOC-deprotection of ε-amine of lysine and *N*-terminal amine exposed two other functionalities that made the other two amide bonds with the anchored **TA** after exposure to PyBOP activator. Screening of the bicyclic library against tumor necrosis factor-α (TNF-α) identified two bicyclic peptides **C1** and **C2** with $K_d = 0.45$ μM and $K_d = 1.6$ μM, respectively. The **C1** exhibited no binding to TNF-α in monocyclic form and a $K_d > 10$ μM when not cyclized. The **C1** also could protect the cells from TNF-α-induced cell death, increasing the LD_{50} of TNF-α from 0.46 ng/mL to 1.8 ng/mL.

Inouye and co-workers scanned 12 variants of di-carboxylic acid linchpins to study the effect of the length of the linker on stabilization of the peptides cross-linked at *i*, *i*+4, *i*, *i*+7 and *i*, *i*+11 (Figure 1.4B depicts peptides with lysine residue at *i*, *i*+11).[51] They found peptides cross-linked at *i*, *i*+11 with flurorenyl- and naphtyl- dicarboxylic acid exhibited the highest helicity (70% and 75% helicity, respectively) at 25 °C.

McDowell and co-workers used alkanediamines linchpins ($NH_2X_nNH_2$, n = 3-5) to crosslink between glutamic acids located at *i*, *i*+7 (Figure 1.4C).[52] A 1,5-pentanediyl linker increased the helicity of the linear peptide (Ac-TN<u>E</u>D-LAARR<u>E</u>QQ) from 20% to 100% as measured by circular dichroism (CD). Propanediyl and butandiyl linkers increased the helicity to 84% and 63% respectively.

The above strategy mandates simultaneous pre-activation of both side chain carboxylates, making such strategy unreliable. Raines and co-workers developed a reaction that bypasses this problem by using a diazonium compound with exclusive reactivity towards native carboxylate nucleophile in the presence of water, amines of Lys, thiols of Cys and phenolic hydroxyl groups of Tyr (Figure 1.4D).[53] Although Raines did not explicitly describe a bis-diazonium linchpin, we anticipate that extension of mono-acylation to bis-acylation is feasible and can result in a bio-reversible cyclization (Figure 1.4E).

1.2.2. Linchpins that Undergo S$_N$2 Nucleophilic Substitution

Thioether bond formation is a fast and selective reaction between cysteines and halo-acetamide or halo-benzyl as common regents. The reaction with cysteine has kinetics with k up to $16 - 20$ M^{-1}s^{-1}. It is accelerated at pH > 7.5 because the thiolate anion (pK_a of cysteine is $7 - 10$) is >10,000 times more reactive nucleophile than the thiol.[54] The resulting thioether can mimic disulfide bonds, but is not susceptible to undesired reduction and exchange. Replacing disulfides with more stable thioether bonds while preserving the activity of the peptides has been addressed by many research groups over the past 20 years. Di-bromomethylbenzene (**DBMB**) is one of the first thiol-reactive linchpins used to form cyclized peptides from unprotected peptides, in aqueous conditions, paving the way to chemical post-translational modifications of peptide libraries.
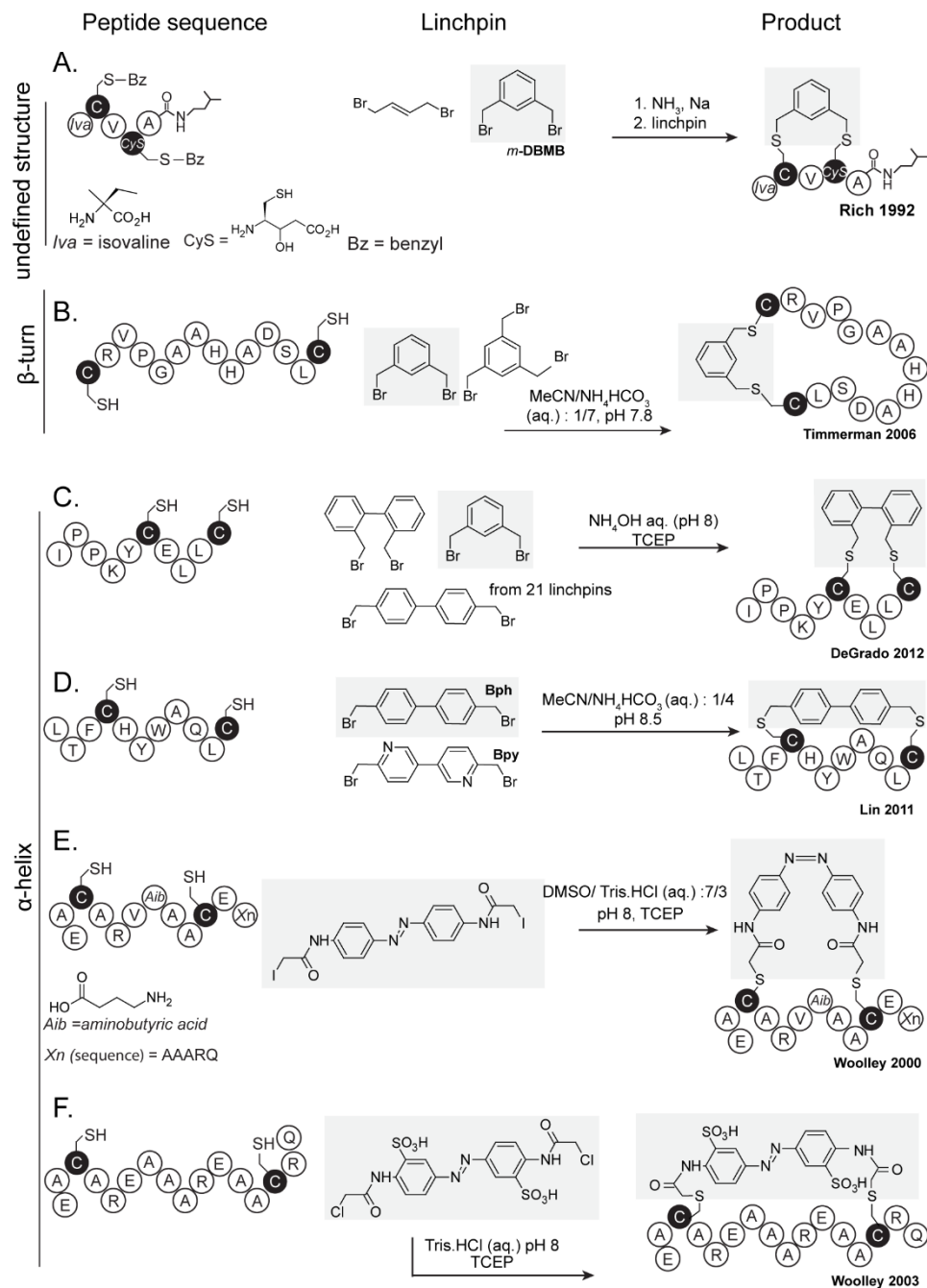
Figure 1.5. Representative alkyl-halide and benzyl halides used as linch-pins.

In an early study, Rich and co-workers used cyclization of an analogue of pepstatin (pepsin protease inhibitor) with allylbromide and bromomethyl benzenes

(**BMB**) to produce new pepsin inhibitors (Figure 1.5A).[55] Unfortunately, the new cyclized peptides did not show improved inhibition comparing to the linear one. This observation shows that simple constraint of the peptide may not be sufficient. The peptide has to be constrained in an active conformation. Another challenge is that the pepsin recognizes multiple conformations and such promiscuous recognition might be the reason why enforced depletion of conformational pool made no difference in binding affinity.

Timmerman and co-workers used **BMB** derivatives to scan the activity of an overlapping set of a 12-residue fragment of follicle-stimulating hormone (FSH-β, Figure 1.5B). They showed that cyclization of the sequence CRVP-GAAHHADSLC with *m*-**DBMB**, *o*-**DBMB** or *p*-**DBMB** resulted in higher affinity (K_d = 2-3 μM), while disulfide-cyclized peptide exhibited K_d = 130 μM and linear peptide had K_d > 1 mM.[34] In this case, the linkers replace a disulfide bond at *i*, *i*+13 that stabilizes the shape of the loop of a β-turn. They also found *o*-**DBMB** and *p*-**DBMB** cyclize faster, most likely because the first thioether formation activates the remaining **BMB**.

DeGrado and co-workers later expanded this approach to 21 different alkyl-halides including **BMB**s (Figure 1.5C) to staple a peptide at *i*, *i*+4 to identify the optimal linchpin for stabilization of the helicity of the peptide Ac-IPPKYCELLC.[39] *m*-**DBMB** cyclized peptide showed the highest helix stabilization based on the shape of the peaks in CD spectrometry experiments. Varying the position of cysteines in the chain (while keeping them at *i*, *i*+4) also changed the helical content of

the peptide, although the authors did not quantify helicity for any of these peptides.[39] **m-DBMB** cyclized peptide exhibited K_i = 11 μm for Kalpain-1 protein (a calcium-regulated cysteine protease) while the unmodified linear peptide showed lower inhibition (K_i <100 μM). These results show that not only the structure of linchpin is important in the stabilization of the helix, but careful positioning of the cysteines is required for optimal results.

Lin and co-workers used a di-bromomethylbiphenyl (**Bph**) and di-bromo-methylbipyridine (**Bpy**) to staple a peptide with cysteines at *i*, *i*+7 (Figure 1.5D). The linear peptide (LTF<u>C</u>HYWAQL<u>C</u>S, a dual inhibitor of p53-Mdm2/Mdmx interactions) showed IC$_{50}$ = 500 nM. After cyclization with **Bph** and **Bpy**, IC$_{50}$ decreased to 22 ± 4.0 nM and 14 ± 2.0 nM, respectively.[56] In addition, modification of the linear peptide resulted in a 6-fold increase in cell permeability after cyclization with **Bph** (6 fold) and **Bpy** (5 fold).
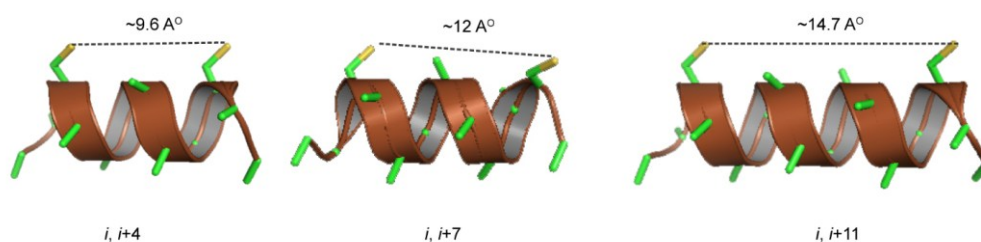
Figure 1.6. Schematic description of location of cysteines in a peptide at *i*, *i*+4, *i*, *i*+7, *i*, *i*+11 to form a stable α-helix after cyclization with linchpins. (Figures generated in PyMol for visualization only).
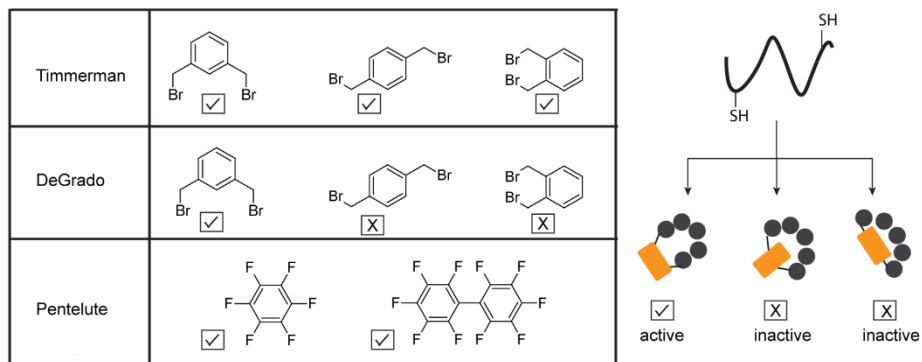
Figure 1.7. Linchpin scanning using a portfolio of cysteine reactive linchpins. One of them has an optimal geometry to maintain the helix in an active folded form. We only included the linchpins for which binding data for the corresponding cyclized peptide was available.

When positioned at i, $i+4$, i, $i+7$ and i, $i+11$, amino acids side chains locate at the same face of an α-helix (Figure 1.6). Therefore, to promote an α-helix structure, the linchpin should connect the amino acids at this position. The end-to end of the linchpin plays an important role here. For example, DeGrado found no enhanced helix stabilization by using **Bph** to connect cysteines at i, $i+4$, while Lin showed that the same linker resulted in active peptide when connecting cysteines at i, $i+7$ (Figure 1.7). In another example, DeGrado reports stabilization of α-helix when ***m*-DBMB** connects cysteines at i, $i+4$ (Figure 1.5C), while the same linker can stabilize a β-turn, if connecting cysteines at i, $i+13$ (Figure 1.5B).

While many examples focus on stabilization of α-helices, cysteine reactive linchpins can constrain other conformations such as β-turn or even unstructured

loops in larger polypeptides. For example, Kritzer and co-workers devised a loop finder to identify bioactive loops (Figure 1.8).[57] They then used a linchpin scanning approach to identify the linchpin structure that captures the loop in an active conformation.
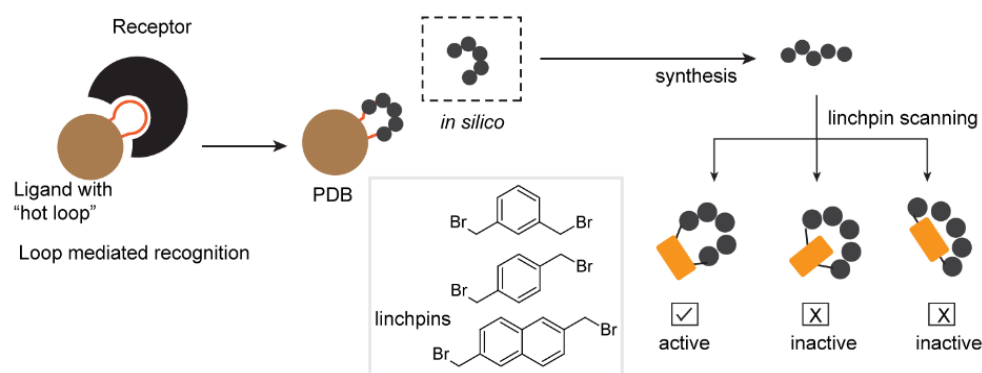
Figure 1.8. Loopfinder coupled to linchpin scanning strategy pursued by Kritzer and co-workers.

All linchpins described above are exclusively used to constrain the structure of the peptide. Linchpins, however, can also accommodate reactive moieties for downstream functionalization. Dawson and co-workers used dichloroacetone (**DCA**) to constrain the peptide and to simultaneously install a ketone functionality. The latter can be used for bioorthogonal labeling of the peptide or a second cyclization (Figure 1.9B).[58] For example, the authors reacted cyclized peptides with aminooxy containing substrates – such as biotin, alexafluor 647 or FLAG epitope (peptide sequence: DYKDDDDK), etc. – in the presence of 100 mM aniline at pH 4.5. In most cases, the oxime ligation proceeded in satisfactory yields (>85%) after a 16 h reaction. The authors attempted to use a **DCA** linchpin to form a bicyclic product through reaction of *N*-terminal aminooxy group in the peptide with the ketone;

however, formation of a bicyclic product was underwhelming and yielded only
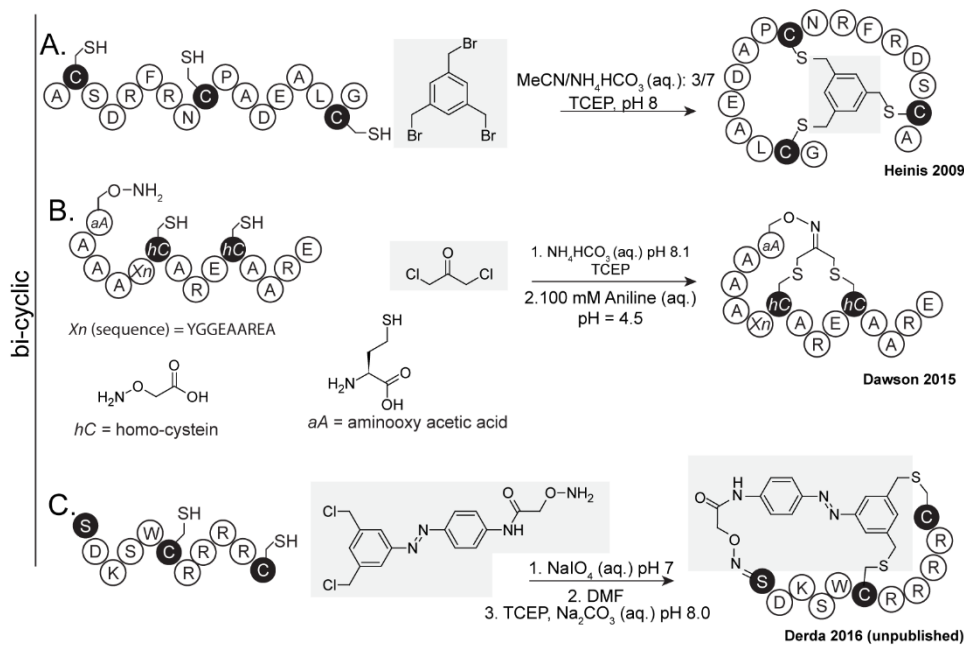
Figure 1.9. Selected alkyl-halide and benzyl halides used as linchpins for bicyclization.

24% of isolated product after 24 h. They later combined **DCA** and a linchpin scanning approach to show that the peptide sequence Ac-KETAA*hC*KFE*hC*QHMDS (where *hC* designates homocysteine) is active after cyclization with the linchpin, whereas neither the analog sequences with natural cysteines, nor its **DCA** conjugate were active.

In the canonical linchpin-scanning approach, it is not always known which conformation should be trapped or even what sequence is a useful binder. Heinis, Winter and co-workers used tri(bromomethyl)benzene (**TBMB**) to cyclize the peptide libraries at three cysteines to form libraries of bicyclic peptide with the structure of CX_6CX_6C displayed on phage (Figure 1.9A).[30] Screening of the library

20

against human plasma protease kallikerin identified peptides that could inhibit the protease activity with IC_{50} =1.7 – 39 nm only in bicyclic form. Non-modified peptides showed at least 250-fold decrease in activity. Heinis and co-workers subsequently combined library screening with linchpin scanning approach. They observed that in a library, different linchpins synergized with different set of sequences, and productive ligands for the same target contained a unique combination of peptide sequence and linchpins.[59]

All the linchpins mentioned above have static structures. Woolley and co-workers has used iodoacetamide and chloroacetamide functionalities to conjugate light-responsive azobenzene (**Azb)** linkers to peptides with cysteines at i, i+4, i, i+7 and i, i+11 (Figure 1.5E, 1.5F). These linkers allowed control of the helical content of the peptides with light (as described in the subsequent section "**1.3. Dynamic control of peptide conformation by light-responsive linchpins**").

In Chapter Four of this thesis, we describe a tridentate linchpin for double cyclization of a peptide with general sequence SX$_m$CRRRRC, where m = 1- 4. We first attached a tridentate azobenzene (**Azb**) linchpin to the serine-derived aldehyde at the *N*-terminus of the peptide and then formed two thioether bonds after reduction of the disulfide bond in the peptide (Figure 1.9C).

1.2.3. Linchpins that Undergo S$_N$Ar Substitution

Modification of Cys in proteins *via* S$_N$Ar reaction is well known. For example, 1-chloro-2,4-dinitro-benzene (**CDNB**) was commonly used for modification of
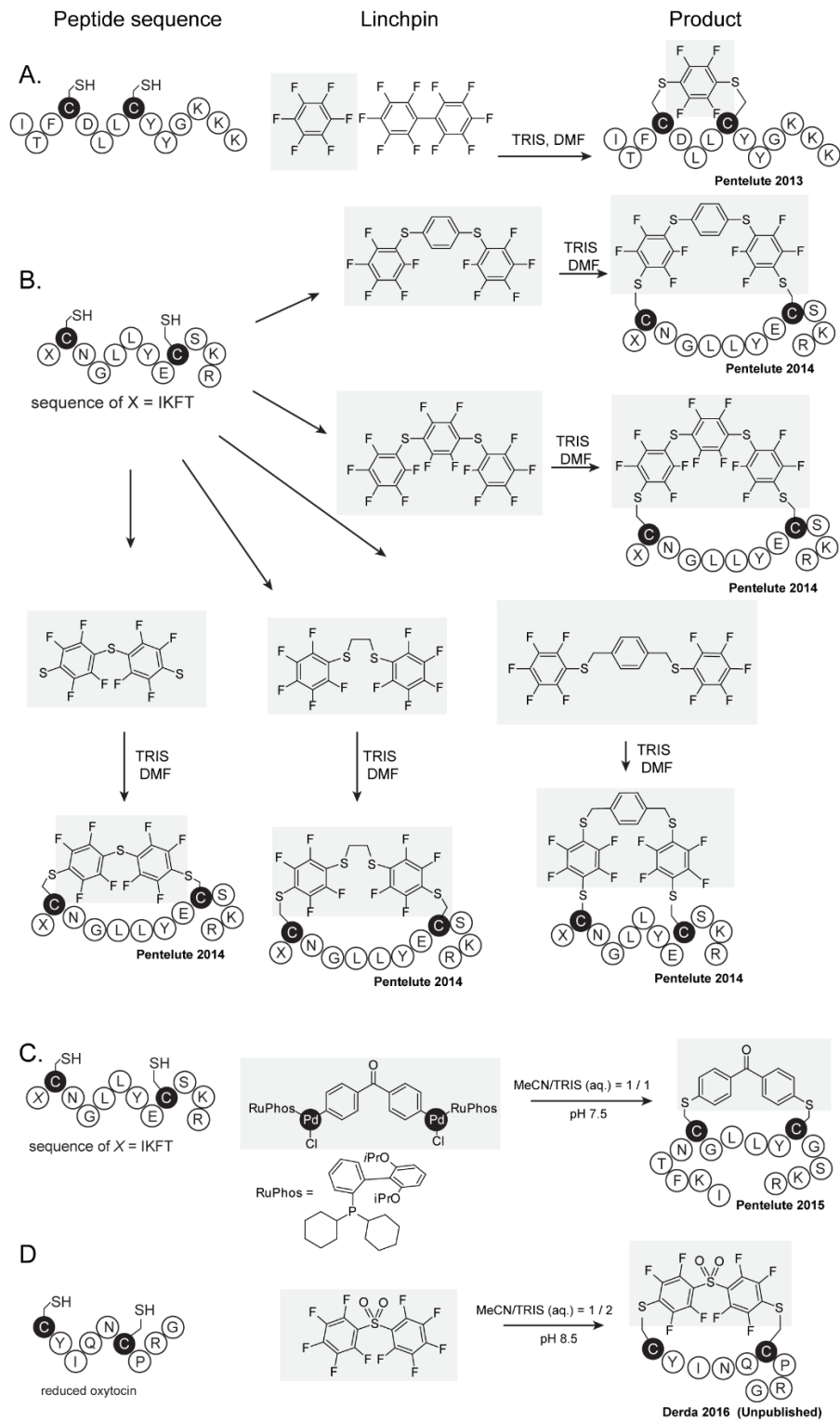
Figure 1.10. Linchpins that cross-link peptide *via* S$_N$Ar reaction with cysteines.

proteins in the mid-20th century.[60] High reactivity of perfluoro benzenes in S_NAr towards hydroxide, amine and thiolate nucleophiles was also noted by Tatlow as early as 1959.[61] Pentelute and co-workers built on these approaches to show that perfluorobenzene and its derivatives can serve as bidentate electrophiles for cyclization of peptides *via* S_NAr mechanism.[36] They first used hexafluorobenzene (**HFB**) and decafluorobiphenyl (**DFB**) to staple helical peptides at *i*, *i*+4 (Figure 1.10A).[36] Crosslinking the **C-CA** (peptide sequence ITF<u>C</u>DLL<u>C</u>YYGKKK, an HIV-1 capsid assembly polyprotein binder) with **HFB** increased the helicity from 16% to 53%, while the same stapling with longer linchpin **DFB** resulted in only 36% increase. Both stapled peptides showed increased binding to **C-CA** as determined by surface plasmon resonance. They also qualitatively showed that a mixture of trypsin and chymotrypsin degraded the linear peptide in 3 h, while no significant cleavage was observed for the **HFB**-stapled peptide and only partial degradation was observed for the **DFB**-stapled peptide. This example reiterated that peptide cyclization can increase the stability of the peptides against protease activity.

Pentelute and co-workers then tested five other variants of perfluoroarenes (**fAr**) on peptides with cysteines positioned at *i*, *i*+1 to *i*, *i*+14 (Figure 1.10B).[29] The stapling of cysteines at *i*, *i*+1, *i*, *i*+2, *i*, *i*+3 with **DFB** occurred in modest yield (7 – 52%), while more flexible **fAr**-linchpins exhibited somewhat higher yields (>59%) in reactions with all the peptides when stapling at *i*, *i*+1 through *i*, *i*+14. The authors hypothesized that conversions observed in these reactions reflect compatibility be-

tween end-to-end distance of the linchpin and the Cys-to-Cys distance in the peptide. Since all reactions were performed in pure DMF, the relevance of these observations to peptide conformation in H_2O is not clear.

As an alternative to perfluorenyl S_NAr linchpins, Pentelute and Buchwald developed a palladium-aryl reagent to arylate the unprotected thiols of cysteines in peptides and proteins (Figure 1.10C). The rate of arylation reactions was comparable to the thiol-maleimide Michael addition at pH 7.5 ($>10^3-10^4$ $M^{-1}s^{-1}$), although the authors did not measure the kinetics explicitly. Use of 5% organic solvent (MeCN) was sufficient for an efficient S_NAr substitution with labeling agents such as biotin and fluorescein; however, the authors had to use 50% of MeCN for cross-linking the cysteine with a bidentate linchpin to cyclize the peptide. The peptide sequence IKFTN<u>C</u>GLL<u>C</u>YESLR (10 µM) was stapled with 4,4'-dichlorobenzo-pheneone-palladium complex (20 µM) resulting in the formation of an arylated macrocycle in 10 min.[62]

Due to the limited solubility of most S_NAr reagents in water, S_NAr reactions should be conducted in pure organic solvent or a mixture of water with a high percentage of organic solvents. This solvent is needed simply to dissolve sufficient amount of perfluoroarene.[29,36] To overcome this limitation, Derda Group performed SAR analysis of thiol-perfluoroarene reactions in water and identified decafluoro-biphenylsulfone (**DFS**) as an effective reagent for cyclization of peptides (Figure 1.10D) with the sequence $X_mCX_nC\underline{X}_l$ (n = 1–15). The reaction proceeds rapidly in solutions containing as low as 30% acetonitrile or 5% DMF in aqueous Tris buffer with quantitative conversion in 5–20 min. To date, this reaction is the fastest S_NAr

reaction of perfluoroarenes and thiols with the rate of up to 180 $M^{-1}s^{-1}$. They found that the rate of the reaction is modulated by the sequence of the peptide. Positively charged amino acids such as Arg in the vicinity of cysteine increases the rate by increasing the basicity of the thiol and by stabilizing the carbanion intermediate. They used **DFS** to cyclize peptide hormones such as oxytocin, urotensin II, salmon calcitonin, melanine concentrating hormone, somatostatin and atrial natriuretic factor in which the number of amino acids between cysteines varied between 4–15.

1.2.4. Linchpins that Undergo Michael Reaction

Maleimides can undergo Michael addition with the thiols to result in a succinimidyl adduct.[63,64] Several types of bis-maleimides cross-linkers are commercially available. However the resulting product can undergo retro-Michael reaction and the succinimide moiety is prone to hydrolysis.[65] Another drawback is simultaneous formation of two stereocenters , and up to four diastereomers.

Pingoud and co-workers used an azobenzene bis-maleimide to attach two cysteins at *i*, *i*+14 in a single chain variant of PvuII (**scPvuII**) restriction enzyme containing a total of 157 amino acid residues. (Figure 1.11A).[66] The **scPvuII** consists of two identical halves that make it possible to include an azobenzene at each half, after substitution of Tyr49 and Asn62 with cysteines. They found that the doubly stapled *cis*-enzyme shows up to 16-fold higher activity compared to its *trans* variant. The *cis*-azobenzene can promote the formation of β-turn while the *trans* azobenzene distorts it.
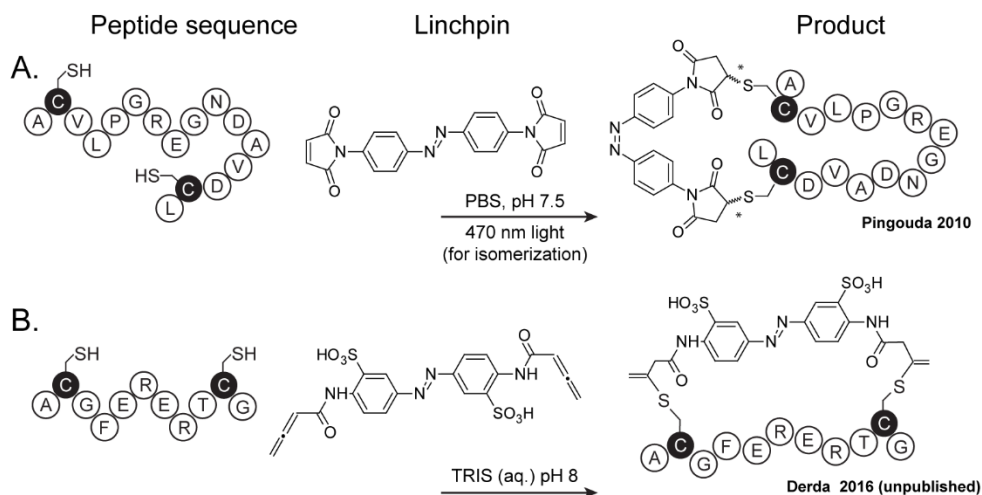
Figure 1.11. Michael acceptors used as peptide cyclization linchpins

In Chapter Three of this thesis we describe cyclization of a peptide sequence A<u>C</u>-GFERERT<u>C</u>G (a streptavidin binding peptide precursor) through Michael addition of the thiols to the LR-linchpin 3,3'-bis(sulfonato)-4,4'-bis(buta-2,3-dienoylamido)azobenzene (**BSBDA**). We show that the reaction proceeds with rate constant $k = 30$ M^{-1}s^{-1} at room temperature to produce the cyclic product (Figure 1.11B). We also used **BSBDA** to cyclize the peptide sequence NH$_2$-A<u>C</u>PARSPLE<u>C</u>GGG displayed on the coat protein of the M13 phage. The reaction yields 98% of cyclized phage in only 20 minutes.

Unlike classical maleimide adducts, allenamide adducts do not form chiral centers. They are also postulated to be more stable towards undergoing retro-Micheal addition and hydrolysis (personal communication with T. Loh, Nanyang Technological University), but the experimental evidence for the stability of such products is limited.

26

1.2.5. Linchpins that Undergo Azide-Alkyne [3+2] Cycloadditions

The copper catalyzed [3+2] cycloaddition of azide and alkyne (CuAAC) is a popular bioconjugation method because it is orthogonal to all unprotected side chains of natural amino acids.[67] After its introduction by Sharpless and Meldal[68] in 2002 many researchers have used this method to modify peptides, proteins, mult-protein complexes and cells.[69] Azidoalanines and propargyl glycines are used in CuAAC mediated cross-linking of synthetic peptides. The same amino acids can be also incorporated in ribosomally synthesized proteins through metabolic label-ing.[70] However, incorporation of unnatural amino acids in ribosomally synthesized peptides could be challenging.

Bong and co-workers used the CuAAC reaction to conjugate di-alkyne link-ers to peptides containing azido side groups as orthogonal handle (Figure 1.12A).[71] Placing the azides at i, $i+4$ resulted in helical stabilization by the linker hexa-1,5-diyne, but not with three other rigid structures ortho-diethynylbenzoic acid, meta-diethynylbenzoic and dipropargyl glycine linker.

Spring and co-workers used poly-arginine substituted dialkynyl benzenes to synthesize cyclic analogs of p53 protein, termed **SP1-SP5**, at i, $i+7$ (Figure 1.12B).[46] While all of the stapled peptides showed higher binding affinity (K_d = 6.7–44.3 nM) to MDM protein (comparing to non-stapled, K_d = 483 nm), only **SP5** with three Arg substitution could penetrate the cell. Peptides containing less than 3 arginine residues were ineffective penetrating the cell. *In vitro* assays showed that binding of **SP5** to MDM resulted in 7-fold increase in p53 activation.
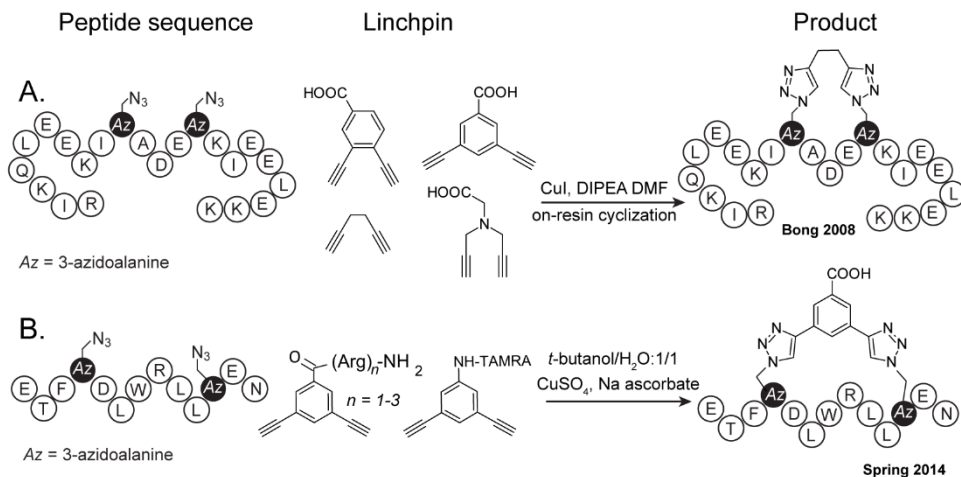
Figure 1.12. Linchpins using CuAAC to form cyclized peptides

1.2.6. Linchpins that Use Other Chemistries

In the past decades, many chemistries have been developed for chemical modification of unprotected amino acids in peptides and proteins.[54,72] However, examples of cyclization of peptides with chemistries other than the mentioned before are rare. Below, we review several linchpins that provide unique advantages not provided by conventional linchpin chemistry.

One such example is selective conversion of amino acids to orthogonal handles such as conversion of cysteine to dehydroalanine (**Dha**).[42] Michael addition of thiols to **Dha** results in formation of two stereoisomers. Webb and co-workers used this reaction to generate eight diastereomers of a bicyclic peptide in one reaction. They used 1,3,5-benzenetrimethanethiol as a C3$_V$ symmetric tridentate linchpin to cyclize a peptide through Michael addition to dehydroalanine produced out of cys-

teines in a linear peptide (Figure 1.13).[73] Cyclization of the sequence A*Dha*S-DRFRN*Dha*PADEAL*Dha*G resulted in a mixture of stereoisomers of **PK15**, a bicyclic peptide reported by Heinis *et al.* to bind tightly to plasma kallikrein protein (Figure 1.13).[30] Binding of the mixture of peptide stereoisomers (IC_{50} = 238 nM) was not higher than the original **PK15** (IC_{50} = 2.7 nM) in which all of the cysteine side chains were consisted of L amino acids. Synthesizing individual peptide isomers, showed no binding improvement (IC_{50} = 240 nM for the most potent binder) for any other diastereomers of **PK15**, but LLL-isomer.
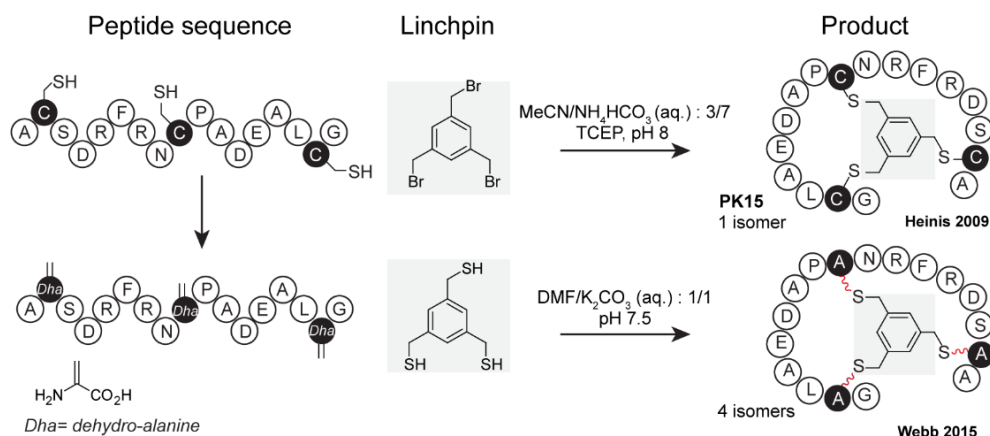
Figure 1.13. Two different C3$_V$ symmetric linchpins to make bicyclic peptides. S$_N$2 substitution of three thiols with bromobenzyls results in a single stereoisomer, while Michael addition of three thiols to **Dha**, produces eight stereoisomers.

Another example is a linchpin that uses reversible covalent bonds to cyclize peptides. Horne and co-workers studied the cyclization of peptides by forming ox-

Figure 1.14. A) oxime bond forming linchpins. B) template oriented cycliza-
tion of peptides: in the absence of receptor, the cyclization is incomplete, how-
ever, in the presence of receptor, the cyclization of peptide with *p*-phtaldehyde
results in exclusively cyclized peptide.

ime and hydrazide bonds (Figure 1.14). They incorporated aldehyde reactive un-
natural amino acids, containing hydroxylamine and hydrazide at *i*, *i*+4 and *i*, *i*+11
of a 17-mer peptide. They found that *m*-phtaldehyde can stabilize the helix when
conjugated through oxime bond at *i*, *i*+4 position, while stapling with three other

aldehydes (*p*-phtaldehyde, o-phtaldehyde and 4,4'-biphenyldicarbaldehyde) shows less stabilization. Conjugation through hydrazide bond resulted in less helicity comparing to oxime product.[47] These results are similar to the study by DeGrado which shows only *m*-**DBMB** can stabilizes the α-helix structure, while *o*-**DBMB** and *p*-**DBMB** do not show such stabilization.

The same group, in a separate work, showed that stapling the sequence Ac-WEEWD*Z*EIN*Z*YTKLIHKLIRE (where *Z* is an unnatural aminooxy carrying amino acid) with 1,4-phtaldehyde can be promoted in the presence of its receptor (HIV gp41-5 protein), suggesting that the receptor can act as a template to enhance macrocyclization (Figure 1.14B).[74] Unfortunately, this stapled peptide did not show increased helicity or improved affinity towards the protein. Binding affinity increased (4.5 fold) only after stapling with 2,5-thiophenedicarboxaldehyde which suggests the effect of the structure of linchpin on the binding, since this cyclization did not increase the helical content of the peptide as well. This work, however, suggests the potential of cyclization by using target receptor as templates for imprinting of a productive conformer. In such molecular imprinting techniques, the recognition sites in the protein can even dictate stereochemistry in cases the product of cyclization results in a mixture of isomers.[75]

The cross-linkers that react with identical functional groups in peptides are $C2_V$ or $C3_V$ symmetric. Asymmetric linchpins for cyclization of peptides require two or more sequential reactions in orthogonal conditions. The Fasan Group developed a strategy to cyclize peptides fused to *N*-terminal cysteine of intein GyrA, using two orthogonal reactive groups (Figure 1.15). Amber suppression was used

31

to genetically encode para-acetylphenylalanine (**pAcF**) in the peptide sequence which provides the ketone to react with hydroxylamine end of the linchpin. Upon oxime bond formation between the peptide and the linker, splicing of the intein completes the cyclization (Figure 1.15). They showed disruption of the interaction between p53 and HDM2 by a 12-mer peptide that was cyclized at *i, i*+11 position.[76] Although selective, this method has three potential drawbacks: (i) an unnatural amino acid has to be incorporated in the peptide sequence. Insertion of a serine at the *N*-terminus and selective oxidation, however, can overcome this problem. (ii) long reaction time (30 h vs 2 h for **TBMB**) (iii) demand for further workup such as capping the cysteine to avoid side product in downstream workflow.
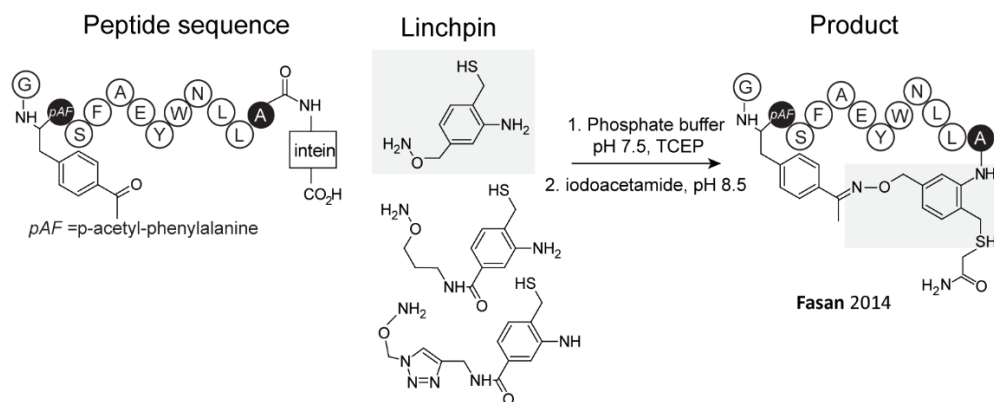
Figure 1.15. Non-symmetrical linchpins developed by Fasan and co-workers to cyclize a peptide after intein splicing.

1.3. Dynamic Control of Peptide Conformation by Light-Responsive Linchpins

Light-responsive (LR) linchpins have been used for dynamic control of the receptor ligand interaction.[77-79] LR-linchpins usually consist of substituted LR-"cores" that change conformation in response to irradiation with light. The conformational change (switching) can be a result of *cis*/*trans* isomerization (e.g. azobenzenes),[80] or formation/opening of a covalent bond (e.g. diarylethenes,[81-83] spiropyranes[84,85]). The electronic properties of substituents of the core often influence its photo-switching properties (e.g., switching wavelength, thermal relaxation half-life, photostationary state, etc.). These properties have to be taken into account when designing the reactive handles for the cyclization. Reactive groups can be more than two bonds apart from the photo-switching core to minimize the disturbance of photochemical properties. Alternatively, they can be designed to provide desired electronic properties after the reaction. For example, Hopmann and co-workers incorporated a fluorinated **Azb**-amino acid in calmodulin (CaM) protein. This moiety can undergo an intramolecular S_NAr reaction with a cysteine inserted in a mutant of the CaM. Such reaction, changed the maximum absorption of the **Azb** from 326 nm to 340 nm and improved the performance of the photoswitch (Figure 1.16).[86]

Azobenzenes and diarylethenes are $C2_V$ symmetric and are suitable for cross-linking of two reactive groups in peptides. Asymmetric LR-linchpins with two orthogonally reactive groups are also known. In Chapter Four of this thesis, I describe

a C2v symmetric linchpin for reaction with three amino acids functionalities to form a LR- bicyclic product.
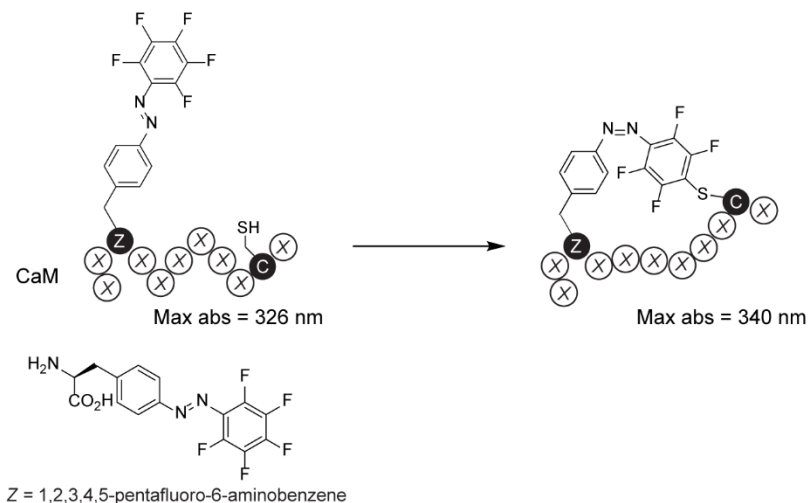
Figure 1.16. Photoswitching can be changed after bio-conjugation.

In this thesis, I will focus on **Azb** as the core of the photoswitchable linchpins for modulation of macrocyclic peptides and their application to form LR-peptides with single or multiple loops.

Although photo-switching ligands have been known since the 1950s, the history of LR-linchpins starts with a report of Woolley and co-workers who synthesized an azobenzene linchpin.[38] The sequence Ac-EA*C*ARV*Aib*AA*C*EAAARQ (where **Aib** is aminobutyric acid) modified by this linchpin at *i*, *i*+7 through alkylation of thiols, increased the helical content at 11 °C from 25% to 48% after light-induced isomerization to *cis* form, while decreasing it to 12% at dark photostationary state.

Woolley and co-workers subsequently released over 20 publications describing various upgrades to the properties of LR-linchpins (Figure 1.17, for in depth review see reference [77]). Notable structural modifications were (i) introduction of water solubilizing sulfonic acid groups in position meta to azobenzene (this location provided a minor interference with electronic properties of LR-linchpin). (ii) expansion of the span of LR-linchpins to 33 Å to identify optimal length and substituents that yield the most significant change upon the irradiation.[87-89] (iii) replacing the iodoacetamide group by a more chemo-selective chloroacetamide reactive group: despite the decrease in the rate of substitution, chloroacetamide was the functionality of choice in over 20 papers, because it can bypass unwanted reactions with amines exhibited by iodoacetamide.[89-92]

One of the many examples is the report by Woolley and Allemann that stapled the **HDH3** (a mutant of engrailed homeodomain binding peptide) with ***trans*-BSBCA** linker to increase its binding to DNA from K_d = 200 nM to K_d = 7.5 nM. After irradiation, the same cyclic peptide shows lower binding affinity (K_d = 140 nm).[93] Using **BSBCA**, LR-peptide ligands have been synthesized for various targets such as streptavidin,[28,94] PDZ domain of human tyrosine-phosphatase,[95] β-adaptin,[96] and RNA[97]. **BSBCA** has been used in 18 publications by 9 different research groups.

Although highly useful, the reported synthesis of **BSBCA** has an overall low yield (1.5% - 10%).[28,89,98] The substitution of the chloroacetamide group by thiols is also slow (k = 0.03 M^{-1}s^{-1}). In the third chapter of this thesis we describe a method to synthesize a new LR-linchpin (**BSBDA**) based on the **Azb** structure of
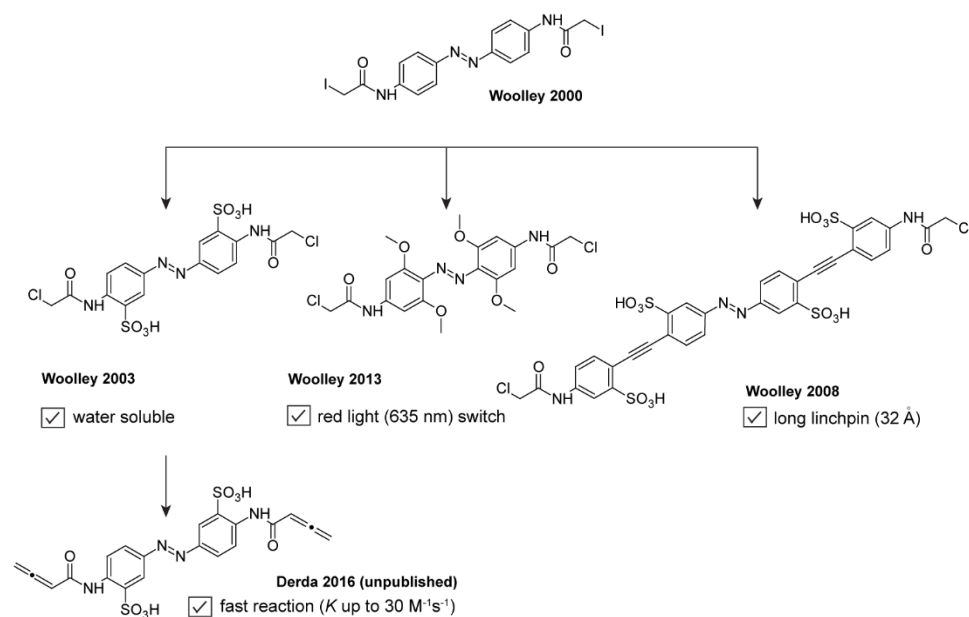
Figure 1.17. Azb-based linchpins developed by Woolley and Derda.

BSBCA. **BSBDA** is substituted with alleneamide functionality that can react up to 500-1000 times faster (k = 15-30 $M^{-1}s^{-1}$) than chloroacetamide with thiols. We describe a new route towards synthesis of water soluble **Azb** linkers which increases the yield of the final product to 39%.

Wooley and others described several SAR studies showing the interplay of substitution position (ortho, meta, para) and switching properties. These results can guide the design of LR-linchpins that have both optimal reactive groups and desired switching properties. Photochemical properties of **Azb** cyclized peptide can be controlled by substitution of **Azb** core with different chemical groups – such as amines, alkyls, amides, etc. – which affects the thermal relaxation half-life (0.25 ms – 43 h)[99,100] and various switching wavelengths (320 nm – 670 nm).[101,102] The Woolley

Group showed photo-switching of a fluorescent **Azb** cross-linked peptide, after injection to zebra fish embryo, by measuring the quenching of fluorescent label after *trans*-*cis* isomerization by red light (635 nm).[103] Isomerization by red light can provide a platform for photo-switchable peptide drugs, because of higher penetration of red light through the skin to the tissues.

1.4. Cyclization of Peptides with Linchpins for Selection Based Discovery of Ligands

Screening of a large library of peptides has become a major method for identification of lead compounds in drug discovery. However, libraries consisting linear peptides often yield binders with poor binding, pharmacokinetic and pharmacodynamic properties. Cyclization of the library members can build in desired properties in the binders identified by screening, including the potency of the binders,[30] proteolytic stability[36,46] cell permeability[26,36] and light-responsiveness.[28] We were the first group to show that LR-genetically encoded libraries of phage can be generated by conjugating **BSBCA** to the context of billions of peptides.[28] The ligands that are identified out of such LR-libraries show different binding to streptavidin in the presence and absence of light. Differences in K_d between *cis* and *trans* isomers of the ligand can reach 4.5 fold. A few months later, Heinis and co-workers used the same approach to find ligands with smaller ring size for the same target.[94]
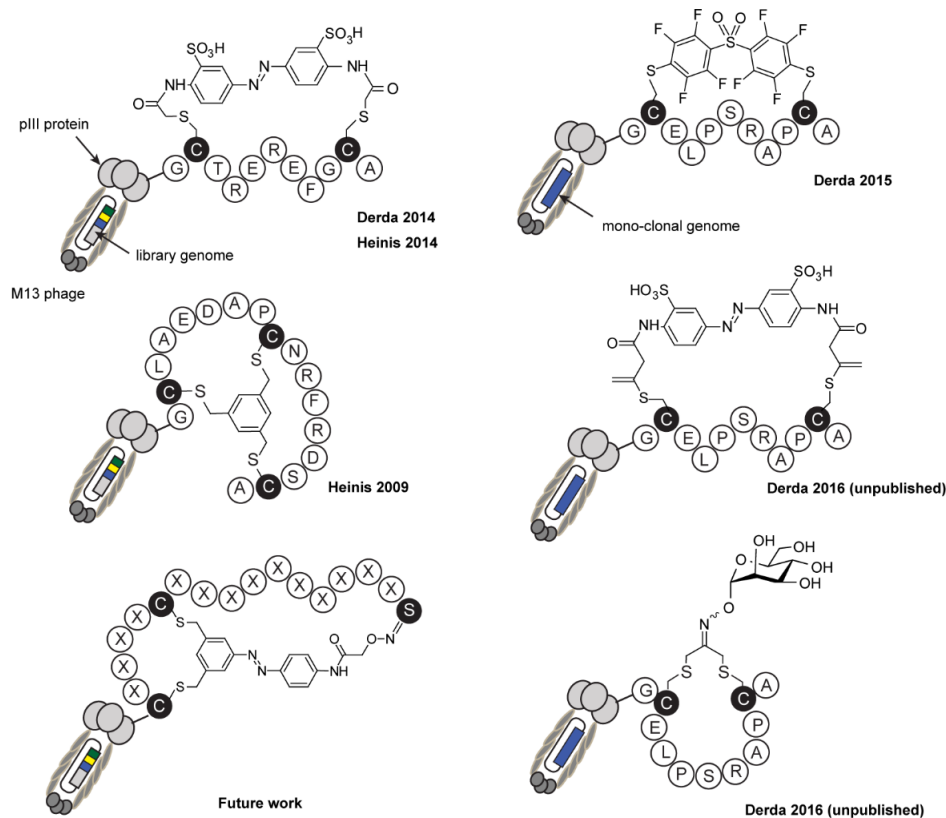
Figure 1.18. Linchpins cyclized peptides displayed on the coat protein of bacteriophage.

1.5. Scope of the Thesis

In this thesis I describe methods to cyclize linear peptide with both known and novel LR-linchpins with the goal of accelerated discovery of LR-macrocycles. In the second chapter, I describe a general method to identify LR-ligands by screening LR-phage display libraries. Since we used a library consisted of natural amino acids, the method can be adapted to any type of peptide libraries. The third chapter of this thesis describes the synthesis of **BSBDA**, a modified version of **BSBCA** LR-linchpin with increased reactivity towards thiols. We anticipate that the **BSBDA**

will be used widely in future research because of higher reaction rates and efficient synthesis. In the fourth chapter of this thesis, I describe synthesis of a novel LR-linker for generation of bicyclic LR peptides. The linker can react with peptides with a general sequence of SX_nCX_nC, after oxidation of N-terminal serine, to yield bicyclic product. The method is valuable because it adds a second ring to a LR-macrocycle which can increase the strain of the peptide or can be reserved for specific sequences such as polyarginine cell permeable motifs.

Chapter 2 Discovery of Light-Responsive Ligands through Screening of a Light-Responsive Genetically-Encoded Library[*]

2.1. Introduction

Discovery of light-responsive (LR) proteins has led to an explosive rise of "optogenetics", a technique that permits control of diverse cells with spatial and temporal resolution.[104,105] LR small-molecule inhibitors could provide a complementary tool for investigating cell function as they could be rapidly introduced into any organism without genetic manipulations.[106,107] Light-responsiveness has been successfully incorporated into natural proteins, peptides and small molecule ligands by rational design.[108-110] These ligands have been used to dissect biological systems with unprecedented spatial and temporal resolution to probe physiological responses that cannot be investigated using genetic approaches.[111-114] Although light-responsive ligands are powerful tools, design of such molecules is a difficult process that hinges on the availability of information about the structure of both ligand and protein target.[106] On the other hand, *in vitro* selection and evolution is a powerful approach that could yield ligands for targets that have no defined natural lig
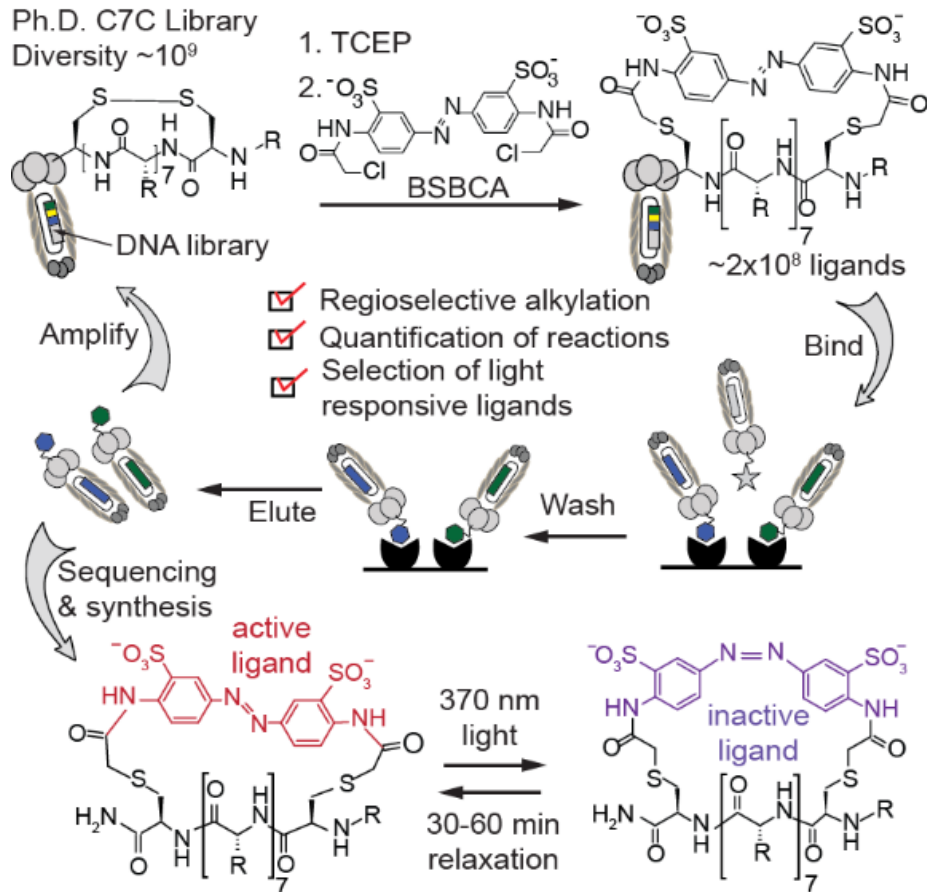
Figure 2.1. Synthesis of the chemically modified genetically-encoded library and general strategy for the selection of LR-ligands.

ands.[115] Herein, we describe the identification of LR-ligands through *in vitro* selection from a genetically-encoded library of LR-macrocycles. We synthesized this library from a genetically-encoded linear peptide library. (Figure 2.1)

Our selection procedure is based on phage display technology that makes it possible to trace and amplify any peptide sequence made of natural amino acids, even after chemical modification of the peptide.[115] In this technique, cloning a library of random nucleotide sequences in the gene encoding the phage coat protein

leads to expression of the random peptides as fusion to *N*-terminus of the coat protein pIII.[116-118] Libraries with complexity of $> 10^9$ members can be produced or purchased from the commercial sources. Functional peptide ligands can be identified from these libraries through rounds of selection (panning) and re-amplification of the phage library. To allow the selection of moieties that cannot be encoded by natural amino acids, we and others have used chemical post-translational modification (cPTM) of phage displayed peptides to yield a genetically encoded library of peptide-derivatives (reviewed in references [119,120]).

Examples of phage libraries modified by cPTM are alkylation of CX_6CX_6C libraries to create bicyclic peptides,[121,122] oxidation and oxime ligation of NH$_2$-(S/T)X_6 libraries to yield glycosylated peptides,[123] hydrazone ligation,[124] intein chemistry,[125] and modification of the Cys-containing peptides to introduce glycans[126] or other known ligands.[127] Non-covalent grafting *via* coiled-coil heterodimer have also been used to select ligands from modified phage libraries.[128] The aim of many chemical modifications is to constrain the peptides by a synthetic linker to decrease conformational flexibility of the ligands and increase their stability to peptidases (for a recent review see[129]). Chemical cyclization has been used extensively in mRNA-displayed peptide libraries.[130-136] In this work, we used cPTM to constrain a linear peptide sequence displayed on phage with a well-characterized LR azobenzene (Azb) linker, developed by Woolley and co-workers.[98] This linker has an end-to-end distance of 16.5–18 Å in the *trans* conformation and 11–16 Å in the *cis* conformation; change in the conformation of the Azb is known to change the conformation of the macrocycle constrained by this linker.[137]

To allow for site-specific cyclization, we used a phage display library in which a heptamer random peptide sequence is flanked by two cysteines. The reduction of the disulfide bond and cyclization by 3,3'-bis(sulfonato)-4,4'-bis-(chloroacetamido) azobenzene (BSBCA) yielded a library of peptide macrocycles constrained by Azb. Alkylation of Cys proceeds with high selectivity in the context of functionality-rich phage particles,[121,122,126,136] and removes the need for unnatural bio-orthogonal handles such as seleno-cysteines[138] or phenylalanine-azide.[139] The resulting macrocycles have relatively small size and their conformation could change significantly in response to light-driven *cis/trans* isomerization of the linker (see below). Two classes of LR-ligands can be selected from this library: (i) ligands that have low binding affinity in the ground state and increase binding affinity upon irradiation ("light-activated ligands");[140-142] (ii) ligands that bind to the receptor in their ground state (dark), and lose their binding in response to light ("light-inhibited ligands")[143,144]. In this work, we perform selection with ligands in their ground state and identified "light-inhibited ligands". By performing the screening in the presence of light, it might also be possible to identify "light-activated ligands". Four months after we published this manuscript, Heinis and co-workers described similar selection that identified "light-activated ligands".[94] Selection of ligands from chemically-modified phage libraries is similar to selection from unmodified phage libraries.[121] It involves rounds of (i) chemical modification of phage library, (ii) binding to target, (iii) elution of binders, and (iv) re-amplification (Figure 2.1). Repeating steps 1 – 4 enriches the binding phage. After rounds of selection, sequencing the DNA of the phage reveals the identity of amino acids within the modified

peptide. Screening of the libraries that are modified through cPTM, however, has one caveat: modification of the peptide sequences has to have high yield. Selection from sparsely modified library is unlikely to yield desired results. We recently demonstrated that the yield of cPTM could be systematically optimized on phage or libraries of phage.[123] In this chapter we use a similar technique to optimize the yield of synthesis of genetically-encoded LR-libraries with $\sim10^9$ members, and used these libraries to select target-binding ligands that can be reversibly turned on or off by visible light.

2.2. Results and Discussion

2.2.1. Modification of Phage-Displayed Peptides by Azb Linker

We used a commercially-available M13 phage library of Cys-flanked random heptapeptides (Ph.D.™ C7C, New England Biolabs) in which the thiols are oxidized to form a cyclic disulfide during the expression in the *E. coli* host.[115] Prior to alkylation, the disulfide bond has to be reduced to generate free thiols (Figure 2.2A). The regioselectivity of this reduction is important because M13 phage has four more disulfides inside pIII protein and disruption of these "internal" disulfides leads to loss of phage viability.[145] Immobilized tris(2-carboxyethyl)phosphine (iTCEP) effectively reduced only sterically-accessible disulfide on pIII, while leaving the sterically-hindered internal disulfides intact (Figure 2.2C, Figure 2.3A, 2.3F). Infectivity of phage remained $\sim$90% in the presence of iTCEP (Figure 2.2D), whereas non-immobilized TCEP decreased phage viability to $\sim$60% in 30 min (Figure 2.2E).
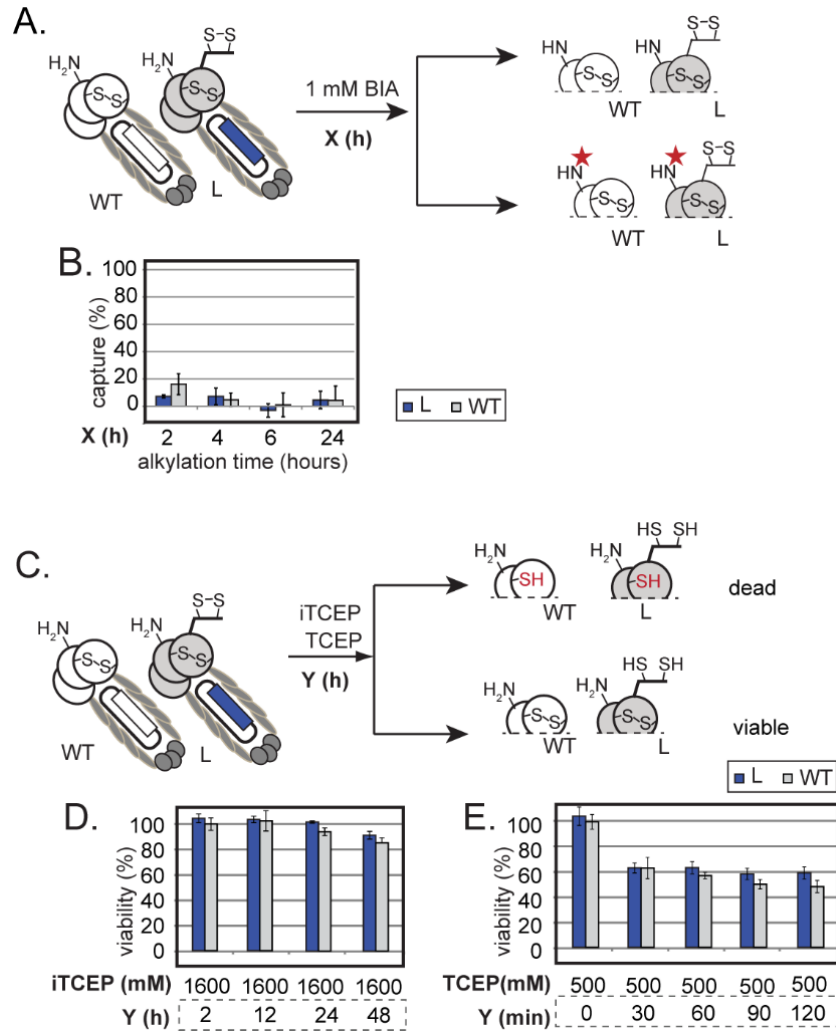
Figure 2.2. Quantification of viability of phage after disulfide reduction. (A) Reaction of the phage with biotinylating reagents in the absence of reducing agent. (B) No biotinylation was observed on either clones in the absence of reducing agent. (C) Introducing the reducing agent can cleave the exposed disulfides as well as internal disulfides that are vital for phage infectivity (D) Using the iTCEP, the phage remains infective through 48 h of reduction in 4°C. (E) Treating the phage with soluble TCEP dropped the viability to 60% in 30 min. (The use of soluble TCEP up to 500 μM was necessary to result in effective reduction of phage (~99%).
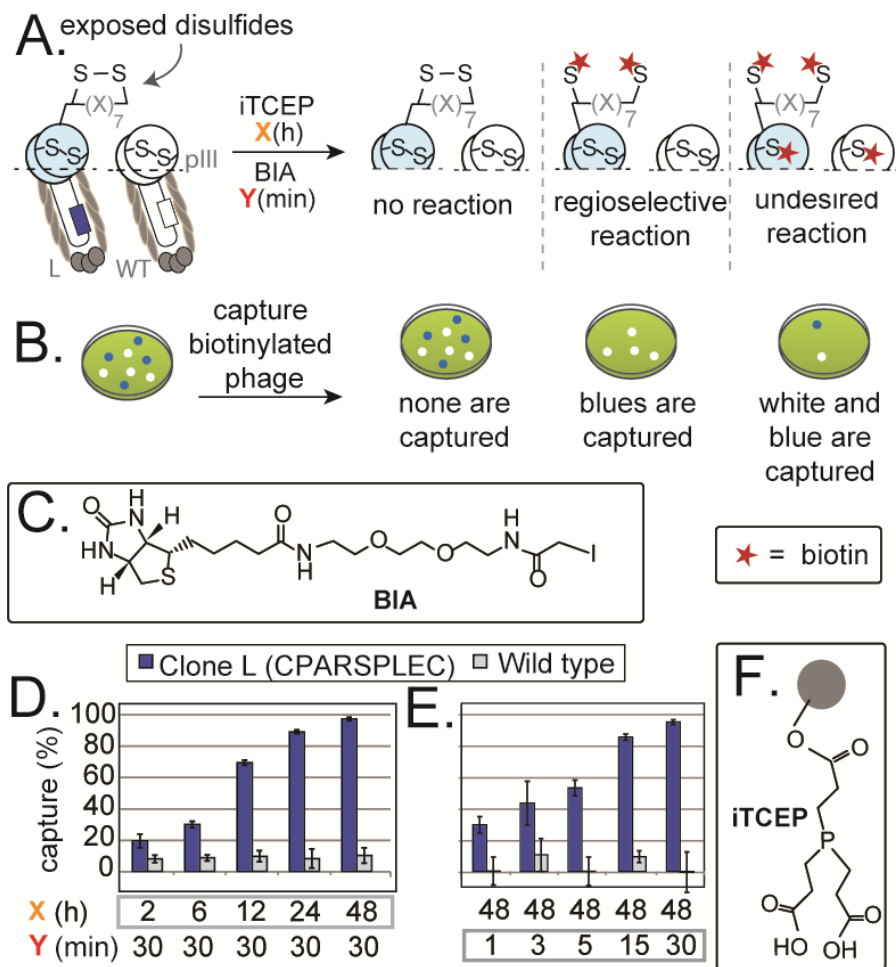
Figure 2.3. Quantification of reduction of exposed disulfides on phage (A) We optimized the regioselective alkylation of phage-displayed peptide using a mixture of phage types: Clone L (ACPARSPLEC) and WT. (B) Scheme describes the outcome of incomplete reduction (no reaction), regioselective and non-regioselective reductions. Capture of biotinylated phage by streptavidin and titration of the remaining (non-biotinylated) phage allows quantifying the yield and regio-selectivity of the reduction. Clone L and WT form blue and white plaques, respectively, in X-gal agar. (C) Structure of thiol-reactive agent **BIA**.

Figure 2.3. (continued)

(D) Reduction of Clone L proceeds to~ 99% in 48 h. (E) Biotinylation of reduced phage reaches ~99% in only 30 min. (F) Structure of the reducing agent iTCEP.

To quantify the yield and regioselectivity of reactions, we used a previously published method based on the labeling of the reactive functionalities by a "capture agent".[123]

Here, we used biotin-PEG$_2$-iodoacetamide (**BIA**) (Figure 2.3C) to label reduced thiols and quantified biotinylation by counting the number of phage before and after capture with streptavidin (Figure 2.3B). To assess the specificity, we performed all reactions on a mixture of two phage clones: (1) A phage displaying AC-PARSPLEC sequence on pIII and containing LacZ reporter in the genome (L-clone) and (2) an insert-free and reporter-free wild type (WT) M13 phage. L and WT phage formed blue and white plaques, respectively, in bacterial lawn in the presence of the colorigenic galactosidase substrate X-gal. Counting the number of blue plaques before and after reaction allowed us to determine the yield of reduction (and biotinylation) of "external" disulfides present only in L-phage. Tracking the white plaques allowed quantifying the reaction of "internal" disulfides present in both L-phage and WT-phage (Figure 2.3B)
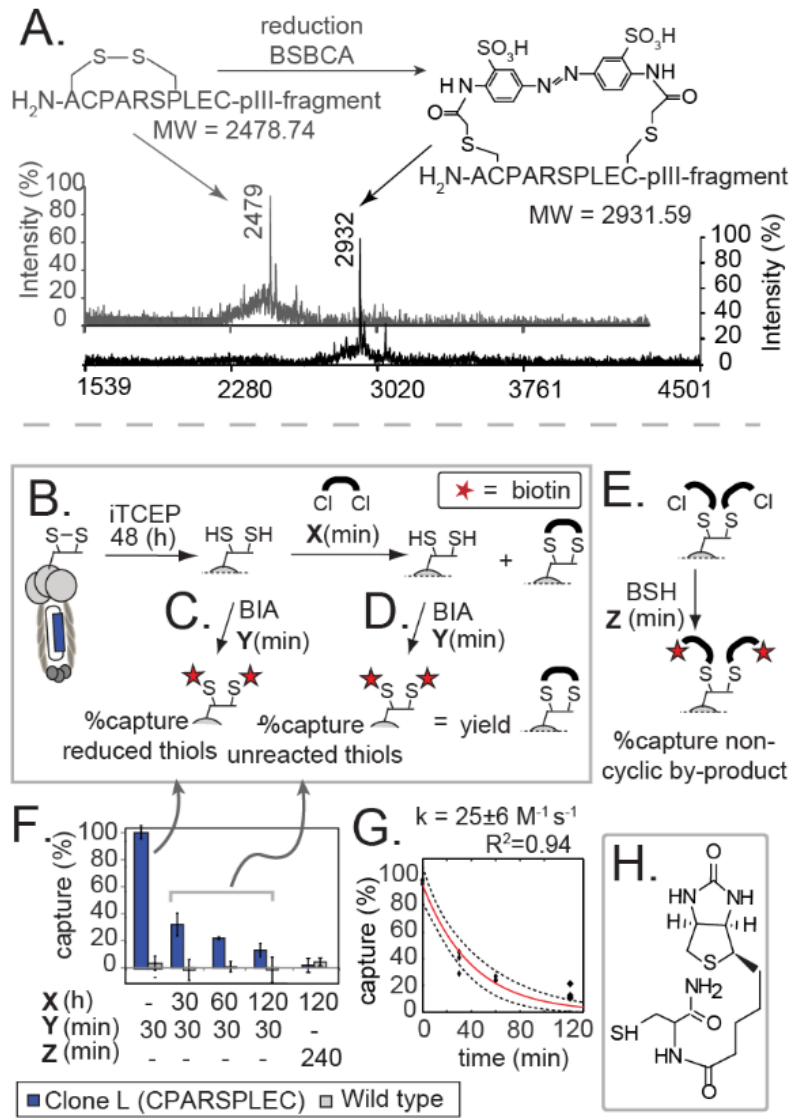
Figure 2.4. Alkylation of the representative peptide and the phage with BSBCA

(A) Mass spectra of peptide (ACPARSPLECGGGSAETVESC(Cam)LAKS) before and after modification with BSBCA. This synthetic peptide is identical to the sequence displayed on the *N*-terminus of pIII protein of phage Clone L. A 453 Da difference in mass shows that the major product is the cyclized peptide. (B) Reducing the phage yields the precursor for alkylation reaction by photo-switch (BSBCA). (C) The efficiency of reduction can be measured through alkylation by **BIA**.

Figure 2.4. (continued)

(D) The reduced phage that did not react with photo-switch linker (BSBCA) can also be quantified by **BIA**. The difference between (D) and (C) is the yield of alkylation by BSBCA. (E) Non-cyclized phage could react with biotin-thiol (**BSH**). (F) Quantification of the alkylation of Clone L by BSBCA. (G) The rate of reaction of phage with BSBCA is similar to the rate of reaction of corresponding peptide with BSBCA (See Figure 2.5). (H) Structure of **BSH**.

We observed that the yield of reduction of L phage by iTCEP reached ~99 % after 48 hours (Figure 2.3D). Alkylation of the reduced phage by **BIA** was fast, yielding 99% biotinylated phage in ~30 min (Figure 2.3E). WT phage clones present in the same solution were not significantly biotinylated (grey bars in Figure 2.3D, 2.3E). Reduction and alkylation, thus, occurred exclusively on AC-PARSPLEC sequence and internal cysteines in pIII protein and non-Cys nucleophiles were not alkylated (Figure 2.3D, Figure 2.2D). As expected, we observed no significant biotinylation in the absence of the reducing agent (Figure 2.2A, 2.2B).

We used a synthetic fragment of pIII peptide as a model to show that alkylation of ACPARSPLEC sequence by BSBCA photo-switch yields the desired cyclic product (Figure 2.4A). High yield of cyclization was consistent with the previous work by Woolley and co-workers who confirmed that reaction of bis-thiol peptides with BSBCA yielded predominantly cyclic peptides.[98] To quantify the yield of alkylation on phage we used the following "pulse-chase" approach: Reduced phage was "pulsed" with 1 mM BSBCA for a specific time and "chased" with an

excess of **BIA** to cap the unreacted thiols by biotin (Figure 2.4B-D). Decrease in biotinylation upon exposure to BSBCA indicated that the free thiols on pIII-displayed peptide reacted with the linker (Figure 2.4F). The rate constant of this process for two different peptides displayed on phage was 17-25 $M^{-1}min^{-1}$; this value was similar to the rate of the reaction between BSBCA and reduced peptides measured by HPLC (Figure 2.4G, Figure 2.5). Non-cyclized peptide can also be detected on phage by exposing the phage to a biotin-thiol conjugate (**BSH**) (Figure 2.4H), which could cap any unreacted chloroacetamide moieties (Figure 2.4E). We observed that the reaction with **BSH** was minimal, indicating that the majority of BSBCA-alkylated peptides are in the cyclic state (Figure 2.4F).

2.2.2. Affinity Selection of LR-Library

To demonstrate the application of BSBCA-modified library, we performed three rounds of affinity selection (panning) against streptavidin as a model substrate. Prior to each round of selection, we reduced the library using iTCEP, alkylated it with BSBCA and quantified the yield of modification on the modified library of peptides using the same **BIA/BSH** capture as described above. Quantification suggested that only ~50% of the clones incorporate BSBCA linker after two hours of reaction; the majority of the reacted phage were in a cyclized state (Figure 2.6A, 2.6B). The 40–50% yield was consistently observed at each step of the selection process (Figure 2.6C). After each round of selection, we amplified the phage in mono-disperse emulsions to avoid enrichment of phage clones with growth preferences
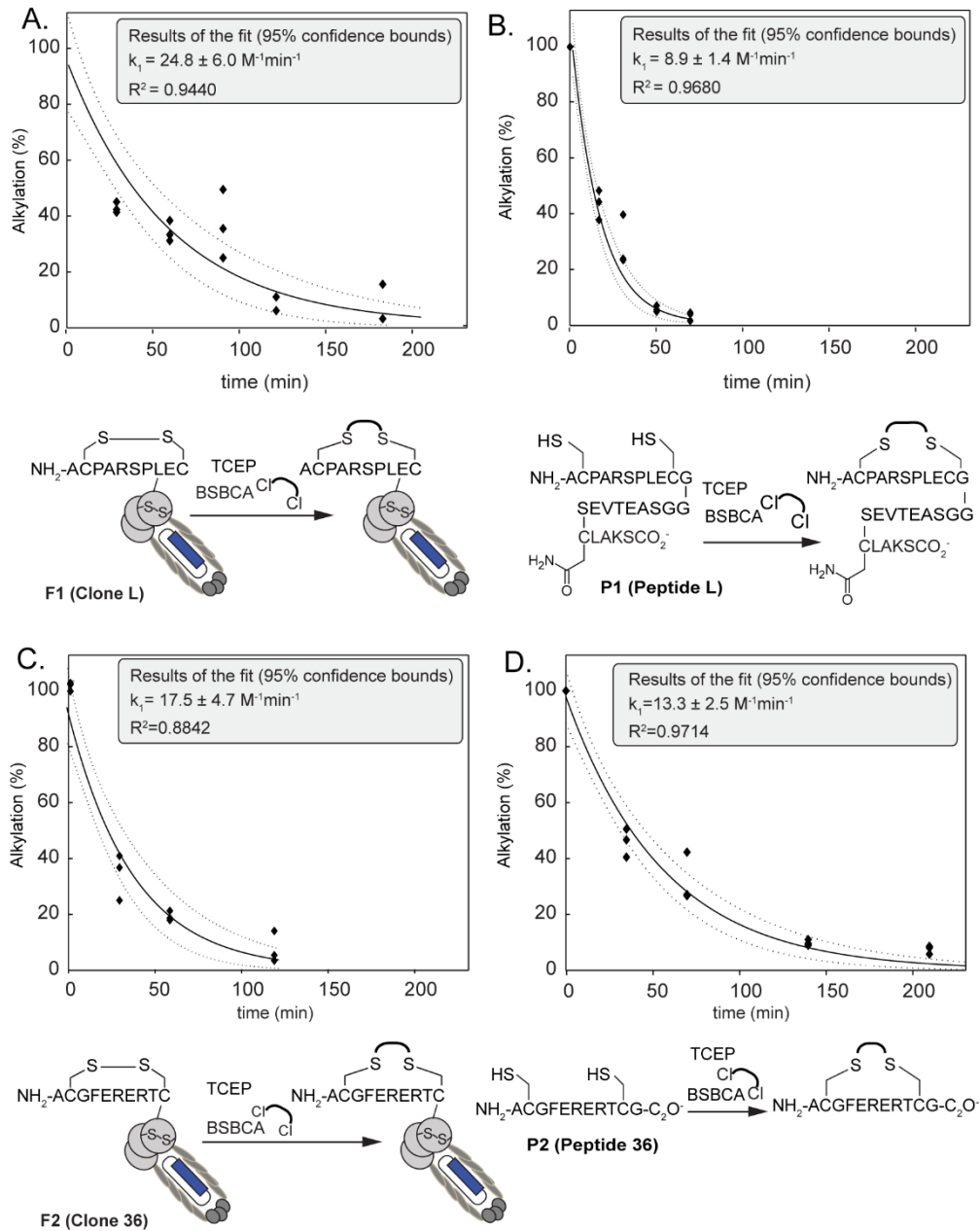
Figure 2.5. Kinetics of reaction of phage and peptide with photo-switch linker (BSBCA). Solid line shows the fit. Dashed line shows 95% confidence bounds. (A) Kinetics of reaction of phage F1 (clone L, 10^{11} PFU/mL) with BSBCA (1 mM). (B) Kinetics of reaction of P1 (peptide L, 0.2 mM) with BSBCA (2 mM).

Figure 2.5. (continued)

(C) Kinetics of reaction of phage F2 (clone 36, 10^{11} PFU/mL) with BSBCA (1 mM). (D) Kinetics of reaction of P2 (peptide 36, 0.6 mM) with BSBCA (2 mM). In all fits, we used pseudo-first order kinetics, in which the rate $d[P]/dt = k[BCBCA][P]$, could be approximated as $d[P]/dt = k'[P]$, where $k'=k[BSBCA]_0$ is a pseudo-first order rate constant, k is a true second-order rate constant and [P] is concentration of phage or peptide. This approximation assumes that concentration of BSBCA does not change over the course of the reaction, which is true when BSBCA is present in a large excess compared to peptide of phage. The reported rate constants are second-order rate constants (k).

(Figure 2.7).[146,147] As a control, we performed three rounds of panning with unmodified C7C library using the same target.

As modification of the library was only ~50%, selection from such a library could yield ligands that bind to the target without BSBCA modification. Therefore, after the three rounds of selection we randomly picked 80 clones and performed sequencing to reveal the sequence of randomized peptides on the phage. Out of 80 clones, 32 clones contained no insert. We tested binding of the other 48 clones to streptavidin before and after BSBCA modification. The screen did not converge on one peptide sequence; however, we used ELISA to identify three clones that exhibited binding only after modification with BSBCA linker (Figure 2.8-2.9). A control panning of the unmodified library converged on sequence CTSHQPRVC (**p81**) with a well-known streptavidin binding motif.

Figure 2.6. Modification of the phage library and peptides with BSBCA. (A) Efficiency of coupling and cyclization of BSBCA linker to naive C7C library.

