Targeting Mutant TERT Promoter in Glioblastoma to inhibit Replicative Immortality

Amanda Rose Metell
Plymouth, Massachusetts

B.S. Chemistry and History, Randolph-Macon College, 2017

A Thesis presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Master of Science

Department of Chemistry

University of Virginia
November, 2019

Abstract

Glioblastoma (GBM) is the most common and malignant brain tumor, with a life expectancy after prognosis of only 15 months. Current therapies for this malignant disease are largely invasive and ineffective. Telomerase is an enzyme that is responsible for maintaining the length of telomeres on DNA, which allows for the ongoing replication of cells. While there are a number of potential targets for therapeutics for GBM, targeting TERT (catalytic subunit of telomerase) seems very promising since mutations in the promoter of TERT increase its expression and are observed in 83% of primary GBMs. Targeting TERT in GBM cells would alter telomerase activity, denying the cancer cells replicative immortality. The promoter mutation in TERT in GBMs results in the loss of binding to the ETS (E26 transformation-specific) family transcription factor, ELF1/2, with replacement by the ETS family transcription factor, GA-binding protein (GABP) which results in increased TERT expression. GABP is a heterodimeric transcription factor consisting of a GABPα DNA binding subunit and a GABPβ transactivation subunit. Since GABPα and GABPβ must bind to each other to increase TERT level, this project aimed to develop a protein-protein interaction inhibitor of GABPα and GABPβ as a therapy for GBM. In this study, the behavior of GABPα and GABPβ was explored with and without the addition of small molecules. First, a FRET assay was developed to monitor the binding by fusing the GFP derivatives, Cerulean and Venus, to the relevant portions of GABPα and GABPβ. With this assay a binding constant as well as IC_{50} values were generated for the proteins and compounds. To confirm which proteins the compounds were binding, a

mutated form of GABPβ (absence of cysteines) was expressed and purified. Lastly, GABPβ, upon the addition of several compounds, was analyzed through ^{15}N-^{1}H HSQC NMR experiments.

The work in this study has identified several compounds with IC_{50} values of less than 1 mM. These compounds represent the building blocks of a potent inhibitor against GABPα and GABPβ. In this study, it was also shown that the cysteine reactive compounds are not binding to the cysteine located on GABPβ. Lastly, this study shows that based on the ^{15}N-^{1}H HSQCs of compounds when bound to GABPβ, the spectrum becomes more resolved, insinuating that these compounds stabilize GABPβ upon binding, or that GABP undergoes a conformational change upon binding to compounds.

This work is dedicated to my mother and father,

My sister,

My grandparents,

My friends and family,

The lab members who have helped me grow as a scientist,

And to Morgan and Cooper, who have never stopped supporting

me with long hikes and unconditional love.

Chapter 1: Introduction

1.1 Significance

Malignant gliomas are the most common fatal cancers that originate in glial cells in the central nervous system. Out of all gliomas, glioblastoma multiforme (GBM) is the most common and malignant brain tumor, with a life expectancy after prognosis of only 15 months. This aggressive cancer affects nearly 200,000 patients each year, with a five-year survival rate of only 10%.[1] Therapies for this malignant disease including surgery, radiation, and chemotherapy, may slow the progression of the cancer and reduce signs and symptoms. However, these therapies are largely invasive and ineffective. Typically, these methods offer a median survival time of only 15 months.[2] There needs to be a drive for novel therapies for the treatment of GBM.

GBM occurs due to the deregulation of factors that control tumor cell growth and survival. Replicative senescence is dependent on a limited number of cell divisions. Human telomeres are repeat sequences of TAGGG/CCCTAA on the ends of chromosomes. Telomerase is an enzyme that is responsible for maintaining the length of telomeres on these chromosomes, which allows for the ongoing replication of cells. In the majority of healthy somatic cells, telomerase is silenced which leads to typical replicative senescence due to the telomeres shortening after multiple rounds of cell division.[3] While there are a number of potential targets for therapeutics for GBM, targeting telomerase reverse transcriptase (TERT) seems very promising. In a study where 78 GBM tumors were examined, 83% were found to have mutations in the promoter of TERT (C228T and

C250T) which increased its expression (Figure 1).[4] Targeting TERT in GBM cells would alter the telomerase activity, denying the cancer cells replicative immortality. Since TERT is the catalytic subunit of telomerase, there is promise in targeting it to alter activity.


Figure 1: Distribution of TERT mutations (red) and other genetic alterations in 51 GBMs. Each column represents one GBM.[4]

In recent studies, it has certainly become clear that many different types of tumors (particularly GBMs) rely on TERT activation to enable unlimited proliferation. Activation of TERT is the rate-limiting step when producing activated telomerase. In a recent study, mutations in the promoter region of TERT were found to be the most common mutation in over 50 cancer types.[5] Research today is shining more light on how TERT becomes activated, likely due to the recruitment of transcription factors that do not normally regulate gene expression.

Specifically in GBMs, the promoter mutation in TERT results in the loss of binding to the ETS (E26 transformation-specific) family transcription factor, ELF1 or ELF2, with replacement by the ETS family transcription factor, GA-binding protein (GABP) which results in increased TERT expression (Figure 2).[6] More specifically in this study, they conducted a proteome-wide approach to identify the dynamic relationship when GABP and ELF1 are in vitro with the TERT promoter region. It was found in this study that when

GABP took on a heterotetrameric form, it became a transcriptionally active form out-competing ELF1 to bind to the promoter region of TERT. In recent years, it has become clear that targeting specific transcription factors with small molecules proves advantageous in many cancers and other diseases. Targeting protein-DNA and protein-protein interactions with small-molecule inhibitors is becoming more common with research today.
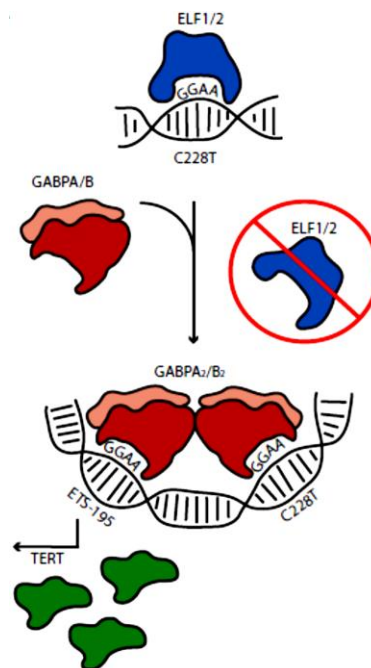
Figure 2: Schematic diagram that shows when there is a mutation in the promoter region of TERT, there is a change from ELF ½ to GABP leading to increased TERT expression.[6]

The focus of this thesis is to produce a protein-protein interaction inhibitor effective with the GABP complex, offering a novel therapy for GBM. This study utilizes 2D NMR studies as well as fluorescence resonance energy transfer (FRET) assays to characterize the binding of small molecules to the GABP subunits. Since FRET relies on the energy transfer

between donor and acceptor fluorophores to be within 10 nanometers from each other, this assay could be utilized to monitor the binding between GABPα and GABPβ. Because GABPα and GABPβ have been fused with their respective fluorophores, when Cerulean is attached to GABPα, and Venus is attached to GABPβ, the energy transfer between the two fluorophores corresponds to the binding of GABPα and GABPβ. By measuring the emission of Cerulean and the emission of Venus, a FRET ratio is produced which is characteristic of the binding between GABPα and GABPβ. To monitor where specific cysteine-reactive compounds were binding, a cysteine-free GABPβ was produced and also tested with this FRET assay. To rule out false positives from the FRET assay, chemical shift perturbations in an ^{15}N-^{1}H heteronuclear single quantum coherence nuclear magnetic resonance (HSQC NMR) spectra can be monitored. This is a well-validated approach to confirm binding to a target when taking NMR of proteins upon the addition of compounds. Since the protein is isotopically labeled with ^{15}N, the amide NH resonances are visualized, ideally producing a spectrum that has a well-dispersed peak for each amino acid in the protein.

1.2 Telomeres

Telomeres are relatively short DNA sequences that reside on the ends of chromosomal DNA. These DNA sequences are quite simple with repeat units of AGGGTT around the size of a few thousand base-pairs.[7,8] Telomeres serve several functions to the chromosomal DNA. While their main purpose is to serve as a buffer during DNA replication, they also serve to stabilize the chromosomal structure.[9,10] Due to the presence of telomeres,

chromosomal DNA can be replicated completely without any loss of terminal bases at the 5' end of each DNA strand. The DNA polymerase machinery is not able to completely replicate an entire strand of DNA because it is only able synthesize DNA from the 5' end to the 3' end. During DNA replication, there is a lagging strand that replicates short strands of 5' to 3' DNA while it moves down the DNA strand from the 3' to 5' end. Because DNA polymerase requires short RNA primers to begin its function, there are many RNA primers dispersed throughout the lagging strand upon completion of DNA replication. At this point, DNA polymerase and DNA ligase replace these RNA primers with DNA fragments and fuse all of the bases together. However, DNA polymerase can only begin its function if there is another DNA strand in front of the gap where the RNA primers were. In almost all instances on the lagging strand, this is true, except on the end, where the last RNA primer was made. So that RNA primer is removed and the daughter DNA strand is slightly shorter than the parent strand (Figure 3).

Thus, telomeres serve as a buffer for the inability to completely replicate a DNA strand, so during each round of replication, a small portion of the telomere is removed, versus an integral part of the chromosomal DNA. The leading strand of DNA is not affected by this phenomenon because the origin of replication starts in the middle of the chromosome, so during replication there are always two lagging strands and 2 leading strands. The beginning of the leading strand that serves as an RNA primer, will be replicated as the last Okazaki fragment of the lagging strand on the same DNA strand.
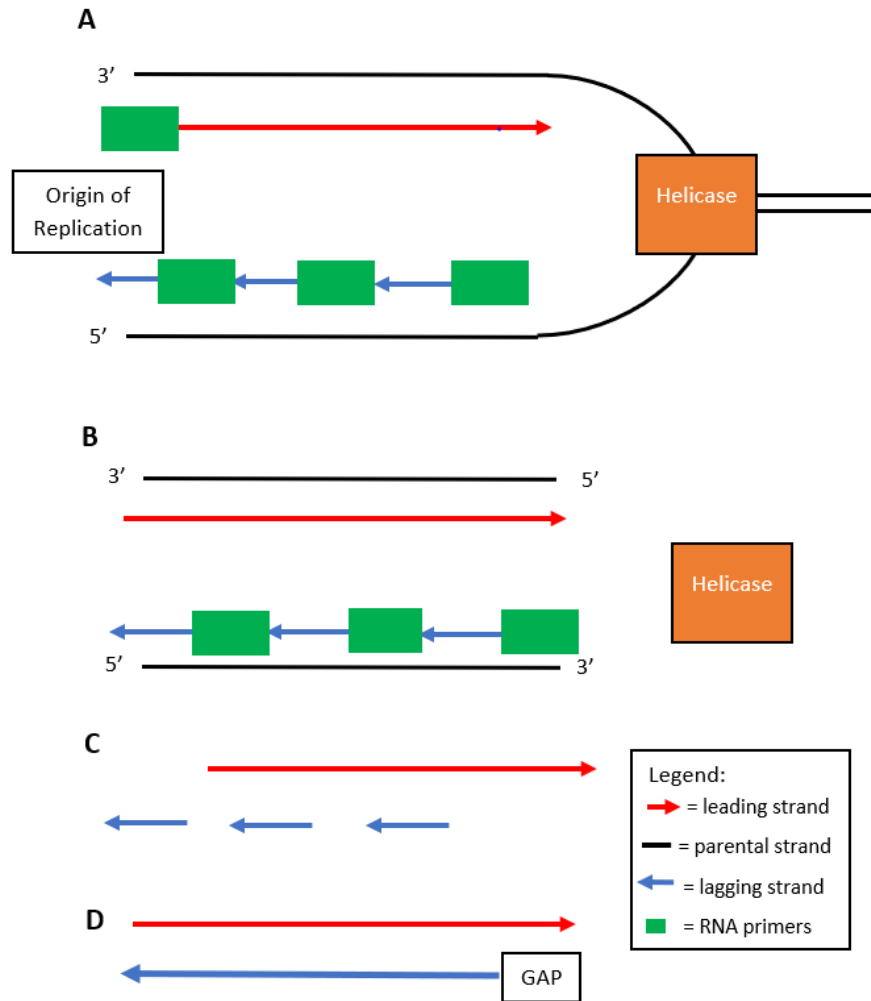
Figure 3: How the lagging strand of DNA shortens during DNA replication. (A) The origin of replication begins in the middle of a DNA strand with the enzyme, helicase, separating the two DNA strands as it moves to the end of the chromosome. RNA primers supersede the synthesis of the leading and lagging strand because DNA polymerase needs an RNA primer to function. Since DNA polymerase only operates in the 5' to 3' direction, the lagging strand makes small portions of replicated DNA called Okazaki fragments. (B) Helicase has reached the end of the chromosome, and the length of each DNA strand has been replicated. (C) RNA primers are removed, leaving gaps within the newly replicated DNA. (D) The gaps are filled in by DNA polymerase. However, DNA polymerase can only function in between newly synthesized DNA fragments. In almost all instances on the lagging strand, this is true. However, it does not happen on the end. So, when the last RNA primer is removed, the daughter DNA strand has a gap and is slightly shorter than the parent strand. The leading strand of DNA has no gap because the beginning of the leading strand that serves as an RNA primer, will be replicated as the last Okazaki fragment of the lagging strand on the same DNA strand.

1.3 Telomerase and TERT

Telomerase is an enzyme composed of the protein TERT and an internal RNA template (TERC), responsible for elongating the telomeres in the 5' to 3' direction. TERT copies the TERC template and adds deoxyribonucleotides in a tandem repeat, characteristic of the telomeric units to the ends of telomeres.[11,12,13] Telomerase is typically silenced in normal somatic cells, allowing the cells to reach senescence and die with age. Before cell death, the chromosomes suffer from becoming shorter due to the absence of telomeres and continued replication. Telomerase is responsible for maintaining telomere length and thus long-term cell viability.[14] In general, telomerase activity in somatic cells correlate with the over-proliferation of cells in many types of cancer. Moreover, immortality of somatic cells requires the presence of this enzyme to maintain the length of telomeres for proper DNA replication.[15]

It has recently been discovered that mutations in the promoter region of TERT are found in many types of tumors, particularly melanomas.[16,17] More specifically, mutations C228T and C250T accounted for about 98.3% of the mutations in the TERT promoter regions.[4] In all of the tumors that have been examined for TERT promoter region mutations thus far, glioblastoma is the cancer that has the most, 83% (Figure 4). Because of this high correlation between TERT promoter region mutations and glioblastomas, it is possible that this mutation is responsible for the high telomerase activity in primary glioblastomas. Due to this, targeting the results of these promoter mutations may prove to provide a novel cure for glioblastomas.
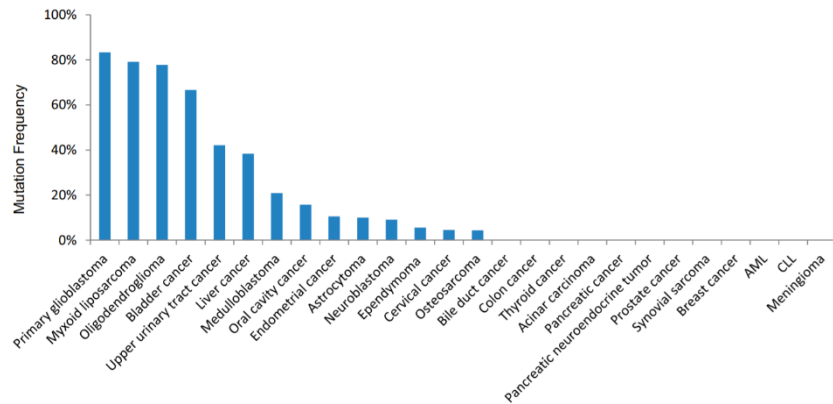
Figure 5: Bar graph depicting the amount in percentage of tumors that had a mutation in the promoter region of TERT.[4]

1.4 GABP

As the name (GA-binding protein) implies, the GABP complex binds to DNA sequences containing mainly the nucleotides guanine (G) and adenine (A) which is a characteristic for ETS (E26 transformation specific) transcription factors.[18,19] There are 28 distinct human ETS factors which all share an evolutionarily conserved DNA binding domain (DBD) of 85 amino acids.[20,21] GABPα has been shown to be essential for several cellular functions. Because it contains the DBD, its inactivation prevents the interaction with DNA, leading to inactivation of the entire GABP complex. This was shown to result in early embryonic lethality in mice.[22,23] On the other hand, only preventing the assembly of the GABPα2β2 tetramer has little impact on cell cycle or apoptosis.[22] Thus, tetramer-forming GABPβ isoforms are seen as interesting therapeutic targets. It has been shown that targeting these isoforms can regulate the self-renewal of leukemic stem cells, without perturbing the survival of hematopoietic stem cells.[24]

Interestingly, the tetrameric GABP complex is unique among ETS factors, since it is an obligate heterodimeric protein composed of two structurally dissimilar subunits, GABPα and GABPβ.[25] GABPα contains an ETS DBD, while the transcriptional activation domain (TAD) is encoded by GABPβ. Consequently, both subunits are required for the transcription of target genes.[26] A crystal structure clearly shows the interaction between GABPα and GABPβ and their interaction with DNA (Figure 5). GABPα (blue) contains the DBD, which is characteristic for ETS transcription factors.[23,25] GABPβ (red) contains the TAD, but does not interact with DNA. GABPβ has also been shown to enhance the binding of GABPα to DNA by relieving auto-inhibition.[27] The amino terminus of GABPβ is responsible for the interaction with GABPα, and in one isoform, the carboxyl-terminus contains a leucine-zipper domain that leads to homodimerization. As a result, a heteromeric α2β2 tetramer is formed (Figure 6). The tetrameric complex is only formed when bound to DNA, whereas only the GABPα/GABPβ heterodimer can form in solution.[26]
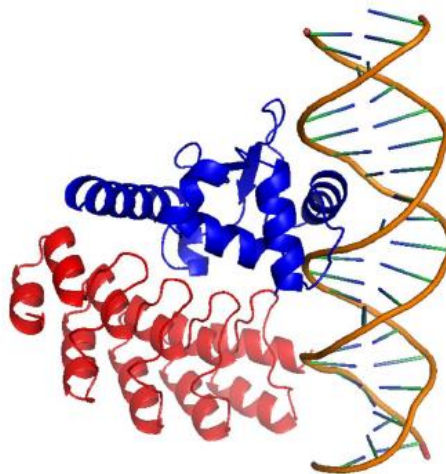
Figure 5: Crystal structure of GABPα and GABPβ bound to DNA (PDB code: 1AWC). Ribbon diagrams of GABPα (blue) and GABPβ (red) are shown together with a stick-model of DNA.
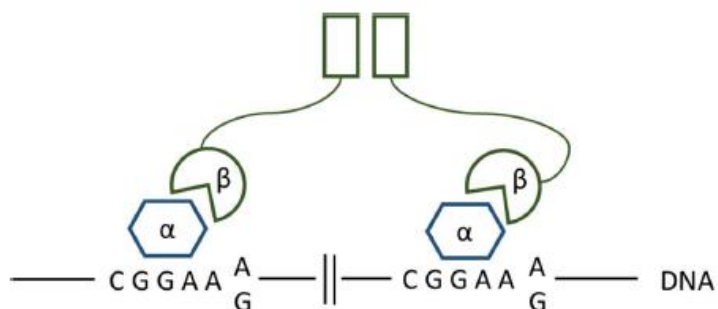
Figure 6: Model of complex formed by GABP and DNA. GABP binds to DNA sequences containing mainly the nucleotides adenine (A) and guanine (G). The α subunit is responsible for DNA binding and for recruitment of the β subunit. The β subunit does not bind to DNA, but interacts with and enhances the α subunit to bind to DNA.

1.4.1 Inhibition of GABP expression

Since GABP is driving the increased TERT expression and thus cellular immortalization, there is utility in the development of a drug that can target this protein for GBM therapy. Preliminary data showed that the inhibition of GABP expression in TERT promoter mutant cells led to a significant inhibition of GBM growth and glioblastoma stem cell survival. This data supports the hypothesis that the inhibition of GABP is a promising route that can be used to inhibit the expression of TERT in GBMs. Moreover, GABPα displays the canonical ETS domain fold but with a C-terminal extension that mediates binding to GABPβ whereas the GABPβ interaction domain is comprised of a series of ankyrin repeats. The binding surface on GABPβ displays a number of significant cavities that should be good sites for small molecule binding.[27,28] A careful mutagenesis study of the interface on GABPβ was carried out to identify critical hotspot residues, with a GABPβ phenylalanine 136 to alanine mutation resulting in a 1000-fold reduction in binding.[29] Since

GABPα and GABPβ must bind to each other to increase TERT expression, this project proposes to develop a protein-protein interaction inhibitor of GABPα and GABPβ as a therapy for GBM.

