# INTRAMOLECULAR INTERACTIONS OF SELENOMETHIONINE IN PROTEINS:

# EXPANDING THE APPLICATIONS OF $^{77}$SE NMR IN BIOLOGICAL MACROMOLECULES

by

Michael Boeri

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Chemistry and Biochemistry

Summer 2021

Copyright 2021 Michael Boeri
All Rights Reserved

**INTRAMOLECULAR INTERACTIONS OF SELENOMETHIONINE IN PROTEINS:**

**EXPANDING THE APPLICATIONS OF $^{77}$SE NMR IN BIOLOGICAL MACROMOLECULES**

by

Michael Boeri

Approved: _____
Sharon Rozovsky, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Brian J. Bahnson, Ph.D.
Chair of the Department of Chemistry and Biochemistry

Approved: _____
John Pelesko, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____
Louis F. Rossi, Ph.D.
Dean of the Graduate College and Vice Provost for Graduate and Professional Education

## ACKNOWLEDGEMENTS

When I look back and reflect on the path that led me to writing this, I realize that without the overwhelming support, generosity, love, and patience of so many people given to me along the way, I without doubt would not have made it to this point. I want to give special thanks to Dr. Sharon Rozovsky for taking a chance and bringing me into your lab under unusual circumstances. Your mentorship was vital to my success, and I will always remember the lessons you taught me to be a good researcher. "An hour at your desk saves you a week in the lab."

I would be remiss if I did not give thanks to the late Dr. Douglas Cerasoli, my mentor who, through his unmatched generosity, allowed me to start this journey while continuing my work in his lab at USAMRICD. I must also give thanks to Dr. Shane Kasten, who picked up the pieces during a difficult time and assumed the role of mentorship which allowed me to continue my pursuit towards this degree.

To my lab members, past and present: Dr. Maggie Chen, Dr. Oshini Ekanayake, Dr. Rujin Cheng, and Shiping Xu. Thank you all for your support, guidance, and friendship.

To my parents, thank you for your love and encouragement during this process, you both were always there to remind me that if this was easy, everyone would do it. Finally, I want to thank my wife Chelsea for giving me the inspiration to enroll in this program, the encouragement and support to carry on, and the drive to make it through. I'll forever be thankful for her subtle yet unmistakable gift of a GRE study guide as a first

step to beginning my graduate program. I am so grateful for her patience and love throughout this entire process.

## TABLE OF CONTENTS

LIST OF TABLES……..……...…....vii
LIST OF FIGURES…….……..…...viii
LIST OF ABBREVIATIONS…….…….…..xi
ABSTRACT……..……xii

Chapter

1. SELENIUM IN BIOLOGY……...……...……..1

1.1 History and Occurrence of Selenium…….....……1
1.2 Selenium vs. Sulfur…… ……...…….....2
1.3 Biological Function of Selenium……4
1.4 The $^{77}$Se Atom…… ……...……5

REFERENCES…… …….……....8

2. SOLUTION $^{77}$Se NMR OF GB1 SeM VARIANTS……...……..12

2.1 Background……...……...…12
2.2 Methods……13

2.2.1 *Materials*……13
2.2.2 *Gene, Plasmid, Cloning, & Site-Directed Mutagenesis*……...……14
2.2.3 *Protein Expression: Natural Abundance $^{77}$Se*….. 14
2.2.4 *Protein Expression*: *$^{77}$Se Enriched*…….....…..…..15
2.2.5 *Protein Purification*…… ……..…...……...…16
2.2.6 *Sample Preparation*……...……18
2.2.7 *Electrospray Ionization Mass Spectrometry*…..…..19
2.2.8 *Data Acquisition & Analysis*…… ….……19

2.3 L5SeM…….……20
2.4 I6SeM……...……22
2.5 V29SeM……...……...…….23
2.6 A34SeM……...……...…….25
2.7 V39SeM……...……...…….26
2.8 V54SeM……...……...…….28
2.9 Conclusions……..30

REFERENCES…… ……34

3. 2-D NMR OF GB1 VARIANTS……35

3.1 Background……..35
3.2 Methods……35
3.3 $^{1}$H-$^{13}$C HSQC Spectra of GB1 SeM Variants……...37
3.4 $^{1}$H-$^{15}$N HSQC Spectra of GB1 SeM Variants……...43
3.5 Structural Perturbations in GB1 SeM Variants……50