Figure 2.6. (continued)

(B) Composition of the library after modification can be represented as a stacked bar. (C) Stacked bars show composition of the library before each round of the selection. Non-cyclized library was not detected in any of the rounds of modification and therefore is not shown. (D) Synthesis of LR-ligand **L36**. Precursor ACGFERERTCG peptide was made by solid phase synthesis.
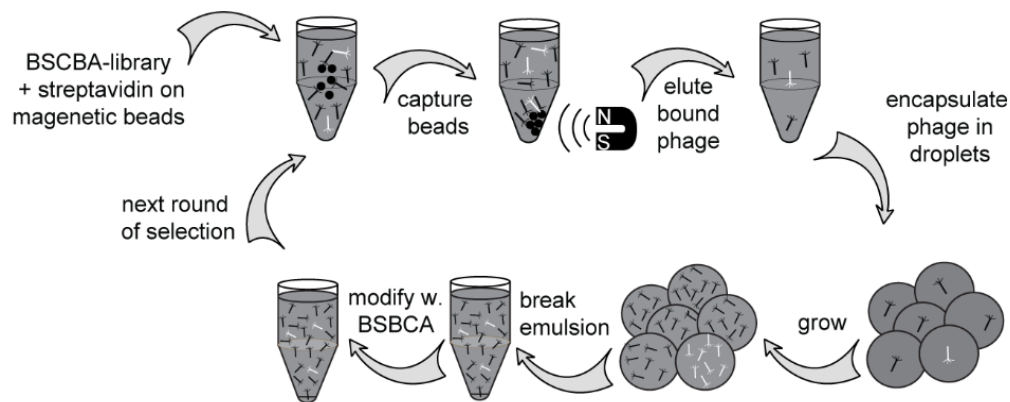
Figure 2.7. Screening procedure for LR-library. We synthesized the LR-library before each round of panning. Amplification of the phage in droplets can prevent outgrow of the clones that have growth preference.
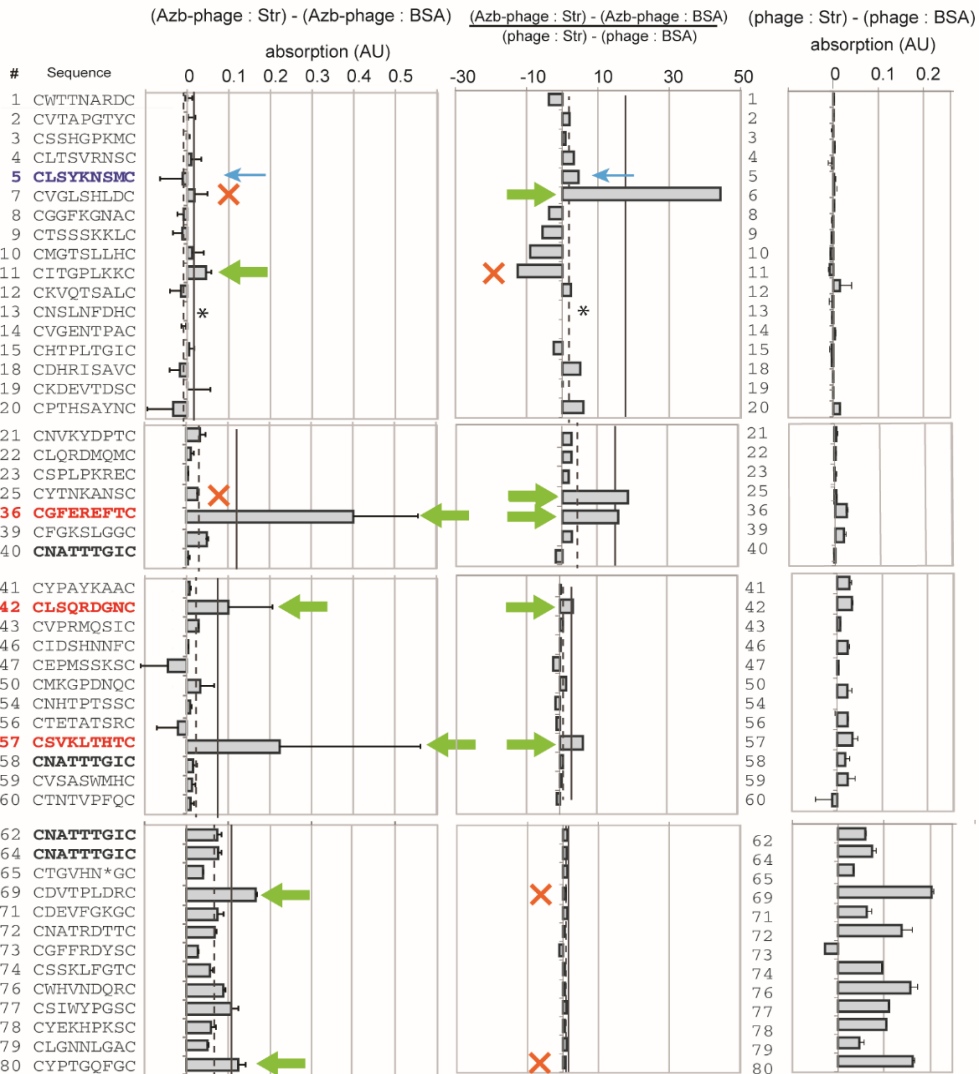
Figure 2.8. ELISA results of the selection against streptavidin. We randomly picked 80 phage clones, measured their binding to streptavidin-coated plate using phage ELISA and sequenced the DNA of all clones. Out of 80 clones, 32 clones contained no insert and were removed from analysis. Four clones displayed the same sequence CNATTTGIC (bold). All ELISA experiments were performed in duplicates.

(Column 1): We determined binding of BSBCA-modified phage (Azb-phage) to streptavidin-coated plate and subtracted binding of the same phage to BSA-

Figure 2.8. (continued)

streptavidin and subtracted binding of the same phage to BSA coated plate (column 3). Column 2 displays the ratio of the values from column 1 and column 3 (i.e., increase in binding upon modification). We used two criteria to select the candidate clones. First: binding of modified phage to streptavidin (column 1) has to be higher than plate average (AVE) + standard deviation (STD) (green arrows in column 1). Second: increase in binding upon modification (column 2) has to be higher than AVE + STD (green arrows in column 2). Five clones satisfied one criteria but not the other (as indicated by a red cross). Three clones satisfied both criteria and were tested in the downstream assays. Their sequences are coloured red. We also used CLSYKNSMC clone (blue arrow) as negative control in the downstream assays. The outliers that exhibited the signal outside of AVE ± (3 × STD) were not used in the calculation of the population average. These clones are indicated by (*).

2.2.3.   Characterization of Binding by ESI-MS Binding Assay.

We synthesized the peptides displayed by positive clones (designated **p36**, **p42** and **p57**, Table 2.1) and modified them with BSBCA to yield ligands designated as **L36**, **L42** and **L57** (Figure 2.6D, Figure 2.17). Using ESI-MS binding assay,[148,149] we quantified the dissociation constant (K_d) of the interaction of these ligands to streptavidin. The K_d value was calculated from the intensity of streptavidin tetramer [S_4] and its complex with the ligand [S_4L] (Figure 2.10A).[150] In line
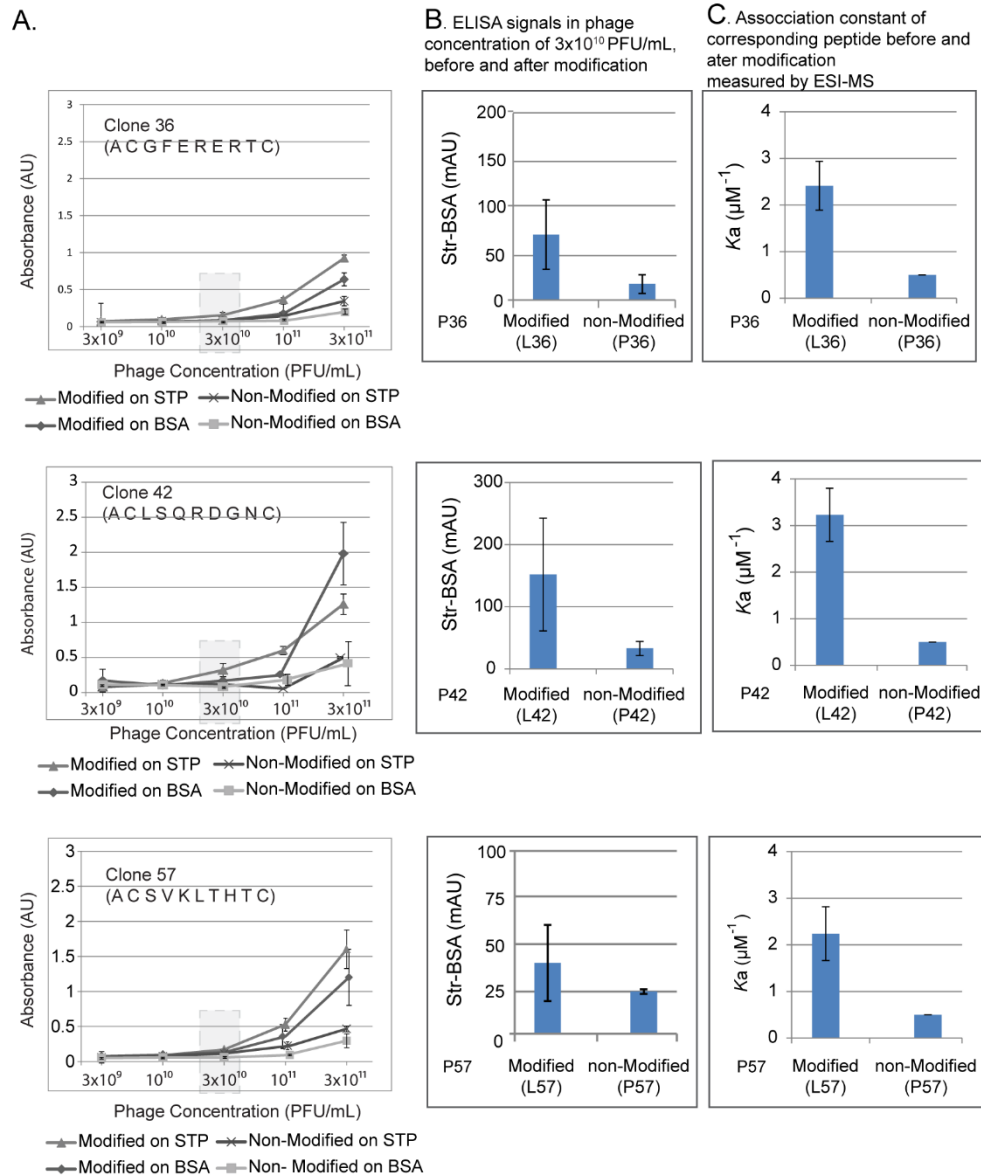
Figure 2.9. Binding of three positive clone to streptavidin (STP) and BSA before and after modification. Each clone (3×10^{12} PFU/mL, 500 μL) was alkylated with BSBCA and purified by precipitation from PEG/NaCl and re-dissolved in 200 μL of ddH$_2$O. We added a serial dilution of the modified and non-modified phage in binding buffer (0.1 % Tween 20, 0.05% BSA in PBS) to
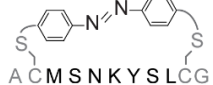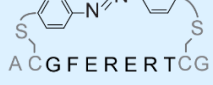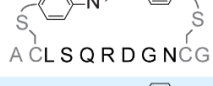
Figure 2.9. (continued)

plates pre-coated with streptavidin and performed ELISA as described in section **2.4.7**. All the tests were run at least in triplicates. A) ELISA signal (absorption of peroxidase substrate) measured for different concentration of phage. B) Comparison of ELISA signal for modified and non-modified phage on streptavidin and BSA coated wells that contained 3×10^{10} PFU/ml of phage C) The association constant of the corresponding peptide measured by ESI-MS. Correlation of ELISA results and MS results shows that the ELISA is the most accurate at phage concentration of 3×10^{10} PFU/ml, while signal at higher concentration of phage deviates from ESI-MS due to non-specific binding of phage.

with the observations in ELISA, only BSBCA-cyclized peptides **L36**, **L42** and **L57** exhibited binding to streptavidin and none of the non-BSBCA modified peptides (**p36**, **p42** and **p57**) had any detectable binding. (Table 2.1, Figure 2.11A). In contrast, peptide **p81**, which was selected from unmodified library, exhibited similar binding to streptavidin, both in the absence of BSBCA modification and in the presence of the modification (Table 2.1). To confirm that BSBCA does not promote non-specific interactions, we tested the peptide **p5** (CLSYKNSMC) identified as a "non-binder" in ELISA (Figure 2.8, clone 5). Neither **p5** nor its BSBCA-derivative **L5** exhibited any significant interaction with streptavidin in ESI-MS assay (Table 2.1).

All BSBCA-cyclic peptides underwent *cis-trans* isomerization upon irradiation with 370 nm light as confirmed by a rapid decrease in absorbance at 370 nm

Table 2.1. Dissociation constant (K_d) for the complexes of streptavidin and ligands measured by ESI-MS at pH 6.8 and 25℃.

Non-Modified Peptide			BSBCA Modified Peptide (light-responsive)				
Peptide	Structure	K_d^t (µM)	Ligand	Structure	K_d^{dark} (µM)*	K_d^{light} (µM)§	ΔK_d¶
P5	A C M S N K Y S L C G	> 2000	L5	A C M S N K Y S L C G	> 2000	> 2000	N/A
P36	A C G F E R E R T C G	> 2000	L36	A C G F E R E R T C G	452 ± 63	> 2000	>4.5
P42	A C L S Q R D G N C G	> 2000	L42	A C L S Q R D G N C G	368 ± 103	814 ± 55	2.2
P57	A C S V K L H T H C G	> 2000	L57	A C S V K L H T H C G	506 ± 89	1223 ± 201	2.4
P81	A C T S H Q P R V C G	158 ± 76	L81	A C T S H Q P R V C G	229 ± 21	239 ± 41	1.0

K_d^{dark} is the apparent K_d claculated from ESI-MS data for a solution before irradiation

K_d^{light} is the apparent K_d claculated from ESI-MS data for a solution after irradiation

ΔK_d is calculated as K_d^{light}/K_d^{dark} which reflects the apparent decrease in binding.

(Figure 2.12B). This process was reversible and thermal relaxation of the *cis*-form back to *trans*-form followed first-order kinetics with half-lives of 37.3 ± 1.4 min (**L36**), 40.1 ± 0.6 min (**L42**), and 32.5 ± 3.9 min (**L57**) (Figure 2.13, Table 2.2). To measure the change in binding upon irradiation, we irradiated the ligands with 370 nm light for 10 minutes, added streptavidin solution and incubated the mixture under irradiation for another 30 min (Figure 2.14). Then, we measured the ratio of [S_4] to [S_4L] by ESI-MS within 1–2 minutes after irradiation (i.e., significantly faster than the *cis/trans* relaxation time). All ligands exhibited a significant change in K_d value upon irradiation (Table 2.1). **L36** showed the most dramatic change
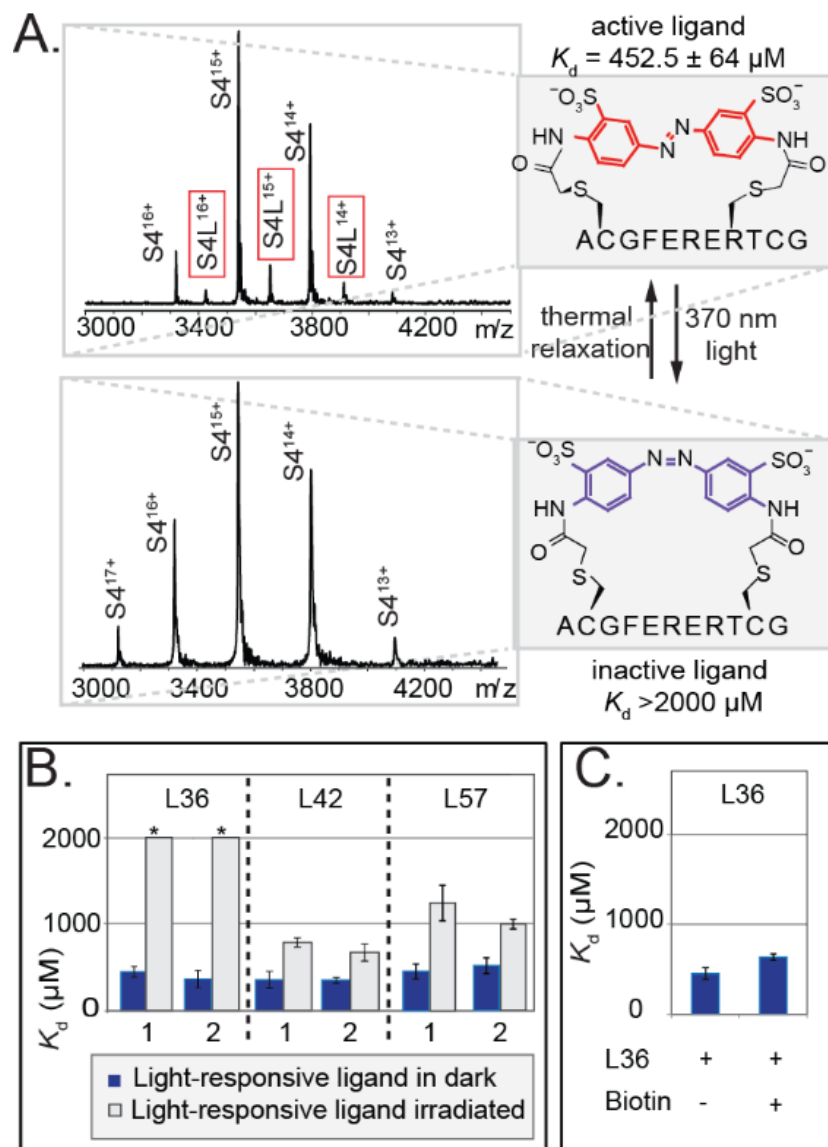
Figure 2.10. (A) ESI-MS analysis of streptavidin (S₄) in the presence of **L36** (L). No complex ions (S₄+L) were observed in the spectrum after irradiation with 370 nm light. (B) Binding assay for light-responsive ligands showed that iterative irradiation of **L36** and **L42** did not cause a significant change in photo-switching ability of the ligands. The asterisk (*) indicates that no complex was observed. (C) Binding of **L36** to streptavidin does not change significantly in the presence of biotin.
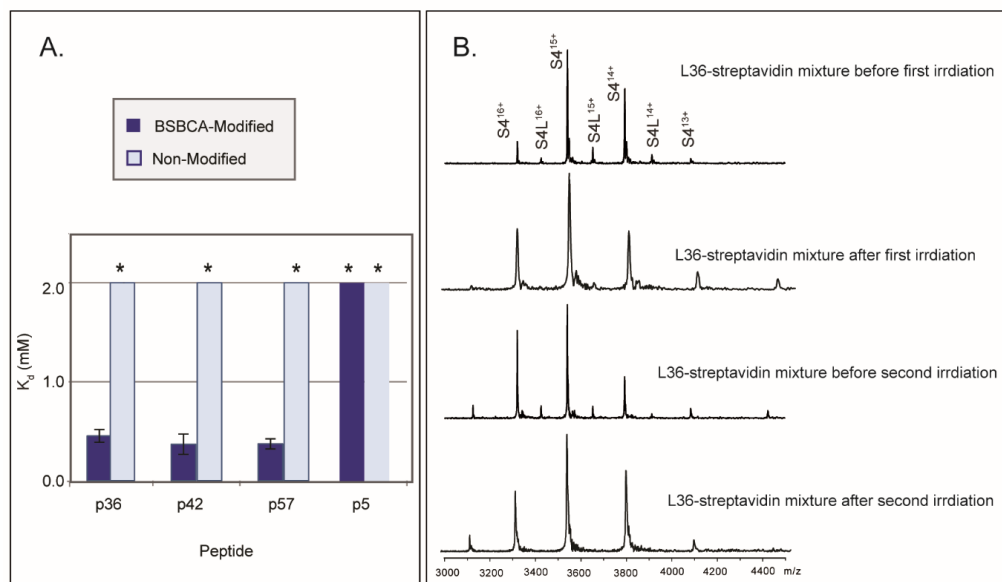
Figure 2.11. Binding data and MS for cyclized and non-cyclized peptides. (A) Dissociation constant (K_d) for the interaction between modified and non-modified peptides and streptavidin before and after irradiation with 370 nm light, measured by the direct ESI-MS binding assay. The * shows that no complex was observed. (B) Detection of **L36**-streptavidin complex by ESI-MS before and after one or two rounds of irradiation with 370 nm light.

from 452 μM to > 2 mM (Figure 2.10B). The decrease in K_d induced by irradiation was reversible and the complex between the ligand and the protein re-appeared after 30 min incubation of the solution in the dark. The binding could be turned "ON" or "OFF" for multiple cycles (Figure 2.10B, 2.11B, Table 2.2). Reversibility of the binding confirmed that the decrease in binding upon irradiation was not caused by photo-damage or any other non-specific processes. Ligand **L81**, synthesized from a peptide that was selected from unmodified library, exhibited no
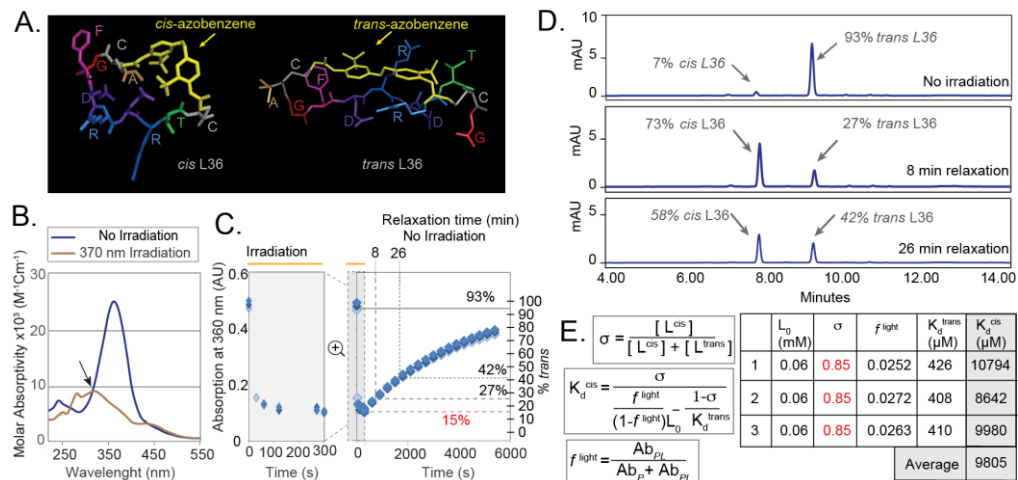
Figure 2.12. Studies on light-responsiveness of **L36**. A) Representative structures of **L36** demonstrate substantial effect of *trans* and *cis* states on peptide conformation. (B) The UV-visible spectrum of **L36** before and after irradiation. The isosbestic point (315 nm) is indicated by an arrow. (C) Kinetics of isomerization of *trans* **L36** to *cis* isomer and its relaxation back to *trans* form. The percentage of the *trans* isomers was measured by HPLC and was extrapolated to be 15% at the time of irradiation. (D) HPLC traces of the **L36** in various *cis/trans* ratios. (Top) dark-adapted ligand; (middle) **L36** injected after irradiation; (bottom) **L36** after irradiation and relaxation for 18 minutes. The time indicated on the traces is a combination of relaxation time and retention time. (E) The K_d of the **L36** in *cis* state can be calculated from σ and f^{light} and K_d of **L36** in *trans* state.

change in K_d upon irradiation (Table 2.1). This observation suggested that BSBCA cyclization of the binders selected from non-BSBCA modified libraries might not be a viable strategy for selection of LR-binders.
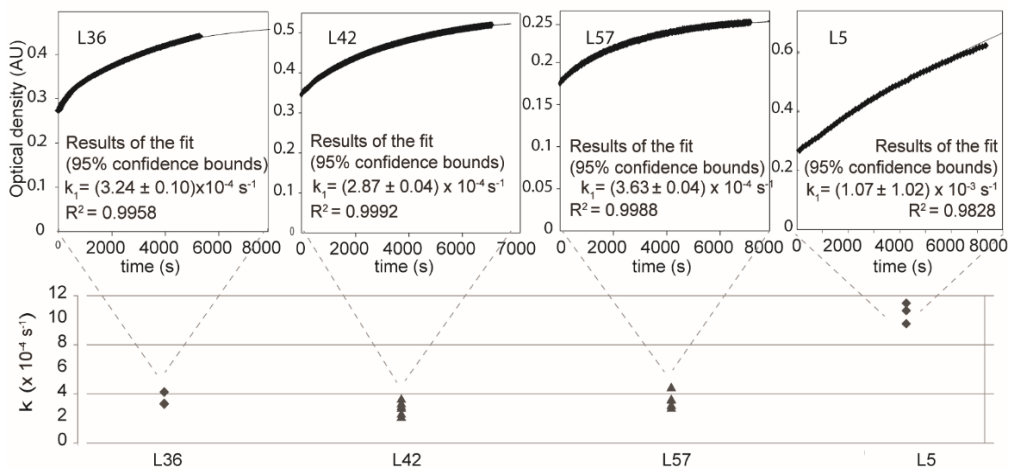
Figure 2.13. Kinetics of thermal relaxation for **L36**, **L42**, **L57** and **L5** after 15 min irradiation with 370 nm light. Each fit represents one typical experiment. (see Table 2.2)
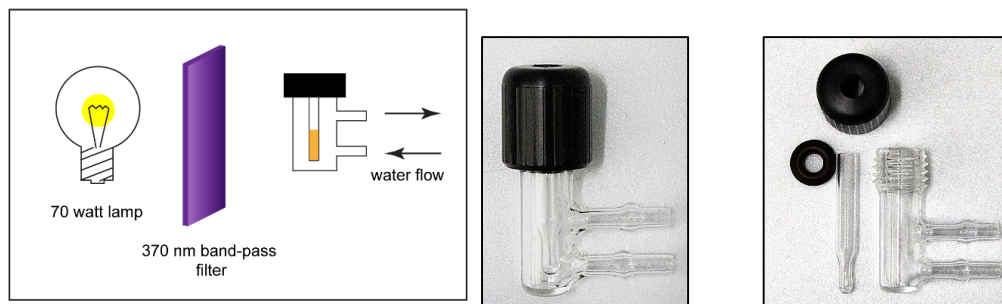
Figure 2.14. Setup for irradiation of the LR-ligands in constant temperature.

To understand the structural basis for change in K_d, we modeled the structure of *cis*-**L36** and *trans*-**L36** ligands using Amber 11 (Figure 2.12A, 2.15). Molecular dynamics (MD) of *trans*-isomer revealed lack of defined secondary structure and dihedral angles extracted from MD were consistent with experimental J^3-coupling constants (6–8 Hz, Table 2.3). Several weak long-range NOEs observed

Table 2.2. Rate constants for thermal relaxation of **L36**, **L43**, **L57** and **L5** after 15 min irradiation with 370 nm light. The values are average of at least 3 independent experiments. (See Figure 2.13)

Ligand	L36	L42	L57	L5
K_I(min^{-1})	$(3.10 \pm 0.12) \times 10^{-4}$	$(2.88 \pm 0.41) \times 10^{-4}$	$(3.60 \pm 0.46) \times 10^{-4}$	$(2.45 \pm 0.65) \times 10^{-4}$
Half-life (min)	37.3 ± 1.4	40.1 ± 6.0	32.5 ± 3.9	10.9 ± 0.58
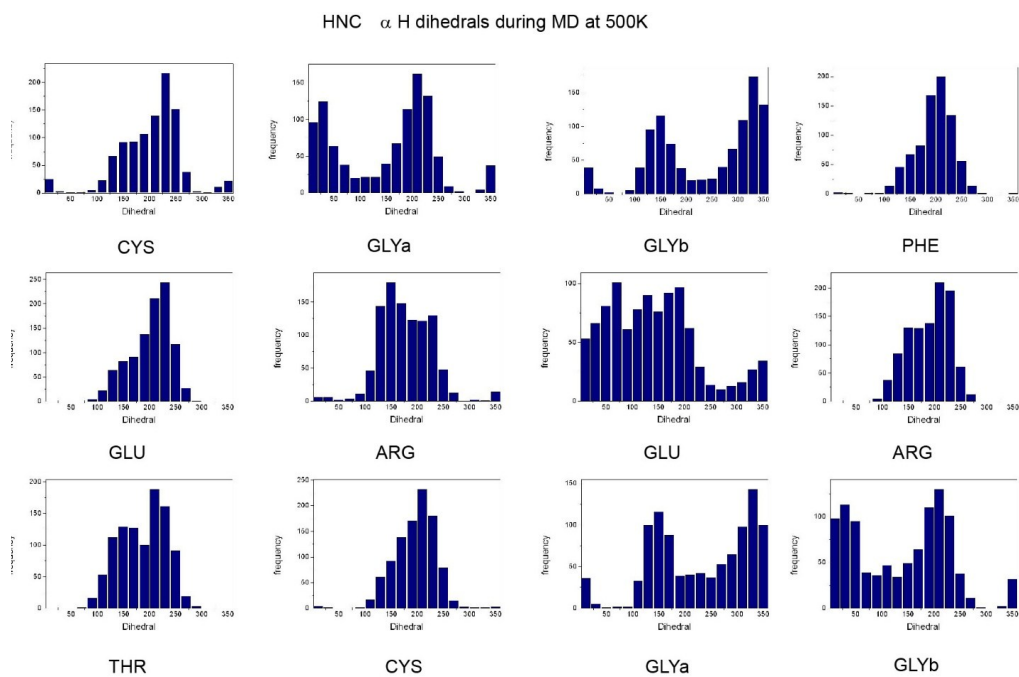
Figure 2.15. Distribution of dihedral angles of amino acids in **L36** predicted by Amber 11. (see text for details)

Table 2.3. Coupling constants of Hα in amino acid constituents of **L36** calculated from Amber 11 and obtained from ^{1}H-NMR.

	Calculated	Experimental
CYS2	6.5	6.98
GLY3a	6.5	5.56
GLY3b	6.0	
PHE4	7.6	7.25
GLU5	6.8	
ARG6	7.2	
GLU7	5.8	
ARG8	7.3	
THR9	6.9	
CYS9	7.4	7.17
GLY10a	6.0	5.94
GLY10b	6.3	

by NMR correlated with highly populated MD trajectories, in which the arginine moieties approach aromatic rings of azobenzene. Change in shape from *cis* isomer to *trans* (Figure 2.16) can explain the loss of binding to streptavidin. Modeling also suggested that **L36** is too bulky for biotin-binding pocket. Indeed, we observed experimentally that biotin did not compete with **L36** for binding (Figure 2.10C). The differences in shape prompted us to separate *cis* and *trans* isomers by HPLC and quantify the *cis/trans* ratio (σ) upon irradiation. Both isomers have the same molar extinction coefficient (ε) at 315 nm (Figure 2.12B) and we used this wavelength to quantify the isomers before and after irradiation (Figure 2.12D). To account for partial thermal relaxation of the *cis*-**L36** during HPLC run, we varied the relaxation time and used first order kinetics to extrapolate the amount of *cis* **L36** during the irradiation (85%, red line, Figure 2.12C). We then used equilibrium-
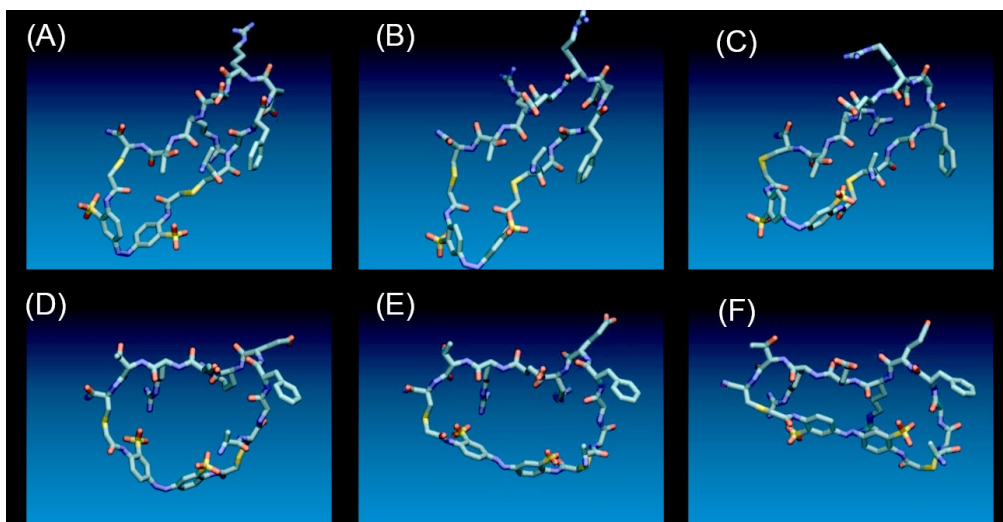
Figure 2.16. Snapshots of relaxation of *cis*-**L36** to *trans*-**L36**, modeled by Amber molecular dynamics and recorded in PyMol.

Table 2.4. Apparent dissociation constant (K_d) for the complexes of streptavidin and ligands **L36**, **L42** and **L57** before and after one or two rounds of irradiation with 370 nm light, measured by ESI-MS at pH 6.8 and 25℃.

Ligand	L36	L42	L57
K_d **before first irradiation (μM)**	452.5 ± 63.7	368.4 ± 103.2	449.5 ± 82.4
K_d **after first irradiation (μM)**	2000*	814.7 ± 55.1	1223.3 ± 201.6
K_d **before second irradiation (μM)**	409.3 ± 114.7	364 ± 33.8	506.1 ± 89.77
K_d **after second irradiation (μM)**	2000*	690.3 ± 101.5	987.3 ± 55.1

The * shows that no complex was observed by ESI-MS.

binding model to extrapolate the K_d of the *cis* isomer (Figure 2.12E). The calculated $^{cis}K_d$ of the **L36** was 9805 µM, or 22 fold higher than the $^{trans}K_d$ of **L36** (Figure 2.12E). Incomplete isomerization ($\sigma = 85\%$), however, masked this difference and yielded only a 4.5- fold apparent change in K_d (Table 2.1, Table 2.4).

The mechanism of incomplete isomerization is known to be caused by the overlap of n$\rightarrow$ π* absorption band for two isomers.[137] Azobenzenes with isomerization efficiencies above 85% have been reported.[151,152] Despite limited isomerization, Azb-containing ligands have been used for photo-regulation in studies *in vivo* and *in vitro*. Ligand binding and DNA transcription has been successfully regulated by ligands tethered to proteins or DNA by Azb linker with ratio of cis-isomer in 50-85% range in photostationary state after irradiation.[153-155] An Azb-derivative of propofol,[156] with the EC_{50} of 6.1 µM and < 85% of *cis* isomer at photostationary state after irradiation, has been used to trigger movement in live zebrafish. Azb-derivatives of disodium cromoglycate (DSCG) with 10–270 µM potency and < 60% of *cis* isomer at photostationary state triggered photo-induced histamine release in mast cells.[157] Photo-regulation of *in vitro* translation was achieved using an Azb-surfactant (azoTAB) with low DNA-binding affinity (1–3 mM)[158] and 70% *cis* isomer at photostationary state.[159] Both the groups of Baigl[158] and Trauner[156] employed ligands that were binding ("ON") in the dark state and turned "OFF" by irradiation; the light-control was the most efficient when the concentration of ligand was similar to the K_d of the protein-ligand complex. Although **L36** exhibits incomplete isomerization (85%) and low "dark" affinity (452 µM), it is a promising tool for light-control studies. The cyclic peptide **L36** may not penetrate the membrane,

but it can be used to actuate or cluster cell surface receptors in cell lines that contain streptavidin-receptor fusion.[160]

2.3. Conclusion

Light-control of protein-ligand interaction has a long history, but design principles for light-responsive ligands have been established only for a few classes of ligand-receptor pairs. To date, most of the LR-ligands have been developed based on rational modification of a well-defined natural ligand.[110,137,157] In the absence of previously available lead compounds, the design of a ligand that exhibits different binding properties in response to light could be challenging. Here we used chemical modification on a 10^9 membered phage display library to produce a LR-library of macrocyles. The *in vitro* selection of the library yielded a "lead" structure that exhibited over 20-fold difference between $^{cis}K_d$ and $^{trans}K_d$. We believe that this approach will seed discovery of LR-ligands for diverse protein targets and provides lead compounds that will enable photo-regulation *in vitro* and *in vivo*.

2.4. Materials and Methods

Proton and carbon nuclear magnetic resonance spectra (^{1}H-NMR and ^{13}C-NMR) were acquired on an Agilent/Varian VNMRS two-channel 500 MHz or Agilent/Varian Inova two-channel 400 MHz spectrometer. The reference line for internal standard was defined as solvent chemical shift (D_2O or DMSO-d_6). ^{1}H-NMR data are reported as follows: chemical shift (ppm), multiplicity (s = singlet, d =

doublet, dd = doublet of doublets, ddd = doublet of doublet of doublets, t = triplet, m = multiplet), coupling constants (Hz) and integration. High resolution ESI mass spectra were recorded by Agilent Technologies 6220 oaTOF. MALDI mass spectra were recorded on AB Sciex Voyager Elite MALDI. High performance chromatography was performed on a Waters system equipped with a UV/Vis Waters detector. The absorption of the ligands was measured on a Molecular Devices Me2 UV/Vis spectrophotometer or a Nanodrop ND1000 UV/Vis spectrophotometer.

EZ-Link biotin-PEG$_2$-iodoacetamide (**BIA**) was purchased from ThermoScientific. TCEP·HCl was purchased from Sigma. Sodium 4-nitroaniline-2-sulfonate and 2-aminobenzensulfonic acid were purchased from TCI Chemicals. Ph.D.™ -C7C library lot# 0061010 was purchased from New England Biolabs. All peptides were purchased from Lifetein and were purified by HPLC. Streptavidin coated magnetic beads were purchased from Bioclone. All other chemicals were purchased from Sigma and were used without prior purification.

2.4.1. Synthesis of LR-Library

To a 99 µL solution of Ph.D. C7C phage display library (NEB, 5 ×10^{11} PFU/mL in 50 mM Tris, pH 8.5) we added a suspension of 25 µL of iTCEP and stirred the mixture for up to 48 h at 4 °C. Then we added a solution of BSBCA (1 µL of 100 mM) and allowed the phage library to react with BSBCA for 2 h. To the reaction mixture, we added a solution of PEG/NaCl (40 µL). We incubated the mixture on ice for 2 h, pelleted the phage by spinning at 14,000 rpm and dissolved the pellet in 50 µL of PBS (pH 7.4) for quantification of streptavidin binding or for selection.

2.4.2. Selection and Amplification of Azb-Modified Phage Library

The BSBCA-modified library (see above) was incubated with PEG/NaCl on ice for 2 hours, centrifuged for 15 min at 14,000 rpm to pellet the phage. The pellet was dissolved in 2% BSA in PBS and incubated in this buffer for 1 h. Streptavidin magnetic beads were blocked with 2 % BSA in PBS for 1 h, then mixed with phage and shaken for 1 h in the dark. The beads were washed 10 times with 0.1% Tween-20/TBS (TBST) to remove the unbound phage. The beads were incubated with 200 µL of elution buffer (0.2 M Glycine, 1 mg/mL BSA, pH 2.2) for 9 min and then neutralized with 30 µL of 1 M Tris, pH 9.3. The eluted phage was mixed with 140 µL of overnight culture of *Escherichia coli* and mixed with 2.5 mL LB medium and amplified in monodisperse emulsions (for details see[161]). Briefly, the mixture was aspirated in a 3 mL BD syringe and passed through a microfluidic flow-focusing device to encapsulate the phage and bacteria in monodisperse droplets suspended in a solution of perfluoro surfactant in HFE-7500. The phage was allowed to grow in emulsion for 5 h at 37 °C. The emulsion was destabilized by the addition of 1% Krytox in HFE-7500. The amplified phage was recovered from the aqueous phase by PEG/NaCl precipitation and used in the next round of modification and selection. After three rounds of selection, 80 clones of phage were amplified and tested for binding using ELISA (Figure 2.8).

2.4.3. Measurement of Binding of Peptides to Streptavidin (S_4).

Aliquots of the stock solutions were mixed to achieve the desired concentrations of S_4 (6.8 µM), non-modified or BSBCA-modified peptide (60 µM) in ammonium acetate buffer (10 mM, pH 6.8). The mixture was incubated at r.t. for 30 min and then analyzed by ESI-MS. To measure binding in the presence of light, the solution was irradiated with a 70 Watt metal–halide lamp (Venture Lightning, 22466) equipped with a 370 nm band pass filter (Edmund optics) in a temperature-controlled chamber (Figure 2.14) and analyzed by ESI-MS within 1–2 minutes of irradiation, 30 minutes after irradiation (relaxation studies) or after repeated irradiation (reversibility studies, Figure 2.10B).

2.4.4. Kinetics of Isomerization of L36

Kinetics of the isomerization was measured as a change in absorption at 360 nm during or after irradiation of **L36** (60 µM in PBS) using Molecular Devices M2^e or Thermo NanoDrop spectrometers. To quantify *cis*/*trans* ratio, a 60 µM solution of the **L36** in NH$_4$Ac buffer (50 mM, pH=7.4) was injected in HPLC before irradiation, <1 min after irradiation and 18 min after irradiation. The isomers were separated on a Waters Symmetry C18 column (3.5 µm, 4.6×75 mm) using a binary phase of NH$_4$Ac 50 mM buffer and acetonitrile, with a linear gradient of 10-30% acetonitrile over 10 min, and monitored at 315 nM using a 2489 Waters detector.

2.4.5. Synthesis of BSBCA and Intermediates
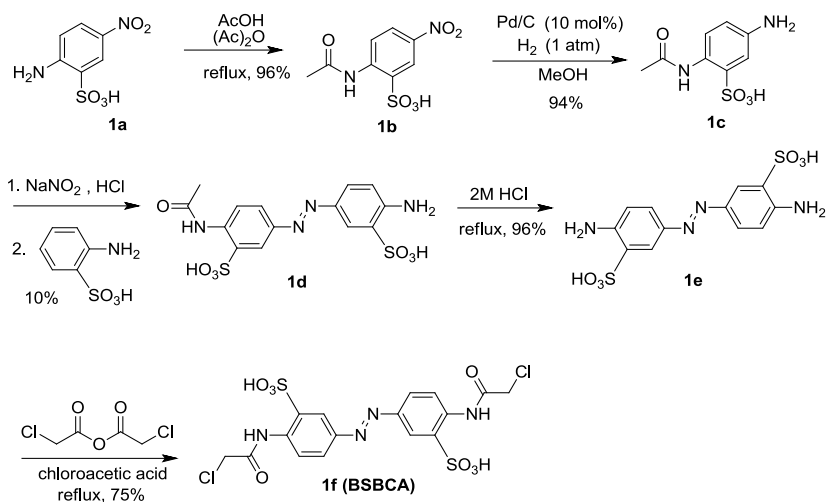
Scheme 2.1. Synthesis of BSBCA. Previous synthesis of BSBCA (**1f**) reported by Burns *et. al*[98] utilized oxidation of a symmetric intermediate (**1c** in the scheme). We designed an alternative strategy for the synthesis that utilizes coupling of two independent intermediates. The route could be potentially used to design BSBCA variants with unequal substituents on two rings. The synthesis starts from protection of the amine functionality of **1a** to obtain **1b**, followed by hydrogenation of the nitro group to yield **1c** in high yield. Diazotization of **1c**, followed by coupling with 2-aminosulfonic acid yielded **1d**. Hydrolysis of acetyl protecting group yielded the **1e**. Reaction of **1e** with chloroacetic anhydride in chloroacetic acid resulted in the formation of BSBCA (**1f**).

Synthesis of **1b**

72
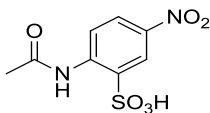
To a solution of sodium 4-nitroaniline-2-sulfonate (0.96 g, 4 mmol) in glacial acetic acid (20 mL), acetic anhydride (837.1 mg, 8.2 mmol) was added and the reaction was stirred at 105 °C for 3 h. The reaction was then cooled down to 65 °C and the solvent was evaporated *in vacuo*. The resulting white solid (999.5 mg, 96%) was dried for at least 3 h and used in next step without further purification. **^{1}H-NMR** (499.8 MHz, DMSO): δ 10.67 (s, 1 H), 8.56 (d, J = 9.0 Hz, 1 H), 8.45 (dd, J = 3.0 Hz, J = 0.5 Hz, 1 H), 8.22 (ddd, J = 0.5 Hz, J = 3.0 Hz, J = 9.0 Hz, 1 H) 1.95 (d, J = 3.0 Hz, 3 H) ppm; **^{13}C-NMR** (125.7 MHz, DMSO-d_6) δ$^+$ 169.1, 141.6, 141.3, 135.8, 126, 123, 119.9, 25.5 ppm.

HRMS-ESI (*m/z*): calc. for $C_8H_8N_2O_6S$ (M-H)$^-$: 259.0024; found: 259.0031; 0.3 ppm

Synthesis of **1c**
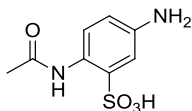
To a solution of **1b** (0.91 g, 3.5 mmol) in methanol (50 mL), Pd/C 10% (0.037g, 0.35 mmol) was added and the mixture was hydrogenated under 1 atm of H_2 at r.t. for 3 h. The reaction mixture was then filtered through Celite and the solvent was evaporated under reduced pressure to yield **1c** as a pale yellow powder (0.757g, 93%). The product was used in the next step without further purification.

^{1}H-NMR (499.8 MHz, DMSO-d_6): δ 9.98 (s, 1 H), 7.87 (d, J = 8.5 Hz, 1 H), 6.95 (d, J = 2.5 Hz, 1 H), 6.45 (dd, J = 2.5 Hz, J = 8.5 Hz, 1 H), 4.89 (s, 2 H), 1.95 (s, 3

H) ppm; ¹³**C-NMR** (125.7 MHz, DMSO) δ 166.7, 144.01, 136.7, 125.2, 121.3, 114.8, 112.8, 25.2 ppm.

HRMS-ESI (*m/z*): calc. for $C_8H_{10}N_2O_4S$ (M-H)⁻: 229.0283; found: 229.0287; -0.6ppm

Synthesis of **1d**
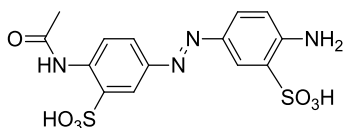
To an ice-cold solution of **1c** (0.23 g, 1.0 mmol) in distilled water (2 mL) we added 2 mL of 2 M HCl. An ice cold solution of NaNO$_2$ (69.2 mg, 1.0 mmol) in water (2 mL) was added dropwise over 5 min to the reaction mixture and the reaction was stirred vigorously at 0 °C for 1 h. We added sodium 2-Aminbenzensulfonate (196.15 mg, 1.0 mmol) to 3 mL of sodium acetate 3 M (pH 7) and stirred the solution on ice bath for 15 min. To this solution, we added ice cold diazo salt dropwise over 10 min and adjusted the final pH to 5. The mixture was stirred at 0 °C for 2 h followed by overnight stirring in r.t. The crude product was purified by preparative HPLC on a C-18 silica gel column (Varian Dynamax 250 × 21.4mm). The bright orange solution containing pure product was concentrated *in vacuo* to yield a dark purple solid (41 mg, 10%)

¹**H-NMR** (499.8 MHz, DMSO-*d*$_6$): δ 10.53 (s, 1 H), 8.44 (d, *J* = 9.0 Hz, 1 H), 8.07 (d, *J* = 2.5 Hz, 1 H), 8.02 (d, *J* = 2.5 Hz, 1 H), 7.77 (dd, *J* = 9.0 Hz, *J* = 2.0 Hz, 1 H), 7.75 (d, *J* = 8.5 Hz, 1 H), 6.74 (dd, *J* = 9.0 Hz, *J* = 2.0 Hz, 1 H), 2.07 (s, 3H)

ppm; ^{13}C-NMR (125.7 MHz, DMSO-d_6) δ 176.8, 148.5, 148.2, 142.7, 135.0, 125.9, 127.6, 125.7, 125.4, 125.2, 121.2, 118.1, 117.8, 20.5 ppm.

HRMS-ESI (*m/z*): calc. for $C_{14}H_{14}N_4O_7S_2$ (M-2H)$^{2-}$: 206.0079; found: 206.0073; -3.2ppm

Synthesis of **1e**
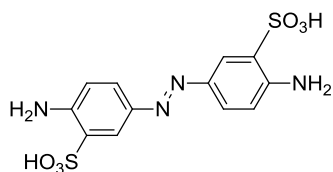
To a solution of **1e** (334 mg, 1.0 mmol) in water (1 mL), we added HCl (0.55 mL, 37.5% aqueous solution) and the mixture was stirred at 105 °C for 2 h. The resulting purple precipitate was filtered and dried in high vacuum. The solid was then dissolved in 1 mL of water and the pH of the solution was adjusted to 8.5 with 0.1 M NaOH. The solvent was evaporated to yield **1e** as a dark orange solid (0.357 g, 96%). ^{1}H-NMR (499.7 MHz, DMSO-d_6): δ 7.93 (d, *J* = 2.5 Hz, 1 H), 7.53 (dd, *J* = 8.5 Hz, *J* = 2.5 Hz, 1 H), 6.69 (d, *J* = 8.5 Hz, 1 H), 6.18 (s, 2 H) ppm; ^{13}C-NMR (125 MHz, DMSO) δ 149.9, 144.6, 132.6, 127.4, 124.2, 117.9

HRMS-ESI (*m/z*): calc. for $C_{12}H_{10}N_4O_6S_2$ (M-2H)$^{2-}$: 185.0026; found: 185.0021; -2.63 ppm

Synthesis of **1f (BSBCA)**
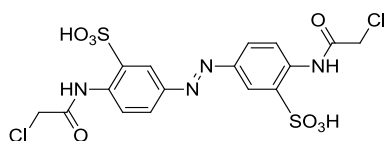
To a mixture of chloroacetic acid (496 mg, 5.2 mmol) and chloroacetic anhydride (899 mg, 5.2 mmol) in a 10 mL flask, we added **1e** (0.039 g, 0.21 mmol). The mixture was sealed with a white rubber stopper and purged with nitrogen. Upon heating to 78° C, the mixture melted down and the resulting solution was stirred for 8 h in the dark. The reaction was then cooled to r.t. and the solid was washed three times with 10 mL of dichloromethane and once by acetone/dichloromethane (1/1). BSBCA was collected as an orange solid (83.5 mg, 75%). The product was used without further purification. **^{1}H-NMR** (499.7 MHz, D_2O): δ 8.25 (d, 1 H, $J = 2.5$ Hz), 8.21 (d, 1 H, $J = 9.0$ Hz), 7.91 (dd, 1 H, $J = 9.0$ Hz, $J = 2.5$ Hz), 4.39 (s, 2 H) ppm; **^{13}C-NMR** (125.3 MHz, D_2O) δ 168.6, 148.9, 136.4, 134.5, 127.0, 124.5, 122.7, 41.49 ppm

HRMS-ESI (*m/z*): calc. for $C_{16}H_{12}N_4O_8S_2Cl_2$ $(M-2H)^{2-}$: 260.9742; found: 260.9545; 1.19 ppm

2.4.6. Synthesis of BSH (Procedure).

The biotin cysteine conjugate (**BSH**) was synthesized on resin by standard solid phase amide coupling. In the first step, Fmoc-Cys(Trt)-OH (0.56 g, 0.96 mmol) was added to 0.3 g of deprotected rink amide resin (0.24 mmol, suspension in DMF), followed by HBTU (0.36 g, 0.96 mmol) and DIPEA (0.25 g, 1.92 mmol). The reaction was allowed to proceed for 2 h. The resin was washed with DMF (3 × 3 mL) and Fmoc was removed by treating with 20% (v/v) piperidine in DMF. The resin was washed with DMF (3 × 3 mL). In the second step, a DMF solution of biotin (0.23 g, 0.96 mmol), HATU (0.36 g, 0.96 mmol) and DIPEA (0.25 g, 1.92 mmol)

was added to the resin. The reaction was allowed to proceed for 2 h, and the resin was washed with DMF (5 × 3 mL), CH_2Cl_2 (3 × 3 mL) and dried in vacuo. The **BSH**-containing resin (100 mg sample) was then treated with 2 mL of 92.5:2.5:2.5:2.5 (v/v/v/v) TFA:TIS:EDT:H_2O for 2 h. The resin was washed with TFA (500 µL). The volume of the combined TFA solutions was reduced under a stream of N_2 to ~500 µL. Cold ether (8 mL) was added to the TFA solution upon vigorous stirring, and the mixture was kept on ice for 2 h. The precipitate was collected by centrifugation, and washed with cold ether (3 × 10 mL). The precipitate was dried in vacuo and resulting solid was used without further purification.

The yield (45.0%) was calculated using the absolute loading of the resin (0.8 mmol/g), and the molecular weight of **BSH** (346.4 g/mol). **^{1}H-NMR** (400 MHz, DMSO): δ 7.89 (d, J = 10.5 Hz, 1 H), 6.38 (s, 1 H), 6.32 (s, 1 H), 4.30 – 4.26 (m, 2 H), 4.12 – 4.09 (m, 1 H), 3.10 – 3.05 (m, 1 H), 3.27 (s, 2 H), 2.80 (dd, J = 6.5 Hz, J = 15.5 Hz, 1 H), 2.74 (dd, J = 6.5 Hz, J = 11.0 Hz, 1 H), 2.63 (ddd, J = 10.0 Hz, J = 10.0 Hz, J = 3.0 Hz, 1 H), 2.54 (d, J = 15.5 Hz, 1 H), 2.18 (t, J = 10.0 Hz, 1 H), 2.14 (t, J = 9.0 Hz, 2 H), 1.60 – 1. 29 (m, 8 H) ppm; **^{13}C-NMR** (100 MHz, DMSO-d_6) δ 172.8, 172.5, 163.2, 61.5, 59.7, 55.7, 52.1, 41.5, 35.5, 28.6, 28.4, 25.6 ppm **HRMS-ESI** (*m/z*): calc. for $C_{13}H_{22}NaO_3S_2$ (M+Na)$^+$: 369.1026; found: 369.1019; -1.81 ppm
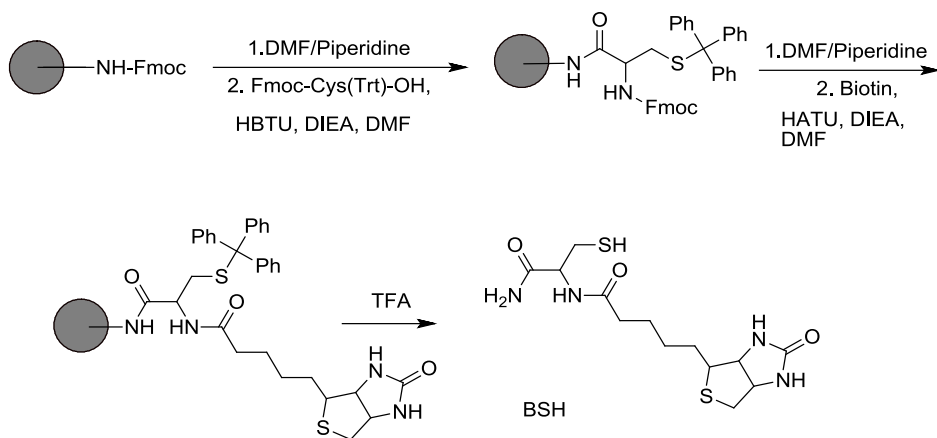
Scheme 2.2. Synthesis of **BSH**

2.4.7. General Procedure for Measurement of Phage Binding by ELISA.

A 150 μL solution of streptavidin (10 μg/mL in 0.1 M NaHCO$_3$, pH 8.5) was added to all wells of a 96-well plate (Costar, 3370) and the plate was incubated at 4 °C overnight. The wells were blocked with 2% BSA (in PBS pH 7.4) for 1 h, followed by washing with 3 × 200 μL of 0.1% Tween 20 in PBS (PBST). Individual clones of amplified phage (5 × 10^{10} PFU/mL, 180 μL) were added to the wells and the wells were incubated on a rocker for 1 h in the dark. The unbound phage was then washed away with 3 × 200 μL of PBST and wells were incubated with 1:1000 dilution of anti M13 antibody-HRP conjugate monoclonal antibody (GE Healthcare) in 2% BSA solution in PBS. After 1 h incubation in dark, the excess antibody was washed away with 4 × 200 μL PBST and 50 μL of HRP substrate (1-Step™ Turbo TMB, Thermo Scientific) was added to each well. The reaction was quenched after 2 h with 50 μL of 1 M H$_2$SO$_4$ and the absorbance was measured at 450 nm using a plate reader.

2.4.8. Typical Procedure for Synthesis of LR-Peptides.

To a 76 µL of a solution of 4 mM peptide 36 (sequence ACGFERERTCG, Lifetein) in Tris 50 mM (pH 8.5), we added 8 µL of 100 mM TCEP·HCl in Tris 50 mM (pH 8.5) and 16 µL of 100 mM BSBCA in DMSO. The mixture was incubated in dark and r.t. up to 4 h. The product was purified over semi-preparative HPLC and the structure of peptide was confirmed by MALDI-MS.
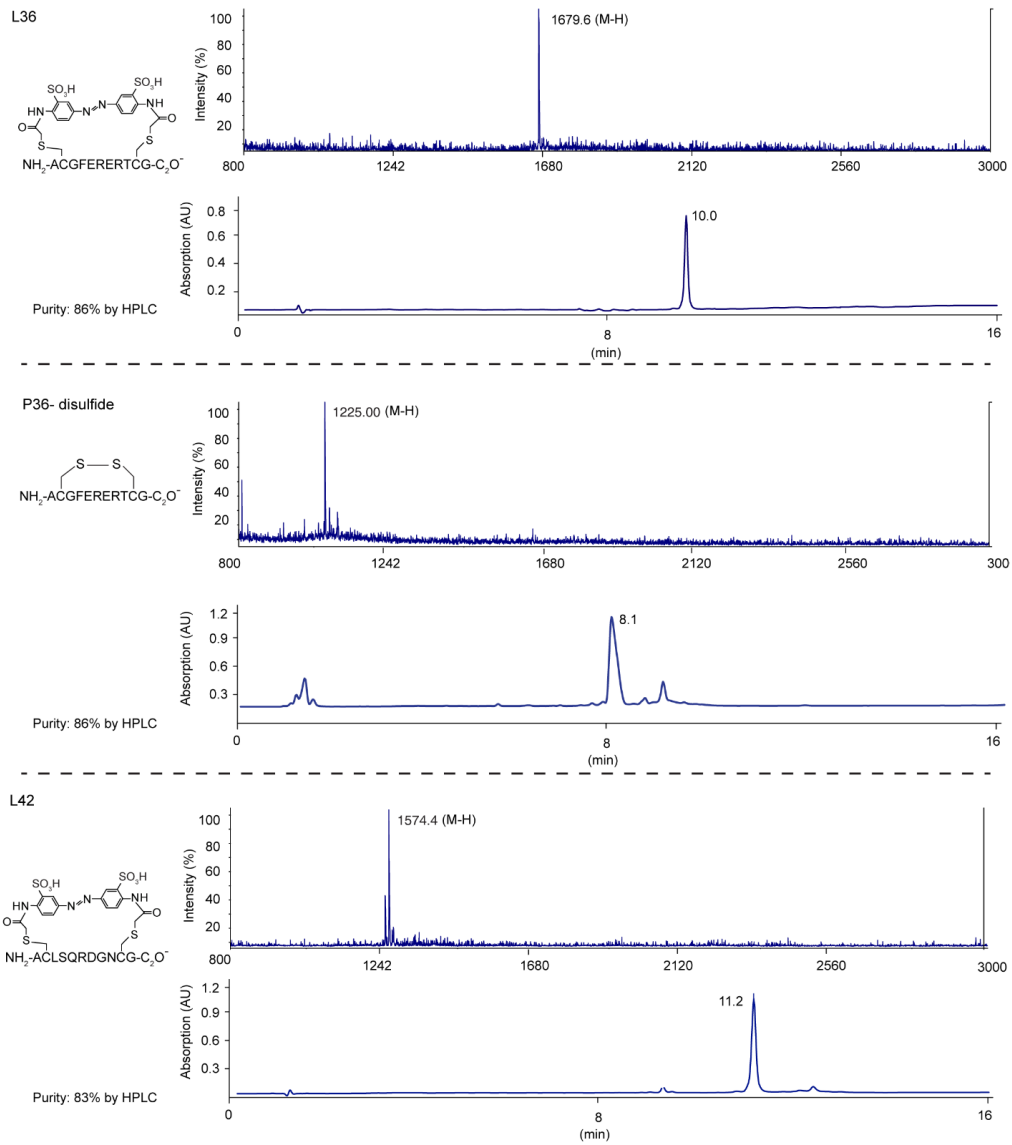
Figure 2.17. Characterization and purity of the peptides used in ESI-MS Binding assay.
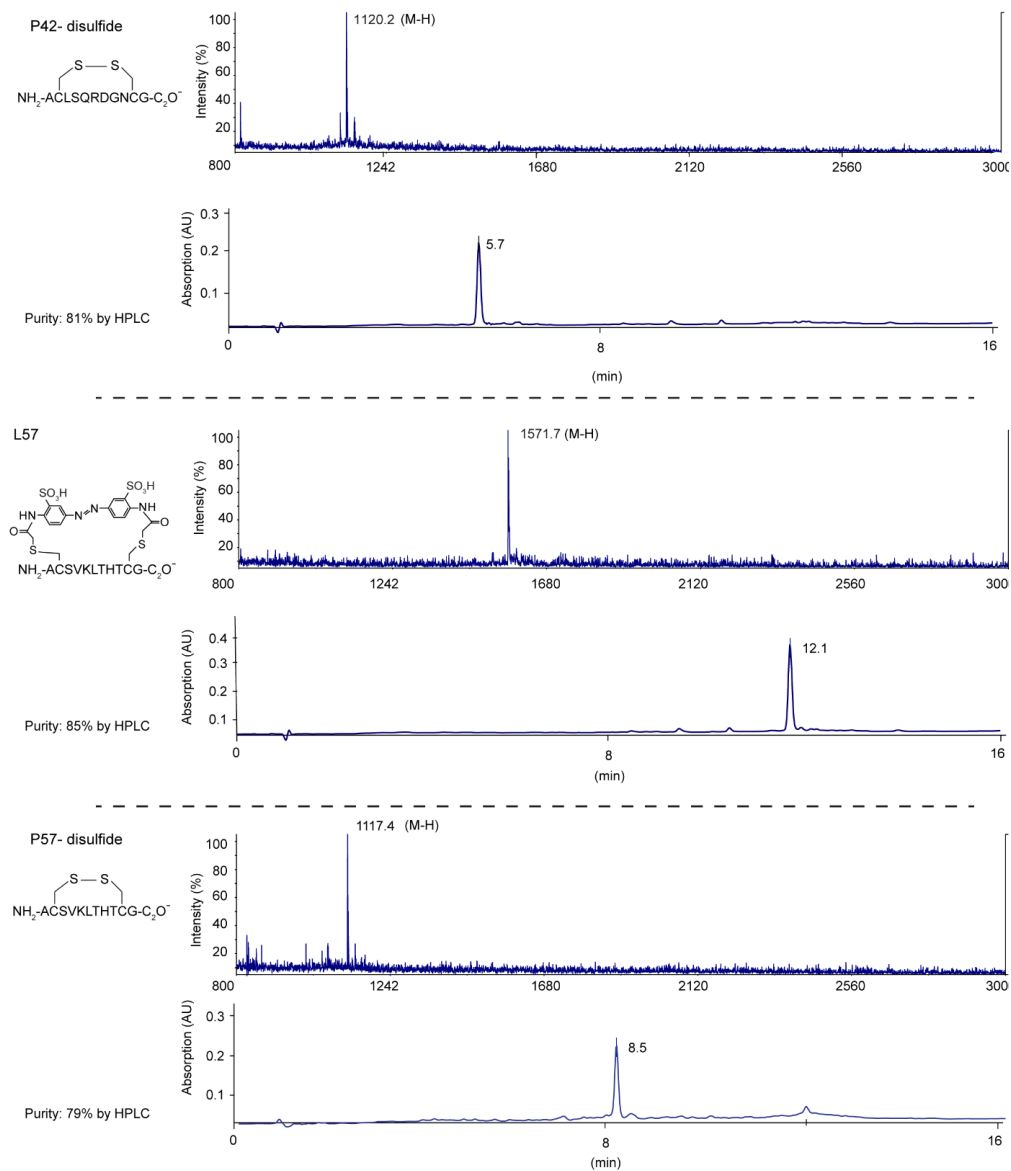
Figure 2.17. (continued) Characterization and purity of the peptides used in ESI-MS Binding assay
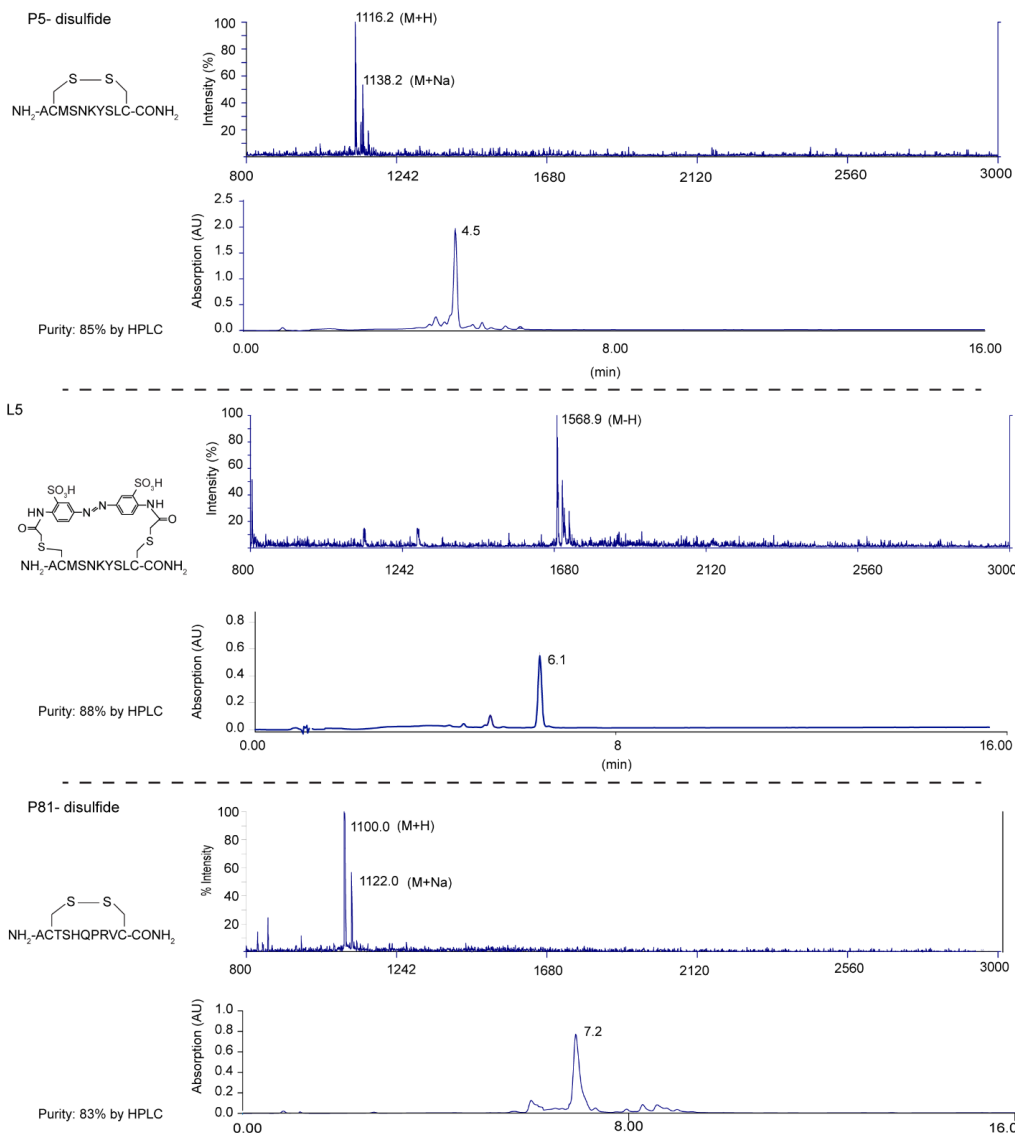
Figure 2.17. (continued) Characterization and purity of the peptides used in ESI-MS Binding assay
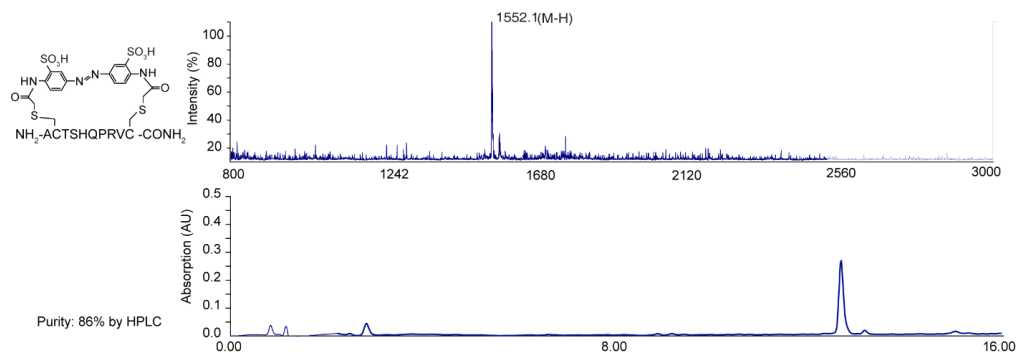
Figure 2.17. (continued) Characterization and purity of the peptides used in ESI-MS Binding assay.

2.4.9. Kinetics of Thermal Relaxation for L36, L42 and L57

The absorption at 360 nm was recorded on a molecular devices Me^2 spectrometer using 50 µL of 60 µM concentration of the ligand after irradiation, in a 384-well plate.

2.4.10. Measurement of Binding of Non-Modified Peptide to Streptavidin (S4).

Aliquots of the stock solutions were mixed to achieve the desired concentrations of S_4 (6.8 µM), non-modified peptide (60 µM) and ammonium acetate (10 mM, pH 6.8). The mixture was incubated at r.t. for 30 min and then analyzed by ESI-MS.[148,162]

2.4.11. Measurement of Binding of BSBCA-Modified Peptide to Streptavidin (S_4).

A. **Binding in the absence of light.** Aliquots of the stock solutions were mixed in dark to achieve the desired concentrations of S_4 (6.8 µM), azobenzene modified peptide (60 µM) and ammonium acetate (10 mM). The mixture was incubated in dark at r.t. for 30 min and then analyzed by ESI-MS.

B. **Binding in the presence of light.** A solution of azobenzene modified peptide (60 µM) in ammonium acetate buffer (10 mM, pH 6.8) was added to a 300 µL glass insert and irradiated with a 70 watt metal–halide lamp (Venture Lightning, 22466) equipped with a 370 nm band pass filter (Edmund optics) for 10 min. A flow of r.t. water was used to prevent any potential change in temperature of solution upon irradiation (Figure 2.14). An aliquot of streptavidin was then added to the solution to yield a final concentration of 10 µM S_4. The mixture was irradiated for an additional 30 min and analyzed by ESI-MS immediately.

C. **Binding after thermal relaxation**. The irradiated mixture from step B was then kept in the dark for 30 min and then analyzed by ESI-MS.

D. **Binding after second irradiation**. The irradiated mixture from step C was then irradiated for 30 min and then analyzed by ESI-MS

2.4.12. Calculations of the Dissociation Constant of the *cis*-Form of the Strep-

tavidin-Binding Ligands.

The apparent dissociation constants in the presence of irradiation (Table 2.4) do not represent true dissociation constant for *cis* form because *cis*-to-*trans* isomerization is incomplete. The K_d of the *cis*-form could be derived from the values in Table 2.4, *cis*/*trans* ratio of ligands upon irradiation, and the following equations, which assume that system reached equilibrium in the presence of light. This assumption is satisfactory because we allow the protein-ligand complex to form for 30 minutes in the presence of light.

1. When the solution is incubated in dark, ligand exists only in its *trans* form. For a 1:1 protein-ligand complex (eq 1), the dissociation equilibrium constant (K_d^t) for *trans* ligand (L^t) and protein (P) is calculated using eq 2:

$$P + L^t \rightleftharpoons PL^t \tag{1}$$

$$K_d^t = \frac{[P][L^t]}{[PL^t]} \tag{2}$$

Here [P], [L^t] and [PL^t] are the equilibrium concentrations of free P, free L^t and PL^t complex, respectively. The equilibrium constant can be derived from the fraction of the protein ligand complex (f^d), which can be observed directly by ESI-MS and calculated using eq 3:

$$f^d = \frac{[PL^t]}{P_0} = \frac{Ab_{PL}}{Ab_P + Ab_{PL}} \tag{3}$$

Here P_0 is the initial concentration of protein, Ab_P and Ab_{PL} are the normalized abundance of P and PL ions in mass spectra. Combination of mass balance for ligand and protein and equation (3), yields:

$$[PL^t] = f^d P_0 \tag{4a}$$

$$[P] = P_0 - [PL^t] = (1 - f^d)P_0 \tag{4b}$$

$$[L^t] = L_0 - [PL^t] = L_0 - f^d P_0 \tag{4c}$$

where L_0 is the initial concentration of ligand. Substituting equations 4a-4c into eq 2 yields:

$$K_d^t = \frac{(1-f^d)(L_0 - f^d P_0)}{f^d} \tag{4}$$

When the solution is irradiated, ligand exists as an equilibrium mixture of *cis* form (L^c) and *trans* form (L^t). Both forms can bind to protein. These processes are represented in equations 5a-c.

$$P + L^t \rightleftharpoons PL^t \tag{5a}$$

$$P + L^c \rightleftharpoons PL^c \tag{5b}$$

$$L^t \rightleftharpoons L^c \tag{5c}$$

The dissociation equilibrium constant (K_d^t *or* K_d^c) for *trans* ligand (L^t) or *cis* ligand (L^c) and protein (P) is calculated using equation 6a or equation 6b:

$$K_d^t = \frac{[P][L^t]}{[PL^t]} \tag{6a}$$

$$K_d^c = \frac{[P][L^c]}{[PL^c]} \tag{6b}$$

86

Here [P], [L^t], [PL^t], [L^c], and [PL^c] are the equilibrium concentrations of free P, free L^t, PL^t complex, free L^c and PL^c complex, respectively. When [L] >> [P], the variables [L^c] and [L^t] could be estimated as:

$$[L^c] = \sigma[L] = \sigma L_0; \quad [L^t] = (1 - \sigma)L_0 \tag{7}$$

Where L_0 is the initial concentration of ligand and σ is the ratio of *cis* isomer to total ligand at photostationary state: $\sigma = [L^c]/([L^c] + [L^t])$.

From mass balance and eqs 6a-6b, the initial concentration of protein (P_0) can be expressed as:

$$P_0 = [P] + [PL^t] + [PL^c] = [P] + \frac{[P][L^t]}{K_d^t} + \frac{[P][L^c]}{K_d^c} \tag{8}$$

Substitute eq7 into eq 8 and rearrange to get [P]:

$$[P] = P_0 / (1 + \frac{\sigma L_0}{K_d^c} + \frac{(1-\sigma)L_0}{K_d^t}) \tag{9}$$

In this case, the binding fraction after irradiation (f^l) is calculated using eq10:

$$f^l = \frac{Ab_{PL}}{Ab_P + Ab_{PL}} = 1 - \frac{[P]}{P_0} = 1 - 1/(1 + \frac{\sigma L_0}{K_d^c} + \frac{(1-\sigma)L_0}{K_d^t}) \tag{10}$$

To estimated K_d^c, equation 10 can be rearrange to yield the following:

$$K_d^c = \frac{\sigma}{\frac{f^l}{(1-f^l)L_0} - \frac{(1-\sigma)}{K_d^t}} \tag{11}$$

To use equation (11), we measured the fraction of the ligand bound in the dark (f^d) and fraction of ligand bound after irradiation with light (f^l) by ESI-MS (note: f^l was converted to K_d^t using equation 4). We also measured the maximum ratio of

the *cis* conformer to total ligand at photostationary state (σ) by HPLC (Figure 2.12). All measurements were performed on the same day using the same concentration of ligand and the same irradiation source.

2.4.13. General Method for Calculations of Molecular Dynamics of Ligand L36

The structure of photo-switchable peptide was parameterized using mixed parameters compatible with Amber 11 as following. Amber ff03 parameters were assigned to peptide residues and the gaff parameters were used for the prosthetic fragment. Force field parameters for the azobenzene fragment were adopted from the recent publication.[163] Partial charges (RESP) for the gaff parameterized fragment were derived using R.E.D. program.[164] Simulated annealing (2000 cycles) was performed first to obtain energy minimized starting coordinates according to the following protocol. The molecule was heated to 3000 K during 1 ps, equilibrated at this temperature for 1 ps and the temperature was lowered to 0 K over 18 ps. The results were clustered by RMSD and the representative structure for the most populated cluster was chosen. This structure contained only trans peptide bonds for all amino acids as well as trans configuration for the azobenzene fragment.

Molecular Dynamics (MD) was conducted in Amber 11 on a CentOS 6.0 computer equipped with an Intel Core 2 Duo E8400 CPU and an MSI N480GTX M2D15 GPU card. Amber 11 was compiled with GPU's CUDA instruction set using version 4.0.17 of the NVIDIA CUDA toolkit and GCC 4.4.4, and pmemd.cuda was

invoked using the SPDP (single precision calculation, double precision accumulation) model.[165] The molecule was placed in periodic box (truncated octahedron) neutralized with NaCl and saturated with TIP3P water with 12 Å offset. After minimization for 500 steps, the system was heated to 300 K equilibrated for 10 ps with controlled pressure and temperature. 40 ns production run was performed at 1 atm., 300 K using the Langevin heat bath with 1 ps^{-1} collision frequency. Non-bonded interactions were evaluated with 12 Å cutoff and the SHAKE constraint system for bonds to hydrogen was enabled. Long-range electrostatics was treated with particle-mesh Ewald (PME) periodic boundary conditions. Trajectories were saved every 100 ps and analyzed using VMD. The conformational space sampled by the molecule was quite large (backbone RMSD=2.8 Å with the average frame as the reference) indicating substantial flexibility.

Chapter 3 Rapid Peptide Cyclization by 3,3'-bis(Sulfonato)-4,4'-bis(Buta-2,3-dienoylamido)Azobenzene (BSBDA) for Generation of Light Responsive Macrocycles.

3.1. Introduction

Light-responsive (LR) ligands have found important applications in probing biological pathways such as cell signalling,[166] quorum sensing,[167] regulation of enzyme activity,[168-170] gene expression,[171] etc. LR-peptide derivatives can be made by attaching an LR-scaffold, such as azobenzenes (Azb), to the peptide backbone.[77,78,172] Irradiation of LR-macrocycles triggers the *trans-cis* isomerization in the Azb ring, which in turn perturbs the secondary structure of the peptide backbone and leads to an increase or decrease of the binding affinity to the target (Figure 3.1A).

Since its first report in 2003,[89] the 3,3'-bis(sulfonato)-4,4'-bis(chloroacetamido)azobenzene (BSBCA, Figure 3.1B) has been used in multiple publications as a precursor of LR-macrocyclic peptides.[93-95,97,173,174] The advantages of BSBCA are water solubility which makes the bio-conjugation possible in water and robust and efficient isomerization by visible light. We[173] and others[94] have demonstrated that screening of BSBCA-cyclized peptide libraries can yield LR-macrocyclic binders for the model target streptavidin.

Despite the wide utility of the BSBCA linker, the reaction of the free thiols in the peptide with chloroacetamide functionality is slow. Optimization of the
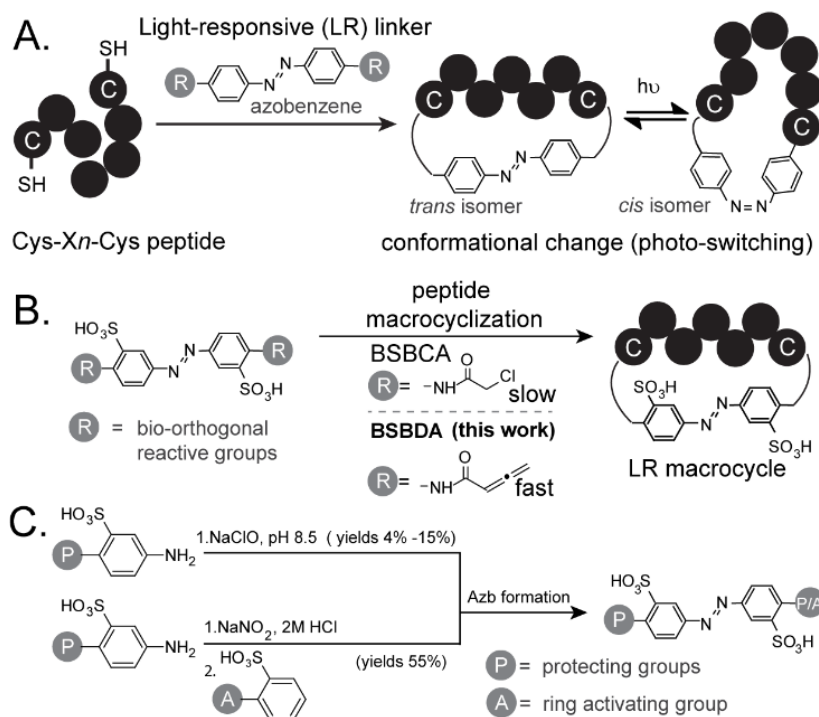
Figure 3.1. Strategies for generation of LR-peptides. A) LR-cyclic peptides can be made by conjugation of an Azb linker to the peptide backbone. Resulting LR-macrocycle can change conformation upon trans-cis isomerization of Azb. B) New water soluble LR-linkers can be made by derivatization of the Azb core **6**. C) We show increased yield of Azb core (**6**) through and efficient Azb synthesis pathway.

peptide conjugation chemistry mandates development of new Azb derivatives. In this chapter we describe the synthesize of the water soluble azobenzene core (**6**) substituted with allenamide functionality (Figure 3.1B), which has demonstrated to have higher rate and chemo-selectivity than the conventional haloacetamide reactive handles.[175] We also describe the first example of cyclization of a peptide

through reaction of the free thiols with allenamide functionalities on BSBDA linker. The reaction of glutathione with allenamide of BSBDA proceeded 500-1000 fold faster than the same reaction with chloroacetamide and peptide cyclization with BSBDA outcompeted the cyclization with BSBCA. We confirmed the structure of the resulting LR-macrocycle with nuclear magnetic resonance (NMR) spectroscopy and demonstrated that the peptides can be switched to the *cis* isomer after 2 min of irradiation with 365 nm light. The new linker BSBDA can replace the BSBCA in future syntheses of LR-peptides, because fast cyclization of the peptide is important for performing the chemical reaction on peptides fused to proteins[34,176] or displayed on bacteriophage.[94,173] In the latter case, increased reaction rate maximizes exposure to alkylating agents and can minimize the decrease in the viability (or infectivity) of the bacteriophage.[30,173]

3.2. Result and Discussion

The previously published syntheses of BSBCA[89,173,177] resulted in poor yields of final product (1.5-9%) because of inefficient formation of diazo bond of azobenzene (Scheme 3.1). Furthermore, purification of the intermediates mandated the use of high performance liquid chromatography (HPLC) equipped with a C18 column. This approach is not favourable because it is difficult to scale up. The low efficiency of the syntheses of the Azb core (**6**) in previous reports,[89,177] prompted us to develop a new synthesis pathway which improved the yield of this key intermediate (Figure 3.1C). Two important changes in the synthetic route resulted in
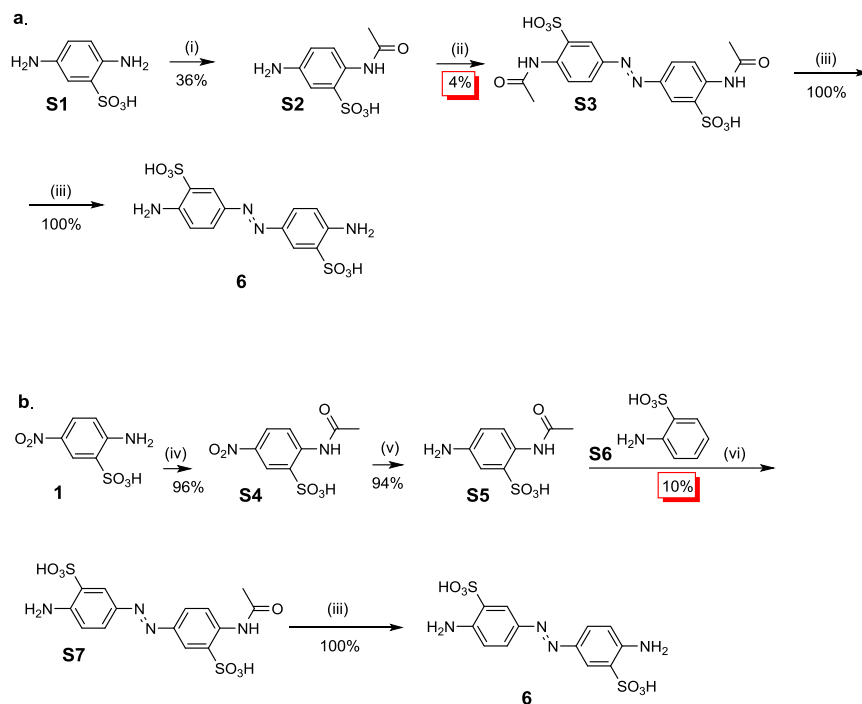
Scheme 3.1. Previously published syntheses of **6** by Zhang *et al.* and Jafari *et.al.*
a) The original synthetic route to **6** by Zhang *et al.* (i) Acetic acid, acetic anhydride, 2 h, reflux. (ii) NaClO, pH 8.5, 3 days, 4 °C, dark. (iii) 2 M HCl, 2 h reflux. b) Synthesis of **6** through diazo coupling by Jafari *et al.*[28] (iv) acetic anhydride, acetic acid, 2 h, reflux. (v) H_2, Pd/C, MeOH, 3 h. (vi) 1. 1 M HCl/MeOH = 1/1, NaNO$_2$, 0 °C, 1 h. 2. Sodium 2-aminobenzenesulfonate (aq), pH 8.5, overnight. Highlighted steps exhibit low yields and/or difficulties in purification of the product

significant improvements of the isolated yield: (i) we integrated a hydrophobic protecting group within the starting material to improve the purification of the negatively charged intermediates on silica. (ii) in the key synthetic step – diazo coupling – we used an activated nucleophile to maximize the yield of this reaction.