1.4.2 Using Small Molecules for Inhibition of Transcription Factors

Transcription factors are the controls a cell can use to regulate specific gene expression, which then determines a specific phenotype. To initiate transcription, transcription factors and RNA polymerase II will assemble on the promoter of the gene.[30] In cancer, it is frequent that transcription factors become overactive which provides a promising target for drug therapy. Ideally, a small molecule that disrupts the transcription factor activity will, in turn, disrupt the growth of cancer cells. In humans, it is estimated that between 5-10% of all gene-coding is performed by transcription factors.[31] This provides many opportunities for transcription factors to have overexpressed activity and thus, possible negative phenotypic outcomes.

Targeted small-molecule therapies offer a better alternative to the traditional cytotoxic chemotherapy. Some successful examples of small molecular therapies include using anti-oestrogens and antiandrogens to treat breast and prostate cancers, retinoic acid to treat acute promyelocytic leukaemia, and imatinib to treat chronic myeloid leukaemia.[32-34] In recent years, small molecules have been utilized to modify transcription factor activities, which has the potential to alter various types of cancer.[35] Ideally, small organic molecules will be cell-permeable and be able to specifically inhibit one transcription factor. One successful way of utilizing small molecules is by targeting ligand-dependent

transcription factors. Since these transcription factors are already dependent on a small molecule, fashioning a similar molecule to out-compete the physiological ligands proves a somewhat easier task. However, these small molecule inhibitors are only applicable to transcription factors dependent on binding to small molecule ligands for activity.[36,37]

In other examples, small molecules may be advantageous if they can bind to the DNA binding domain of the transcription factor. Ideally this would directly inhibit the transcription factor from binding to the DNA promoter region, thus altering the ability of the transcription factor to function. One recent example of this is the binding of a small molecule to the DNA binding domain B-ZIP proteins. These proteins are from a class of transcription factors that mediate a variety of signaling pathways that were identified as potential molecular targets for diseases including cancer and diabetes.[38] One problem with targeting the DNA binding domain is that a domain can adopt a different structure when it binds to DNA. So, it is difficult to fashion a small molecule to target that domain when it is relatively unstructured when not bound to DNA.

Similarly, nuclear hormone receptors also have a DNA binding domain. But they also contain a ligand binding domain which binds to hormones that recruit coactivator molecules to alter its activity to regulate gene expression.[39-41] Thus, it is easy to fashion a small molecule to target this ligand binding domain, since it is already structured to bind tightly to a different small molecule (hormones).[35] For example, it was shown that a small molecule inhibitor is potent against the interaction of thyroid hormone receptors and its coactivator.[42] These researchers were able to discover the first small molecule that could suppress transcription activity by irreversibly inhibiting coactivator binding. For this study,

it was relatively simple to screen compounds that resembled the structure of thyroid hormone when testing against the thyroid hormone receptor.

Small molecules can also be used against protein-protein interactions. Frequently transcription factors will be comprised of several proteins to complete their function. Fashioning small molecule inhibitors of these protein-protein interactions is another viable method to control the activity of the transcription factors. However, synthesizing a protein-protein interaction inhibitor proves to be much more difficult because the area to target on the protein binding interface is typically flat with little areas of binding pockets. Moreover, crystal structures of single proteins might not show an accurate structure of the protein when it is bound to other proteins, or it just may not show any deep and penetrable pockets for small organic molecules.[43] Other problems arise because small organic molecules are frequently not selective for one particular protein, but instead bind to several proteins, eliciting an unwanted response.

However, despite all of the challenges that arise when aiming to synthesize a protein-protein interaction inhibitor, there have been many successes. Chromosomal mutations that alter the Mixed Lineage Leukemia (MLL) gene lead to more aggressive types of leukemia. These mutations produce MLL transcription factor fusions that drive aberrantly high expression of target genes leading to disease.[44] To bind to DNA, there are two regions of MLL: an N-terminal motif that binds to coactivators, and a CXXC domain that binds to non-methylated cytosine guanine dinucleotide sites.[45] Recent research has developed small molecule protein-protein interaction inhibitors of the N-terminal motif and the coactivator menin. These small molecule inhibitors were successful in that they reduced

the amount of MLL binding to chromatin and reduced the expression of genes that are

drivers of this type of leukemia.[46]

1.5 Aim of Thesis

The goal of this project was to develop a novel therapy for GBM by (i) further

studying the structure of GABPα and GABPβ with standard triple resonance nuclear

magnetic resonance (NMR) experiments (HSQC, HNCACB, CBCA(CO)NH) to assign the

resonances in the spectrum to specific amino acids, (ii) using synthetic organic chemistry

methods to synthesize small molecule protein-protein interaction inhibitors, and (iii) testing

those small molecule inhibitors in a FRET molecular assay with fluorescently labeled

GABPα and GABPβ, and through chemical shift perturbations in ^{15}N-^{1}H heteronuclear

single quantum coherence NMR (HSQC) spectra.

Although there is already a crystal structure of the GABP complex, this is a crystal

structure of mus musculus GABPα and GABPβ already bound to DNA. It would be very

advantageous to have more extensive structural data on free GABPβ, free GABPα, the

GABPα- GABPβ complex, and each of those upon the addition of compounds. Seeing how

these two proteins interact before binding to DNA will allow for facile construction of an

inhibitor to disrupt this protein-protein interaction. Triple NMR methods make it possible

to show which amino acids each compound is interacting with. By expressing and purifying

the relevant domains of GABPα and GABPβ with ^{15}N labeling to record high quality NMR

spectra, we can confirm whether fragments are binding or not, as well as identify regions

of the protein where the compound binds elucidating whether these compounds operate in an allosteric or competitive manner.[47]

In-lab organic synthesis provides a more cost-efficient and quicker method then waiting for synthetic chemical companies to synthesize custom compounds. In addition, by synthesizing compounds, we can alter functional groups, allowing us to further explore the most ideal/potent compound to interact with GABPα and GABPβ.

Since fluorescence resonance energy transfer (FRET) relies on the energy transfer between donor and acceptor fluorophores within 10 nanometers from each other, this assay could be utilized to monitor the binding between GABPα and GABPβ. Because GABPα and GABPβ have been fused with their respective fluorophores, with Cerulean attached to GABPα and Venus attached to GABPβ, the energy transfer between the two fluorophores corresponds to the binding of GABPα and GABPβ. By measuring the emission of Cerulean and the emission of Venus, a FRET ratio is produced which is characteristic of the binding between GABPα and GABPβ. For HSQC NMR studies, protein is isotopically labeled with ^{15}N, thus, the amide NH resonances are visualized. Ideally, this produces a spectrum that has a well-dispersed peak for each amino acid in the protein. Because of lethality in the inhibition of GABPα, there was more of a focus to find compounds that bind to GABPβ, and therefore these NMR studies have been optimized for GABPβ.

References

1.   Wen, P. Y. and Kesari, S. Malignant gliomas in adults. *New England Journal of Medicine* **2008**, 359 (1), 492-507.

2.   Parsons, D.W., Joes, S., Zhang, X., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I., Gallia, G.L., Olivi, A., Mclendon, R., Rasheed, B.A., Keir, S., Nikolskaya, T., Nikolsky, Y., Busam, D.A., Tekleab, H., Diaz, L.A., Hartigan, J., Smith, D.R., Strausberg, R.L., Marie, S.K.N., Shinjo, S.M.O., Yan, H., Riggins, G.J., Bigner, D.D., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V.E., Kinzler, K.W. An Integrated Genomic Analysis of Human Glioblastoma Multiforme. *Science* **2008**, 312 (5897), 1807-1812.

3.   Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C., Morin, G.B., Harley, C.B., Shaw, J.W., Lichtsteiner, S., Wright, W.E. Extension of Life-Span by Introduction of Telomerase into Normal Human Cells. *Science*, **1998**, 279 (5349), 349-352.

4.   Killela, P.J.Reitman, Z.J., Jiao, Y., Bettegowda, C., Agrawal, N., Diaz, L.A., Friedman, A.H., Friedman, H., Gallia, G.L., Giovanella, B.C., Grollman, A.P., He, T., He, Y., Hruban, R.H., Jallo, G.I., Mandahl, N., Meeker, A. K., Mertens, F., Netto, G.J., Rasheed, B.A., Riggins, G.J., Rosenquist, T.A., Schiffman, M., Shih, L., Theodorescu, D., Torbenson, M.S., Velculescu, V.E., Wang, T., Wentzensen, N., Wood, L.D., Zhang, M., McLendon, R.E., Bigner, D.D., Kinzler, K.W., Vogelstein, B., Papadopoulos, N., Yan, H. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proceedings of the National Academy of Sciences of the United States of America*, **2013**, 110 (15), 6021-6026.

5.   Bell, R. J., Rube, H.T., Xavier-Magalhães, A., Costa, B.M., Mancini, A., Song, J.S., and Costello, J.F. Understanding TERT Promoter Mutations: A Common Path to Immortality. *Molecular Cancer Research* **2016,** 14 (4), 315-323.

6.   Makowski, M. M., Willems. E., Fang, J., Choi, J., Zhang, T., Jansen, P.W., Brown, K.M., and Vermeulen, M. An interaction proteomics survey of transcription factor binding at recurrent TERT promoter mutations. *Proteomics* **2016,** 16 (3), 417-426.

7.   Murray, A.W., Szostak, J.W. Construction of artificial chromosomes in yeast. *Nature* **1983**, 305 (1), 189-193.

8.   Brown, W.R.A., MacKinnon, P.J., Villasanté, A., Spurr, N., Buckle, V.J., Dobson, M.J. Structure and polymorphism of human telomere-associated DNA. *Cell* **1990**, 63 (1), 119-32.

9.   Blackburn, E.H., Szostak, J.W. The Molecular Structure of Centromeres and Telomeres. *Annual Review of Biochemistry* **1984**, 53 (1), 163-194.

10.   Zakian, V.A. Structure and function of telomeres. *Annual Review of Genetics* **1989**, 23 (1), 579-604.

11.   Cong, Y.S., Wright, W.E., Shay, J.W. Human telomerase and its regulation. *Microbiology and Molecular Biology Reviews* **2002**, 66 (3), 407-425.

12.   Harrington L. Biochemical aspects of telomerase function. *Cancer Letters* **2003**, 194 (2) 139-154.

13.   Yamaguchi, H., Calado, R.T., Ly, H., Kajigaya, S., Baerlocher, G.M., Chanock, S.J., Lansdorp, P.M., Young, N.S. Mutation in TERT, the Gene for Telomerase Reverse Transcriptase, in Aplastic Anemia. *New England Journal of Medicine* **2005**, 352 (1), 1413-1424.

14.   Blackburn, E.H. Structure and function of telomeres. *Nature* **1991**, 350 (6319), 569-573.

15.   Shay, J.W., Bacchetti, S. A Survey of Telomerase Activity in Human Cancer. *European Journal of Cancer* **1997**, 33 (5), 787-791.

16.   Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll,

I., Nagore, E., Hemminki, K., Schadendorf, D., Kumar, R. TERT promoter mutation in familial and sporadic melanoma. *Science* **2013**, 339 (6122), 959-961.

17. Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G.V., Chin, L., Garraway, L.A. Highly recurrent TERT promoter mutations in human melanoma. *Science* **2013**, 339 (6122), 957-959.

18. Thompson, C. C., Brown, T., and McKnight, A. S. L., Convergence of ETS- and notch-related structural motifs in a heteromeric DNA binding complex. *Science* **1991,** 253 (5021), 762-768.

19. Triezenberg, S. J., LaMarco, K. L., and McKnight, S. L. Evidence of DNA: protein interactions that mediate HSV-1 immediate early gene activation by VP16. *Genes & Development* **1988,** 2 (1), 730-742.

20. Bassuk, A. G. and Leiden, J. M. The role of ETS transcription factors in the development and function of the mammalian immune system. *Advances in Immunology* **1997**, 64 (1), 65-104.

21. Dittmer, J. and Nordheim, A. ETS transcription factors and human disease *Biochimica et Biophysica Acta* **1998,** 1377 (2), F1-F11.

22. Ristevski, S., O'Leary, D. A., Thornell, A. P., Owen, M. J., Kola, I., and Hertzog, P. J. The ETS transcription factor GABPalpha is essential for early embryogenesis. *Molecular Cellular Biology* **2004,** 24 (13), 5844-5849.

23. Xue, H. H., Bollenbacher, J., Rovella, V., Tripuraneni, R., Du, Y.B., Liu, C.Y., Williams, A., McCoy, J. P., and Leonard, W. J. GA binding protein regulates interleukin 7 receptor alpha chain gene expression in T cells. *Nature Immunology* **2004,** 5 (10), 1036-1044.

24. Yu, S., Jing, X., Colgan, J. D., Zhao, D.M., and Xue, H. H. Targeting tetramer-forming GABPβ isoforms impairs self-renewal of hematopoietic and leukemic stem cells. *Cell Stem Cell* **2012,** 11 (2), 207-219.

25. Sharrocks, A. D., Brown, A. L., Ling, Y., and Yates, P. R. The ETS-domain transcription factor family. *The International Journal of Biochemistry & Cell Biology* **1997,** 29 (12), 1371-1387.

26. LaMarco, K., Thompson, C. C., Byers, B. P., Walton, E. M., and McKnight, S. L. Identification of ETS- and notch-related subunits in GA binding protein. *Science* **1991,** 253 (5021), 789-792.

27. Batchelor, A. H., Piper, D. E., de la Brousse, F. C., McKnight, S. L., and Wolberger, C. The structure of GABPalpha/beta: an ETS domain- ankyrin repeat heterodimer bound to DNA. *Science* **1998,** 279 (5353), 1037-1041.

28. Chinenov, Y., Henzl, M., and Martin, M. E. The alpha and beta subunits of the GA-binding protein form a stable heterodimer in solution.: Revised model of heterotetrameric complex assembly. *Journal of Biological Chemistry* **2000,** 275 (1), 7749-7756.

29. Desrosiers, D. C. and Peng, Z. Y. A binding free energy hot spot in the ankyrin repeat protein GABPβ mediated protein-protein interaction. *Journal of Molecular Biology* **2005,** *354 (2)*, 375-384.

30. Berg, T. Inhibition of transcription factors with small organic molecules. *Current Opinion Chemical Biology* **2008**, 12 (4), 464-471.

31. Arndt, H.D. Small molecule modulators of transcription. *Angewandte Chemie* **2006**, 45 (28), 4552-4560.

32. Huang, M.E., Ye, Y.C., Chen, S.R., Chai, J.R., Lu, J.X., Zhoa, l., Gu, L.J., Wank, Z.Y. Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. *Blood* **1988,** 72 (2), 567-572.

33. O'Brien, S.G., Guilhot, F., Larson, R.A., Gathmann, I., Baccarani, M., Cervantes, F., Cornelissen, J.J., Fischer, T., Hochhaus, A., Hughes, T., Lechner, K., Nielsen, J.L., Rousselot,

P., Reiffers, J., Saglio, G., Shepherd, J., Simonsson, B., Gratwohl, A., Goldman, J.M., Kantarjian, H., Taylor, K., Verhoef, G., Bolton, A.E., Capdeville, R., Druker, B.J., Investigators, I.R.I.S. Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. *New England Journal of Medicine*, **2003**, 348 (1), 994-1004.

34.   Druker, B.J., Guilhot, F., O'Brien, S.G., Gathmann, I., Kantarjian, H., Gattermann, N., Deininger, M.W., Silver, R.T., Goldman, J.M., Stone, R.M., Cervantes, F., Hochhaus, A., Powell, B.L., Gabrilove, J.L., Rousselot, P., Reiffers, J., Cornelissen, J.J., Hughes, T., Agis, H., Fischer, T., Verhoef, G., Shepherd, J., Saglio, G., Gratwohl, A., Nielsen, J.L., Radich, J.P., Simonsson, B., Taylor, K., Baccarani, M., So, C., Letvak, L., Larson, R.A., IRIS Investigators, Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *New England Journal of* Medicine, **2006**, 355 (1), 2408-2417.

35.   Bushweller, J.H. Targeting transcription factors in cancer – from undruggable to reality. *Nature* **2019**.

36.   Aranda, A., Pascual, A. Nuclear Hormone Receptors and Gene Expression. *Physiological Reviews* **2001**, 81 (3), 1269-1304.

37.   McKenna, N.J., O'Malley, B.W. Combinatorial Control of Gene Expression by Nuclear Receptors and Coregulators. *Cell* **2002**, 108 (4), 465-474.

38.   Rishi, V., Potter, T., Laudeman, J., Reinhart, R., Silvers, T., Selby, M., Stevenson, T., Krosky, P., Stephen A.G., Acharya. A., Moll, J., Oh, W.J., Scudiero, D., Shoemaker, R.H., Vinson, C. A high-throughput florescence-anisotropy screen that identifies small molecule inhibitors of the DNA binding of B-ZIP transcription factors. *Analytical Biochemistry* **2005**, 340 (2), 259-271.

39.   Mangelsdorf, D.J. Thummel, C., Beato, M., Herrlich, P., Schuts, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., Evans, R.M. The Nuclear Receptor Superfamily: The Second Decade. *Cell* **1995**, 83 (6), 835-839.

40.   Robinson-Rechavi, M., Escriva Garcia, H., Laudet, V. The nuclear receptor superfamily. *Journal of Cellular Science* **2003**, 116 (1), 585-586.

41.   Pawlak, M., Lefebvre, P., Staels, B. General molecular biology and architecture of nuclear receptors. *Current Topics in Medicinal Chemistry* **2012**, 12 (6), 486-504.

42.   Arnold, L.A., Estébanez-Perpiñá, E., Togashi, M., Jouravel, N., Shelat, A., McReynolds, A.C., Mar, E., Nguyen, P., Baxter, J.D., Flettweick, R.J., Webb, P., Guy, R.K. Discovery of Small Molecule Inhibitors of the Interaction of the Thyroid Hormone Receptor with Transcriptional Coregulators. *Journal of Biological Chemistry* **2005**, 280 (52), 43048-43055.

43.   Arkin, M.R., Wells, J.A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nature Reviews* **2004**, 3 (1), 301-317.

44.   Slany, R. When epigenetics kills: MLL fusion protein in leukemia. *Hematological Oncology* **2005**, 23 (1), 1-9.

45.   Xu, C., Liu, K., Lei, M., Yang, A., Li, Y., Hughes, T., Min, J. DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. *Structure* **2018**, 26 (1), 85-95.

46.   Grembecka, J., He, S., Shi, A., Purohit, T., Muntean, A.G., Sorenson, R.J., Showalter, H.D., Murai, M.J., Belcher, A.M. Hartley, T., Hess, J.L., Cierpicki, T. Menin-MLL inhibitors reverse oncogenic activity of MLL fusion proteins in leukemia. *Nature Chemical Biology* **2012**, 8 (3), 277-284.

47.   Yuan, X., Davydova, N., Conte, M.R., Curry, S., Matthews, S. Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein. *Nucleic Acids Research* **2002**, 30 (2), 456-462.

Chapter 2: Methods

### 2.1 Fluorescence Resonance Energy Transfer

Fluorescence resonance energy transfer is a process by which energy is transferred from one fluorophore (the donor) to another (the acceptor). This process allows for an accurate measurement between 10 and 100 Å between two molecules. If the donor and acceptor are within 40-60 Å of each other, then they are at the distance where exactly half of the energy is transferred between the fluorophores, known as the Förster radius.[1] When the donor fluorophore is in an excited state, it passes this energy to the acceptor fluorophore through non-radiative dipole-dipole coupling. The efficiency of FRET is inversely proportional to the inverse sixth power of the distance between the two fluorophores using the equation:

$$E = \frac{R_0^6}{R_0^6 + R^6}$$

where E is efficiency, R_0 is the Förster radius and R is the distance between the donor and acceptor fluorophore.[2,3] Because FRET is so sensitive, it is a good physical process to use if you want to measure changes in the proximity of small molecules.

Since the desired biomolecules to study do not typically have an active fluorophore, it is standard practice to attach a type of green fluorescent protein (GFP) to the molecules of study. Since there are many analogues of GFPs, all with different excitation and emission wavelengths, it is best to choose a set of GFPs that have sufficient overlap in fluorophore one's emission wavelength and fluorophore two's excitation wavelength. However, you do not want so much overlap, where fluorophore one's excitation overlaps with either the excitation or emission of

fluorophore 2. Therefore, you are able to get reasonable separation in emission spectra between donor and acceptor fluorophores for obtaining accurate measurements.[4] Ideal overlap between two GFP fluorophores is visualized in Figure 7.
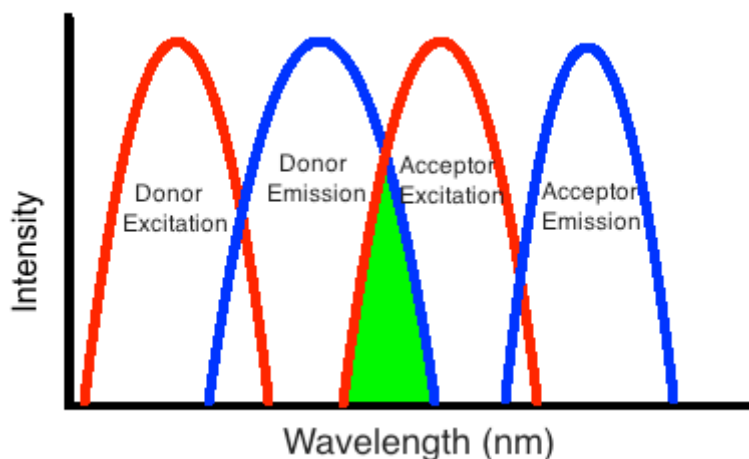
Figure 7: Ideal wavelength overlap needed for FRET. In this image, the area shaded in green represents the wavelength overlap of the donor emission and the acceptor excitation. In this depiction, it is clear to see that the donor excitation does not overlap with the acceptor excitation nor the acceptor emission.