3.5.1 *L5SeM Perturbations*……51
3.5.2 *I6SeM Perturbations*……...……..53
3.5.3 *V29SeM Perturbations*……...55
3.5.4 *A34SeM Perturbations*……...57
3.5.5 *V39SeM Perturbations*……...59
3.5.6 *V54SeM Perturbations*……...63

3.6 $^{1}$H-$^{77}$Se HMBC Spectra of GB1 SeM Variants……..…65

REFERENCES…… ……..72

4. CONCLUSIONS……..…… ……...73

4.1 Conclusions and Future Work.....……..73

REFERENCES…… ……..77

## LIST OF TABLES

Table 1. Physicochemical properties of sulfur and selenium……..…6

Table 2. Local environment and intramolecular interactions of GB1 SeM variants…….13

Table 3. Defined growth medium for natural-abundance SeM incorporation……...17

Table 4. Calculated and observed molecular masses of the GB1 variants……18

Table 5. Solution $^{77}$Se NMR properties of GB1 L5SeM…….……..21

Table 6. Solution $^{77}$Se NMR properties of GB1 I6SeM……22

Table 7. Solution $^{77}$Se NMR properties of GB1 V29SeM…….24

Table 8. Solution $^{77}$Se NMR properties of GB1 A34SeM…….26

Table 9. Solution $^{77}$Se NMR properties of GB1 V39SeM…….27

Table 10. Solution $^{77}$Se NMR properties of GB1 V54SeM……...29

## LIST OF FIGURES

Figure 1. $^{77}$Se spectra of L5SeM at 277 K, 293 K, and 303 K……..21

Figure 2. $^{77}$Se spectra of I6SeM at 277 K, 293 K, and 303 K……...……23

Figure 3. $^{77}$Se spectra of V29SeM at 277 K, 293 K, and 303 K……25

Figure 4. $^{77}$Se spectra of V39SeM at 277 K, 293 K, and 303 K……28

Figure 5. $^{77}$Se spectra of V54SeM at 27 7K, 293 K, and 303 K……30

Figure 6. $^{77}$Se spectra of all six GB1 SeM variants at 293 K…….33

Figure 7. $^{1}$H-$^{13}$C HSQC spectrum of L5SeM…….38

Figure 8. $^{1}$H-$^{13}$C HSQC spectrum of I6SeM……..39

Figure 9. $^{1}$H-$^{13}$C HSQC spectrum of V29SeM……..40

Figure 10. $^{1}$H-$^{13}$C HSQC spectrum of A34SeM……41

Figure 11. $^{1}$H-$^{13}$C HSQC spectrum of V39SeM……42

Figure 12. $^{1}$H-$^{13}$C HSQC spectrum of V54SeM……43

Figure 13. $^{1}$H-$^{15}$N HSQC spectrum of WT GB1……44

Figure 14. $^{1}$H-$^{15}$N HSQC spectrum of L5SeM……..45

Figure 15. $^{1}$H-$^{15}$N HSQC spectrum of I6SeM……...46

Figure 16. $^{1}$H-$^{15}$N HSQC spectrum of V29SeM……....47

Figure 17. $^{1}$H-$^{15}$N HSQC spectrum of A34SeM……48

Figure 18. $^{1}H$-$^{15}N$ HSQC spectrum of V39SeM……49

Figure 19. $^{1}H$-$^{15}N$ HSQC spectrum of V54SeM……50

Figure 20. Overlay of $^{1}H$-$^{15}N$ HSQC spectra of WT GB1 and L5SeM……52

Figure 21. Proximity of leucine-5 to perturbed residues in L5SeM……..53

Figure 22. Overlay of $^{1}H$-$^{15}N$ HSQC spectra of WT GB1 and I6SeM……..54

Figure 23. Proximity of isoleucine-6 to perturbed residues in I6SeM……...55

Figure 24. Overlay of $^{1}H$-$^{15}N$ HSQC spectra of WT GB1 and V29SeM……..56

Figure 25. Proximity of valine-29 to perturbed residues in V29SeM……57

Figure 26. Overlay of $^{1}H$-$^{15}N$ HSQC spectra of WT GB1 and A34SeM……..58

Figure 27. Proximity of alanine-34 to perturbed residues in A34SeM……..59

Figure 28. Overlay of $^{1}H$-$^{15}N$ HSQC spectra of WT GB1 and V39SeM……..61

Figure 29. Proximity of valine-39 to perturbed residues in V39SeM (1)……..62

Figure 30. Proximity of valine-39 to perturbed residues in V39SeM (2)……..63

Figure 31. Overlay of $^{1}H$-$^{15}N$ HSQC spectra of WT GB1 and V54SeM……..64

Figure 32. Proximity of valine-54 to perturbed residues in V54SeM……65

Figure 33. $^{1}H$-$^{77}Se$ HMBC spectrum of natural abundance $^{77}Se$ L5SeM……..66

Figure 34. $^{1}H$-$^{77}Se$ HMBC spectrum of natural abundance $^{77}Se$ I6SeM……...67

Figure 35. $^{1}H$-$^{77}Se$ HMBC spectrum of natural abundance $^{77}Se$ V29SeM……68