We used dodecanoyl chloride to protect the amine in the first step to yield **2** (Figure 3.2A). The resulting compound **2** was hydrogenated in 96% yield to give the compound **3** which served as the precursor for diazotization. The choice of the nucleophile is critical for diazotization. Presence of an electron donating group such as an amine in the para position to the nucleophilic carbon enhances the efficiency of the diazo coupling. Unfortunately, unprotected and non-substituted 2-ABS exhibited poor yield. We did not observe any reaction of **7** (Figure 3.2B) with the diazo salt of **3**, confirming the known observation that amides are not suitable in diazotization, thus ruling out any amide-forming protecting groups.

Protecting 2-aminobenzenesulfonic acid (ABS, **8**) with para-methoxybenzyl (PMB), enhanced the reaction efficacy. PMB-2-ABS (**4**) reacted with the diazo salt of **3** to produce the compound **5** in 55% isolated yield. Aside from increased nucleophilicity of **4**, another reason of higher isolated yield in comparison to previous reports was the simplified purification of the product **5**. The acetyl protected compound **S7** (Scheme 3.1B) exhibited broad tailing on the normal phase silica column during purification, while the key intermediate **5** can be efficiently purified on the same column with DCM/MeOH = 3/1 solvent as eluent. Deprotection by refluxing in 2 M HCl, in the presence of anisole as scavenger, yielded core **6** in 85% yield. We attempted to boost the nucleophilicity of the amine by introducing di-substitution on the amine with 1,2-dibromomethyl benzene (Figure 3.2B, **9**). However, we did not observe any reaction between **9** and the diazo salt of **3**.
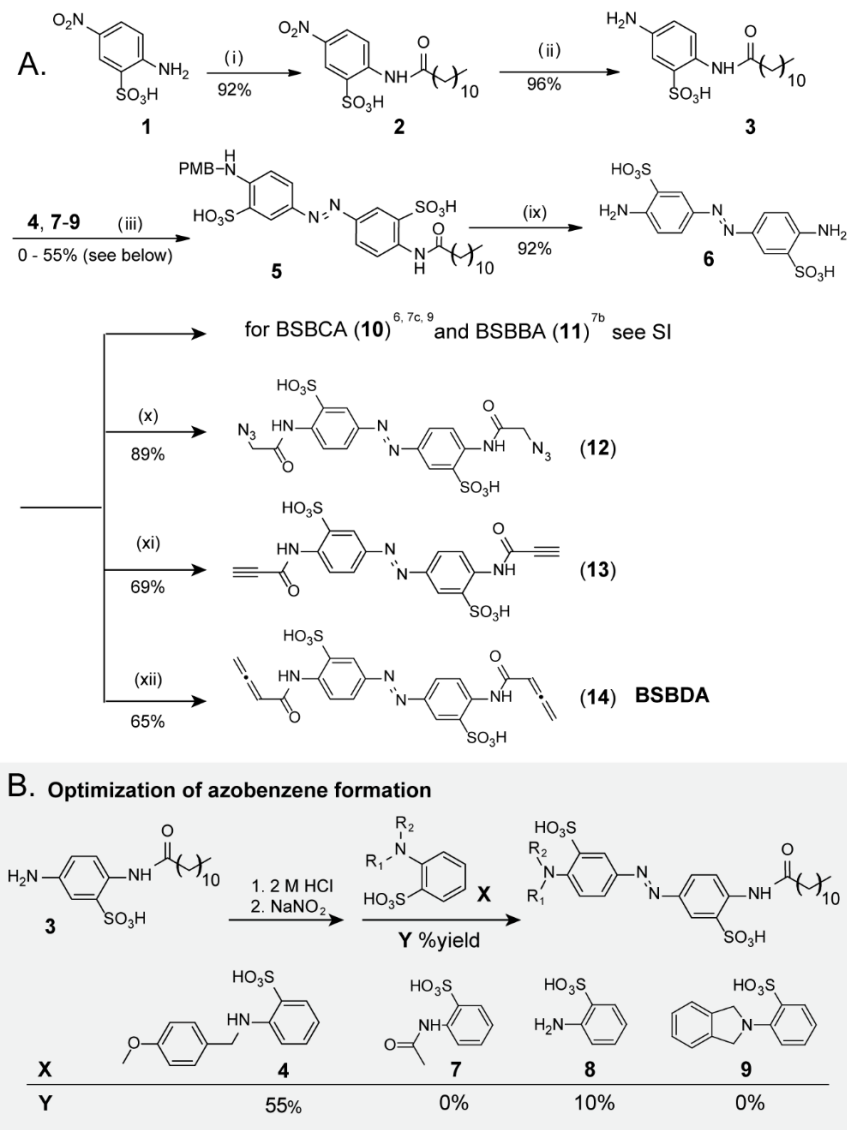
Figure 3.2. Synthesis of BSBDA and other Azb derivatives. A) (i) Dodecanoyl chloride, DMF, 0 °C to 110 °C, 4 h. (ii) Pd/C 10 mol%, H_2, 1 atm, 2 h. (iii) 2.0 M HCl/MeOH (1/1), $NaNO_2$, 0 °C, 1 h, then 1.2 eq of nucleophile **4**, **7-9** in Na_2CO_3 solution, pH 5, 18 h. (ix) 2 M HCl, anisole, 5 eq, reflux, 2 h. (x) azido-acetyl chloride, DMF, DIPEA, 0 ° C to r.t., 115 min. (xi) propiolic chloride, 1.0 M aq. Na_2CO_3, 90 min. (xii) 3-butynoyl chloride, aq. $NaHCO_3$, 60 min.

Figure 3.2. (continued)

B) Reaction of **3** with different nucleophiles to test the efficiency of the diazo coupling.

Using azobenzene **6**, we synthesized five azobenzene linkers (Figure 3.2A, **10-14**) that can be used to cyclize the peptides using different bio-conjugation strategies such as thiol-haloacetamide nucleophilic substitution,[94,173,178] copper catalysed azide-alkyne cycloaddition (CuAAC)[179] and thiol-alleneamide addition.[175] The cyclization of the peptides has been shown using thio-haloacetamide[173,176,177] and click chemistry.[179-181] Recently Abbas *et al.* showed modification of the cysteine containing peptides with allenamide containing scaffolds through Michael addition;[175] but to the best of our knowledge, peptide cyclization through Michael addition to allenamide has not been reported.

Since kinetics of the alleneamide-thiol reaction have not been characterized in the original report,[175] we measured the kinetics of the reaction of BSBDA with glutathione. Incubation of BSBDA with 160 μM GSH at pH = 8.0 resulted in the formation of di-substituted product in 20 min (Figure 3.3). Fitting the kinetics in a consecutive first order plot (Figure 3.4A, 3.4B), yielded the rate constant $k_1 = 30.2$ $M^{-1}s^{-1}$ and $k_2 = 16.1$ $M^{-1}s^{-1}$, respectively. In contrast, the rate constant of the reaction of GSH with chloroacetamide is slower with $k = 0.03$ $M^{-1}s^{-1}$ at pH 8.0.[182]

To compare the relative rates of reaction in controlled conditions, we investigated the cyclization of a streptavidin binder precursor **P36** (100 μM) with
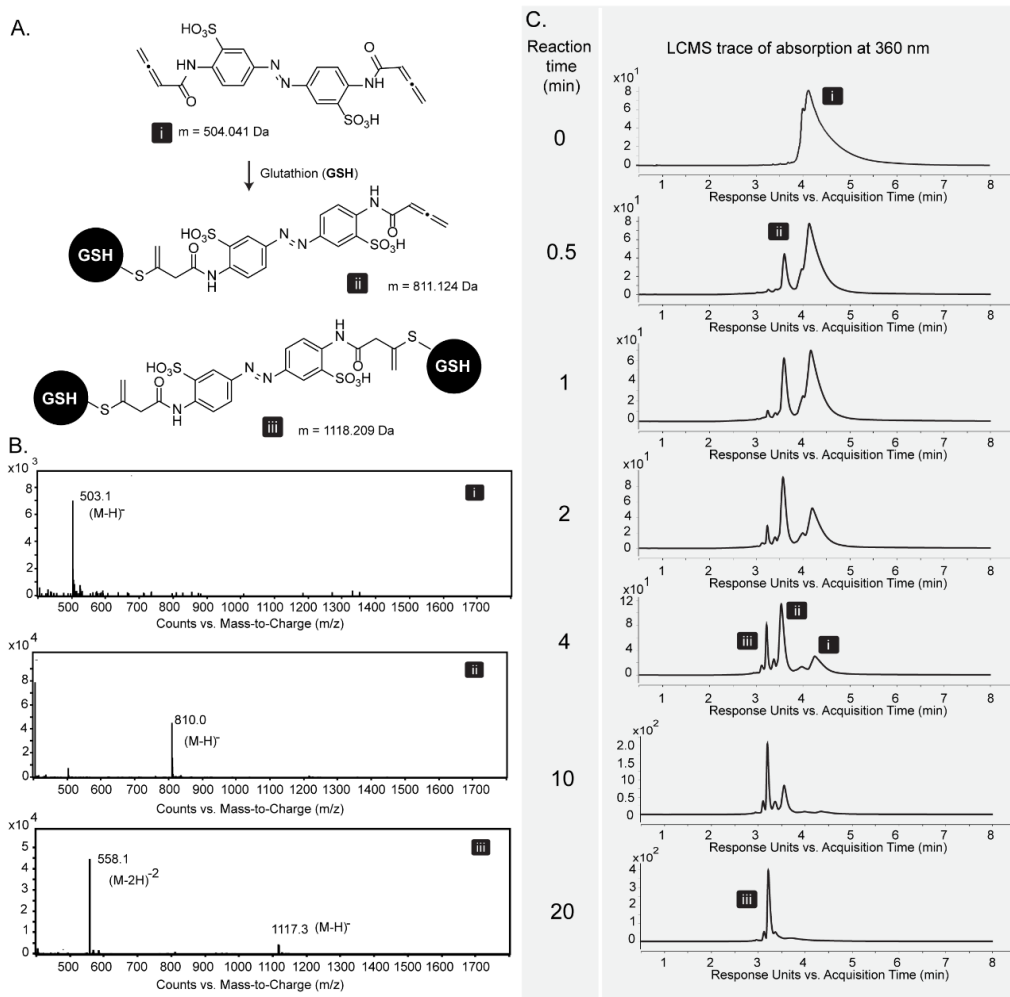
Figure 3.3. Kinetics of reaction of BSBDA with GSH. A) Reaction of BSBDA (**14**) with glutathione can result in formation of mono- and di-substituted product. B) Mass spectra of BSBDA (i), mono-substituted product (ii) and di-substituted product (iii). C) LCMS trace of the reaction of BSBDA with GSH shows the consumption of BSBDA and formation of di-substituted product in 20 min.

BSBDA (160 µM) in the presence of BSBCA (160 µM). The linear peptide reacted with BSBDA to form **16** with > 85% conversions in two minutes, while no significant reaction with BSBCA was observed in 10 min (Figure 3.6). The reaction of 5
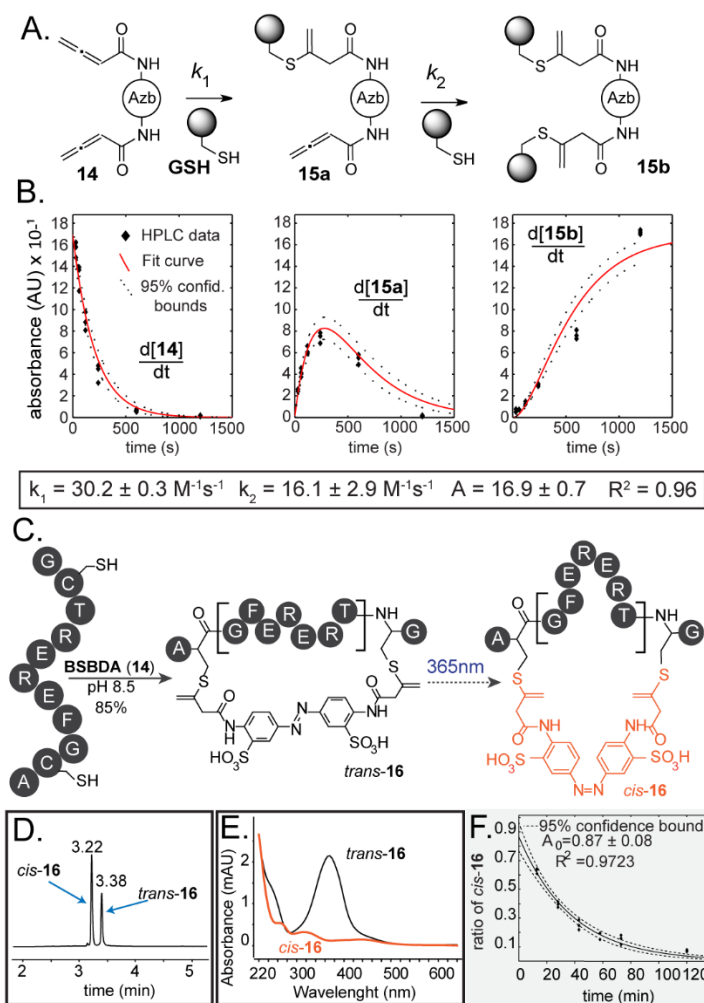
Figure 3.4. Reaction of BSBDA with GSH and **P36**. A) Reaction of the BSBDA (20 μM) with GSH (160 μM) results in production of the mono- and finally di-substituted product in 20 min (see Figure 3.3). B) Fitting the kinetics data into a consecutive first order model determined the $k_1 = 30.2$ M^{-1}s^{-1} and $k_2 = 16.1$ M^{-1}s^{-1}. C) Schematic reaction of the peptide **P36** with BSBDA yields *trans*-**16** which can be isomerized to *cis* form by 365 nm light. D) The *trans*-**16** and the *cis*-**16** can be separated on C18 column (see Figure 3.5 for

Figure 3.4. (continued)

complete HPLC traces). E) Changes in UV spectrum upon isomerization of the *trans*-**16** to *cis* isomer. F) Measuring the ratio of the *trans* isomer at 15 min intervals after irradiation can be used to extrapolate the ratio of the isomers at t = 0 min.
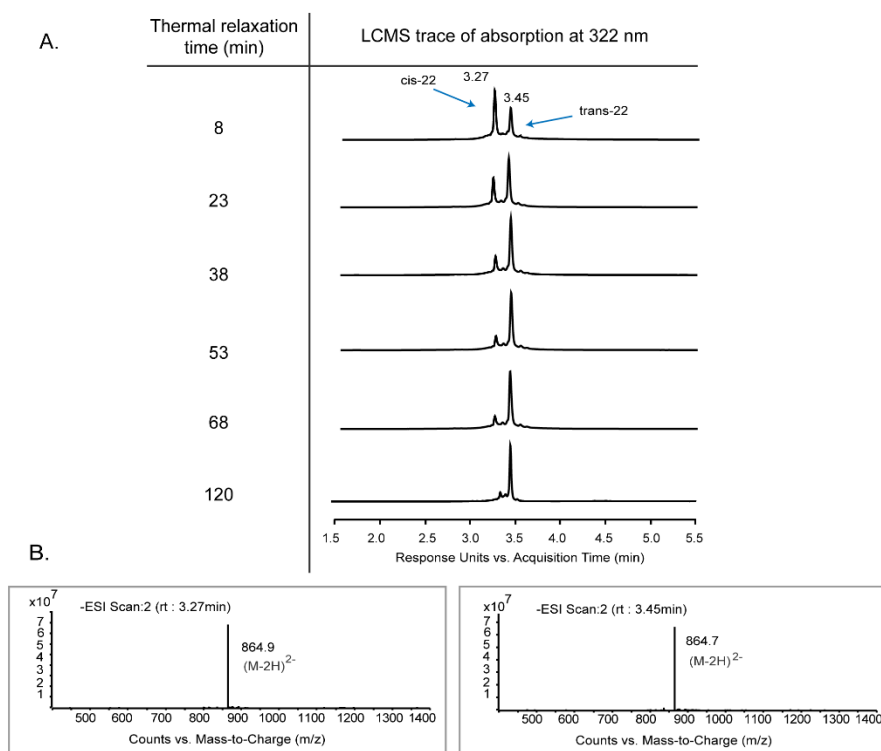
Figure 3.5. Kinetics of thermal relaxation of **16.** A) *Cis*- and *trans*-**16** can be isolated on C8 column using LCMS instrument after thermal relaxation. Changes in absorption of **16** at 322 nm (isosbestic point) 8, 23, 38, 53, 68 and 120 min of thermal relaxation after irradiation with 365 nm LED. B) Mass spectra of *cis* and *trans*-**16**.
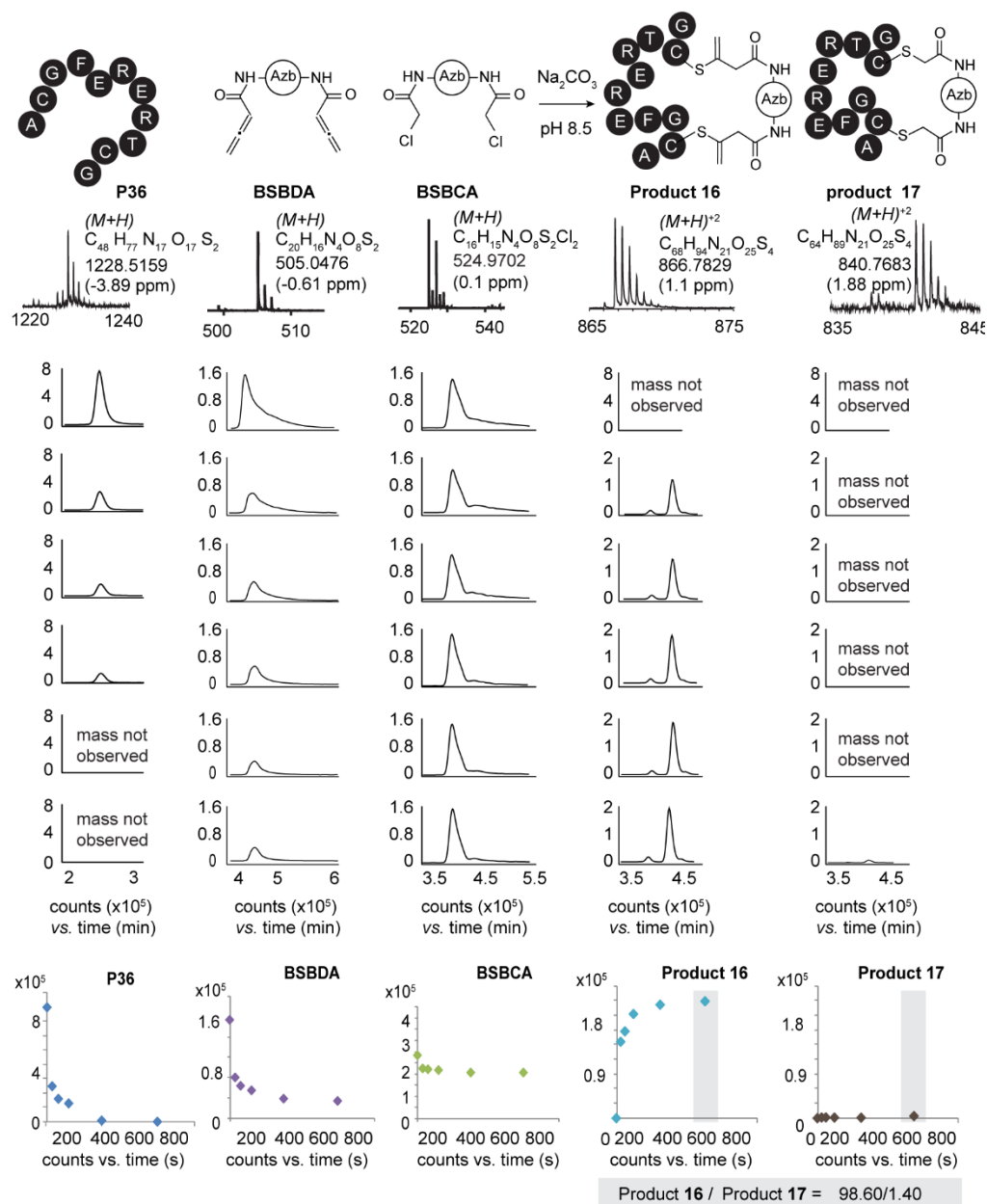
Figure 3.6. Cyclization of **P36** (100 µM) with BSBDA (160 µM) in the presence of BSBCA (160 µM). Over 85% of the product **16** was formed in 2 min. All of the peptide was consumed in 5 min. The ratio of product **16** to product **17** was 98.60/1.40 after 10 min.

mg of **P36** with BSBDA, resulted in 82% isolated yield (Figure 3.4C). The cyclized peptide **16** showed absorption spectra with maximum absorption at 370 nm, characteristic of peptide macrocycles containing similar azobenzene structure. Peptide **16** successfully switched to the *cis* isomer upon irradiation with a 365 nm LED, as indicated by a decrease in the 370 nm absorption (Figure 3.4E). Measuring the amount of *cis* and *trans* isomers by HPLC at 322 nm (isosbestic point) at 15 min intervals, determined the ratio of *cis* isomer to total ligand at photostationary state to be (87 ± 8.0) % (Figure 3.4D- 3.4F, Figure 3.5).

We confirmed the cyclic structure of **16** by NMR spectroscopy (Table 3.1). The distinctive signals for the vinylic hydrogens appeared as four singlets in the region of 5.2 ppm - 5.7 ppm (Figure 3.7A). The resolution of the geminal hydrogens by 0.2 ppm made it possible to deconvolute the structure of **16**, because vinylic hydrogens experience a different microenvironment in the context of a cyclic peptide, although the BSBDA skeleton is symmetric. The <u>T</u>ransverse <u>R</u>otating <u>F</u>rame <u>O</u>verhauser <u>E</u>ffect <u>S</u>pectroscop<u>y</u> (TROESY) experiment showed through-space coupling of "inner" vinylic hydrogens (Figure 3.7A, H5, H5') with the methylene (Figure 3.7A, H_6, H_7, $H_{6'}$, $H_{7'}$) of BSBDA moiety, while the "outer" vinylic hydrogens (Figure 3.7A, H_1, $H_{1'}$) show through space coupling with methylene hydrogens of the cysteines (Figure 3.7A, H_3, H_4, $H_{3'}$, $H_{4'}$).

The $^3J_{\text{HN-HA}}$ of the amino acids in the peptide averaged around 6 Hz (Figure 3.7B, 3.7C), which shows an unstructured backbone in **16**. This result
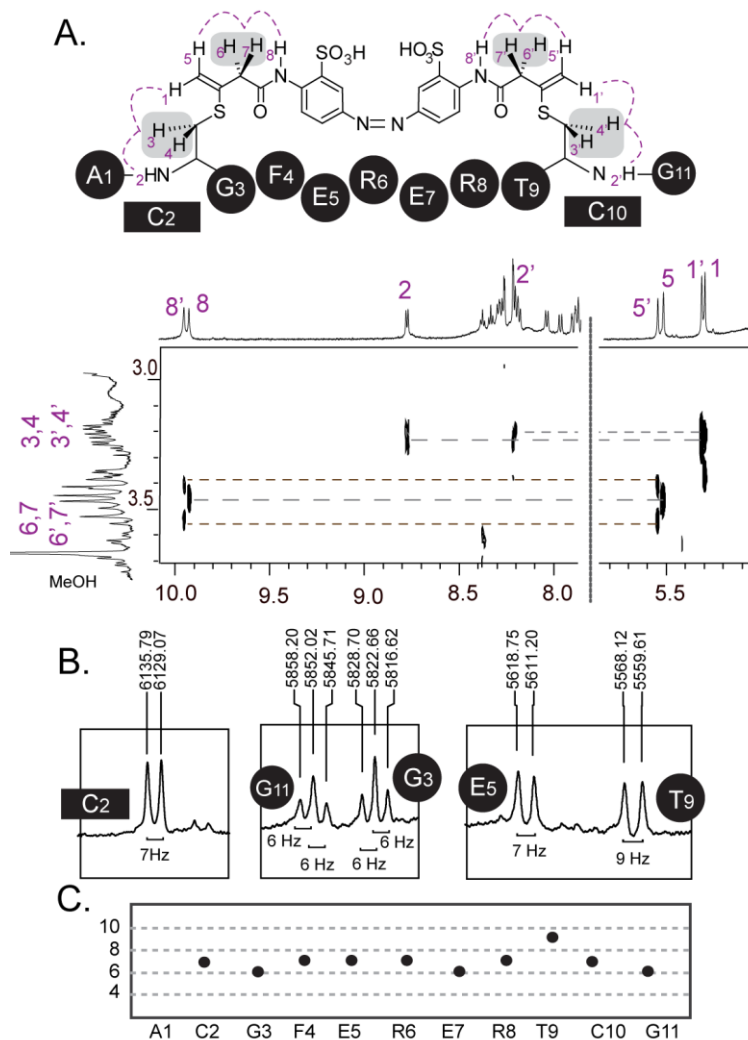
Figure 3.7. NMR spectrum of **16**. A) TROESY cross peaks of the macrocycle **16** shows through-space coupling of the Cys-methylenes (H_3, H_4, $H_{3'}$, $H_{4'}$) with "outer" vinylic hydrogens (H_1, $H_{1'}$) and cysteines' amide hydrogens (H_2, $H_{2'}$) and also coupling of methylene of the BSBDA (H_6, H_7, $H_{6'}$, $H_{7'}$) with "inner" vinylic hydrogens of BSBDA (H_5, $H_{5'}$) and amide-hydrogens of the BSBDA (H_8, $H_{8'}$) at the other side. B) $^3J_{NH\text{-}HA}$ coupling of amino acids in **16** is 6-9 Hz which is typical of an unstructured peptide backbone. NMR was recorded on 700 MHz Agilent system. C) $^3J_{NH\text{-}HA}$ for all amino acids present in *trans*-**16**.

102

is similar to our previous observations that showed **P36** has no defined structure upon its cyclization with BSBCA in *trans* conformer.[173] Since our focus in this work was on development of LR-macrocycle, we did not further study binding of **16** to streptavidin.

The fast cyclization reaction of BSBDA prompted us to study the same reaction with a peptide displayed on coat protein of phage M13, because chemical modification of phage displayed peptides and libraries can be used for discovery of new LR-macrocycles.[183-185] To track the reaction on bacteriophage, we used a "pulse-chase" method that was developed in our lab. A mixture of phage displaying sequence ACPARSPLEC on pIII protein and wild type (WT) clone of phage displaying no peptide was mixed with 1 mM tris(2-carboxyethyl)phosphine (TCEP) to reduce the displayed disulfide and then pulsed with BSBDA for 10, 20 and 30 minutes. Chasing with biotin-PEG2-iodoacetamide (**BIA**) decorated all non-reacted thiols with biotin (Figure 3.8A). Capture of the biotinylated phage with streptavidin coated beads revealed the percentage of biotinylated phage at various times of the reaction. In the absence of BSBDA, 70% of the phage can be biotinylated. The fraction decreased to 40%, 3% and <2% after 10 and 30 min of pre-exposure to BSBDA (Figure 3.8B). We observed that more than 46% of the reactive phage was capped with BSBDA in only 10 min which plateaued at 97% and 98% in 20 and 30 min respectively (Figure 3.8C). We observed only a minor biotinylation on the WT phage which indicated that the TCEP and **BIA** only
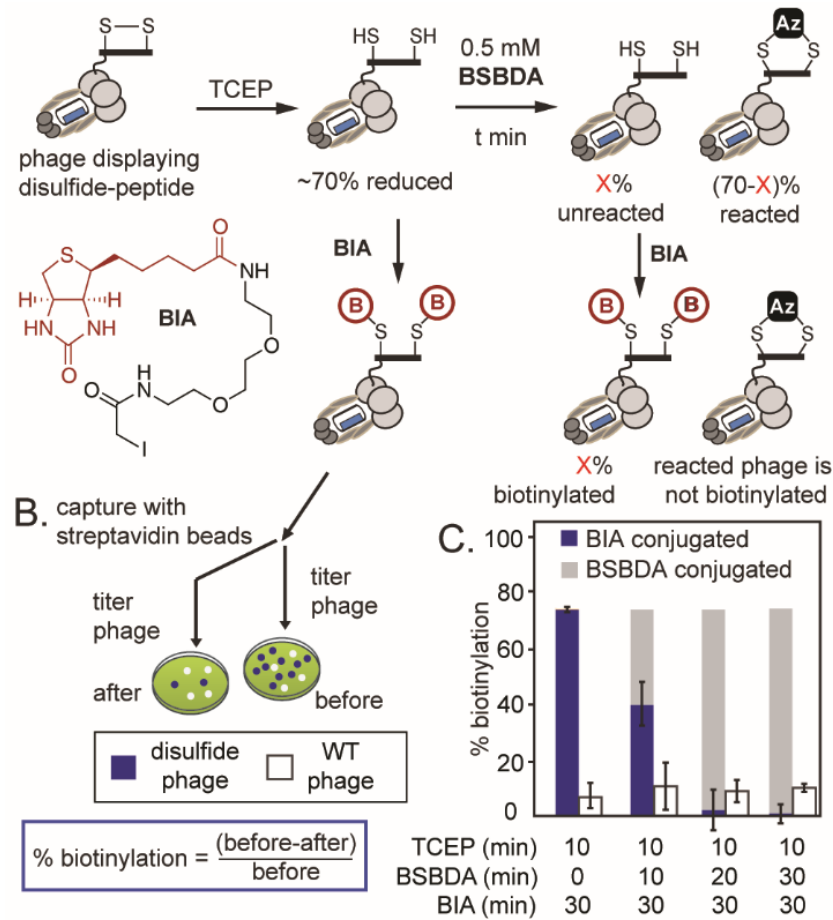
Figure 3.8. Reaction of ACPARSPLEC peptide sequence displayed on bacteriophage M13 with BSBDA. A) Pulse-chase strategy was used to biotinylate the free thiols before and after reaction with BSBDA. B) Reduction of the population of blue phage after capture determines the percentage of biotinylated phage. C) Capture results indicate that about 98% of the phage reacted with BSBDA in 30 minutes. Blue bar shows biotinylated phage and grey bars show BSBDA reacted phage. White bars show WT phage.

react with the displayed peptide. These observations, indirectly, also rule out non-specific reactivity of the thiol reactive BSBCA.

3.3. Conclusion

In summary, we demonstrated an efficient strategy to synthesize the 3,3′-bis(sulfonato)-4,4′-bis-(amino)azobenzene (**6**) precursor to produce LR-linkers for the synthesis of LR- peptides. We also showed fast cyclization of an 11-mer peptide through thiol-allenamide reaction of BSBDA with cysteine containing peptides which can be a suitable strategy to form peptide macrocycles with added properties such as light-responsiveness as well as bioavailability.[18]

3.4. Materials and Methods

3.4.1. General Information

Chemical reagents and solvents were purchased from Sigma-Aldrich or TIC chemicals unless otherwise noted. 3-butynoic acid,[175] 2-azidoacetic acid[186] were synthesized according a previously published protocols and their spectroscopic characteristics matched to the previously reported. 3-butyne-1-ol was purchased from AKScientific. **BSBCA (10)** was synthesized from 3,3'-bis(sulfonato)-4,4'-bis(chloroacetamido)azobenzene (**6**) in 89% yield according to previous protocols and the spectroscopy characteristics matched to the previously reported compound.[177] Peptide sequence ACGFERERTCG was purchased from Lifetein and was purified by HPLC before use. Proton (^{1}H-NMR) and Carbon (^{13}C-NMR) nuclear magnetic resonance spectra were recorded on an Agilent/Varian VNMRS two channel 500 MHz or Agilent/Varian Inova two-channel 400 MHz spectrometer. The chemical shifts are given in part per million (ppm). The solvent peak was used

as a reference. The following abbreviations were used: s, singlet; d, doublet; t, triplet; m, multiplet; b, broad; d, doublet of doublet of doublets. High resolution ESI mass spectra (HRMS) were recorded by Agilent Technologies 6220 oaTOF. Preparative HPLC was conducted using Waters 1525 Binary pump equipped with a Waters Symmetryprep 19×50 mm C18 Columns and Waters 2489 UV detector. We used following method for HPLC purification unless otherwise noted: A gradient of solvent A (0.1% TFA in MQ water) and solvent B (0.1% TFA in MeCN) was run at a flow rate of 8 mL/min (0-2 min 5% B; 2-18 min 5%→50% B; 18-21 min 50%→100% B; 21-24 100% →2% B). LCMS data was obtained on Agilent Technologies 6130 LCMS. A gradient of solvent A (MQ water) and solvent B (MeCN/H_2O 95/5) was run at a flow rate of 0.5 mL/min (0-0.4 min 5% B; 0.4-5 min 5%→60% B; 5-5.5 min 60%→100% B; 5.50-7.50 100% B, 7.50-11 min 100%→5% B).

3.4.2. Measuring the Kinetics of Reaction of Glutathione (GSH) with BSBDA

To 150 μL of 0.1% TFA in H_2O solution, we added 2 μL of a fresh solution of 16 mM GSH in water, followed by 10 μL of 100 μM solution of BSBDA in H_2O. The reaction was triggered by addition of 25 μL of 500 mM MOPS buffer (pH 8.5) to adjust the pH of the mixture to 8.0 and the reaction was mixed by pipetting. We added 30 μL of reaction mixture to 5 μL of glacial acetic acid to stop the reaction at 0.5, 1, 2, 4, 10 and 20 min.
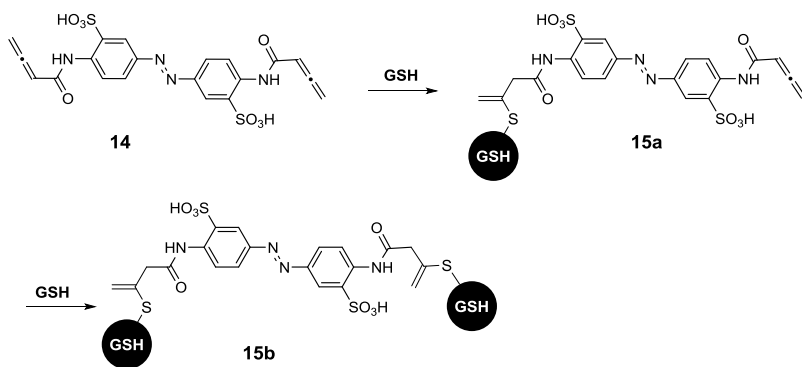
The data was processed using MatLab to fit the reading to consecutive pseudo first order kinetic process:

$$14 \xrightarrow{k1} 15a \xrightarrow{k2} 15b$$
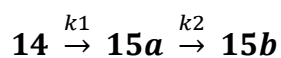
$A_t = A_0 e^{k_1 t}$

$B_t = \frac{k_1 A_0}{k_2 - k_1} \left(e^{-k_1 t} - e^{-k_2 t} \right)$

$C_t = A_0 \left[1 + \frac{1}{k_1 - k_2} \left(k_2 e^{-k_1 t} - k_1 e^{-k_2 t} \right) \right]$

where A_t, B_t and C_t are area under the curve (AUC) for **14**, **15a** and **15b**, respectively, at time t. A_0 is initial AUC and the k_1 and k_2 are the rate constants of the first and second reaction. (Figure 3.3)

3.4.3. Competitive Reaction of P36 with BSBCA and BSBDA

To 10 µL of a 600 µM solution of **P36** in aqueous 1% TFA, we added 10 µL of a fresh solution of 1 mM BSBDA in water and 10 µL of a fresh solution of 1 mM BSBCA in water. The reaction was triggered by addition of 30 µL of 500 mM

Na$_2$CO$_3$ buffer (pH 8.5) to adjust the pH of the mixture to 8.5 and the reaction was mixed by pipetting. We added 10 µL of reaction mixture to 5 µL of glacial acetic acid to stop the reaction at 0.5, 1, 2, 4, 10 and 20 min and injected the solution to the LC/HRMS system as described above.

3.4.4. Measuring the Ratio of *cis*-16 to total-16 at Photostationary State

We prepared a 0.2 mM solution of **16** in 50 mM NH$_4$OAc (pH 7.2) in water. We transferred 100 µL of this solution to a 300 µL glass vial insert and irradiated the solution with 365 nm LED for 2 min. A 10 µL of the solution was then injected to LCMS system to separate the *cis* and *trans* isomers on C8 column at 15 minutes intervals. The absorption of each isomer at 322 nm (isosbestic point) showed the amount of each isomer in the solution. The data was fitted into first order kinetics: $A_t = A_0 + A(1-e^{-kCt})$ using MatLab software. The A_0 was then interpolated to give the amount of *cis* and *trans* isomer at t$_0$ (Figure 3.5).

3.4.5. Measuring the Biotinylation of the Phage with BIA

To 96.5 µL of a solution of 10^{11} pfu/mL of phage (sequence AC-PARSPLEC/WT = 5/1) in 50 mM Tris buffer (pH 8.5), we added 2.5 µL of 20 mM TCEP followed by 1 µL of 100 mM BIA. We incubated this reaction for 30 min at r.t. and then stopped the reaction by diluting it 10^7 times in PBS. The reaction was split in two aliquots of A and B. We added aliquot A to 10 µL of pre-washed strep-tavidin coated magnetic beads (Promega) and agitated the mixture for 10 min at

1000 rpm using IKA MS 3 digital orbital shaker. Then we captured the beads on a magnetic stand and titered the supernatant of A and the aliquot B. The amount of biotinylation was calculated as:

%Biotinylated phage = [(titer of B – titer of A) / titer of B] × 100%

3.4.6. Measuring the Reaction of the Phage with BSBDA

To 96.5 µL of a solution of 10^{11} pfu/mL of phage (sequence AC-PARSPLEC/WT = 5/1) in 50 mM Tris buffer (pH 8.5), we added 2.5 µL of 20 mM TCEP followed by 2.5 µL of 20 mM BSBDA. We incubated this reaction for 30 min at r.t. and then added 1 µL of a 100 mM solution of BIA in DMF. We stopped the reaction by diluting it 10^7 times in PBS. The reaction was split in two aliquots of A and B. We added aliquot A to 10 µL of pre-washed streptavidin coated magnetic beads (Promega) and agitated the mixture for 10 min at 1000 rpm using IKA MS 3 digital orbital shaker. Then we captured the beads on a magnetic stand and titered the supernatant of A and the aliquot B. The amount of biotinylation was calculated as:

%Biotinylated phage = [(titer of B – titer of A) / titer of B] × 100%

Reacted Phage = 100% - %biotinylated phage

3.4.7. Synthesis of BSBDA and Intermediates

Synthesis of 2-(dodecanoylamino)-5-nitrobenzenesulfonic acid (**2**)
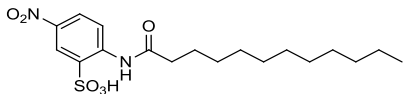
To a solution of sodium 2-amino-5-nitrobenzenesulfonate (1.20 g, 5.0 mmol, TCI Chemicals) in ice-cold dry DMF, we added lauroyl chloride (1.1 mL, 5.5 mmol, TCI Chemicals) drop wise. The reaction was warmed up to room temperature and then stirred at 110 °C for 4 h. The reaction was cooled down to r.t., the solvent was evaporated under vacuum and the resulting crude mixture was re-dissolved in 50 mL of DCM. The DCM solution was washed with 25 mL of 2 M HCl, organic layer was separated and the DCM was removed by rotary evaporator. The resulting vis-cose liquid was purified by column chromatography and the product was eluted using DCM/MeOH = 8/2. The solvent was removed by rotary evaporator and the resulting pale yellow oil was incubated at room temperature under vacuum for 24 h to yield compound **9** as pale yellow solid (1.89 g, 92%). **^{1}H-NMR** (399.8 MHz, DMSO) δ 10.66 (br, 1 H, NH), 8.56 (d, J = 9.2 Hz, 2 H), 8.43 (d, J = 2.8 Hz, 2 H), 8.19 (d, J = 2.8 Hz, J = 9.2 Hz, 2 H), 2.34 (t, J = 7.6 Hz, 2 H), 1.59 (m, 2 H), 1.2 (m, 16 H), 0.81 (t, J = 6.8 Hz, 3 H) ppm. **^{13}C-NMR** (100.5 MHz, DMSO) δ 171.8, 141.4, 141.4, 135.6, 125.9, 122.9, 119.5, 38.1, 31.7, 29.5, 29.4, 29.2, 28.9, 25.1, 225.5, 14.3 ppm. **HRMS-ESI** (*m/z*)**:** (M-H)⁻ calc. for: $C_{18}H_{27}N_2O_6S$: 399.1595, ob-served: 399.1594, 0.29 ppm.

Synthesis of 5-amino-2-(dodecanoylamino)benzenesulfonic acid (**3**)
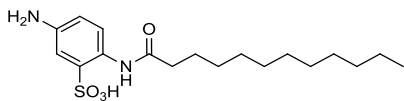
To a solution of **2** (1.2 g, 3 mmol) in 100 mL of MeOH, we added Pd/C (100 mg, 0.1 mmol) with vigorous stirring. We put a septum on the flask and passed hydrogen through a needle for 2 min. Then we attached a balloon on the septum, filled it with hydrogen (~2 L) and let the reaction mixture to stir for 2 h. After completion of the reaction, we filtered the reaction mixture through celite and removed the solvent by rotary evaporator. The resulting solid was purified on silica by column chromatography and eluted by DCM/MeOH = 8/2 to yield **3** as pale yellow solid (1.09g, 96%). **^{1}H-NMR** (399.8 MHz, DMSO) δ 10.0 (s, 1 H), 7.99 (s, J = 8.4 Hz, 2 H), 7.1 (d, J = 2.4 Hz, 2 H), 6.61 (dd, J = 2.4 Hz, J = 8.4 Hz, 2 H), 2.20 (t, J = 7.2 Hz, 2 H), 1.57 (t, J = 6.8 Hz, 2 H), 1.27-1.17 (m, 16 H), 0.85 (t, J = 6.4 Hz, 3 H) ppm. **^{13}C-NMR** (100.5 MHz, DMSO) δ 170.0, 141.2, 136.5, 126.8, 121.3, 116.4, 114.3, 55.5, 46.2, 38.2, 31.7, 29.5, 29.4, 29.4, 29.2, 29.2, 29.0, 25.5, 22.5, 14.4, 9.0. **HRMS-ESI** (*m/z*): (M-H)$^-$ calc. for: $C_{18}H_{29}N_2O_4S$: 369.1854, observed: 369.1847; 1.84 ppm.

Synthesis of Sodium 2-[(4-methoxybenzyl)amino]-benzenesulfonate (**4**)
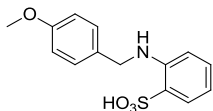
To 3.0 mL of anisaldehyde (30 mmol, TCI chemicals), we added 1.54 g (10 mmol) of 2-aminobenzenesulfonic acid. The resulting suspension was stirred overnight at 110 °C in the presence of activated molecular sieves. We then cooled down the reaction mixture to r.t. and diluted it with 30 mL DCM. The suspension was vacuum filtered through Whatman grade 1 filter paper and the yellow precipitate was washed 5 times with DCM. We then collected the solid residue and re-suspended it in 50 mL MeOH. The suspension was stirred on ice for 30 min and then 370 mg (10 mmol) of sodium borohydride was added in four portions. We stirred the suspension for 2 h on ice followed by 6 h at r.t. After quenching the excess of sodium borohydride with slow addition of ice-cold water to the mixture, we added HCl until the pH of the solution reached ~1 (as determined by universal pH paper). Cooling the solution on ice resulted in precipitation of white solid flakes which were collected by vacuum filtration. Drying the solid *in vacuo* yielded the product **4** as white solids (1.9 g, 65% over two steps). **^{1}H-NMR** (300.0 MHz, DMSO) δ 7.47 (dd, $J = 2.4$ Hz, $J = 10$ Hz, 1 H), 7.27 (dd, $J = 2.8$ Hz, $J = 8.8$ Hz, 2 H), 7.01 (t, $J = 8.8$ Hz, 1 H), 6.87 (dd, $J = 2.8$ Hz, $J = 8.8$ Hz, 2 H), 6.74 (t, $J = 7.6$ Hz, 1 H), 6.49-6.43 (m, 2 H), 4.24 (d, J = 7.6 Hz, 2 H), 3.72 (s, 3 H) ppm. **^{13}C-NMR** (100.5 MHz, DMSO) δ 158.6, 145.2, 132.0, 131.2, 130.4, 128.8, 127.6, 114.6, 114.2, 111.0, 55.5, 46.4 ppm. **HRMS-ESI** (*m/z*): (M-H)$^-$ calc. for: $C_{14}H_{14}NO_4S = 292.0649$, observed: 292.0649; -0.5 ppm.

Synthesis of 3,3'-bis(sulfonato)-(4-methoxybenzyl)amido-(4'-dodecanoylamido)-azobenzene (**5**)
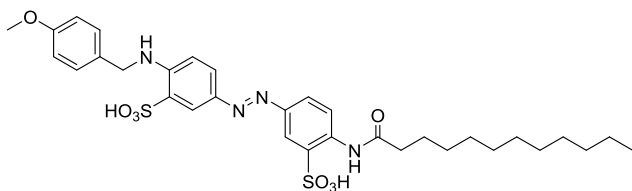
To an ice cold solution of 2-amino-5-nitrobenzenesulfonate (**3**) (1.85 g, 5 mmol) in 2 M HCl (4 mL) and MeOH (50 mL), we added 5 mL of a solution of 1 M NaNO$_2$ drop-wise. Stirring this solution for 1 h yielded the diazonium salt of **3**. We dissolved the PMB-2-ABS (**4**, 1.7 g, 6 mmol) in 2 mL of 2 M aqueous NaOH solution and diluted this solution by adding 20 mL of MeOH. This methanolic solution of **4** was stirred on ice for 20 min. We then added the diazonium salt suspension to the ice cold solution of **4** slowly with vigorous stirring. After stirring for 15 min, we added 100 mL of ice-cold DCM to the mixture and continued stirring vigorously on ice bath for 1 h. We noticed that adding the DCM was critical possibly because of higher solubility of the fatty acid protected diazonium salt in DCM. A 10 cm magnetic stir bar and high stirring speed (2000 rpm) were used to ensure proper mixing of the aqueous and organic phase. We then warmed up the reaction mixture to room temperature and increased the pH to 5.0 by slowly adding saturated solution of NaHCO$_3$ to the reaction. The reaction turned pale orange in 1 h and deep red after 18 h of stirring. We transferred the reaction mixture to a separatory funnel, separated the orange-colored organic layer, extracted the aqueous layer twice with 25 mL of DCM, combined organic layers and then removed the DCM by rotary evaporator. The resulting red crude solid was purified by column chromatography

113

on silica and the product eluted in DCM/MeOH = 3/1. Removing of the solvent by a rotary evaporator yielded compound **5** as a dark orange solid (1.8 g, 55%). **^{1}H-NMR** (399.8 MHz, DMSO) δ 10.6 (s, 1 H), 8.48 (d, J = 8.4 Hz, 1 H), 8.12 (d, J = 2.0 Hz, 1 H), 8.08 (d, J = 2.0 Hz, 1 H), 7.78 (dd, J = 8.8 Hz, J = 2.0 Hz, 1 H), 7.72 (dd, J = 8.8 Hz, J = 2.0 Hz, 1 H), 7.31 (d, J = 8.4 Hz, 2 H), 6.92 (d, J = 8.4 Hz, 2 H), 6.70 (d, J = 8.8 Hz, 1 H), 4.42 (s, 2 H), 3.74 (s, 3 H), 2.31 (t, J = 7.6 Hz, 2 H), 1.62 (t, J = 7.6, 2 H), 1.29-1.14 (m, 16 H), 0.84 (t, J = 6.8, 3 H) ppm. **^{13}C-NMR** (100.5 MHz, DMSO) δ 171.2, 158.8, 147.9, 147.2, 142.0, 136.8, 136.1, 131.5, 1313.1, 128.9, 127.7, 125.4, 121.1, 120.2, 119.3, 114.4, 110.9, 55.5, 46.2, 38.1, 29.5, 29.4, 29.4, 29.3, 29.2, 29.2, 29.0, 25.3, 22.5, 14.4 ppm. **HRMS-ESI** (*m/z*): (M-2H)$^{2-}$ calc. for: $C_{32}H_{40}N_4O_8S_2$: 336.1149, observed: 336.1154; -1.52 ppm.

Synthesis of 3,3′-bis(sulfonato)-4,4′-bis(amino)azobenzene (**6**)
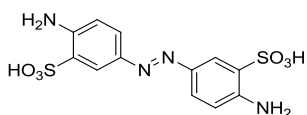
To a suspension of 1.0 g (1.5 mmol) of compound **5** in 30 mL of 30% aqueous methanol, we slowly added 8 mL of 37% HCl in H_2O followed by 0.81 mL of anisole. The resulting solution was heated to reflux for 2 h at 110 °C and then cooled down to r.t. The solvent was removed using rotary evaporator and the resulting solid was washed 5 times with 50 mL dichloromethane. Recrystallization of the product in EtOH/H_2O = 95/5 yielded the product as dark purple crystals (520 mg, 92%). **^{1}H-NMR** (400.0 MHz, D_2O) δ 7.99 (s, 2 H), 7.54 (d, J = 8.4 Hz, 2 H), 6.82

(d, $J = 8.4$ Hz, 2 H) ppm. **^{13}C-NMR** (125.7 MHz, 0.1% TFA in D$_2$O) δ 147.9, 143.8, 126.5, 126.2, 124.7, 118.9 ppm. **HRMS-ESI** (*m/z*): (M-H) calc. for: C$_{12}$H$_{11}$N$_4$O$_6$S$_2$, 371.0125, observed: 371.0135; -2.42 ppm.

Synthesis of 2-(acetylamino)benzenesulfonic acid (**7**)
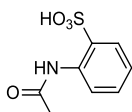
To a solution of 2-aminobenzene sulfonic acid (0.7 g, 4 mmol) in 50 mL of acetic acid, we slowly added 0.36 mL of acetyl chloride (5 mL) and stirred the resulting mixture for 2 h at 100 °C. Removing the solvent by rotary evaporator yielded product **7** as a pale pink solid (0.86 g, quantitative). **^{1}H-NMR** (400.0 MHz, DMSO) δ 10.31 (s, 1 H), 8.21 (d, $J = 8.0$ Hz, 1 H), 7.62 (dd, $J = 7.6$ Hz, $J = 1.6$ Hz, 1 H), 7.24 (dt, $J = 8.4$, $J = 1.6$ Hz, 1 H), 6.96 (dt, $J = 1.6$ Hz, $J = 7.6$ Hz, 1 H), 1.99 (s, 3 H) ppm. **^{13}C-NMR** (100.6 MHz, DMSO) δ 168.1, 135.4, 130.2, 127.3, 122.7, 120.15, 25.3 ppm. **HRMS-ESI** (*m/z*): (M-H) calc. for C$_8$H$_8$NO$_4$S: 214.018, observed: 214.0176; 1.85 ppm.

Synthesis of 2-(1,3-dihydro-2*H*-isoindol-2-yl)benzenesulfonic acid (**9**)
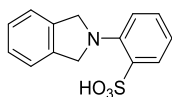
To a suspension of 2-aminobenzene sulfonic acid (0.34 g, 2 mmol) in 15 mL DMF, we added of dibromomethylbenzene (0.58 g, 2.2 mmol) and NaOH (100

mg, 2.2 mmol). We warmed up the mixture to 90 ° C using oil bath and then stirred

the resulting solution for 1 h. After removal of the solvent by rotary evaporator we

purified the product by column chromatography. The product was eluted by

DCM/MeOH = 95/5. Removing the solvent by rotary evaporator yielded the prod-

uct as grey powder (0.37 g, 67.55%). **¹H-NMR** (400.0 MHz, DMSO) δ 7.85 (dd, *J*

= 2.4, *J* = 5.6 Hz, 1 H), 7.65 (dd, *J* = 2.4 Hz, *J* = 5.6 Hz, 1 H), 7.56 (m, 2 H), 7.43

(m, 4 H), 5.20 (s, 4 H) ppm. **¹³C-NMR** (100.5 MHz, DMSO) δ 139.3, 137.6, 134.4,

132.4, 130.1, 129.3, 128.2, 123.4, 121.6, 63.4 ppm. **HRMS-ESI** (*m/z*): (M-H) calc.

for $C_{14}H_{12}NO_3S$: 274.0543, observed: 274.0541; -0.99 ppm.

Synthesis of 3,3'-bis(sulfonato)-4,4'-bis(bromoacetamido)azobenzene (**11**)
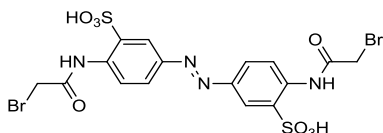
To an ice cold solution of 3,3′-bis(sulfonato)-4,4′-bis(amino)azoben-

zene (**6**, 50 mg, 0.13 mmol) in 5 mL of dry DMF, we added 35 μL (0.39 mmol) of

bromoacetyl bromide, followed by addition of 135 μL of DIPEA (0.78 mmol) with

vigorous stirring. The solution was warmed up to r.t. in 15 min and was stirred for

an additional 30 min in that temperature. The progress of the reaction was moni-

tored by LCMS. The reaction mixture was purified by HPLC on a C18 silica col-

umn. The product was eluted in H_2O/MeCN:70/30. After collecting the fractions,

the solvent volume was reduced to 1.0 mL on a rotary evaporator and the solution

was lyophilized to yield product **11** as a dark red powder (63 mg, 79%). We note

that the same reaction did not yield the desired product when bromoacetyl chloride was used as the acylating agent, because the chloride ion displaced the bromide during the course of the reaction. **^{1}H-NMR** (400.0 MHz, D_2O) δ 8.30 (d, $J = 2$ Hz, 2 H), 8.20 (d, $J = 8.8$ Hz, 2 H), 7.97 (d, $J = 8.8$ Hz, $J = 2$ Hz, 2 H), 4.25 (s, 4 H) ppm. **^{13}C-NMR** (125.6 MHz, D_2O) δ 168.6, 149.1, 136.7, 134.9, 127.0, 124.9, 122.8, 30.2 ppm. **HRMS-ESI** (*m/z*): (M-H) calc. for: $C_{16}H_{13}Br_2N_4O_8S_2$: 610.8547, observed: 610.8555; 1.3 ppm.

Synthesis of 3,3'-bis(sulfonato)-4,4'-bis(azidoacetamido)azobenzene (**12**)
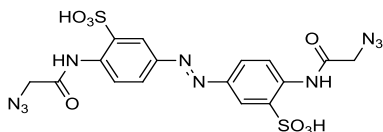
To an ice cold solution of azidoacetic acid (0.16 mL, 2.1 mmol) in 0.4 mL of chloroform, we added 1.8 mL (2.1 mmol) oxalyl chloride. After stirring for 15 minutes at 0 °C, we added 10 μL of DMF to the mixture. The reaction mixture was stirred for 15 min on ice, followed by warming up to r.t. and stirring for an additional 90 minutes until the effervescence was stopped. We then added this solution to an ice cold solution of 3,3′-bis(sulfonato)-4,4′-bis(amino)azobenzene (**6,** 50 mg, 0.13 mmol) in 5 mL of dry DMF followed by addition of 0.7 mL of DIPEA. The reaction was warmed up to r.t. and stirred for 15 min. The reaction mixture was purified by HPLC on a C18 silica column. The product was eluted in H_2O/MeCN = 70/30. After collecting the fractions, the solvent volume was reduced to 1.0 mL on a rotary evaporator and the solution was lyophilized to yield product **12** as a dark

red solid (64 mg, 89%). **^{1}H-NMR** (400.0 MHz, DMSO) δ 11.06 (s, 2 H), 8.55 (d,

J = 8.8 Hz, 2 H), 8.20 (d, J = 2.4 Hz, 2 H), 7.96 (dd, J = 8.8 Hz, J = 2.4 Hz, 2 H),

4.25 (s, 4 H) ppm. **^{13}C-NMR** (125.7 MHz, DMSO) δ 166.1, 146.8, 136.8, 136.4,

126.0, 120.16, 119.4, 52.3 ppm. **HRMS-ESI** (*m/z*): (M-H) calc. for:

$C_{16}H_{13}N_{10}O_8S_2$: 537.0365, observed: 537.0369; -0.86 ppm.

Synthesis of 3,3'-bis(sulfonato)-4,4'-bis(prop-2-ynoylamido)azobenzene (**13**)
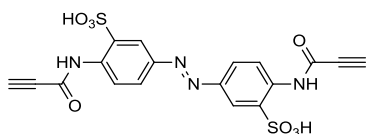
To 0.10 mL of ice cold propiolic acid (0.01 g, 1.4 mmol), we slowly added

0.11 mL (0.16 g, 1.4 mmol) of oxalyl chloride at r.t. After stirring for 15 min, we

cooled down the reaction to 0 °C on ice bath and added 10 μL of dry DMF. The

reaction mixture was stirred on ice for 15 min and then for an additional 45 min at

r.t. until effervescence stopped. We then added this solution drop-wise, using a Gil-

son P200 pipette, to a solution of 3,3′-bis(sulfonato)-4,4′-bis(amino)azobenzene (**6**,

20 mg, 0.05 mmol) in 2 mL of 1.0 M sodium carbonate in water, at r.t. over 15 min.

We stirred the reaction mixture for an additional 30 min at r.t The reaction mixture

was purified by HPLC on a C18 silica column. The product was eluted in

H_2O/MeCN = 60/40. After collecting the fractions, the solvent volume was reduced

to 1.0 mL on a rotary evaporator and the solution was lyophilized to yield product

as a yellow-orange solid (17.4 mg, 69%). **^{1}H-NMR** (400.0 MHz, DMSO) δ 8.45

(d, J = 8.8 Hz, 2 H), 8.20 (d, J = 2.4 Hz, 2 H), 7.98 (dd, J = 8.8 Hz, J = 2.4 Hz, 2

H), 4.57 (s, 2 H) ppm. **^{13}C-NMR** (125.7 MHz, D$_2$O, 10% DMSO-d_6) δ 152.8, 149.2,

136.2, 134.8, 127.0, 125.2, 122.9, 79.1, 77.3 ppm. **HRMS-ESI** (*m/z*): (M-H) calc.

for: C$_{18}$H$_{11}$N$_4$O$_8$S$_2$: 475.0024, observed: 475.0035; 2.4 ppm.

Synthesis of 3,3'-bis(sulfonato)-4,4'-bis(buta-2,3-dienoylamido)azobenzene (**14**)
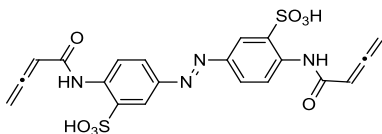
To a suspension of 3-butynoic acid (250 mg, 3 mmol) in 0.5 mL chloroform, we

slowly added 0.26 mL of oxalyl chloride. After stirring for 15 min at r.t., we cooled

down the reaction to 0 °C on ice bath and added 20 µL of dry DMF. The reaction

mixture was stirred on ice for 15 min and then for an additional 90 min at r.t. until

effervescence stopped and solids dissolved completely. We then added this solution

drop-wise, using a Gilson P200 pipette to a solution of 3,3′-bis(sulfonato)-4,4′-

bis(amino)azobenzene (**6,** 100 mg, 0.27 mmol) in 20 mL of saturated sodium bi-

carbonate in water, at r.t. over 15 min. The reaction was stirred for an additional 30

min at r.t. The reaction mixture was washed with 2 × 20 mL DCM, concentrated to

5 mL on a rotary evaporator and purified by HPLC on a C18 silica column. The

product was eluted in H$_2$O/MeCN = 70/30. After collecting the fractions, the sol-

vent volume was reduced to 1.0 mL on a rotary evaporator and the solution was

lyophilized to yield product **14** as a yellow-orange solid (88 mg, 65%). **^{1}H-NMR**

(399.8 MHz, D$_2$O) δ 8.25 (d, *J* = 2.4 Hz, 2 H), 8.23 (d, *J* = 9.2 Hz, 2 H), 7.85 (dd,

J = 2.4 Hz, *J* = 9.2 Hz, 2 H), 5.86 (t, *J* = 6.4 Hz, 2 H), 5.58 (d, *J* = 6.4 Hz, 4 H)

ppm. **^{13}C-NMR** (125.7 MHz, D$_2$O) δ 213.8, 167.0, 148.3, 137.3, 133.4, 127.1, 123.4, 122.8, 91.5, 83.2 ppm. **HRMS-ESI** (*m/z*): (M-H) calc. for C$_{20}$H$_{15}$N$_4$O$_8$S$_2$: 503.0345, observed: 503.0337; 1.57 ppm.

3.4.8. Synthesis of Cyclized Peptide 16

To a 1 mL solution of 5.5 mg of peptide 36 (with amino acid sequence of NH$_2$-ACGFERERTCG-CO$_2$H, 0.004 mmol) in Tris buffer (50 mM, pH 8.5) we added 2.2 mg of BSBDA (**14**), followed by 40 μL of 100 mM aqueous TCEP solution. The solution was incubated for 15 minutes in dark at r.t. before purification by HPLC over C18 column. Collected fractions were concentrated to 1 mL using rotary evaporator to yield the product as a yellow powder (4.2 mg, 82%). **HRMS-ESI** (*m/z*): (M-2H)$^{-2}$ calc. for C$_{68}$H$_{91}$N$_{21}$O$_{25}$S$_4$: 864.7694, observed: 864.7715; 2.37 ppm.

Table 3.1. NMR signal of protons in the cyclized peptide **16.** The hydrogens highlighted in grey in table were exchanged with deuterium and hence were not detectable. Chemical shifts are reported in ppm. 1D-NMR, TOCSY, COSY and ROESY were run using a 3.5 mM peptide solution sample in H$_2$O/D$_2$O (90/10) supplemented with 0.1% trifluoroacetic acid. For the structure of the peptide and hydrogen numbering see Scheme 3.2.

Residue	NH (#)	NH (δ)	Hα (#)	Hα (δ)	Hβ (#)	Hβ (δ)	Hγ (#)	Hγ (δ)	Hδ (#)	Hδ (δ)
Ala	1,2	---	6	4.1	3,4,5	1.48	---	---	---	---
Cys	7	8.77	8	4.65	92,93	3.22	---	---	---	---
Gly	9	8.33	10,11	3.95	---	---	---	---	---	---
Phe	12	7.64	13	4.35	14,15	2.82	---	---	---	---
Glu	21	8.04	22	4.12	23,24	1.89, 1.78	25,26	2.23	---	---
Arg	28	7.87	29	4.13	30,31	1.56, 1.65	32,33	1.4	34,35	2.98
Glu	40	8.18	41	4.22	42,43	1.9, 2.0	44,45	2.4	---	---
Arg	47	8.2	48	4.27	49,50	1.62, 1.73	51,52	1.41	53,54	2.96
Thr	59	7.96	60	4.25	62	4.2	63-65	1.1	---	---
Cys	66	8.21	67	4.65	68,69	3.21	---	---	---	---
Gly	70	8.37	71,72	3.67	---	---	---	---	---	---
BSBDA	78	9.95	---	---	---	---	---	---	---	---
BSBDA	87	9.93	---	---	---	---	---	---	---	---

Residue	functionality	H (#)	H (δ)
Arg	guanidine	55-58, 36-39	---
Phe	aromatic	16, 17, 18, 19, 20	7.15, 7.15, 7.17
BSBDA	aromatic	79-81, 83-85	8.28, 8.27, 8.21, 7.88, 7.85
BSBDA	alkene	74,75	5.51, 5.54
BSBDA	alkene	90,91	5.29,5.31
BSBDA	sulfonate	82, 86	---
BSBDA	methylene	76,77	3.40, 3.45
Gly	carboxylic acid	73	---
Thr	Hydroxyl	61	---
Glu	carboxylic acid	27,46	---
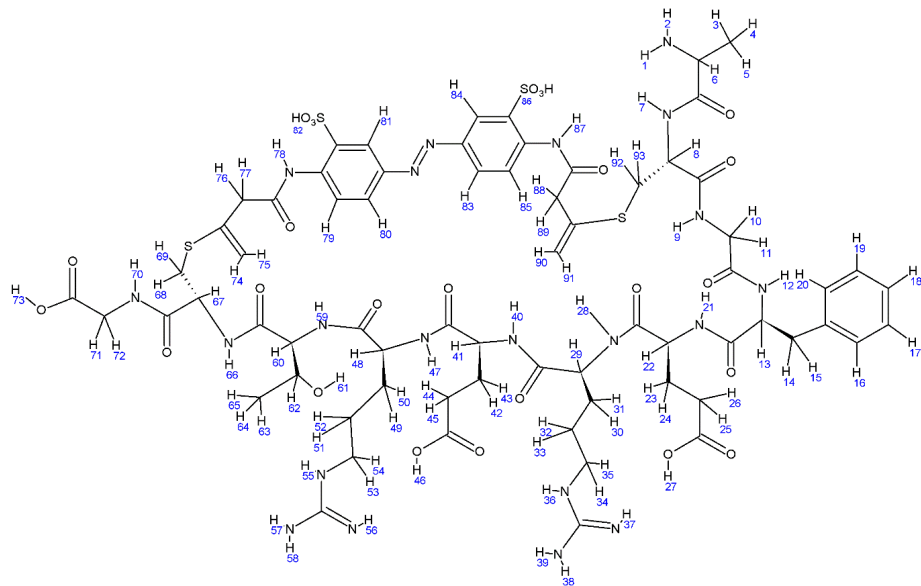
Scheme 3.2. Structure of **16** with explicit hydrogens.

Chapter 4 Light-Responsive Bicyclic Peptides.

4.1. Introduction

In this chapter, we describe a method for synthesis of light-responsive (LR) bicyclic macrocycles from linear peptides composed of 20 natural amino acids. We developed a tridentate C2-symmetric _hydroxyl_ _amine_ and _di-chlorobenzene_ containing _azobenzene_ (HADCAz) LR-linker with two orthogonally reactive functionalities (chlorobenzyl and hydroxylamine) to convert a linear unprotected peptide to a bicyclic peptide in a one-pot two–step reaction. This linker reversibly isomerizes from _trans_ to _cis_ form upon irradiation with blue light (365 nm). The resulting bicyclic peptide contains two loops of amino acids, one of which is constrained with an azobenzene moiety that can change conformation in response to visible light (Figure 4.1A). We used the second static ring to incorporate a constrained cyclic cell permeable peptide sequence.

Peptides and their chemical derivatives constitute many of the biologically active compounds such as hormones, neurotransmitters, toxins and therapeutic agents. There are more than 60 FDA approved peptide therapeutics in the market and more than 140 peptide candidates in clinical trials.[6] Cyclic and bicyclic peptides are an important subclass of peptide therapeutics and biological probes, because they can have high binding affinity,[121,187] increased stability to enzymatic hydrolysis,[121,129] and improved permeability through the cell membrane.[35,188] Macrocycles have shown significant therapeutic potentials[189-191] and 23 out of the 60 FDA approved peptide drugs have cyclic structure.[10]

Figure 4.1. The strategy and linchpins for synthesis of cyclic and bicycle peptides with dynamic and/or static surface. A) Scheme of the synthesis of bicyclic LR-peptides from an unprotected linear peptide made of 20 natural amino acids. B) thioether bond forming linchpins used for generation of monocyclic (**DBMB**) bicyclic (**TBMB**) and monocyclic LR-peptides. C) A bicyclic LR-peptide may contain a static and a dynamic surface, each targeting a specific binding function.

Macrocyclization of peptide chains with small molecule linkers or "linchpins" (see chapter 1) is convenient because the linchpin can contain one or more orthogonal reactive groups that can react with the side chain of unprotected natural amino acids in the peptides. The use of peptides composed of unprotected natural

amino acids and translationally made polypeptide sequences is attractive because it allows modification of peptides of natural origin such as expressed proteins, phage-displayed peptide libraries, mRNA-displayed libraries, etc. In addition to constraining the peptide, synthetic linchpins can also introduce new properties such as extended half-life,[13] cell penetration,[192] and light-responsiveness.[173,177]

To form n cycles in the peptide, $n+1$ points should be connected through a linchpin with $n+1$ reactive functionalities. For example, di- and tri-bromomethyl benzene (**DBMB** and **TBMB** respectively, Figure 4.1B) are classical $C2_V$ and $C3_V$ symmetrical linchpins that have been used to convert linear peptide chains with 2 or three cysteines to mono- and bicyclic peptides.[34] Bidentate $C2_V$ symmetric linchpins that contain LR-moieties, such as azobenzene, have also been used to make LR-cyclic peptides macrocycles. An example of such linchpins is BSBCA, Figure 4.1B).[93,173,177] Devising a LR-linchpin with three attachment point is not trivial, because all functionalities that undergo light-induced isomerization – such as azobenzenes, spyropyrans and diarylethenes – have lower $C2_V$ symmetry. Heinis and co-workers reported a synthesis of a putative $C3_V$-symmetric linchpin for modification of peptides with three cysteines.[193] The authors, however, admitted that the use of two tandem Azb functionalities in a linchpin was not optimal because these moieties isomerize independently, yielding a poor light-switching performance. Use of a non-$C3_V$ symmetric linchpin with three identical reactive groups is impractical because its reaction with the peptide yields 3 or more constitutional isomers.

A viable topological solution for the synthesis of LR-bicycles uses a C2$_V$ symmetric linchpin with 2+1 reactive groups that target 2 and 1 orthogonally-reactive functionalities in a peptide side chain. Two types of LR-linchpins (LRL) can be designed: (i) an LRL with two dynamic loops, both containing an Azb photoswitch and (ii) an LRL with one dynamic loop and one static loop. In this chapter we designed a C2$_V$ linchpin that can make the latter bicycle with dynamic and static loops. We envision that one peptide ring can dynamically change its affinity to the receptor of interest, while the second ring can be reserved to introduce other properties to the molecule, such as-cell permeability.

4.2. Results and Discussion

In our design, the LR-linchpin should meet three criteria: 1) it includes functionalities that can selectively react with two orthogonally-reactive handles present in or derived from natural amino acid side-chains. Herein, we focus on thiol functionalities of cysteine and glyoxal functionality produced from *N*-terminal serine by mild oxidation with sodium periodate.[41,194] 2) The linker should have a half-life of thermal relaxation of more than 10 min after irradiation to provide sufficient time for potential biological studies, such as ligand–receptor interactions. 3) LRL should switch in response to visible light (λ_{max} of > 360 nm), because most biological systems are sensitive to light with λ below 350 nm.[77]

We synthesized four different azobenzene cores (Figure 4.2, **1a-1d**) which carried the suitable functionalities (-OH, -NH$_2$, -COOH) for further installation of
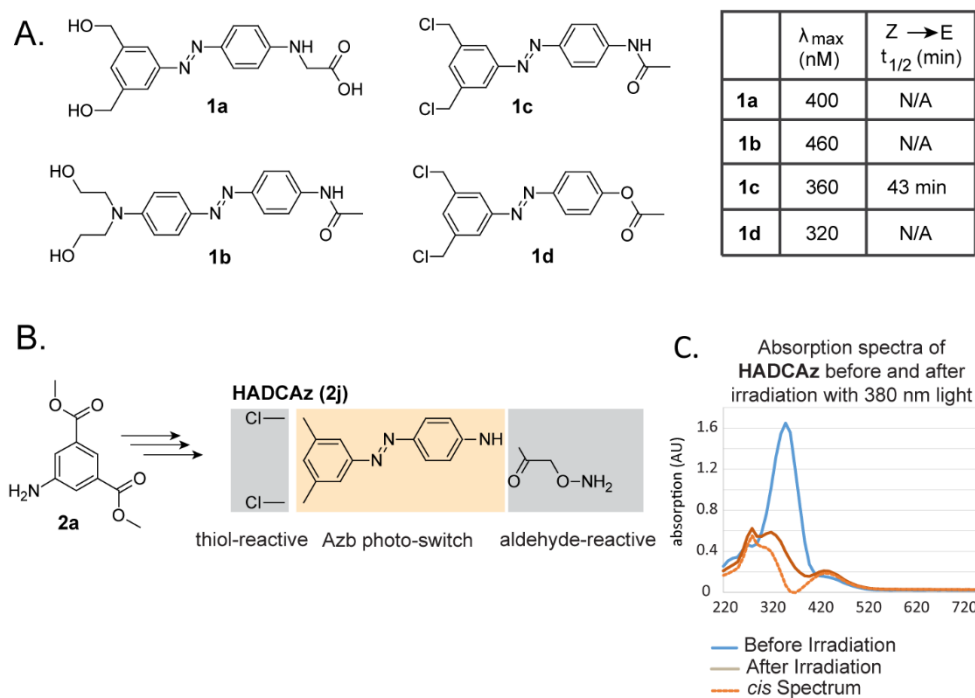
Figure 4.2. Synthesis of Azb cores and LR-linchpin HADCAz. A) Potential Azb precursors to bicyclic macrocycles and their light-induced isomerization properties. *B)* Minimalistic design of the light-responsive linker HADCAz for making bicyclic macrocycles (See Scheme 4.1 for complete synthetic route to HADCAz). C) Absorption spectra of HADCAz before and after irradiation with 380 nm light.

cysteine and glyoxal-reactive groups. We did not observe any changes in spectra after irradiation of compounds **1a** and **1b** with 380 nm and 470 nm LEDs, respectively, most likely due to the fast thermal relaxation of the azobenzene (Figure 4.3A, 4.3B). This is consistent with other published studies,[77] that found single or double alkylation of one or both aromatic amine at either end of the Azb to result

in fast switching. Compound **1c** exhibited desired properties: its *trans* to *cis* isomerization was induced by irradiation with 380 nm and thermal relaxation of *cis* form back to *trans* occurred with a half-life of 43 min (Figure 4.3C).
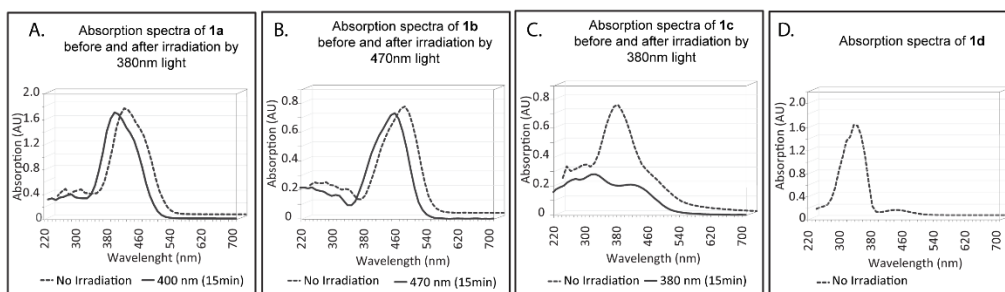
Figure 4.3. A-C) Absorption spectra of compounds **1a-1c** before and after irradiation. D) Absorption spectra of compound **1d**.
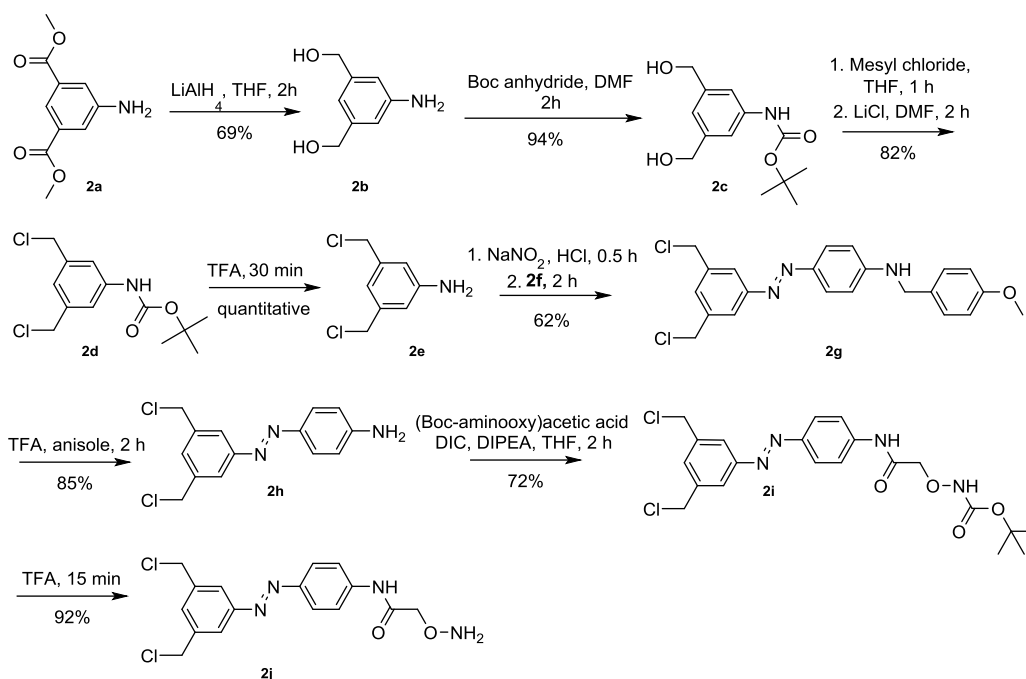
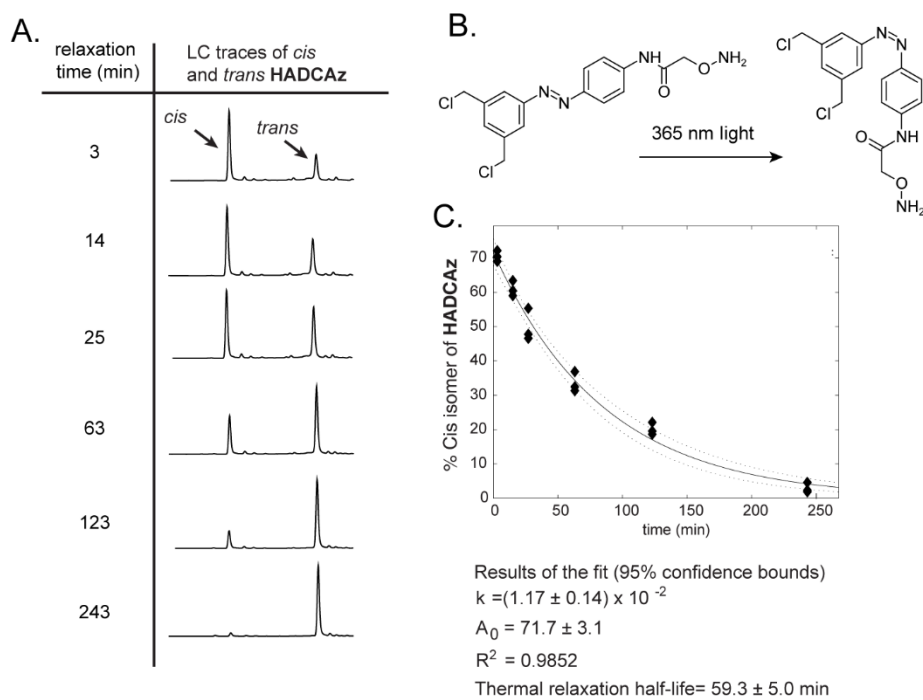**Scheme 4.1.** Synthetic pathway for HADCAz.

Figure 4.4. Thermal relaxation HADCAz. A) LCMS traces of thermal relaxation of HADCAz after irradiation by 365 nm light. B) Scheme of isomerization of *trans*-HADCAz to *cis* form. C) Fitting thermal relaxation data to first order kinetic determined the *cis*-ratio to be 71% at photostationary state after irradiation and thermal relaxation half-life to be 59 min.