DNA sequences coding for the relevant portions of GABPα and GABPβ have been cloned into vectors containing one of two green-fluorescent protein (GFP) fluorophores, Cerulean and Venus, as well as a His6-tag (Figure 8). The GFPs Cerulean and Venus were selected for this study because Cerulean has an excitation maximum at 433 nm and an emission maximum at 475 nm, while Venus as an excitation maximum at 515 nm and an emission maximum at 528 nm (Figure 9). This gives an ideal overlap in the emission of Cerulean (donor) and the excitation of Venus (acceptor), while not having overlap between Cerulean's excitation with Venus's excitation or emission. The polyhistidine tag was used for ease of purification through affinity purification methods. The relevant GABP regions used were the domains of

GABPα and GABPβ that interact with each other. For GAPBα, it was residues 316-434, and for GABPβ it was residues 1-167. The relevant portions of GABPα were cloned into a pET22b-Cerulean vector. For GABPβ, the relevant portions were cloned into a pET22b-Venus vector. In addition, this process was repeated for both relevant portions of GABPα and GABPβ into a pET22b vector containing only a His6-tag (no fluorophore-containing GFP). These unlabeled vectors were constructed to be used as controls for competition assays, as well as NMR studies. All vectors were validated through sequencing. After expression and purification of these constructs, the resulting proteins were also validated through mass spectrometry (Figure 10). A schematic representing the FRET competition assays in this study is represented in Figure 11.
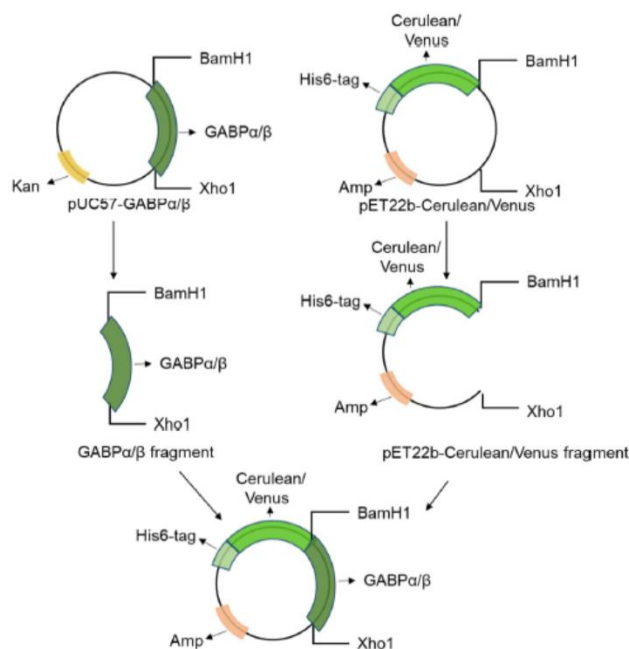
Figure 8. Schematic of the generation of Cerulean-GABPα and Venus-GABPβ constructs. DNA sequences encoding amino acid residues for the relevant portions of GABPα and GABPβ (residues 316-434 and residues 1-167 respectively) were cloned into a pET22b vector containing either the Cerulean or Venus fluorophore.
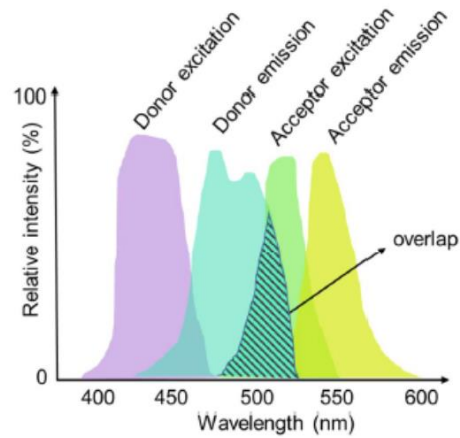
Figure 9: Excitation and emission spectra of Cerulean (donor) and Venus (acceptor). Cerulean has an excitation maximum at 433 nm and an emission maximum at 475 nm. FRET measurements require an overlap between the donor emission and the acceptor excitation (striped region). Venus's excitation maximum is at 515 and its emission maximum is measured at 528 nm.
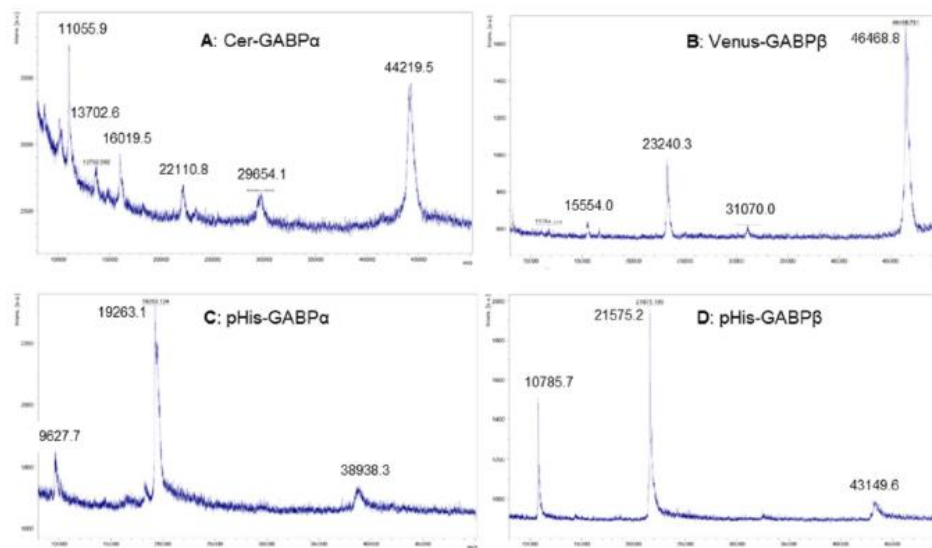
Figure 10: Mass spectrometry spectra of Cerulean GABPα, Venus GABPβ, pHis-GABPα, and pHis-GABPβ.
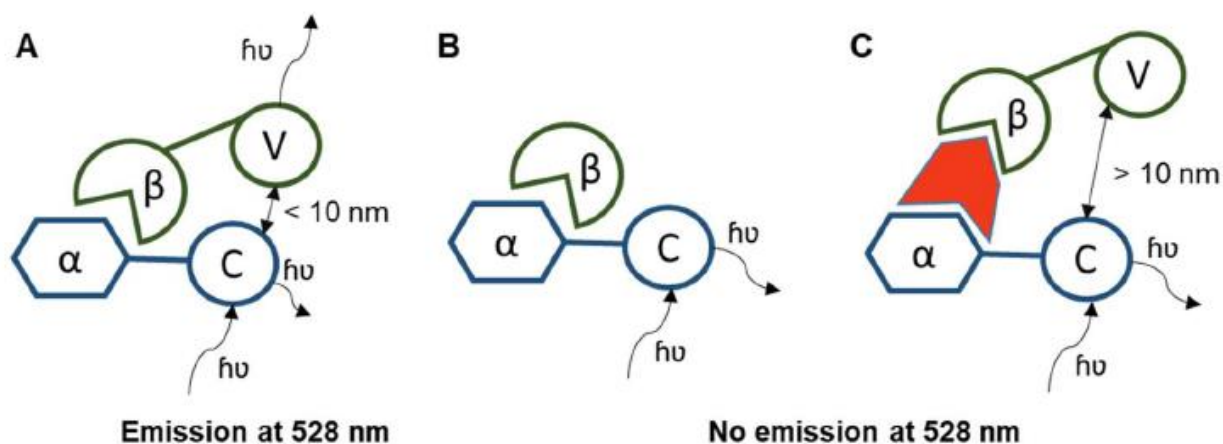
Figure 11: Illustration of the principle of competition assays and drug screens. A: The binding of Cerulean-GABPα (C-α) and Venus-GABPβ (V-β) leads to an energy transfer between the two fluorophores Cerulean (C) and Venus (V), resulting in an increasing emission at 528 nm. B: The binding of pHis-GABPβ (β) interrupts the binding between the two fluorescent proteins and thereby their energy transfer. C: A small molecule inhibits the association of GABPα and GABPβ. As soon as the distance between the two fluorophores is bigger than 10 nm, the energy transfer is interrupted. Consequently, the ratio of acceptor and donor emission wavelengths corresponds to the association of the two fluorescent proteins.

2.2 Nuclear Magnetic Resonance

2.2.1 Isotopic Labeling

Since NMR utilizes isotopes with an odd number of protons and neutrons, to be able to visualize protein structure with NMR, proteins must be isotopically labeled. Frequently, proteins are supplemented with ^{15}N for visualizing with NMR. Proteins can also be labeled with ^{13}C. In some cases, it is also advantageous to label protein with ^{2}H. Larger molecular weight proteins (MW > 20 kDa) suffer from low signal-to-noise ratios due to increased relaxation rates caused by slower tumbling of protein in solution. Moreover, with so many signals from the larger proteins, spectral resolution decreases.[5] To combat this, deuterium labeling improves the signal-to-noise ratio by suppressing spin

diffusion and decreasing relaxation rates for ^{13}C and ^{15}N spins. However, because of cost, most studies first label proteins with ^{15}N for NMR studies.

In this study, NMR was utilized for biological screening applications as well as briefly for NMR assignment and structure determination. These studies often require a large amount of isotopically-labeled protein which is a costly procedure in time and money. Proteins are expressed in the bacterial host, E. coli, and are grown in minimal media. Minimal media contains all of the ingredients necessary to supplement the growth for the bacteria cells. However, ammonium chloride, supplemented with ^{15}N is used.[6] What results is protein with all nitrogen isotopically labeled with ^{15}N. When analyzing with NMR, what results is a well-dispersed 2D spectrum with a signal for the amide backbone NHs as well as side chain NHs. Protein minimal media can also be supplemented with ^{13}C-labeled glucose and D_2O to grow a protein with multiple isotopically-labeled atoms. In this study, it was found that GABPβ has limited stability. Therefore, the time from the beginning of purification to concluding NMR studies was minimized as much as possible.

2.2.2 Heteronuclear Single Quantum Coherence

Heteronuclear single quantum coherence (HSQC) is a type of NMR experiment. This technique involves transferring magnetization from a proton to a second nucleus that is not a proton. For the purposes of this study, we studied the interactions between proton nuclei and ^{15}N nuclei. This transfer is achieved through an insensitive nuclei enhanced by polarization transfer (INEPT) step. INEPT transfers nuclear spin polarization from spins

with large Boltzmann coupling to those with low Boltzmann coupling. This gives a much more resolved 2D spectrum and enhances sensitivity.

The theory behind NMR revolves around the phenomenon that each nucleus has a nuclear spin magnetic moment. Each nucleus will have a different nuclear spin magnetic moment based on its environment, which gives rise to different signals in an NMR spectrum. Each nucleus acts like a small magnet with its own magnetic dipole or moment. NMR studies how this small magnetic moment interacts with a strong applied magnetic field. In an NMR sample, there is a large amount of magnetic spins, but what is observed on the spectrometer is an average of all of those small magnetic spins (Figure 12).[7] However, when the sample is first put into the NMR, there is no bulk magnetization because the average of all the random magnetic moments would be zero. After waiting a few moments and letting the sample come to equilibrium, a Boltzmann distribution is produced which selects those magnetization moments for a specific energy with orientations for a lower energy, which would be the magnetic moments that are aligning with the strong magnetic field (Figure 13).
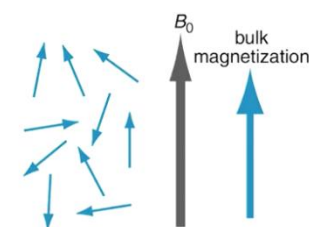
Figure 12: Average magnetic spin, comprised of small magnetic moments in an NMR sample. Small blue arrows represent the magnetic spins of all of the nuclei in an NMR sample. The large blue arrow represents the bulk magnetization which is the sum of the individual magnetic moments. B_0 is the strong magnetic field the NMR sample is placed in.
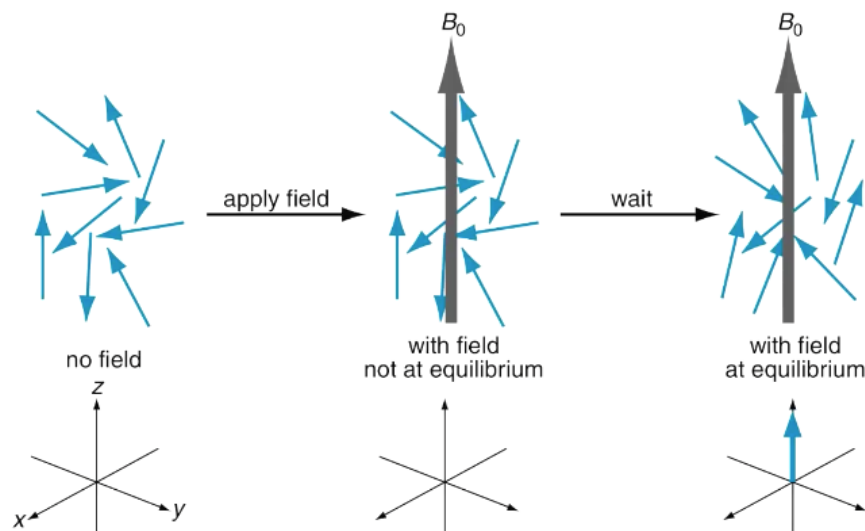
Figure 13: After waiting for a Boltzmann distribution, which selects those magnetization moments for a specific energy with orientations for a lower energy. In the first/left column, there is an NMR sample with no magnetic field applied. In the second/middle column, a magnetic field is applied but the magnetic moments average to 0, so there is no bulk magnetization from the NMR sample. In the third/right column, equilibrium has been met and via a Boltzmann distribution, those magnetization moments for a specific energy with orientations for a lower energy have been selected. So, the overall bulk magnetization is aligned in the Z-axis and with the applied strong magnetic field.

However, nothing can be observed when the bulk magnetization is in the Z-axis. So, to observe signals from the sample, radiofrequency (RF) pulse along the X-axis is applied. Since the strong applied magnetic field (B_0) is so large, to overcome this and push the bulk magnetization away from B_0, a weaker radiofrequency pulse along the X-axis in waves that oscillates at the Larmor frequency is applied. The Larmor frequency is the rate at which the small magnetic moments in each nuclei precess about the external magnetic field (Figure 14). By applying the RF pulse at the Larmor frequency, enough magnetic strength is garnered and the bulk magnetization is pushed away from the Z-axis (external magnetic field).
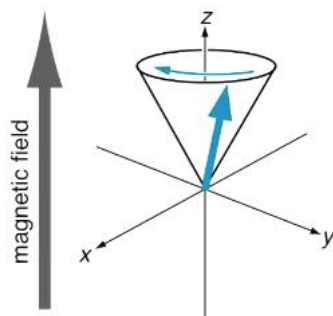
Figure 14: The Larmor frequency. The bulk magnetization will precess about the strong external magnetic field with a particular precessional frequency. This is referred to as the Larmor frequency.

To detect the sample, the Larmor frequency is observed because it is a value that changes over time. A coil is then placed over the sample, revolving around the X-axis. The precessing bulk magnetization cuts through the coil and induces a current which is what an NMR instrument detects. From having the bulk magnetization cut through the X-axis at a specific frequency that changes over time, a sine graph is produced which is called the free-induction decay/signal (FID) (Figure 15). This signal is digitized and then Fourier-transformed. The Fourier transform is a tool that breaks a waveform (NMR signal) into an alternate representation, characterized by sine and cosine functions.[8] In practice, this method takes the waveform of a FID and transforms it into a sum of sinusoidal functions.
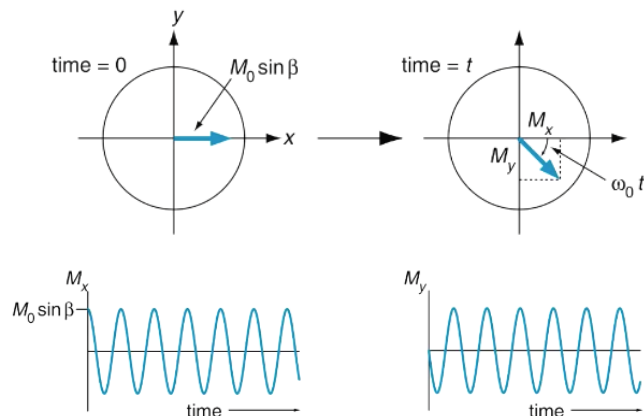
$$M_x = M_0 \sin\beta \cos\omega_0 t \qquad M_y = -M_0 \sin\beta \sin\omega_0 t$$

Figure 15: The free-induction signal (FID). The bulk magnetization cutting through the X-axis produces a frequency of signals that results in a sine graph, or the free-induction signal.

A frequent process that NMR experiments employ is a spin-echo (Figure 16). In a spin-echo, there is usually a 90° RF pulse, then a delay. After another 180° RF pulse, there is another delay of the same time. Then the signal observed is recorded. At the end of this sequence, no matter what the relaxation time or offset it, the magnetization vector is always in the same place. It is as if the magnetization has not precessed at all.[7] Essentially, this experiment refocuses the spins of the nuclei that have been scattered by constant field distortions and inhomogeneities, which can cause the spins to precess at different rates.
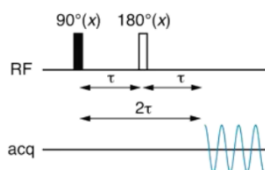
Figure 16: Spin-echo NMR experiment pulse sequence. In a spin echo, there is first, a 90° RF pulse (black box). Then there is a relaxation time (τ). Then a 180° RF pulse (white box) is applied, and then another relaxation time of the same length is applied. Then the signal observed is recorded.

For an HSQC, the process starts with an initial 90° pulse on the proton. Then in period A (Figure 17), which is a spin echo, a 180° pulse is applied to both spins of the proton and nitrogen right in the middle of the first relaxation time, τ. The offsets of both nuclei's spins will be refocused, but the coupling between them will evolve throughout the experiment. Coherence is then transferred directly to the ^{15}N nucleus by simultaneous 90° pulses on each nucleus. Period B is a spin echo in which a 180° pulse is applied to just the proton. Thus, the offset of the proton is refocused, and only the offset of spin 2 affects the evolution. Then there is a 90° pulse on both spins to transfer the magnetization back to the proton via an antiphase magnetization. Antiphase magnetization is observed when two coupled spins have free evolution. Antiphase peaks cannot be converted to in-phase peaks with phase correction after acquiring data. The final spin echo (period C), converts this antiphase term into an in-phase proton magnetization. After a delay, the observed signal is acquired. [9] After the Fourier transformation, what results is a 2D spectrum with a peak for each interaction between a ^1H and a ^{15}N. Ideally, there are well-dispersed peaks for each backbone NHs and sidechain NHs when taking an HSQC on a protein.

This method can be employed upon the addition of compounds that bind to a protein. After assigning amino acids to each peak (see next section below) it is easy to see which amino acids a compound is binding to, based on chemical shift perturbations of the resulting amino acid peaks. This is a very sensitive method to detect perturbations in the environment of backbone amides due to ligand binding. Moreover, it is frequent to see perturbations in the side-chain peaks, but this could also be due to a conformational change in the protein from ligand binding. Since there are chemical shift perturbations from both the ^1H and ^{15}N aspects

of a peak, a simple equation is used to calculate an overall change in chemical shift. The absolute values of chemical shifts for ^{1}H and ^{14}N are combined in this equation:

$$\Delta \delta \, (^{15}N + {}^1H) = | \, (\Delta \delta \, {}^{15}N \, / \, 4.69) + \Delta \delta \, {}^1H \, |$$

Although chemical shift perturbations are all relative to each protein and ligand system, it is typically accepted that chemical shift perturbations ($\Delta \delta \, (^{15}N + {}^1H)$) that are over 0.1 are worth noting as a possible ligand-binding site.[10]
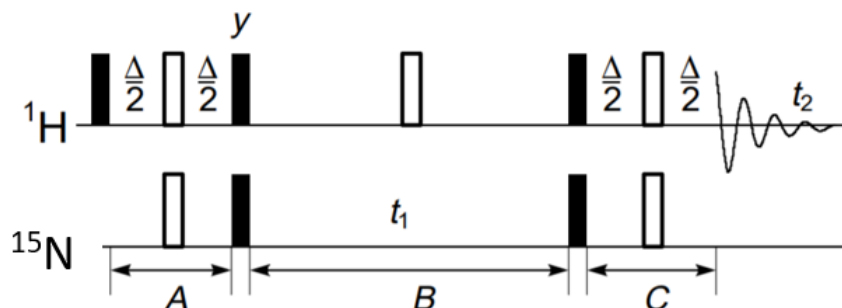
Figure 17: HSQC NMR experiment pulse sequence. Filled rectangles represent 90° pulses and open rectangles represent 180° pulses. The time delay, τ, is set to $1/(2J_{HN})$, which is 1 over 2 x the coupling of H to N.