Figure 36. $^{1}H$-$^{77}Se$ HMBC spectrum of natural abundance $^{77}Se$ A34SeM……69

Figure 37. $^{1}H$-$^{77}Se$ HMBC spectrum of natural abundance $^{77}Se$ V39SeM……70

Figure 38. $^{1}H$-$^{77}Se$ HMBC spectrum of natural abundance $^{77}Se$ V54SeM……71

## LIST OF ABBREVIATIONS

| | |
|---|---|
| NMR | Nuclear magnetic resonance |
| SeM | Selenomethionine |
| Met | Methionine |
| GB1 | β1 Immunoglobulin Binding Domain of Protein G |
| HSQC | Heteronuclear Single Quantum Coherence |
| HMBC | Heteronuclear Multiple Bond Coherence |
| WT | Wild-type |
| UTR | Untranslated region |
| mRNA | Messenger ribonucleic acid |
| GPx | Glutathione peroxidase |
| TrxR | Thioredoxin reductase |
| ppm | Parts per million |
| LB | Line broadening |
| ssNMR | Solid-state nuclear magnetic resonance |
| CSA | Chemical shift anisotropy |
| PDB | Protein Data Bank |
| SNR | Signal-to-noise ratio |

## ABSTRACT

Selenium plays an essential role in the proper function of human physiology in the form of the uncommon amino acid selenomethionine (SeM), in which the sulfur atom in the side chain of methionine (Met) is replaced by a selenium atom in the same position. In an effort to better understand the dynamics of how selenium plays an essential role in proper physiological function, it was first necessary to understand the fundamental interactions that selenium experiences within its local environment in proteins. Selenium, in the form of its $^{77}$Se isotope, exists in nature at 7.6% natural abundance and has an intrinsic spin of ½, which makes it an ideal candidate for nuclear magnetic spectroscopy (NMR). NMR serves as a powerful tool for measuring a diverse range of chemical properties, including the response of a nucleus to its local environment through a number of different interactions. These interactions, such as solvent effects and various intramolecular interactions, can be effectively introduced via a rationally designed model system. This approach was carried out through the incorporation of a selenomethionine residue at six different positions within the well-characterized model protein, the β1 immunoglobulin binding domain of Protein G (GB1). Each of the selenomethionine substitutions were made at a location within the tertiary structure of GB1 that promoted a different chemical environment in order to better understand how selenium responds to these local environments. These six selenomethionine-containing GB1 variants were then evaluated using a range of one-dimensional and two-dimensional NMR studies, which yielded a broad library of information that provided a new understanding of how selenium interacts in different chemical environments.

# Chapter 1

# SELENIUM IN BIOLOGY

## 1.1 History and Occurrence of Selenium

Selenium is a nonmetal atomic element which is located in group 16, period 4 on the periodic table. It was first described by the Swedish chemist Jöns Jacob Berzelius in 1817 while he was working at a sulfuric acid processing plant. [1-3] Berzelius noticed a reddish sediment that had accumulated at the bottom of a vessel in which sulfuric acid was made, initially believing the substance to be a form of the element tellurium due to its characteristic odor; however, upon closer analysis, he determined that the substance was a new element. Berzelius made the observation that this new element had properties that were intermediate between sulfur and tellurium. He named the element selenium after the Greek moon goddess Selene, as the closely-related tellurium was named after the Earth.