Presence of amide linkage in **1c** was crucial for optimal light-response because core **1d** in which amide was replaced by ester bond showed a maximum absorption at 320 nm which is in the bio-incompatible UV region (Figure 4.3D). Based on structure activity relationship of aforementioned azobenzenes, we believed that arrangement of functional groups in azobenzene **1c** yielded near-optimal photochemical properties. Installing desired functional groups onto **1c** Azb core yielded a <u>h</u>ydroxyl <u>a</u>mine, <u>di</u>-<u>c</u>hlorobenzene containing <u>az</u>obenzene (HADCAz)
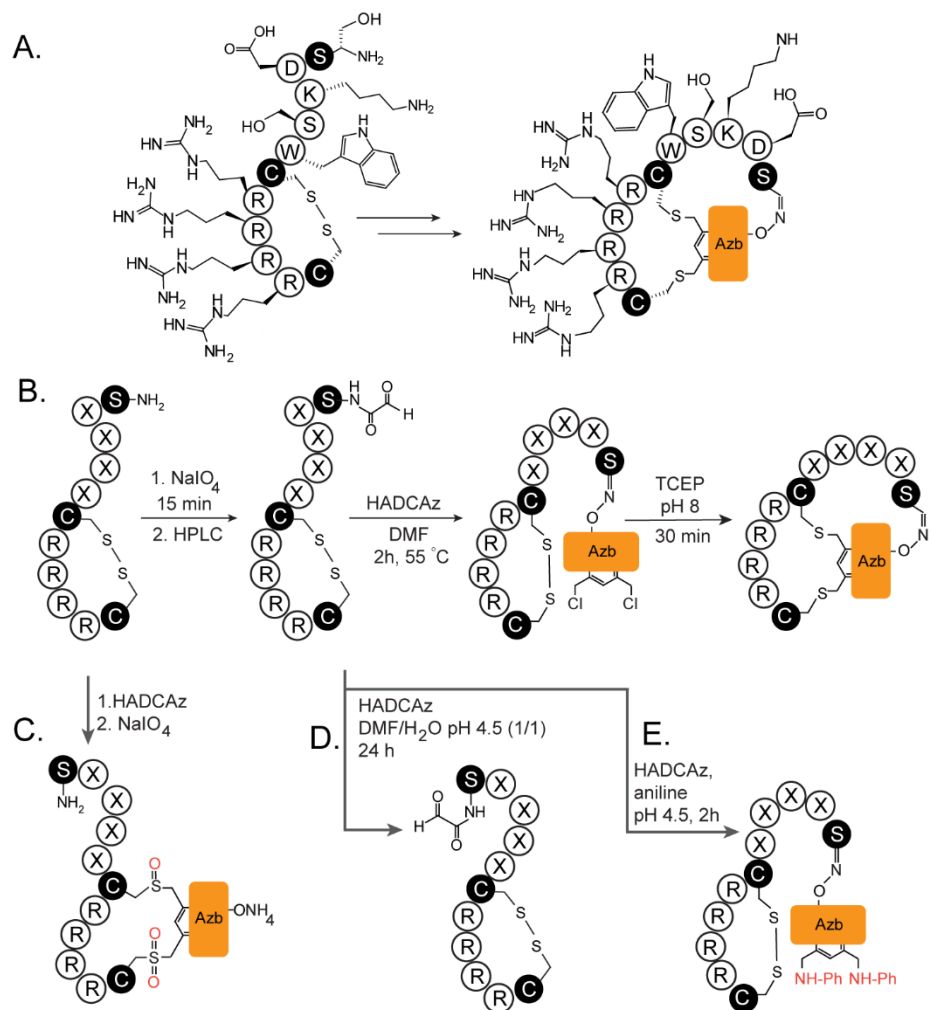
Figure 4.5. Synthetic pathways for generation of LR-bicyclic peptides. A) Site-selective reaction of peptide SDKSWCRRRRC with HADCAz results in formation of bicyclic product while leaving all reactive side chain functionalities intact. B) Successful route to bicyclization starts from oxidative conversion of N-terminal serine to glyoxal followed by reaction with HADCAz to form an oxime bond; subsequent reduction of the disulfide bond with TCEP triggers the intramolecular formation of two thioether bonds to close both cycles. C) The alternative unsuccessful routs to the same product. Formation of the first cycle

Figure 4.5. (continued)

through thioether bonds is not a suitable first step because subsequent periodate oxidation results in the oxidation of thioethers. D) No reaction was observed between the peptide and HADCAz in aqueous conditions. E) Addition of aniline as catalyst results in undesired substitution of chlorine by aniline.

from diethyl-5-(amino)isophtalate in 18.5% overall yield in seven steps (Scheme 4.1). HADCAz exhibited maximum absorption at 350 nm which is close to its parent Azb core **1c**. HADCAz was successfully switched to 71% *cis* isomer at photostationary state after irradiation with 365 nm light and relaxed thermally to *trans* isomer with a half-life of ~60 min (Figure 4.4).

To explore the utility of HADCAz in bicyclization peptides, we investigated the reaction of HADCAz with peptide sequences $SX_nCRRRRC$ where X_n was W, SW, SKD, or DKSW (Figure 4.5A) which contained major reactive functionalities in natural amino acids (amine, hydroxyl, carboxylic acid, indole). *N*-terminal serine can be converted to glyoxal whereas cysteines flank a tetra-arginine sequence that has been used by Pei and co-workers to enhance cell permeability of bi-cyclic peptides.[192]

Oxidation of the disulfide peptide (Figure 4.5B) by adding 2 equivalents of sodium periodate (25 mM) at pH 7.0 yielded a peptide glyoxal with intact disulfide bond. After purification, we lyophilized the peptide glyoxal and then dissolved it in a DMF solution of HADCAz (1.2 eq) to form the oxime bond. Reaction was complete in 2 h at 55 °C (Figure 4.6, 4.7A) and required up to 24 h of incubation
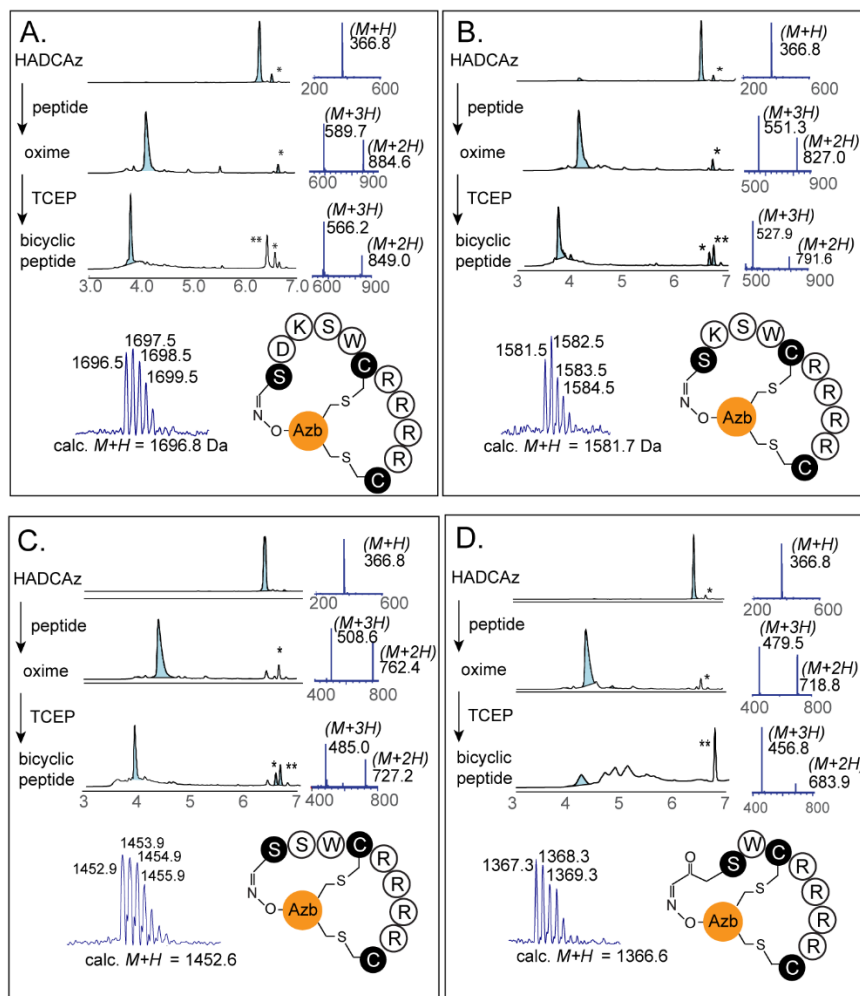
Figure 4.6. LCMS traces for confirming the synthesis of bicyclic peptides **3a-3d**. MALDI-MS shows the isotopic pattern for the product (for high resolution MS see Figure 4.13). * indicates HADCAz adduct of reaction with formaldehyde (DMF impurity). ** indicates HADCAz byproducts with unidentified strucutre.

(Figure 4.7B, 4.7C) at room temperature (r.t). Our attempt to perform oxime ligation in water or even water-organic mixtures such as 50% aqueous DMF (pH 4.5) was   unsuccessful   and   yielded   no
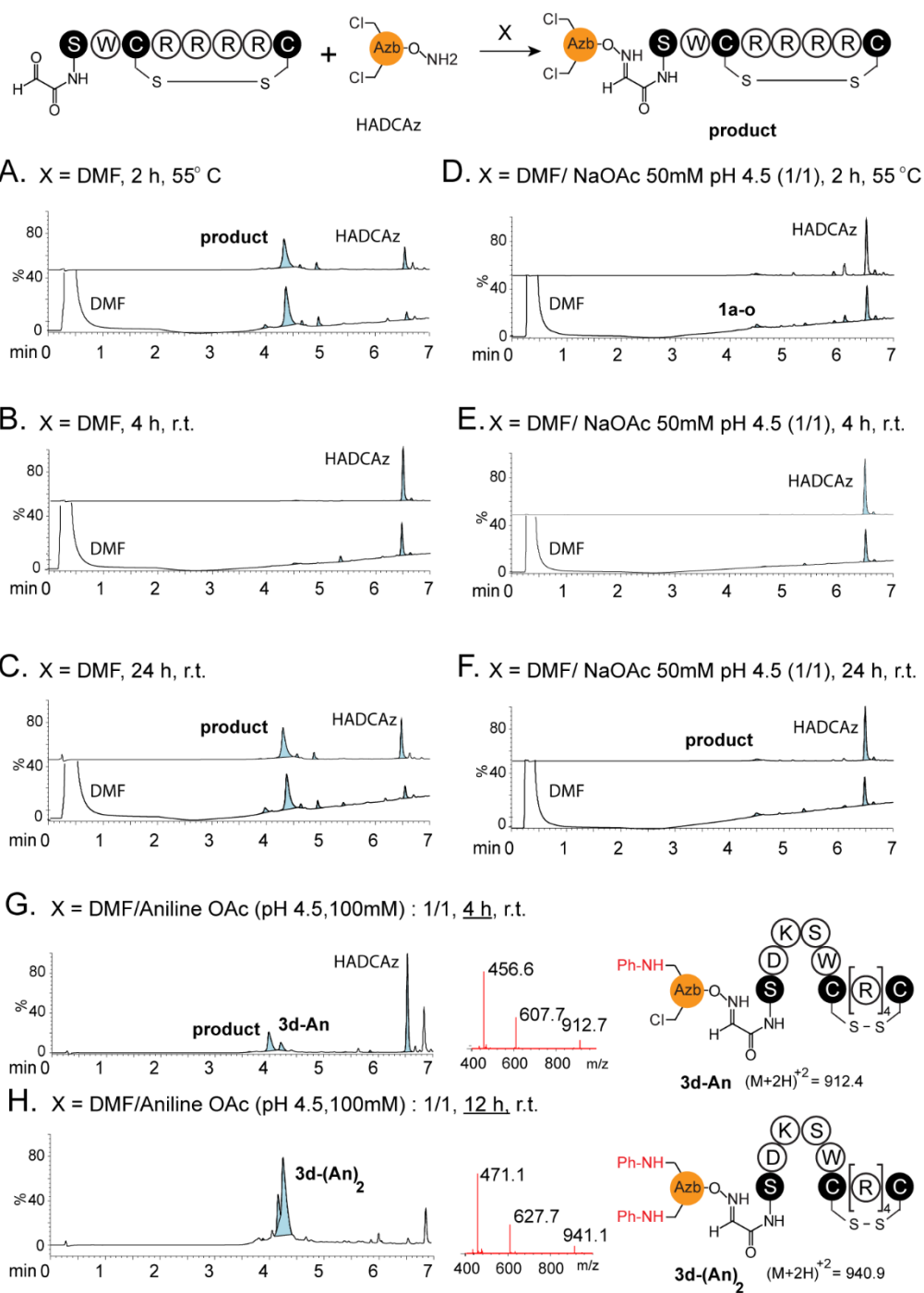
Figure 4.7. Optimization of the formation of oxime bond in the absence and the presence of aniline catalyst (SDKSWRRRRC peptide sequence were used in the latter case). A-F) Formation of products is undetectable in solvents containing

Figure 4.7. (continued)

water, but is occurs efficiently in pure DMF. G-H) Addition of aniline resulted

in undesired substitution of one or both chlorides in the linker.

apparent conversion (Figures 4.7D-4.7F). We note that the reaction could not take

advantage of aniline catalyst,[123] because addition of 100 mM aniline to the reaction

resulted in nucleophilic substitution of benzyl chlorides of HADCAz by aniline

(Figure 4.7G, 4.7H). Once conditions for oxime bond formation was optimized,

the subsequent bicyclization was relatively straightforward. Introducing 3 equiva-

lents of TCEP to the reaction mixture resulted in double intramolecular nucleophilic

substitution of both chlorine atoms with the thiols. Bicyclization occurred effi-

ciently for three of the peptides with the structure of SX_nCR_4C, where X_n is SW,

KSW and DSKW, to result in 74%, 63% and 68% isolated yield (Figure 4.6A-C).

The structures were analyzed by NMR spectroscopy. The isolated yield for the pep-

tide sequence SWCRRRRC was lower (11%) and reaction contained a mixture of

unidentified byproducts (Figure 4.6D). The low yield of the reaction for this se-

quence highlights the lower limit of the bicyclization. The distance between cyste-

ine and glyoxal functionality has to be at least two amino acids. To investigate the

upper limit, we foresee studying the bicyclization with several sequences SX_nCR_4C

where X_n is 10-20 amino acid peptides with known potent binding affinity to intra-

cellular targets (such as Notch, P53, etc.)

We found that making the first cycle by reaction of free thiols with chloro-

benzyls was not a suitable strategy, because the resulting sulfides are significantly

	Sequence	Isolated yield (%)	Relaxation half-life (min)	*cis*-isomer at photo-stationary state(%)
3a	SWCRRRRC	11	144.3 ± 18.7	86.2 ± 2.0
3b	SSWCRRRRC	74	243.1 ± 33.4	91.0 ± 1.9
3c	SKSWCRRRRC	63	117.2 ± 21.0	81.9 ± 1.5
3d	SDKSWCRRRRC	68	29.5 ± 1.3	74.9 ± 3.3

Table 4.1. Isolated yield, thermal relaxation half-life and ratio of *cis*-isomer at photostationary state after irradiation for peptides **3a-3d**.
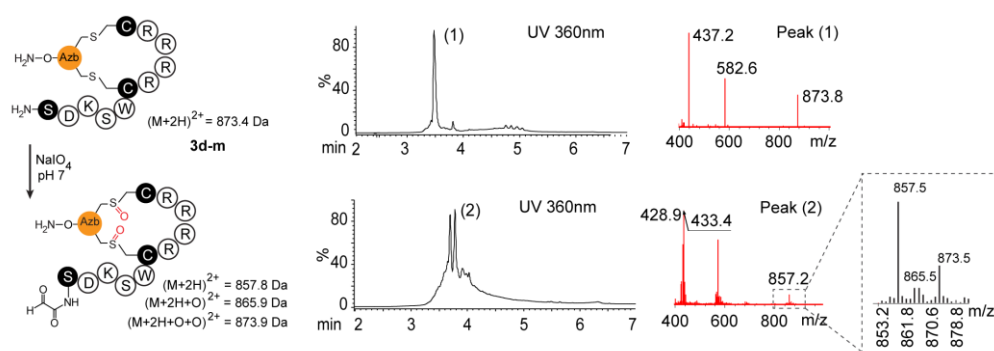
Figure 4.8. Oxidation of **3d-m** results in a mixture of byproducts including oxidized sulfides.

more susceptible to periodate oxidation than the disulfides.[195,196] We observed adducts with +16 Da and +32 Da higher in the mass spectrum, confirming the oxidation of sulfides (Figure 4.8).

Bicyclic products **3a-3d** successfully switched to *cis* isomer after irradiation with 365 nm light, as indicated by a decrease in their absorption at 350 nm (Figure 4.9). We determined the ratio of *cis*-isomer to *trans*-isomer for each peptide in time intervals after irradiation by LCMS (Figure 4.10A) and fitting the data to a first order kinetic model determined the content of the *cis* isomer at at photostationary
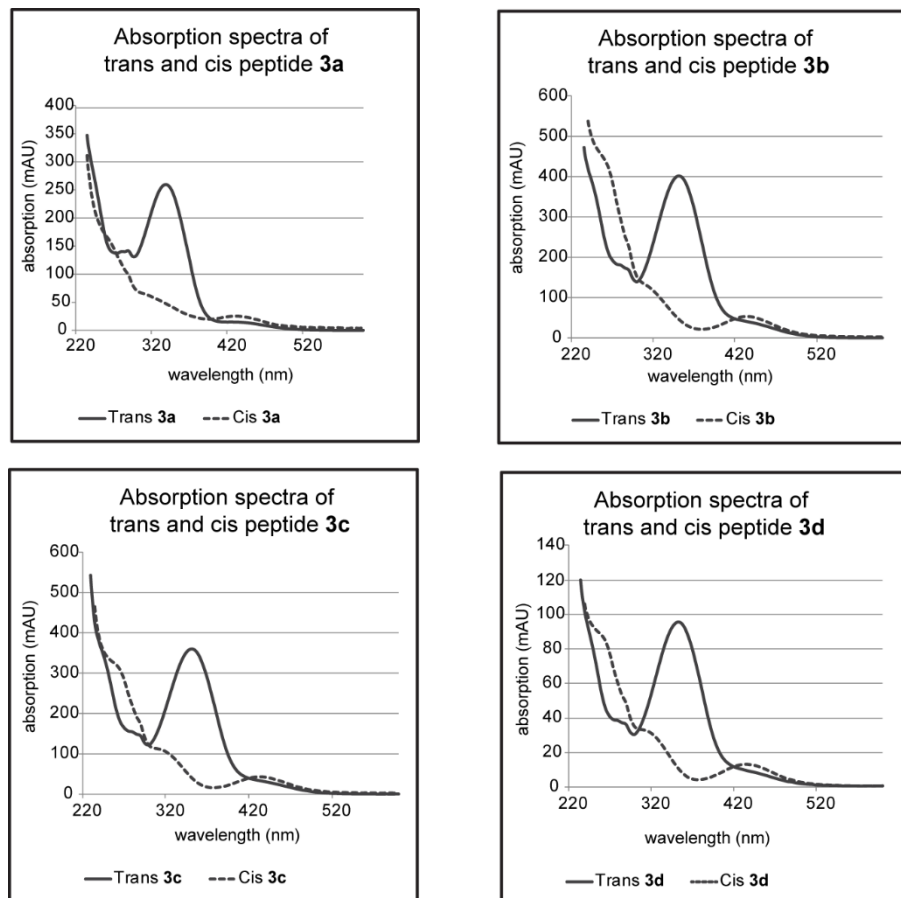
Figure 4.9. Absorption spectra of **3a-3d** before and after irradiation with 365 nm light.

state (Figure 4.10B). Interestingly we found that the ratio of *cis*-isomer at state for the bicyclic peptides **3a-3c** is higher than the HADCAz (Table 4.1). The thermal relaxation half-life for **3a**, **3b** and **3c** was 144, 243 and 117 min, respective which is higher than the HADCAz (59 min). The ratio of *cis*-isomer at photostationary state for **3d** was 75% which was close to the *cis*-ratio of HADCAz; however, its thermal relaxation dropped to 29 min. The higher *cis*-ratio and the relaxation half-life of products could be because of higher strain in the ring in **3a-3c**. We suspect
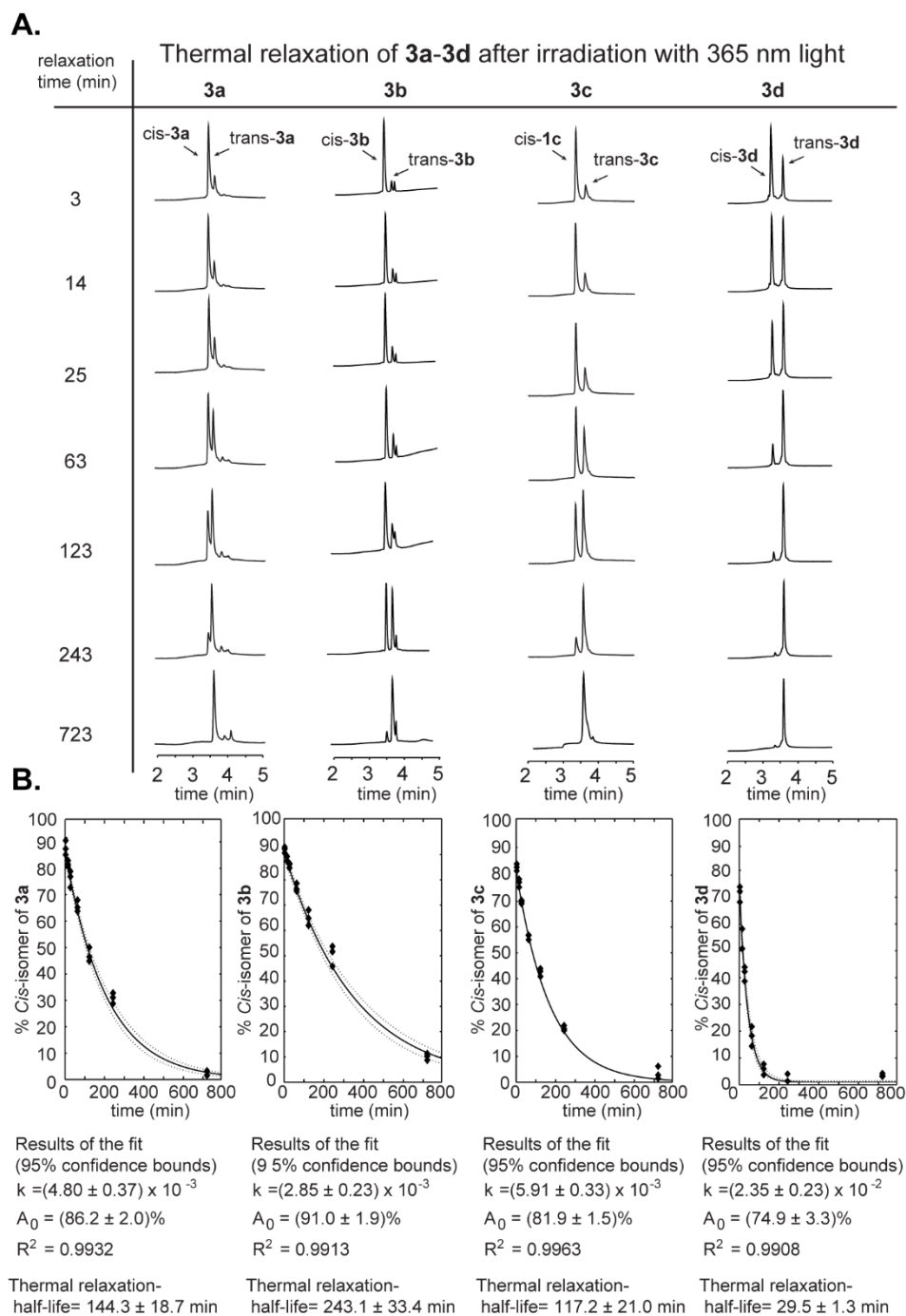
Figure 4.10. Kinetics of thermal relaxation for **3a-3d**. A) LCMS traces for thermal relaxation of **3a-3d** after irradiation with 365 nm light. B) fitting the kinetic data in a first order kinetic model determined the relaxation rate, *cis*-ratio at photostationary state (A_0) and relaxation half-life.

that this modulation of photo-switching could also be sequence-specific but the evaluation of different sequences extends beyond the scope of this chapter.

4.3. Conclusion and Future Studies

In conclusion we developed a C2v symmetric, tridentate LR-linchpin that can react with peptides of structure SX_mCX_nC to form LR-bicyclic macrocycles. The reaction occurs site-specifically on unprotected peptides consisted of only natural amino acids and the cyclization results in satisfactory yields for the peptides that contain a medium sized ring (n >1). The dynamic loop of the bicyclic products (SX_nC), contained an azobenzene moiety which was successfully switched to *cis*-isomer, in response to 365 nm light, for n = 1- 4. The number of amino acids in the loop modulates the switching properties of the product. The static ring of the molecule harbors a tetra arginine sequence which has been shown to increase the permeability of bicyclic peptides into mammalian cells. The platform we described in this chapter can be used for dynamic regulation of cell function *via* binding to intracellular signalling components. For example, Schepartz and co-workers showed that the dimerization of intra-cellular domain of epidermal growth factor receptor (EGFR) can be inhibited by a cyclic cell-permeable hydrocarbon stapled peptide, derived form the intra-cellular domain of this protein (residues 645−662). This peptide inhibits the coiled-coil interaction that is necessary for EGFR activation.[197] We envision that a bicyclic LR-inhibitor based on peptide of Schepartz can not only enter the cell, but also change its inhibition ability in response to temporal and spatial modulation of light. Such modulation can potentially control EGFR activity on

a single cell level. In future projects, we foresee synthesizing a cell-permeable LR-peptide derived from the same sequence and assessing its potentials to inhibit EGFR activation in EGFR-reporter cell lines such as A549-EGFR-GFP and modulate this function in response to light. These and other bicyclic LR-inhibitors, will expand the growing toolbox of chemical optogenetics[78] and photo-pharmacology,[78] and permit the investigation of cell signaling pathways that are not possible to study using conventional tools of chemical genetic.

4.4. Materials and methods

4.4.1. Synthesis of 1a-1d, HADCAz and Intermediates
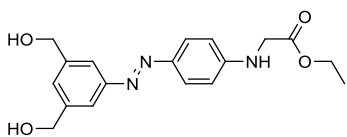
Synthesis of **1a**

To an ice cold solution of 450 mg (2.4 mmol) of compound **2e** (see the following) in 10 mL EtOH, we added 2 mL of 2 M HCl and stirred the solution on ice for 30 min. To this mixture we added a solution of 165 mg (2.4 mmol) of NaNO$_2$ in water drop wise with vigorous stirring for 1 h to form the diazo salt. Then we added the diazo salt solution drop wise to an ice cold solution of 537 mg (3 mmol) of *N*-phenyl-ethylglycine in 50 mL of EtOH/2 M aqueous glycine (0.2 M, pH 5) = 1/1. After stirring for 15 min at this temperature, the solution was warmed up to r.t.

139

and stirred for an additional 2 h. Then we evaporated the solvents on a rotary evaporator and dissolved the solids in a 100 mL mixture of DCM/2 M HCl = 1/1. The organic phase was separated and the aqueous phase was extracted twice with 50 mL of DCM. The organic phases were combined and the solvent was removed on a rotary evaporator. The resulting solid was purified on silica by column chromatography (eluted in hexane/EtOAc = 85/15) to yield the azobenzene **1c** as orange crystals (620 mg, 32.5%). **^{1}H-NMR** (399.8 MHz, CDCl$_3$) δ 7.91 (d, *J* = 8.4 Hz, 2 H), 7.86 (s, 2 H), 7.47 (s, 1 H), 6.71 (d, *J* = 8.4 Hz, 2 H), 4.67 (s, 4 H), 4.29 (q, *J* = 7.2 Hz, 2 H), 4.02 (s, 2 H), 1.33 (t, *J* = 7.2 Hz, 3 H) ppm. **^{13}C-NMR** (100.5 MHz, CDCl$_3$) δ 170.1, 152.7, 150.6, 144.6, 139.0, 129.4, 126.2, 122.2, 112.8, 61.7, 45.26, 45.2, 14.2 ppm. **HRMS-ESI** (*m/z*): (M+H)$^+$ calc. for C$_{18}$H$_{19}$Cl$_2$N$_3$O$_2$: 380.0927, observed: 380.0918.1295; 2.23 ppm.

Synthesis of **1b**
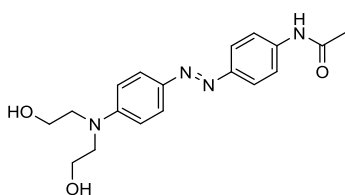
To an ice cold solution of 4-aminoacetanilide (182 mg, 1.2 mmol) in 4 mL of 1 M HCl 1M, we added 1.2 mL of a 1 M solution of NaNO$_2$ drop-wise. After stirring the reaction mixture on ice for 1 h, we slowly added the mixture to an ice cold solution of *N*-phenyldiethanolamine (253 mg 1.4 mmol) in EtOH/H$_2$O = 1/1 with vigorous stirring. We stirred the solution for 3 h, then removed the solvent on a

rotary evaporator and purified the products over silica by column chromatography. The product was eluted by DCM/MeOH = 4/1 to yield the compound **1d** as orange solids (60%, 288 mg). **^{1}H-NMR** (399.8 MHz, CDCl$_3$) δ 10.2 (s, 1 H), 7.72 – 7.68 (m, 6 H), 6.80 (d, *J* = 8.8 Hz, 2 H), 3.76 (t, *J* = 5.6 Hz, 4 H), 4.82 (t, *J* = 5.2, 2 H), 3.60 – 3.51 (m, 4 H), 2.10 (s, 3 H) ppm. **^{13}C-NMR** (100.5 MHz, DMSO) δ 169.0, 151.0, 148.4, 142.8, 141.1, 125.0, 123.0, 119.6, 111.7, 58.6, 53.7, 24.6 ppm. **HRMS-ESI** (*m/z*): (M+H)$^+$ calc. for C$_{18}$H$_{22}$N$_4$O$_3$: 343.1765, observed: 343.1763; 0.47 ppm.

Synthesis of **1c**
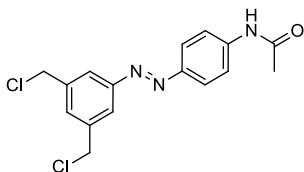
To 100 mg (0.34 mmol) of a solution of compound **2h** in dichloromethane, we added 0.2 mL of acetic anhydride and stirred the solution for 6 h at r.t. The solvent and excess acetic anhydride were then removed by a rotary evaporator and the compound **1f** was obtained as orange solids (125 mg, 100%). **^{1}H-NMR** (400.0 MHz, CDOD3/CDCl$_3$: 3/1) 7.88 (d, *J* = 8.8 Hz, 2 H), 7.79 (s, 2 H), 7.74 (d, *J* = 8.8 Hz, 2 H), 7.48 (s, 1 H), 4.71 (s, 4 H), 2.17 (s, 3 H) ppm. **^{13}C-NMR** (100.6 MHz, CDOD3/CDCl$_3$: 3/1) δ 171.7, 154.2, 149.9, 144.1, 142.7, 128.4, 124.6, 120.9, 120.8, 64.7, 24.0 ppm. **HRMS-ESI** (*m/z*): (M+H)$^+$ calc. for: C$_{16}$H$_{17}$Cl$_2$N$_4$O$_2$: 367.0723, observed: 367.0723; -0.22 ppm.

Synthesis of **1dA**
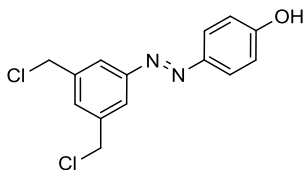
To an ice cold solution of **2e** (190 mg, 1 mmol) in THF, we added 2 mL of 2 M HCl. Then we added 1 mL of a 1 M solution of NaNO$_2$ in water dropwise. After stirring on ice for 1 h we slowly added this mixture to an ice cold solution of phenol (113 mg, 1 mmol) in saturated NaHCO$_3$. The reaction mixture was stirred at this temperature for 30 min and then was warmed up to r.t., then stirred at r.t. for another 2 h. Then the solvent was removed on a rotary evaporator and the resulting mixture was purified over silica by column chromatography (eluted in hexane/EtOAc = 9/1) to yield **1gA** as orange crystals (94 mg, 32%). **^{1}H-NMR** (399.8 MHz, CDCl$_3$) δ 7.90 (dd, J = 8.8 Hz, J = 2 Hz, 2 H), 7.87 (d, J = 0.8 Hz, 2 H), 7.52 (d, J = 0.8 Hz, 1 H), 6.96 (dd, J = 8.8 Hz, J = 2 Hz, 2 H), 4.69 (s, 4 H) ppm. **^{13}C-NMR** (399.8 MHz, CDCl$_3$) δ 158.6, 153.1, 147.0, 139.1, 130.1, 125.2, 122.6, 116.0, 45.4 ppm. **HRMS-ESI** (*m/z*): (M+H)$^+$ calc. for C$_{14}$H$_{13}$N$_2$Cl$_2$O: 295.0399, observed: 295.0399; 0.46 ppm.

Synthesis of **1d**
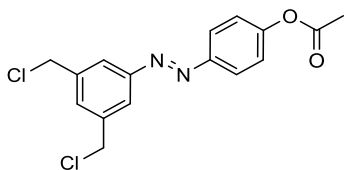
To a solution of **1dA** (60 mg, 0.2 mmol) in dichloromethane, we added acetyl chloride (212 µL, 0.3 mmol) followed 100 µL of pyridine. The reaction was then warmed to r.t. and stirred for 1 h at this temperature. After the completion of the reaction, the solvent was evaporated on a rotary evaporator and the resulting solid was purified on silica by column chromatography. The product eluted in hexane/EtOAc = 9/1 and solvent was removed on a rotary evaporator to yield **1d** as orange crystals (62 mg, 91%). **¹H-NMR** (499.8 MHz, CDCl$_3$) δ 7.98 (d, J = 8.8, 2 H), 7.92 (s, 2 H), 7.57 (s, 1 H), 7.29, (d, *J* = 8.8. Hz, 2 H), 4.70 (s, 4 H), 2.37 (s, 3H) ppm. **¹³C-NMR** (125.7 MHz, CDCl$_3$) δ 169.0, 153.0, 153.0, 150.0, 139.2, 130.8, 124.3, 122.9, 122.3, 45.4, 21.2 ppm. **HRMS-ESI** (*m/z*): (M+H)$^+$ calc. for: C$_{16}$H$_{15}$Cl$_2$N$_2$O$_2$: 337.0505, observed: 337.0502,-0.75 ppm.

Synthesis of 3,5-bis(hydroxymethyl) aniline (**2b**)
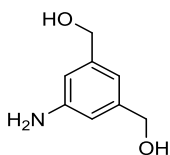
We synthesized 3,5-bis(hydroxymethyl) aniline using a previously published method.[198] To an ice cold solution of diethyl-5-(amino)isophtalate (**2a**) (5.23 g, 25 mmol, TCI Chemicals) in THF (300 mL) we slowly added an ice cold suspension of LiAlH$_4$ (2.77g, 75 mmol) in THF (200mL) with vigorous stirring. After stirring for 30 min at this temperature, the solution was warmed up to r.t. and was stirred for an additional 2 h. Upon the completion of the reaction, the excess LiAlH$_4$ was quenched by slowly adding 60 mL ethyl acetate and stirring for 1 h, followed

143

by 20 mL of methanol. Then we added 50 mL of a solution of saturated NH4Cl to hydrolyze the aluminum salt. The resulting mixtures was filtered off and the filtrate was concentrated on a rotary evaporator. The resulting oil was further purified by column chromatography on silica. The product eluted in DCM/MeOH = 3/1 to give compound **2b** in as a pale yellow solid (2.65 g, 69%). **¹H-NMR** (400.0 MHz, CDOD₃): δ 6.67 (s, 1 H) 6.61 (s, 2 H), 4.46 (s, 4 H) ppm. **¹³C-NMR** (100.6 MHz, CDOD₃): δ 147.3, 142.3, 115.2, 113.0, 63.5 ppm. **HRMS-ESI** (*m/z*): calc. for $C_8H_{12}NO_2$ (M+H)⁺: 154.0868; found: 154.0861; 0.89 ppm

Synthesis of **2c**
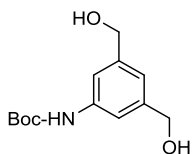
To a solution of 3,5-bis(hydroxymethyl) aniline (compound **2b**, 7 g, 45.5 mmol) in 200 mL DMF, we added di-*tert*-butyl dicarbonate (10.9 g, 50 mmol) with vigorous stirring. After stirring for 2 h, the solvent was removed on a rotary evaporator and the crude solid was purified over silica by column chromatography. The product was eluted in DCM/ MeOH = 85/15 to give the **2c** as white solid (10.3 g, 94%). **¹H-NMR** (399.7 MHz, CDOD₃): δ 7.31 (s, 2 H), 6.99 (s, 1 H), 4.54 (s, 4 H), 1.50 (s, 9 H). **¹³C-NMR** (100.5 MHz, CDOD₃): δ 153.9, 142.1, 139.2, 119.3, 115.9, 79.4, 63.7, 27.3 **HRMS-ESI** (*m/z*): calc. for $C_{13}H_{19}NO_4$ (M+H)⁺: 271.1652; found: 241.1649; 0.33 ppm.

Synthesis of **2d**
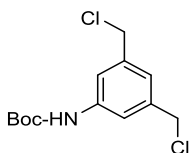
To an ice cold suspension of compound **2c** (7g, 27.6 mmol) in 200 mL THF, we slowly added methanesulfonyl chloride (5.3 mL, 69 mmol), followed by triethylamine (7.6 mL, 55 mmol). The mixture was stirred at 0 ºC for 30 min followed by stirring at room temperature for an hour. To the resulting solution, we slowly added a solution of LiCl (2.8 g, 69 mmol) in 100 mL DMF and the reaction mixture was stirred for 2 h at r.t. After completion of the reaction, the solvent was evaporated and the resulting mixture was dissolved in 200 mL of a 1:1 mixture of DCM and 0.1 M aqueous solution of $NaHCO_3$ in water. The organic phase was separated and the aqueous phase was extracted twice by 100 mL of DCM. The organic phases were combined and the solvent was removed on a rotary evaporator. The resulting crude mixture was purified over silica by column chromatography. The product eluted in Hexane/EtOAc = 4/1 to yield **2d** as a fluffy powder (7.6 g, 82%). **^{1}H-NMR** (399.7 MHz, CDCl$_3$): δ 7.39 (d, J = 1.2 Hz, 2 H), 7.09 (d, J = 1.2 Hz, 1 H), 6.65 (broad, 1 H), 4.53 (s, 4 H), 1.53 (s, 9 H) ppm. **^{13}C-NMR** (100.5 MHz, CD$_3$OD): δ 152.6, 139.2, 138.9, 123.0, 118.35, 81.1 ,45.7, 28.3 **HRMS-ESI** (*m/z*): ESI (m/z): calc. for $C_{13}H_{17}Cl_2NO_2$ Na (M+Na)$^+$: 312.0529; found: 312.0527; 0.65 ppm

Synthesis of **2e**
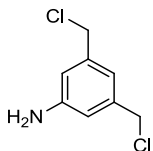
To 1.17 g (4.0 mmol) of compound **2d,** we added 6 mL of TFA at room temperature, with vigorous stirring. After 30 min, the TFA was removed on a rotary evaporator and the resulting solid was used with no further purification (0.71 g, quantitative). **¹H-NMR** (399.8 MHz, CDCl₃): δ 6.79 (s, 1 H), 6.69 (s, 1 H), 4.50 (s, 4 H) 3.69 (br, 2 H) **¹³C-NMR** (100.5 MHz, CDCl₃): δ 147.2, 139.1, 118.7, 115.0, 46.0 ppm. **HRMS-ESI** (*m/z*): (M+H)⁺ calc. for $C_8H_{10}Cl_2N$: 190.0185, observed: 190.0184; 0.3ppm

Synthesis of **2f**
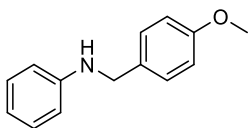
To a solution of 5.1 mL (55 mmol) of aniline in 100 mL methanol, we added 6.1 mL (55 mmol) of 4-methoxy benzaldehyde with vigorous stirring. After reflux-ing overnight, the reaction was cooled down to 0 ºC and NaBH₄ (1.9 g, 50 mmol) was added in small portions while stirring vigorously. We stirred the reaction for 30 min on ice and followed by 1 h at room temperature. Then, we evaporated the solvent and re-dissolved the solid in a mixture of DCM/0.1 M HCl = 1/1. Then separated the organic phase and removed the solvent on a rotary evaporator. The

crude product was purified by column chromatography over silica. The product was eluted in hexane/EtOAc = 85/15 to yield compound **2f** (10.4 g, 88% over two steps) as white solid. **^{1}H-NMR** (399.8 MHz, CDCl$_3$): δ 7.34 (dd, J = 6.4 Hz, J = 2 Hz, 2 H), 7.23 (2 H) 6.93 (dd, J=6.4 Hz, J=2 Hz, 2 H), 6.77 (1 H), 6.68 (2 H) 4.30 (s, 2 H), 3.99 (br, 1 H), 3.85 (s, 3 H) ppm. **^{13}C-NMR** (100.5 MHz, CDCl$_3$): δ 158.9, 148.3, 138.5, 129.3, 128.9, 117.5, 114.0, 112.9, 55.3, 47.8 ppm. **HRMS-ESI** (*m/z*): (M+H)$^+$ calc. for C$_{14}$H$_{15}$NO: 214.1226, observed: 214.1223; 1.67 ppm.

Synthesis of **2g**
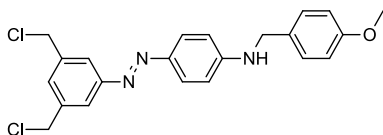
To an ice cold solution of 760 mg (4 mmol) of compound **2e** in 4 mL methanol, we added, 4 mL of 2 M HCl and the solution was stirred on ice for 30 min. To this mixture we added a solution of 276 mg of NaNO$_2$ (4 mmol) in water drop wise with vigorous stirring for 1 h to form the diazo salt. Then we added the diazo salt solution drop wise to an ice cold solution of **2f** in 50 mL of THF and the pH was adjusted to 5 with saturated solution of NaHCO$_3$. After stirring for 1 h at this temperature, the solution was warmed up to r.t. and further stirred overnight. Then we removed the solvents on a rotary evaporator and dissolved the solids in a 100 mL mixture of DCM/brine = 1/1. The organic phase was separated and the aqueous layer was extracted twice with 50 mL of DCM. The organic layers were combined

and the solvent was removed on a rotary evaporator. The resulting solid was purified on silica by column chromatography and the product was eluted in hexane/EtOAc = 3/1 to yield the azobenzene **2g** as orange crystals (1.03 g, 62%). **^{1}H NMR** (399.9 MHz, CDCl$_3$) δ 7.85 (2 H), 7.83 (d, J = 2 Hz, 2 H), 7.45 (1 H), 7.29 (dd, J = 6.8 Hz, J = 2.4 Hz, 2 H), 6.91 (dd, J = 6.8 Hz, J = 2.4 Hz, 2 H), 6.69 (d, J = 8.8 Hz, 2 H), 4.65 (s, 4 H), 4.35 (s, 2H), 3.81 (s, 3 H) ppm. **^{13}C-NMR** (100.6 MHz, CDCl$_3$) δ 159.1, 153.5, 151.2, 144.6, 138.9, 130.2, 129.2, 128.8, 125.6, 122.3, 114.2, 112.5, 55.3, 47.3, 45.6 ppm. **HRMS-ESI** (*m/z*): (M+H)$^+$ calc. for C$_{22}$H$_{20}$Cl$_2$N$_3$O:412.0978, observed: 412.0982; 0.89 ppm.

Synthesis of **2h**
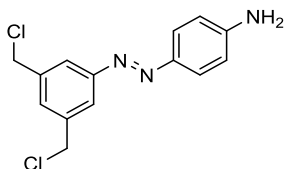
To 1.0 g mg (2.4 mmol) of compound **2g**, we added 500 μL of anisole, followed by 5 ml of TFA and the resulting mixture was stirred in dark for at least 2 h at r.t. After completion of the reaction, the solvent was removed on a rotary evaporator and the resulting solid was re-dissolved in MeOH and neutralized by saturated NaHCO$_3$ solution. The solvent was evaporated on a rotary evaporator, the solids were dissolved in DCM/brine = 1/1 and the organic phase was separated. We removed the DSM on a rotary evaporator and purified the crude mixture by column chromatography over silica. The product eluted in hexane/EtOAc = 65/75 to yield

compound **2h** as an orange solid (600 mg, 85 %). **¹H-NMR** (399.8 MHz, CDCl₃) δ 7.843 (s, 2 H), 7.83 (d, J = 8.5 Hz, 2H) 7.48 (s, 1 H), 6.76 (d, J = 8.5 Hz, 2H), 4.68 (s, 2 H) ppm. **¹³C-NMR** (100.5 MHz, CDCl₃) δ 153.5, 150.1, 145.4, 138.9, 129.5, 125.4, 122.4, 114.6, 45.5 ppm. **HRMS-ESI** (*m/z*): (M+H)⁺ calc. for C₁₄H₁₃Cl₂N₃:294.0559, observed: 294.0557; 0.87 ppm.

Synthesis of **2i**
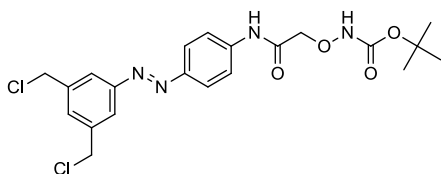
To a solution of 70 mg (0.24 mmol) of compound **2h** in 25 mL of THF, we added (BOC-aminooxy)acetic acid (92 mg, 0.48 mmol) followed by DIC (60 mg, 0.48 mmol) and DIPEA (81 µL, 0.48 mmol) with vigorous stirring. The reaction was stirred for 2 h and then was washed with 20 mL of water followed by 20 mL of brine. The organic phase was separated and the solvent was removed on a rotary evaporator. Resulting solid was purified over silica by column chromatography and the product eluted in hexane/EtOAc = 6/4 to yield compound **2i** as an orange solid (80 mg, 72%). **¹H-NMR** (399.8 MHz, CDCl₃) δ 10.6 (br, 1 H) 7.97-7.88 (m, 7 H), 7.5 (s, 1 H) 4.68 (s, 4 H), 4.49 (s, 2 H), 1.53 (s, 9 H) ppm. **¹³C-NMR** (100.5 MHz, CDCl₃) δ 167.5, 158.5, 153.2, 148.8, 141.0, 139.0, 130.4, 124.1, 122.7, 129.9, 83.9, 77.1, 45.4, 28.1 ppm. **HRMS-ESI** (*m/z*): (M+H)⁺ calc. for C₂₁H₂₄Cl₂N₄O₄: 467.1247, observed: 467.1244; 0.5 ppm.

Synthesis of **2j (HADCAz)**
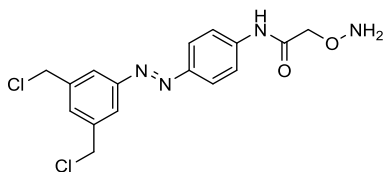
To 150 mg of the compound **2i**, we added 2 mL of TFA and stirred the mixture for 15 min. The TFA was then removed on a rotary evaporator and the resulting compound was purified on silica by column chromatography with no further workup. The product eluted in hexane/EtOAc = 7/3 to yield compound **2j** as orange solid (108 mg, 92%). **^{1}H-NMR** (399.8 MHz, CDCl$_3$) δ 7.94 (dt, J = 8.8 Hz, J = 2 Hz, 2 H), 7.90 (d, J = 1.6 Hz, 2 H), 7.83 (dt, J = 8.8 Hz, J = 2 Hz, 2 H), 7.61 (t, J = 1.6 Hz, 1 H), 4.74 (s, 4 H), 4.68 (s, 2 H) ppm. **^{13}C-NMR** (100.6 MHz, CDOD$_3$) δ 168.6, 154.4, 150.3, 142.2, 141.2, 132.0, 125.0, 123.6, 121.3 73.1, 46.0 ppm. **HRMS-ESI** (*m/z*): (M+H)$^+$ calc. for C$_{16}$H$_{17}$Cl$_2$N$_4$O$_2$: 367.0723, observed: 367.0723, -0.22 ppm.

4.4.2. General Procedure for Synthesis of Peptides Using FMOC Chemistry

We synthesized the peptides using Rink Amide AM resin (Chempep) and standard solid phase amide coupling protocol. Briefly, we transferred 250 mg (0.24 mmol, 0.96 mmol/g) into a PolyPrep® chromatography column. We set up the column on a manifold vacuum and equipped it with a three-way stopcock that allows draining of the solvent by vacuum filtration. Agitation of the resin was done by purging nitrogen gas as described by Verdine and co-workers.[199] We added 3 mL

of DCM to the dry resin and let it swell for 15 min. We drained the DCM and re-suspended the resin in 3 mL of DMF for 15 min. We drained the DMF and then deprotected the amine with 20% (v/v) piperidine in DMF (4 mL) for 1 min. We repeated the de-protection for another 9 min using fresh 20% (v/v) piperidine in DMF (4 mL). We then washed the resin with DMF (4 x 4 mL). We dissolved Fmoc-protected amino acid (0.96 mmol, 4 eq.) in 3 mL of DMF (4 mL) and then added HBTU (363 mg, 0.96 mmol, 4 eq) to the solution until dissolved. We added this mixture to resin followed by 0.3 mL of DIPEA (1.92 mmol, 8 eq). After 30 minutes of mixing, we removed the reagents by vacuum filtration and washed the resin with DMF (4 x 4 mL). We repeated the Fmoc-deprotection, amide coupling, and wash-ing steps to elongate the sequence up to the *N*-terminal residue. After Fmoc de-protection, we washed the resin with DMF (5 x 4 mL), followed by DCM (5 x 4 mL). We left the resin on the manifold for 10 min to dry on vacuum. We added the cleavage cocktail (4 mL, TFA/H$_2$O/TIPS/Phenol/EDT, 90/2.5/5/5/2.5 (v/v/v/w/v)) to the dried resin and rocked the column for 2 h to cleave the peptide. We collected the flow through from the column washed the resin was with TFA (1 mL). We added combined cleavage mixture drop-wise to 40 mL of ice cold diethyl ether in a 50 mL polypropylene centrifuge tube (Falcon, Thermo Fisher). We incubated the mixture on ice for 30 min. Then we separated the precipitates by centrifugation (5 min, 3000 rpm). We decanted the supernatant and washed the precipitates with cold diethyl ether (2 × 40 mL). After air drying the peptide, we dissolved it in 40 mL of 50% aqueous MeCN solution and adjusted the pH to 8 by aqueous ammonia. We then added 500 µL of DMSO to this solution and stirred the mixture for 24 h at

room temperature in the presence of air to form the disulfide bonds. The solution was then concentrated to 5 mL by speedvac and was purified by HPLC. Addition of 10% acetic acid was often necessary to dissolve the peptide, before purification.

For HPLC purification, we injected the peptide solution into a semi preparative RP-HPLC system equipped with C18 column (Waters Symmetryprep 19 × 50 mm). A gradient of solvent A (MQ water, 0.1% (v/v) TFA) and solvent B (MeCN, 0.1% (v/v) TFA) was run at a flow rate of 8 mL/min (0-2 min: 2% B; 2-18 min: 2→50% B). We used same method for all the purifications with HPLC, unless otherwise noted. The fractions corresponding to the main peak were collected. We removed the acetonitrile by evaporation using a speedvac. Lyophilizing the aqueous fraction yielded the peptide as white powder. Identity and purity of the peptides, as confirmed by LCMS (Figure 4.11). MALDI-MS data confirmed the presence of monocyclic disulfides. No dimers or trimers were found.
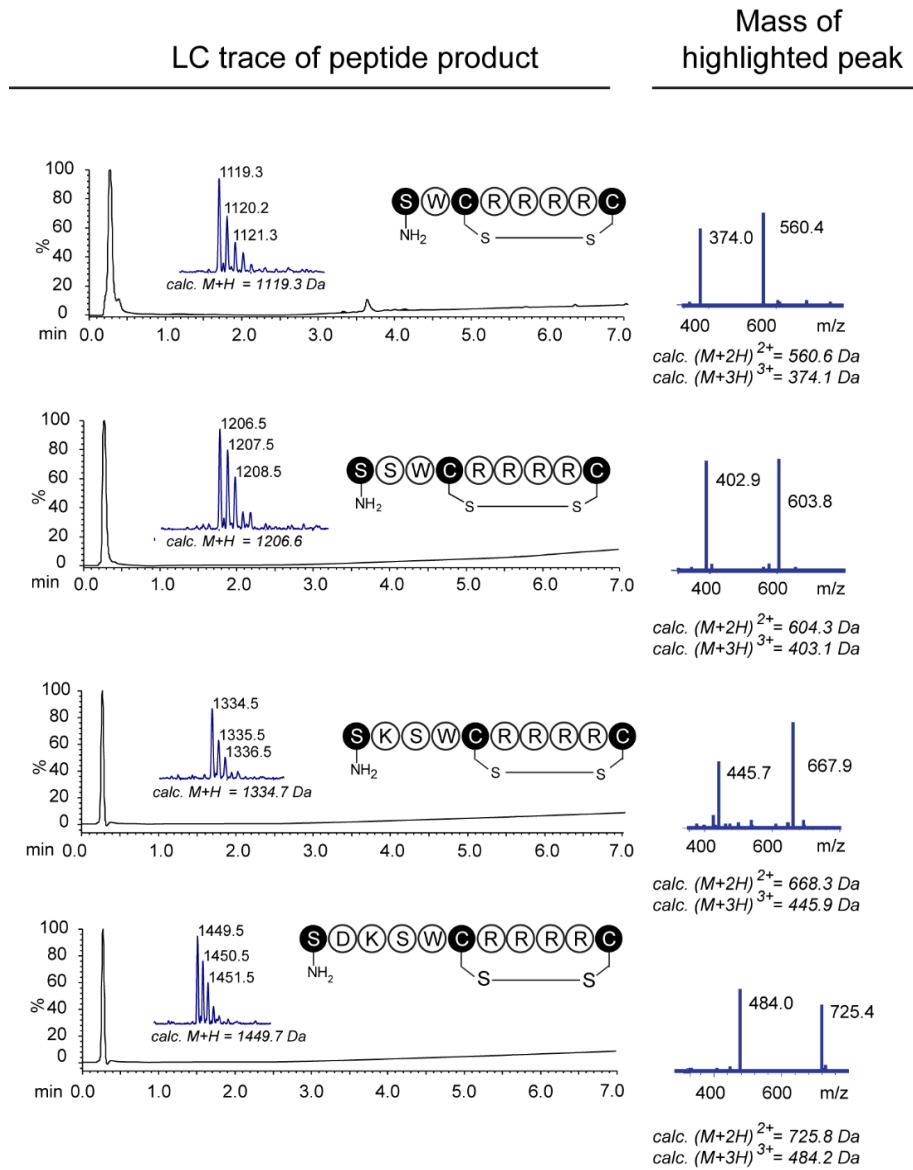
Figure 4.11. LCMS traces of disulfide peptides. MALDI–MS results shows isotope pattern.
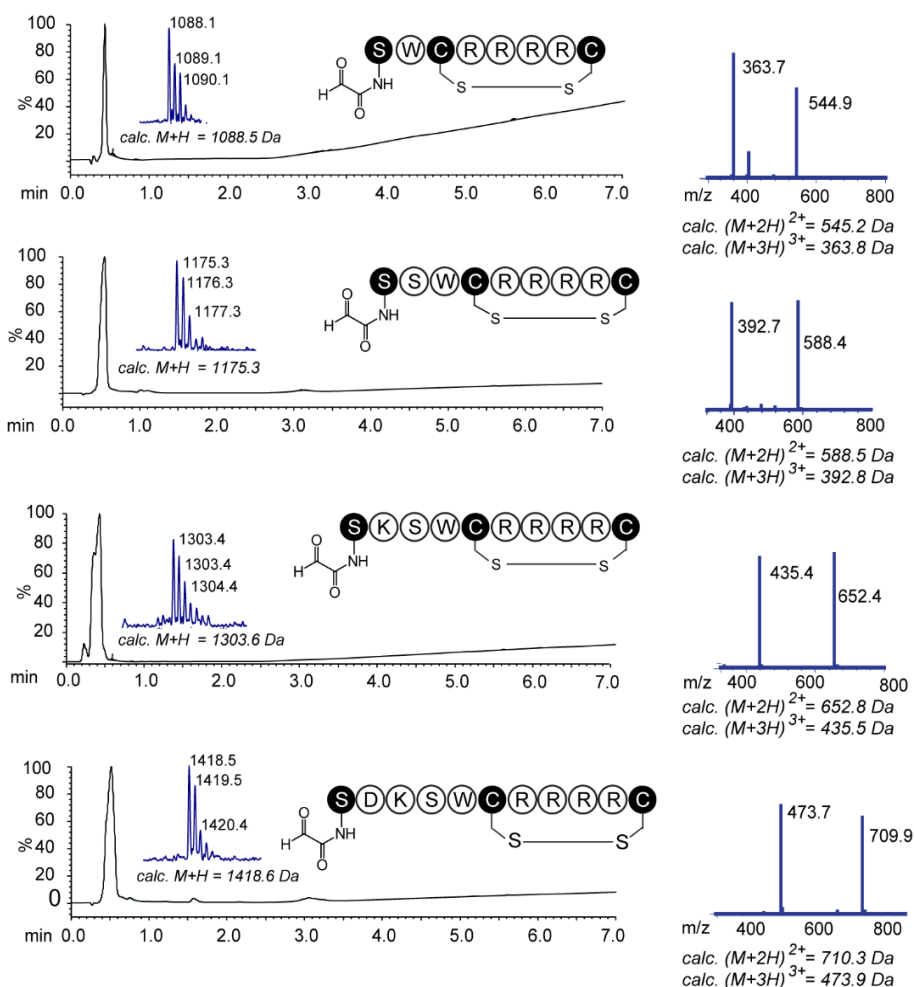
Figure 4.12. LCMS traces of the peptides after oxidation of the *N*-terminal serine with sodium periodate. MALDI–MS results shows isotope pattern

4.4.3. General Procedure for Periodate Oxidation of Serine in Peptide Sequence SX_nCRRRRC

To an ice cold solution of 10 mg peptide in 10 mL of 100 mM phosphate buffer (pH 7.4) we added a solution of NaIO$_4$ (2 eq) in one portion. The reaction mixture was incubated on ice for 10 min and at r.t. for an additional 10 min and

then was purified with HPLC on C18 silica. In some cases, we observed precipitation. In this case, the solution was acidified to pH 5 with 10% AcOH to dissolve the peptide and then was immediately injected to HPLC (Figure 4.12).

4.4.4. General Procedure for Synthesis of Bicyclic Peptide

To a solution of 5 mg of oxidized peptide in 500 µL DMF, we added 100 µL of a solution of HADCAz (1.2 eq) in DMF. We incubated the reaction for 2 h on a 55 °C bath and then diluted it to 2.5 mL with DMF. We further diluted the mixture by adding 2.5 mL of H_2O and then added 1.2 equivalent of TCEP solution (25 mM, pH 8.0) in two portions in 15 min intervals. The pH of the mixture was carefully adjusted to 8 (as determined by pH paper) and the reaction mixture was incubated at r.t. for 45 min. We then acidified the reaction to pH 5 with 10% AcOH solution and purified the product by HPLC. The identity of the peptides was confirmed by HRMS and/or NMR (Figures 4.13, Appendix 5).

4.4.5. Thermal Relaxation Kinetics and *cis*-Ratio of HADCAz and Bicyclic Peptides at Photostationary State

To quantify the *cis*-ratio of azobenzene at photostationary state after irradiation, we injected 5 µL of a 100 µM solution of the Azb containing compound to HPLC. We repeated the injections 0, 11, 22, 60, 120, 240 and 720 min after irradiation by 365 nm light. The absorption of the peaks was measured at 420 nm (isosbestic point).
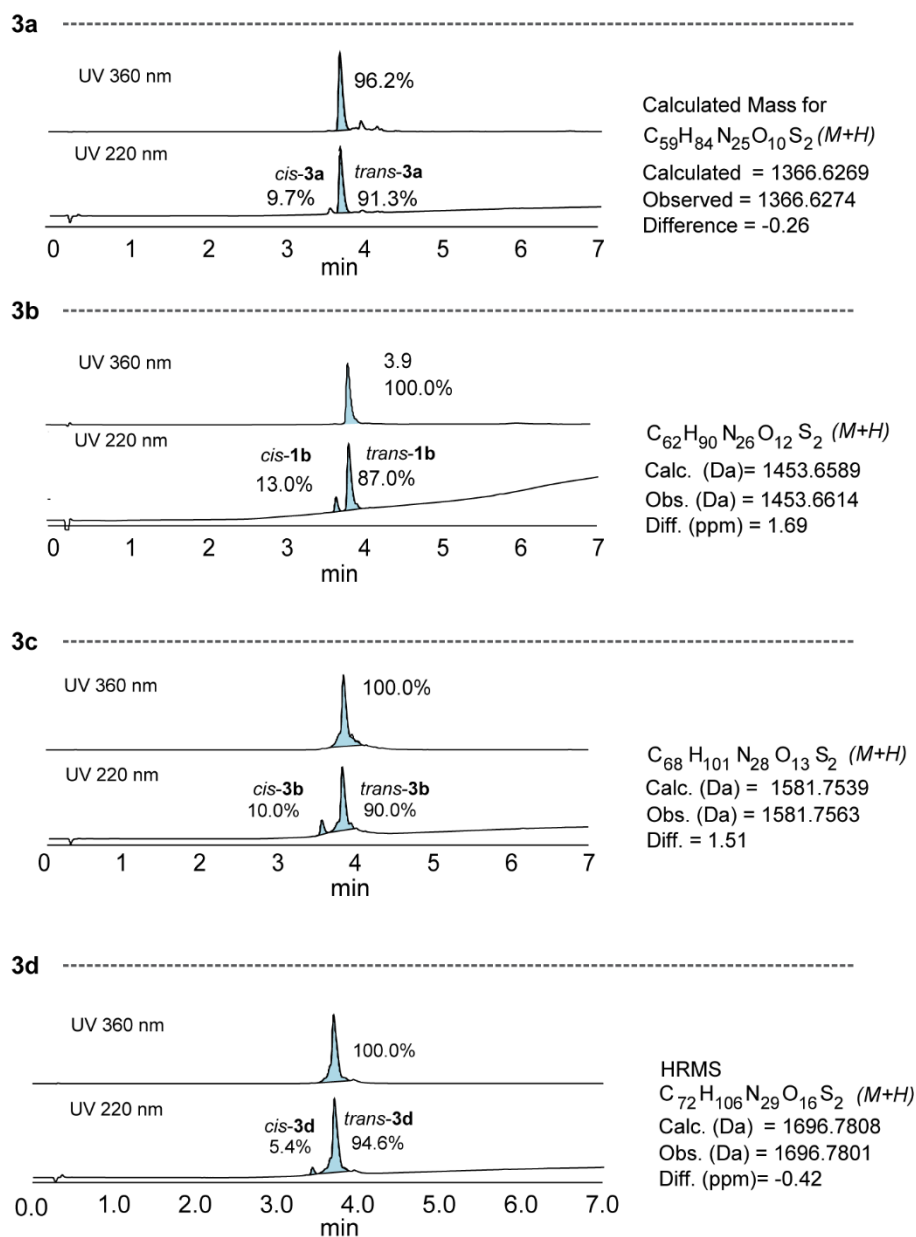
Figure 4.13. LCMS traces of purified peptide **3a-3d** and HRMS characterization. The *cis*-isomer is visible when measured at 220 nm, but is not detectable at 360 nm chromatogram, because of low absorption of **Azb** at 360 nm for *cis* form.

Chapter 5 Conclusion

5.1 Summary of the Thesis

Light-responsive ligands have been used for the study of molecular pathways in the cells, because their function can be controlled in high temporal and spatial resolution. Photo-pharmacology is a promising field in therapeutics represented by development of azobenzene (Azb) conjugated LR-analogs of clinically used drugs such as quinolones,[200] propofol[201] and colchicine.[202] However, design of LR-ligands is a complex procedure, because it needs available information about the structure of the receptor, the SAR of the ligand and the mechanism of binding. Devising a photo-switch in a known ligand without perturbing its binding properties is not trivial. Screening of LR-libraries can be a suitable alternative to conventional rational design, because it surpasses the demand for available structural information of receptor and/or iterative multi-step synthesis of ligands to find the best analog with no significant compromise of the binding.

In chapter 2 of this thesis, I pioneered the discovery of LR-ligands through screening of genetically encoded libraries.[28] I provided a method for chemical modification of a commercially available phage display library by the LR-linchpin BSBCA, which results in a genetically encoded library of LR-peptides. The quality of the library can be measured by quantification of the yield of the chemical modification on the peptides displayed on phage coat protein. I demonstrated that screening of such a library can result in identification of light-responsive macrocyclic peptides that reversibly decrease their binding properties in response to blue light.

The reversibility of the change in binding is important, because it provides the possibility of iteration in activation/inactivation downstream events. In chapter 3, I described the synthesis of BSBDA, an alleneamide substituted azobenzene linchpin, based on BSBCA, which reacts with K up to 30 $M^{-1}s^{-1}$ with thiols to produce LR-macrocyclic peptides. The high reactivity of BSBDA makes it a good candidate to replace the BSBCA in future studies, especially in chemical modification of phage displayed libraries, where limited exposure to reagent is essential to preserve the function of the target structure. In chapter 4, I described the synthesis of a LR-linchpin that can react with peptides containing natural amino acids to produce bicyclic LR-scaffolds. Such bicyclic ligands contain a static surface and a dynamic LR-moiety, each of which can exhibit a specific function such as binding, cell permeation, hydrophobicity, etc. Simple one-pot reaction on peptides makes this method suitable for further application on proteins; although certain limitations, such as non-aqueous reaction conditions, should be overcome prior to adapting the method for generation of bicyclic LR-phage display libraries.

5.2 Future Directions

LR-ligands can be used to investigate the effect of activation of receptors in the asymmetric division of cancer stem cells (CSCs). CSCs are a small population of cancer cells that have the potential to self-renew or to differentiate into non-tumorogenic cells.[203] Failure in asymmetric division could result in renewal of cancer stem cells instead of differentiation.[204] The cell can divide asymmetrically, because a gradient of cell-fate determinants is present in the cytoplasm.[205] However,

the study of such division—by inducing or inhibiting the effect of cell-fate determining material—is difficult, because it requires activation or inhibition of proteins in specific locations of the cell. By locally controlling the function of specific proteins in the cell surface using LR-ligands, we can investigate the mechanisms of asymmetric control of the cells.

Epidermal growth factor receptor (EFGR) is a protein where activation has been related to increase in population of CSCs.[206] This transmembrane tyrosine kinase protein produces an active homo-dimer after binding of its native ligand, epidermal growth factor (EGF), to a binding pocket in the extracellular domain.[207] To locally inhibit the activation of EGFR, two classes of LR-inhibitors can be envisioned: 1) inhibitors that block the binding of EGF to EGFR in the extracellular domain of the cell; 2) inhibitors that block the function of EGFR by inhibiting the dimerization of intracellular domain of the receptor protein. LR-inhibitors of EGFR can inhibit the binding of the native ligand EGF, until pulses of light are applied to a specific part of the cell—which results in a decrease in binding of the inhibitor. The LR-dissociation of the inhibitor can result in local activation of the receptor on the cell in the presence of EGF, and subsequently a potentially differentiated daughter cell after division.

The former class of the LR-inhibitors can be identified by screening a LR-library of peptides displayed on phage (Figure 5.1) against the extracellular domain of the EGFR immobilized on solid phase, such as magnetic beads. In a similar strategy, the panning procedure can be performed using cells that highly express the EGFR. As described in chapter 2, a single selection can result in identification of
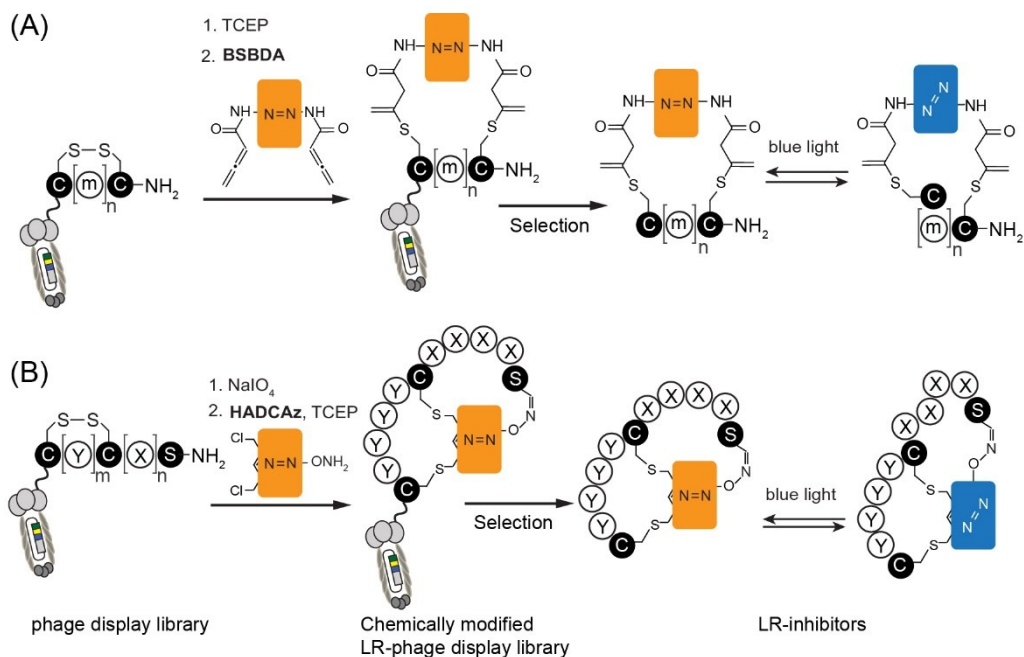
Figure 5.1. LR-monocyclic (A) and bicyclic (B) inhibitors can be identified by screening of chemically modified phage display libraries.

multiple LR-inhibitors with different properties such as binding affinity and relaxation half-life. In the latter class, as described in detail in chapter 4, we will synthesize a bicyclic cell-permeable[31] LR-peptide that can inhibit the coiled-coil formation of the EGFR intra-cellular domain (Figure 5.2).[197] The static surface of the bicyclic peptide can lead it through the cell membrane and the dynamic part can inhibit the EGFR dimerization, until the pulses of light are applied. Measurement of the extent of stabilization of the α-helix by circular dichroism in the resulting bicyclic LR-ligand can provide a preliminary vision of the success in inhibition, because the peptide inhibitor should have a helical structure for maximum inhibition effect. The LR-inhibition/activation can be tested by various methods. We will use the
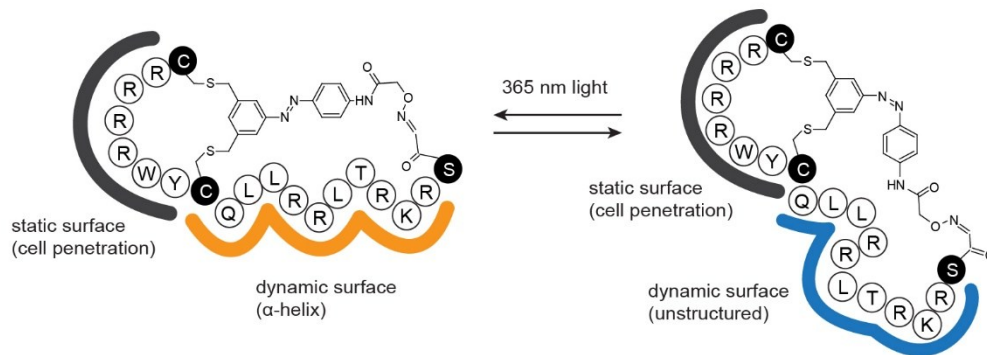
Figure 5.2. A bicyclic LR-inhibitor for intracellular inhibition of EGFR. Cell permeable static surface can lead the peptide through the cell-membrane, while the dynamic moiety changes the binding affinity to EGFR in response to blue light.

A549 lung cancer cells which express a GFP tagged variant of EGFR. LR-activation of the inhibited EGFR can be observed by fluorescent microscopy, because such activation results in internalization of the fluorescent protein into the cell.

References

(1)   Peptide Therapeutics Market - Global Industry Analysis, S., Share, Growth, Trends and Forecast 2014 - 2020 In *Transparency Market Research*; www.transparencymarketresearch.com/pressrelease/peptide-therapeutics-market.htm, last acccessed: 12/2015.

(2)   The-Pharmaceutical-Industry-in-Figures. *European Federation of Pharmaceutical Industries and Associations* **2015**.

(3)   Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. *Chemical Biology & Drug Design* **2013**, *81*, 136.

(4)   Sato, A. K.; Viswanathan, M.; Kent, R. B.; Wood, C. R. *Curr. Opin. Biotechnol.* **2006**, *17*, 638.

(5)   Kaspar, A. A.; Reichert, J. M. *Drug Discovery Today* **2013**, *18*, 807.

(6)   Fosgerau, K.; Hoffmann, T. *Drug Discovery Today* **2015**, *20*, 122.

(7)   McGregor, D. P. *Curr. Opin. Pharmacol.* **2008**, *8*, 616.

(8)   Carton, J. M.; Strohl, W. R. In *Introduction to Biological and Small Molecule Drug Research and Development*; Ganellin, R., Jefferis, S. R., Eds.; Elsevier: Oxford, 2013, p 127.

(9)   Antosova, Z.; Mackova, M.; Kral, V.; Macek, T. *Trends Biotechnol.* **2009**, *27*, 628.

(10)   Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatisky, M. *Drug Discovery Today* **2010**, *15*, 40.

(11)   Smart, A. L.; Gaisford, S.; Basit, A. W. *Expert Opin. Drug Deliv.* **2014**, *11*, 1323.

(12)   Uhlig, T.; Kyprianou, T.; Martinelli, F. G.; Oppici, C. A.; Heiligers, D.; Hills, D.; Calvo, X. R.; Verhaert, P. *EuPA Open Proteoms* **2014**, *4*, 58.

(13)   Hackett, M. J.; Zaro, J. L.; Shen, W.-C.; Guley, P. C.; Cho, M. J. *Adv. Drug Del. Rev.* **2013**, *65*, 1331.

(14)   Kurtzhals, P. *Endocrinol. Metab. Clin. North Am.* **2007**, *36*, 14.

(15)   Elbrond, B.; Jakobsen, G.; Larsen, S.; Agerso, H.; Jensen, L. B.; Rolan, P.; Sturis, J.; Hatorp, V.; Zdravkovic, M. *Diabetes Care* **2002**, *25*, 1398.

(16)   White, C. J.; Yudin, A. K. *Nat Chem* **2011**, *3*, 509.

(17)   Liskamp, R. M. J.; Rijkers, D. T. S.; Bakker, S. E. In *Modern Supramolecular Chemistry*; Wiley-VCH Verlag GmbH & Co. KGaA: 2008, p 1.

(18)   Bock, J. E.; Gavenonis, J.; Kritzer, J. A. *ACS Chem. Biol.* **2013**, *8*, 488.

(19)   Rezai, T.; Bock, J. E.; Zhou, M. V.; Kalyanaraman, C.; Lokey, R. S.; Jacobson, M. P. *J. Am. Chem. Soc.* **2006**, *128*, 14073.

(20)   Yudin, A. K. *Chem. Sci.* **2015**, *6*, 30.

(21)   White, T. R.; Renzelman, C. M.; Rand, A. C.; Rezai, T.; McEwen, C. M.; Gelev, V. M.; Turner, R. A.; Linington, R. G.; Leung, S. S. F.; Kalgutkar, A. S.; Bauman, J. N.; Zhang, Y.; Liras, S.; Price, D. A.; Mathiowetz, A. M.; Jacobson, M. P.; Lokey, R. S. *Nat. Chem. Biol.* **2011**, *7*, 810.

(22)   http://circlepharma.com 2015.

(23)   Miller, S. J.; Blackwell, H. E.; Grubbs, R. H. *J. Am. Chem. Soc.* **1996**, *118*, 9606.

(24)   Walensky, L. D.; Kung, A. L.; Escher, I.; Malia, T. J.; Barbuto, S.; Wright, R. D.; Wagner, G.; Verdine, G. L.; Korsmeyer, S. J. *Science* **2004**, *305*, 1466.

(25)   Traboulsi, H.; Larkin, H.; Bonin, M.-A.; Volkov, L.; Lavoie, C. L.; Marsault, É. *Bioconjug. Chem.* **2015**, *26*, 405.

(26)   Qian, Z.; Liu, T.; Liu, Y.-Y.; Briesewitz, R.; Barrios, A. M.; Jhiang, S. M.; Pei, D. *ACS Chem. Biol.* **2013**, *8*, 423.

(27)   Lajoie, M. J.; Söll, D.; Church, G. M. *J. Mol. Biol.*

(28)   Jafari, M. R.; Deng, L.; Kitov, P. I.; Ng, S.; Matochko, W. L.; Tjhung, K. F.; Zeberoff, A.; Elias, A.; Klassen, J. S.; Derda, R. *ACS Chem. Biol.* **2014**, *9*, 443.

(29)   Zou, Y.; Spokoyny, A. M.; Zhang, C.; Simon, M. D.; Yu, H.; Lin, Y.-S.; Pentelute, B. L. *Org. Biomol. Chem.* **2014**, *12*, 566.

(30)   Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. *Nat. Chem. Biol.* **2009**, *5*, 502.

(31)   Lian, W.; Upadhyaya, P.; Rhodes, C. A.; Liu, Y.; Pei, D. *J. Am. Chem. Soc.* **2013**, *135*, 11990.

(32)   Chua, K.; Fung, E.; Micewicz, E. D.; Ganz, T.; Nemeth, E.; Ruchala, P. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 4961.

(33)   Bashiruddin, N. K.; Nagano, M.; Suga, H. *Bioorg. Chem.* **2015**, *61*, 45.

(34)   Timmerman, P.; Beld, J.; Puijk, W. C.; Meloen, R. H. *ChemBioChem* **2005**, *6*, 821.

(35)   Chu, Q.; Moellering, R. E.; Hilinski, G. J.; Kim, Y.-W.; Grossmann, T. N.; Yeh, J. T. H.; Verdine, G. L. *MedChemComm* **2015**, *6*, 111.

(36)   Spokoyny, A. M.; Zou, Y.; Ling, J. J.; Yu, H.; Lin, Y.-S.; Pentelute, B. L. *J. Am. Chem. Soc.* **2013**, *135*, 5946.

(37)   Renner, C.; Moroder, L. *ChemBioChem* **2006**, *7*, 868.

(38)   Kumita, J. R.; Smart, O. S.; Woolley, G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3803.

(39)   Jo, H.; Meinhardt, N.; Wu, Y.; Kulkarni, S.; Hu, X.; Low, K. E.; Davies, P. L.; DeGrado, W. F.; Greenbaum, D. C. *J. Am. Chem. Soc.* **2012**, *134*, 17704.

(40)   Ban, H.; Nagano, M.; Gavrilyuk, J.; Hakamata, W.; Inokuma, T.; Barbas, C. F. *Bioconjug. Chem.* **2013**, *24*, 520.

(41)   Geoghegan, K. F.; Stroh, J. G. *Bioconjug. Chem.* **1992**, *3*, 138.

(42)   Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernandez-Gonzalez, M.; Bernardes, G. J. L.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G. *Chem. Sci.* **2011**, *2*, 1666.

(43)   Levengood, M. R.; van der Donk, W. A. *Nat. Protocols* **2007**, *1*, 3001.

(44)   Dirksen, A.; Dawson, P. E. *Bioconjug. Chem.* **2008**, *19*, 2543.

(45)   Galonić, D. P.; van der Donk, W. A.; Gin, D. Y. *Chem. - Eur. J.* **2003**, *9*, 5997.

(46)   Lau, Y. H.; de Andrade, P.; Quah, S.-T.; Rossmann, M.; Laraia, L.; Skold, N.; Sum, T. J.; Rowling, P. J. E.; Joseph, T. L.; Verma, C.; Hyvonen, M.;

Itzhaki, L. S.; Venkitaraman, A. R.; Brown, C. J.; Lane, D. P.; Spring, D. R. *Chem. Sci.* **2014**, *5*, 1804.

(47)   Haney, C. M.; Horne, W. S. *J. Pept. Sci.* **2014**, *20*, 108.

(48)   Albrecht, M. In *Highlights in Bioorganic Chemistry*; Wiley-VCH Verlag GmbH & Co. KGaA: 2005, p 31.

(49)   Tegoni, M. *Eur. J. Inorg. Chem.* **2014**, *2014*, 2177.

(50)   Toniolo, C. *CRC Crit. Rev. Biochem.* **1980**, *9*, 1.

(51)   Fujimoto, K.; Kajino, M.; Inouye, M. *Chem. - Eur. J.* **2008**, *14*, 857.

(52)   Phelan, J. C.; Skelton, N. J.; Braisted, A. C.; McDowell, R. S. *J. Am. Chem. Soc.* **1997**, *119*, 455.

(53)   McGrath, N. A.; Andersen, K. A.; Davis, A. K. F.; Lomax, J. E.; Raines, R. T. *Chem. Sci.* **2015**, *6*, 752.

(54)   Lundblad, R. L. *Chemical Reagents for Protein Modification*; Fourth ed.; CRC Press, 2014.

(55)   Szewczuk, Z.; Rebholz, K. L.; Rich, D. H. *Int. J. Pept. Protein Res.* **1992**, *40*, 233.

(56)   Muppidi, A.; Wang, Z.; Li, X.; Chen, J.; Lin, Q. *Chem. Commun.* **2011**, *47*, 9396.

(57)   Gavenonis, J.; Sheneman, B. A.; Siegert, T. R.; Eshelman, M. R.; Kritzer, J. A. *Nat. Chem. Biol.* **2014**, *10*, 716.

(58)   Assem, N.; Ferreira, D. J.; Wolan, D. W.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2015**, 54, 8665-8668.

(59)   Chen, S.; Bertoldo, D.; Angelini, A.; Pojer, F.; Heinis, C. *Angew. Chem. Int. Ed.* **2014**, *53*, 1602.

(60)   Gold, A. M. *Biochemistry (Mosc.)* **1968**, *7*, 2106.

(61)   Godsell, J. A.; Stacey, M.; Tatlow, J. C. *Nature* **1956**, *178*, 199.

(62)   Vinogradova, E. V.; Zhang, C.; Spokoyny, A. M.; Pentelute, B. L.; Buchwald, S. L. *Nature* **2015**, *526*, 687.

(63)   Moore, J. E.; Ward, W. H. *J. Am. Chem. Soc.* **1956**, *78*, 2414.

(64)   Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. *Chem. - Asian J.* **2009**, *4*, 630.

(65)   Fontaine, S. D.; Reid, R.; Robinson, L.; Ashley, G. W.; Santi, D. V. *Bioconjug. Chem.* **2015**, *26*, 145.

(66)   Schierling, B.; Noël, A.-J.; Wende, W.; Hien, L. T.; Volkov, E.; Kubareva, E.; Oretskaya, T.; Kokkinidis, M.; Römpp, A.; Spengler, B.; Pingoud, A. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 1361.

(67)   T., H. G. *Bioconjugate Techniques* Third ed.; Academic Press, 2013.

(68)   Tornøe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67*, 3057.

(69)   Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. *Bioconjug. Chem.* **2010**, *21*, 1912.

(70)   Link, A. J.; Mock, M. L.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2003**, *14*, 603.

(71)   Torres, O.; Yüksel, D.; Bernardina, M.; Kumar, K.; Bong, D. *ChemBioChem* **2008**, *9*, 1701.

(72)   Hermanson, G. T. In *Bioconjugate Techniques (Third edition)*; Hermanson, G. T., Ed.; Academic Press: Boston, 2013, p 1.

(73)   Morrison, P. M.; Foley, P. J.; Warriner, S. L.; Webb, M. E. *Chem. Commun.* **2015**, *51*, 13470.

(74)   Haney, C. M.; Horne, W. S. *Org. Biomol. Chem.* **2015**, *13*, 4183.

(75)   Whitcombe, M. J.; Alexander, C.; Vulfson, E. N. *Synlett* **2000**, 911.

(76)   Smith, J. M.; Frost, J. R.; Fasan, R. *Chem. Commun.* **2014**, *50*, 5027.

(77)   Beharry, A. A.; Woolley, G. A. *Chem. Soc. Rev.* **2011**, *40*, 4422.

(78)   Fehrentz, T.; Schönberger, M.; Trauner, D. *Angew. Chem. Int. Ed.* **2011**, *50*, 12156.

(79)   Velema, W. A.; Szymanski, W.; Feringa, B. L. *J. Am. Chem. Soc.* **2014**, *136*, 2178.