2.2.3 CBCANH / HNCACB

While HSQC's explore the coupling between ^{1}H and ^{15}N, this does not give the resonance assignment of the protein and only allows you to see which amino acids are perturbed when performing a drug screen. Triple resonance experiments must be performed to assign an amino acid to each peak in the HSQC. To do this, the protein of interest must be isotopically labeled with ^{15}N and ^{13}C. In this NMR experiment,

magnetization is transferred from $^1H\alpha$ and $^1H\beta$ to $^{13}C\alpha$ and $^{13}C\beta$. And then magnetization

is transferred from $^{13}C\beta$ to $^{13}C\alpha$. Then it is transferred to $^{15}N^H$ and then to $^1H^N$ for detection.

Thus, for each NH group, there are two Cα and Cβ peaks. The chemical shift is evolved

simultaneously on $^{13}C\alpha$ and $^{13}C\beta$, so these appear in one dimension. The other two

dimensions are $^{15}N^H$ and $^1H^N$. A schematic for the entire magnetization transfer is

visualized in Figure 18.[11]
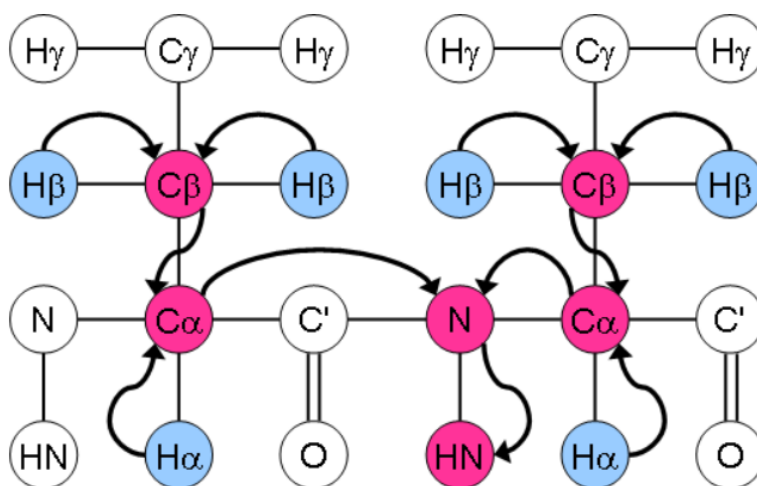
Figure 18: Magnetization transfer in a CBCANH / HNCACB NMR experiment. This represents a section of the amino acid backbone of a protein. The left half of the image represents one amino acid, while the right half represents an adjacent amino acid.

What results from these experiments are images of "strips" where four peaks are

visible: two peaks are from one residue, and two peaks are from an adjacent residue. In the

adjacent residue, the peaks are typically much weaker. To identify these, another NMR

experiment can be run, CBCA(CO)NH. In this experiment, the Cα and Cβ peaks are

actually in opposite phases, thus making it much easier to identify which carbon is alpha

and beta. By matching these strips of four peaks together, it is then straightforward to assign each peak in the HSQC spectrum to a specific amino acid, provided that an accurate amino acid sequence for the protein of interest exists.[12]

2.3 Mutated GABPβ

While NMR offers a way to rule out false positives from the FRET assay, another experiment can be run, where mutated forms of GABP are grown to see how it affects the FRET assay. This offers an alternative way to rule out false positives from the assay, as well as confirm where compounds are binding on the protein.

Recently, the irreversible kinase inhibitors afatinib (targeting EGFR) and ibrutinib (targeting Bruton's tyrosine kinase) were approved by the FDA for non-small cell lung cancer and chronic lymphocytic leukemia. These inhibitors covalently react with cysteines on the target kinase in each case. Cysteines possess unique reactivity due to their nucleophilic thiol group. However, despite its reactivity, cysteine is one of the least common amino acids found in proteins. These two properties together make cysteines a viable and popular covalent drug target.[13-15] Covalent drugs are inhibitors that are designed to bind covalently and irreversibly to their target protein. These inhibitors will typically have a functional group with low chemical reactivity so that it will bind specifically and quickly to its targeted nucleophilic residue on the protein.[16] This directly contrasts reversible inhibitors that are in equilibrium with their targets, constantly binding and unbinding. Some advantages with covalent inhibitors are that they can provide a high level of potency and selectivity. In addition, since covalent inhibitors are irreversible, the

restoration of the pharmacological activity after the inhibitor binds requires re-synthesis of the protein target. This allows for lower dosing and longer duration of action of the drug.[17] This covalent addition to the protein increases the potency of the inhibitor. Moreover, it can increase the duration of inhibition because recovery of activity requires synthesis of new protein.

As there are cysteine residues located at the interface for binding of GABPα to GABPβ, a 1,000-compound library of cysteine reactive molecules (Enamine) was screened for inhibition of GABPα-GABPβ binding. This screen identified two compounds that show significant inhibition ($IC_{50} < 150\ \mu M$) using the FRET assay. Other targets with this library were screened and these two compounds were not active in the other screens, suggesting they have selectivity for GABP.

In order to rule out false-positives from the FRET assay drug screen, it is essential to confirm binding of these compounds to one of the two target proteins (GABPα or GABPβ). There are several methods to do this. Growing a mutated form of GABPα or GABPβ where a cysteine has been mutated to an alanine would be one way to visualize where cysteine reactive compounds are binding. Since the binding surface displays a number of significant cavities that should be good sites for small molecule binding, and only contains one cysteine, a mutated form of GABPβ where the one cysteine was mutated to an alanine was expressed and purified. This mutant form of GABPβ was validated through sequencing. Since previous studies showed that the inhibition of GABPα leads to early embryonic lethality[18,19] there is more of a focus to find a target that binds to GABPβ, and therefore it is essential to confirm whom these targets are binding.

2.3 Experimental Methods and Materials

2.4.1 Materials

Unless otherwise noted, commercially obtained reagents were used as received. Synthetic reactions were done under an atmosphere of argon. Progress of reactions was monitored by TLC performed on Analtech 250-micron silica gel GF plates visualized with 254 nm UV light and also by mass spectrometry using a Waters single-quadrupole LCMS. All compounds were purified on Biotage Isolera Four Flash Chromatography system using SNAP cartridges. [1]H spectra were recorded on Bruker NMR spectrometers operating at 800 MHz in $CDCl_3$ and DMSO-d_6, with TMS as internal standard. Chemical shift values are reported in δ ppm units. Mass spectra were recorded on a Micromass AutoSpec Ultima Magnetic sector mass spectrometer in positive ESI mode.

Cell lines used were E. coli K12, Turbo competent cells from New England Biolabs (Ipswich, MA, USA) and Arctic Express Cells from Agilent Technologies (Santa Clara, CA, USA. Enzyme materials that were used included quick ligase buffer, quick ligase, Smartcut buffer, BamH1, and Xho1. All were purchased from New England BioLabs (Ipswich, MA, USA). Fragment library screening was done with Ro1500 Diversity Maybridge Fragment Library from Thermo Fisher Scientific (Waltham, MA, USA) and a customized fragment library from ChemDiv (San Diego, CA, USA).

For growing unlabeled protein, terrific broth, purchased from Fisher Bioreagents (Hampton, NH, USA) was used. For growing isotopically-labeled proteins, a minimal media was used. This consisted of 45 mM Na_2HPO_4, 15 mM KH_2PO_4 9 mM NaCl, 0.3 mM

Na_2SO_4 1 mM $MgSO_4$ 0.3 mM $CaCl_2$ 0.5% (w/v) glucose, 0.1% (w/v) biotin, 0.1% (w/v) thiamin, 19 mM $^{15}NH_4Cl$, and 0.5% (v/v) ^{15}N Bioexpress growth media (Cambridge Isotope Laboratories, Tewksbury, MA, USA).

For protein purification, lysis, wash, and elution buffers were utilized. Unless otherwise stated, all buffers were kept at pH: 7.5 and degassed before use. The lysis buffer consisted of 600 mM KCl, 50 mM tris HCl (pH: 8.0), and 0.01% (w/v) NaN_3. The protein wash buffer included 1 M KCl, 50 mM Tris/HCl, pH 8.0), and 0.01% (w/v) NaN_3. The protein elution buffer was 600 mM KCl, 50 mM tris HCl (pH: 8.0), 300 mM imidazole, and 0.01% (w/v) NaN_3. For FRET assays, the buffer used was 50 mM Na_2HPO_4/NaH_2PO_4 (pH 7.9), 300 mM NaCl, 1 mM, EDTA, and 1 mM DTT. For protein NMR studies, buffer used was 300 mM arginine, 300 mM glutamic acid, and 1 mM DTT.

For gel electrophoresis staining, Coomassie Brilliant Blue solution was used which consisted of 45% (v/v) methanol, 10% (v/v) acetic acid, and 0.25% (w/v) Coomassie Brilliant Blue. To load samples onto the gel, a loading buffer was used which consisted of 50% (v/v) distilled water, 12.5% (v/v) 0.5 M tris HCl (pH:6.8), and 10% (v/v) glycerol, 20% (v/v) of 10% (w/v) SDS, 5% v/v 0.8 M DTT, and 0.25% (w/v) bromophenol blue. Running buffer that was used was 0.328 (w/v) tris HCl base, 14.41% (w/v) glycine, and 1% (w/v) SDS. This buffer was then diluted by a factor of 10 for use.

2.4.2 Cloning, Expression and Purification of GABP

Transformation: Competent E. coli cells (stored in -80°C) were thawed on ice for 5-30 minutes. Into the competent cells, 50 ng of desired plasmid (GABP) was added (To

increase efficiency in Arctic Express cells, 2 µL of 10% (v/v) of β-mercaptoethanol was added). The plasmid and E. coli mixture was incubated on ice for 5-30 minutes. This mixture was then heat-shocked in a 42°C water bath for 45 seconds. The mixture was then placed back on ice for 5 minutes to reduce any damage to the cells. To this mixture, 300 µL of SOC broth (New England Biolabs, Ipswich, MA, US) was added and incubated for 1 hour at 37°C. The mixture was then spread on LB agar plates containing ampicillin. These plates were then grown upside-down overnight in a 37°C incubator.

Plasmid Extraction: E. coli cells containing a desired plasmid were centrifuged (4000 rpm, 20 min, 4°C). After the supernatant was discarded, the cells were resuspended in 250 µL of resuspension buffer. After adding 250 µL of lysis buffer, the tubes were lightly inverted and incubated at room temperature for 5 minutes. After adding 350 µL of neutralization buffer and inverting lightly, the cell fragments were centrifuged. (13,200 rpm, 10 min, 4°C). The supernatant was transferred to *QIAprep* 2.0 Spin Columns (*QIAGEN*, Netherlands). The resulting plasmids were washed with 750 µL of wash buffer. The plasmids were then eluted from the columns by adding 20 µL of nuclease-free water (*Thermo Fisher Scientific*, US) and centrifuging (13,200 rpm 1 min, 4°C).

Plasmid Digestion: To a solution of 23 µL of plasmid, 3 µL of CutSmart Buffer (*New England Biolabs*, US), 2 µL of BAMH1 (*New England Biolabs*, US), and 2 µL of Xho1 (*New England Biolabs*, US) were added. This mixture was added to a thermocycler (*Bio-Rad Laboratories*, USA) which heated the mixture to 37°C for 1 hour and then cooled to 4°C. To this mixture, 10 µL of DNA Loading Buffer was added. The mixture was then run through a 1% (w/v) agarose gel containing 0.01% ethidium bromide. The gel

containing DNA and a 1 kb DNA ladder (*New England Biolabs*, US) was run (100V, 1 hour). The DNA bands were cut out of the gel and extracted at 50°C with 600 µL of solubilization buffer (*QIAGEN*, Netherlands). After adding 200 µL of 2-propanol, the mixture was added to a *QIAprep* 2.0 Spin Column (*QIAGEN*, Netherlands) and centrifuged (13,200 rpm 1 min, 20°C). The DNA was washed with 750 µL of wash buffer and eluted with 20 µL of nuclease-free water (*Thermo Fisher Scientific*, US) and centrifuging (13,200 rpm 1 min, 4°C).

Ligation: A mixture containing 5 µL of ligase buffer, 2 µL insert, 2 µL of vector, and 1 µL quick ligase was incubated at room temperature for 10 minutes. To a tube containing 50 µL of Arctic Express (DE3) E. coli Competent Cells (*Agilent*, USA), 2 µL of the ligation mixture was added. This mixture was incubated at 4°C for 30 minutes and then heat shocked at 42°C for 45 seconds. After another incubation of 5 minutes at 4°C, 300 µL of SOC Outgrowth Medium (*New England Biolabs*, US) was added. The transformed cells were then incubated on LB agar plates containing 0.01% (w/v) ampicillin at 37°C until colonies began to form.

Protein Expression: A single colony of cells was added to a 100 mL solution of terrific broth media containing 0.01% (w/v) ampicillin. After incubating at 37°C with shaking overnight, the cells were added to 1000 mL of terrific broth media containing 0.01% (w/v) ampicillin. The cells were grown at 37°C until they reached OD_{600}=0.75-0.90. After the cells were cooled down to 10°C, 0.25 g of isopropyl β-D-1-thiogalactopyranoside (IPTG) (*Gold Biotechnology*, US) was added. The cells induced for 16-18 hours and were then centrifuged (4,000 rpm 25 min, 4°C). After the supernatant was discarded, the cell

pellet was resuspended in lysis buffer containing a Protease Inhibitor Cocktail tablet (*Roche, Switzerland*). The cells were then lysed using a cell disruptor instrument (*Avestin*, Canada) with a pressure at about 12,500 psi. Cell contents were removed through centrifugation (35,000 rpm 45 min, 4°C).

His-tag Purification: Lysed and centrifuged protein supernatant was loaded onto nickel-nitrilotriacetic acid (Ni-NTA) agarose columns. After letting the supernatant load through gravity, unspecific bound proteins were removed by washing until the OD_{600} was under 0.1. To elute the protein, a solution containing 50 µL of 1 M dithiothreitol (DTT) and 100 µL 0.5 ethylenediaminetetraacetic acid (EDTA), 2 mL of elution buffer was placed under the column. The column was then washed with elution buffer with the elution collected in the DTT and EDTA mixture underneath, until the OD_{600} was under 0.1. To the eluted mixture, 0.02% (v/v) of Pierce Universal Nuclease (*Thermo Fisher Scientific*, US) was added to degrade any remaining nucleic acids.

Size Exclusion Purification: A column packed with Sephacryl S-100 High Resolution Size Exclusion Chromatography Resin (*GE Healthcare*, US) was flushed with a degassed binding buffer and was attached to an AKTAprime plus instrument (*GE Healthcare*, US). The protein was concentrated to 2 mL and loaded onto the instrument. Fractions were collected and analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

SDS-PAGE: Into the wells of a Mini-PROTEAN® TGX™ Precast Gel (*Bio-Rad*, US) 10 µL of protein mixture was added (10 µL Laemmli Buffer, and 15 µL desired

protein). The gel was run (275 V, 15 min) with a protein marker (2-212 kDa, *New England Biolabs*, US). The gel was stained with Coomassie Brilliant Blue for analysis.

Mass-Spectrometry: All samples were analyzed in the Biomolecular Analysis Facility (*University of Virginia*, US) by utilizing matrix-assisted laser desorption ionization (MALDI) and a time-of-flight spectrometer. Sinapinic acid was used as an acidic matrix, and bovine-serum albumin (BSA) was used as a standard. When reacting with compounds, protein samples were equilibrated with compounds for 2 hours and then desalted through PD MiniTrap G-25 columns (GE Healthcare, Chicago, IL, US) before mass analysis.

GABPβ Mutant	Forward Primer	Reverse Primer
C139A	5'- CCCAGAGCAAGTTTGCGA AGACCGCCTTCGAC -3'	5'- GTCGAAGGCGGTCTTCG CAAACTTGCTCTGGG -3'

Table 1: Primers used for the creation of GABPβ cysteine-free mutant used in this study.

2.4.3 FRET Assay

For all FRET assays, a PHERAstar FSX Microplate Reader (*BMG Labtech*, US) was utilized, as well as nonbinding surface plates (Corning, Corning, NY, US). All samples were incubated for 30 minutes at room temperature in the dark when solutions were made, and again once the plates were loaded. For all assays, light at 430 nm was applied as well as emission measured at both 470 nm and 530 nm. For K_D determinations, 12-point serial dilutions were made with a starting protein concentration of 500 nM of GABPα and 500 nM of GABPβ. For IC_{50} determinations, 17-point serial dilutions were made with a protein

concentration of three times the K_D and a starting compound concentration of 5000 µM dissolved in dimethyl sulfoxide (DMSO). The amount of DMSO solution in protein solution was kept at 5% (v/v). For FRET-based single point drug screens, using a precision pipetting system (*BioTek* Instruments, US), 5 µL of the Maybridge drug fragments were added to 95 µL of protein held at 3x the K_D. For the ChemDiv library, 1 µL of the fragments were mixed with 19 µL of protein solution by the precision pipetting system.

2.4.4 Acquiring NMR

GABPβ was grown in minimal media supplemented with ^{15}N-labeled ammonium sulfate. After the protein was expressed and purified (see above), it was concentrated to an amount above 150 µM. In all NMR spectra of compounds, DMSO solutions were added (5% v/v) to a final concentration of 5000 µM. All samples contained D_2O (5% v/v). All spectra were taken with shaped Shigemi tubes and a Bruker Avance III 800 with 5mm HCN Zpfg cryoprobe (*Bruker,* US) for 2 hours at 37°C. Data was processed with NMR Pipe (*NIST*, US) and visualized with CCP NMR.

2.4.5 Mutated Proteins

Quick Change Reaction: The desired protein plasmid (GABPβ) was thawed on ice. Nuclease-free deionized water was added to the ordered primers until a concentration of 100 µM was reached. A mix of forward primer (0.5%, v/v), reverse primer (0.5%, v/v), GABPβ plasmid (10%, v/v), dimethyl sulfoxide (DMSO) (5%, v/v), nuclease-free deionized water (34%, v/v), and Quick-Load Taq 2X Master Mix (New England Biolabs,

Ipswich, MA, US) (50%, v/v) was made. This mixture was put into a Bio-Rad T100 Thermal Cycler (Bio-Rad, Hercules, CA, US). This mixture was then heated to 98°C for 30 seconds. Then the mixture completed 25 cycles of 98°C (10 seconds), 52-72°C (20 seconds), and 72°C (10 minutes). The temperature was then kept at 72°C for 2 minutes and then held at 4°C overnight. All mutated plasmids were then transformed, digested, and analyzed for sequencing to verify that the mutation was successful. Then the plasmids were expressed and purified based on the methods described above.

2.4.6 Organic Synthesis

AM-2-37
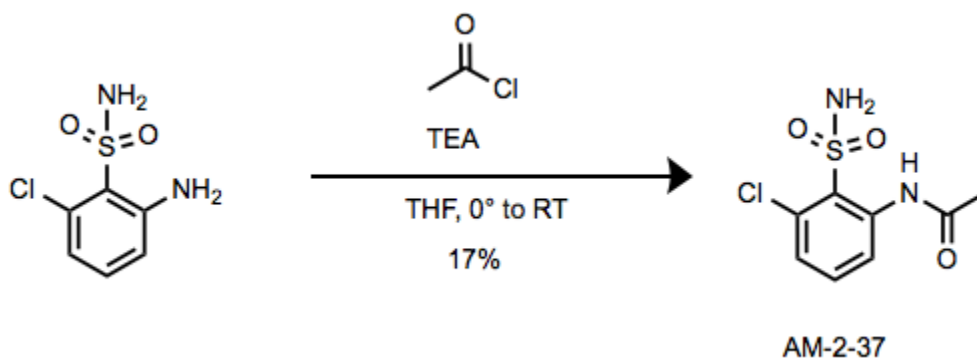
2-amino-6-chlorobenzene-1-sulfonamide was purchased from Enamine (Monmouth, NJ, USA). To a solution of 2-amino-6-chlorobenzene-1-sulfonamide (73 mg, 0.35 mmol, 1 equivalent) and acetone (20 mL) at 0°C, triethylamine (74 μL, 0.53 mmol, 1.5 equivalent) was added. After stirring for 20 minutes at room temperature, acetyl chloride (38 μL, 0.53 mmol, 1.5 equivalents) was added. This stirred for 3 hours while letting the temperature rise to room temperature. The reaction then stirred overnight. The

reaction was quenched with deionized water and extracted with diethyl ether. Residual solvent was removed under reduced pressure and the yellow oil was purified by flash chromatography (12.6 mg, 17% yield).[20] **1H NMR** (DMSO-$d6$, 25 °C, 800 MHz): δ 10.39 (s, 1H); 8.29 (dd, J= 11.2, 1.6 Hz, 1H); 7.93 (s, 2H); 7.52 (t, J= 10.9 Hz, 1H); 7.36 (dd, J= 10.6, 1.68 Hz, 1H); 2.11 (s, 3H) ppm.

AM-2-37

AM-1-143
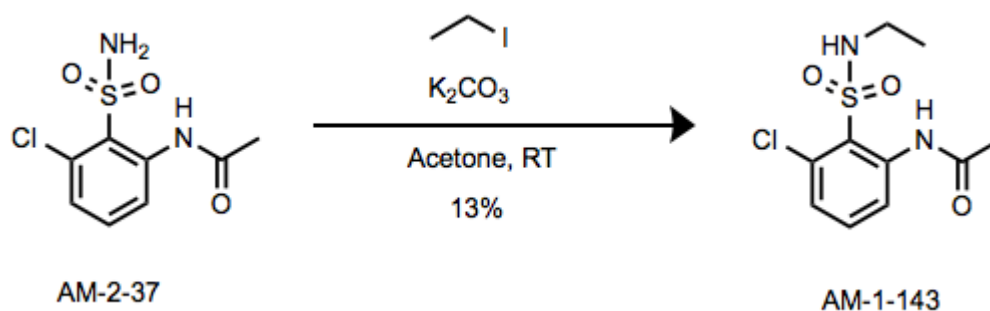
To a solution of AM-2-37 (10.4 mg, 41.9 µmol, 1 equivalent) and acetone (3 mL), K_2CO_3 (8.69 mg, 62.9 µmol, 1.5 equivalent) was added. After stirring for 5 minutes at room temperature, iodoethane (3.37 µL, 41.9 µmol, 1 equivalent) was added. This stirred at room temperature overnight. Residual solvent was removed under reduced pressure and the yellow oil was purified by flash chromatography (1.5 mg, 13% yield).[21] **1H NMR** (DMSO-$d6$, 25 °C, 800 MHz): δ 10.54 (s, 1H); 8.71 (dd, J= 11.6, 1.8 Hz, 1H); 7.43 (t, J= 10.7 Hz, 1H); 7.23 (dd, J= 10.6, 1.6 Hz, 1H); 3.04 (q, J= 8.1 Hz, 2H); 2.23 (s, 3H); 1.15 (t, J= 9.7 Hz, 3H) ppm.

AM-1-153
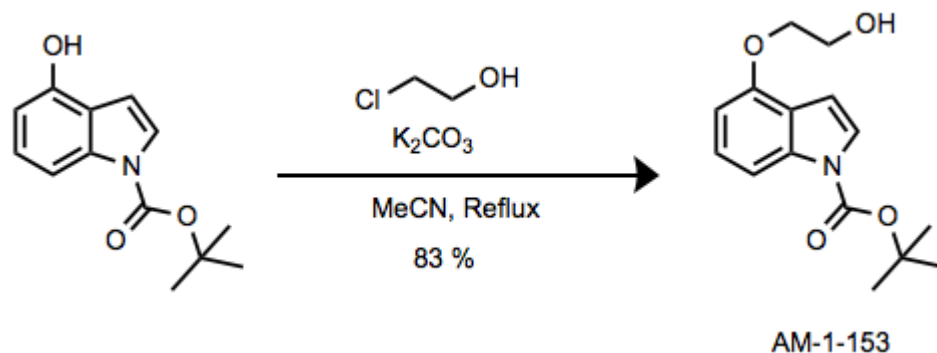
N-boc-4-hydroxyindole was purchased from 1-Click Chemistry (Kendall Park, NJ, USA). To a solution of n-boc-4-hydroxyindole (0.233 g, 1 mmol, 1 equivalent) and MeCN (10 mL), K_2CO_3 (0.7 g, 5 mmol, 5 equivalents) was added. After stirring for 1 hour at room temperature, iodoethane (70 µL, 1 mmol, 1 equivalent) was added. This stirred at reflux overnight. Residual solvent was removed under reduced pressure and the crude oil was purified by flash chromatography (0.1742 g, 63% yield).[22] **1H NMR** (DMSO-d6, 25 °C, 800 MHz): δ 7.78 (s, 1H); 7.51 (s, 1H); 7.22 (t, J= 8.1 Hz, 1H); 6.68 (d, J= 7.9 Hz, 2H); 4.23 (t, J= 4.4 Hz, 2H); 4.04 (t, J= 4.5 Hz, 2H); 1.67 (s, 9H) ppm.

AM-1-153   AM-2-3
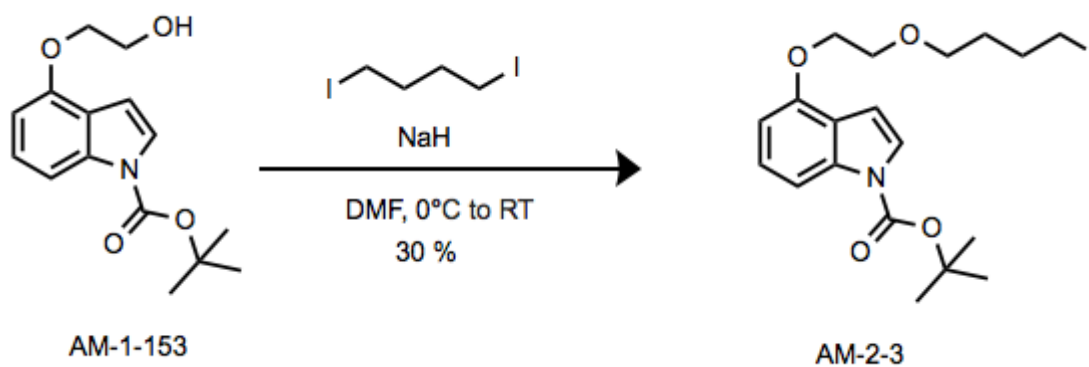
To a solution of AM-1-153 (0.2672 g, 1 mmol, 1 equivalent) and DMF (10 mL) at 0°C, NaH (0.7 g, 5 mmol, 5 equivalents) was added. After stirring for 30 minutes at 0°C,

1,4-diiodobutane (254 µL, 2 mmol, 2 equivalents) was added. This stirred for 3 hours while letting the temperature rise to room temperature. Residual solvent was removed under reduced pressure and the crude oil was purified by flash chromatography (0.1328 g, 30% yield).[23] **1H NMR** (DMSO-*d*6, 25 °C, 800 MHz): δ 7.76 (d, *J*= 10.6 Hz, 1H); 7.49 (d, *J*= 5.0 Hz, 1H); 7.20 (t, *J*= 10.9 Hz, 1H); 6.71 (dd, *J*= 4.9, 1.0 Hz, 1H); 6.64 (d, *J*= 10.3 Hz, 1H); 4.50 (t, *J*= 6.2 Hz, 2H); 4.31 (t, *J*= 6.8 Hz, 2H); 4.01 (t, *J*= 8.5 Hz, 3H); 3.21 (t, *J*= 9.2 Hz, 3H); 1.91 (m, *J*= 9.5 Hz, 2H); 1.75 (d, *J*= 10.6 Hz, 2H); 1.67 (s, 9H) ppm.
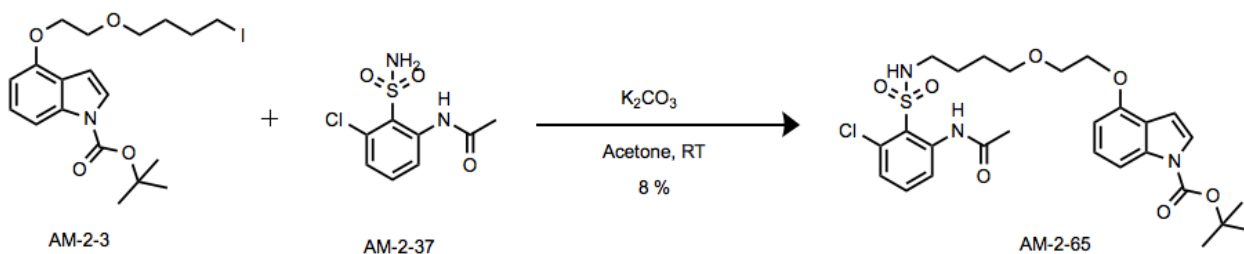
To a solution of AM-2-37 (40 mg, 0.161 mmol, 1 equivalent) and acetone (3 mL) at 0°C, K₂CO₃ (35 mg, 0.2093 mmol, 1.3 equivalents) was added. After stirring for 5 minutes at room temperature, AM-2-3 (0.1 g, .2093 mmol, 1.3 equivalents) was added. This stirred overnight at room temperature. The reaction was quenched with deionized water and extracted with diethyl ether. Residual solvent was removed under reduced pressure and the crude oil was purified by flash chromatography (6 mg, 8% yield).[21] **1H NMR** (DMSO-*d*6, 25 °C, 800 MHz): δ 8.09 (s, 1H); 7.64 (t, *J*= 10.9 Hz, 1H); 7.48 (dd, *J*= 10.6, 1.3 Hz, 1H); 7.25 (d, *J*= 4.2 Hz, 1H); 7.11 (d, *J*= 11.2 Hz, 1H); 7.04 (t, *J*= 10.5 Hz, 1H); 6.52 (d, *J*= 7.7 Hz, 2H); 6.39 (dd, *J*= 4.2, 1.0 Hz, 1H); 4.40 (t, *J*= 5.8 Hz, 2H); 4.27

(t, *J*= 5.9 Hz, 2H); 4.17 (t, *J*= 9.1 Hz, 2H); 3.27 (t, *J*= 9.2 Hz, 2H); 2.29 (s, 9H) ppm; 1.18

m, *J*= 9.6 Hz, 2H); 1.70 (m, *J*= 10.2 Hz, 2H); 1.43 (s, 9H) ppm.

References

1. Sekar, R.B., Periasamy, S. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *Journal of Cell Biology* **2003**, 160 (5), 629-633.
2. Clegg, R.M. Fluorescence resonance energy transfer. In *Fluorescence Imaging Spectroscopy and Microscopy*. Wang, X.F., Hermna, B., Eds. John Wiley and Sons Inc. New York, New York, **1996**, pp 179-251.
3. Förster, T. Delocalized excitation and excitation transfer. In *Modern Quantum Chemistry*. Sinanoglu, O., Eds. Academic Press, New York, New York, **1965**, pp 93-137.
4. Pollok, B.A., Heim, R. Using GFP in FRET-based applications. *Trends in Cellular Biology* **1999**, 9 (1), 57-60.
5. Sattler, M., Fesik, S.W. Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure* **1996**, 4 (11), 1245-1249.
6. Marley, J., Liu, M., Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR* **2001**, 20 (1), 71-75.
7. Keeler, J. The Vector Model. In *Understanding NMR spectroscopy*. John Wiley and Sons Inc. New York, New York, **2011**, pp 25-46.
8. Keeler, J. The Fourier Transformation. In *Understanding NMR spectroscopy*. John Wiley and Sons Inc. New York, New York, **2011**, pp 47-64.
9. Mandal, P.K., Majumdar, A. A Comprehensive Discussion of HSQC and HMQC Pulse Sequences. *Concepts in Magnetic Resonance* **2004**, 20A (1), 1-23.
10. Yuan, X., Davydova, N., Conte, M.R., Curry, S., Matthews, S. Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein. *Nucleic Acids Research* **2002**, 30 (2), 456-462.
11. Grzesiek, S., Bax, A. An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins. *Journal of Magnetic Resonance* **1992**, 99 (1), 201-207.
12. Grzesiek, A., Bax, A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* **1992**, 114 (16), 6291-9293.
13. Copeland, R.A., Pompliano, D.L., and Meek, T.D. Drug-target residence time and its implications for lead optimization. *Nature Reviews Drug Discovery* **2006** 5 (1), 730–739.
14. Potashman. M.H. and Duggan. M.E. Covalent Modifiers: An Orthogonal Approach to Drug Design. *Journal of Medicinal Chemistry* **2009**, 52 (1), 1231–1246.
15. Smith, A.J., Zhang, X.Y., Leach, A.G., and Houk, K.N. Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins. *Journal of Medicinal Chemistry* **2009**, 52 (1), 225–233.
16. Singh, J., Petter, R. C., Baillie, T. A, and Whitty, A. The resurgence of covalent drugs. *Nature Reviews Drug Discovery* **2011**, 10 (1), 307–317.
17. Durham, T. B. and Blanco, M. Target Engagement in Lead Generation. Bioorganic *Medicinal Chemistry Letters* **2015**, 25 (1), 998–1008.
18. Ristevski, S., O'Leary, D. A., Thornell, A. P., Owen, M. J., Kola, I., and Hertzog, P. J. The ETS transcription factor GABPalpha is essential for early embryogenesis. *Molecular Cellular Biology* **2004**, 24 (1), 5844-5849.
19. Xue, H. H., Bollenbacher, J., Rovella, V., Tripuraneni, R., Du, Y.B., Liu, C.Y., Williams, A., McCoy, J. P., and Leonard, W. J. GA binding protein regulates interleukin 7 receptor alpha chain gene expression in T cells. *Nature Immunology* **2004**, 5 (1), 1036-1044.

20. Gouliaev, A.H., Larsen, M., Varming, T., Mathiesen, c., Johansen, T. H., Scheel-Kruger, J., Olsen, G.M., Nielsen, E.O. Preparation of benzothiadiazines, quinazolines, and other aryl-fused heterocycles as positive AMPA-receptor modulators for treatment of memory and learning disorders. *Patent Cooperation Treaty* **1999**, WO994245A2.

21. Doherty, J.B., Shu, M., Shen, D.M., Zhang, F. Preparation of 1,2-dihydroquinolin-2-one, 1,2-dihydroquinoxalin-2-one, and 1,2-dihydronapthyridin-2-one derivatives for treating ocular hypertension. *Patent Cooperation Treaty* **2007**, WO2007108968A2.

22. Li, Y., Wu, R., Zhang, Q., Ma, X., Shi, Z., Mao, G., Yang, W., Zhao, G. Bifunctional hydroxy ketone benzoyl formate compounds for photo initiators. *Faming Zhuanli Shenqing* **2014**, CN103709036A.

23. Li, W., Feng, Z., Zhu, H., Yu, R. Berberrubine and Honokiol synergistic compound, preparation method and application. *Faming Zhuanli Shenqing* **2018**, CN108395429A.

Chapter 3: Results

3.1 Fragments that Disrupt the Binding of GABPα to GABPβ

The analysis of FRET data obtained from this study shows that several structures bind to the GABP complex, which inhibits the protein from binding. Initially, two chemical libraries were screened, and based upon those results, compounds with similar structures to those that had potent activity were synthesized or purchased and analyzed for activity with the FRET assay.

Before this assay was used for fragment-based screening, it was used to determine the K_D. The K_D is a dissociation constant that measures the binding capability between two targets. This K_D measurement is essential because it validates that the proteins have been properly purified and are folding correctly to be associated with each other. The ratio of the emission wavelength of Cerulean and the emission wavelength of Venus were plotted against the log (concentrations of the proteins) and fitted with a sigmoidal curve. This gave a K_D measurement for the binding of the two to one another of 3.0 nM ± 0.9, which agrees well with the published K_D determined using isothermal titration calorimetry (ITC) (K_D = 1.7 nM).[1] Since this corresponds well to the previously published K_D of GABPα and GABPβ, it can be assumed that the use of the GFP derivatives Cerulean and Venus does not have an effect on binding of GABPα to GABPβ. The K_D determination was also essential because it determined the proper concentration for the two proteins for subsequent fragment screening. For this, the concentration of the proteins would be held at three times the K_D to ensure maximum dynamic range, but without having the protein concentration

so high that it might inhibit the effects of the compound fragments.

In addition to comparing with published literature, competitive binding experiments were conducted to validate the assay. While following the same assay protocol, fluorescently-labeled GABPα and GABPβ were held at three times the K_D, while a series of concentrations of pHis-GABPβ were added. The unlabeled GABPβ acts as an inhibitor to the fluorescently-labeled GABPβ, out-competing it to bind to fluorescently-labeled GABPα. When this happens, there is no energy transfer, and thus the emission of Venus decreases. By plotting the FRET ratio of Cerulean to Venus emission to the concentration of added pHis-GABPβ, an IC_{50} (concentration of the inhibitor where there is a 50% inhibition) value is determined. With this data, the inhibition constant, K_i, was determined using the Cheng-Prusoff equation.[2] This inhibition constant represents the affinity of pHis-GABPβ and Cerulean-GABPα. Since the dissociation constant of Cerulean-GABPα and Venus-GABPβ (K_D) matches the inhibition constant (K_i) of Cerulean-GABPα and pHis-GABPβ, there is further proof that there is no effect on the binding of GABPα to GABPβ from the two fluorophores.

Fragment screening is an alternative approach that has recently gained popularity in the pharmaceutical industry. Contrary to high-throughput screening, this method employs relatively small molecules. Although decreased in size, these small molecules are highly drug-like, making them good candidates for further development. One of the advantages to this approach is that it offers a wide degree of chemical diversity, but with a more limited number of compounds. Because of this, it increases the likelihood of finding a molecule that can bind to a specific pocket on the protein. Unfortunately, because

these fragments are smaller, this methodology suffers from the obstacle that the hits obtained are typically of weak potency (100s of µM to mM IC_{50} values). Because of this, multiple fragments can be paired and linked together, possibly with cysteine-reactive compounds to decrease the IC_{50} values to levels that would have a potent effect in cells. This method relies upon the principle that multiple small molecule fragments have been shown to bind to the GABP complex in different areas, and thus by linking them together, a larger molecule with a higher potency is produced.

The previously described GABPα-GABPβ FRET assay was used to single-point screen two different fragment libraries of 1,500 (Maybridge library) and 12,000 (ChemDiv Library) compounds for inhibition of GABPα-GABPβ binding (Figure 19). This preliminary single-point screening was completed by Adam Boulton and Sonja Thölmann. For these single-point screens, an average FRET ratio and standard deviation was determined. For the Maybridge library, compounds with FRET ratios above 2.5 standard deviations were flagged. For the ChemDiv library, FRET ratios above three standard deviations were flagged. This identified 23 compounds from the Maybridge library and 93 fragments from the ChemDiv library. Interestingly, most of the hits in the ChemDiv library shared a similar aromatic structure. To validate these hits, multiple-point screens were carried out on these fragments. For the Maybridge library, 12-point screens were carried out, while for the ChemDiv library, 8-point screens were performed. In this secondary screening, three compounds were identified to have IC_{50} values under 1 mM. 38F3 had an IC_{50} of 500 µM, 26C6 had an IC_{50} of 900 µM, and 37A6 had an IC_{50} of 590 µM. All ligand and compound structures, names, and reactivity for the duration of this study can be viewed

in Table 2. Moreover, none of these three compounds were hits in screens carried out with two unrelated targets.