(80)   Merino, E.; Ribagorda, M. *Beilstein J. Org. Chem.* **2012**, *8*, 1071.

(81)   Fukaminato, T.; Hirose, T.; Doi, T.; Hazama, M.; Matsuda, K.; Irie, M. *J. Am. Chem. Soc.* **2014**, *136*, 17145.

(82)   Babii, O.; Afonin, S.; Berditsch, M.; Reißer, S.; Mykhailiuk, P. K.; Kubyshkin, V. S.; Steinbrecher, T.; Ulrich, A. S.; Komarov, I. V. *Angew. Chem. Int. Ed.* **2014**, *53*, 3392.

(83)   Warford, C. C.; Lemieux, V.; Branda, N. R. In *Molecular Switches, Second Edition* 2011; Vol. 1, p 1.

(84)   Chen, L.; Wu, J.; Schmuck, C.; Tian, H. *Chem. Commun.* **2014**, *50*, 6443.

(85)   Klajn, R. *Chem. Soc. Rev.* **2014**, *43*, 148.

(86)   Hoppmann, C.; Maslennikov, I.; Choe, S.; Wang, L. *J. Am. Chem. Soc.* **2015**, *137*, 11218.

(87)   Kuil, J.; van Wandelen, L. T. M.; de Mol, N. J.; Liskamp, R. M. J. *J. Pept. Sci.* **2009**, *15*, 685.

(88)   Zhang, F.; Sadovski, O.; Woolley, G. A. *ChemBioChem* **2008**, *9*, 2147.

(89)   Zhang, Z.; Burns, D. C.; Kumita, J. R.; Smart, O. S.; Woolley, G. A. *Bioconjug. Chem.* **2003**, *14*, 824.

(90)   Beharry, A. A.; Chen, T.; Al-Abdul-Wahid, M. S.; Samanta, S.; Davidov, K.; Sadovski, O.; Ali, A. M.; Chen, S. B.; Prosser, R. S.; Chan, H. S.; Woolley, G. A. *Biochemistry (Mosc.)* **2012**, *51*, 6421.

(91)   Beharry, A. A.; Sadovski, O.; Woolley, G. A. *Beilstein J. Org. Chem.* **2008**, *6*, 4323.

(92)   Blanco-Lomas, M.; Samanta, S.; Campos, P. J.; Woolley, G. A.; Sampedro, D. *J. Am. Chem. Soc.* **2012**, *134*, 6960.

(93)   Guerrero, L.; Smart, O. S.; Woolley, G. A.; Allemann, R. K. *J. Am. Chem. Soc.* **2005**, *127*, 15624.

(94)   Bellotto, S.; Chen, S.; Rentero Rebollo, I.; Wegner, H. A.; Heinis, C. *J. Am. Chem. Soc.* **2014**, *136*, 5880.

(95)   Buchli, B.; Waldauer, S. A.; Walser, R.; Donten, M. L.; Pfister, R.; Blöchliger, N.; Steiner, S.; Caflisch, A.; Zerbe, O.; Hamm, P. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11725.

(96)   Nevola, L.; Martín-Quirós, A.; Eckelt, K.; Camarero, N.; Tosi, S.; Llobet, A.; Giralt, E.; Gorostiza, P. *Angew. Chem.* **2013**, *125*, 7858.

(97)   Mart, R. J.; Wysoczański, P.; Kneissl, S.; Ricci, A.; Brancale, A.; Allemann, R. K. *ChemBioChem* **2012**, *13*, 515.

(98)   Burns, D. C.; Zhang, F.; Woolley, G. A. *Nat. Protocols* **2007**, *2*, 251.

(99)   Chi, L.; Sadovski, O.; Woolley, G. A. *Bioconjug. Chem.* **2006**, *17*, 670.

(100)   Pozhidaeva, N.; Cormier, M.-E.; Chaudhari, A.; Woolley, G. A. *Bioconjug. Chem.* **2004**, *15*, 1297.

(101)   Dong, M.; Babalhavaeji, A.; Samanta, S.; Beharry, A. A.; Woolley, G. A. *Acc. Chem. Res.* **2015**, *48*, 2662.

(102)   Dong, M.; Babalhavaeji, A.; Hansen, M. J.; Kalman, L.; Woolley, G. A. *Chem. Commun.* **2015**, *51*, 12981.

(103)   Samanta, S.; Beharry, A. A.; Sadovski, O.; McCormick, T. M.; Babalhavaeji, A.; Tropepe, V.; Woolley, G. A. *J. Am. Chem. Soc.* **2013**, *135*, 9777.

(104)   Deisseroth, K. *Nat. Met.* **2011**, *8*, 26.

(105)   Miesenböck, G. *Science* **2009**, *326*, 395.

(106)   Mayer, G.; Hechel, A. *Angew. Chem. Int. Ed.* **2006**, *45*, 4900.

(107)   Fehrentz, T.; Schönberger, M.; Trauner, D. *Angew. Chem., Int. Ed.* **2011**, *50,* 12156.

(108)   Erdmann, F.; Zhang, Y. *Mol. Biosyst.* **2010**, *6*, 2103.

(109)   Volgraf, M.; Gorostiza, P.; Numano, R.; Kramer, R. H.; Isacoff, E. Y.; Trauner, D. *Nat. Chem. Biol.* **2006**, *2*, 47.

(110)   Levitz, J.; Pantoja, C.; Gaub, B.; Janovjak, H.; Reiner, A.; Hoagland, A.; Schoppik, D.; Kane, B.; Stawski, P.; Schier, A. F.; Trauner, D.; Isacoff, E. Y. *Nat. Neurosci.* **2013**, *16*, 507.

(111)   Tsai, H.-C.; Zhang, F.; Adamantidis, A.; Stuber, G. D.; Bonci, A.; de Lecea, L.; Deisseroth, K. *Science* **2009**, *324*, 1080.

(112)   Lima, S. Q.; Miesenböck, G. *Cell* **2005**, *121*, 141.

(113)   Wyart, C.; Bene, F. D.; Warp, E.; Scott, E. K.; Trauner, D.; Baier, H.; Isacoff, E. Y. *Nature* **2009**, *461*, 407.

(114)   Dwyer, M. A.; Lu, W.; Dwyer, J. J.; Kossiakoff, A. A. *Chem. Biol.* **2000**, *7*, 263.

(115)   Kehoe, J. W.; Kay, B. K. *Chem. Rev.* **2005**, *105*, 4056.

(116)   Clackson, T.; Hoogenboom, H. R.; Griffiths, A. D.; Winter, G. *Nature* **1991**, *352*, 624.

(117)   Scott, J. K.; Smith, G. P. *Science* **1990**, *249*, 386.

(118)   Barbas Iii, C. F.; Lerner, R. A. *Methods* **1991**, *2*, 119.

(119)   Angelini, A.; Heinis, C. *Curr. Opin. Chem. Biol.* **2011**, *15*, 355.

(120)   Ng, S.; Jafari, M. R.; Derda, R. *ACS. Chem. Biol.* **2011**, *7*, 123.

(121)   Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. *Nat. Chem. Biol.* **2009**, *5*, 502.

(122)   Angelini, A.; Cendron, L.; Chen, S.; Touati, J.; Winter, G.; Zanotti, G.; Heinis, C. *ACS. Chem. Biol.* **2012**, *7*, 817.

(123)   Ng, S.; Jafari, M. R.; Matochko, W. L.; Derda, R. *ACS. Chem. Biol.* **2012**, *7*, 1482.

(124)   Woiwode, T. F.; Haggerty, J. E.; Katz, R.; Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Cwirla, S. E. *Chem. Biol.* **2003**, *10*, 847.

(125)   Scott, C. P.; Abel-Santos, E.; Jones, A. D.; Benkovic, S. J. *Chem. Biol.* **2001**, *8*, 801.

(126)   Arai, K.; Tsutsumi, H.; Mihara, H. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 4940.

(127)   Santoso, B.; Lam, S.; Murray, B. W.; Chen, G. *Bioorg. Med. Chem. Lett.*

(128)   Meyer, S. C.; Shomin, C. D.; Gaj, T.; Ghosh, I. *J. Am. Chem. Soc.* **2007**, *129*, 13812.

(129)   Roxin, Á.; Zheng, G. *Future Med. Chem.* **2012**, *4*, 1601.

(130)   Schlippe, Y. V. G.; Hartman, M. C. T.; Josephson, K.; Szostak, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 10469.

(131)   Li, S. W.; Roberts, R. W. *Chem. Biol.* **2003**, *10*, 233.

(132)   Millward, S. W.; Takahashi, T. T.; Roberts, R. W. *J. Am. Chem. Soc.* **2005**, *127*, 14142.

(133)   Sako, Y.; Morimoto, J.; Murakami, H.; Suga, H. *J. Am. Chem. Soc.* **2008**, *130*, 7232.

(134)   Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R. *Science* **2004**, *305*, 1601.

(135)   Guillen Schlippe, Y. V.; Hartman, M. C. T.; Josephson, K.; Szostak, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 10469.

(136)   Hofmann, F. T.; Szostak, J. W.; Seebeck, F. P. *J. Am. Chem. Soc.* **2012**, *134*, 8038.

(137)   Beharry, A. A.; Woolley, G. A. *Chem. Soc. Rev.* **2011**, *40*, 4422.

(138)   Sandman, K. E.; Banner, J. S.; Noren, C. J. *J. Am. Chem. Soc.* **2000**, *122*, 960.

(139)   Tian, F.; Tsao, M. L.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126*, 15962.

(140)   Tochitsky, I.; Banghart, M. R.; Mourot, A.; Yao, J. Z.; Gaub, B.; Kramer, R. H.; Trauner, D. *Nat. Chem.* **2012**, *4*, 105.

(141)   Andersson, J.; Li, S.; Lincoln, P.; Andréasson, J. *J. Am. Chem. Soc.* **2008**, *130*, 11836.

(142)   Hoppmann, C.; Seedorff, S.; Richter, A.; Fabian, H.; Schmieder, P.; Rück-Braun, K.; Beyermann, M. *Angew. Chem. Int. Ed.* **2009**, *48*, 6636.

(143)   Guerrero, L.; Smart, O. S.; Woolley, G. A.; Allemann, R. K. *J. Am. Chem. Soc.* **2005**, *127*, 15624.

(144)   Hayashi, G.; Hagihara, M.; Nakatani, K. *Chem.- Eur. J.* **2009**, *15*, 424.

(145)   Kather, I.; Bippes, C. A.; Schmid, F. X. *J. Mol. Biol.* **2005**, *354*, 666.

(146)   Derda, R.; Tang, S. K. Y.; Whitesides, G. M. *Angew. Chem. Int. Ed.* **2010**, *49*, 5301.

(147)   Derda, R.; Tang, S. K. Y.; Li, S. C.; Ng, S.; Matochko, W.; Jafari, M. R. *Molecules* **2011**, *16*, 1776.

(148)   El-Hawiet, A.; Kitova, E. N.; Kitov, P. I.; Eugenio, L.; Ng, K. K.; Mulvey, G. L.; Dingle, T. C.; Szpacenko, A.; Armstrong, G. D.; Klassen, J. S. *Glycobiology* **2011**, *21*, 1217.

(149)   Deng, L.; Broom, A.; Kitova, E. N.; Richards, M. R.; Zheng, R. B.; Shoemaker, G. K.; Meiering, E. M.; Klassen, J. S. *J. Am. Chem. Soc.* **2012**, *134*, 16586.

(150)   Kitova, E.; El-Hawiet, A.; Schnier, P.; Klassen, J. *J. Am. Soc. Mass. Spectrom.* **2012**, *23*, 431.

(151)   Bléger, D.; Schwarz, J.; Brouwer, A. M.; Hecht, S. *J. Am. Chem. Soc.* **2012**, *134*, 20597.

(152)   Siewertsen, R.; Neumann, H.; Buchheim-Stehn, B.; Herges, R.; Näther, C.; Renth, F.; Temps, F. *J. Am. Chem. Soc.* **2009**, *131*, 15594.

(153)   Asanuma, H.; Liang, X.; Nishioka, H.; Matsunaga, D.; Liu, M.; Komiyama, M. *Nat. Protoc.* **2007**, *2*, 203.

(154)   Volgraf, M.; Gorostiza, P.; Szobota, S.; Helix, M. R.; Isacoff, E. Y.; Trauner, D. *J. Am. Chem. Soc.* **2006**, *129*, 260.

(155)   Lien, L.; Jaikaran, D. C. J.; Zhang, Z.; Woolley, G. A. *J. Am. Chem. Soc.* **1996**, *118*, 12222.

(156)   Stein, M.; Middendorp, S. J.; Carta, V.; Pejo, E.; Raines, D. E.; Forman, S. A.; Sigel, E.; Trauner, D. *Angew. Chem. Int. Ed.* **2012**, *51*, 10500.

(157)   Velema, W. A.; van der Toorn, M.; Szymanski, W.; Feringa, B. L. *J. Med. Chem.* **2013**, *56*, 4456.

(158)   Estévez-Torres, A.; Crozatier, C.; Diguet, A.; Hara, T.; Saito, H.; Yoshikawa, K.; Baigl, D. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12219.

(159)   Le Ny, A.-L. M.; Lee, C. T. *J. Am. Chem. Soc.* **2006**, *128*, 6400.

(160)   Gotoh, H.; Matsumoto, Y. *Gene* **2007**, *389*, 146.

(161)   Matochko, W. L.; Ng, S.; Jafari, M. R.; Romaniuk, J.; Tang, S. K. Y.; Derda, R. *Methods* **2012**, *58*, 18.

(162)   Deng, L.; Broom, A.; Kitova, E. N.; Richards, M. R.; Zheng, R. B.; Shoemaker, G. K.; Meiering, E. M.; Klassen, J. S. *J. Am. Chem. Soc.* **2012**, *134*, 16586.

(163)   Duchstein, P.; Neiss, C.; Görling, A.; Zahn, D. *J. Mol. Model.* **2012**, *18*, 2479.

(164)   Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. *PCCP* **2010**, *12*, 7821.

(165)   Pierce, L. C. T.; Salomon-Ferrer, R.; Augusto F. de Oliveira, C.; McCammon, J. A.; Walker, R. C. *J. Chem. Theory. Comput.* **2012**, *8*, 2997.

(166)   Tochitsky, I.; Banghart, M. R.; Mourot, A.; Yao, J. Z.; Gaub, B.; Kramer, R. H.; Trauner, D. *Nat Chem* **2012**, *4*, 105.

(167)   Van der Berg, J. P.; Velema, W. A.; Szymanski, W.; Driessen, A. J. M.; Feringa, B. L. *Chem. Sci.* **2015**, *6*, 3593.

(168)   Westmark, P. R.; Kelly, J. P.; Smith, B. D. *J. Am. Chem. Soc.* **1993**, *115*, 3416.

(169)   Ferreira, R.; Nilsson, J. R.; Solano, C.; Andréasson, J.; Grøtli, M. *Sci. Rep.* **2015**, *5*, 9769.

(170)   Szymański, W.; Beierle, J. M.; Kistemaker, H. A. V.; Velema, W. A.; Feringa, B. L. *Chem. Rev.* **2013**, *113*, 6114.

(171)   Deiters, A. *ChemBioChem* **2010**, *11*, 47.

(172)   Mayer, G.; Hechel, A. *Angew. Chem., Int. Ed.* **2006**, *45*, 4900.

(173)   Jafari, M. R.; Deng, L.; Kitov, P. I.; Ng, S.; Matochko, W. L.; Tjhung, K. F.; Zeberoff, A.; Elias, A.; Klassen, J. S.; Derda, R. *ACS Chem. Biol.* **2013**, *9*, 443.

(174)   Nevola, L.; Martín-Quirós, A.; Eckelt, K.; Camarero, N.; Tosi, S.; Llobet, A.; Giralt, E.; Gorostiza, P. *Angew. Chem. Int. Ed.* **2013**, *52*, 7704.

(175)   Abbas, A.; Xing, B.; Loh, T.-P. *Angew. Chem. Int. Ed.* **2014**, *53*, 7491.

(176)   Assem, N.; Ferreira, D. J.; Wolan, D. W.; Dawson, P. E. *Angew. Chem.* **2015**, *127*, 8789.

(177)   Burns, D. C.; Zhang, F.; Woolley, G. A. *Nat. Protoc.* **2007**, *2*, 251.

(178)   Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. *Chem. - Asian J.* **2009**, *4*, 630.

(179)   Lau, Y. H.; Wu, Y.; de Andrade, P.; Galloway, W. R. J. D.; Spring, D. R. *Nat. Protocols* **2015**, *10*, 585.

(180)   Turner, R. A.; Oliver, A. G.; Lokey, R. S. *Org. Lett.* **2007**, *9*, 5011.

(181)   Ingale, S.; Dawson, P. E. *Org. Lett.* **2011**, *13*, 2822.

(182)   Gerwin, B. I. *J. Biol. Chem.* **1967**, *242*, 451.

(183)   Ng, S.; Jafari, M. R.; Derda, R. *ACS Chem. Biol.* **2012**, *7*, 123.

(184)   Francis, M. B.; Bernard, J. M. L. *Front. Microbiol.* **2014**, *5*.

(185)   Lau, Y. H.; de Andrade, P.; Wu, Y.; Spring, D. R. *Chem. Soc. Rev.* **2015**, *44*, 91.

(186)   Shrimp, J. H.; Hu, J.; Dong, M.; Wang, B. S.; MacDonald, R.; Jiang, H.; Hao, Q.; Yen, A.; Lin, H. *J. Am. Chem. Soc.* **2014**, *136*, 5656.

(187)   Löfblom, J.; Feldwisch, J.; Tolmachev, V.; Carlsson, J.; Ståhl, S.; Frejd, F. Y. *FEBS Lett.* **2010**, *584*, 2670.

(188)   Villar, E. A.; Beglov, D.; Chennamadhavuni, S.; Porco Jr, J. A.; Kozakov, D.; Vajda, S.; Whitty, A. *Nat. Chem. Biol.* **2014**, *10*, 723.

(189)   Sewald, N.; Jakubke, H.-D. In *Peptides: Chemistry and Biology*; Wiley-VCH Verlag GmbH & Co. KGaA: 2009, p 63.

(190)   Newman, D. J.; Cragg, G. M. In *Macrocycles in Drug Discovery*; The Royal Society of Chemistry: 2015, p 1.

(191)   Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. *Nat. Rev. Drug Discov.* **2008**, *7*, 608.

(192)   Lian, W.; Jiang, B.; Qian, Z.; Pei, D. *J. Am. Chem. Soc.* **2014**, *136*, 9830.

(193)   Bellotto, S.; Reuter, R.; Heinis, C.; Wegner, H. A. *J. Org. Chem.* **2011**, *76*, 9826.

(194)   Kitov, P. I.; Vinals, D. F.; Ng, S.; Tjhung, K. F.; Derda, R. *J. Am. Chem. Soc.* **2014**, *136*, 8149.

(195)   Hermanson, G. T. In *Bioconjugate Techniques (Second Edition)*; Hermanson, G. T., Ed.; Academic Press: New York, 2008, p 396.

(196)   El-Mahdi, O.; Melnyk, O. *Bioconj. Chem.* **2013**, *24*, 735.

(197)   Sinclair, J. K. L.; Denton, E. V.; Schepartz, A. *J. Am. Chem. Soc.* **2014**, *136*, 11232.

(198)   Zheng, J.; Lin, S.; Jiang, B. W.; Marder, T. B.; Yang, Z. *Can. J. Chem.* **2012**, *90*, 138.

(199)   Kim, Y.-W.; Grossmann, T. N.; Verdine, G. L. *Nat. Protocols* **2011**, *6*, 761.

(200)   Velema, W. A.; van der Berg, J. P.; Hansen, M. J.; Szymanski, W.; Driessen, A. J. M.; Feringa, B. L. *Nat Chem* **2013**, *5*, 924.

(201)   Stein, M.; Middendorp, S. J.; Carta, V.; Pejo, E.; Raines, D. E.; Forman, S. A.; Sigel, E.; Trauner, D. *Angew. Chem. Int. Ed.* **2012**, *51*, 10500.

(202)   Borowiak, M.; Nahaboo, W.; Reynders, M.; Nekolla, K.; Jalinot, P.; Hasserodt, J.; Rehberg, M.; Delattre, M.; Zahler, S.; Vollmar, A.; Trauner, D.; Thorn-Seshold, O. *Cell* **2015**, *162*, 403.

(203)   Clarke, M. F.; Dick, J. E.; Dirks, P. B.; Eaves, C. J.; Jamieson, C. H. M.; Jones, D. L.; Visvader, J.; Weissman, I. L.; Wahl, G. M. *Cancer Res.* **2006**, *66*, 9339.

(204)   Reya, T.; Morrison, S. J.; Clarke, M. F.; Weissman, I. L. *Nature* **2001**, *414*, 105.

(205)   Morrison, S. J.; Kimble, J. *Nature* **2006**, *441*, 1068.

(206)   Ma, L.; Zhang, G.; Miao, X.-B.; Deng, X.-B.; Wu, Y.; Liu, Y.; Jin, Z.-R.; Li, X.-Q.; Liu, Q.-Z.; Sun, D.-X.; Testa, J. R.; Yao, K.-T.; Xiao, G.-H. *FEBS J.* **2013**, *280*, 2027.

(207)   Tomas, A.; Futter, C. E.; Eden, E. R. *Trends Cell Biol.* **2014**, *24*, 26.

Appendix 1

Investigating the origins of discrepancy between yield of alkylation of mono-clonal phage and a library of phage.

Introduction

Chemically modified phage libraries are indispensable tools for identification of the new lead peptide ligands for biologically important targets. We and others have screened such libraries for discovery of peptides ligands with augmented benefits such as high binding affinity[1] and light-responsiveness.[2] In the second chapter of this thesis we described a method to quantify the yield of chemical modification on phage libraries. In order to quantify the modification, we first reduced the peptide disulfide displayed on phage with TCEP and then alkylated it by BIA. Capturing the biotinylated phage with streptavidin (Str) coated magnetic beads, revealed the yield of reduction/alkylation of the thiols. We observed that while the biotinylation of the monoclonal phage was >90%, the yield of reaction for C7C library was consistently lower (~60%).[2] In this appendix, I describe preliminary studies of the origins of discrepancies in reaction of mono-clonal phage with phage library.

Results and discussion

Three hypothesis can justify the low yield of alkylation in libraries. The first hypothesis is that a certain population of the phage does not display the randomized sequence, because the genome of the phage does not contain the

specific genetic insert responsible for expression of the randomized peptide. Such low quality of library can be a result of poor transformation of the plasmids at the library production level. This theory can be validated by sequencing the unreacted population. In chapter 2, we showed that 40% of the sequenced phage did not contain the gene insert. This no-insert population is in accordance with the population of unreacted phage after each round of amplification. (Figure 2.6.)

 We sequenced the library of phage before and capture with Str beads using ion-torrent high throughput sequencing.[3,4] If the un-reactivity of the phage is due to the absence of randomized gene insert, Str capture should deplete the population of peptide displaying phage and enrich the population of no-insert phage.

The result of ion torrent sequencing did not show any no-insert phage (Table 1), because the design of the primers for amplification of phage DNA was not optimal. The DNA of the phage should be amplified by polymerase chain reaction (PCR) before sequencing. We used barcoded primers which made it possible to mix different samples of independent experiments and read the codes in a single run.[4] However, the reverse primer used for PCR only complemented the DNA of the phage with randomized insert (Figure A1.1); therefore the DNA of the no-insert phage did not amplify. We could not confirm the 40% no-insert population we observed among 80 Sanger sequenced clones.

The second hypothesis is that for a certain population of the phage, the disulfide displaying randomized sequence is lost due to hydrolysis or peptidase activity of the enzymes during phage excretion and purification. In this case, since the genome of the non-reacted phage carries the disulfide peptide, amplification
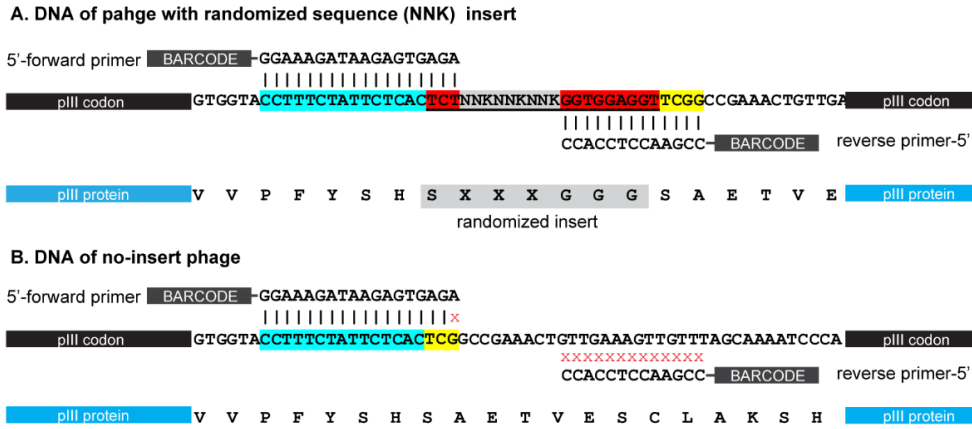
Figure A1.1 A) BARCODEd primers used for phage displaying randomized sequence complemented the sequences flanking the randomized insert. B) The reverse primer did not complement the no-insert gene sequence.

of each clone should yield a monoclonal population of the phage that is reactive towards the alkylating agent.

To investigate reactivity of unreacted phage after amplification, we alkylated the naïve phage library with BIA and captured the biotinylated phage as described in chapter 2. Then we infected the bacteria with the unreacted phage and incubated at 37 °C to grow single colonies. We picked 8 clones and re-amplified each for 4.5 h. After purification, we reduced the disulfides of each clones and reacted them with BIA. We observed that the clones reinstated reactivity with BIA after amplification (Figure A1.2) which indicated that the phage displayed the disulfide peptide. A WT-clone and a no-insert clone (clone 29) showed no significant biotinylation (< 20%) which showed that the alkylation is specific for disulfide displayed peptide. This observation shows that un-reactivity of the phage is not exclusively due to lack of randomized gene insert, but it can also be due to

loss of the reactive peptide sequence of the phage which makes it "temporarily un-reactive": Re-amplification of the same clone restores its reactivity.

Interestingly, the monoclonal phages showed different percentage of alkylation for each clone averaging to 75% ± 25% which is close to the alkylation yield of the library. Clone **R2** and **a5** were the most reactive (>90% alkylation) and clone **a7** was the least (58% alkylation). This observation highlights that the reactivity of a single clone should not be trasnlated as the reactivity of the whole library.
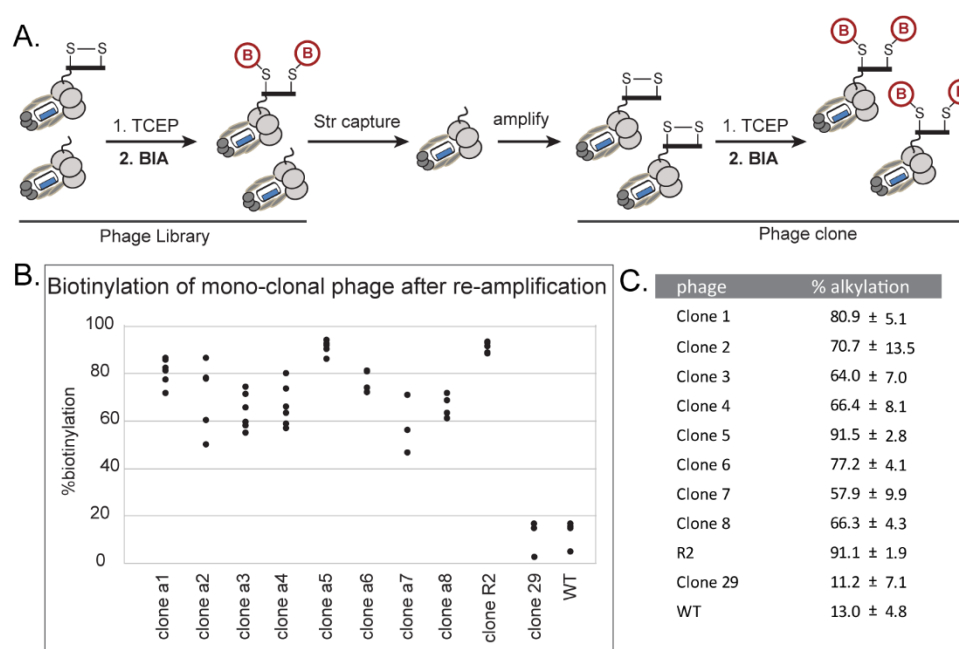
Figure A1.2. A) amplification of single clone in un-reactive phage library can yield monoclonal reactive phage. B, C) Yield of alkylation for 8 unreacted phage clone after re-amplification and three control clones.

The latter observation is suggestive of a third hypothesis: the reaction is sequence specific. While the reaction occurs fast for some sequences (e.g. CPARSPLEC, clone **R2**), it proceeds slower for other sequences. If that is true, elongation of the

174

reaction time should increase the yield to saturation for all clones; however, it also increases the chance of non-specific reaction of amines and other nucleophiles with the BIA.

Sequence specific reactions can be studied in detail by reacting the phage with two probes that use different chemistries: a slow reacting probe (biotin-PEG2-chloroacetamide, $k = 0.03$ $M^{-1}s^{-1}$) and a fast reacting probe (biotin-PEG2-allenamide $k = 30$ $M^{-1}s^{-1}$). The slow reacting probe should show a higher difference in reaction yield in a defined reaction time, while a fast reacting probe should diminish this effect.

Conclusion

We briefly studied the reason behind different reactivity of displayed disulfide peptides in monoclonal phage and phage library. The experiment resulted in mixed data and needs further studies, specifically to find: (i) the percentage of no-insert clones in naïve and non-captured libraries. (ii) sequences of individual reacting clones amplified out of non-reacted library. (iii) rate of alkylation of individual different clones of non-reacted library with other probes. We also suggest deep sequencing of the phage library by Illumina®[3] because of expanded range of sequence reads per run (up to 10^9 sequences).

Materials and Methods

Alkylation of the Phage Library and Clones by BIA

175

A 10 µL solution of 10^{11} pfu/mL of phage was reduced by TECP and alkylated by BIA according to the method described in chapter 2.

Amplification of Single Clones

To 25 mL of LB medium in a 125 mL baffled Erlenmayer flask, we added 200 µL of log phase *E. coli*. We picked a blue plaque of the phage grown on agar lawn with the tip of a 10 µL micro-pipette and transferred it to the bacterial mixtures. The flask was shaken on an orbital shaker for 4.5 h at 37 °C and then bacterial mixture was centrifuged at 4700 rpm for 15 min at 4 °C. The supernatant was carefully transferred into a 50 mL polycarbonate centrifuge tube and 12 mL of PEG/NaCl solution was added. We incubated this phage solution on ice for 2 h and then spun the tube at 14000 rpm for 20 min at 4 °C. We discarded the supernatant and the resulting pellet was re-dissolved in PBS. We transferred the solution into a 1.7 mL plastic tube and at 14000 rpm for 20 min at 4 °C. We collected the supernatant and repeated the centrifugation until no pellet was observed. The phage solution was incubated at -20 for storage.

Table 1. Result of 50 top sequence reads by ion torrent. The second row shows the coverage of the library by this 500 sequences based of the numbers of reads.

% coverage of first 200 sequence	97.4	97.3	97.9		97.9	98.8
sequence	counts	counts	counts	sequence	counts	counts
ACHYHSSKSC	3912	3761	1935	ACHYHSSKSC	1225	480
ACHYHSSKSC	3912	3761	1935	ACIKTSDHWC	1361	343

ACHYHSSKSC	3912	3761	1935	ACSKVHIPLC	959	291
ACIKTSDHWC	3512	3591	1703	ACRTSHNAYC	898	282
ACIKTSDHWC	3512	3591	1703	ACVTRMDMAC	927	190
ACIKTSDHWC	3512	3591	1703	ACTLGQLHSC	738	308
ACSKVHIPLC	3426	3152	1465	ACPPSLPWSC	846	168
ACSKVHIPLC	3426	3152	1465	ACSESPMRYC	607	191
ACSKVHIPLC	3426	3152	1465	ACSAENLHRC	530	184
ACRTSHNAYC	2914	3139	1530	ACDKGNYYMC	511	106
ACRTSHNAYC	2914	3139	1530	ACIPYLTSSC	389	201
ACRTSHNAYC	2914	3139	1530	ACLVSDRFTC	459	85
ACTLGQLHSC	2613	2541	1399	ACHYHSSKSC	3	480
ACTLGQLHSC	2613	2541	1399	ACHYHSSKSC	3	480
ACTLGQLHSC	2613	2541	1399	ACGYHNWMLC	285	130
ACTLGQLHSC	2613	2541	1399	ACPPQFRTPC	355	53
ACPPSLPWSC	2027	2142	983	ACALELANCC	316	91
ACPPSLPWSC	2027	2142	983	ACSMAPGLQC	270	107
ACDKGNYYMC	1689	1639	1046	ACIKTSDHWC	2	343
ACDKGNYYMC	1689	1639	1046	ACRNTEHIPC	276	50
ACDKGNYYMC	1689	1639	1046	ACGLYIWSIC	209	113
ACSESPMRYC	1720	1764	868	ACTLGQLHSC	1	308
ACSESPMRYC	1720	1764	868	ACSKVHIPLC	1	291
ACVTRMDMAC	1662	1729	846	ACSKVHIPLC	1	291
ACVTRMDMAC	1662	1729	846	ACRTSHNAYC	3	282
ACSAENLHRC	1551	1593	803	ACSSEGKVWC	223	51
ACSAENLHRC	1551	1593	803	ACTTMHFTAC	213	51
ACLVSDRFTC	1277	1300	567	ACELLHYLAC	182	77
ACLVSDRFTC	1277	1300	567	ACIPYLTSSC	1	201
ACIPYLTSSC	1183	1237	578	ACVTRMDMAC	2	190
ACIPYLTSSC	1183	1237	578	ACTQVSYMTC	137	53
ACGYHNWMLC	1116	1080	510	ACSAENLHRC	2	184
ACALELANCC	787	871	530	ACSAENLHRC	1	184
ACSMAPGLQC	882	888	335	ACPPSLPWSC	2	168
ACTTMHFTAC	816	833	369	ACLERTPYRC	116	47
ACTTMHFTAC	816	833	369	ACKHSNPSSC	102	43
ACELLHYLAC	647	529	328	ACLPLSLKSC	69	52
ACPPQFRTPC	525	508	343	ACPSTGNYRC	64	45
ACPPQFRTPC	525	508	343	ACLELRNSIC	71	34
ACLELRNSIC	536	571	251	ACHGHSRTNC	101	0
ACTQVSYMTC	401	510	355	ACALELANCC	1	91
ACLERTPYRC	440	546	252	ACLFNAKMQC	81	8
ACSSEGKVWC	550	315	293	ACYDLSCESC	66	23

ACSSEGKVWC	550	315	293	ACADHWASPC	52	36
ACLPLSLKSC	469	438	231	ACMPTQMVQC	82	1
ACPSTGNYRC	380	460	243	ACKMVSVFQC	81	0
ACRNTEHIPC	420	432	196	ACPSTPLLLC	61	19

References

(1)   Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. *Nat Chem Biol* **2009**, *5*, 502.

(2)   Jafari, M. R.; Deng, L.; Kitov, P. I.; Ng, S.; Matochko, W. L.; Tjhung, K. F.; Zeberoff, A.; Elias, A.; Klassen, J. S.; Derda, R. *ACS Chem. Biol.* **2014**, *9*, 443.

(3)   Matochko, W. L.; Chu, K.; Jin, B.; Lee, S. W.; Whitesides, G. M.; Derda, R. *Methods* **2012**, *58*, 47.

(4)   Ernst, A.; Gfeller, D.; Kan, Z.; Seshagiri, S.; Kim, P. M.; Bader, G. D.; Sidhu, S. S. *Mol. BioSyst.* **2010**, *6*, 1782.

Appendix 2

Serendipitous Discovery of Potential LR-Macrocyclic Protein A Binders.

In this this appendix, I describe the serendipitous discovery of Azb cyclized peptide binders that can mimic the Fc region of antibodies and therefore bind to *S*.aureus protein A (**pA**). The sequences of the peptide binders emerged from screening of a library of BSBCA cyclized peptides – displayed on the coat protein of phage – against anti-FLAG antibody immobilized by **pA**. We used a high throughput sequencing to find the sequences of the eluted phage enriched for the protein target. Synthesis of the 81 top sequences on a peptide array on cellulose revealed that none of the peptides bound to antiFLAG-AB, however, a large population of peptide (~30%) showed binding to a secondary antibody probe and **pA**. Herein, I describe the procedure that led to discovery of these peptides and the studies on the binding of such peptides to **pA**.

LR-ligands are important bio-molecular tools, because they can change binding affinity towards bio-targets in high temporal and spatial control.[1,2] We[3] and others[4] have already shown that LR-ligands can be discovered by screening of a genetically encoded LR-library. We envisioned LR-peptide ligands can be used in purification of antibodies (ABs), where light acts as an eluent. Conventional antibody purification uses **pA** – immobilized on stationary phase – as affinity ligand to

concentrate the AB on the affinity column, while impurities are washed off the column with washing solution. The column is then treated with acidic eluent ($\sim$ pH 4) which results in a decrease in affinity of the **pA** to AB (Figure A2.1A). However, the structure of AB is sensitive to pH and low pH can result in deactivation of AB. Affinity chromatography is also used for purification of other proteins containing a fusion such as FLAG, HisTag, Streptag, GST, etc. In this method of purification, the eluent contains competitive binders (FLAG peptide, imidazole, biotin, GSH, etc.) that can replace the tagged protein. The drawback of this method is that the eluted solution contains high concentration of eluent and requires dialysis after purification to remove the small molecules (Figure A2.1B).

LR-affinity ligands, change binding affinity in response to light and therefore, the eluting solution is only consisted of a buffer with suitable pH. This method can be highly important, especially in industrial scale production of therapeutic ABs, where the tolerance for presence of impurities in the final product is close to zero.

We started the screening by cyclizing a phage library of 7-mer disulfide peptides (Ph.D. C7C, NEB) with BSBCA according to the methods described in chapter 2. The resulting LR-library was subjected to four rounds of selection against anti-FLAG-AB that was immobilized on **pA** coated beads (Figure A2.2A). The non-bound phage was washed 10 times and the eluted phage was amplified. We repeated this selection cycle for four times, while modifying the amplified, binder enriched library before each selection cycle. The yield of BSBCA coupling varied between 33% to 60.5% after each modification (Figure A2.2B). We defined the "recovery" as the titer of the phage eluted after the last wash to the titer of the phage
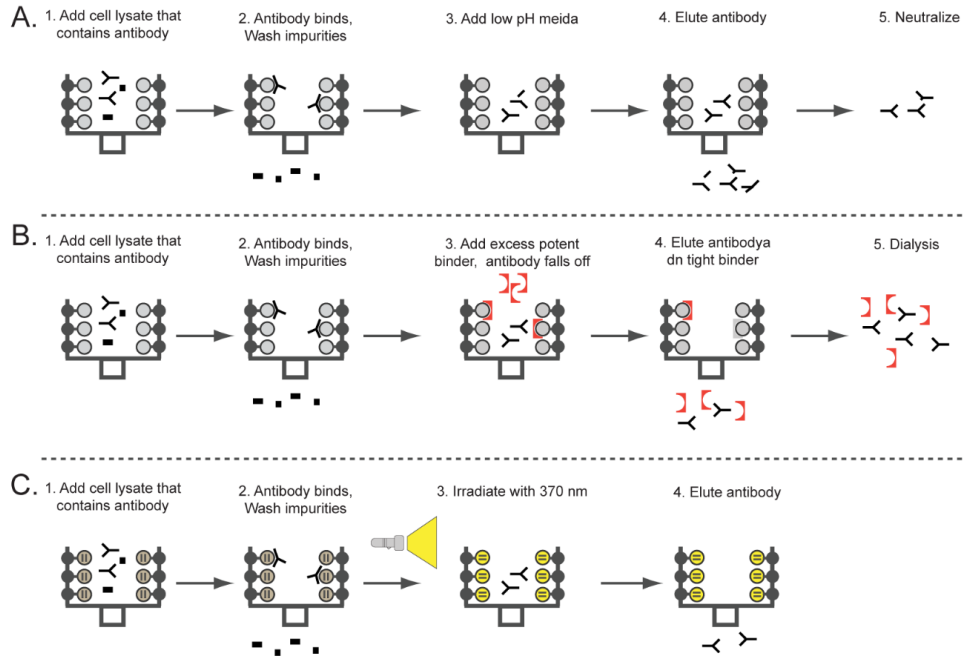
Figure A2.1. A) antibody purification using affinity tag. The AB is eluted in acidic conditions that can result in damage to AB structure and function. B) The AB is eluted by competitive binders and final antibody solution is contaminated with small molecules. C) AB purification with light. The final AB solution is pure AB in suitable buffer.

after elution. Interestingly, the recovery of the phage library increased from 2.2 and 2.4 in first and second round, respectively, to 16.0 in the third round of selection (Figure A2.2C) which indicated the enrichment of the binders. In a parallel control experiment, elution of a monoclonal FLAG displaying phage from antiFLAG-AB coated beads resulted in 55.5% recovery, but elution of same phage clone from **pA** coated beads showed a recovery of only 0.07 (Figure A2.2D). These results indicated that the increase in recovery occurs only if the specific binders are enriched during the screening. Since the fourth round of screening did not show enhanced
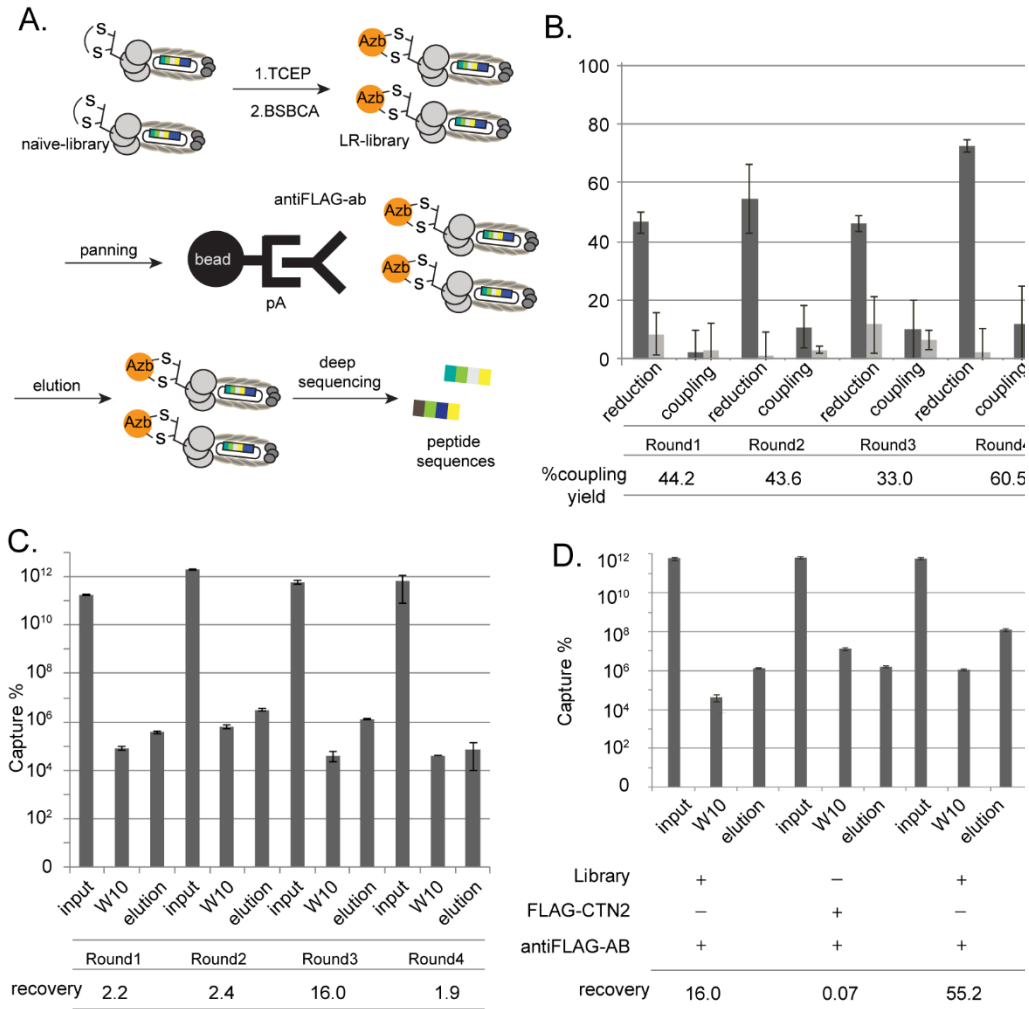
Figure A2.2. A) Scheme of screening procedure against antiFLAG-AB. B) Yield of modification of library with BSBCA before each round of screening. C) Recovery of the phage after each round of selection. D) recovery of the library vs FLAG-phage in the third round of panning.

recovery, we analyzed the sequence of the peptides by high throughput ion torrent deep sequencing.[5]
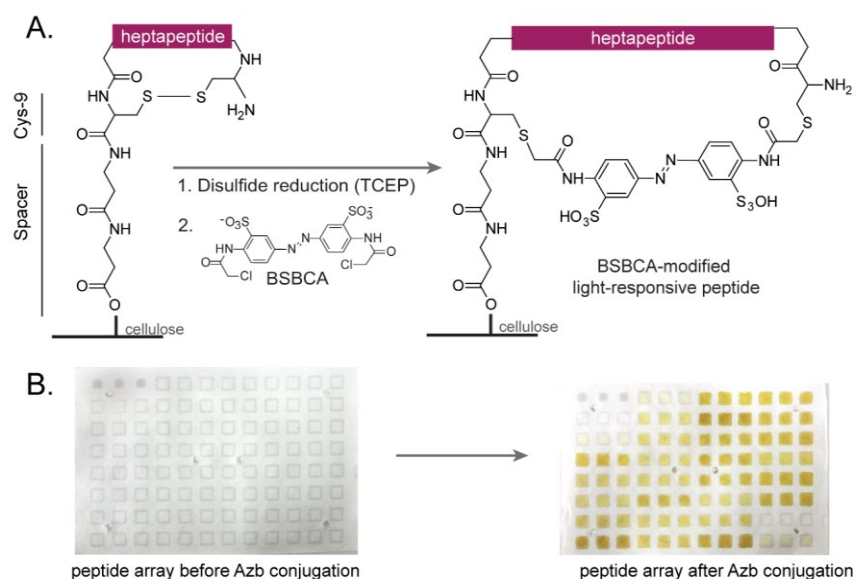
Figure A2.3. A) Cellulose supported Synthetic peptides were conjugated to Azb after disulfide reduction to yield LR-peptide array. B) peptide spots turn yellow which is the typical color of Azb conjugated peptides.

We selected the 81 top sequences from deep sequencing, because it shows the phage population that potentially enriched against the protein target. We synthesized them on a Teflon® barrier coated cellulose support (SPOT synthesis) according to a previously published method (Figure A2.3).[6] Two FLAG sequence (DYKDDDDK) were included in each array as positive controls. A negative control containing only βAla-βAla was also included. (Figure A2.4). The disulfide bond of other peptides was reduced and the peptides were cyclized BSBCA.

To test the binding of the Azb-cyclized peptide, we incubated the array with antiFLAG-AB and then probed the binding by antimouse/HRP conjugate AB. Ad

Array 1

	1-3		4-6		6-8		9-12
1	NONE	PAB6	CNTGSPYEC	PAB14	CTPGHTNRC	PAB22	CEFSKFRSC
2	BLANK	PAB7	CQSGQPALC	PAB15	CAIKNTKSC	PAB23	CPQTNRNMC
3	DYKDDDDKC	PAB8	CTNANHYFC	PAB16	CKDHVTRVC	PAB24	CMARYMSAC
4	PAB1 CAKATCPAC	PAB9	CSISSLTHC	PAB17	CSWNQMRGC	PAB25	CHNEGNRAC
5	PAB2 CIGNSNTLC	PAB10	CMAPDSRVC	PAB18	CAGHNRDRC	PAB26	CTMPLRGHC
6	PAB3 CTVRTSADC	PAB11	CGQSGARFC	PAB19	CMSTGLSSC	PAB27	CGYSSFNRC
7	PAB4 CNWGDRILC	PAB12	CLGSVGRDC	PAB20	CFNDIAKSC		BLANK
8	PAB5 CRGATPMSC	PAB13	CEGQRWMQC	PAB21	CRSANIYTC		CDYKDDDDK

Array 2

	1-3		4-6		6-8		9-12
1	DYKDDDDKC	PAB33	CLKNQSDQC	PAB41	CYSMKYGSC	PAB49	CWSNGQLMC
2	NONE	PAB34	CDRTTRHLC	PAB42	CKLTTQMMC	PAB50	CGGNSNQRC
3	BLANK	PAB35	CIGKSNTLC	PAB43	CQSGTQRAC	PAB51	CVQMPAHSC
4	PAB28 CSTEHTRSC	PAB36	CHPVSGQKC	PAB44	CLPGNPNKC	PAB52	CMNRQDHAC
5	PAB29 CESKYLSSC	PAB37	CSSVTDRWC	PAB45	CMMHNMRHC	PAB53	CVNVSKWQC
6	PAB30 CLNSSQPSC	PAB38	CTHSNRLVC	PAB46	CVNTTNGAC	PAB54	CAKATCLRC
7	PAB31 CLKLGEKWC	PAB39	CSVTNYSSC	PAB47	CWTSGRHAC		BLANK
8	PAB32 CTKQMVAAC	PAB40	CYSRTAPSC	PAB48	CQQTKNYYC		CDYKDDDDK

Array 3

	1-3		4-6		6-8		9-12
1	DYKDDDDKC	PAB60	CSNPLKLDC	PAB68	CPTVKQKWC	PAB76	CTMRSPTGC
2	BLANK	PAB61	CTLMDANHC	PAB69	CSDSYLPGC	PAB77	CNAKHHPRC
3	NONE	PAB62	CTLTMYNRC	PAB70	CSHMEYPRC	PAB78	CSTLKELSC
4	PAB55 CSTMNGKVC	PAB63	CTSYNRMAC	PAB71	CVLKNTSQC	PAB79	CNLATGWSC
5	PAB56 CFSNQHRTC	PAB64	CTKSLAHTC	PAB72	CFNMFSRVC	PAB80	CRTHQTGAC
6	PAB57 CSAKSSLKC	PAB65	CHQGLDTFC	PAB73	CQFTRGAVC	PAB81	CLSTTHMTC
7	PAB58 CAGQKPLCC	PAB66	CRTDAMRNC	PAB74	CGSKNSAVC		BLANK
8	PAB59 CNMQITKGC	PAB67	CNRSLEPWC	PAB75	CRAHSHTLC		CDYKDDDDK

Figure A2.4. List of the sequences and their position on the peptide array

dition of TMB substrate to the array resulted in developing blue color, which indicated the peptides binding to antiFLAG-AB. Only FLAG sequence peptide developed blue color and the signal for other peptides was negligible. However, in a separate experiment we noticed that some of the peptides show high signals after exposure to antimouse/HRP-AB in the absence of antiFLAG-AB (Figure A2.5). We hypothesized that the peptides can mimic the Fc region of the primary antibody and therefore are recognized by an Fc binding secondary AB. This unexpected result can be rationalized by the use of **pA** in the screening procedure to immobilize
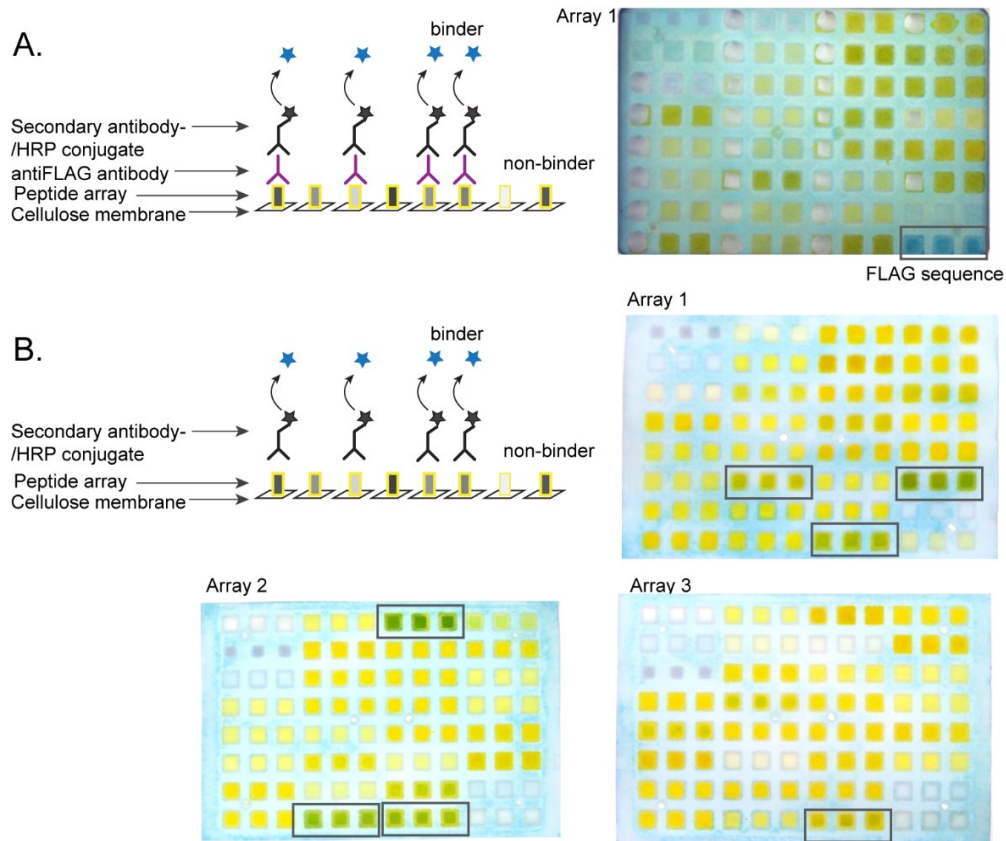
Figure A2.5. A) binding of antiFLAG-AB to the peptides in Array 1, probed by antimouse/HRP-AB. Only FLAG containing sequence developed blue color. B) Binding of antimouse/HRP-AB to the peptides on three arrays. High intensity signals are indicated in a rectangle.

the antiFLAG-AB: the modified library was enriched against **pA** instead of anti-FLAG-AB.

To investigate this hypothesis, we labeled the **pA** with fluorescein (**F-pA**) and incubated the array with **F-pA** (Figure A2.6A). Indeed, we observed significantly high fluorescent signals in ~30% of the peptides, which showed binding to **pA** (Figure A2.6B).
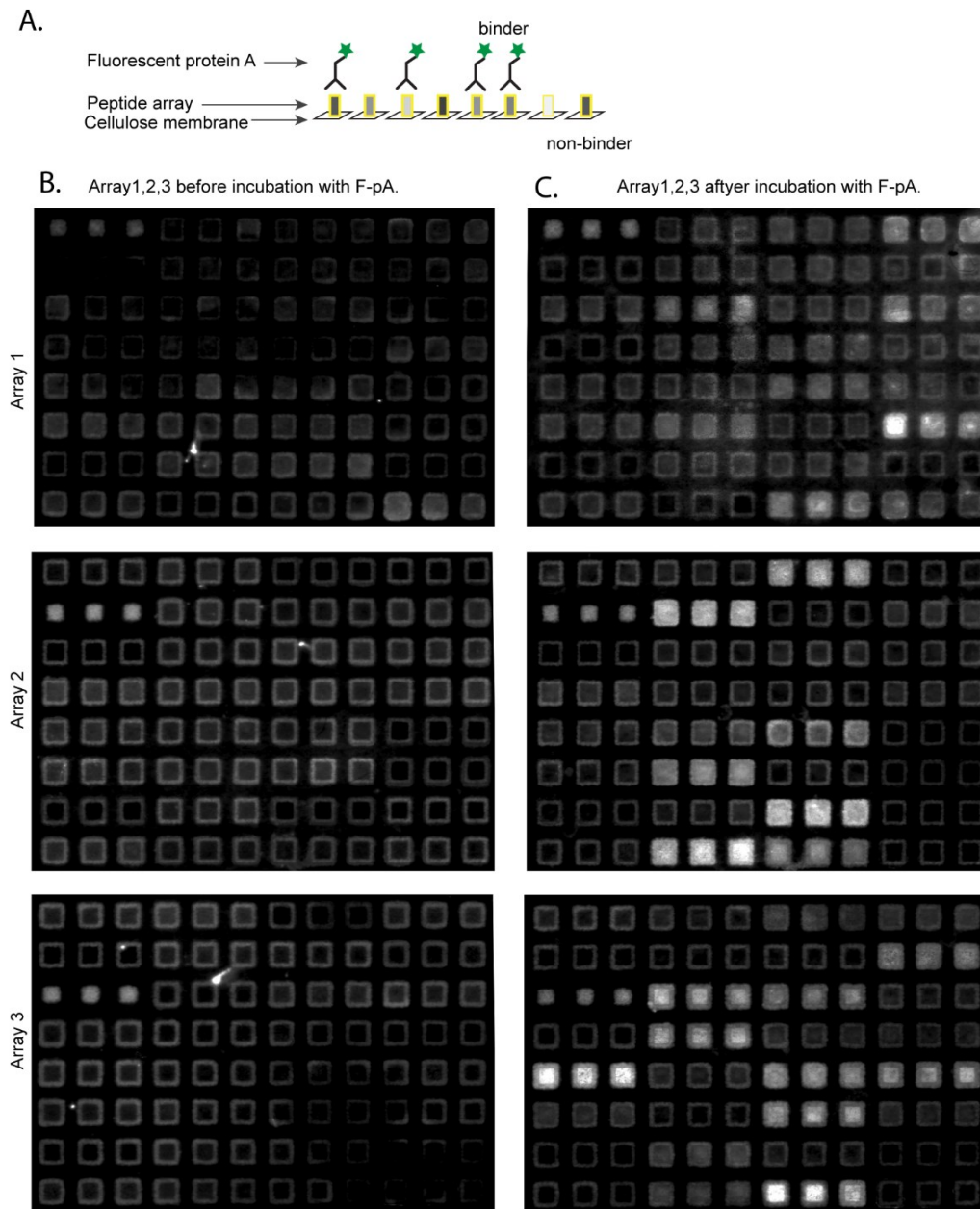
Figure A2.6. A) Scheme of the ELISA experiment to probe **F-pA** binders. 2) Fluorescence of peptide spots after incubation with **F-pA**.
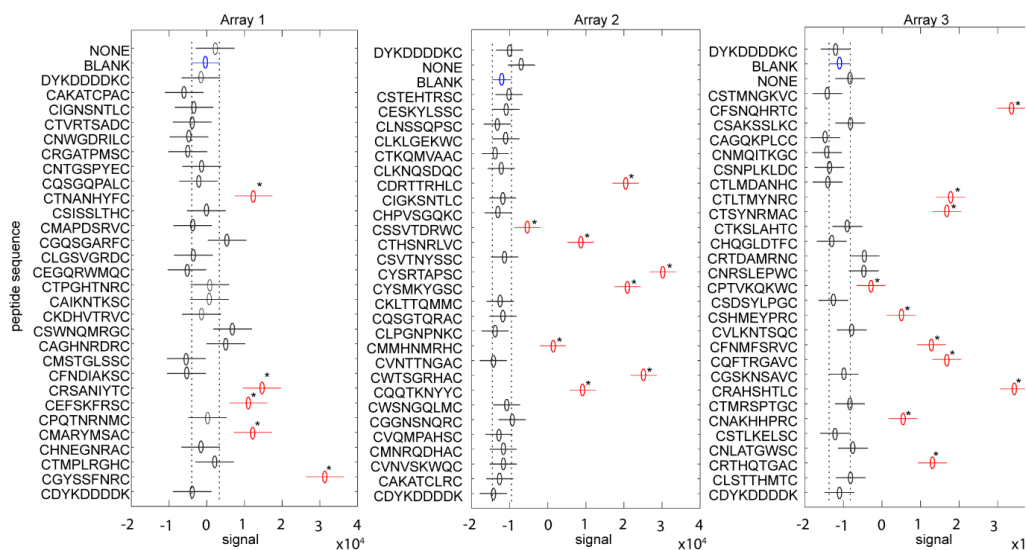
Figure A2.7. ANOVA analysis of 81 disulfide peptides synthesized on paper array after binding to **F-pA**. The signal was calculated as (fluorescence intensity after **F-pA** treatment - fluorescence intensity before **F-pA** treatment). The dotted line shows 95% confidence bound. Signals indicated by asterisk (*) show significant difference comparing to BLANK spot ($p < 0.05$).

We analyzed the binding data by ANOVA (Figure A2.7) and selected the peptides that showed significant difference comparing to the blank SPOT, in each array ($p < 0.05$). The analysis yielded 23 peptide sequences. We selected 10 peptides with the highest signal and synthesized them by solid phase peptide synthesis according to the methods described in chapter 4. We cyclized each peptide by BSBCA and through disulfide bond formation to yield potential LR-**pA** binders (**LPAB**s) or **pA** binders (**PAB**s), respectively. We used an ELISA assay to investigate if the binding of the cyclized peptides to **pA** can inhibit the binding of other ABs. We incubated the **pA** coated wells with different concentrations of **LPAB**s

and **PAB**s for 2 h and then added the antimouse/HRP-AB. After 1 h we added TMB substrate and measured the developed color by spectrophotometer after quenching the reaction by 2 M HCl. We observed no inhibition of AB in the presence of any of the cyclized peptides (Figure A2.8A, A2.8B). A positive control consisted of antiFLAG-AB showed ~50% inhibition at 25 nm concentration, and a negative control with no inhibitor showed same signal similar to all tested **PAB**s and **LPAB**s. These results showed that potential binding of neither **PAB**s or **LPAB**s can result in inhibition of the competitive AB binder.
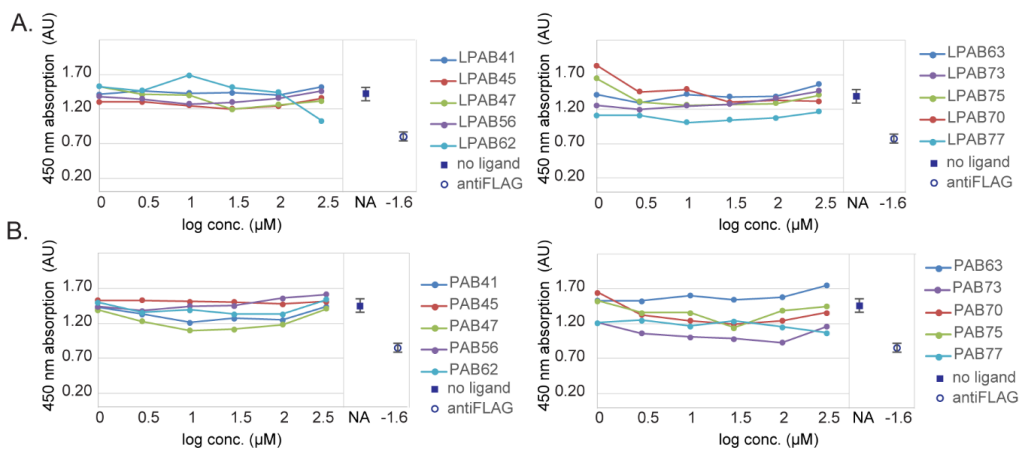
Figure A2.8. ELISA assay showed no significant inhibition of antimouse/HRP AB in the presence of **LPAB**s (A) or **PAB**s (B). AntiFLAG-AB inhibited the binding of competing antibody by ~50% at 25 nm concentration.

To directly measure the binding of the cyclized peptides with **pA**, we performed a fluorescence polarization assay for **LPAB56** which showed the highest signal on the paper array. We conjugated a fluorescein molecule to **LPAB56** at the
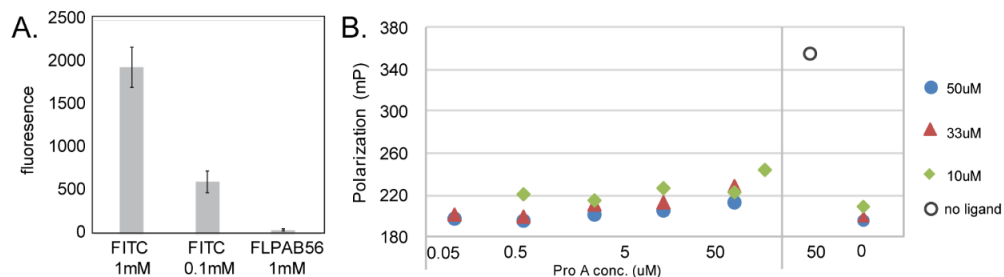
Figure A2.9. A) fluorescence of **FLPAB56** is quenched over 40 times comparing to FITC. B) changes in fluorescence polarization of **FLPAB56** at various concentrations of ligand and receptor **pA**.

N-terminus of the peptide to yield **FLPAB56**. This ligand showed a 40-fold decrease in fluorescence comparing to FITC at the same concentration, due to the quench of fluorescence with Azb (Figure A2.9A). Measuring the polarization of 50 μM and 33 μM of the ligand at various concentrations of **pA** (0.05 – 50 μM) showed no significant change in the polarization. However, a slight increase was observed when the concentration of the ligands was decreased to 10 μM. This observation suggests that high concentration of the ligands may have resulted in receptor depletion which masked the change in polarization for the bound ligand. However, low fluorescence intensity due to Azb quenching prevented measuring the polarization in lower concentrations of the ligand. This problem can be circumvented by inserting a linker between the cyclized ligand and the fluorophore which can potentially result in less quenching of the fluorescence.[4] Another strategy is to use far red fluorophores, because the Azb in BSBCA might be less efficient in quenching such fluorophores.

Conclusion

We described attempts to identify Azb-cyclized ligands that bind to **pA**. The screening was planned to identify antiFLAG-AB binders, however, the library was enriched against a secondary target (**pA**) which was used to immobilize the target protein. These results highlight the importance of "negative screening" in which the library is depleted of **pA** binders by preliminary selection against **pA**. A peptide array based screening revealed the binder sequences that showed high binding for **pA** on the array; however, we were not successful in measuring the binding of the peptides to the **pA** in solution. Redesigning the fluorescent peptide constructs can increase the sensitivity for fluorescent polarization assays for improved binding measurement.

Material and methods

General information

The protein A was purchased from Invitrogen (10-1006, #1133487A). FITC was purchased from Sigma. Anti FLAG antibody was purchased form Genescript (A01429-100). BSBCA was prepared as described in chapter 2. 1-Step Ultra TMB-ELISA Substrate Solution was purchased from Thermo-Fisher. **PAB** ligands were synthesized by SPPS according to the method described in chapter 4. **LAPAB**s were synthesized according to the method described in chapter 2.

Labeling pA with fluorescein

To 1 mL of a 2 mg/mL solution of protein A in 50 mM borate buffer (pH 8.5) we added 1 mg FITC in 50 µL DMF. We incubated this solution in dark at r.t. for 2 h and then dialyzed it against 5 L PBS for 3 × 8 h at 4 °C, using a 3 mL Slide-A-lyzer (Thermo, 10 K cutoff).

BSBCA cyclization of peptides array

We incubated each peptide array for in 5 mM TCEP solution in TRIS pH 8.5 for 20 min. Then we added 2.5 mg of BSBCA to the solution and rocked the array for 2 h in dark. We washed the array with 5 × 30 mL mQ H_2O followed by 3 × 10 mL MeOH. The array was air dried and incubated in -20 °C for long term storage.

Binding of antiFLAG-AB (F-pA) to peptide array.

We incubated the array of cyclized peptide in 10 mL of a 2% solution of BSA in PBS (BSA-PBS, pH 7.4) for 1 h to block the peptides. Then we washed the array with 3 × 30 mL PBS supplemented with Tween®-20 (PBST) for 5 min. We incubated the array with 10 mL of a 5 µg/mL solution of anti FLAG antibody in 2% BSA-PBS. We washed the array with 3 × 30 mL PBST and then incubated it in 10 mL of antimouse/HRP conjugate (1/5000 dilution) in 2% BSA-PBS. The array

was washed with 3 × 30 mL PBST and blotted against filter paper. We placed the array in a container containing 5 mL of a TMB substrate solution and quenched it after 5 min with 5 mL of 2 M H_2SO_4

Binding of fluorescent pA (F-pA) to peptide array.

We incubated the array of cyclized peptide in 10 mL of a 2% solution of BSA in PBS (BSA-PBS, pH 7.4) for 1 h to block the peptides. Then we washed the array with 3 × 30 mL PBS supplemented with Tween®-20 (PBST) for 5 min. We incubated the array with 10 mL of a 10 μg/mL solution of **F-pA** in 2% BSA-PBS for 2 h. We washed the array with 3 × 30 mL PBST and measured the fluorescence by Typhoon FLA 7000 (GE Healthcare Life Sciences) using 473 nm excitation and BPB1 530 nm emission filter. The analysis of the fluorescence was performed by custom made MATLAB scripts.

MATLAB script for quantification of fluorescence on peptide array.

Below is the explanation of some parameters

Script	Explanation
`'image.tif'`	is the name of the image out of Typhoon.
`red = B (:,:);`	reads the file in grayscale.
`w=5;`	specifies that the with of the reading area is 10 px (5px to right, left, up and down)
`for i=1:12`	(12 column format)
`for j=1:8`	(8 rows format)
`ROW = [1 1 1 2 2 2 3 3 3 4 4 4 ];`	indicates the format of replicates

Scripts starts here:

```
clear all
close all

B = imread('image.tif');
red = B (:,:);
imagesc(red);

 [top]=ginput(1);
[bot]=ginput(1);

d1 = ( bot(1) - top(1)) / 11;
d2 = ( bot(2) - top(2)) / 7;

w=5;
value=[];

for i=1:12
    for j=1:8

        C1 = round ( top(1) + (i-1)*d1 );
        C2 = round ( top(2) + (j-1)*d2 );

        row = C2-w : C2+w;
        col = C1-w : C1+w;

        square = red( row, col );
        value(i,j) = mean(mean(square));

        red( row, col )=0;
    end
end

figure(1)
imagesc(red);

value2=value'

ROW = [1 1 1 2 2 2 3 3 3 4 4 4 ];
ID=[];

for i=1:8
    ID = [ ID; ROW+5*(i-1) ];
end


figure(2);
plot(ID,value','ok');
```

end of the script

MATLAB script for ANOVA analysis of the fluorescence results for three arrays

```
clear all;
[read, peptides] = xlsread ('Matlab results for ANOVA.xlsx');
pmemONE = peptides (2:97, 1);
pmemTWO = peptides (2:97, 4);
pmemTHR = peptides (2:97, 8);
rmemONE = read (1:96, 1);
rmemTWO = read (1:96, 4);
rmemTHR = read (1:96, 8);


%[P1,~,stat1] = anova1(rmemONE, pmemONE);
%[p2,~,stat2] = anova1(rmemTWO, pmemTWO);
[p3,~,stat3] = anova1(rmemTHR, pmemTHR);


%multcompare (stat1,0.05)
%multcompare (stat2,0.05)
 multcompare (stat3,0.05)
```

ELISA assay of LPABS and PABS

A 96 well plate was coated by adding 50 µL of 0.1 µg/mL of **pA** in NaHCO$_3$ buffer (pH 8.5) into each well, followed by overnight incubation at 4 °C. The plate was washed with 3 × 150 µL of PBST before adding the ligands. **LPAB** and **PAB** ligands were dissolved in 50 mM NH$_4$CO$_3$ (pH7.4) buffer to make a 1 mM stock solution. The stock solution was serially dilated to make a total concentration of 333, 111, 37, 12.3, 4.1 and 1.4 µM solution. We then transferred 120 µL of the

ligand solution to the **pA** coated well and incubated it for 2 h at r.t. in the dark. We washed the wells with 3 × 150 µL of PBST and added 50 µL of antimouse/HRP conjugate (1/5000 dilution) in 2% BSA-PBS for 30 min. The plate was washed with 3 × 30 mL PBST and 50 µL of TMB substrate solution was added. The reaction was quenched after 5 min with 50 µL of 2 M H_2SO_4 and the absorption of the solution was measured by a spectrophotometer at 450 nm.

References

(1)   Fehrentz, T.; Schönberger, M.; Trauner, D. *Angew. Chem. Int. Ed.* **2011**, *50*, 12156.

(2)   Szymański, W.; Beierle, J. M.; Kistemaker, H. A. V.; Velema, W. A.; Feringa, B. L. *Chem. Rev.* **2013**, *113*, 6114.

(3)   Jafari, M. R.; Deng, L.; Kitov, P. I.; Ng, S.; Matochko, W. L.; Tjhung, K. F.; Zeberoff, A.; Elias, A.; Klassen, J. S.; Derda, R. *ACS Chem. Biol.* **2014**, *9*, 443.

(4)   Bellotto, S.; Chen, S.; Rentero Rebollo, I.; Wegner, H. A.; Heinis, C. *JACS* **2014**, *136*, 5880.

(5)   Matochko, W. L.; Chu, K.; Jin, B.; Lee, S. W.; Whitesides, G. M.; Derda, R. *Methods* **2012**, *58*, 47.

(6)   Deiss, F.; Matochko, W. L.; Govindasamy, N.; Lin, E. Y.; Derda, R. *Angew. Chem. Int. Ed.* **2014**, *53*, 6374.