As there are cysteine residues located at the interface for binding of GABPα to GABPβ, a 1,000-compound library of cysteine reactive molecules (Enamine) was screened for inhibition of GABPα-GABPβ binding. This screen identified two compounds that showed significant inhibition (IC_{50} = 120 µM and 140 µM) using the FRET assay (Figure 20). Other targets with this library were screened and these two compounds were not active in the other screens, suggesting they will have selectivity for GABP.
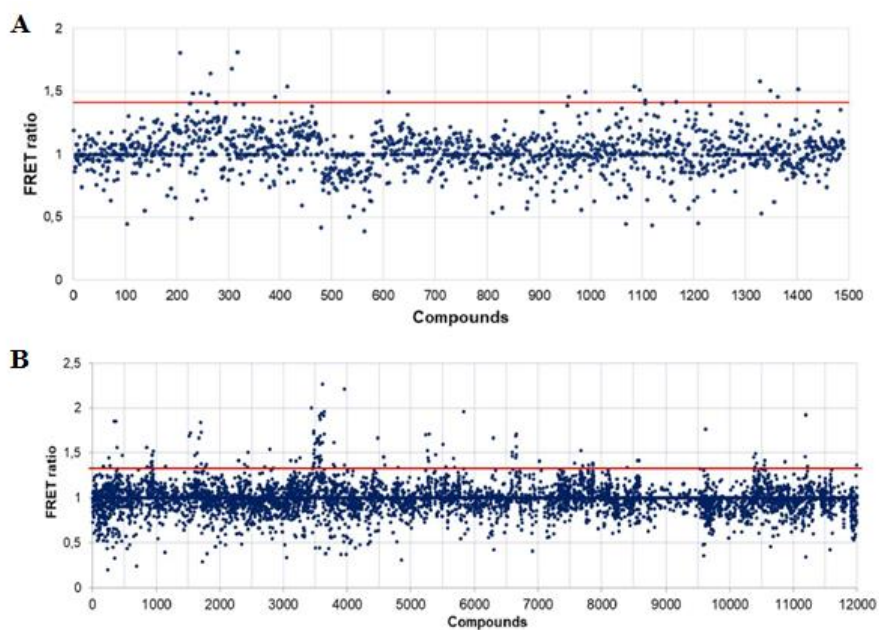
Figure 19: Visual results of the FRET-based single-point screens of the Maybridge (A) and ChemDiv (B) libraries. Compounds on or above the red line were flagged for further screening.
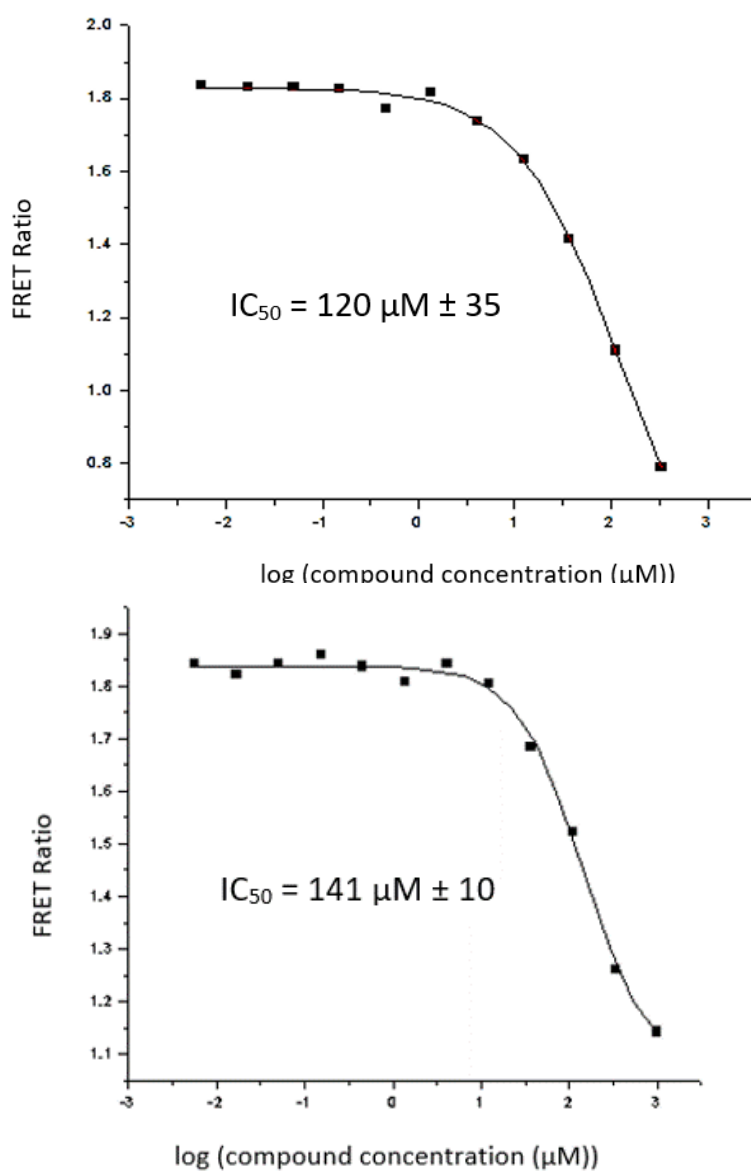
Figure 20: IC$_{50}$ determination of the most active cysteine-reactive compounds: Z1455664051 (top), and AM-2-1 (bottom), identified from the Enamine screen using the FRET assay.

Great effort was spent altering the parent structures of 37A6, 38F3, and 26C6 to see if the potency of the compound would increase. For this part of the study, the compounds that were focused on were those that either had a facile synthesis, or were relatively inexpensive to purchase. Moreover, there was a push to synthesize and purchase compounds that contained sulfonamides. Sulfonamides offer a facile linking point to link together multiple compounds to increase the potency of a drug. In addition, in recent research, sulfonamide-containing compounds have proven to be potent inhibitors of various protein systems. The inhibition of the human carbonic anhydrase isozyme XII (an isozyme associated with human renal cell carcinoma) was investigated with a series of relatively simple sulfonamides. Activity of these compounds went as low as the 11-12 nM range.[3-6]

For 37A6, any alteration to the primary amine significantly decreased the potency (10-fold decrease). Moreover, altering the methyl substituent for an isopropyl group also significantly decreased the potency. However, upon changing this methyl group to an ethyl group, the potency increased from 590 µM (37A6) to 385 µM (AM-3-5). For 38F3, compound Z3487916510 which added a sulfonamide to the parent structure, severely decreased the potency. Moreover, 2C57582 which altered the carboxyl group to an amine saw the potency decrease slightly. Compound 37A6 was the structure that was varied the most due to the combination of its FRET assay potency, as well as its performance in NMR studies (See section 3.3). At first, multiple similar structures of 26C6 (Z1849888669 and QC-3565) were purchased and had their activities tested. While both of these compounds had increased potency when compared with 26C6, the sulfonamide showed better activity.

So, it was then decided to synthesize the sulfonamide version of 26C6 (AM-2-37) which altered the carboxyl group to a sulfonamide group with a primary amine. This severely decreased the potency by a factor of 10. However, upon alkylating this sulfonamide with an ethyl group (AM-1-143), the potency increased greatly (291 µM), even surpassing the activity of 26C6 (900 µM). The study then focused on altering the substituents that directly come off of both amines in the structure AM-2-37. Varying these groups proved to significantly decrease the original activity of the parent structure. However, it was seen that when an isopropyl group is added to the acetyl group, and the sulfonamide amine is left as a primary amine, then the activity increases to 350 µM, the most potent for this type of structure.

In this study, there was also a significant effort to explore the activity of indole-containing compounds on the GABP complex. The indole's structure is found all throughout nature, such as many alkaloids, the amino acid tryptophan, and plant hormones. Because this is such a vastly used structure by nature, there has been a real push to explore the limitations (if any) of indoles. Due to this exploration, it was discovered that indoles are capable of being therapeutic targets such as anti-inflammatories, phosphodiesterase inhibitors, HMG-CoA reductase inhibitors, etc. Many of these target receptors belong to integral membrane G-protein coupled receptors, and thus all possess a conserved binding pocket that ideally binds to indoles.[7] In other recent studies, it was shown that indole-containing drugs can interfere with tubulin function, which offers a broad anti-tumor method for drug discovery.[8] Microtubules are the basic components of maintaining cell structure. Drugs that can inhibit microtubule formation or microtubule polymerization are

of great value to cancer research.[9] Moreover, aroylindoles, arylthioindoles, diarylindoles, and indolylglyoxyamides have shown to inhibit tubulin polymerization.[8] Just over the past decade, countless indole-containing compounds have been reported to have potency inhibiting microtubule activity. Indole-containing drugs have even been previously synthesized as an inhibitor of HIV-1 reverse transcriptase.[10] Cleary, indoles have proven to be a powerful structure, which is the reason there was a drive to test different indole moieties against the GABP complex. Moreover, a mutagenesis study carried out on the interface of GABPβ showed a 1,000-fold reduction in binding to GABPα when there was a Phe136Ala mutation.[1] Clearly, aromaticity plays a crucial role at the binding interface, which is reflective in the cysteine-reactive compound library screen, where one compound (AM-2-1) was identified with high activity. This compound contains an indole, and thus there was a push to screen for more indole-containing compounds.

AM-3-3, AM-3-4, AM-3-6, and AM-3-7 were all tested for activity against the GABP complex. These structures varied in where the methoxy group on the indole was placed. While all of these activities were relatively weaker (1-2 mM range) the indole with the methoxy group in the 5 position (if the amine is assigned position 1) gave the best activity with 1,400 µM (AM-3-3). There was then an effort to see how potent the activity would be if two of these compounds were linked together.
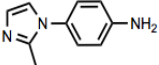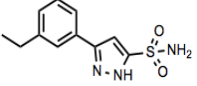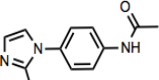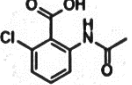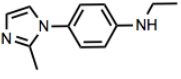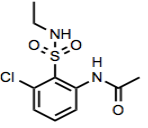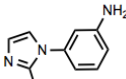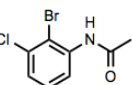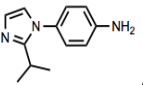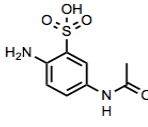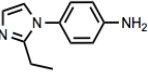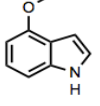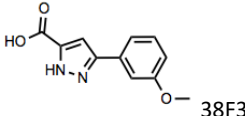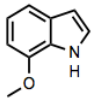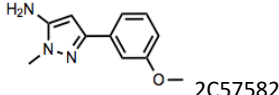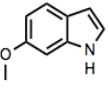
After seeing an increase in potency when the primary amine in the sulfonamide (AM-2-37) is converted to a secondary amine (AM-1-143), it was hypothesized that using this amine as a linking anchor for another potent fragment might afford an even higher potency. Indole-containing compound AM-3-3 was selected as another fragment to link to
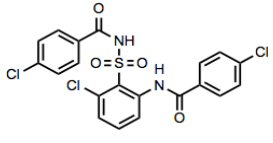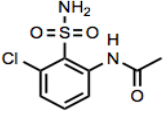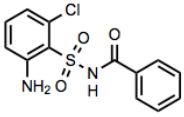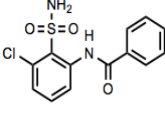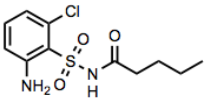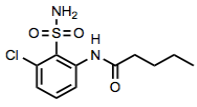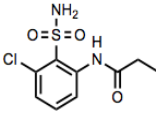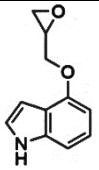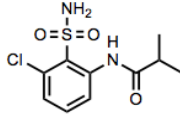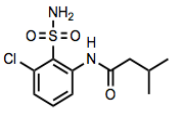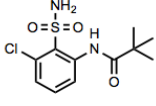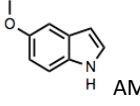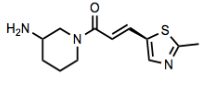
AM-2-37 to increase potency. Since this indole-containing compound is structurally quite different from the 26C6 parent structure, it was hypothesized that these two compounds are binding to different places on the GABPα-GABPβ complex. Therefore, linking these two fragments together should increase the overall potency of the compound.

A proposed molecule that links Z3487916514 and AM-3-3 with an ethylene glycol-like chain was hypothesized to have high activity. AM-2-65 (Figure 21) was synthesized in four steps. Unfortunately, it proved difficult to remove the boc group, that was originally put on to protect the amine group on the indole. New ways to protect this amine group are currently being studied instead of using a boc group. Protecting groups such as benzyl, and benzyl carbamate groups are being explored.
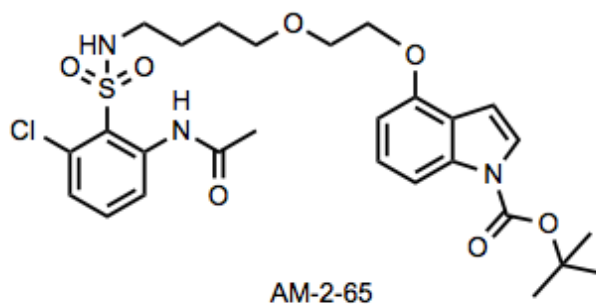
AM-2-65

Figure 21: Structure of AM-2-65. This molecule links AM-2-37 together with an indole containing compound with an ethylene glycol-like linker. It was theorized that this would increase the potency in GABP inhibition.

Name/Structure	IC$_{50}$ (µM)	Name/Structure	IC$_{50}$ (µM)
37A6	590 µM ± 69	Z3487916510	8200 µM ± 130
AM-1-99	2,200 µM ± 350	26C6	900 µM ± 300
AM-1-101	2,000 µM ± 240	AM-1-143	291 µM ± 40
AM-3-1	2,500 µM ± 290	Z1849888669	340 µM ± 53
AM-3-2	3,000 µM ± 1,000	QC-3565	792 µM ± 130
AM-3-5	385 µM ± 23	AM-3-3	1,400 µM ± 80
38F3	500 µM ± 72	AM-3-4	2,000 µM ± 350
2C57582	529 µM ± 45	AM-3-6	1,800 µM ± 500

Structure	IC50	Structure	IC50
HPK-P-14	8200 µM ± 130	AM-2-37	2,700 µM ± 1,000
HPK-NP-2	3,200 µM ± 1,000 Crashed out in aqueous	HPK-P-2	1,800 µM ± 1,000
HPK-NP-3	1,000 µM ± 200 Crashed out in aqueous	HPK-P-3	1,300 µM ± 350 Crashed out in aqueous
HPK-6	1,200 µM ± 350	AM-2-1	140 µM ± 10
HPK-8	1,800 µM ± 700	HPK-P-16	350 µM ± 50
HPK-P-7	450 µM ± 40 Crashed out in aqueous	AM-3-7	1,750 µM ± 130
Z1455664051	120 µM ± 35		

Table 2: Names and structures of ligands that bind to the GABP complex and their associated IC$_{50}$ values. All HPK compounds were synthesized by Hanuman Kalmode.

3.2 Cysteine-Reactive Ligands do not Bind to GABPβ Cysteine

It was assumed that if the cysteine-reactive compounds (Z1455664051 and AM-2-1) were binding to the one cysteine on the GABPβ, then the potency of the drug would drop substantially upon removal of the cysteine. After a GABPβ mutant (C139A) was expressed and purified with a Venus-GFP derivative attached, this protein was used with wild-type GABPα with the Cerulean-GFP derivative attached to test the cysteine-reactive compounds in the FRET assay. After carrying out the same assay with the mutant, the IC_{50} values actually decreased slightly, indicating that the cysteine-reactive compounds do not bind to the cysteine on GABPβ (Figures 22 and 23). It is possible that these compounds could be binding to one of the nine cysteines on alpha or (less likely) might bind somewhere else on either protein. Since previous studies showed that the inhibition of GABPα leads to early embryonic lethality[11,12] there was still more of a focus to find a target that binds to GABPβ.
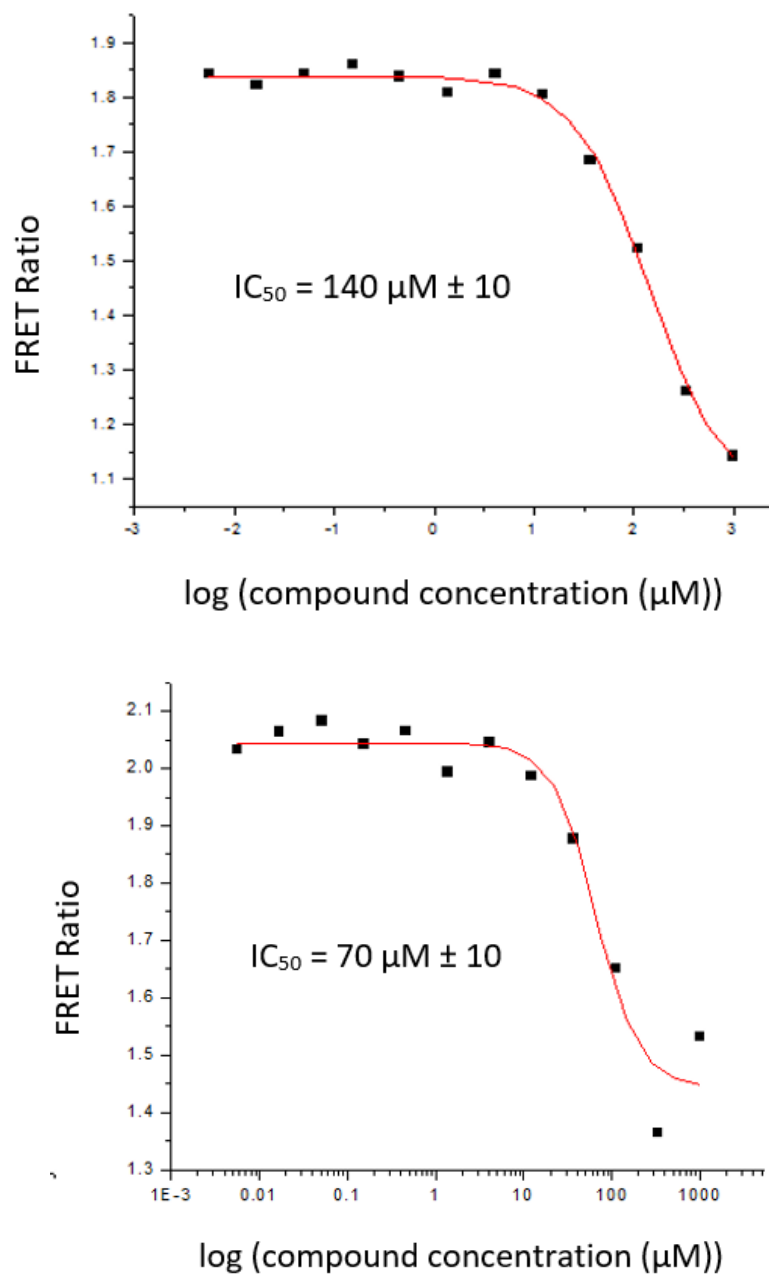
**Figure 22:** IC_{50} determination of wild-type GABPβ (top) and C139A mutant GABPβ (bottom) with AM-2-1.
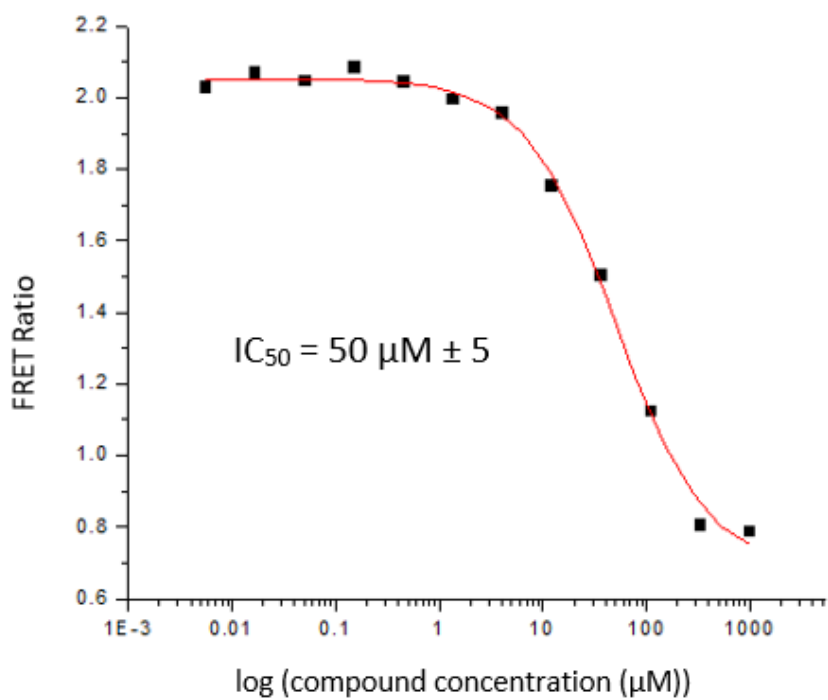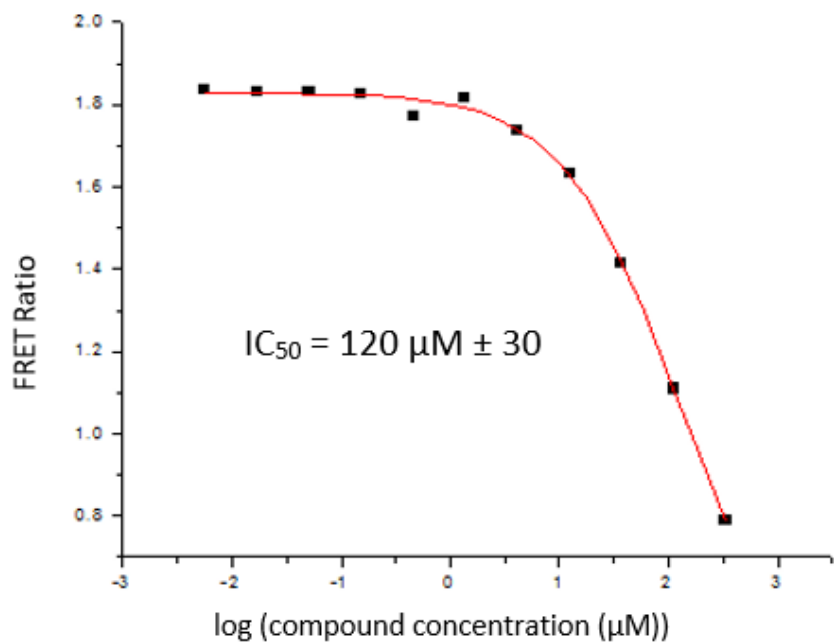
**Figure 23:** IC_{50} determination of wild-type GABPβ (top) and C139A mutant GABPβ (bottom) with Z1455664051.