^{1}H-NMR of **1b**

¹³C-NMR of **1b**

240 220 200 180 160 140 120 100 80 60 40 20 0 ppm

¹H-NMR of **1c**

9.988

7.885
7.868

6.957
6.952
6.470
6.465
6.453
6.448

4.896

3.313

2.531
2.499
2.495
2.491
2.488
2.484
1.954

H₂O

DMSO

1c

| 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | ppm |

0.99

1.00

0.99   1.05

2.03

2.22

0.18

3.10

¹³C-NMR of **1c**

^{13}C-NMR of **1c**

166.757
144.070
136.679
125.178
121.294
114.831
114.825
114.819
112.844
112.840
112.832

39.934
39.923
39.917
39.891
39.870
39.864
39.827
39.782
39.778
39.706
39.700
39.696
39.690
39.661

25.184

1c

240   220   200   180   160   140   120   100   80   60   40   20   0   ppm

¹H-NMR of **1d**

¹H-NMR of **1e**

^{13}C-NMR of **1e**
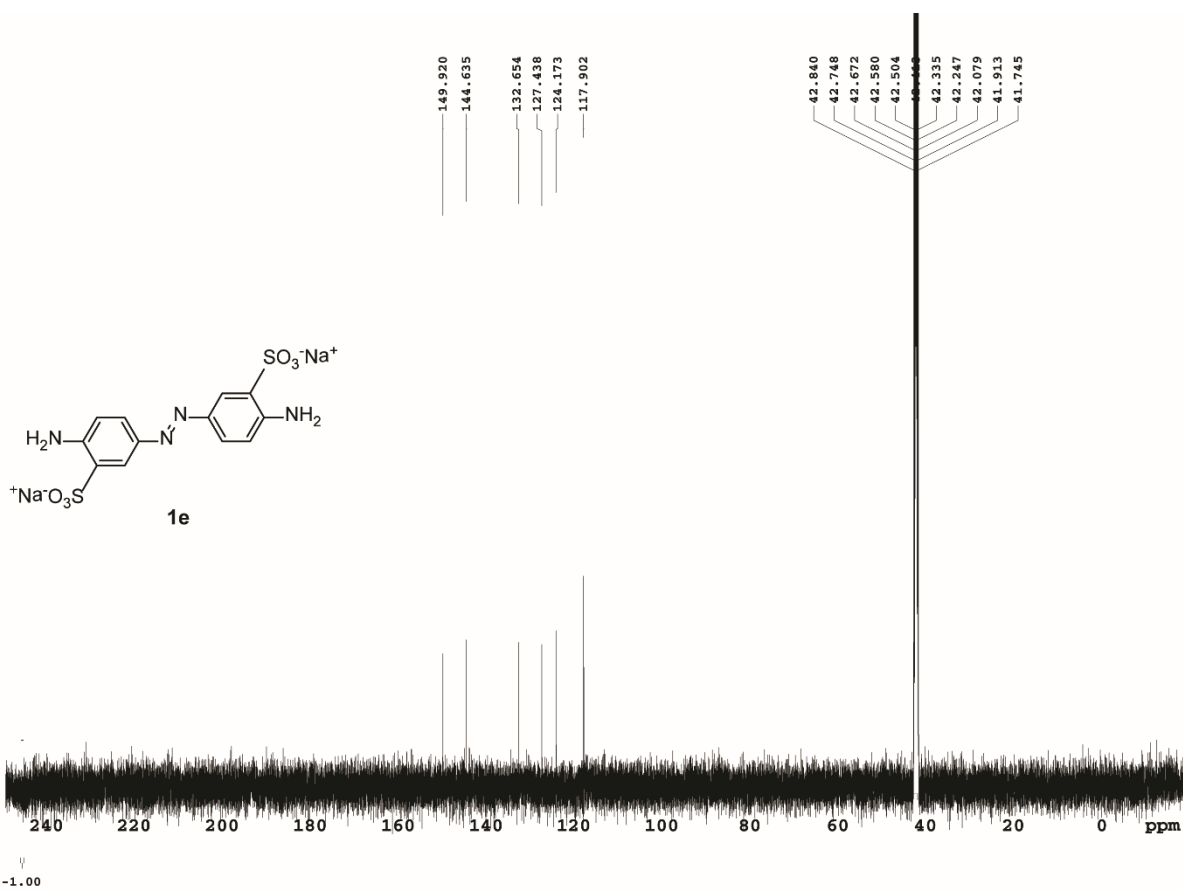
¹H-NMR of **1f**
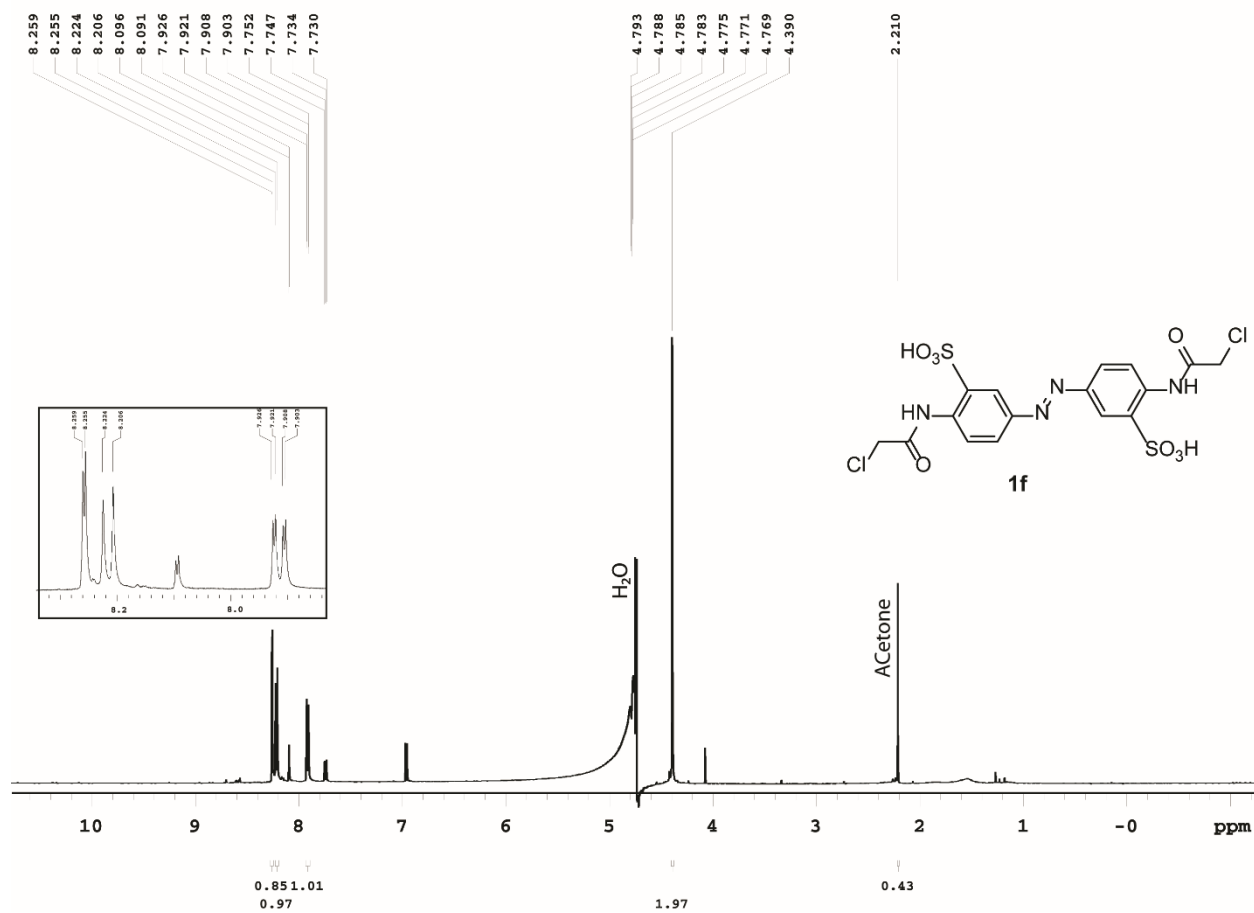
¹³C-NMR of **1f**
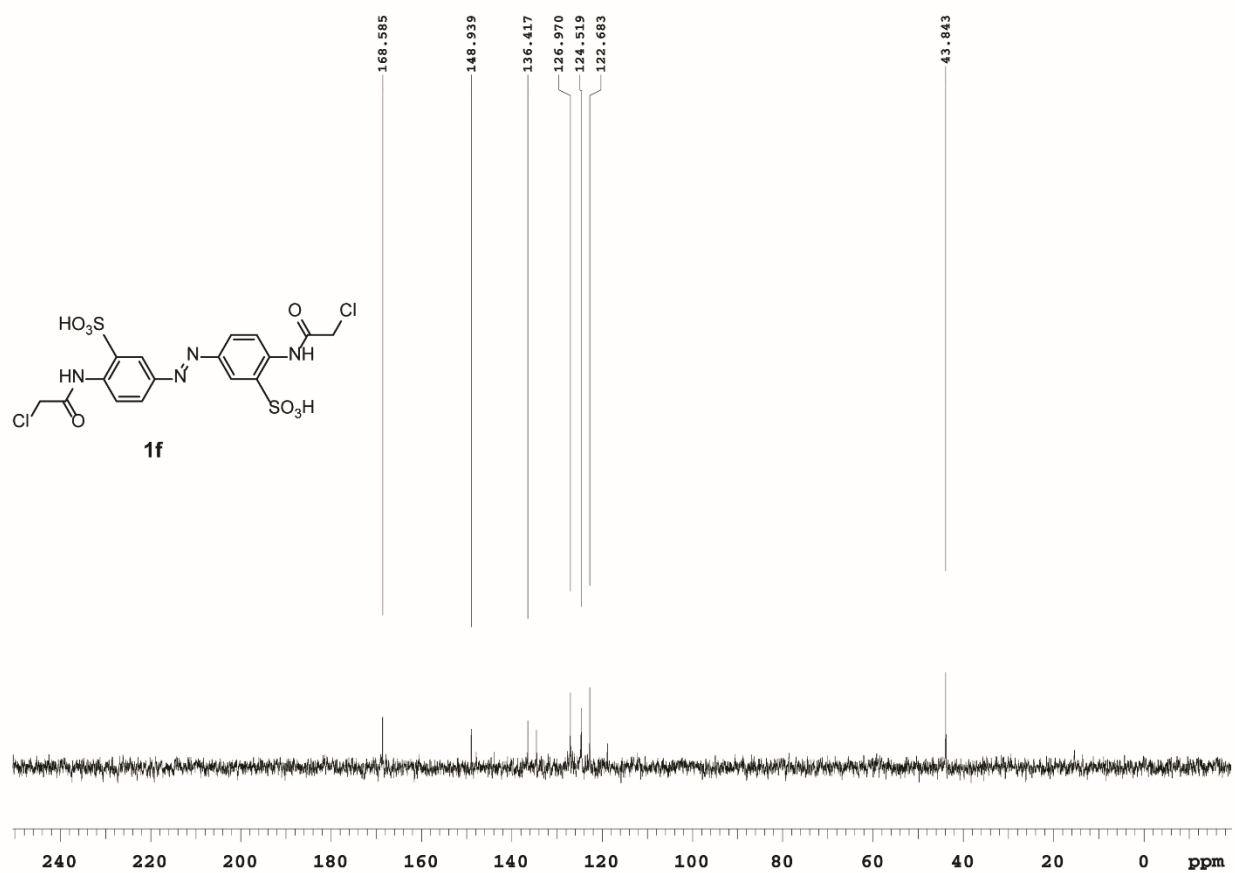
¹H-NMR of **BSH**

¹³C-NMR of **BSH**

207

^{1}H-NMR of **L36**

10.457 8.928 8.567 8.343 8.259 8.061 7.924 7.824 7.515 7.199 7.105 7.039 6.834 6.758 6.412 6.349 5.068 5.004 4.842 4.746 4.666 4.411 4.301 4.221 4.127 4.071 3.913 3.834 3.772 3.664 3.550 3.349 3.204 3.105 3.029 2.923 2.850 2.721 2.622 2.363 2.255 2.058 1.993 1.914 1.756 1.591 1.536 1.480 1.365 1.240 1.147 0.850 0.637 0.000

H2O

DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid)

TOCSY ^{1}H-NMR of **L36**

F2 (ppm)

TOCSY ^{1}H-NMR of **L36 (Cntd.)**

699.740 MHz H1 zTOCSY in d2o
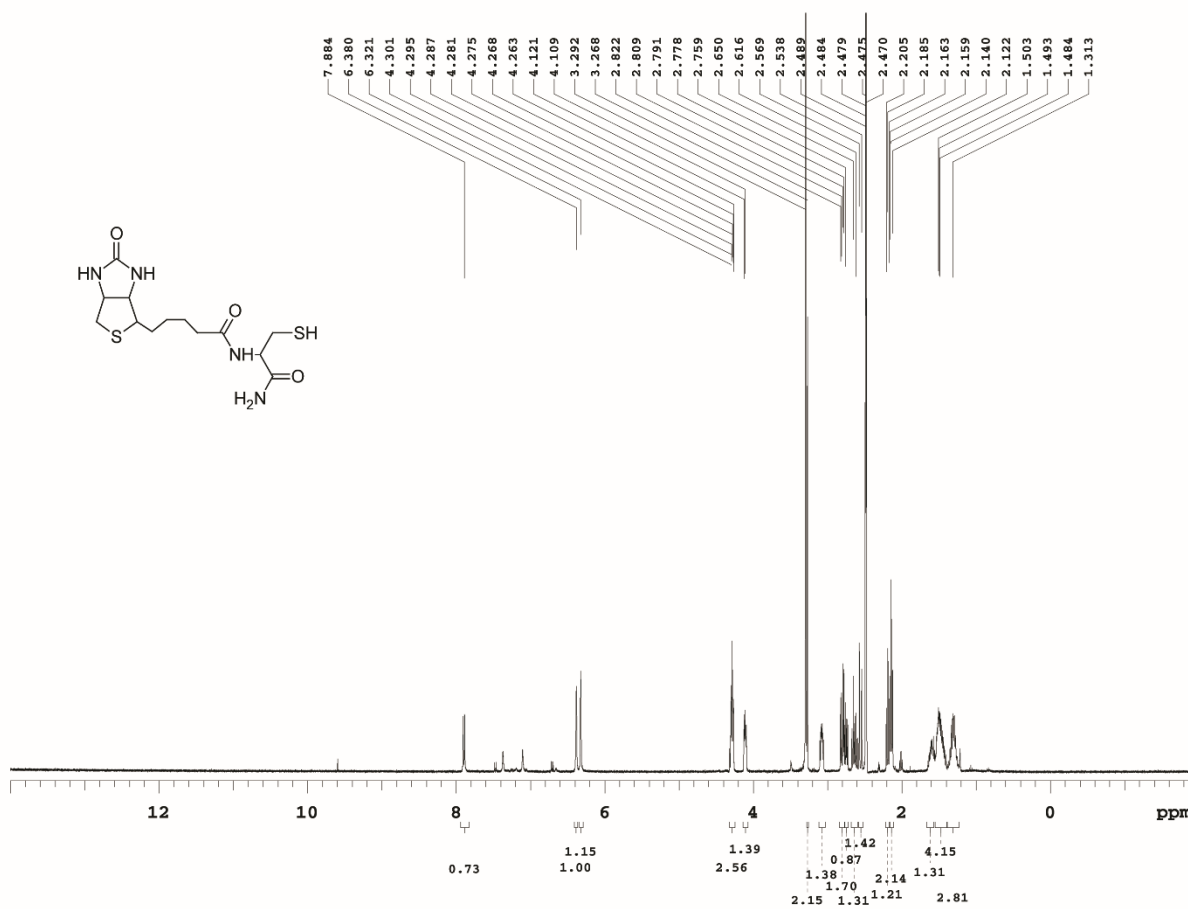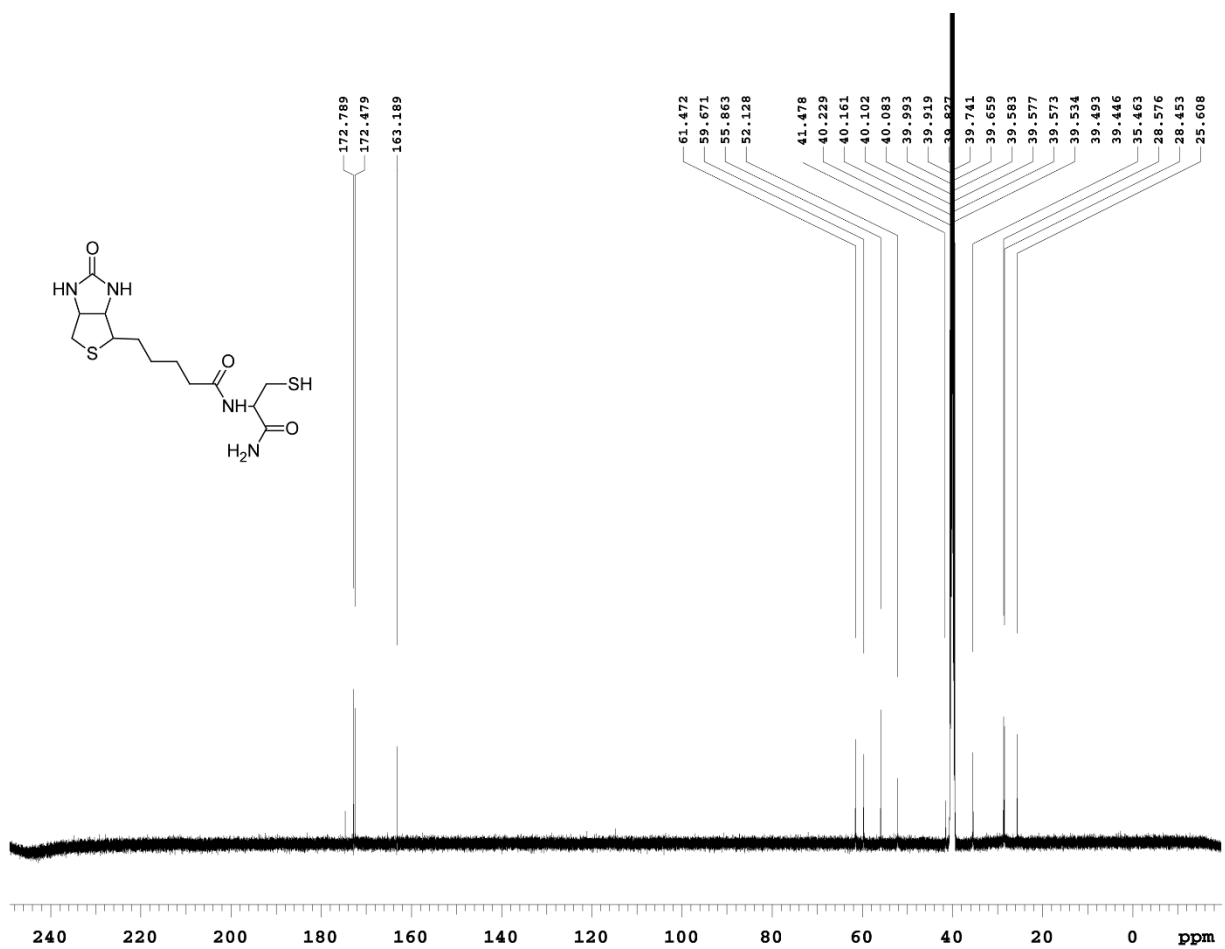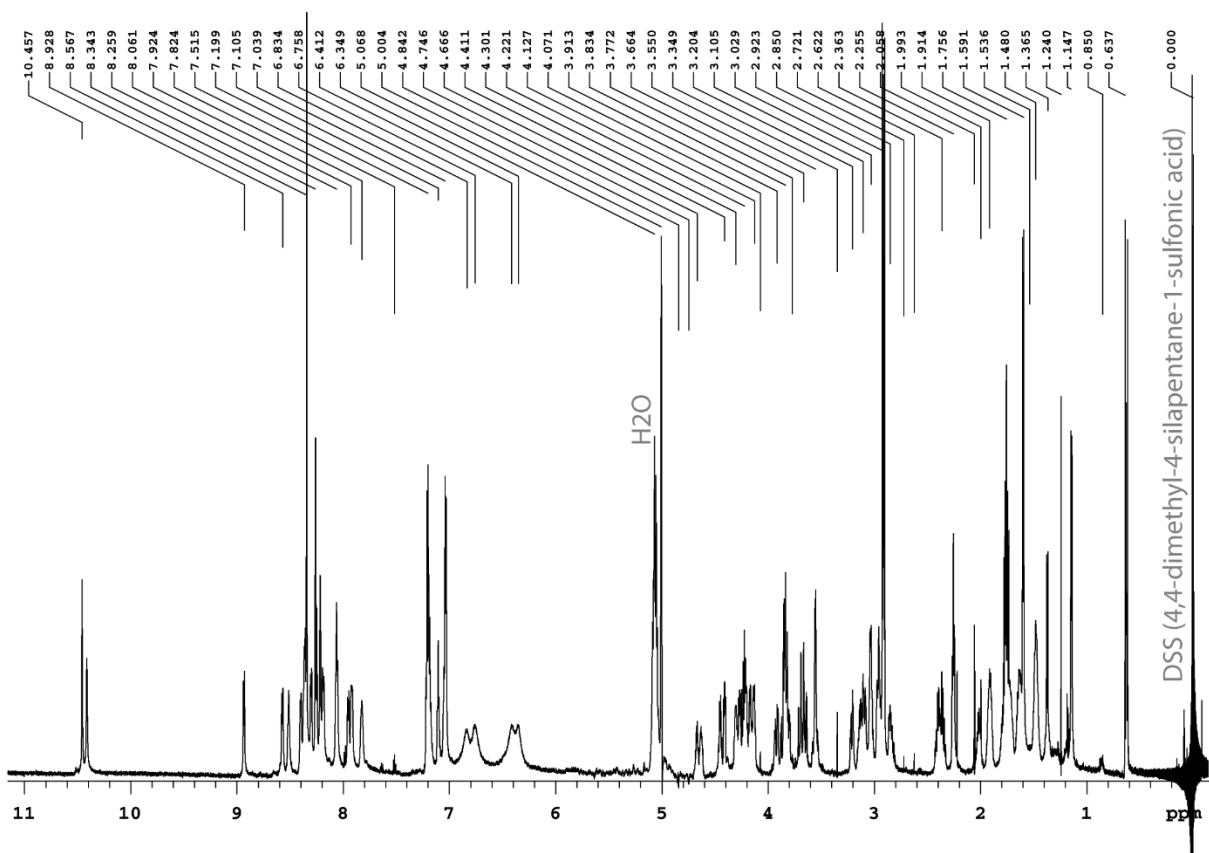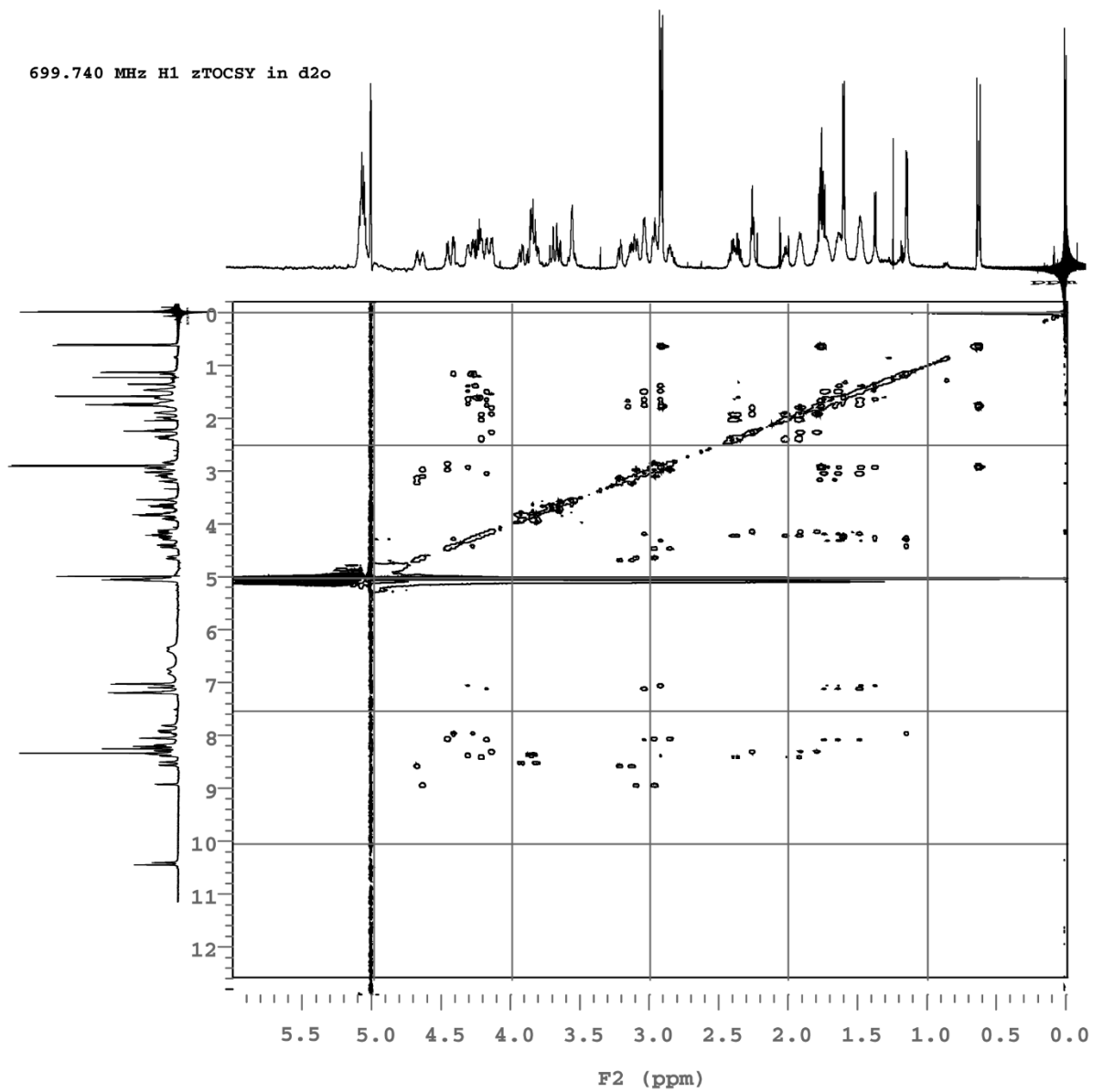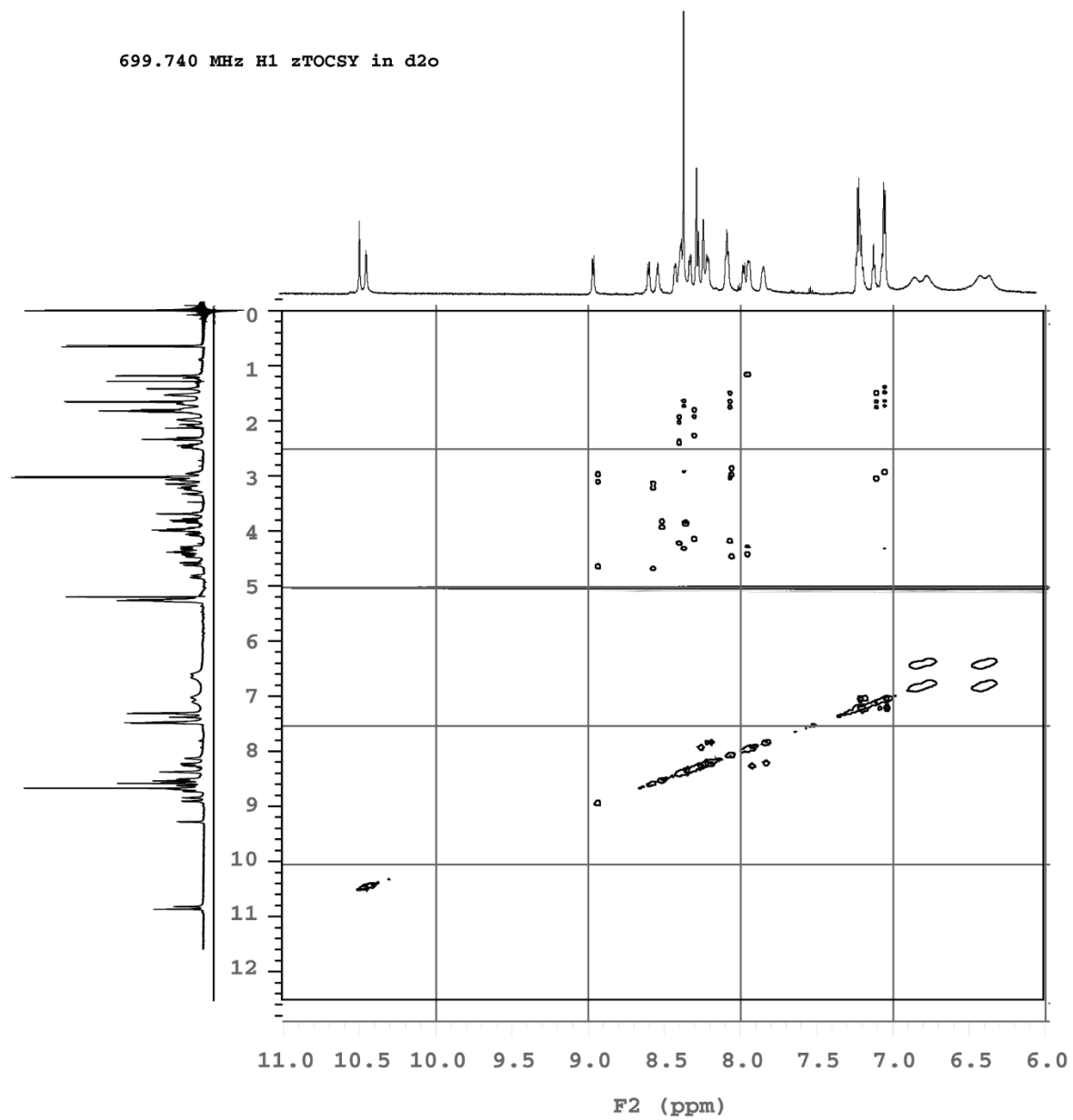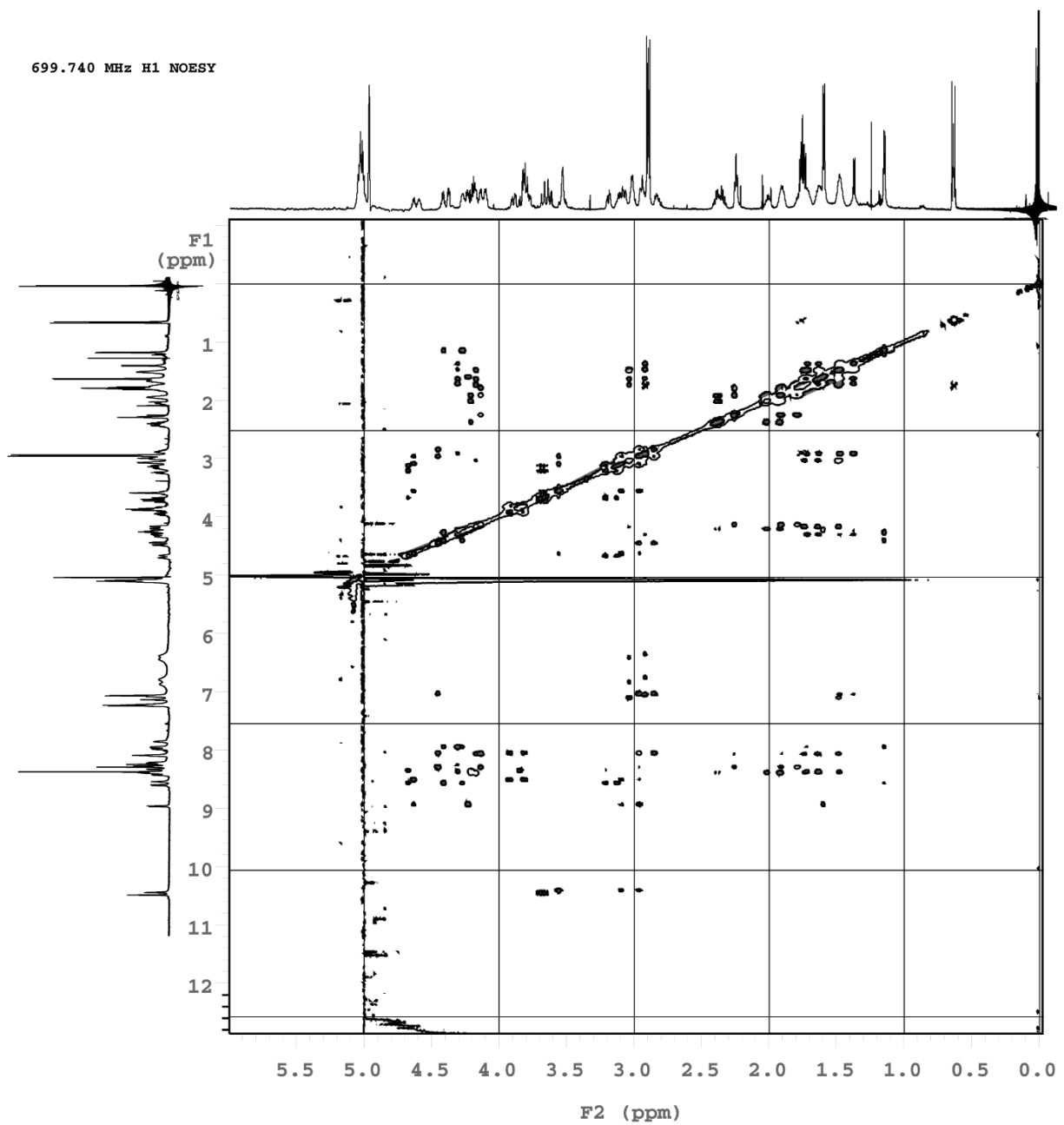
NOESY ^{1}H-NMR of **L36**

NOESY ¹H-NMR of **L36 (Cntd)**

699.740 MHz H1 NOESY
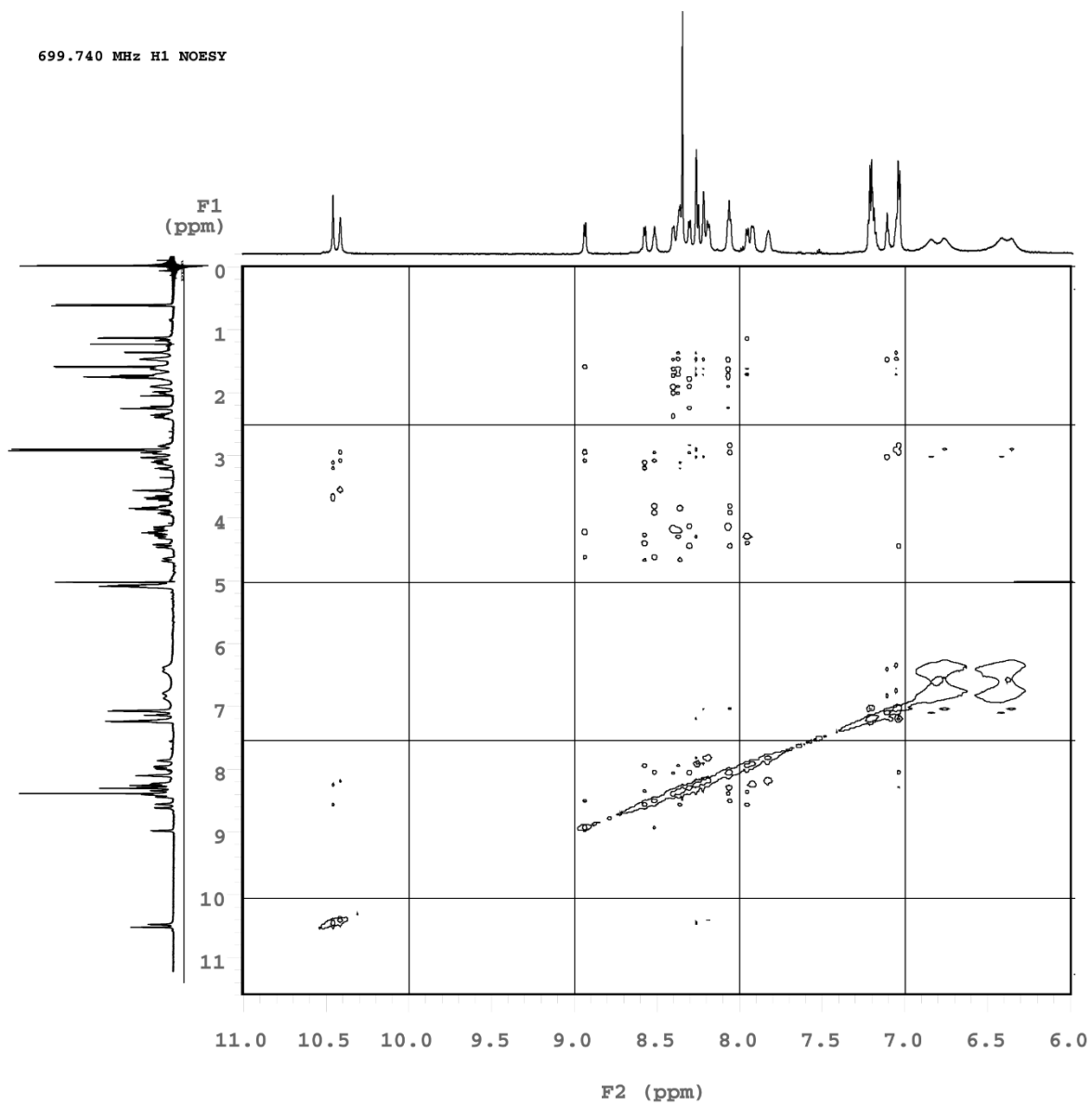
Appendix 4 Supporting Information for Chapter 3

TOCSY-NMR of cyclized peptide ACGFERETCG (**16**)

For numbering of protons, see Table 3.1 and Scheme 3.2.

¹H- NMR of cyclized peptide ACGFERETCG (**16**)
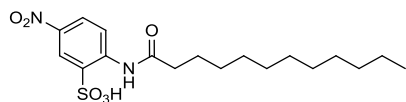
599.927 MHz H1 water_ES in d2o

H2O

MeOH

Pulse Sequence: water_ES

^{1}H-NMR of **2**

399.796 MHz H1 1D in dmso (ref. to DMSO @ 2.49 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe
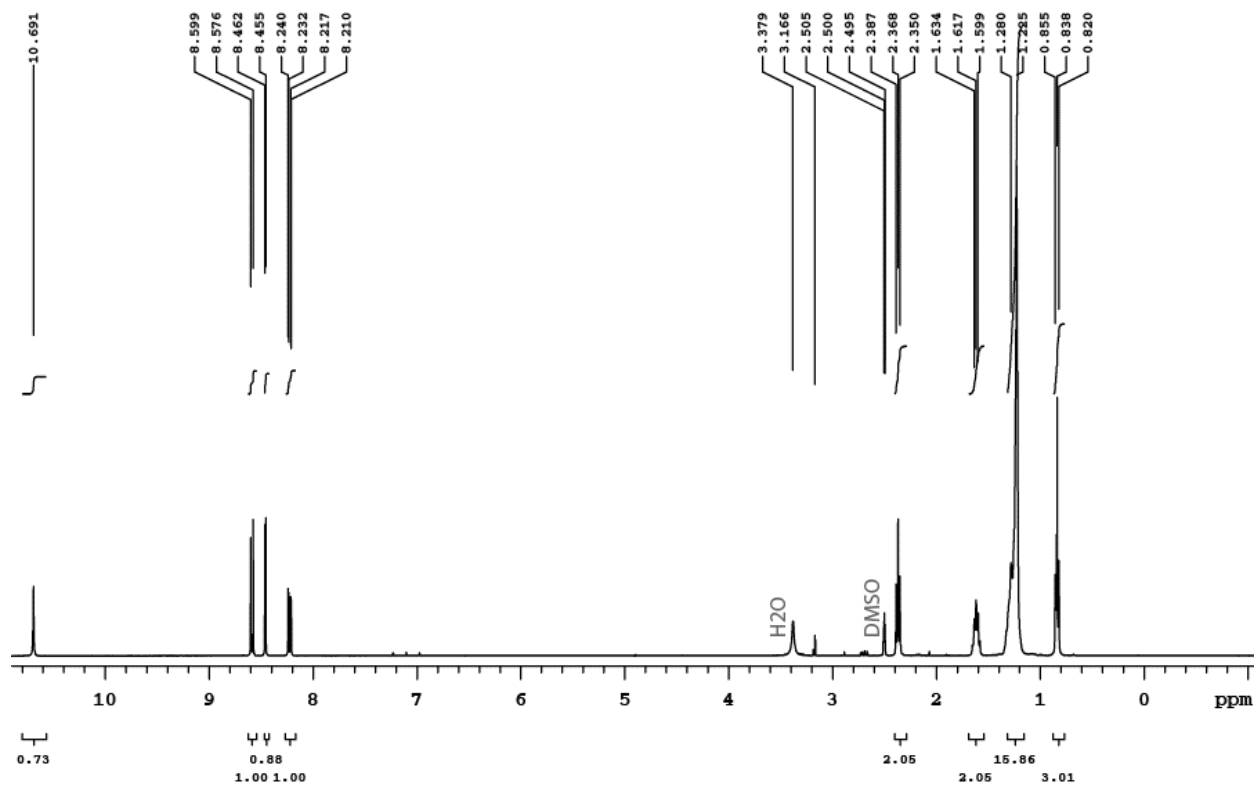
¹³C-NMR of **2**
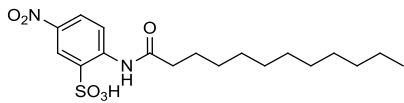
100.540 MHz C13[H1] 1D in dmso (ref. to DMSO @ 39.5 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe
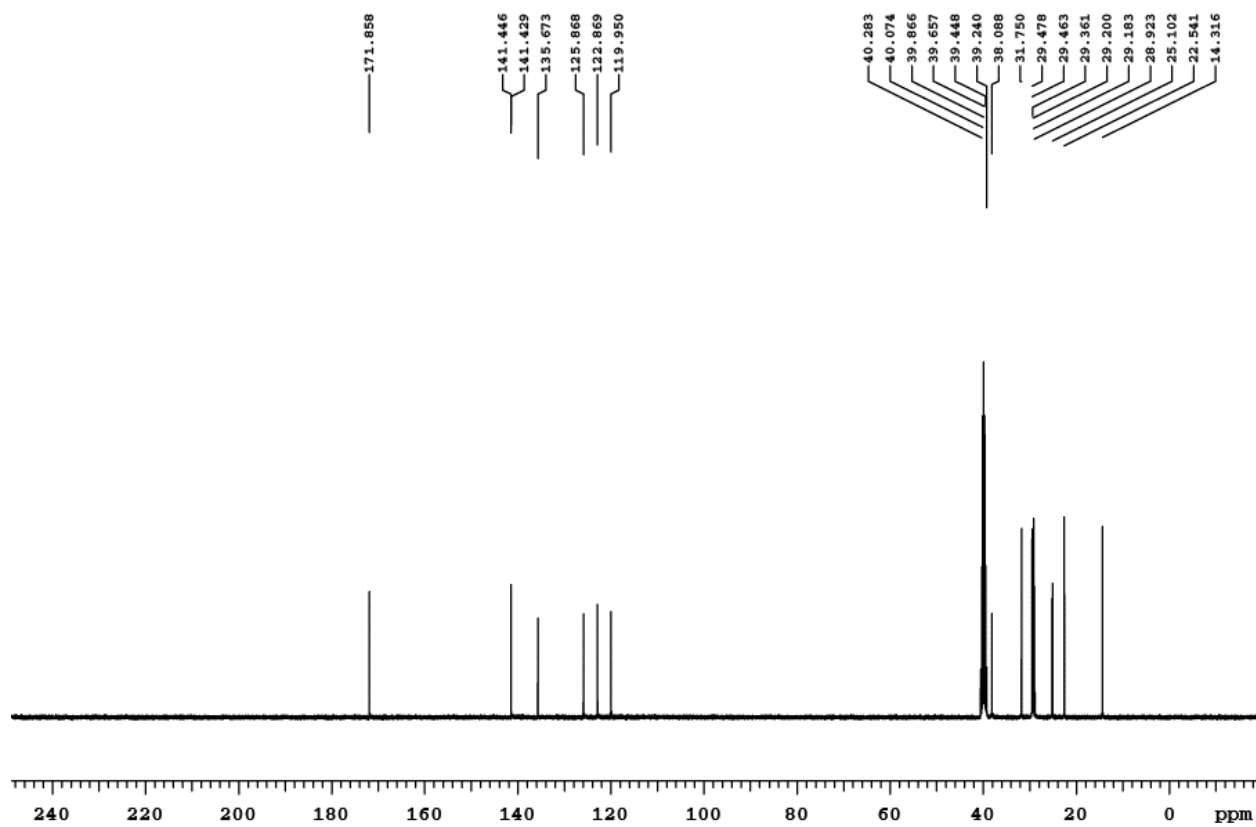
^{1}H-NMR of **3**
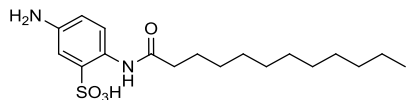
The peaks marked with asterisk (*) show residual ethyl acetate.

399.796 MHz H1 1D in dmso (ref. to DMSO @ 2.49 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe
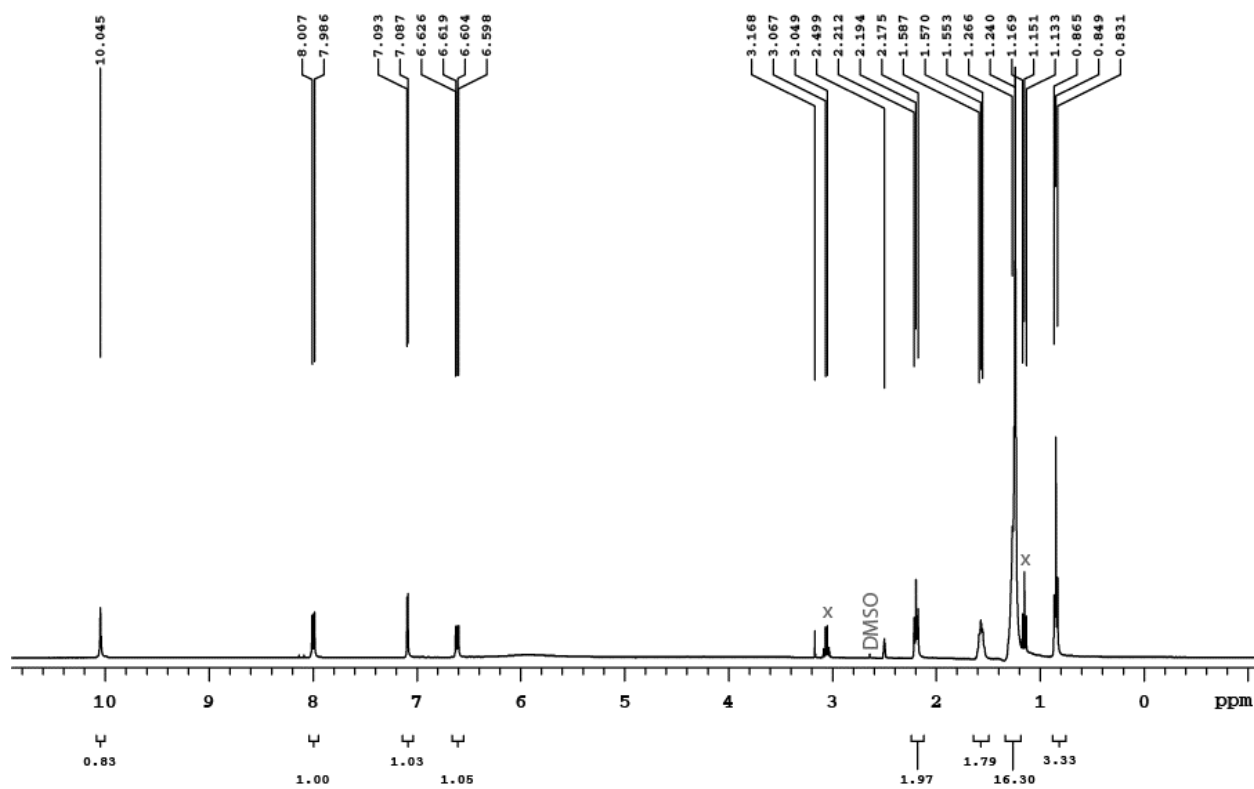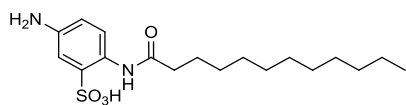
¹³C-NMR of **3**

100.540 MHz C13[H1] 1D in dmso (ref. to DMSO @ 39.5 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe
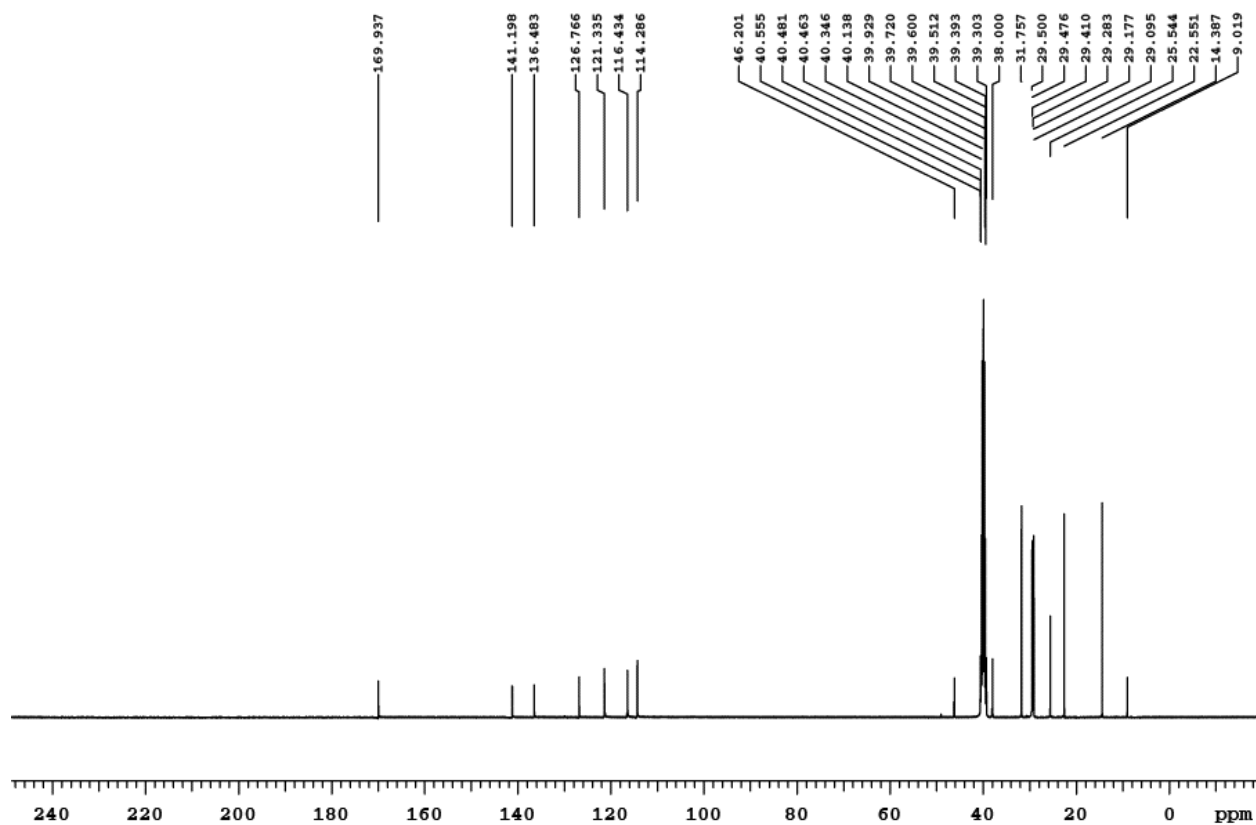
¹H-NMR of 2-[(4-methoxybenzyl)amino]benzenesulfonic acid (**4**)
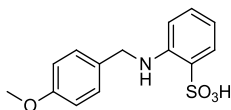
299.973 MHz H1 1D in dmso (ref. to DMSO @ 2.49 ppm), temp 27.0 C -> actual temp = 27.0 C, id300 probe
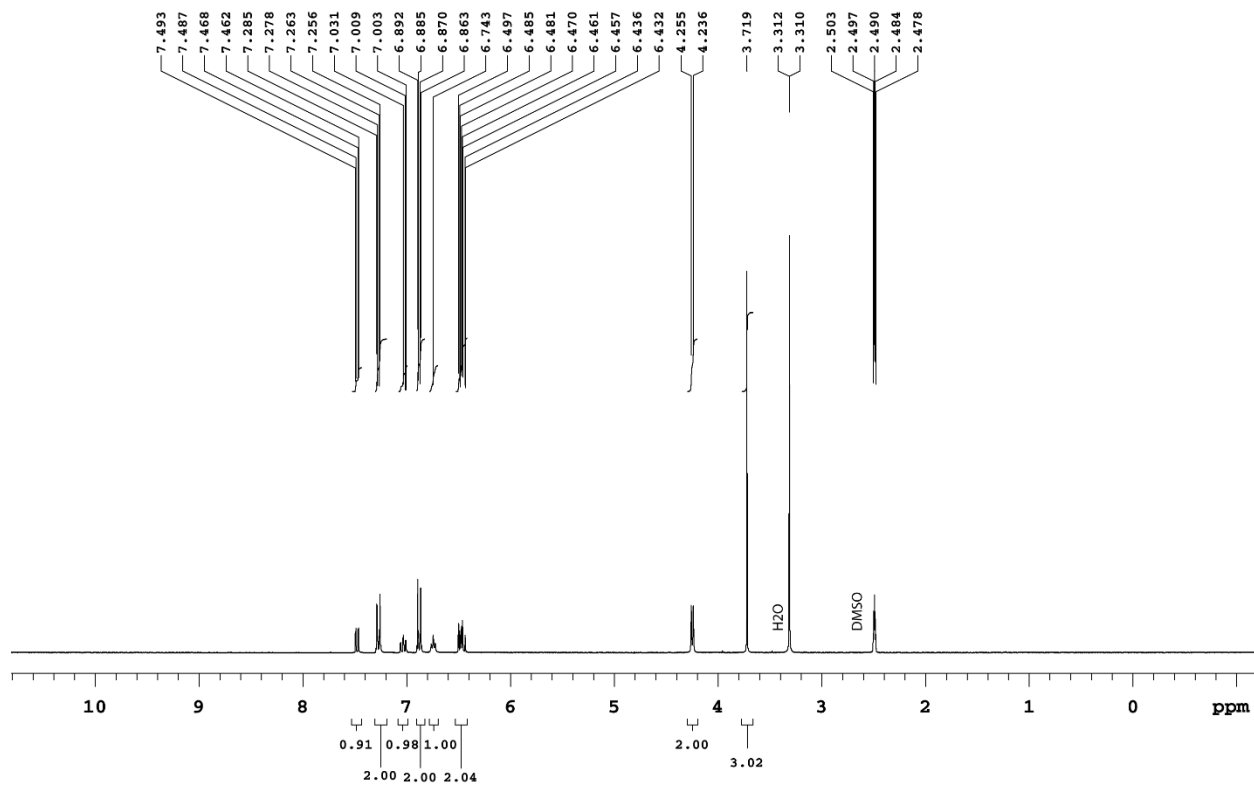
^{13}C-NMR of 2-[(4-methoxybenzyl)amino]benzenesulfonic acid (**4**)
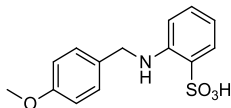
100.540 MHz C13[H1] 1D in dmso (ref. to DMSO @ 39.5 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe

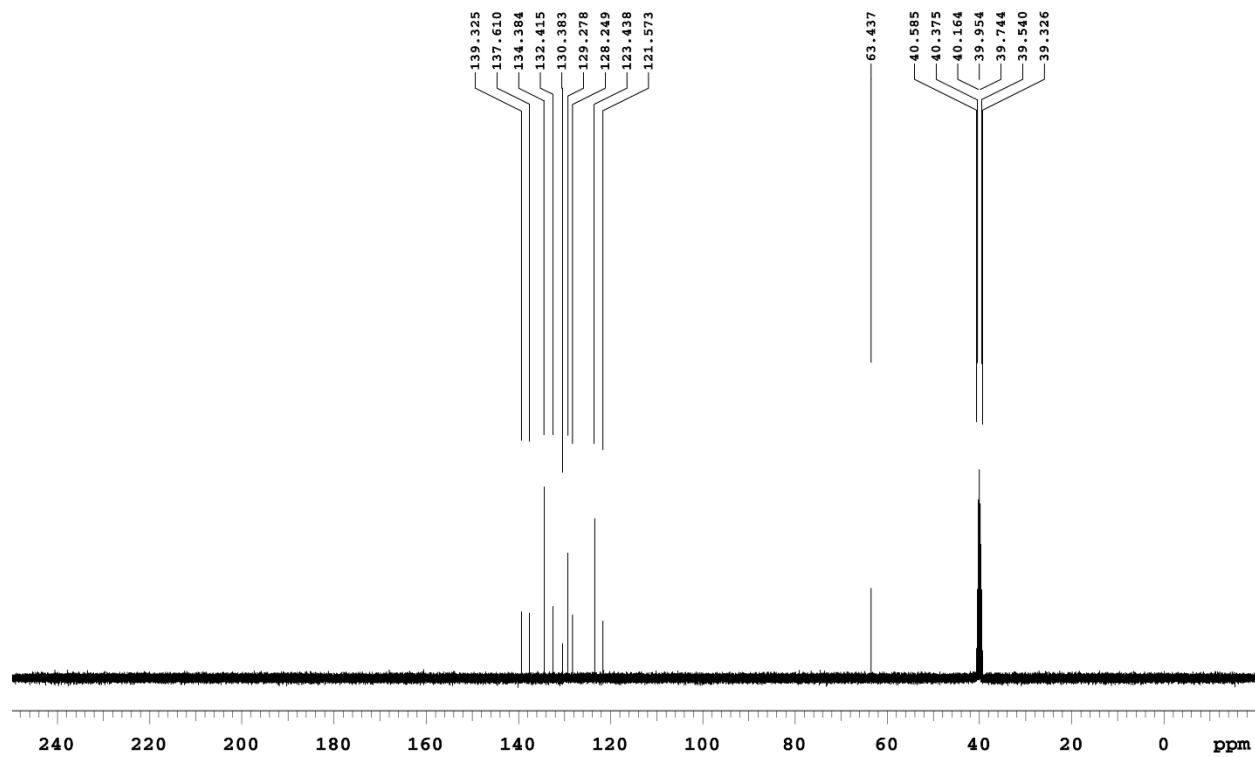158.631
145.193
131.997
131.217
130.375
128.814
127.629
114.592
114.255
111.035
55.488
46.410
40.563
40.354
40.146
39.937
39.728
39.520
39.311
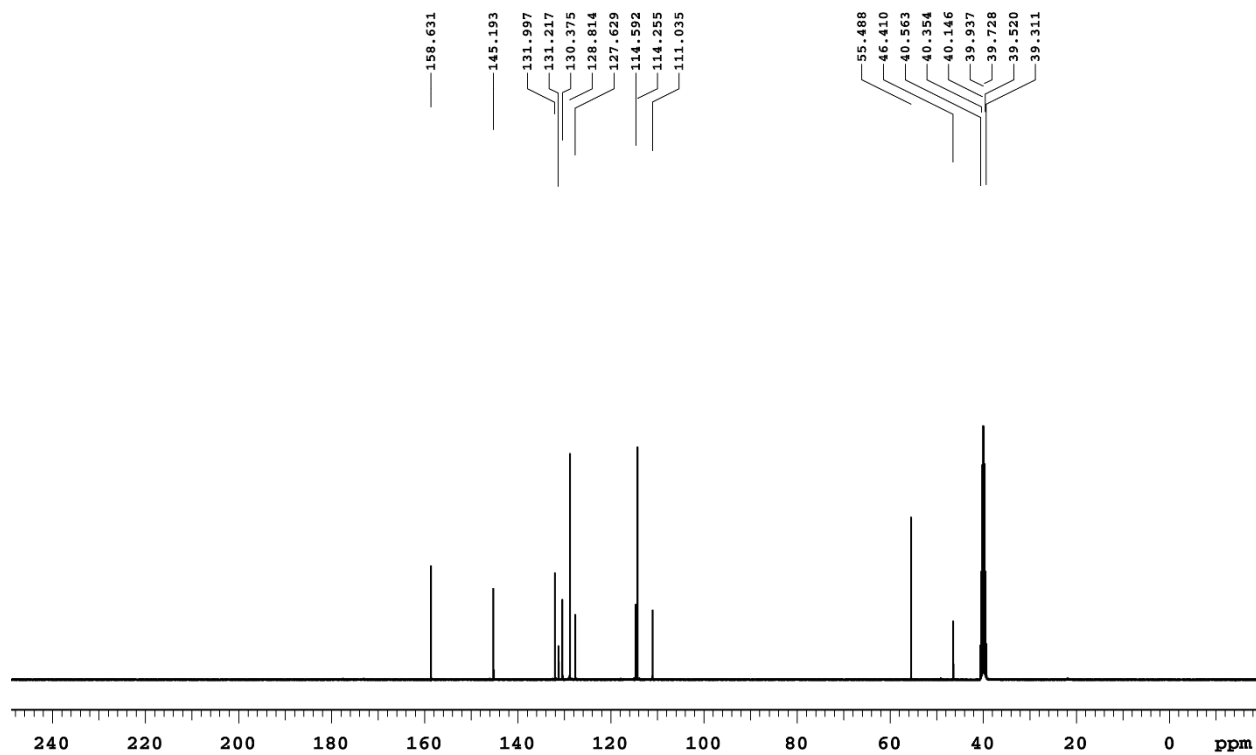
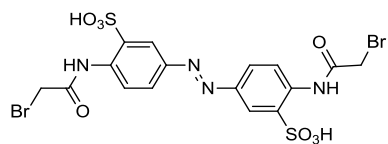^{1}H-NMR of 3,3'-bis(sulfonato)-(4-methoxybenzyl)amido-(4'-dodecanoylamido)-azobenzene (**5**)
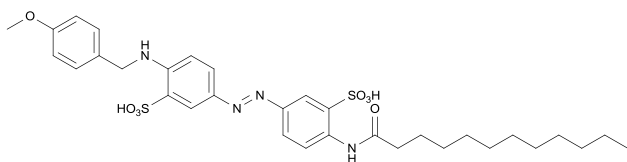
399.796 MHz H1 1D in dmso (ref. to DMSO @ 2.49 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe
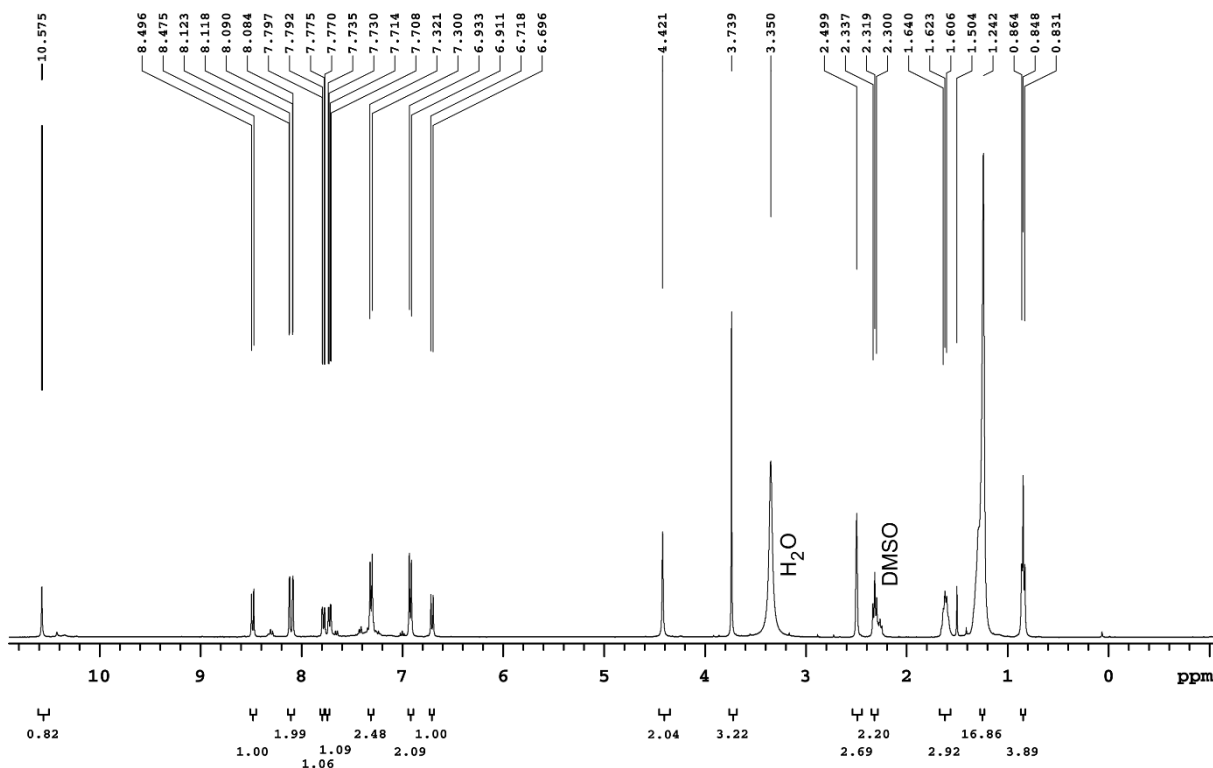
^{13}C-NMR of 3,3'-bis(sulfonato)-(4-methoxybenzyl)amido-(4'-dodecanoylamido)-azobenzene (**5**)
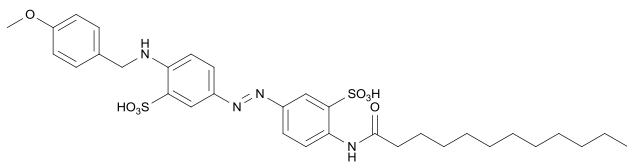
100.540 MHz C13[H1] 1D in dmso (ref. to DMSO @ 39.5 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe

171.193
158.817
147.887
147.214
141.973
136.829
136.133
131.524
131.148
128.898
127.719
125.424
121.130
120.220
119.357
114.402
110.908

55.515
46.236
40.592
40.383
40.174
39.966
39.757
39.548
39.340
38.168
31.744
29.476
29.457
29.371
29.220
29.167
29.005
25.350
22.545
14.404

240 220 200 180 160 140 120 100 80 60 40 20 0 ppm

Pulse Sequence: s2pul

222

¹H-NMR of 3,3′-bis(sulfonato)-4,4′-bis(amino)azobenzene (**6**)
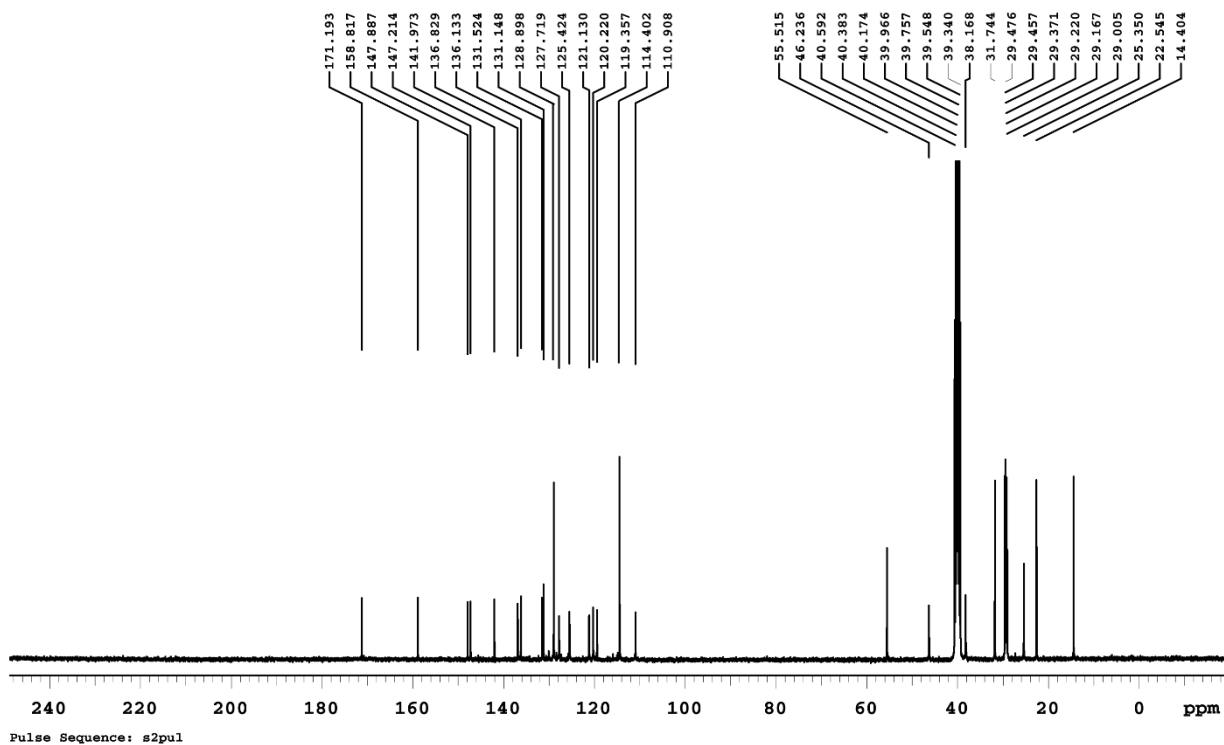
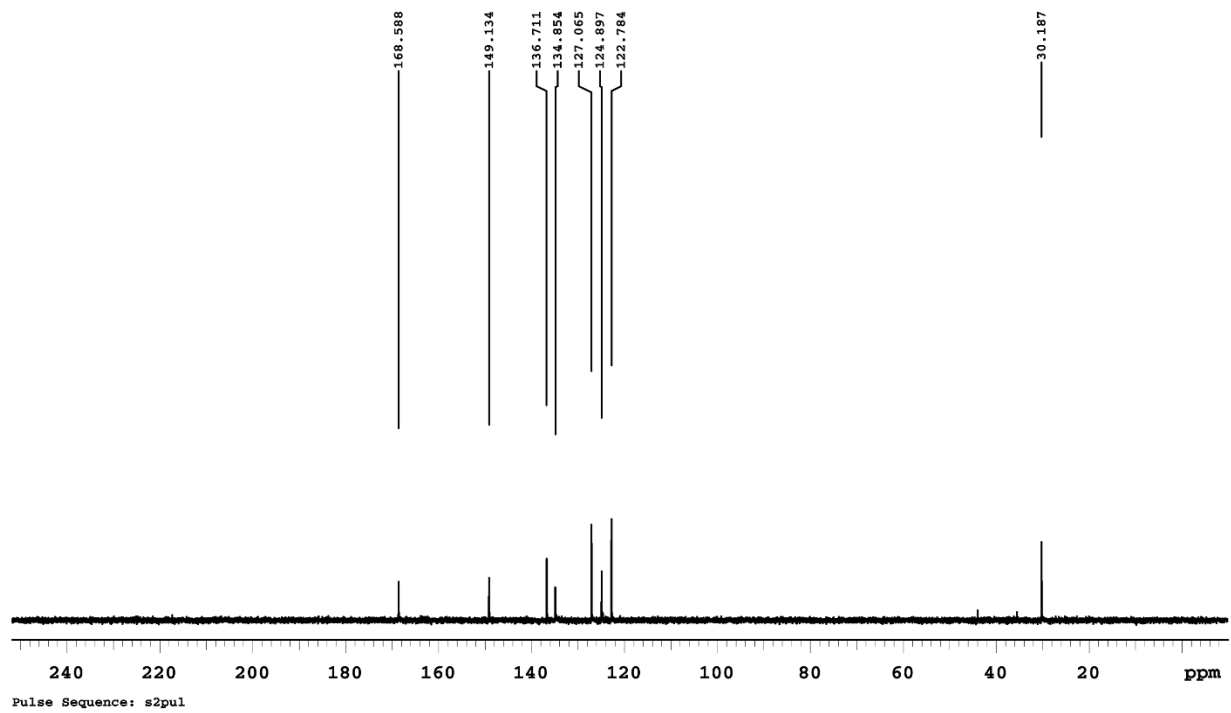^{1}H-NMR of 3,3′-bis(sulfonato)-4,4′-bis(amino)azobenzene (**6**)
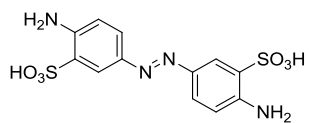
399.985 MHz H1 1D in d2o (ref. to external acetone @ 2.225 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe
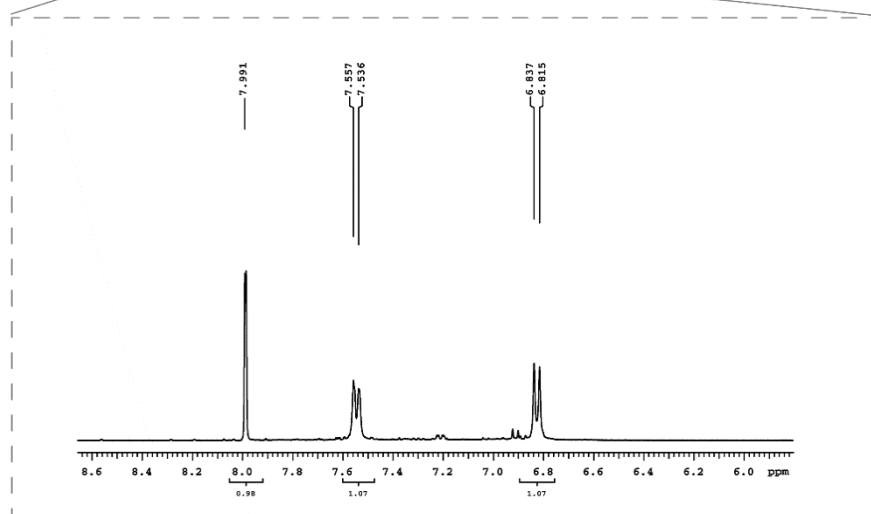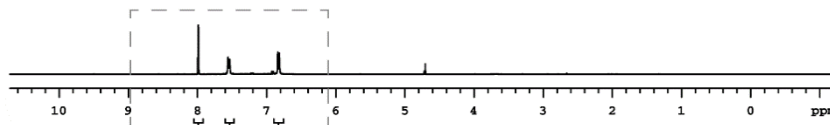
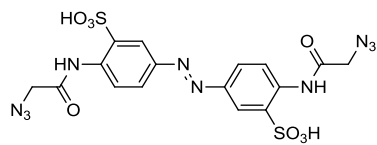¹³C-NMR of 3,3′-bis(sulfonato)-4,4′-bis(amino)azobenzene (**6**)

0.1% TFA in D$_2$O was used as solvent
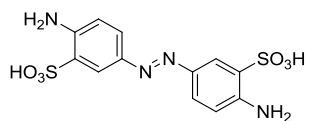
125.691 MHz C13[H1] 1D in d2o (ref. to external acetone @ 31.07 ppm), temp 27.7 C -> actual temp = 27.0 C, colddual probe
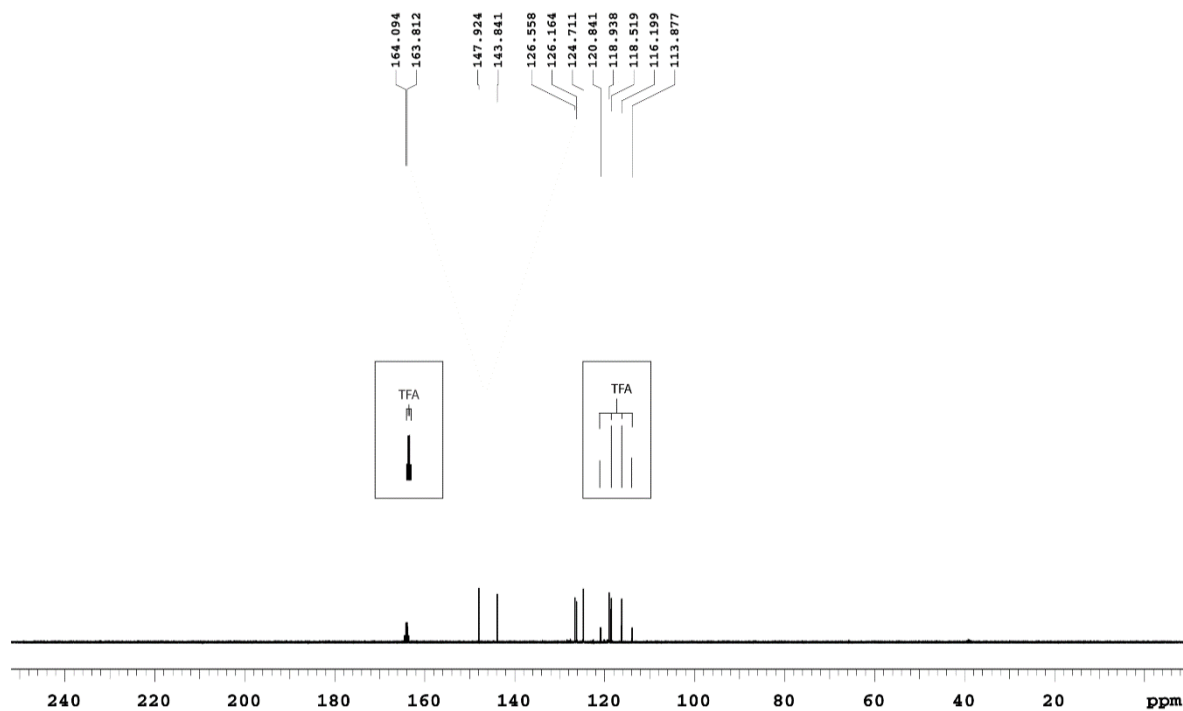
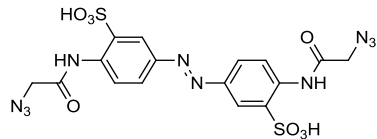# ^{1}H-NMR of 2-(acetylamino)benzenesulfonic acid (**7**)
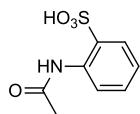
399.986 MHz H1 1D in dmso (ref. to DMSO @ 2.49 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe
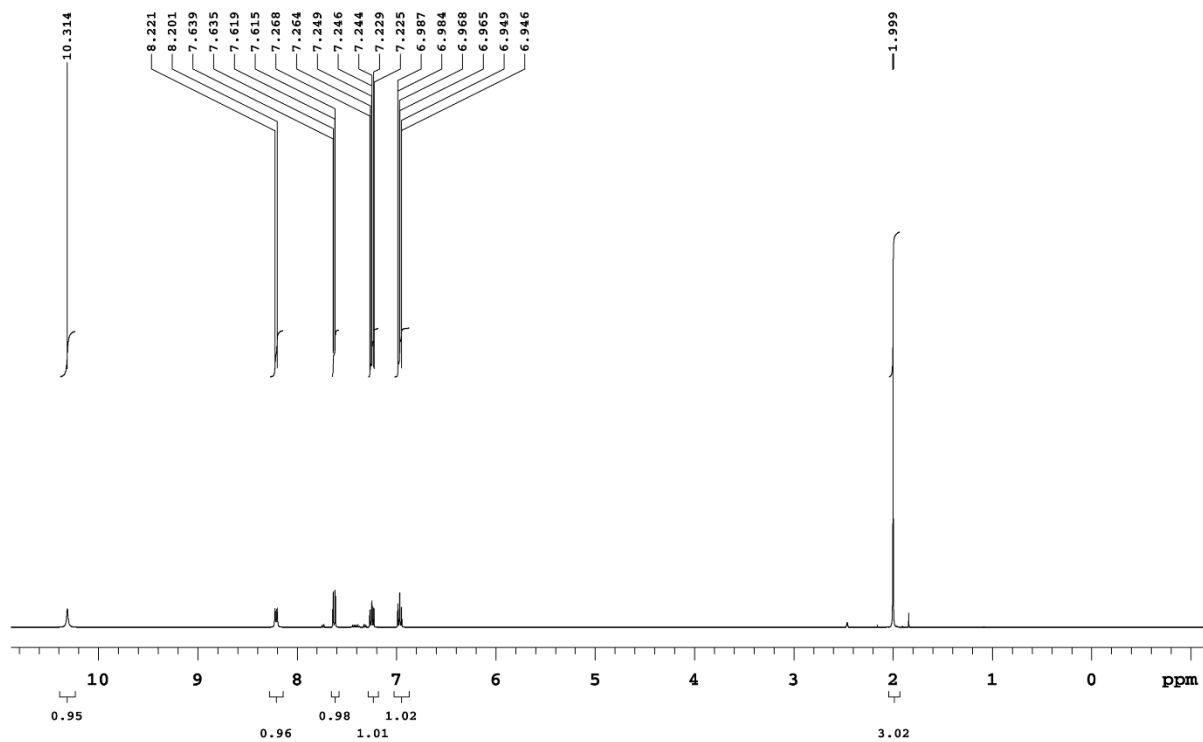
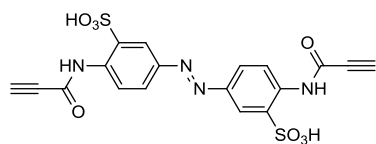# ¹H-NMR of 2-(acetylamino)benzenesulfonic acid (**7**)
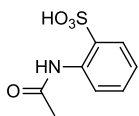
100.588 MHz C13[H1] 1D in dmso (ref. to DMSO @ 39.5 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe

Pulse Sequence: s2pul

^{1}H-NMR of 2-(1,3-dihydro-2H-isoindol-2-yl)benzenesulfonic acid (**9**)

399.986 MHz H1 1D in dmso (ref. to DMSO @ 2.49 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe
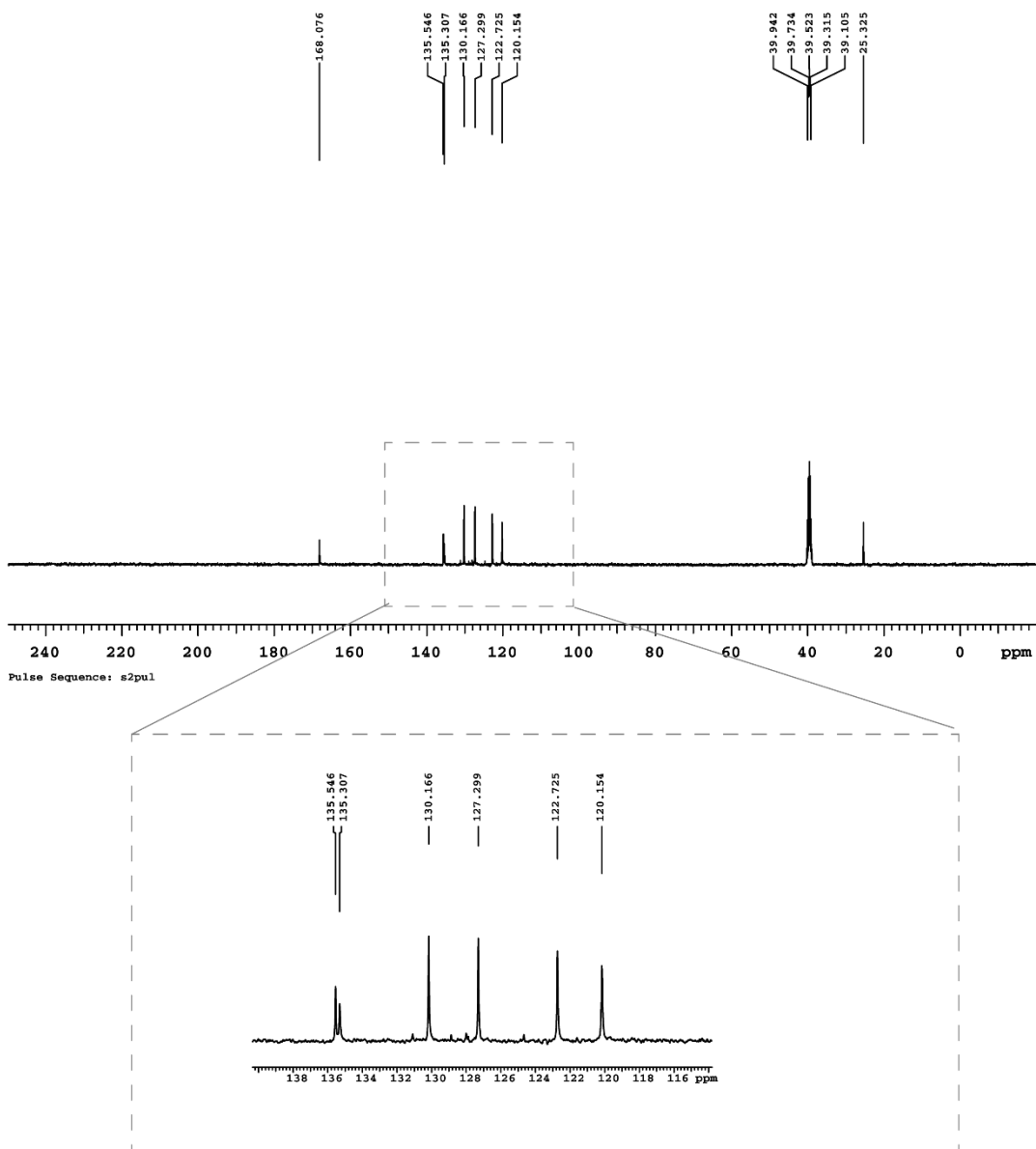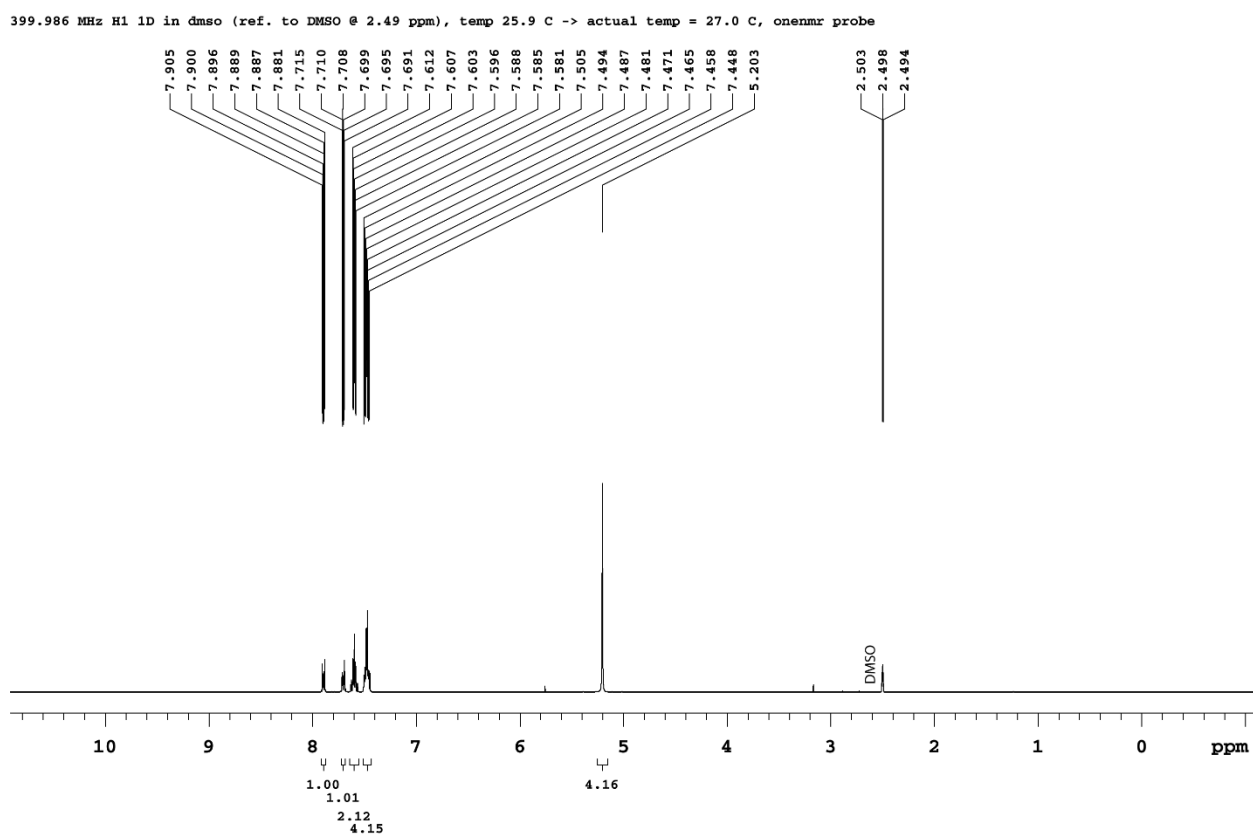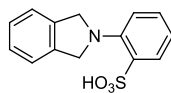
^{13}C-NMR of 2-(1,3-dihydro-2H-isoindol-2-yl)benzenesulfonic acid (**9**)

100.588 MHz C13[H1] 1D in dmso (ref. to DMSO @ 39.5 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe

139.325
137.610
134.384
132.415
130.383
129.278
128.249
123.438
121.573

63.437

40.585
40.375
40.164
39.954
39.744
39.540
39.326

240 220 200 180 160 140 120 100 80 60 40 20 0 ppm
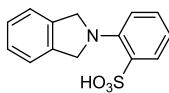
¹H-NMR of 3,3'-bis(sulfonato)-4,4'-bis(bromoacetamido)azobenzene (**11**)
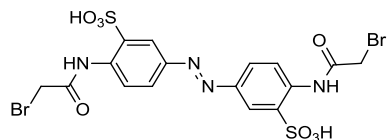
399.985 MHz H1 1D in d2o (ref. to external acetone @ 2.225 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe
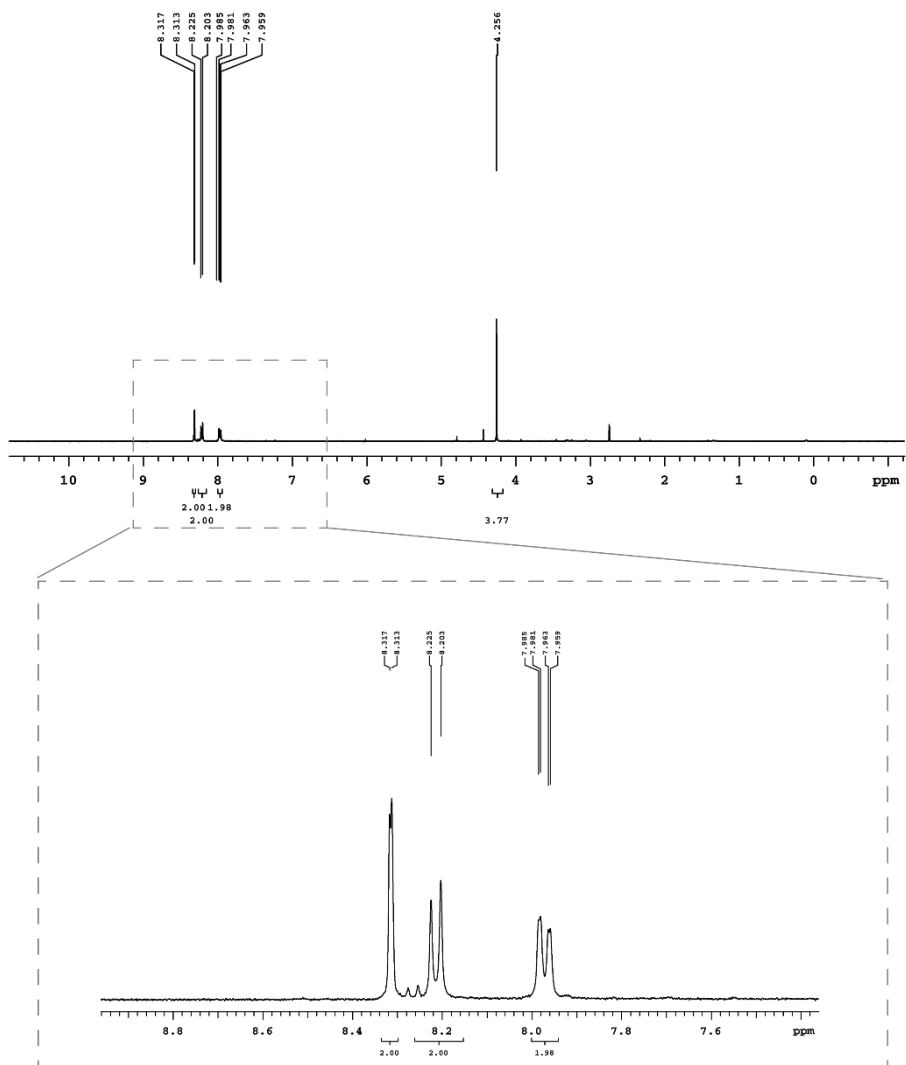
^{13}C-NMR of 3,3'-bis(sulfonato)-4,4'-bis(bromoacetamido)azobenzene (**11**)

125.691 MHz C13[H1] 1D in d2o (ref. to external acetone @ 31.07 ppm), temp 27.7 C -> actual temp = 27.0 C, colddual probe
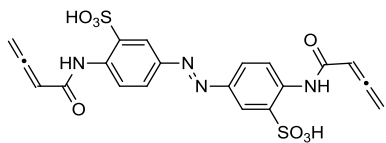
168.588 149.134 136.711 134.854 127.065 124.897 122.784 30.187