Since the mutagenesis study gave unexpected results, we attempted to verify that the cysteine-reactive compounds were indeed binding to GABPα through mass spectrometry analysis. After allowing each compound to come to equilibrium separately with GABPα and GABPβ, each sample was analyzed by mass spectrometry (Figures 24-26). In addition, to elucidate which protein the compounds, 26C6, 38F3, and 37A6, were binding to, these were also allowed to come to equilibrium with GABPβ and then tested for mass spec (Figure 27). In all of these mass spectroscopy plots, the signal-to-noise ratio is too large to determine which protein these compounds are binding to. Therefore, it is necessary to confirm the binding through NMR studies.
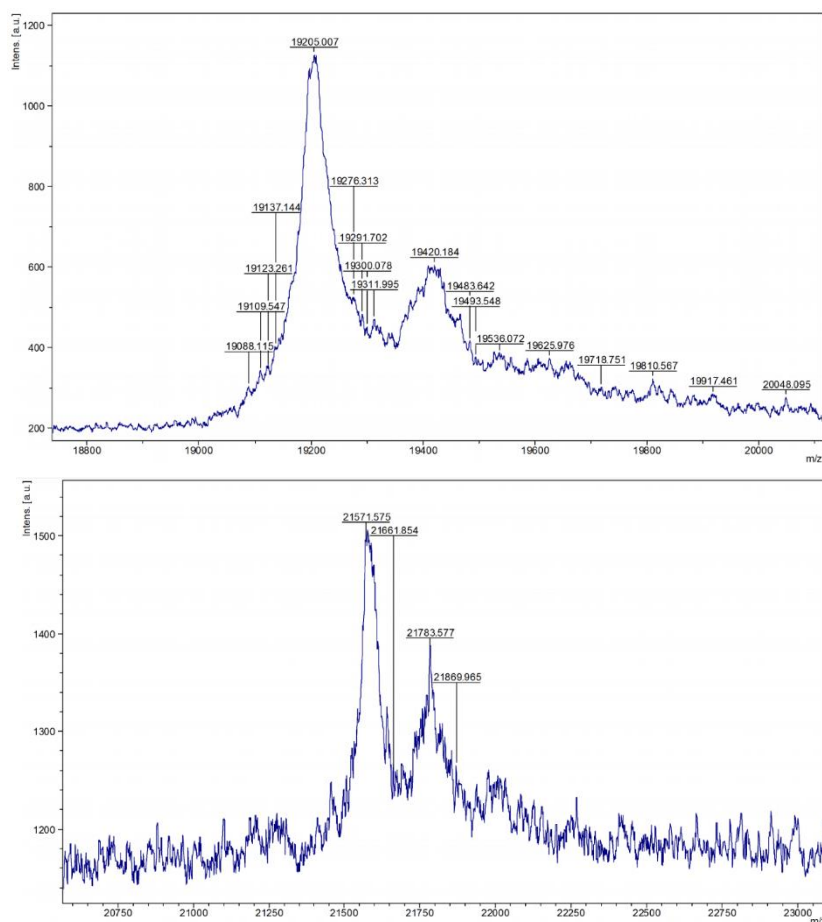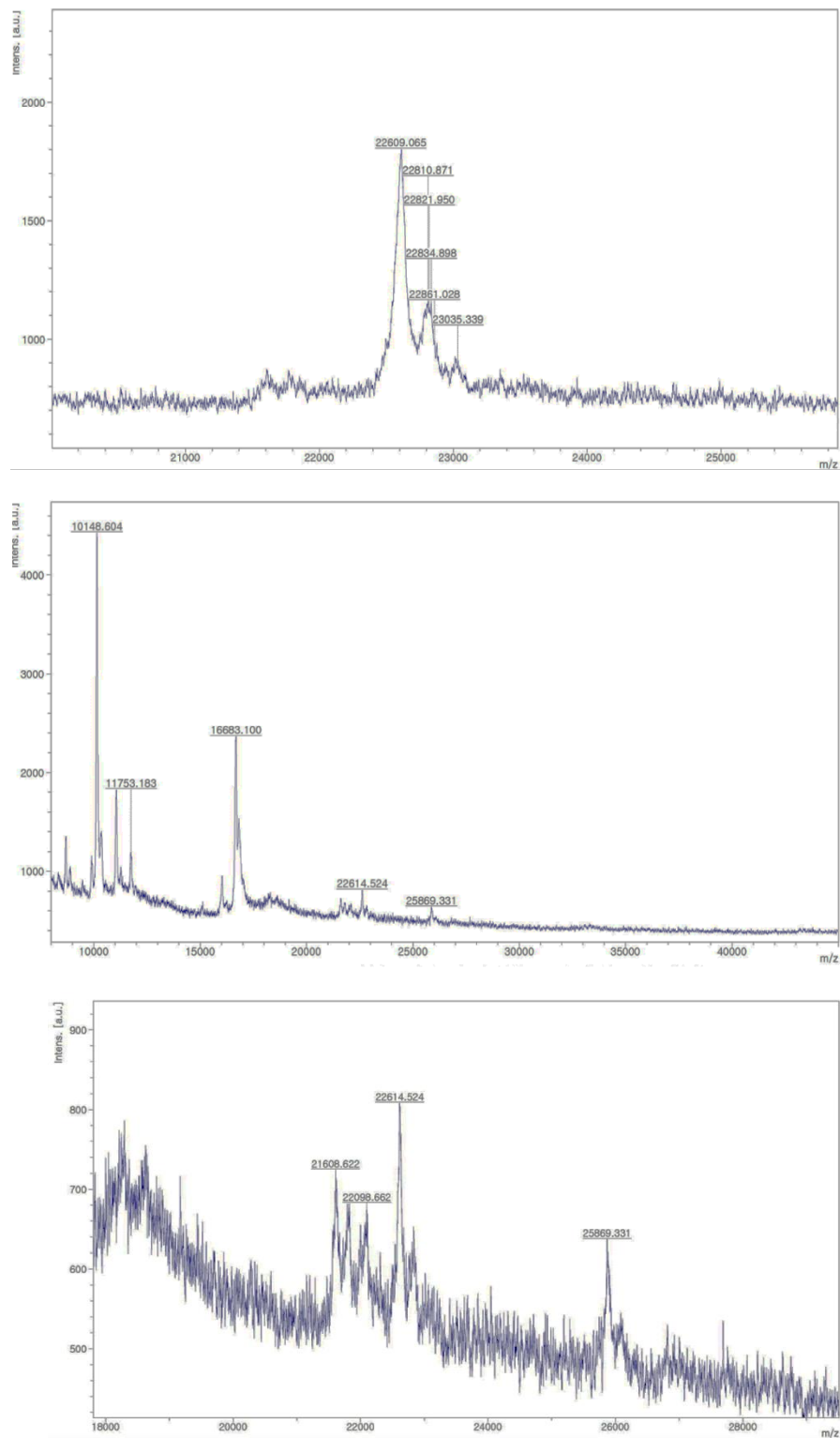

Figure 24: Mass spec results of GABPα (top) and GABPβ (bottom) when reacted with DMSO.
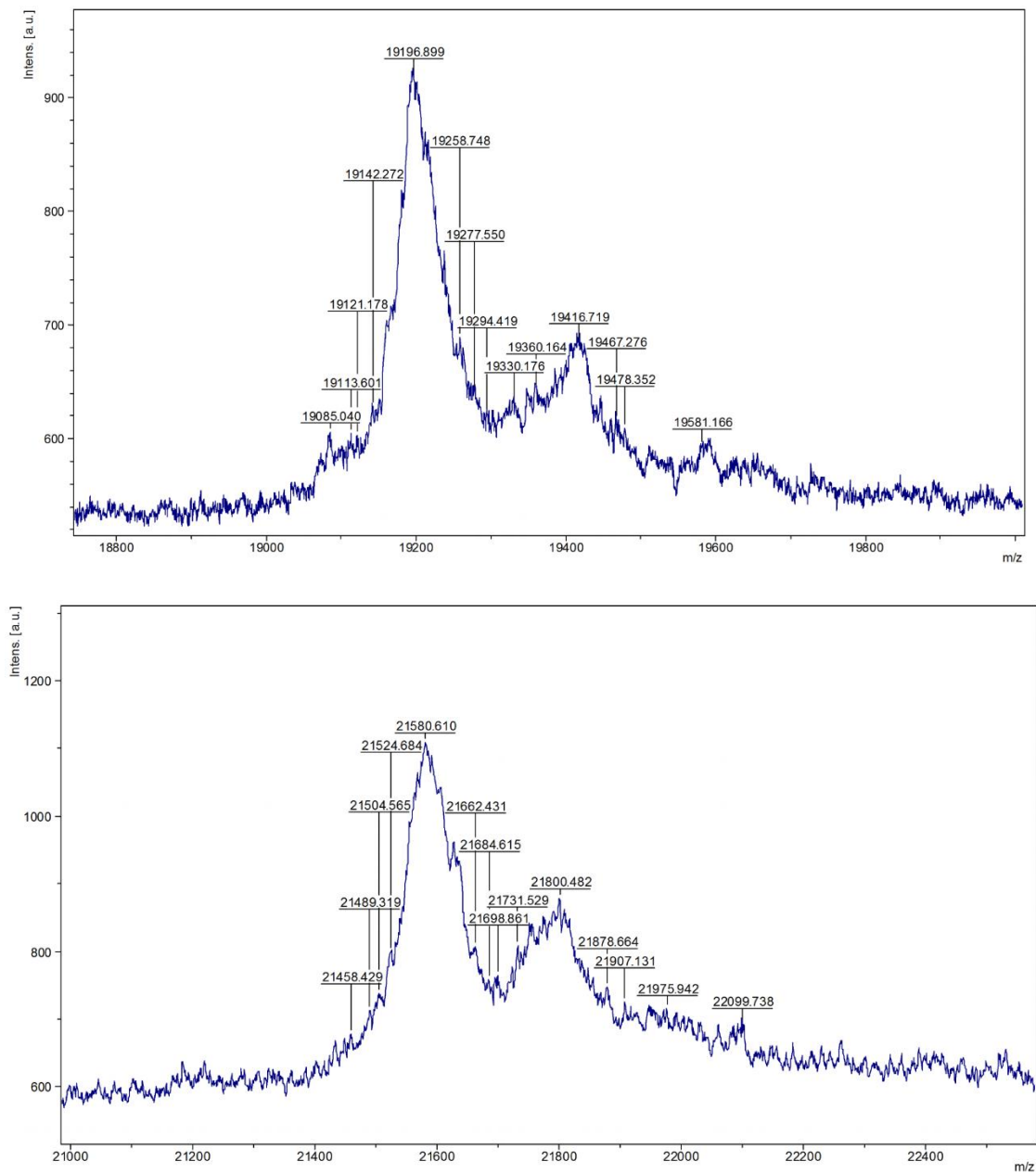
Figure 25: Mass spec results of GABPα (top) and GABPβ (bottom) when reacted with AM-2-1.
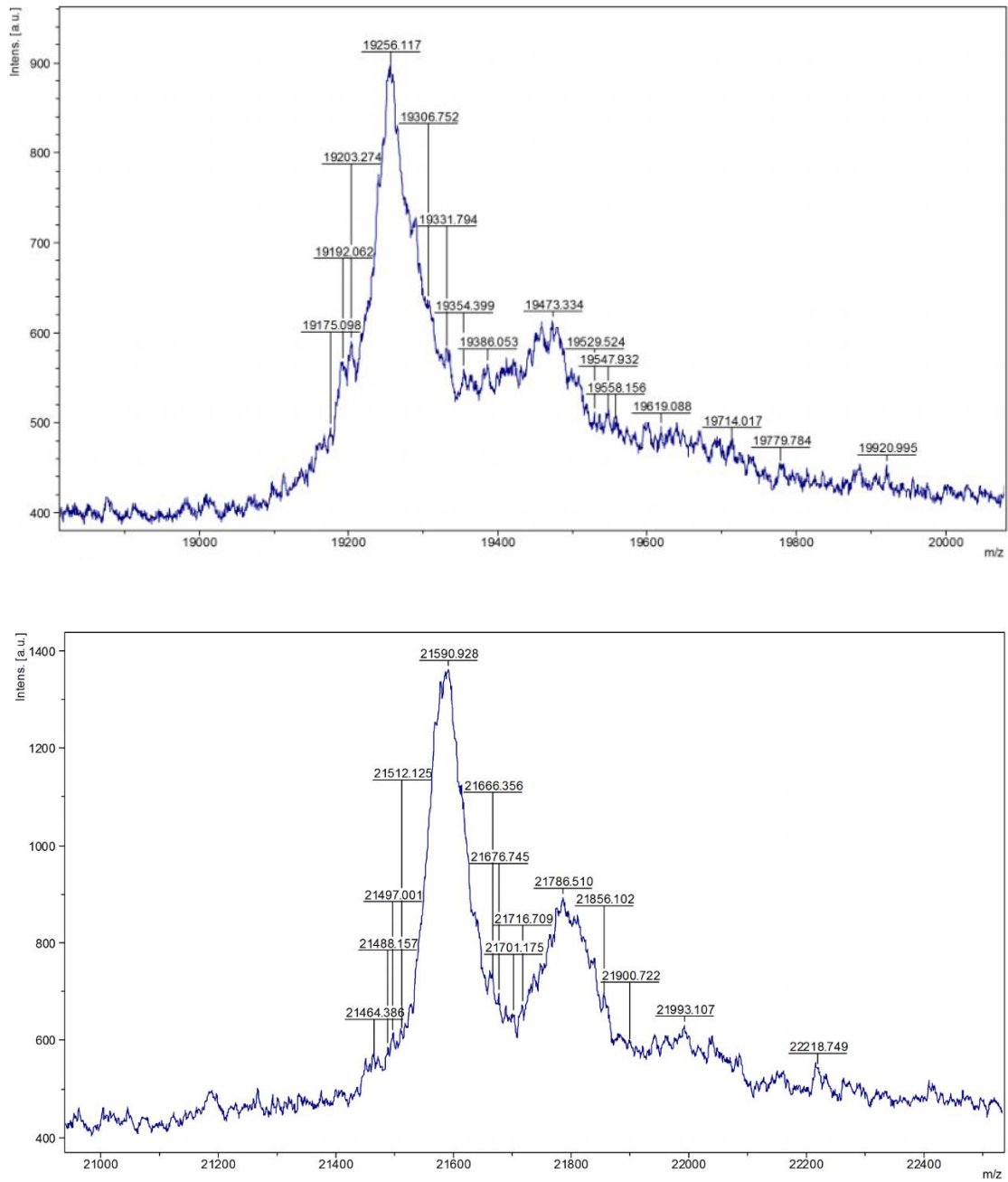
Figure 26: Mass spec results of GABPα (top) and GABPβ (bottom) when reacted with Z1455664051.

Figure 27: Mass spec results of GABPβ when reacted with 26C6 (top), 38F3 (middle), and 37A6 (bottom).

3.3 Stability of GABPβ Increases when in a Buffer with Free-Amino Acids and Upon the Addition of Ligands

For NMR measurements, there was, again, a focus to study GABPβ due to the lethality associated with inhibiting GABPα. pHis- GABPβ was used instead of the Venus labeled protein because this GFP-label adds a significant amount of molecular weight, which can seriously degrade the spectral data. Larger proteins detrimentally affect the NMR spectrum due to line broadening.[13] Moreover, as shown in the FRET assay, the addition of the GFP-label has no effect on the binding of compounds, and therefore should have no effect on NMR spectra upon removal.

Since GABPβ is ill-behaved in solution on its own, there was substantial effort put into finding a suitable buffer that would give a spectrum with usable quality. Unfortunately, high salt content is needed in the solution buffer because without it GABPβ tends to aggregate. However, a higher salt content typically reduces the sensitivity of NMR experiments. Initial NMR spectra of GABPβ was recorded in a buffer that contained 500 mM KCl. It is clear to see in Figure 28 that the high-salt content has decreased the sensitivity so much that the signal-to-noise ratio is too low to distinguish what is background noise and what is protein.

In general, the sensitivity of NMR, particularly with cryogenic probes, is reduced by conductive samples. Although biological samples are not typically conductive, the buffers they are dissolved in usually are. Biological macromolecules are typically more soluble and stable in samples that contain a higher salt content, which decreases NMR sensitivity. It is not particularly the salt concentration that affects the sensitivity, but rather

the conductivity which is a function of the concentration and the mobility of the ions in solution.[14] So, in theory, choosing a salt system with low ionic mobility, will afford an NMR spectrum with higher sensitivity.

Moreover, recent literature has shown that when you use arginine and glutamic acid in lieu of more traditional salts, such as high amounts of sodium chloride, the sensitivity of NMR experiments is increased.[15] This replacement also simultaneously reduces the sample conductivity and improves protein stability. In this study, it was shown that by switching to a salt system of arginine and glutamic acid, the sensitivity of the NMR experiments increased 6-fold. Because these amino acids are charged oppositely, they come together to form a highly-soluble salt in aqueous solution. Arginine and glutamic acid have been shown previously to prevent aggregation in proteins, therefore it is fitting that this would be an ideal salt system for GABPβ.[16-19] Taking this into account it was found that a 200 mM arginine and 200 mM glutamic acid salt solution afforded the best quality NMR spectrum (Figure 29). In addition, more resolved and well dispersed spectra were found to coincide with NMR data recorded at 37°C. Moreover, previous research has noted that when analyzing a protein in a high-salt concentration buffer, it is advantageous to use an NMR tube that is 2 or 3 mm in diameter. This has an intense effect on the sample resistance, resulting in improvements to the signal-to-noise ratio, as well as shorter 90º pulses.[20] Because of this, all NMR spectra were obtained in a Shaped Shigemi Tube (Hampton-Township, PA, US). This NMR tube has an internal sample cavity of 2 mm wide by 20 mm high (rectangle sample cavity).
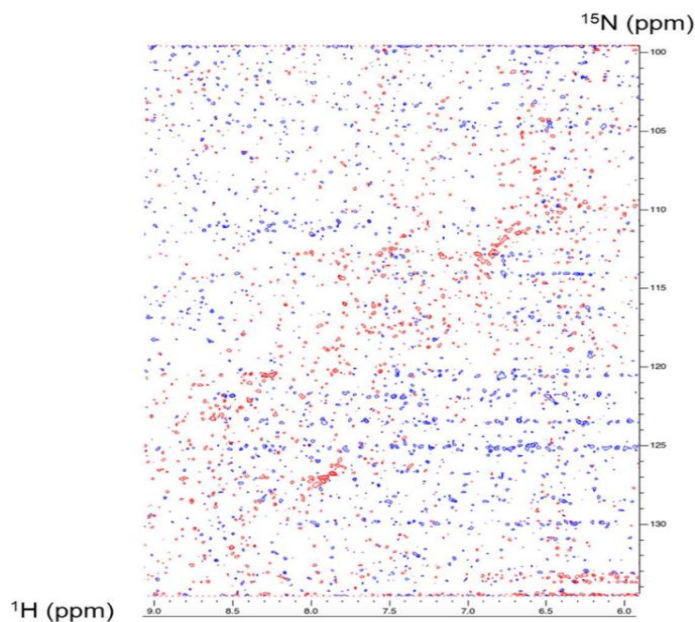
Figure 28: HSQC of GABPβ with DMSO. Buffer used was 500 mM KCl, 50 mM tris, 5 mM DTT, pH: 7.5. The Y-axis represents the ^{15}N chemical shift and the X-axis represents the ^{1}H chemical shift. In a typical HSQC spectrum of a protein, each peak (spot) represents one amino acid. The signal-to-noise ratio is so low on this spectrum that no peaks of amino acids are seen
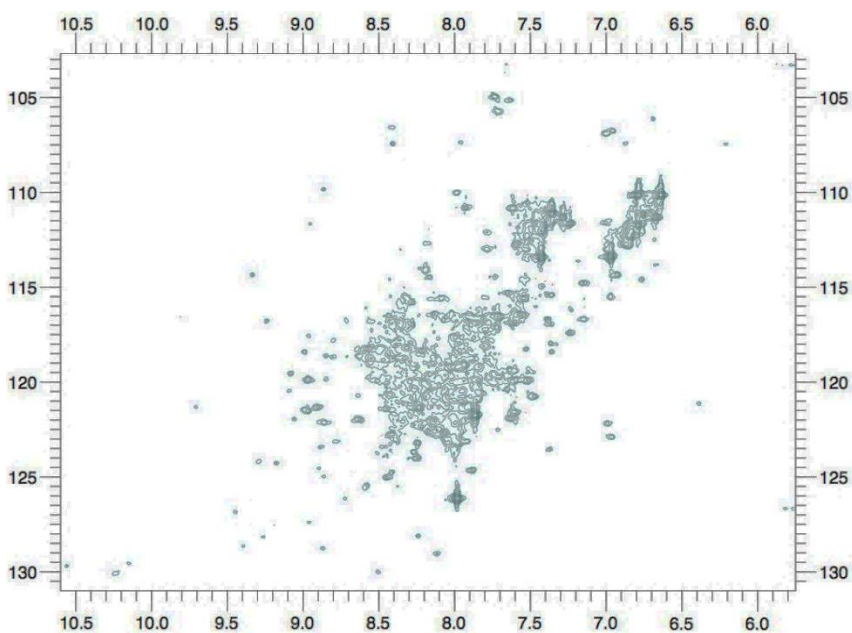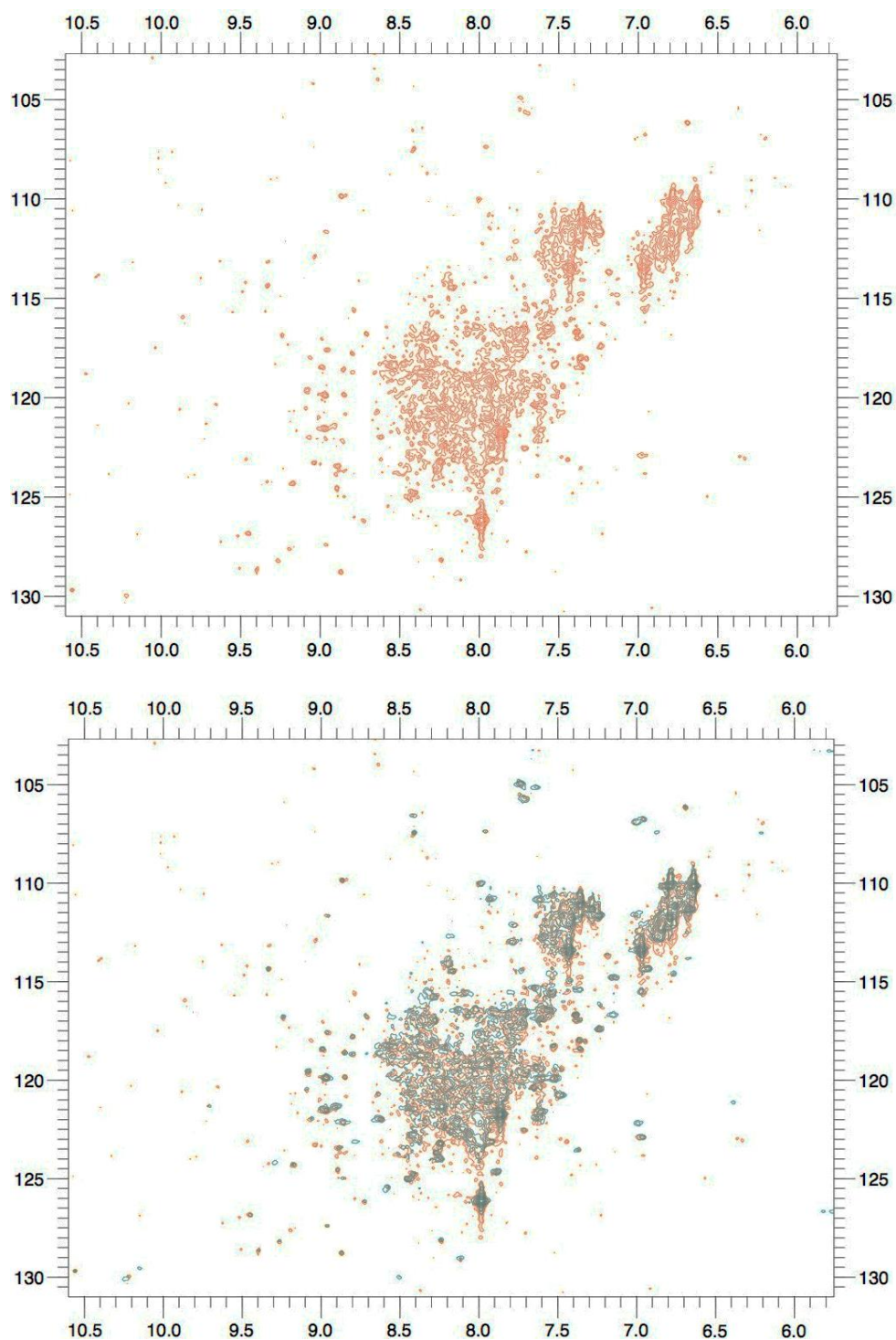

Figure 29: HSQC of GABPβ with DMSO. Buffer used was 200 mM arginine, 200 mM glutamic acid, 5 mM DTT, pH: 7.5. The Y-axis represents the ^{15}N chemical shift and the X-axis represents the ^{1}H chemical shift.

Keeping the experimental conditions mentioned previously the same, NMR spectra have been produced for free GABPβ (Figure 29) and upon the addition of 26C6 (Figure 30), 37A6 (Figure 31), and 38F3 (38F3). Interestingly, upon addition of 26C6 and 38F3, the spectra become much more resolved with peaks that are more dispersed (Figures 30 and Figure 33). This insinuates that these compounds are binding to GABPβ and stabilizing it to get spectra with better resolution. It may be possible then, that GABPβ undergoes a conformational change when it binds to these compounds, and possibly when it binds to GABPα.

Unfortunately, triple resonance NMR experiments were attempted, but the signal-to-noise ratio was too low to be able to assign the backbone on an HSQC spectrum. This is due to the instability of GABPβ in solution at 37°C. Even with the increased stability from the addition of these compounds, GABPβ only lasts in solution for approximately two hours before precipitating out into a white powder in solution. Because of this, triple-resonance NMR studies were unable to be successfully completed during this study. Future research will need to include developing a solution for GABPβ that allows it to remain in solution longer for more extensive NMR experiments, which will allow for the assignment of the backbone. If this is possible, then it will be easy to pinpoint which amino acids the compounds are binding to, due to the chemical shift perturbations.
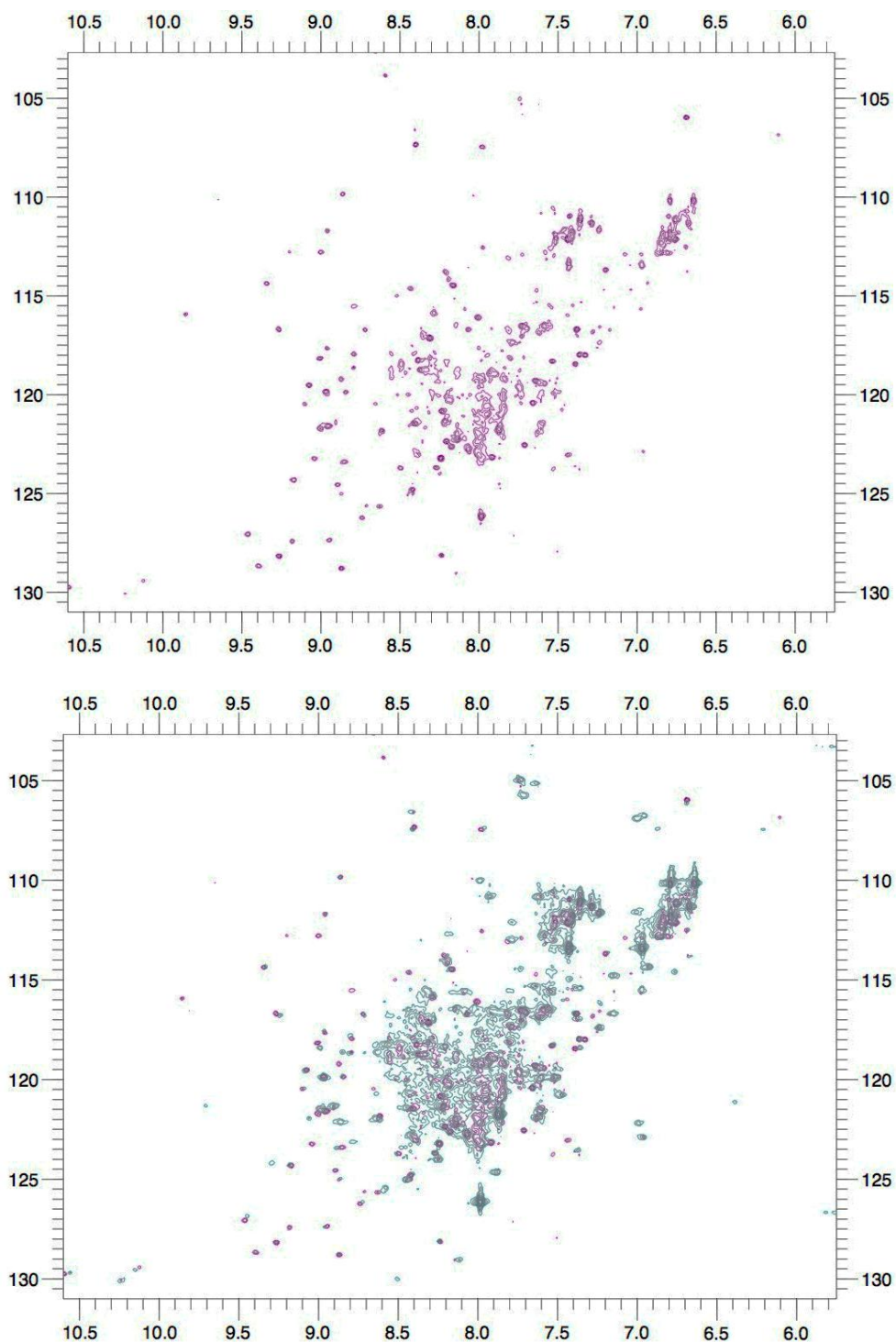
Figure 29: HSQC of GABPβ with 26C6. Buffer used was 200 mM arginine, 200 mM glutamic acid, 5 mM DTT, pH: 7.5. The Y-axis represents the ^{15}N chemical shift and the X-axis represents the ^{1}H chemical shift. The top spectrum is just GABP and 26C6. The bottom spectrum is GABP and 26C6 overlaid on top of the GABP and DMSO spectrum.

Figure 30: HSQC of GABPβ with 37A6. Buffer used was 200 mM arginine, 200 mM glutamic acid, 5 mM DTT, pH: 7.5. The Y-axis represents the ^{15}N chemical shift and the X-axis represents the ^{1}H chemical shift. The top spectrum is just GABP and 37A6. The bottom spectrum is GABP and 37A6 overlaid on top of the GABP and DMSO spectrum.
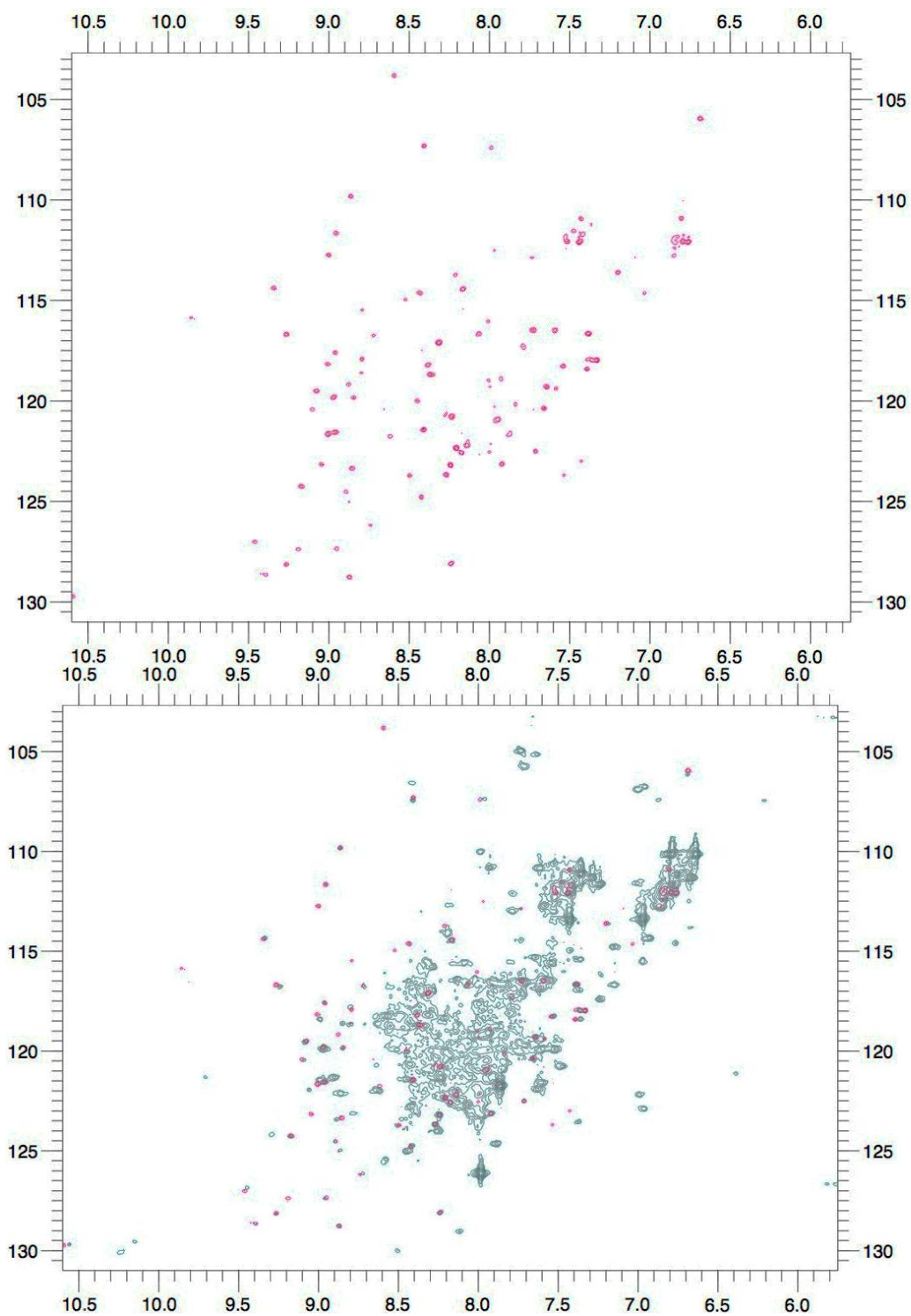
Figure 31: HSQC of GABPβ with 38F3. Buffer used was 200 mM arginine, 200 mM glutamic acid, 5 mM DTT, pH: 7.5. The Y-axis represents the ^{15}N chemical shift and the X-axis represents the ^{1}H chemical shift. The top spectrum is just GABP and 38F3. The bottom spectrum is GABP and 38F3 overlaid on top of the GABP and DMSO spectrum.

3.4 Conclusions

This study has focused on developing a protein-protein interaction inhibitor of GABPα and GABPβ as a therapy for GBM. To date, the results of this research indicate that the fragments and cysteine-reactive compounds, which have been identified with higher potency, have the potential to be the building blocks for a drug therapy for GBM. All of these compounds have IC_{50} values under 1 mM, with some under 150 µM. Many of these compounds have a particular aromaticity to their structure, insinuating that the GABP protein system favors molecules structured in this way. Moreover, after mutating the one cysteine in GABPβ to an alanine, it was shown that the cysteine-reactive compounds are not binding to that cysteine. They are either binding to other amino acids on GABPβ or to one of the nine cysteines on GABPα. Lastly, as shown in the HSQCs, when bound to GABPβ, some compounds are able to give a more resolved spectrum with well-dispersed peaks. This insinuates that free-GABPβ in solution is not stable, or the compounds are eliciting a conformational change that allows for a more resolved spectrum. When combining the potency of several of these fragments (particularly with the indole-containing compounds and variants of 26C6) or with a cysteine-reactive compound, the potency may reach a value where there would be a noticeable inhibition of GABPα binding to GABPβ and thus decrease the expression of TERT. With a decreased TERT expression, it is likely that GBM tumor size would decrease, offering a novel therapy to this malignant disease.

References

1.   Desrosiers, D.C., Peng, Z.Y. A binding free energy hot spot in the ankyrin repeat protein GABPbeta mediated protein-protein interaction. *Journal of Molecular Biology* **2005**, 354 (1), 375-384.

2.   Cheng, Y., Prusoff, W.H. Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50% inhibition (IC_{50}) of an enzymatic reaction. *Biochemical Pharmacology* **1973**, 22 (1), 3099-30108.

3.   Vullo, D., Innocenti, A., Nishimore, I., Pastorekm J., Scozzafava, A., Pastorekova, S., Supuran, C.T. Carbonic anhydrase inhibitors. Inhibition of the transmembrane isozyme XII with sulfonamides—a new target for the design of antitumor and antiglaucoma drugs? *Bioorganic and Medicinal Chemistry Letters* **2005**, 15 (4), 963-969.

4.   Abbate, F., Supuran, C.T., Scozzafava, A., Orioli, P., Stubbs, M.T., Klebe, G. Nonaromatic Sulfonamide Group as an Ideal Anchor for Potent Human Carbonic Anhydrase Inhibitors: Role of Hydrogen-Bonding Networks in Ligand Binding and Drug Design. *Journal of Medicinal Chemistry* **2002**, 45 (17), 3583-3587.

5.   Nishimori, I., Minakuchi, T., Morimoto, K., Sano, S., Onishi, S., Takeuchi, H., Vullo, D., Scozzafava, A., Supuran, C. Carbonic Anhydrase Inhibitors: DNA Cloning and Inhibition Studies of the alpha-Carbonic Anhydrase from Helicobacter pylori, A New Target for Developing Sulfonamide and Sulfamate Gastric Drugs. *Journal of Medicinal Chemistry* **2006**, 49 (6) =, 2117-2126.

6.   Supuran, C.T., Scozzafava, A., Conway, J. Carbonic Anhydrase: Its Inhibitors and Activators. In *CRC Enzyme Inhibitors Series*, CRC Press, New York, New York, **2005**, pp 1-56.

7.   De Sa Alves, F.R., Barreiro, E.J., Manssour Fraga, A. From Nature to Drug Discovery: The Indole Scaffold as a 'Privileged Structure'. *Mini Reviews in Medicinal Chemistry* **2009**, 9 (7), 782-793.

8.   Patil, R., Patil, S.A., Beaman, K.D., Patil, S.A. Indole molecules as inhibitors of tubulin polymerization: potential new anticancer agents, an update (2013-2015). *Future Medicinal Chemistry* **2016**, 8 (11), 1291-1316.

9.   Brancale, A., Silvestri, R. Indole, a core nucleus for potent inhibitors of tubulin polymerization. *Medicinal Research Reviews* **2007**, 27 (2), 209-238.

10.   Williams, T.M., Ciccarone, T.M., MacTough, S.C., Clarene, S.R., Balani, S.K., Condra, J.H., Emini, E.A., Goldman, M.E., Greeniee, W.J. 5-Chloro-3-(phenylsulfonyl) indole-2-carboxamide: a novel, non-nucleoside inhibitor of HIV-1 reverse transcriptase. *Journal of Medicinal Chemistry* **1993**, 36 (9), 1291-1294.

11.   Ristevski, S., O'Leary, D. A., Thornell, A. P., Owen, M. J., Kola, I., and Hertzog, P. J. The ETS transcription factor GABPalpha is essential for early embryogenesis. *Molecular Cellular Biology* **2004**, 24 (1), 5844-5849.

12.   Xue, H. H., Bollenbacher, J., Rovella, V., Tripuraneni, R., Du, Y.B., Liu, C.Y., Williams, A., McCoy, J. P., and Leonard, W. J. GA binding protein regulates interleukin 7 receptor alpha chain gene expression in T cells. *Nature Immunology* **2004**, 5 (1), 1036-1044.

13.   Wagner, G. Prospects for NMR of large proteins. *Journal of Biomolecular NMR* **1993**, 3 (4), 375-385.

14.   Kelly, A.E., Ou, H.D., Withers, R., Dötsch, V. Low-Conductivity Buffers for High-Sensitivity NMR Measurements. *Journal of the American Chemical Society* **2002**, 124 (1), 12013-12019.

15.   Hautbergue, G.M., Golovanov, A.P. Increasing the sensitivity of cryoprobe protein

NMR experiments by using the sole low-conductivity arginine glutamate salt. *Journal of Magnetic Resonance* **2008**, 191 (1), 335-339.

16.   Valente, J.J., Verma, K.S., Manning, M.C., Wilson, W.W., Henry, C.S. Second viral coefficient studies of cosolvent-induced protein self-interaction. *Biophysics Journal* **2005**, 89 (1), 4211-4218.

17.   Vedadi, M., Niesen, F.H., Allali-Hassani, A., Fedorov, O.Y., Finerty, P.J., Wasney, G.A., Yeung, R., Arrowsmith, C., Ball, L.J., Berglund, H., Hui, R., Marsden, B.D., Nordlund, P., Sundstron, M., Weigelt, J., Edwards, A.M. Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proceedings of the National Academy of Sciences* **2006**, 103 (1), 15835-15840.

18.   Blobel, J., Schmidl, S., Vidal, D., Nisius, L., Bernado, P., Millet, O., Brunner, E., Pons, M. Protein tyrosine phosphatase oligomerization studied by a combination of N-15 NMR relaxation and Xe-129 NMR. Effect of buffer containing arginine and glutamic acid. *Journal of the American Chemical Society* **2007**, 129 (1), 5946-5953.

19.   Placzek, W.J., Almeida, M.S., Wuthrich, K. NMR structure and functional characterization of a human cancer-related nucleoside triphosphatase. *Journal of Molecular Biology* **2007**, 367 (1), 788-801.

20.   Voehler, M.W., Collier, G., Young, J.K., Stone, M.P., Germann, M.W. Performance of cryogenic probes as a function of ionic strength and sample tube geometry. *Journal of Magnetic Resonance* **2006**, 183 (1), 102-109.