240   220   200   180   160   140   120   100   80   60   40   20   ppm

Pulse Sequence: s2pul

^{1}H-NMR of 3,3'-bis(sulfonato)-4,4'-bis(azidoacetamido)azobenzene (12)

399.986 MHz H1 1D in dmso (ref. to DMSO @ 2.49 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe
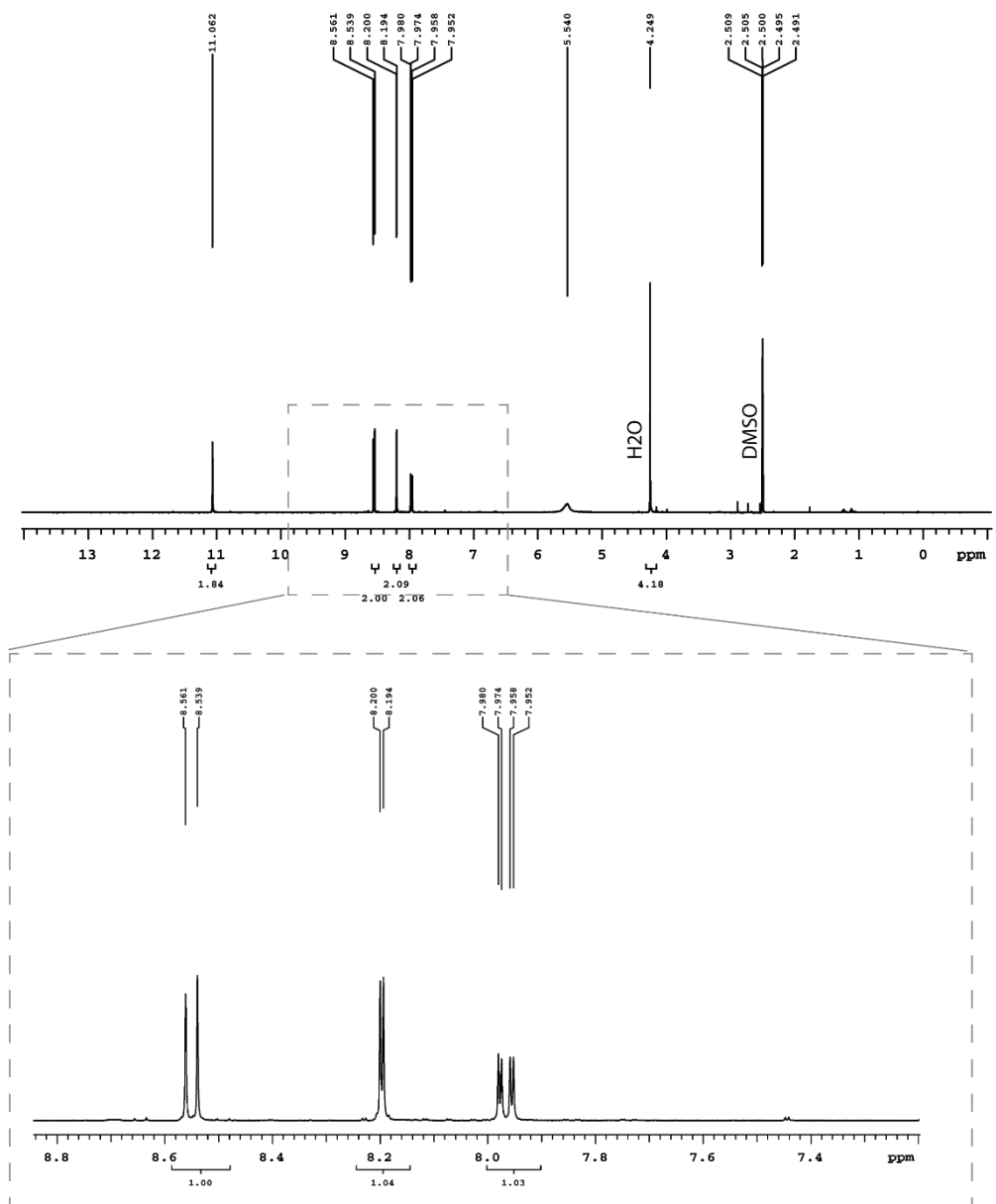
^{13}C-NMR of 3,3'-bis(sulfonato)-4,4'-bis(azidoacetamido)azobenzene (**12**)

125.691 MHz C13[H1] 1D in dmso (ref. to DMSO @ 39.5 ppm), temp 27.7 C -> actual temp = 27.0 C, colddual probe

166.089

146.787

136.781
136.442

126.005
120.167
119.418

52.351
39.994
39.920
39.826
39.752
39.660
39.582
39.492
39.325
39.159
38.991

DMSO

Pulse Sequence: s2pul

240 220 200 180 160 140 120 100 80 60 40 20 ppm
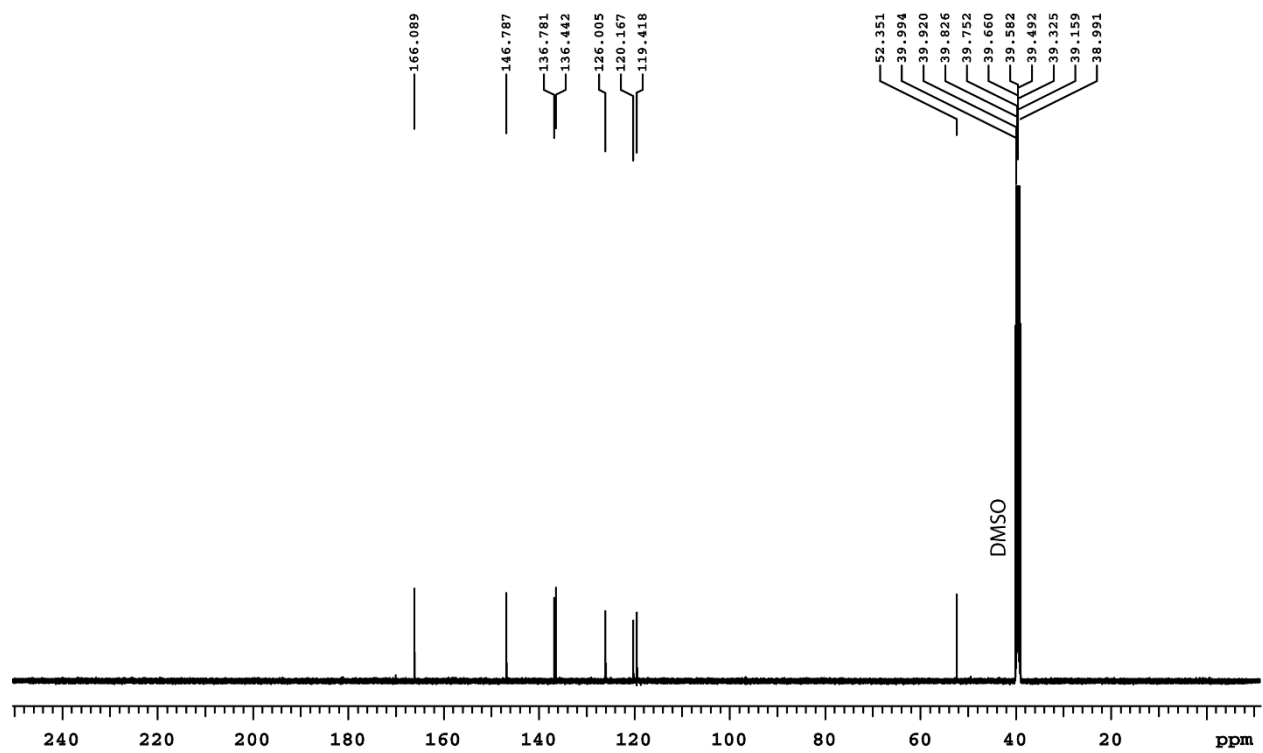
¹H-NMR of 3,3'-bis(sulfonato)-4,4'-bis(prop-2-ynoylamido)azobenzene (**13**)

399.986 MHz H1 1D in dmso (ref. to DMSO @ 2.49 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe
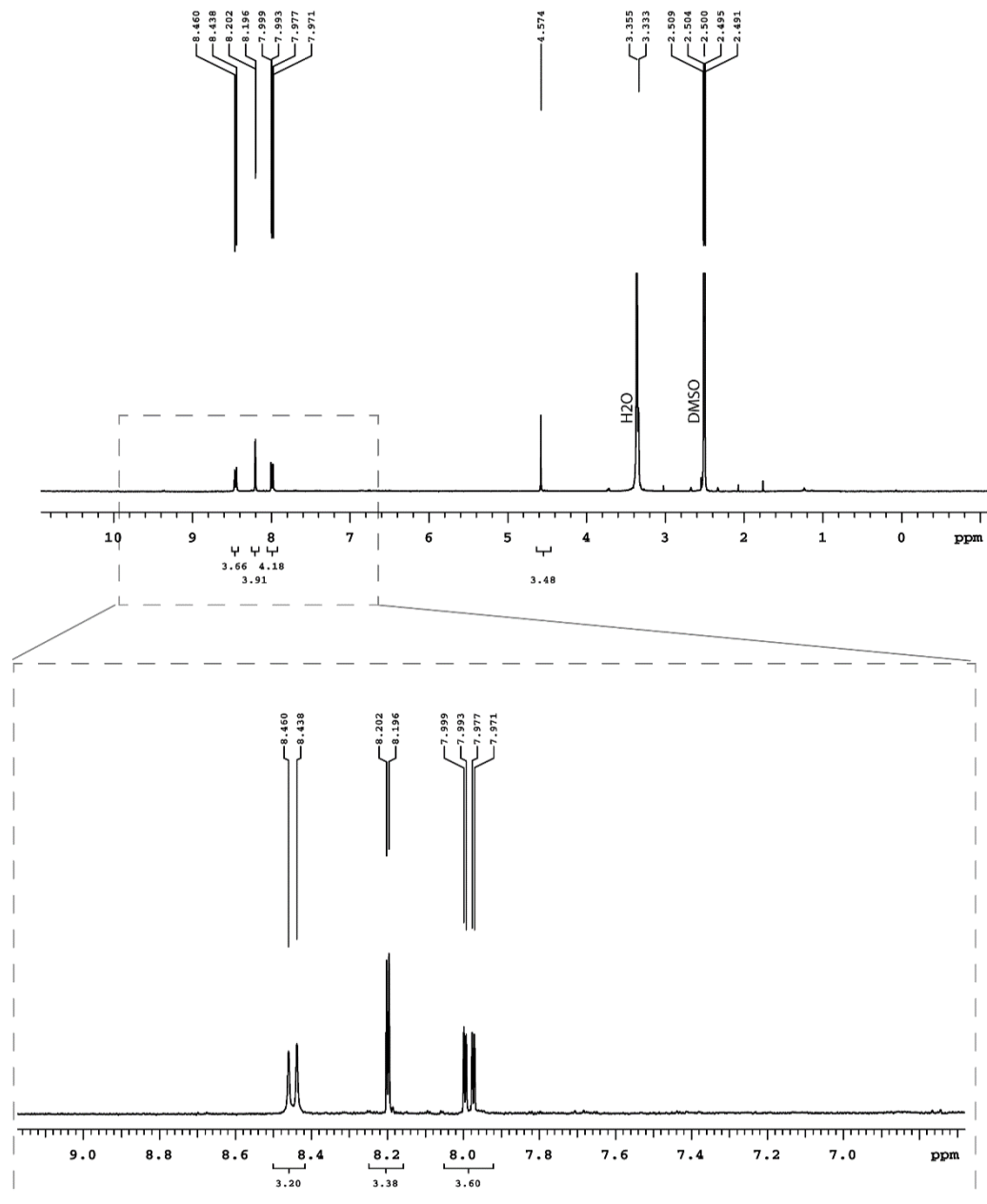
^{13}C-NMR of 3,3'-bis(sulfonato)-4,4'-bis(prop-2-ynoylamido)azobenzene (**13**)

125.691 MHz C13[H1] 1D in d2o (ref. to external acetone @ 31.07 ppm), temp 27.7 C -> actual temp = 27.0 C, colddual probe

152.778
149.254
136.194
134.856
127.055
125.208
122.930
79.086
77.285
39.712

DMSO

240 220 200 180 160 140 120 100 80 60 40 20 ppm

Pulse Sequence: s2pul
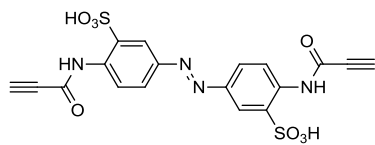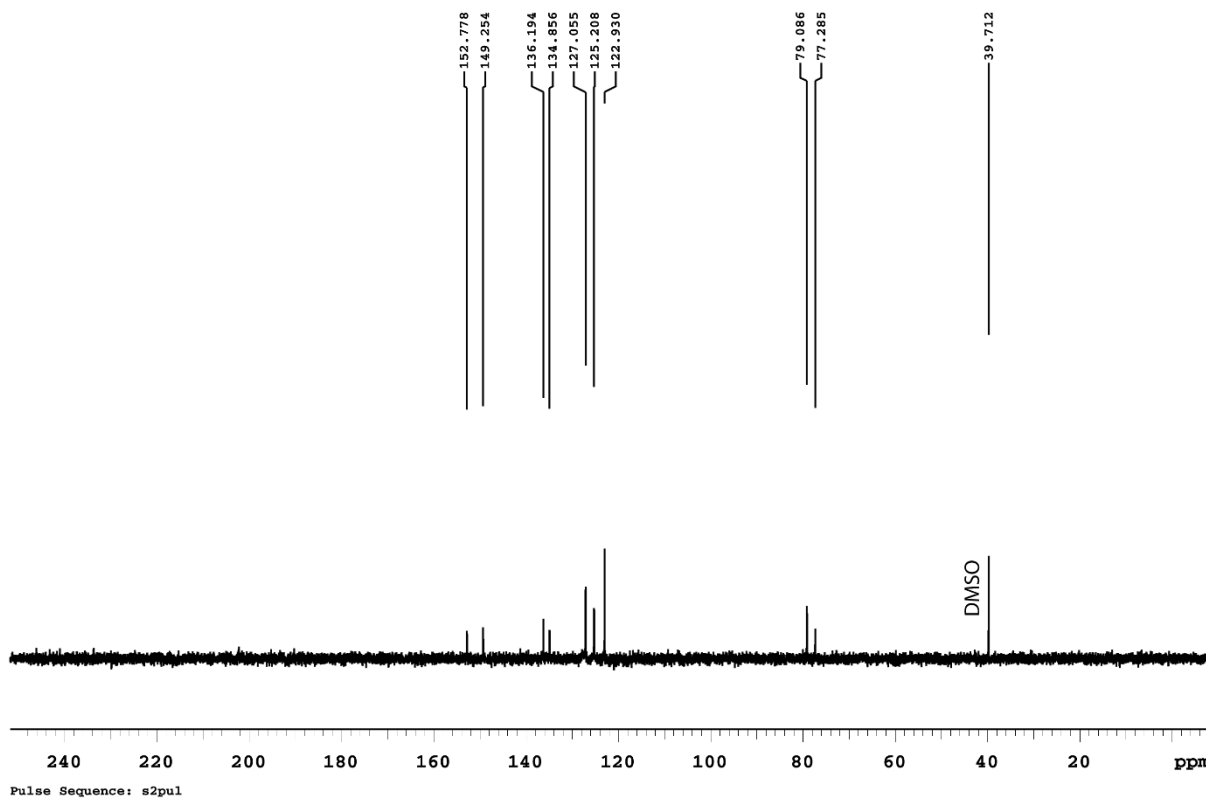
H-NMR of 3,3'-bis(sulfonato)-4,4'-bis(buta-2,3-dienoylamido)azobenzene (**14**)

399.795 MHz H1 1D in d2o (ref. to external acetone @ 2.225 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe
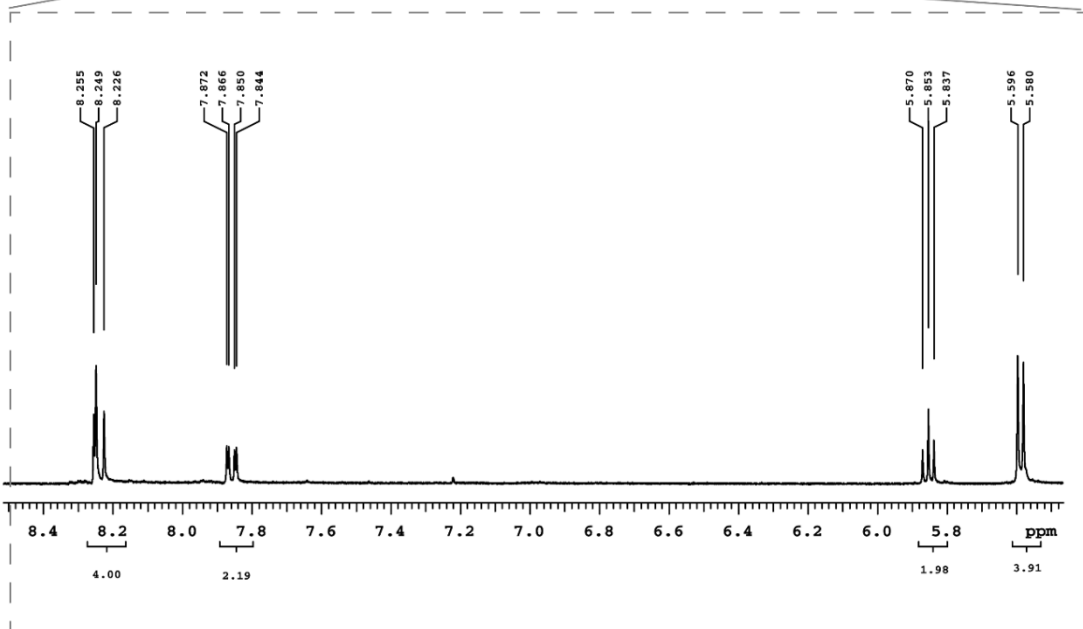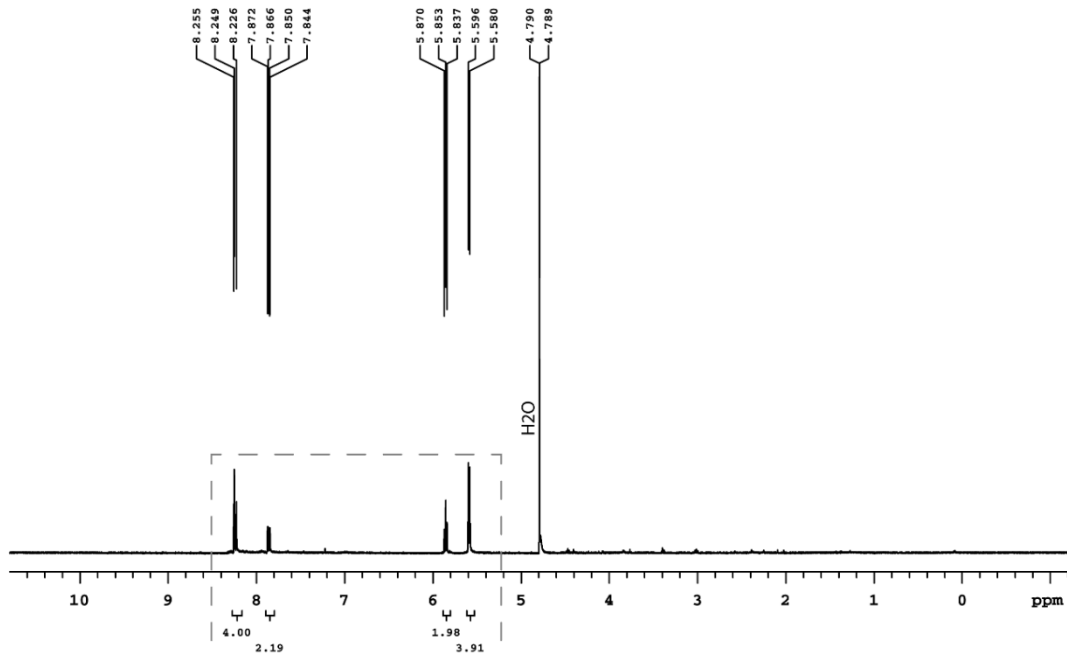
C-NMR of 3,3'-bis(sulfonato)-4,4'-bis(buta-2,3-dienoylamido)azobenzene (**14**)
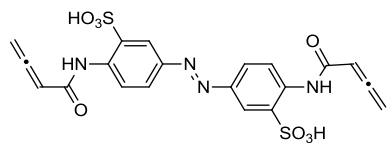
125.691 MHz C13[H1] 1D in d2o (ref. to external acetone @ 31.07 ppm), temp 27.7 C -> actual temp = 27.0 C, colddual probe

Pulse Sequence: s2pul
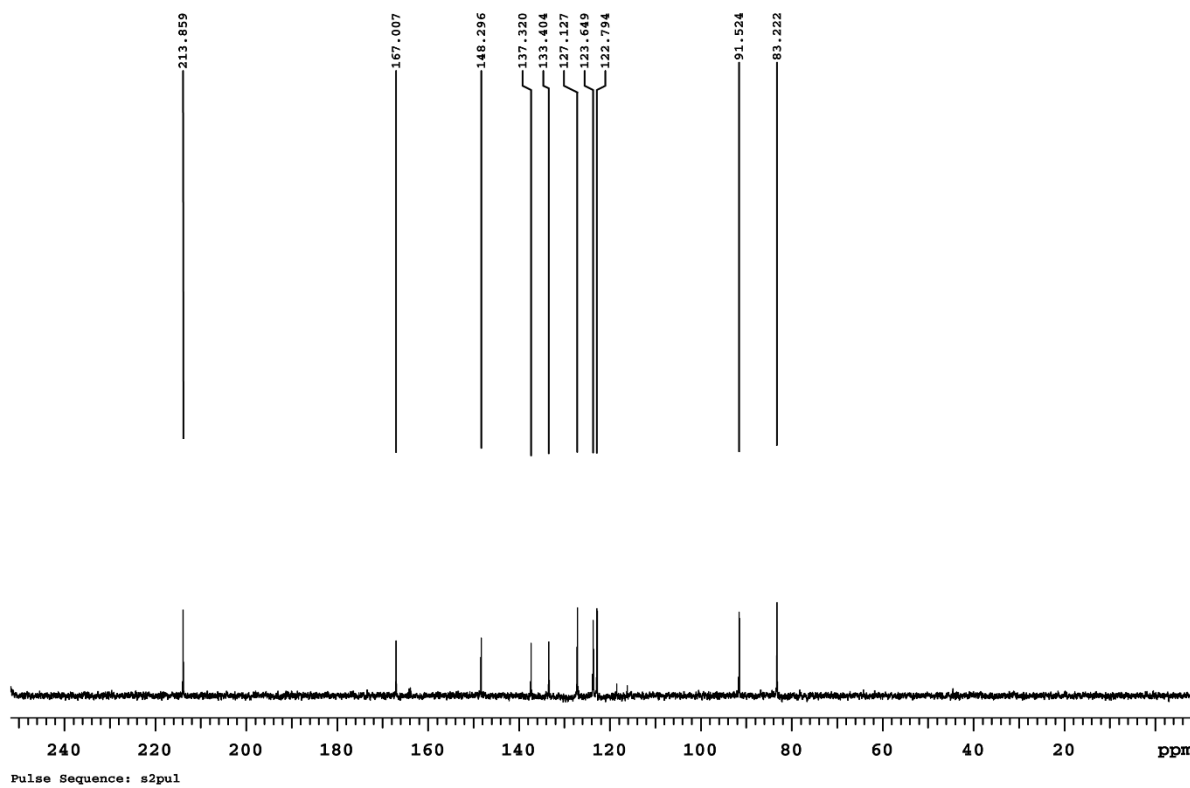
^{1}H-NMR of **1a** (asterisk shows residual EtOAc)

399.794 MHz H1 1D in cdcl3 (ref. to CDCl3 @ 7.26 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe
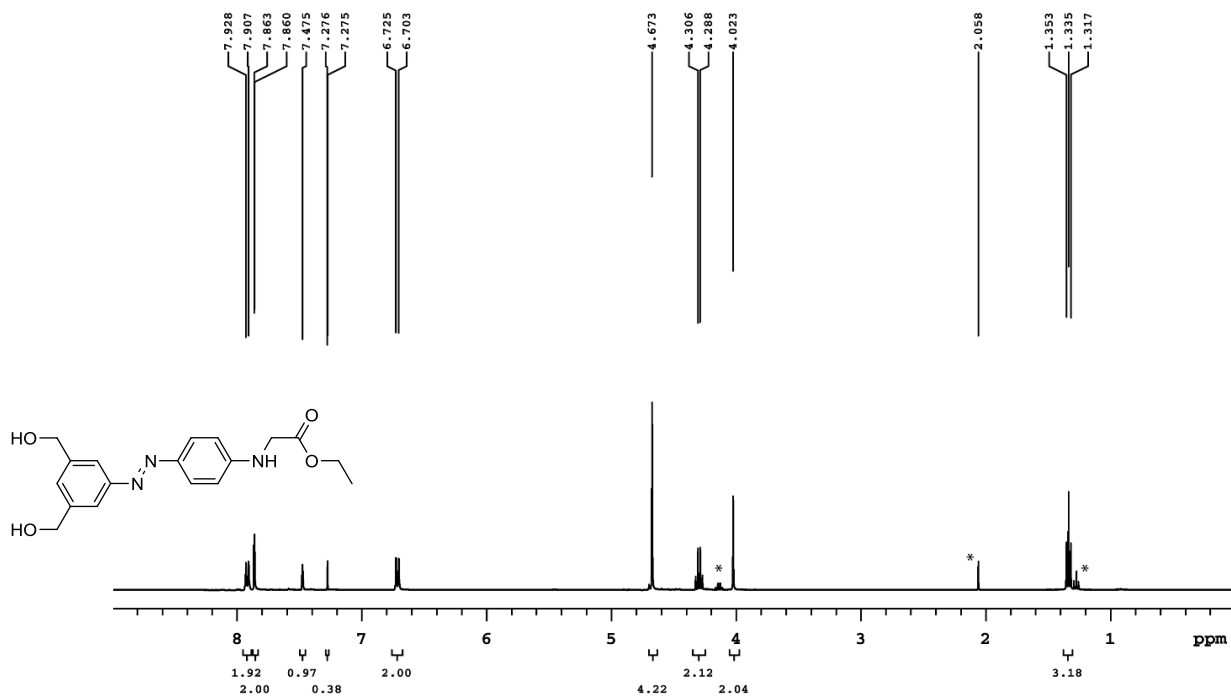
^{13}C-NMR of **1a**

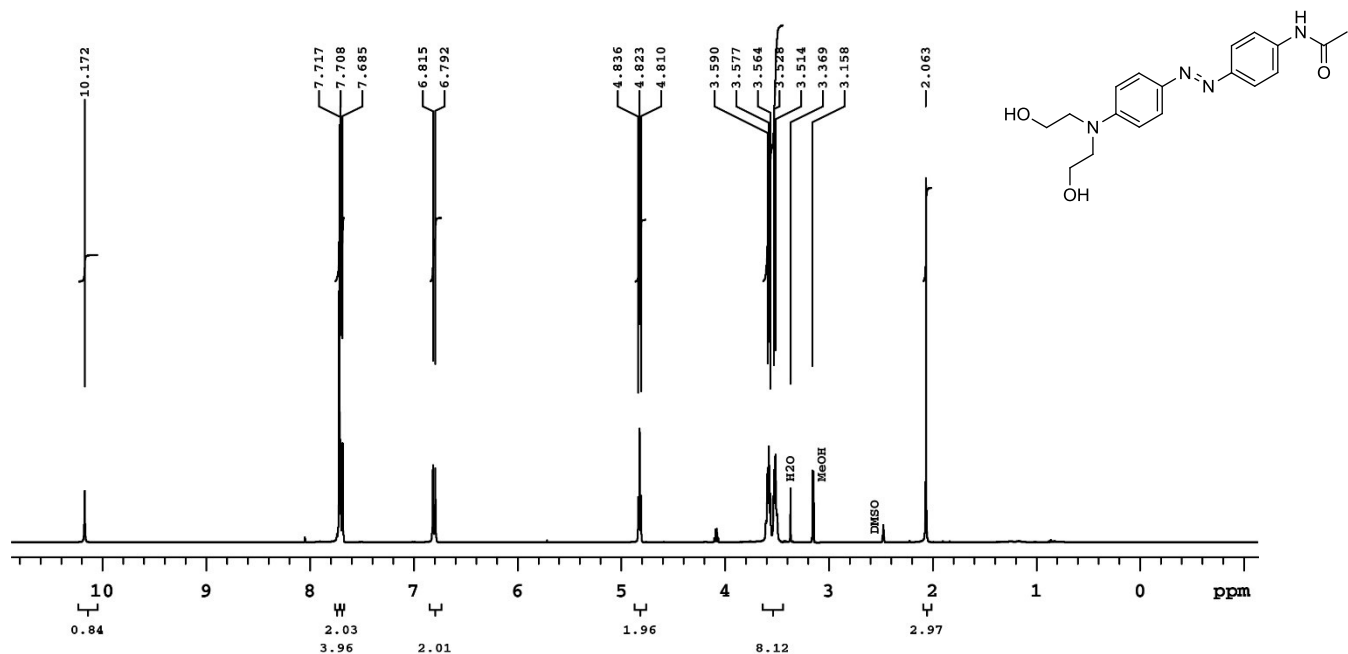100.539 MHz C13[H1] 1D in cdcl3 (ref. to CDCl3 @ 77.06 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe
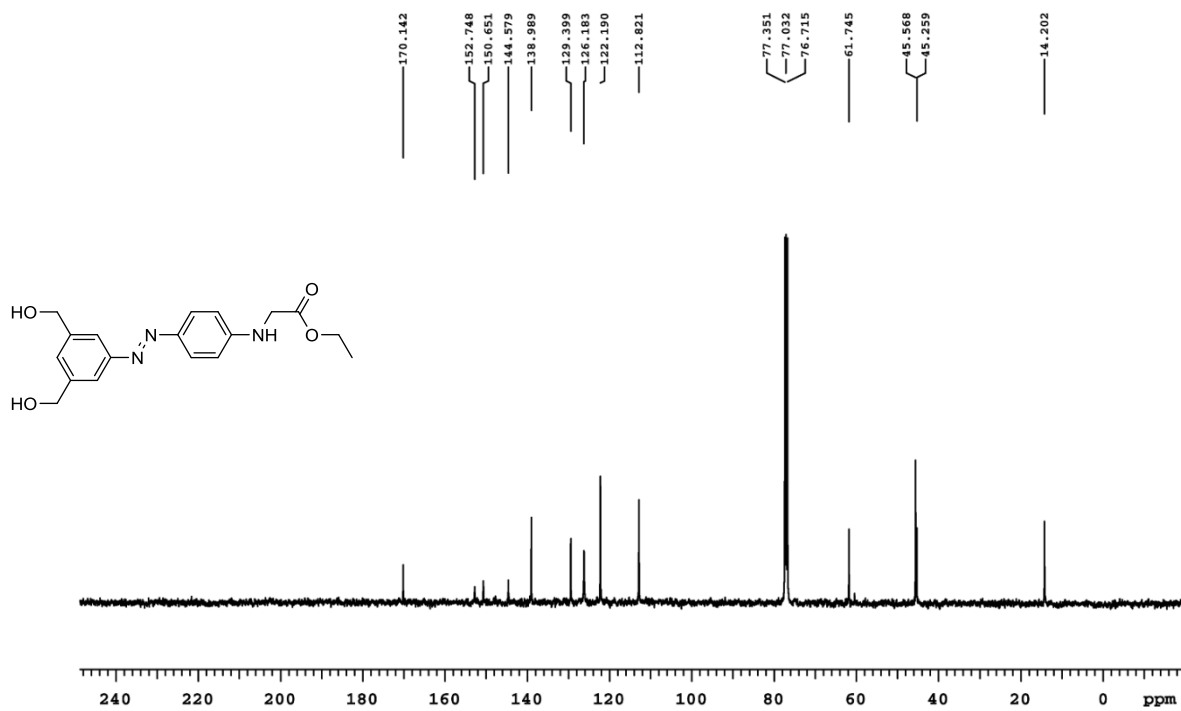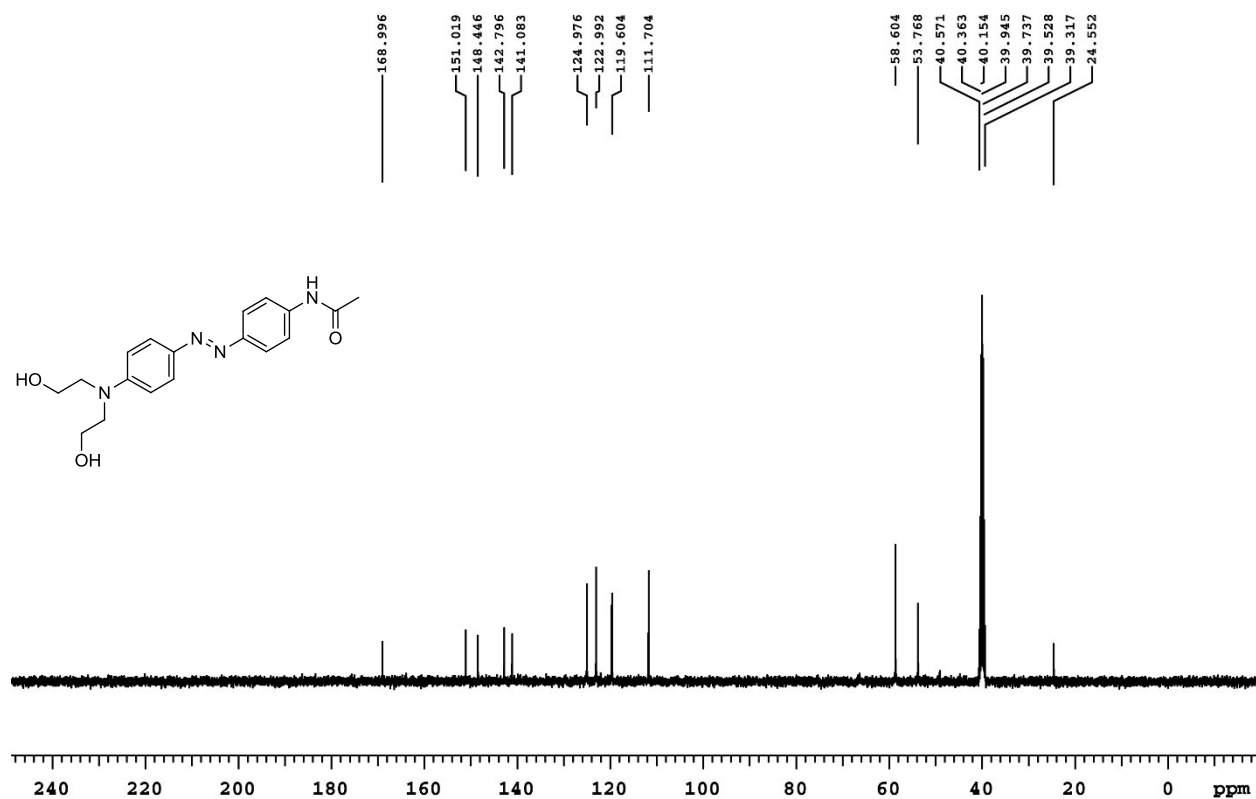
^{1}H-NMR of **1b**

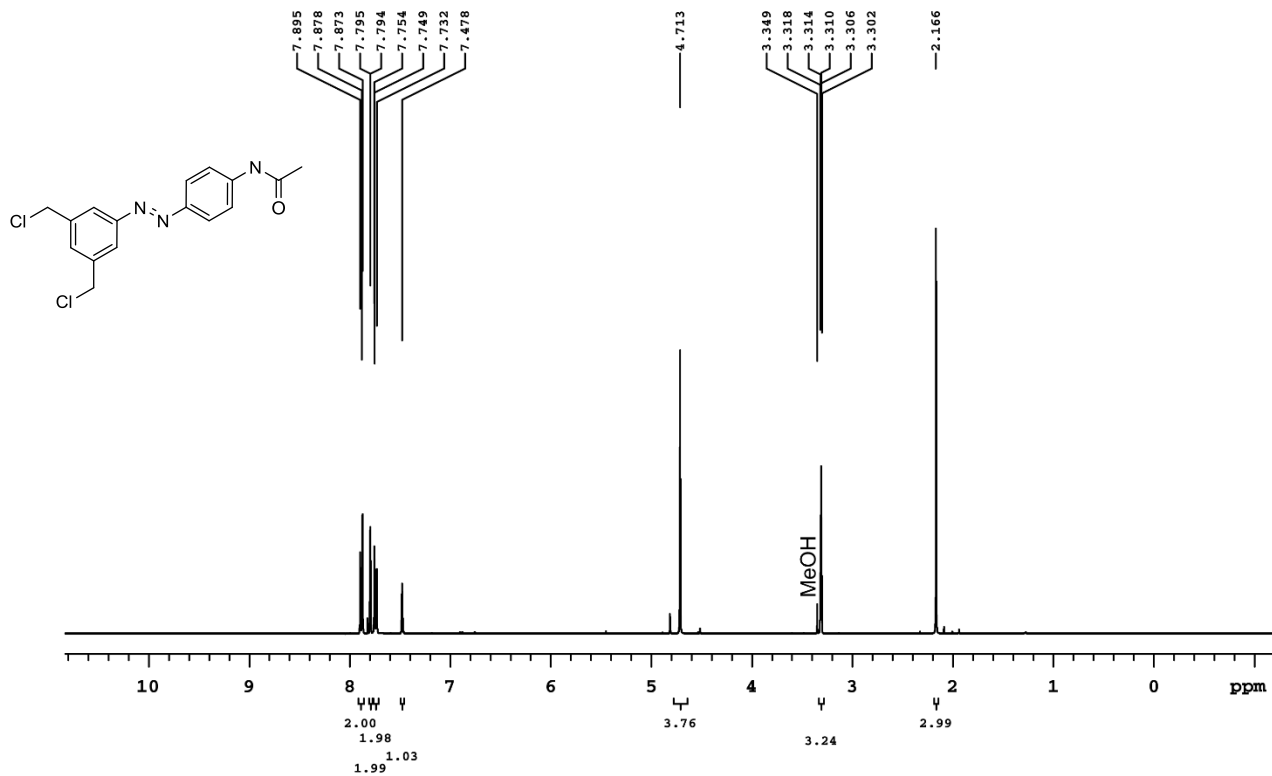399.796 MHz H1 1D in dmso (ref. to DMSO @ 2.49 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe

10.172

7.717
7.708
7.685

6.815
6.792

4.836
4.823
4.810

3.590
3.577
3.564
3.528
3.514
3.369
3.158

2.063

H2O
MeOH
DMSO

| 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | ppm |

0.84   2.03   1.96   2.97
        3.96 2.01   8.12

^{13}C-NMR of **1b**

100.540 MHz C13[H1] 1D in dmso (ref. to DMSO @ 39.5 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe

168.996

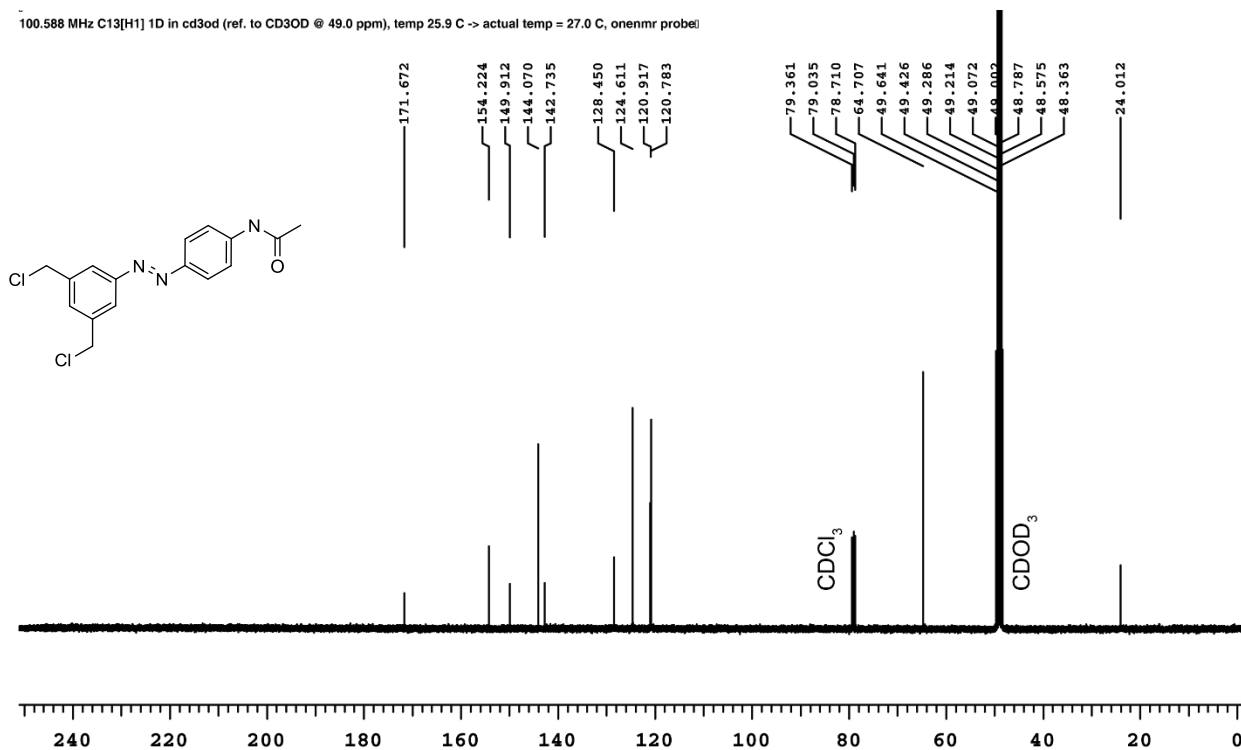151.019
148.446
142.796
141.083

124.976
122.992
119.604
111.704

58.604
53.768
40.571
40.363
40.154
39.945
39.737
39.528
39.317
24.552

| 240 | 220 | 200 | 180 | 160 | 140 | 120 | 100 | 80 | 60 | 40 | 20 | 0 | ppm |

238

^{1}H-NMR of **1c**

399.986 MHz H1 1D in cd3od (ref. to CD3OD @ 3.30 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe

7.895
7.878
7.873
7.795
7.794
7.754
7.749
7.732
7.478

4.713

3.349
3.318
3.314
3.310
3.306
3.302

2.166

MeOH

10   9   8   7   6   5   4   3   2   1   0   ppm

2.00
1.98
1.03
1.99

3.76

3.24

2.99

^{13}C-NMR of **1c**

100.588 MHz C13[H1] 1D in cd3od (ref. to CD3OD @ 49.0 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe

171.672

154.224
149.912
144.070
142.735

128.450
124.611
120.917
120.783

79.361
79.035
78.710
64.707
49.641
49.426
49.286
49.214
49.072
48.787
48.575
48.363

24.012

CDCl$_3$

CDOD$_3$

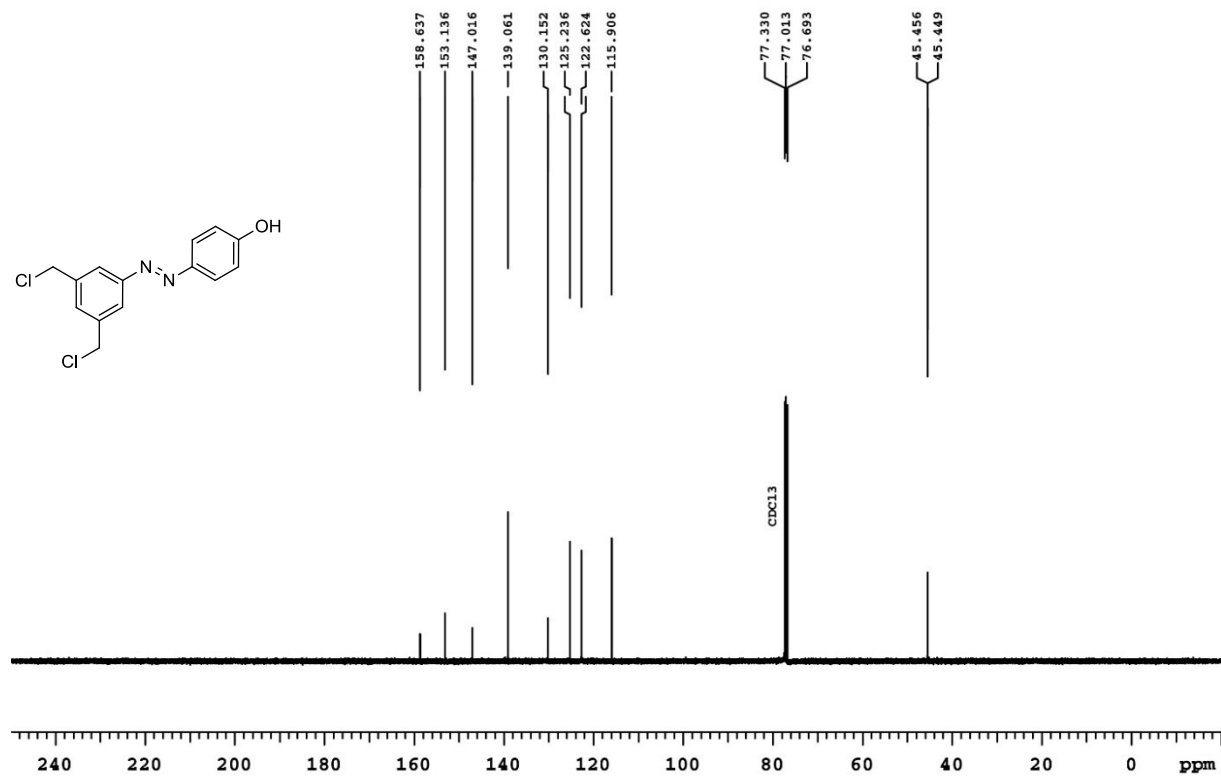240   220   200   180   160   140   120   100   80   60   40   20   0

239

399.794 MHz H1 1D in cdcl3 (ref. to CDCl3 @ 7.26 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe
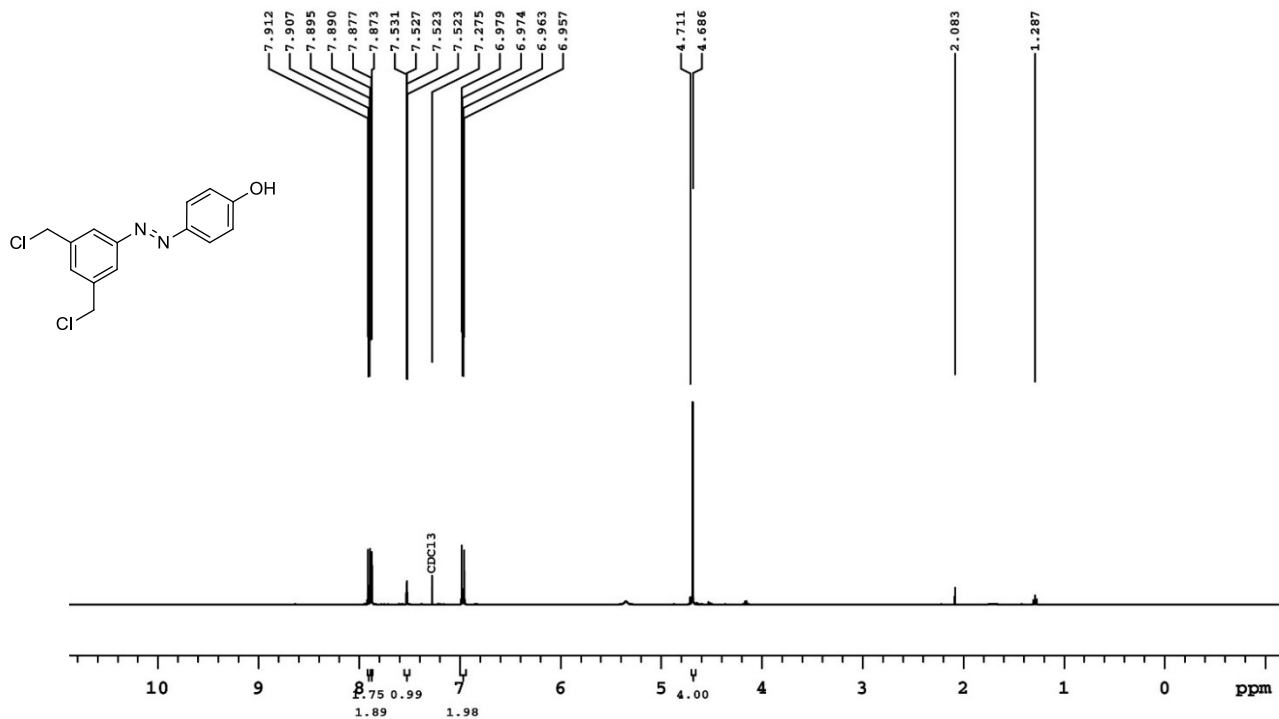
^{13}C-NMR of **1dA**

100.587 MHz C13[H1] 1D in cdcl3 (ref. to CDCl3 @ 77.06 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe

<superscript>1</superscript>H-NMR of **1d**

399.794 MHz H1 1D in cdcl3 (ref. to CDC13 @ 7.26 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe
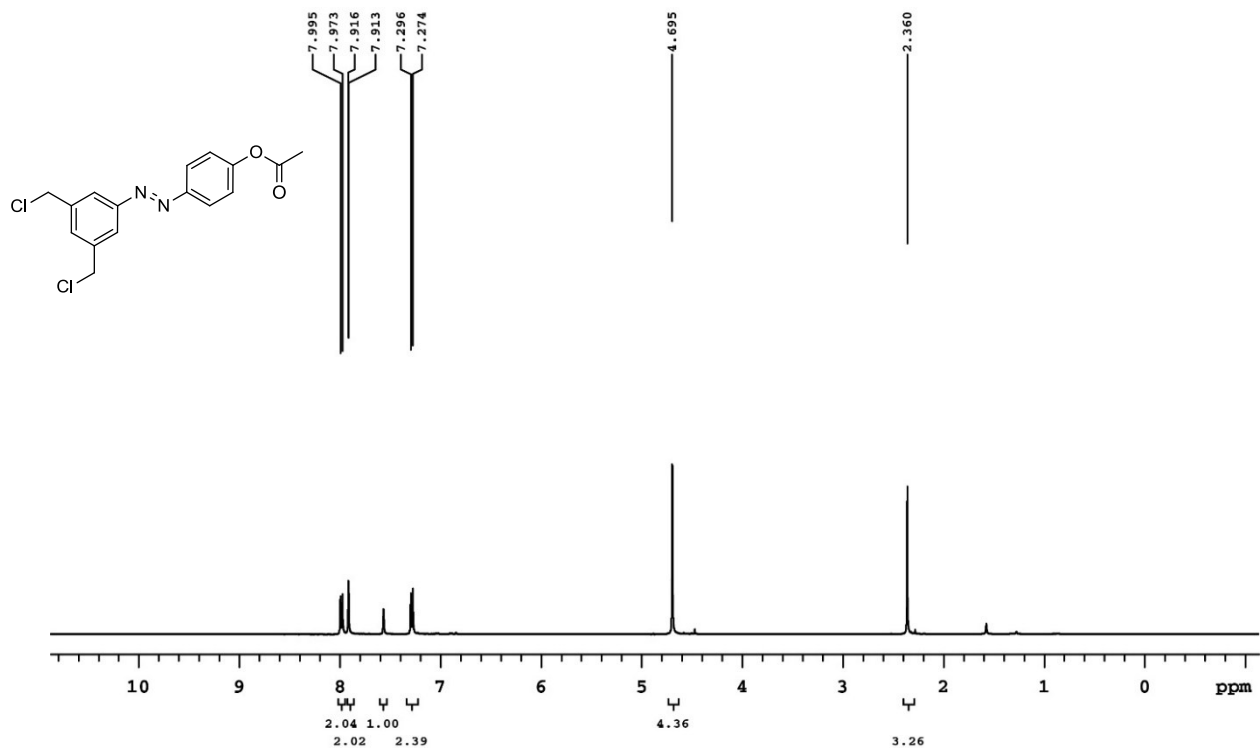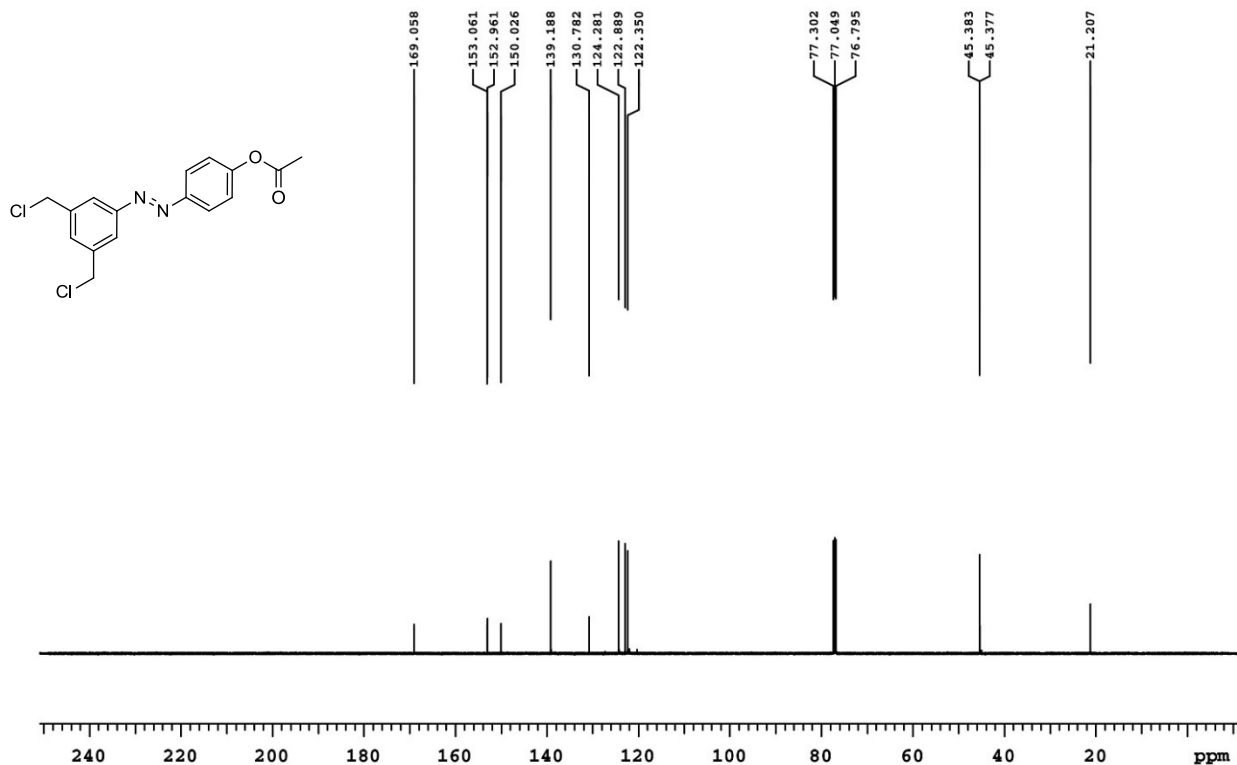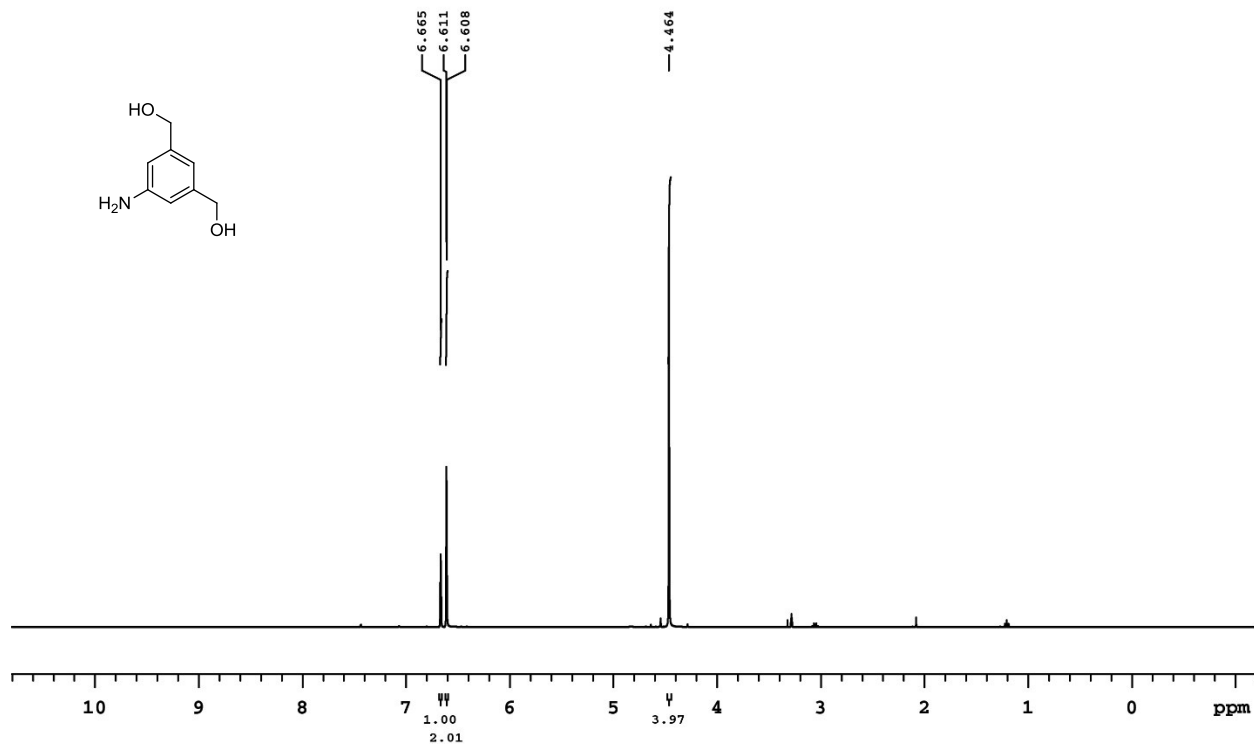

<superscript>13</superscript>C-NMR of **1d**

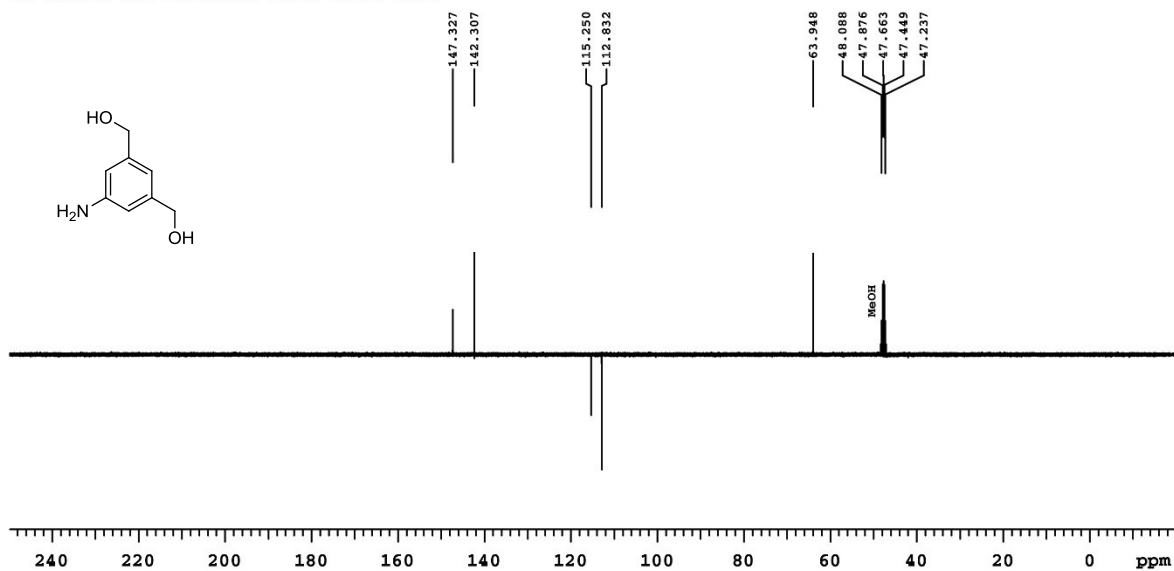125.691 MHz C13[H1] 1D in cdcl3 (ref. to CDC13 @ 77.06 ppm), temp 27.7 C -> actual temp = 27.0 C, colddual probe

¹H-NMR of **2b**

399.986 MHz H1 1D in cd3od (ref. to CD3OD @ 3.30 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe
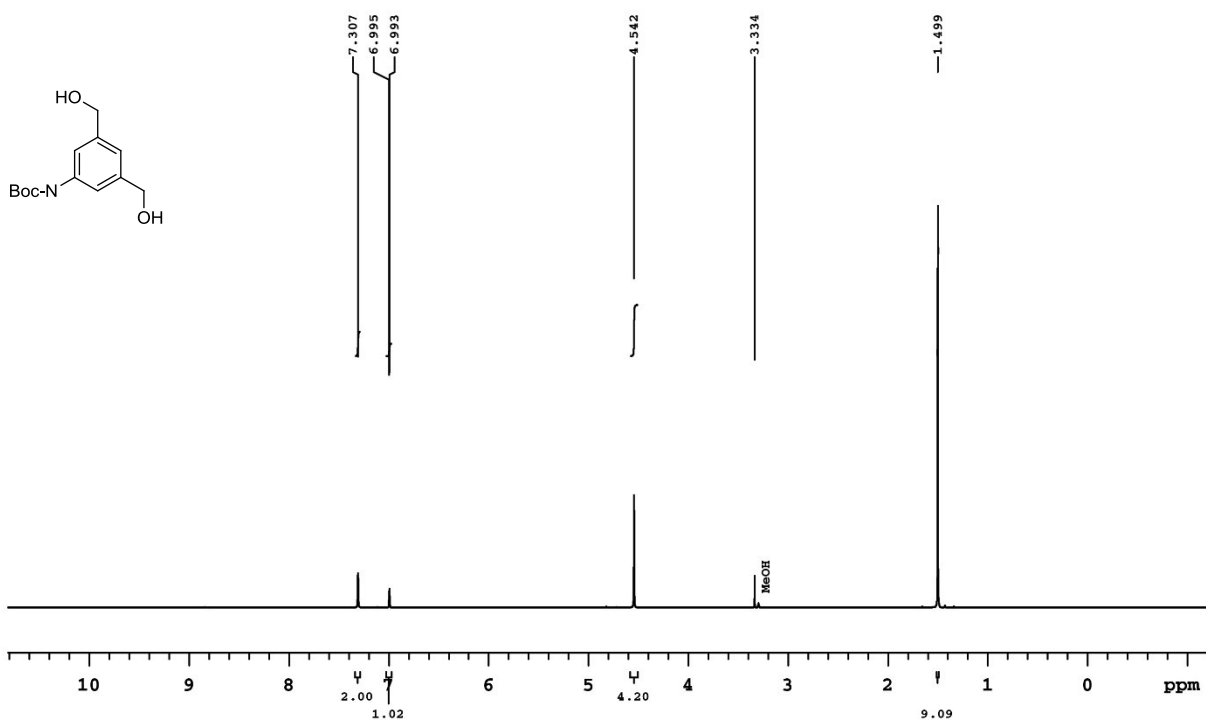
6.665
6.611
6.608

4.464

10   9   8   7   6   5   4   3   2   1   0   ppm

1.00
2.01

3.97

¹³C-NMR of **2b**

100.588 MHz C13[H1] APT_ad in cd3od (ref. to CD3OD @ 49.0 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe
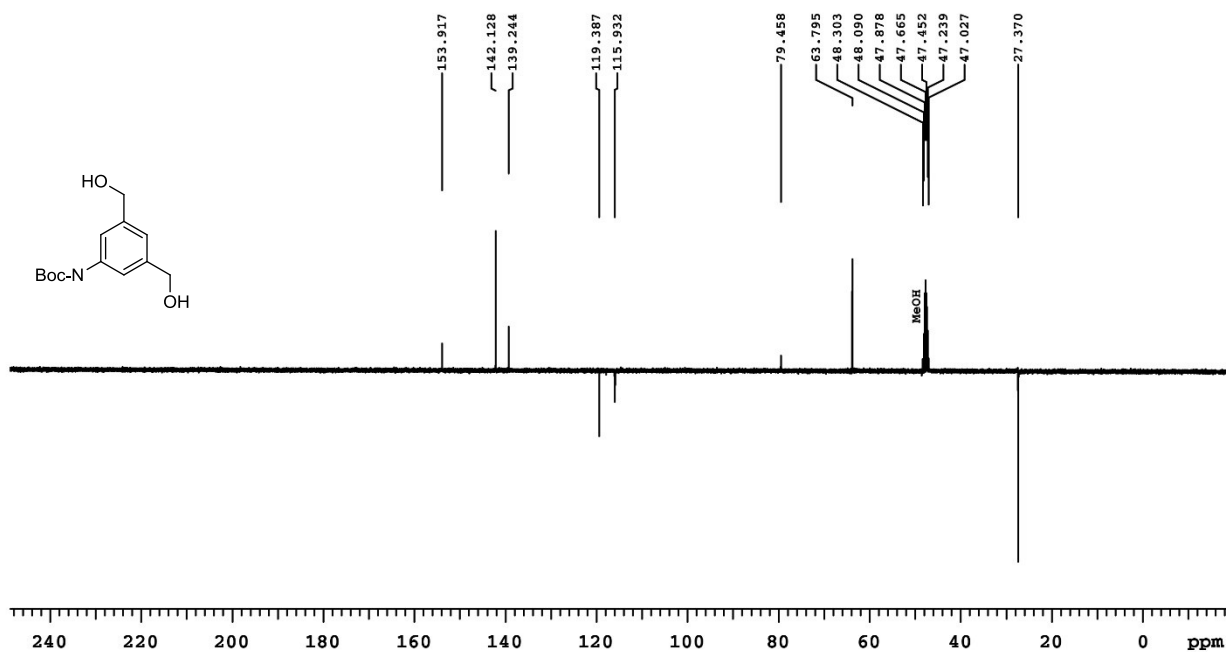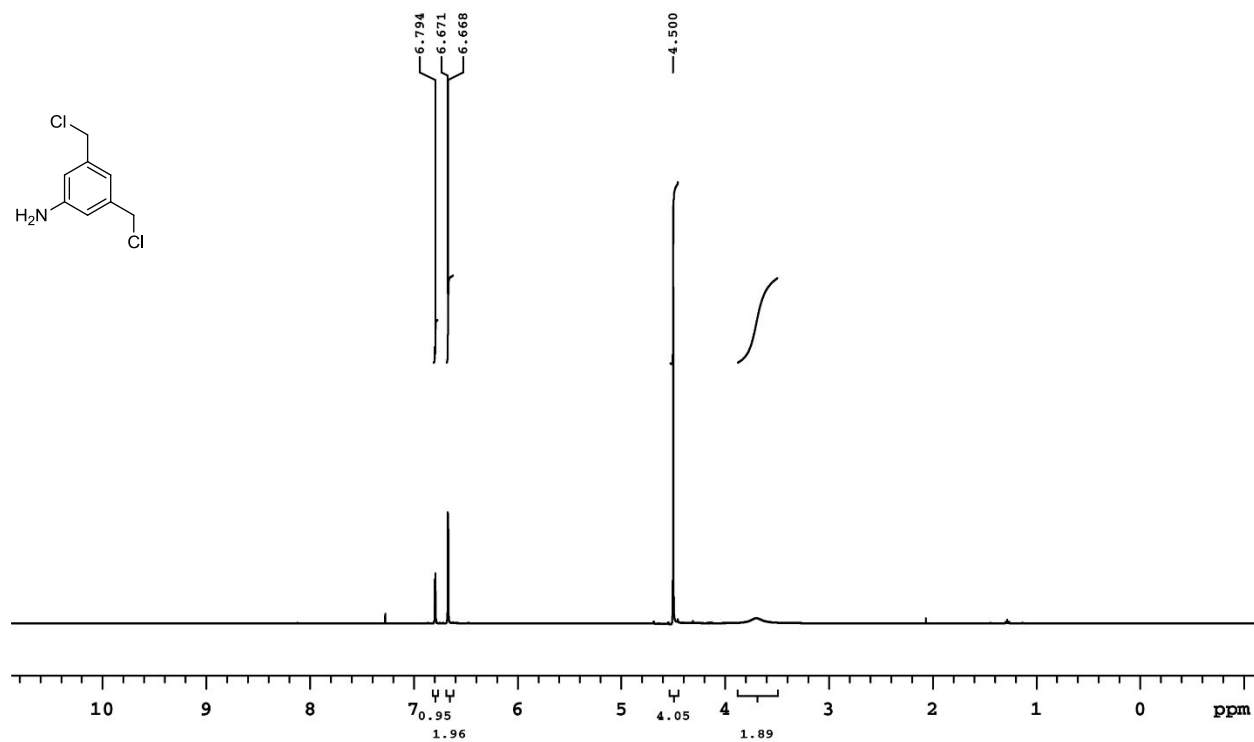C & CH2 same, CH & CH3 opposite side of solvent signal

147.327
142.307
115.250
112.832
63.948
48.088
47.876
47.663
47.449
47.237

MeOH

240   220   200   180   160   140   120   100   80   60   40   20   0   ppm

242

^{1}H-NMR of **2c**

7.307
6.995
6.993
4.542
3.334
1.499

MeOH

10   9   8   7   6   5   4   3   2   1   0   ppm

2.00
1.02
4.20
9.09

^{13}C-NMR of **2c**

153.917
142.128
139.244
119.387
115.932
79.458
63.795
48.303
48.090
47.878
47.665
47.452
47.239
47.027
27.370

MeOH

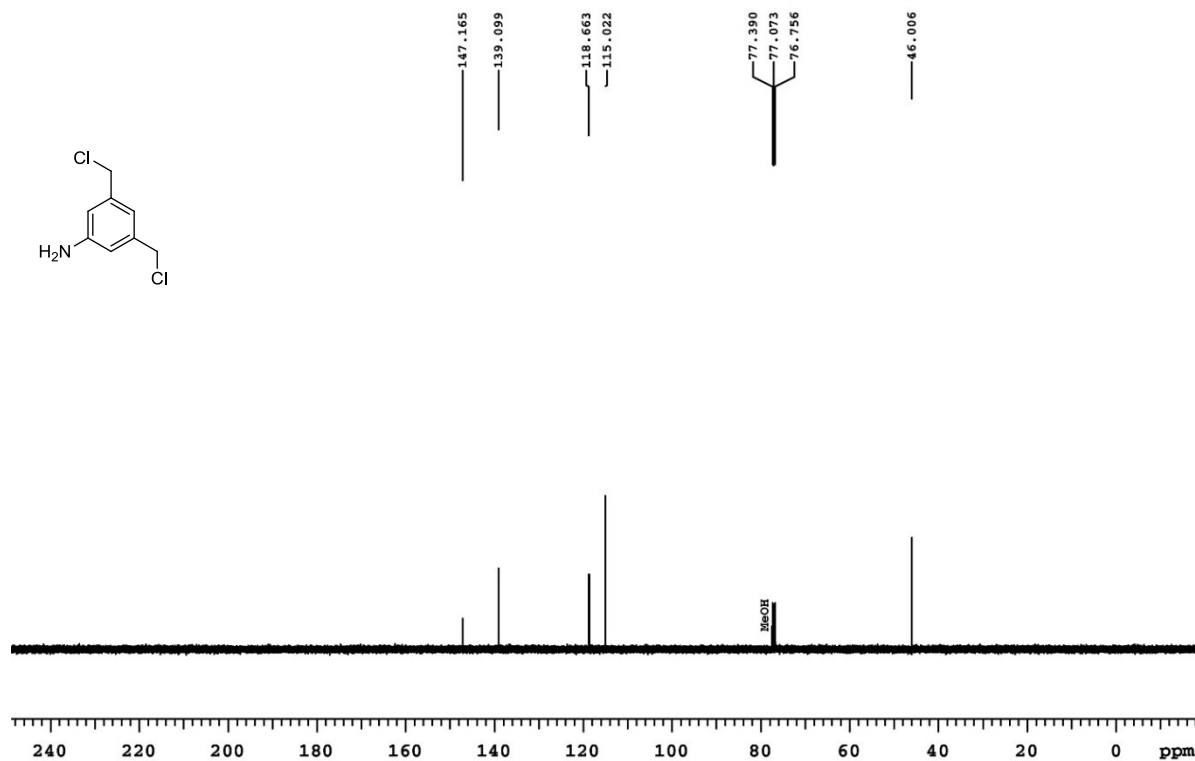240   220   200   180   160   140   120   100   80   60   40   20   0   ppm

243

¹H-NMR of 2d

¹H-NMR of **2d**

399.794 MHz H1 1D in cdcl3 (ref. to CDCl3 @ 7.26 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe

7.395
7.392
7.093
7.090
7.089
7.086
7.085
6.651

4.537

1.534

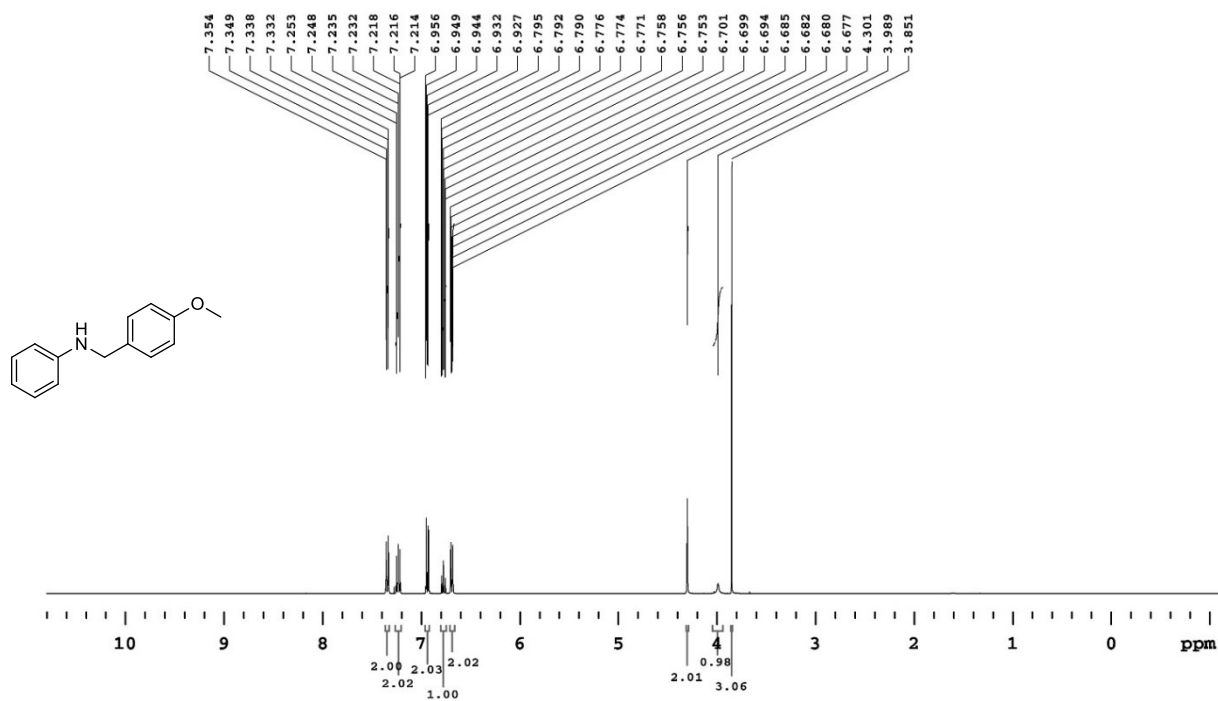2.00 1.00 0.94   4.07   9.23

10   9   8   7   6   5   4   3   2   1   0   ppm

¹³C-NMR of **2d**

100.539 MHz C13[H1] 1D in cdcl3 (ref. to CDCl3 @ 77.06 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe
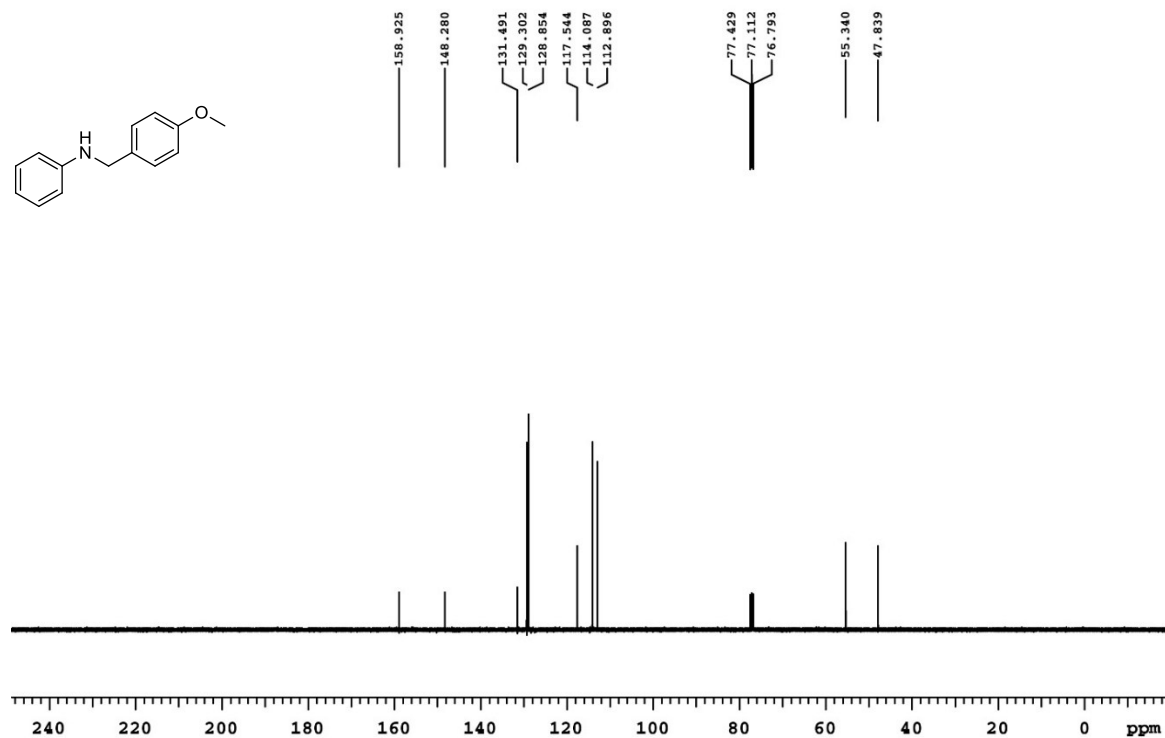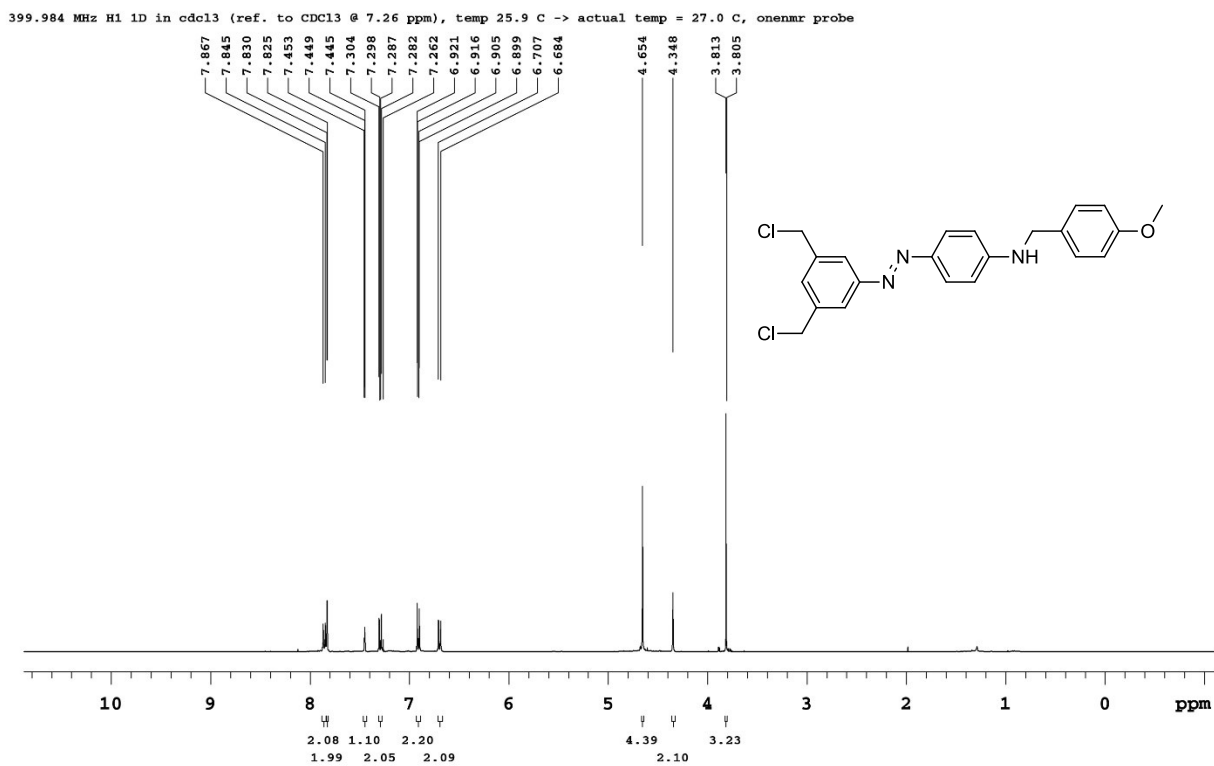
152.588
139.183
138.897
123.041
118.362
81.017
77.382
77.063
76.746
45.699
28.319

CHCl3

240 220 200 180 160 140 120 100 80 60 40 20 0 ppm

¹H-NMR of 2e

¹³C-NMR of 2e

245

^{1}H-NMR of **2f**

7.354, 7.349, 7.338, 7.332, 7.253, 7.248, 7.235, 7.232, 7.218, 7.216, 7.214, 6.956, 6.949, 6.944, 6.932, 6.927, 6.795, 6.792, 6.790, 6.776, 6.774, 6.771, 6.758, 6.756, 6.753, 6.701, 6.699, 6.694, 6.685, 6.682, 6.680, 6.677, 4.301, 3.989, 3.851

2.00 2.02 2.03 2.02 2.01 0.98 3.06
2.02 1.00

^{13}C-NMR of **2f**

158.925, 148.280, 131.491, 129.302, 128.854, 117.544, 114.087, 112.896, 77.429, 77.112, 76.793, 55.340, 47.839

¹H-NMR of **2g**

399.984 MHz H1 1D in cdcl3 (ref. to CDCl3 @ 7.26 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe

7.867
7.845
7.830
7.825
7.453
7.449
7.445
7.304
7.298
7.287
7.282
7.262
6.921
6.916
6.905
6.899
6.707
6.684

4.654
4.348

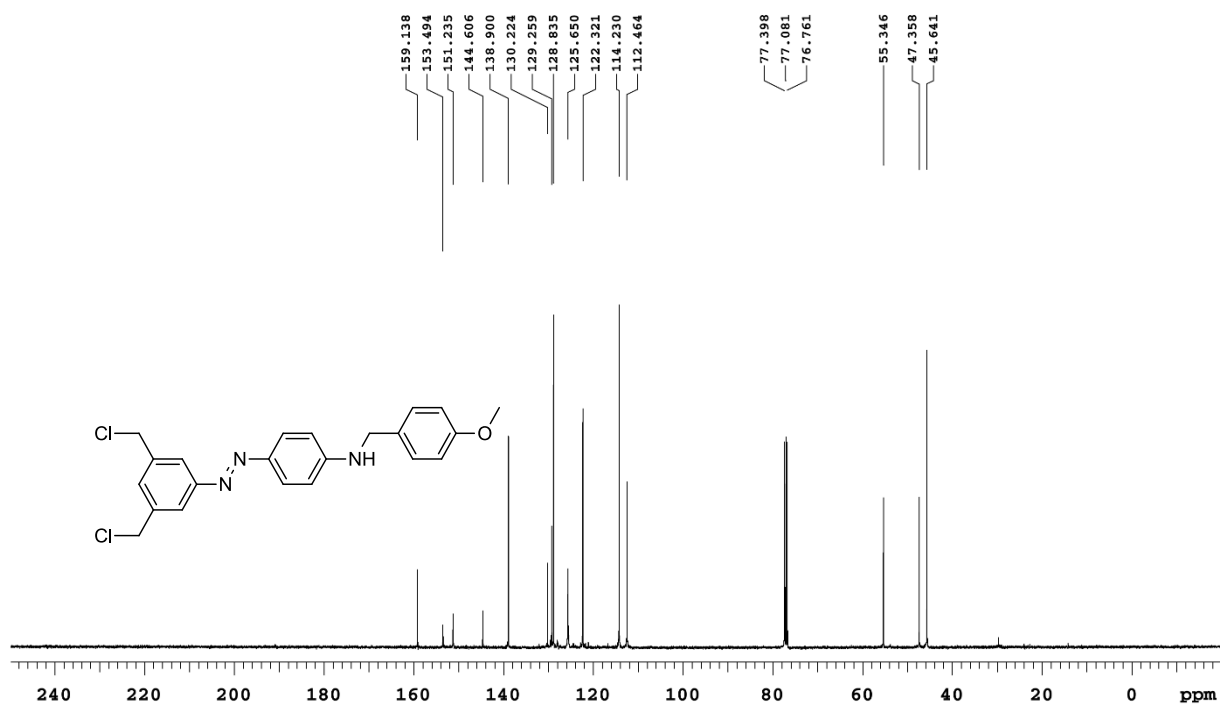3.813
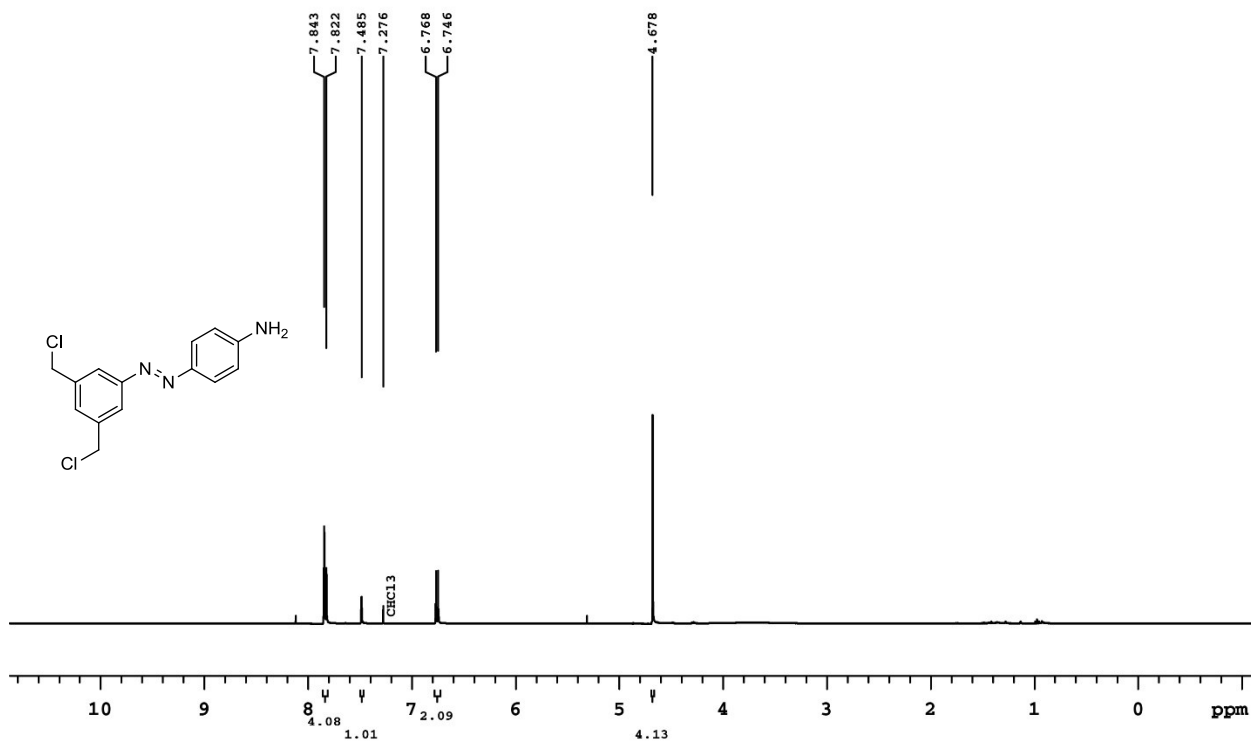3.805

2.08 1.10 2.20
1.99 2.05 2.09

4.39   3.23
   2.10

¹³C-NMR of **2g**

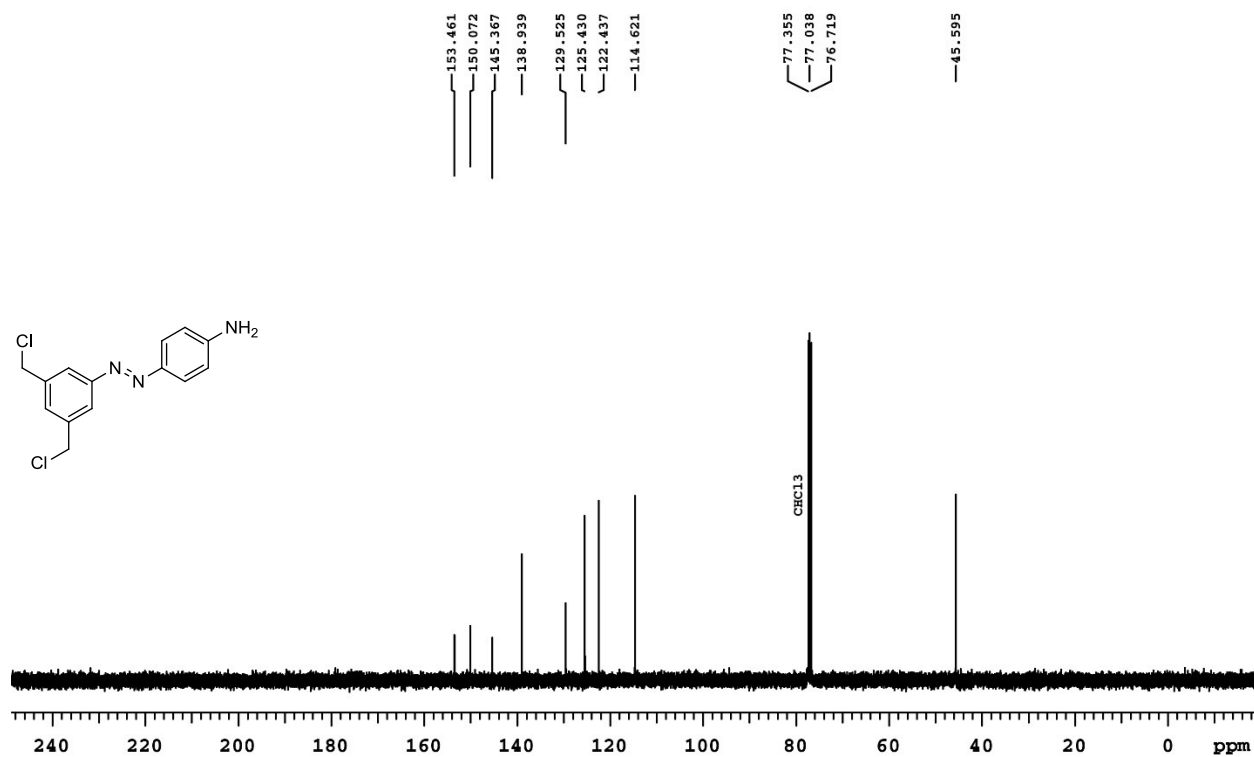100.587 MHz C13[H1] 1D in cdcl3 (ref. to CDCl3 @ 77.06 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe

159.138
153.494
151.235
144.606
138.900
130.224
129.259
128.835
125.650
122.321
114.230
112.464

77.398
77.081
76.761

55.346
47.358
45.641

247

^{1}H-NMR of **2h**

7.843
7.822
7.485
7.276
6.768
6.746
4.678

10   9   8   7   6   5   4   3   2   1   0   ppm
      4.08   2.09   4.13
         1.01

^{13}C-NMR of **2h**

153.461
150.072
145.367
138.939
129.525
125.430
122.437
114.621
77.355
77.038
76.719
45.595

240   220   200   180   160   140   120   100   80   60   40   20   0   ppm

^{1}H-NMR of 2i

399.794 MHz H1 1D in cdcl3 (ref. to CDC13 @ 7.26 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe

7.972
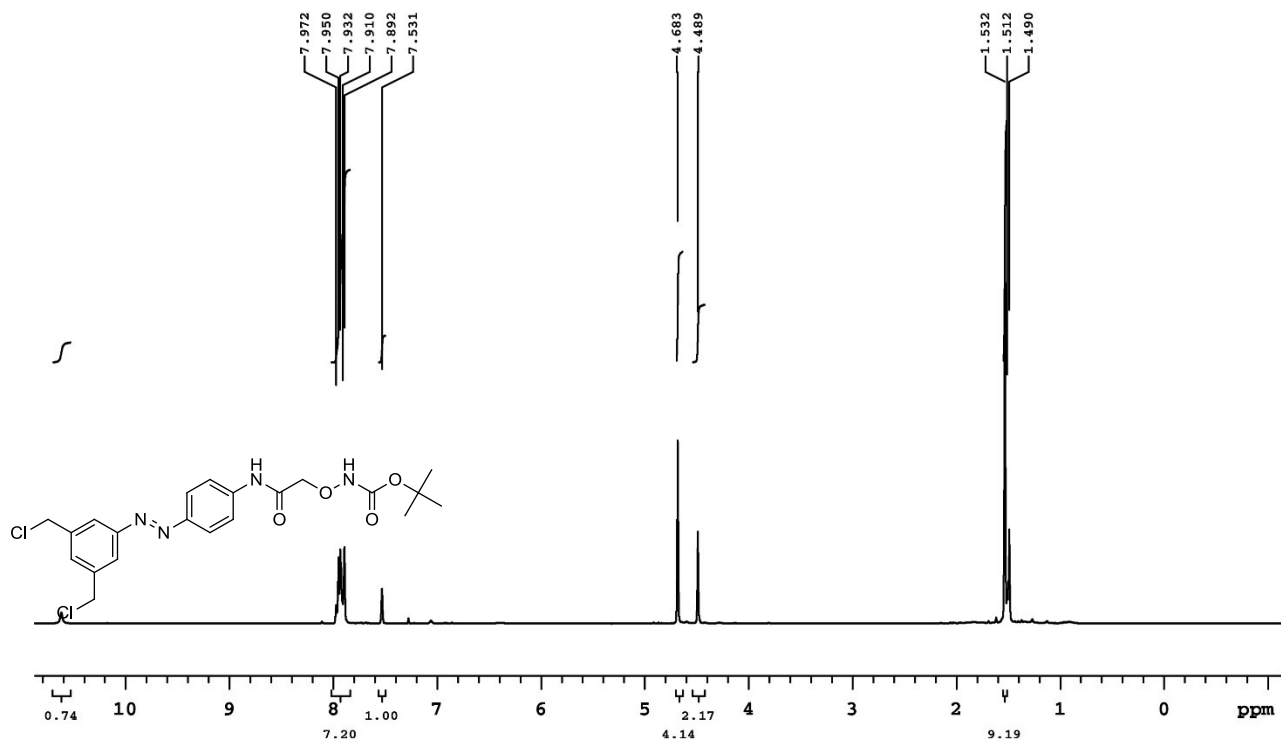7.950
7.932
7.910
7.892
7.531

4.683
4.489

1.532
1.512
1.490

0.74   10   9   8   1.00   7   6   5   2.17   4   3   2   1   0   ppm

7.20   4.14   9.19

^{13}C-NMR of 2i

100.539 MHz C13[H1] 1D in cdcl3 (ref. to CDC13 @ 77.06 ppm), temp 26.5 C -> actual temp = 27.0 C, autoxdb probe

167.540
158.557
153.161
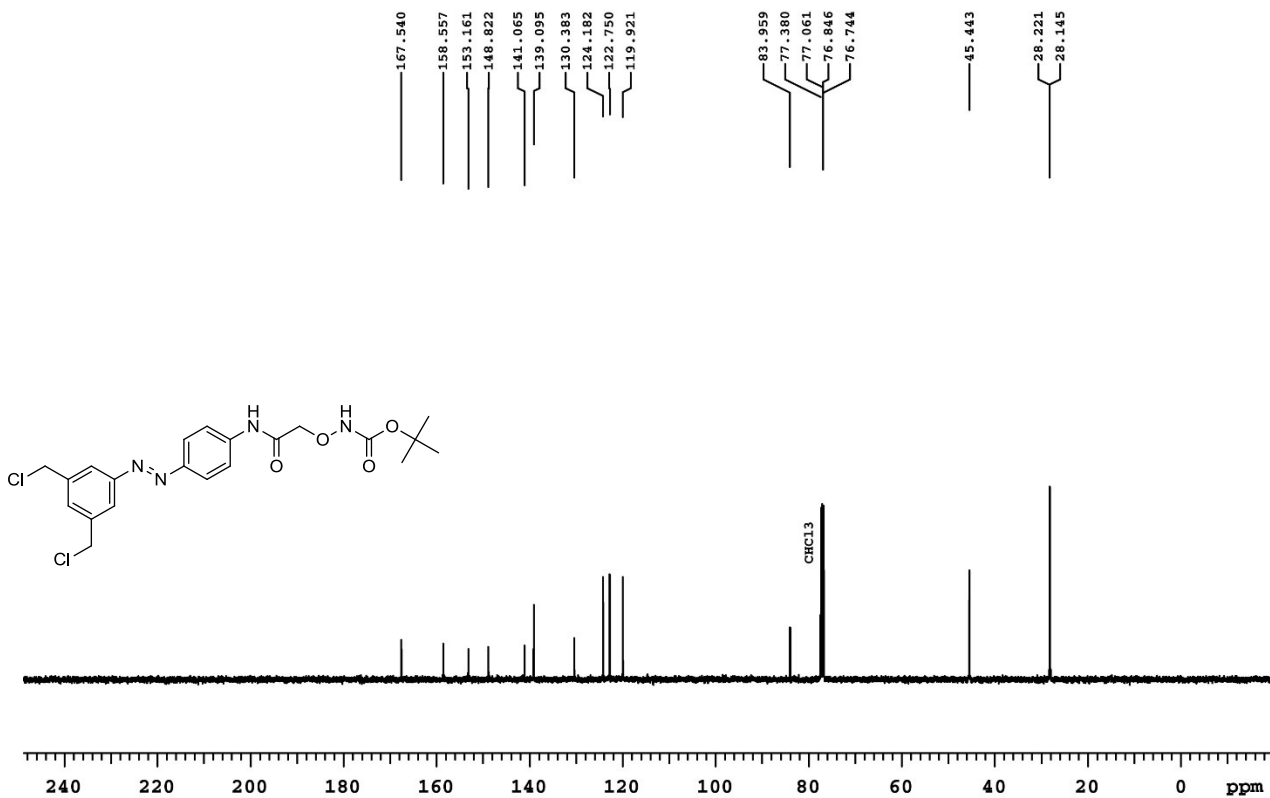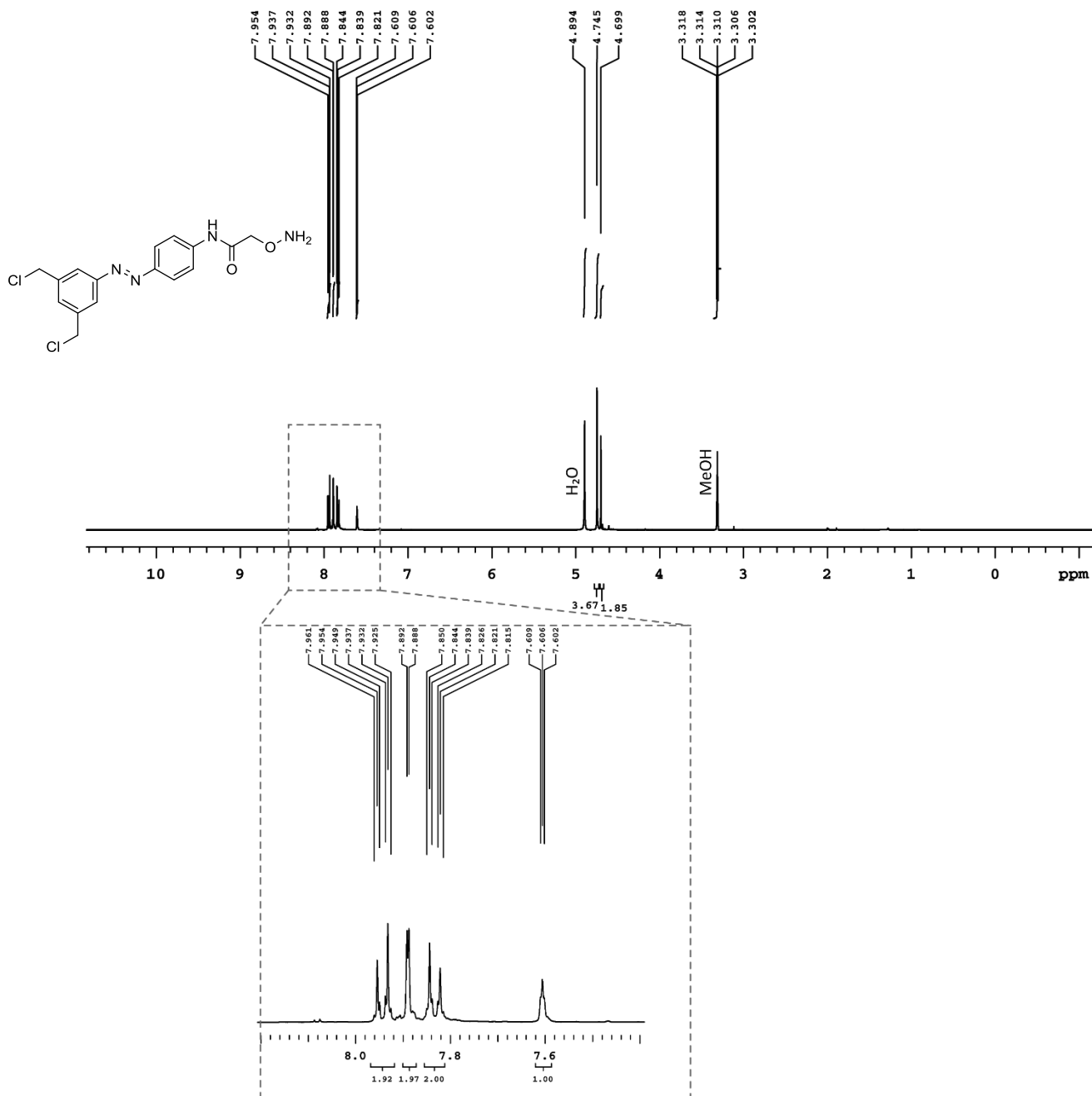148.822
141.065
139.095
130.383
124.182
122.750
119.921

83.959
77.380
77.061
76.846
76.744

45.443

28.221
28.145

CHC13

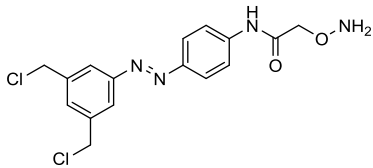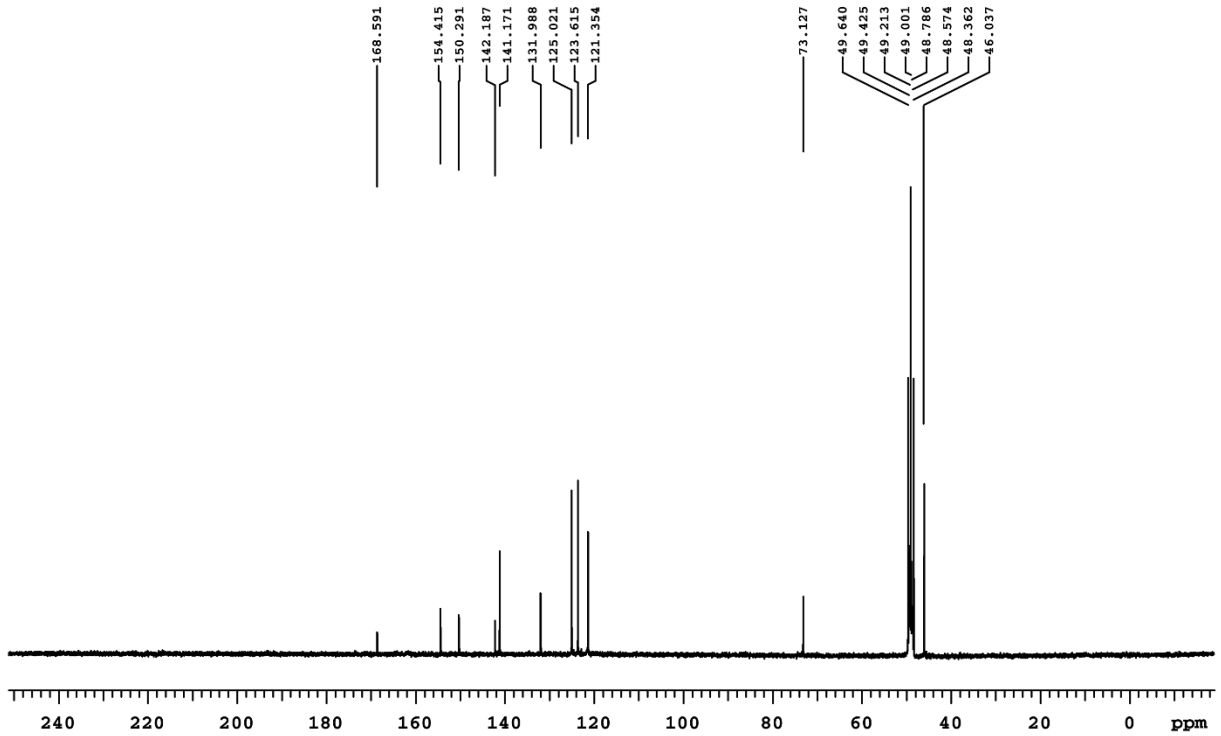240   220   200   180   160   140   120   100   80   60   40   20   0   ppm

^{1}H-NMR of **2j (HADCAz)**

399.986 MHz H1 1D in cd3od (ref. to CD3OD @ 3.30 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe

7.954
7.937
7.932
7.892
7.888
7.844
7.839
7.821
7.609
7.606
7.602

4.894
4.745
4.699

3.318
3.314
3.310
3.306
3.302

H$_2$O

MeOH

10 9 8 7 6 5 4 3 2 1 0 ppm

3.67 1.85

7.961
7.954
7.949
7.937
7.925
7.892
7.888
7.850
7.844
7.839
7.826
7.821
7.815
7.609
7.606
7.602

8.0   7.8   7.6

1.92 1.97 2.00   1.00

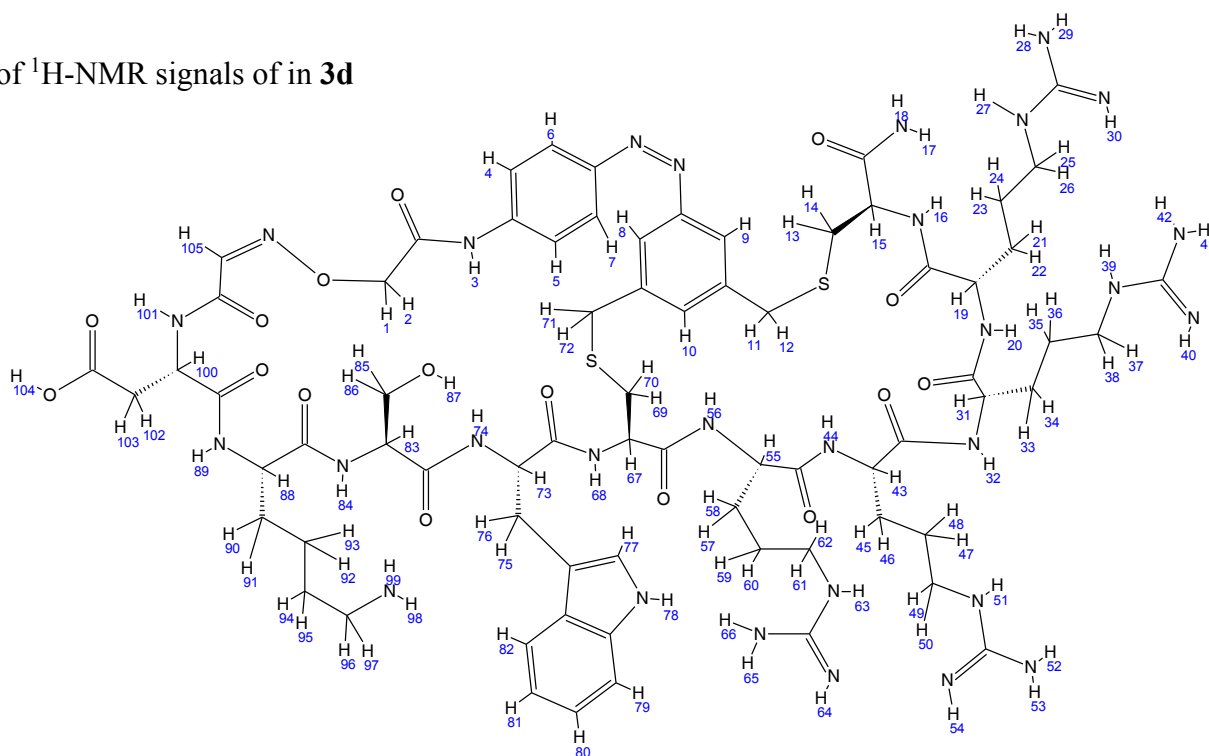## ^{13}C-NMR of **2j (HADCAz)**

100.588 MHz C13[H1] 1D in cd3od (ref. to CD3OD @ 49.0 ppm), temp 25.9 C -> actual temp = 27.0 C, onenmr probe

168.591
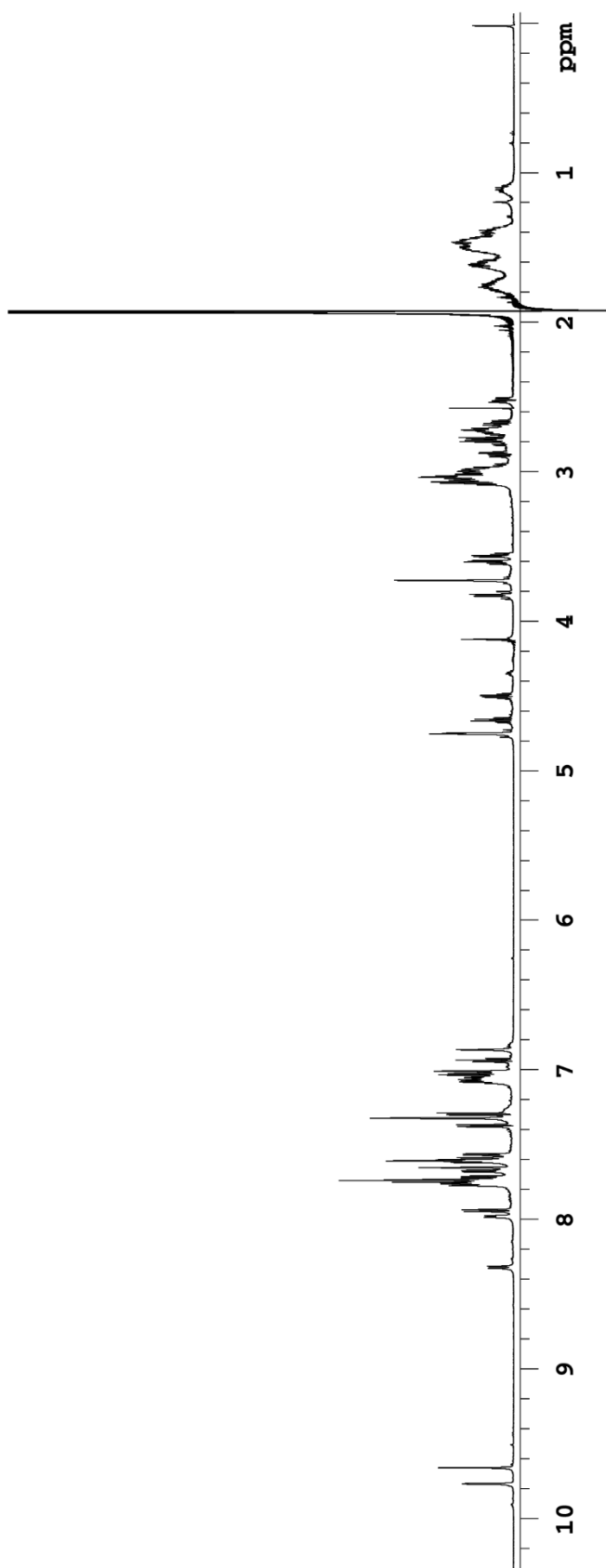154.415
150.291
142.187
141.171
131.988
125.021
123.615
121.354
73.127
49.640
49.425
49.213
49.001
48.786
48.574
48.362
46.037

List of ^{1}H-NMR signals of in **3d**

Residue	NH	ppm	H$_\alpha$	ppm	H$_\beta$	ppm	H$_\gamma$	ppm	H$_\delta$	ppm	H$_\epsilon$	ppm
Asp	101	8.31	100	4.66	102,103	2.81	---	---	---	---	---	---
Lys	89	7.65	88	4.15	90,91	1.53, 1.38	92, 93	1.38	94, 95	2.71	96, 07	3.01
Ser	84	7.76	83	4.2	85,86	3.58	---	---	---	---	---	---
Trp	74	7.56	73	4.34	75,76	3.05, 2.96	---	---	---	---	---	---
Cys	68	7.93	67	4.5	69,70	2.87, 2.66	---	---	---	---	---	---
Arg	56	7.74	55	4.26	57,58	1.7, 1.63	59, 60	1.47	61, 62	3.06	63-66	7.02
Arg	44	7.77	43	4.08	45,46	1.74, 1.61	47, 48	1.46	49, 50	3.01	51-54	7.06
Arg	32	7.68	31	4.12	33,34	1.78, 1.63	35, 36	1.48	37, 38	3.03	39-42	7.05
Arg	20	7.98	19	4.05	21,22	1.78, 1.57	23, 24	1.48	25, 26	3.03	27-30	7.07
Cys	16	7.62	15	4.15	13,14	2.72, 2.53	---	---	---	---	---	---
Trp (Ar)	77	9.79	78	7.01	79	7.32	80	7.03	81	6.93	82	7.38
HADCAz	1,2	4.76	71, 72	3.83	11,12	3.71	---	---	---	---	---	---
	3	9.67	---	---	---	---	---	---	---	---	---	---
	4,5	7.59	6,7	7.73	---	---	---	---	---	---	---	---
	8	7.65	9	7.61	10	7.32	---	---	---	---	---	---
	105	6.86	---	---	---	---	---	---	---	---	---	---

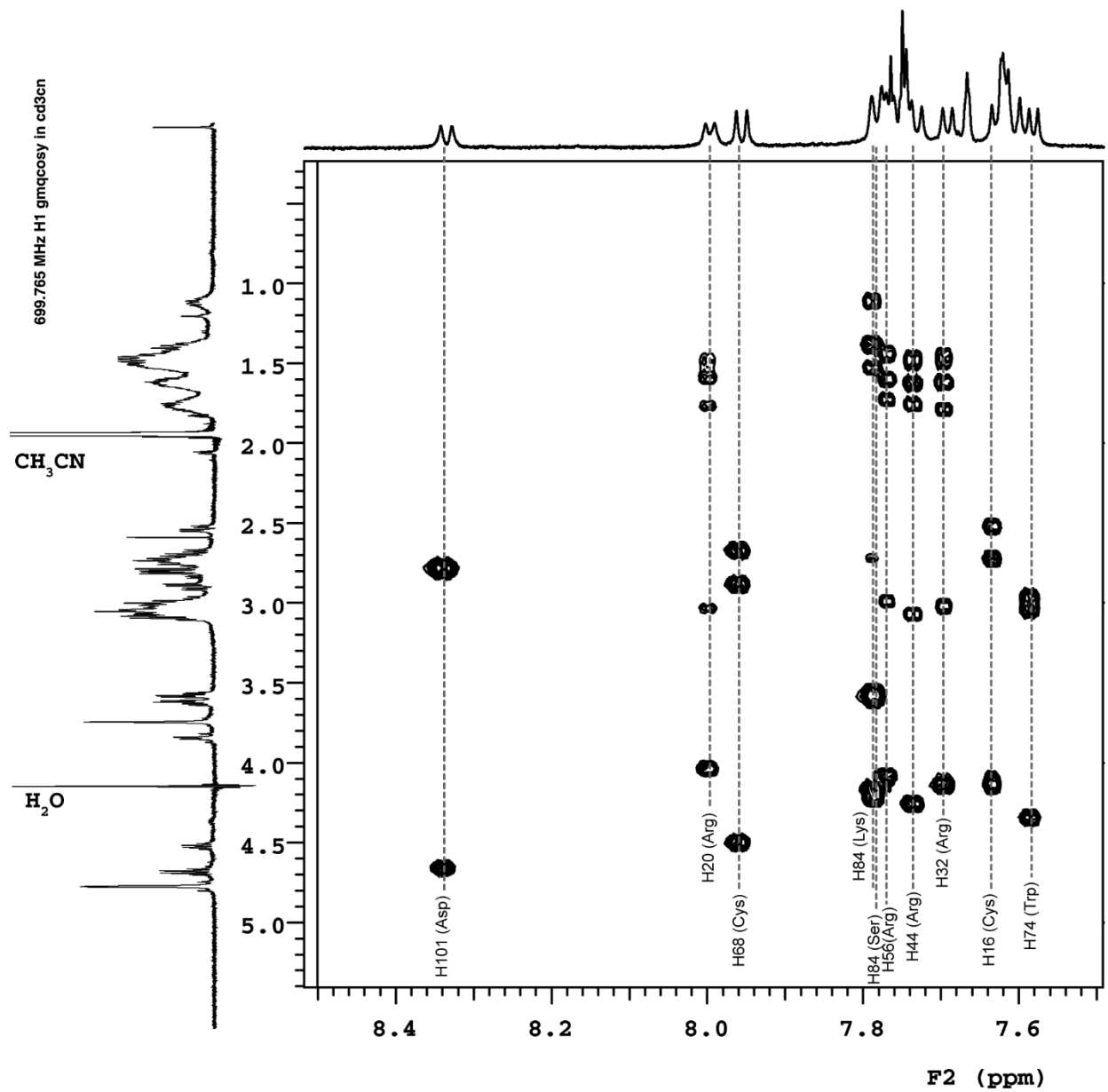699.765 MHz H1 water_ES in cd3cn
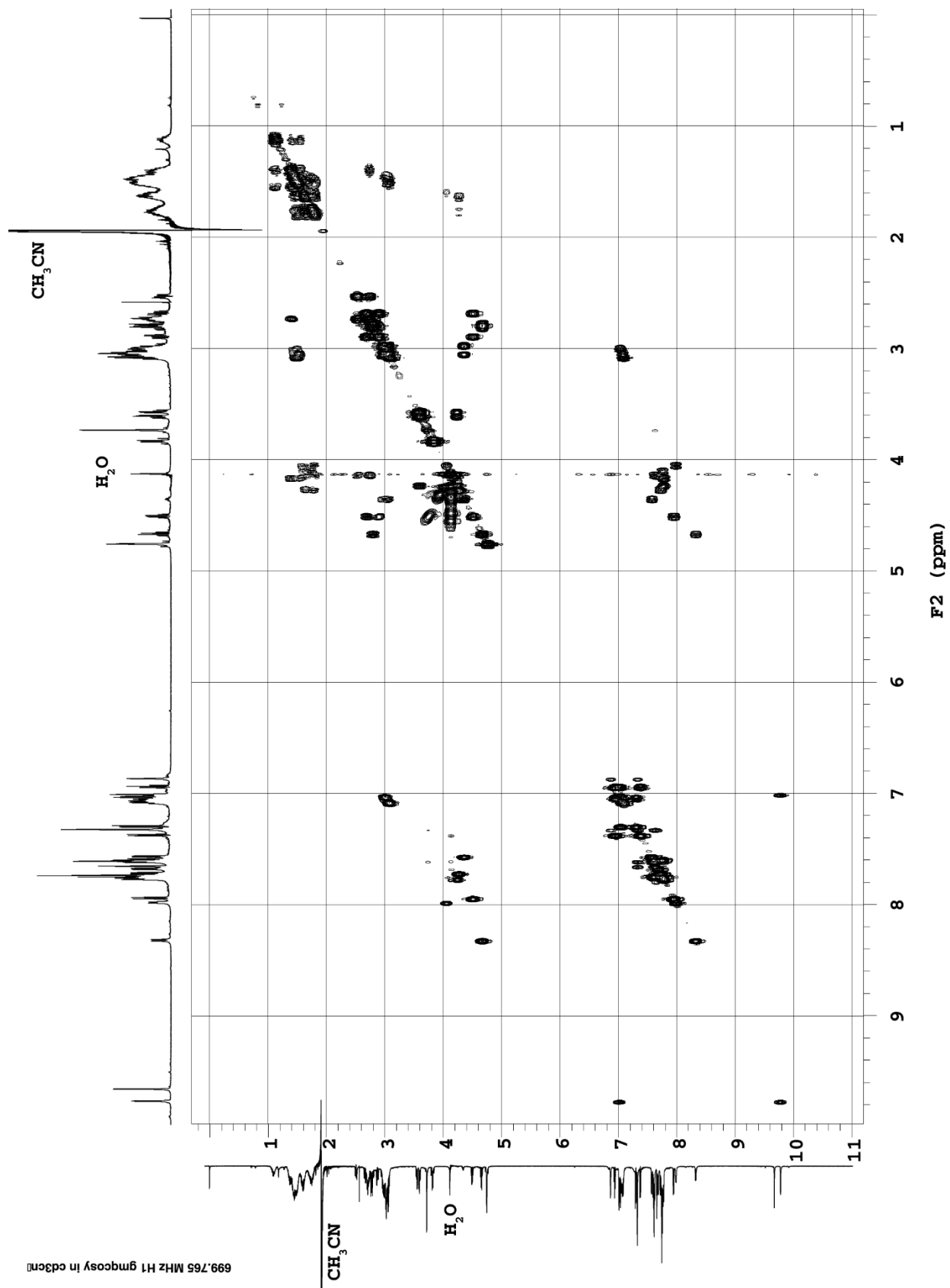
TOCSY⁻-NMR of **3d**

699.765 MHz H1 zTOCSY_ES in cd3cn
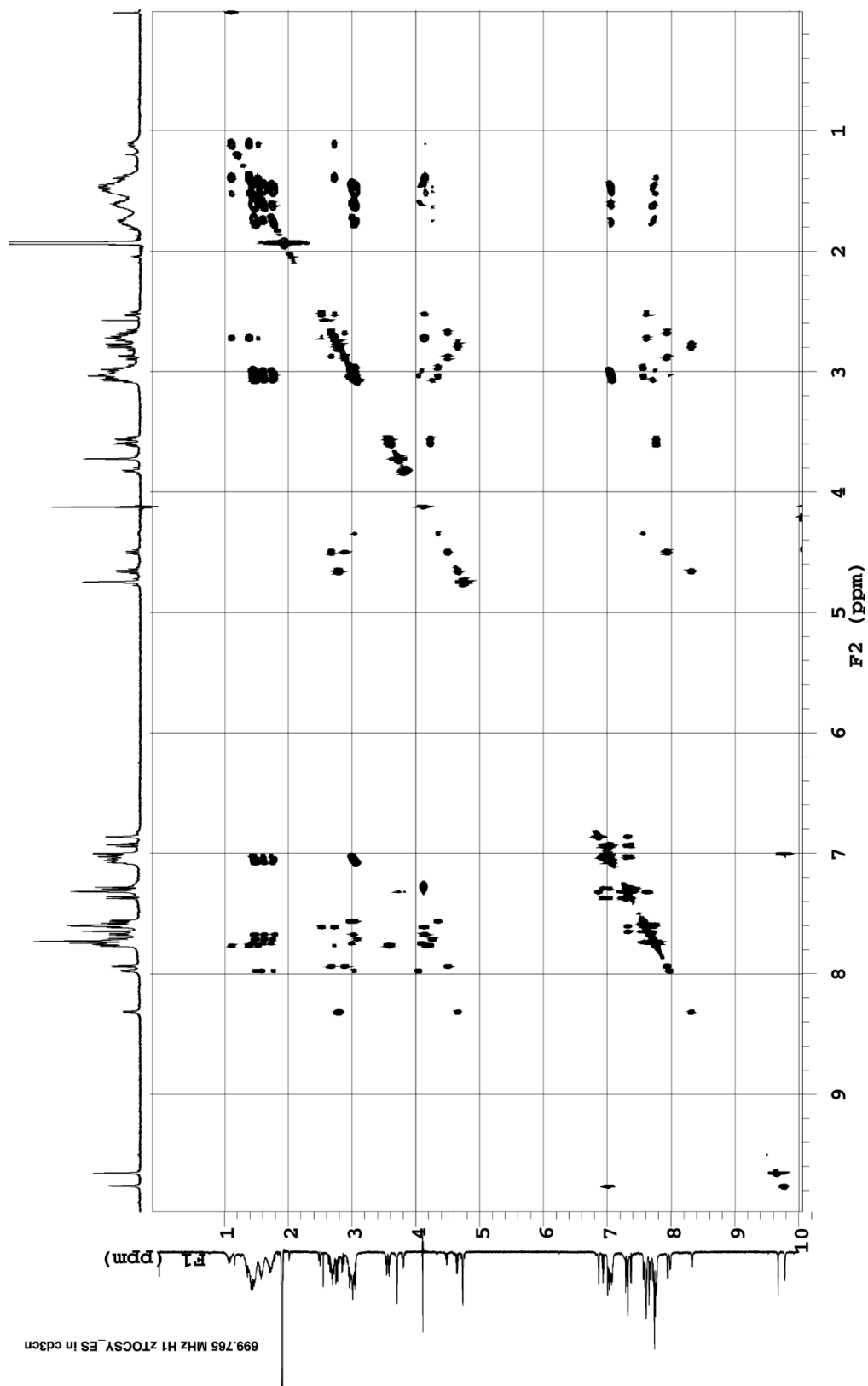
TOCSY-NMR of **3d** (expanded)

COSY-NMR of **3d**
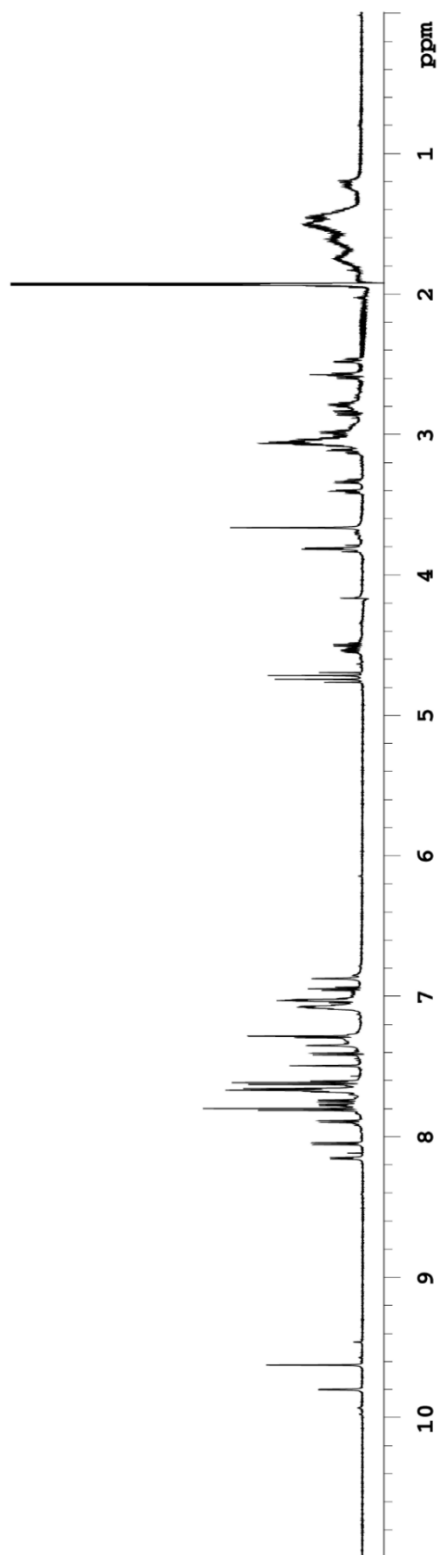
NOESY $^-$-NMR of **3d**

699.765 MHz H1 NOESY_ES in cd3cn

257
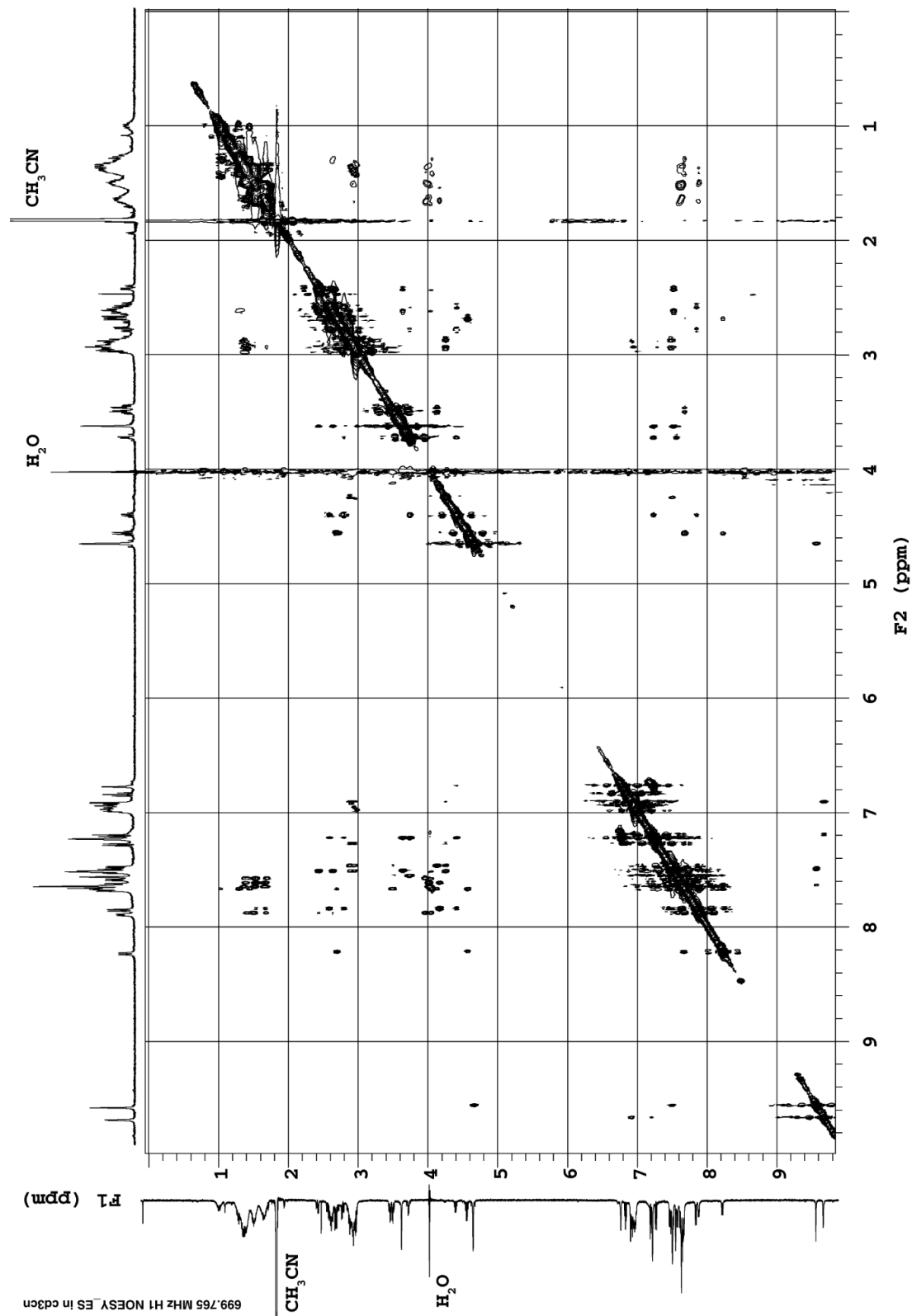
List of ^{1}H-NMR signals of in **3c**

| Residue	NH	ppm	Hα	ppm	Hβ	ppm	Hγ	ppm	Hδ	ppm	Hε	ppm
Lys	89	8.05	88	4.35	90, 91	1.67, 1.54	92, 93	1.24	94, 95	1.54	96, 97	2.82
Ser	84	7.62	83	4.27	85, 86	3.34, 3.43	---	---	---	---	---	---
Trp	74	7.75	73	4.56	75, 76	3.13, 3.08	---	---	---	---	---	---
Cys	68	7.90	67	4.52	69, 70	2.88, 2.60	---	---	---	---	---	---
Arg	56	7.78	55	4.08	57, 58	1.74, 1.62	59, 60	1.43	61, 62	3.01	63-66	7.09
Arg	44	7.66	43	4.12	45, 46	1.80, 1.63	47, 48	1.49	49, 50	3.07	51-54	7.04
Arg	32	7.66	31	4.25	33, 34	1.80, 1.63	35, 36	1.49	37, 38	3.07	39-42	7.08
Arg	20	7.98	19	4.06	21, 22	1.77, 1.57	23, 24	1.50	25, 26	3.07	27-30	7.09
Cys	16	8.13	15	4.24	13, 14	2.59, 2.49	---	---	---	---	---	---
Trp (Ar)	77	9.81	78	7.05	79	7.3	80	7.04	81	6.96	82	7.43
HADCAz	1, 2	4.72	71, 72	3.68	11, 12	3.83	---	---	---	---	---	---
	3	9.63	---	---	---	---	---	---	---	---	---	---
	4,5	7.62	6, 7	7.80	---	---	---	---	---	---	---	---
	8	7.50	9	7.66	10	7.28	---	---	---	---	---	---
	100	6.88	---	---	---	---	---	---	---	---	---	---

¹H-NMR of **3d**

699.765 MHz H1 water_ES in cd3cn

259

TOCSY-NMR of **3c**

TOCSY-NMR of **3c** (expanded)
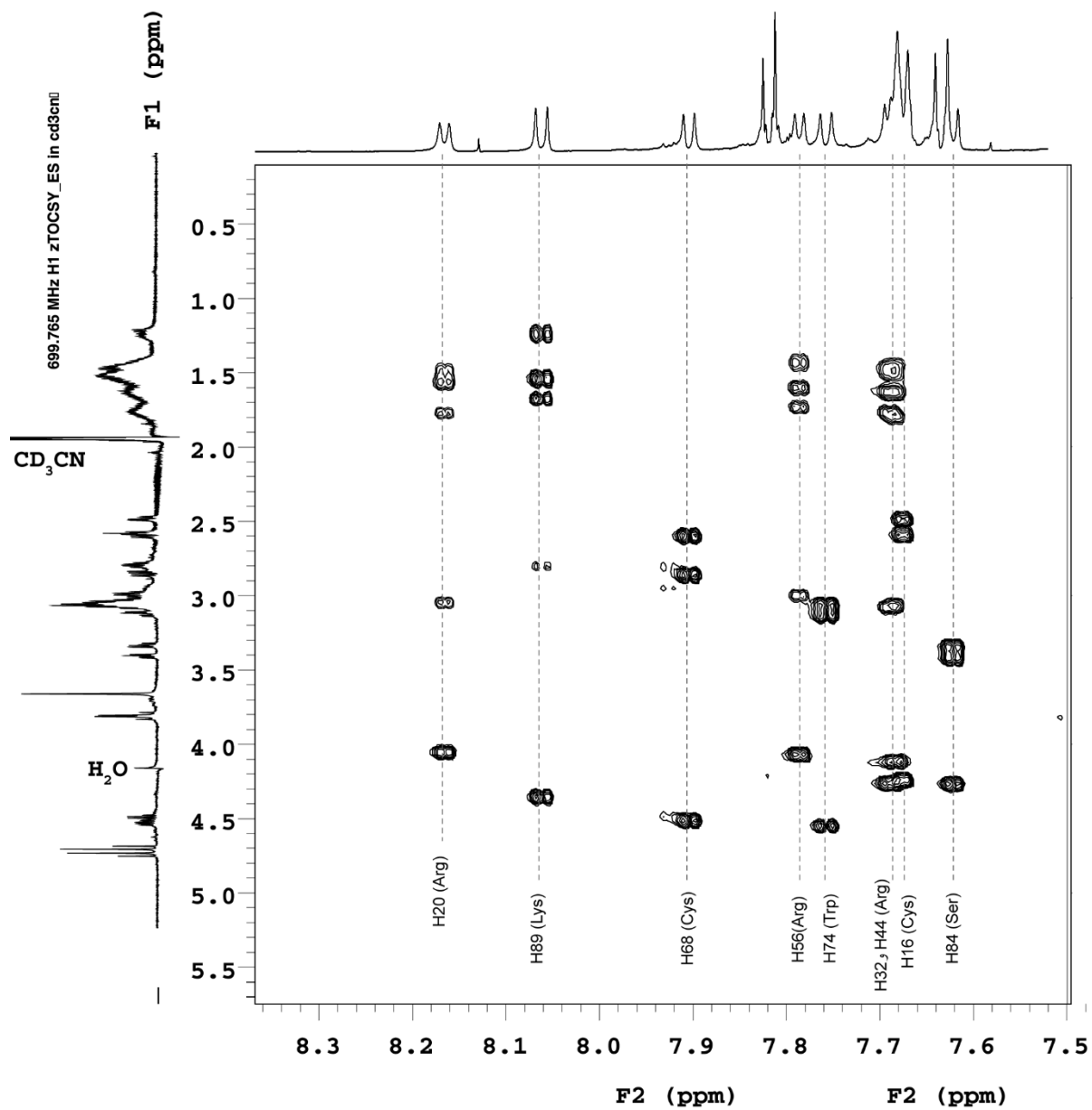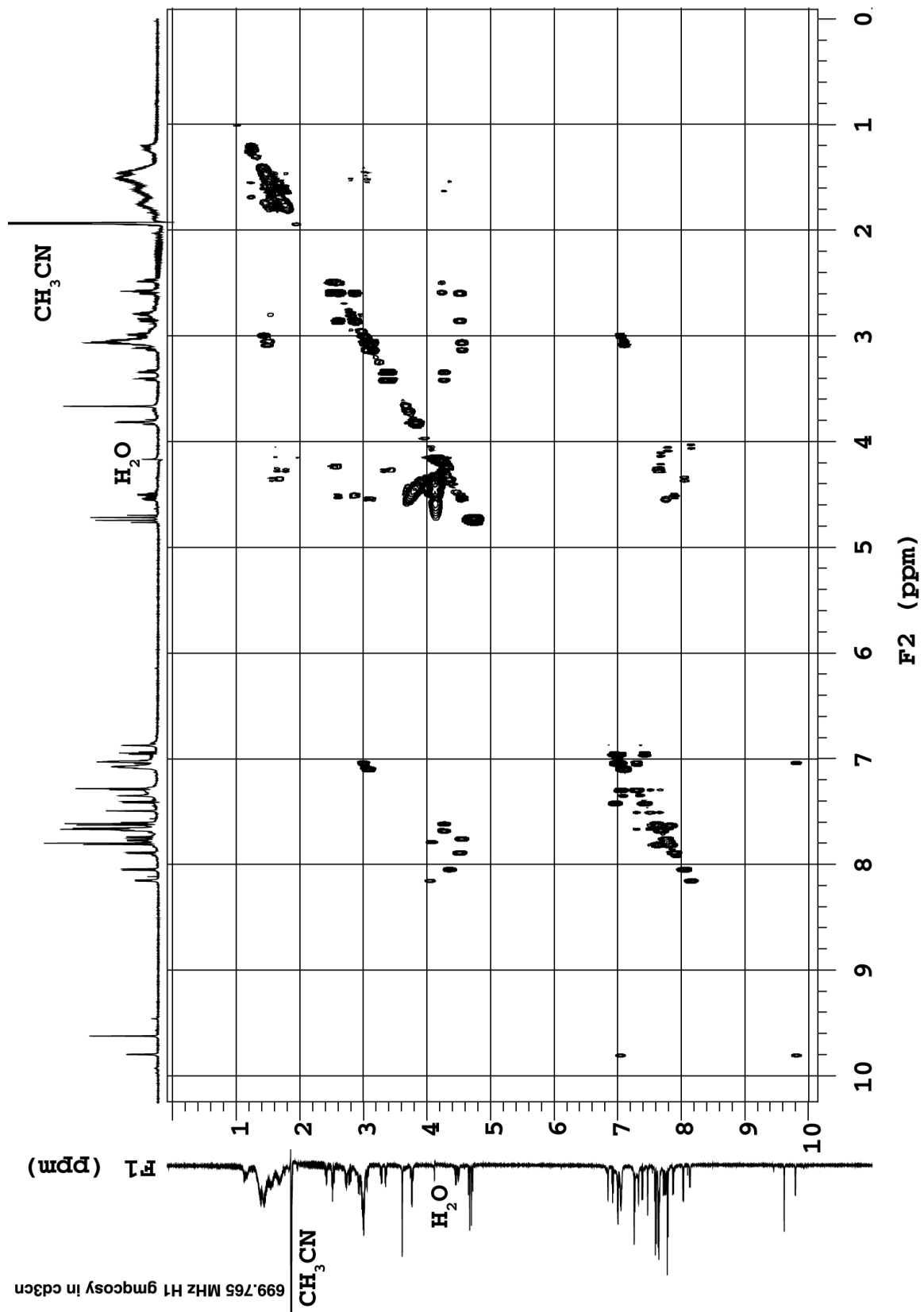
COSY-NMR of **3c**
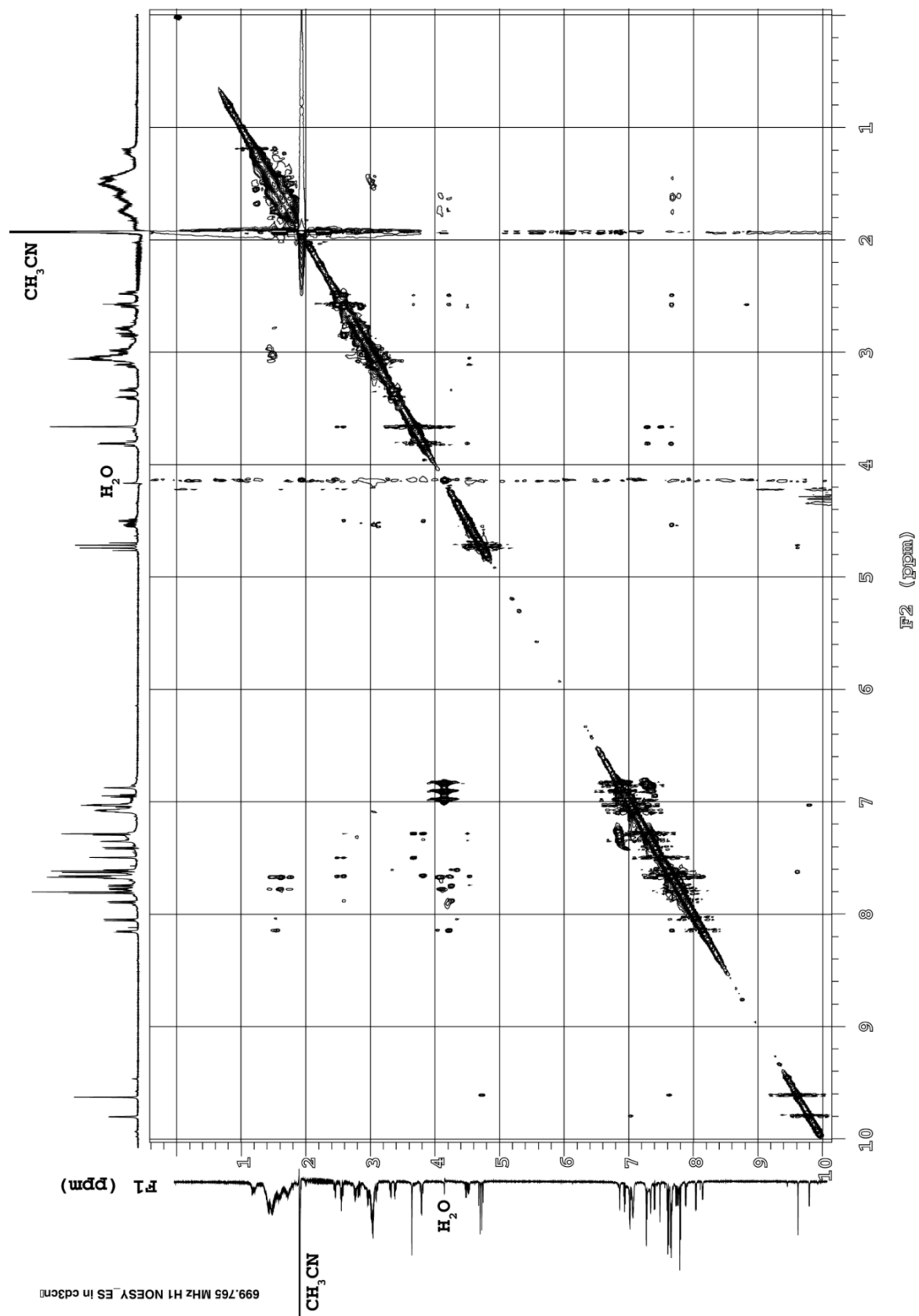
List of ¹H-NMR signals of in **3b**

| Residue	NH	ppm	Hα	ppm	Hβ	ppm	Hγ	ppm	Hδ	ppm	Hε	ppm
Ser	84	7.34	83	4.03	85, 86	3.22	---	---	---	---	---	---
Trp	74	7.55	73	4.67	75, 76	2.87, 3.06	---	---	---	---	---	---
Cys	68	7.98	67	4.50	69, 70	2.82, 2.58	---	---	---	---	---	---
Arg	56	7.65	55	4.26	57, 58	1.75, 1.60	59, 60	1.48	61, 62	3.06	63-66	7.10
Arg	44	7.75	43	4.19	45, 46	2.05, 1.67	47, 48	1.51, 1.52	49, 50	3.08	51-54	7.07
Arg	32	7.67	31	4.10	33, 34	1.76, 1.64	35, 36	1.45	37, 38	2.97	39-42	7.00
Arg	20	8.14	19	4.08	21, 22	1.78, 1.59	23, 24	1.59, 1.50	25, 26	3.05	27-30	7.06
Cys	16	7.68	15	4.09	13, 14	3.09, 2.89	---	---	---	---	---	---
Trp (Ar)	77	9.85	78	6.97	79	7.32	80	7.07	81	6.93	82	7.38
HADCAz	1, 2	4.17	71, 72	3.82	11, 12	3.74	---	---	---	---	---	---
	3	9.69	---	---	---	---	---	---	---	---	---	---
	4,5	7.65	6,7	7.75	---	---	---	---	---	---	---	---
	8	7.64	9	7.53	10	7.32	---	---	---	---	---	---
	88	6.87	---	---	---	---	---	---	---	---	---	---

699.765 MHz H1 zTOCSY_ES in cd3cn

TOCSY‾-NMR of **3b** (expanded)
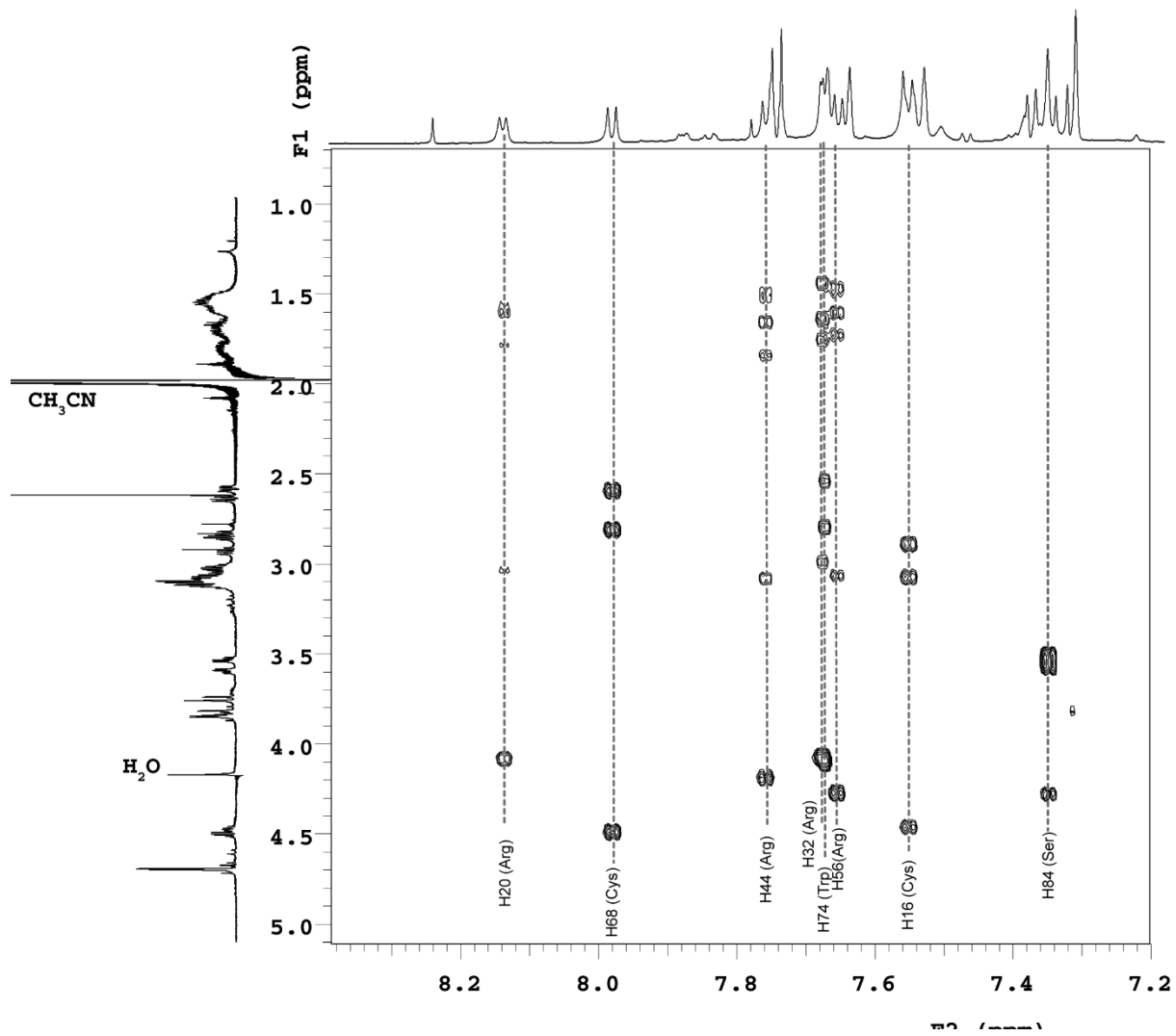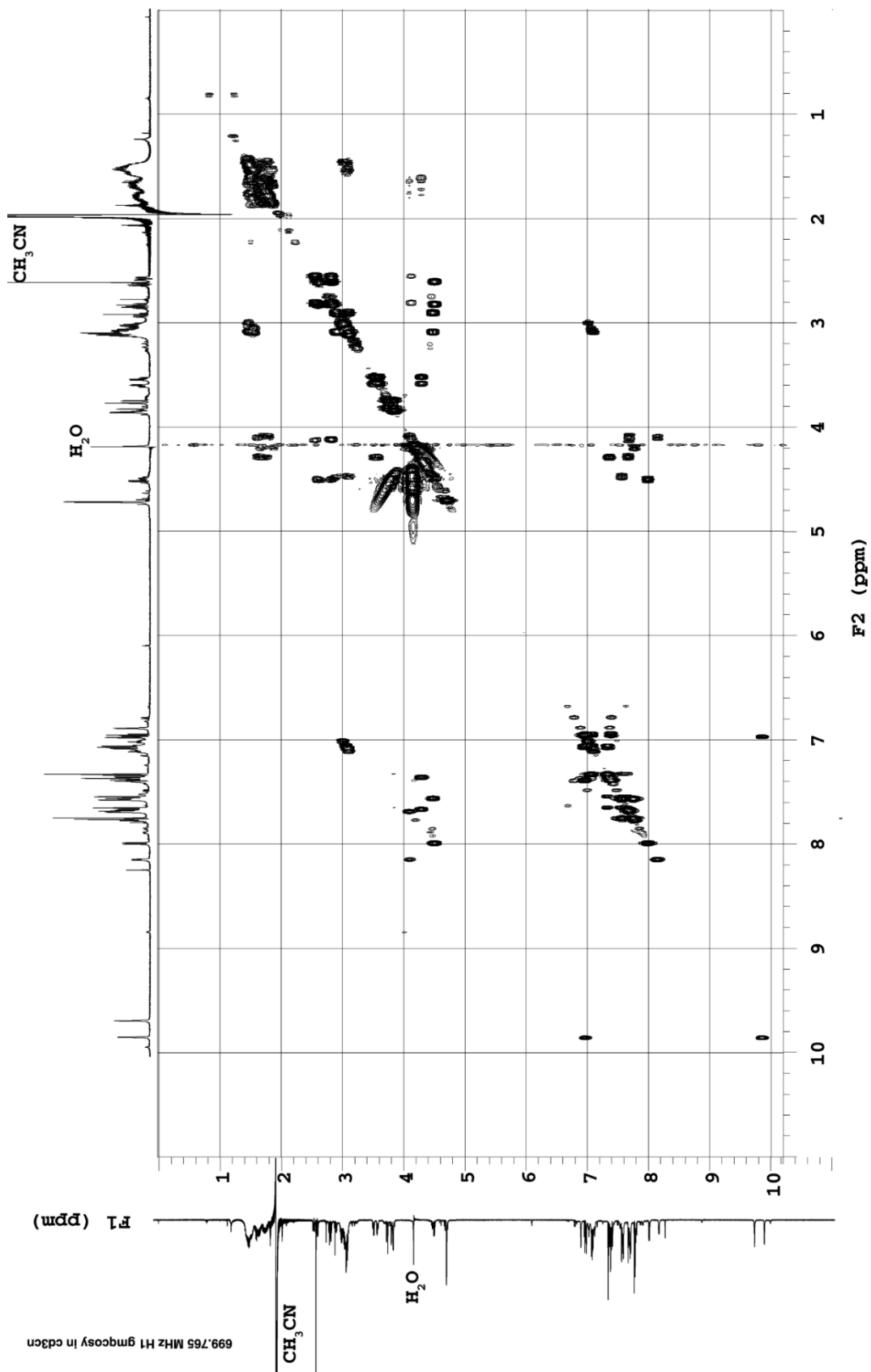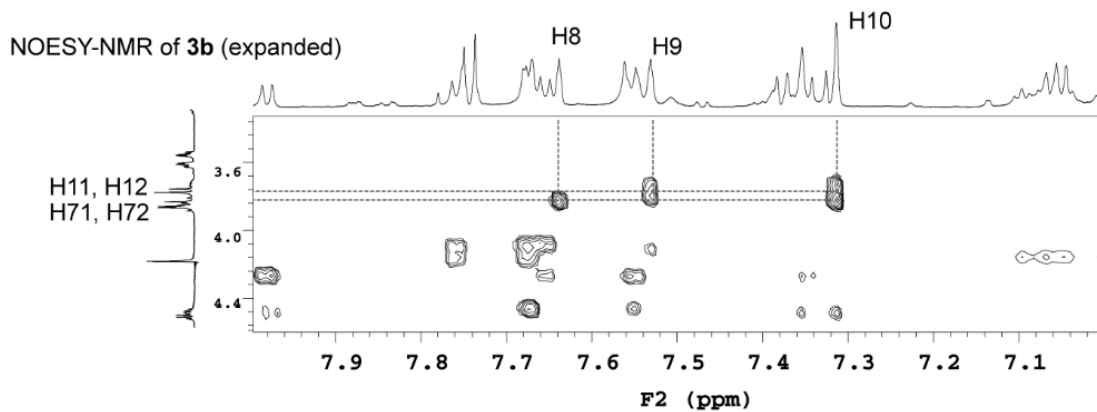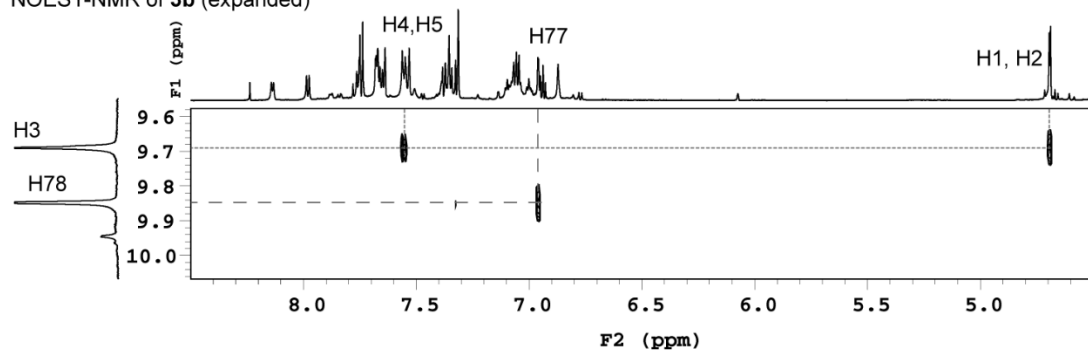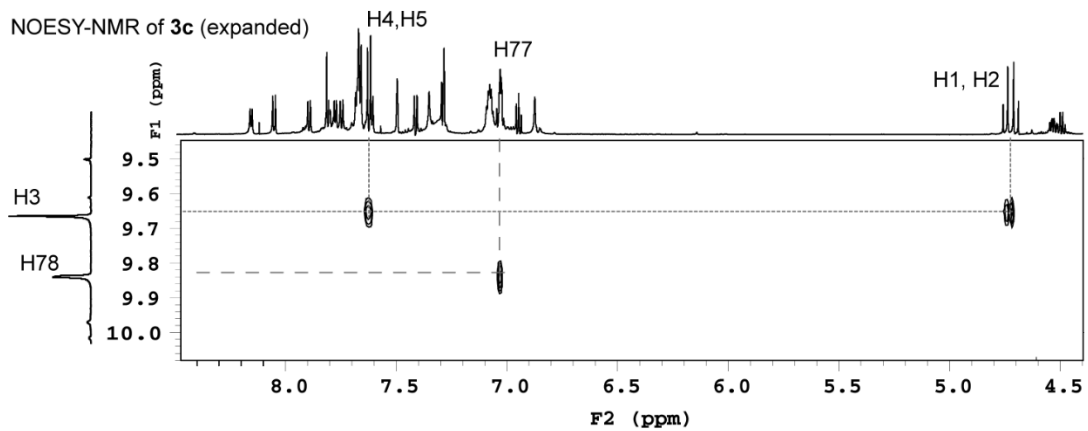
COSY-NMR of **3b**
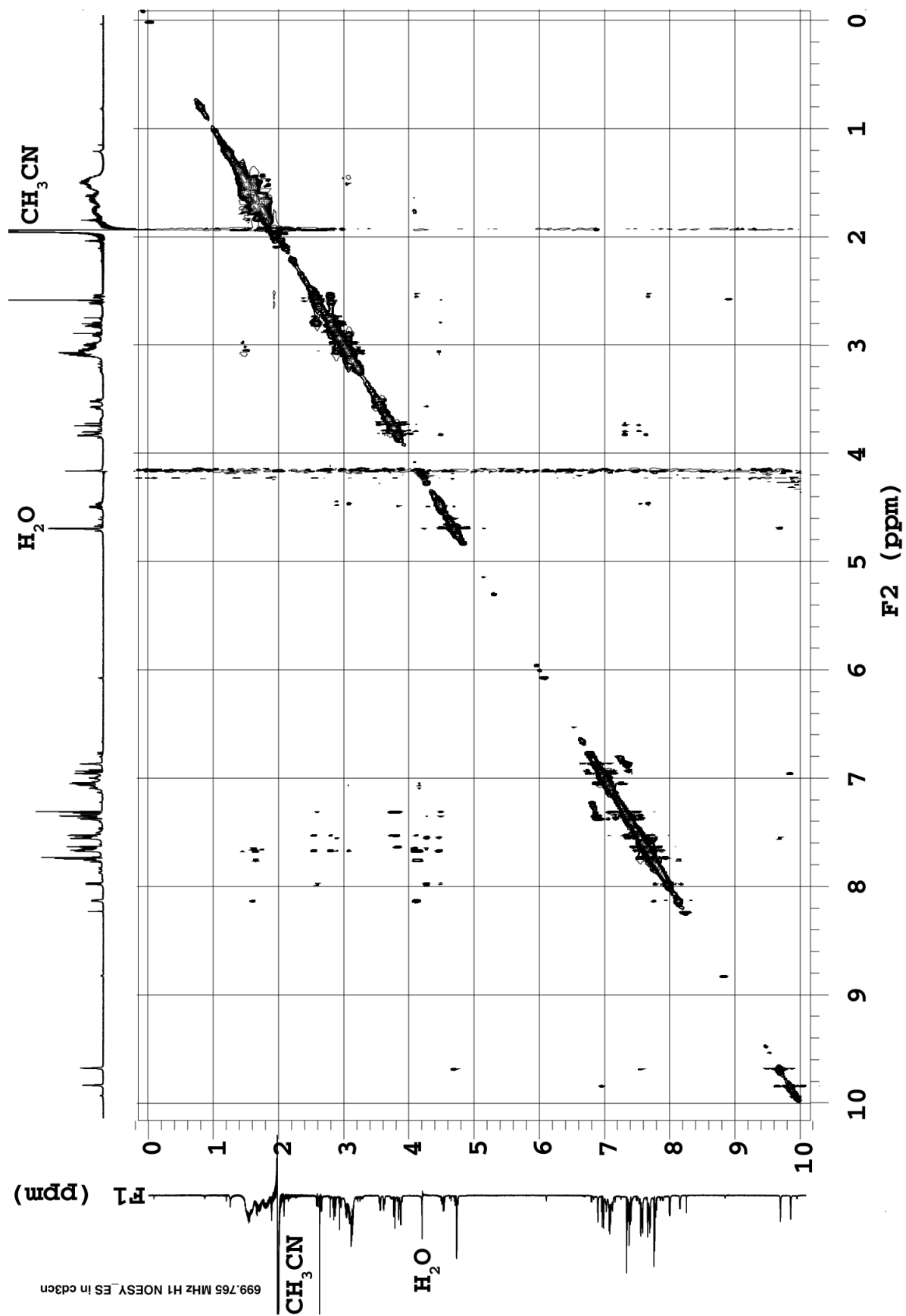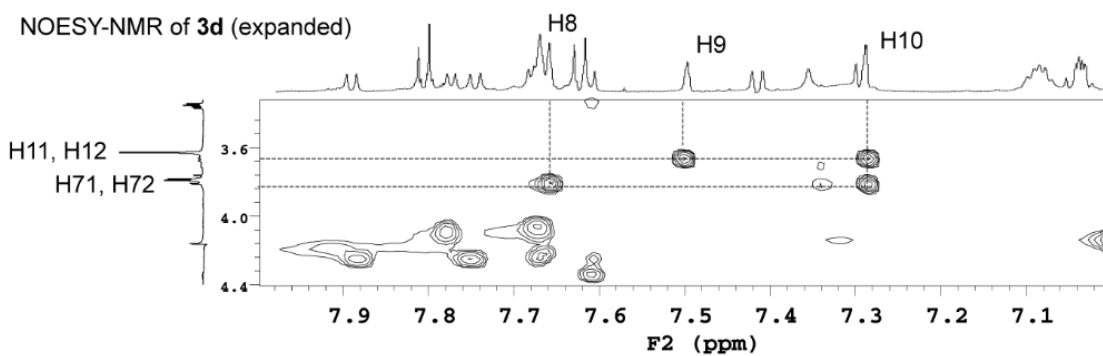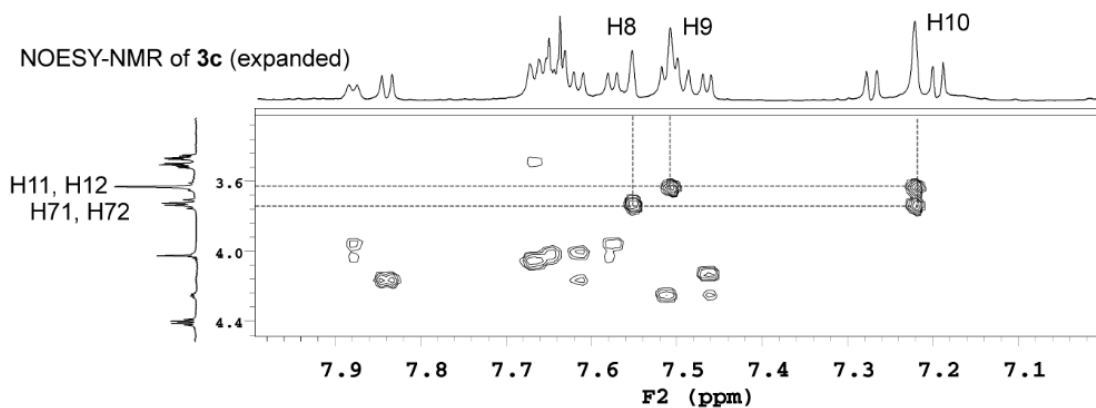
NOESY-NMR of **3b** (expanded)

NOESY-NMR of **3c** (expanded)

NOESY-NMR of **3d** (expanded)

269

NOESY-NMR of **3b** (expanded)

H4,H5   H77   H1, H2

H3
9.6
9.7
H78
9.8
9.9
10.0

8.0   7.5   7.0   6.5   6.0   5.5   5.0

F1 (ppm)

F2 (ppm)

NOESY-NMR of **3c** (expanded)

H4,H5   H77   H1, H2

H3
9.5
9.6
9.7
H78
9.8
9.9
10.0

8.0   7.5   7.0   6.5   6.0   5.5   5.0   4.5

F1 (ppm)

F2 (ppm)

NOESY-NMR of **3d** (expanded)

H4,H5   H77   H1, H2

H3
9.6
9.7
H78
9.8
9.9
10.0

8.0   7.5   7.0   6.5   6.0   5.5   5.0

F1 (ppm)

F2 (ppm)

270

COSY-NMR of **3b** (expanded)

COSY-NMR of **3c** (expanded)

COSY-NMR of **3d** (expanded)

271