Selenium is found in low amounts in the earth's crust at 0.05 parts per million (ppm) by weight, making it the 67$^{th}$ most abundant element in the crust. [4] Rarely is selenium found in nature in its elemental form, existing instead primarily in various metal ores. Some of the various selenium containing ores are ferrselitelite ($FeSe_2$), berzelianite ($Cu_2Se$), umangite ($Cu_3Se_2$), crookesite ($Cu_7$[Tl,Ag]$Se_4$), athabascaite ($Cu_5Se_4$), aguilarite ($Ag_4SeS$), and tiemannite (HgSe). [5] Selenium is obtained as a by-product of mining

other elements, primarily copper, lead, and iron, as there is no mechanism for the profitable mining of selenium.

## 1.2 Selenium vs. Sulfur

Located in group 16 of the periodic table, both sulfur and selenium belong to the chalcogen family, along with oxygen & tellurium ("light chalcogens" ), the radioactive element polonium, and the synthetic element livermorium ("heavy chalcogens"). Each of the four light chalcogens play a role in biology, with oxygen and sulfur assuming more diverse biological roles than selenium and tellurium. Chalcogens all possess six valence shell electrons and are most commonly found in the -2, +2, +4, and +6 oxidation states. In particular, there are a number of similarities between sulfur and selenium in terms of their chemical properties, which manifest themselves in the form of their roles in biology. Sulfur and selenium possess physicochemical similarities such as their ionic radius, electronegativity, and chemical bonding configurations; however, there are some notable differences between the two in a biological context. Physical properties of the sulfur and selenium atoms are summarized in Table 1.

Selenium is an essential micronutrient in animals, meaning that is essential to maintaining functional physiological processes. It is found in biology in the form of the atypical amino acids selenocysteine (Sec, U) and selenomethionine (SeM) in which a selenium atom takes the place of the sulfur atom in the more abundant amino acids cysteine (Cys, C) and methionine (Met, M). Although the structure of the selenium-containing variants of these amino acids share similar physical properties, there are notable differences that give rise to distinct properties. For example, the pKa of cysteine

is 8.3, whereas the pKa of Sec is 5.3, and this decreased electronegativity makes Sec a stronger nucleophile than Cys [6-8]. This marked difference in pKa between Cys and Sec is of particular importance in the physiological context, because at physiological pH, Sec will more frequently be found in the deprotonated selenolate form as opposed to the protonated thiol form in Cys. In addition, the selenol group of Sec is a much stronger nucleophile than its thiol counterpart in Cys, making it a much more effective electron donor in reactions with other atoms. This further magnifies the difference between Sec and Cys in a physiological context, with the deprotonated selenium atom in Sec being significantly more reactive than the sulfur atom in Cys, which remains mostly protonated as a function of its pKa being higher than the physiological pH of 7.4.

One of the many important roles of sulfur in the biological context is the formation of disulfide bonds. Disulfides function both as redox moieties in various proteins, as well as critically important components of protein tertiary structure where they are referred to as "disulfide bridges." Disulfide bonds are typically formed between the sulfur atoms of Cys residues in close proximity, although asymmetric disulfide bonds can occur, such as the intermolecular binding of thiolated ligands, or a single thiolated ligand to a Cys residue. The disulfide bond is sufficiently stronger than transient bonds such as hydrogen bonds, with a bond dissociation energy of 60 kcal/mol; however, this is roughly 40% weaker than the typical carbon-carbon or carbon-hydrogen bond, so it is generally referred to as a "weak bond" in the context of molecular binding [9]. In addition to disulfide bonds, similar analogous linkages also occur in various biological systems. Diselenide bonds are found in some cases where two Sec residues are in close proximity, as is seen in the case of selenoprotein L (SelL) [10]. Finally, selenylsulfide

bonds are formed between a sulfur and selenium atom and are also found in biology, as is seen in the mammalian protein thioredoxin reductase [11].

## 1.3 Biological Role of Selenium

Selenium plays an essential role in biological processes across all forms of life on Earth. Incorporation of Sec and SeM residues into proteins is carried out by two different routes. First, the incorporation of selenocysteine into organisms is done during the process of translation via a unique incorporation method involving a special RNA element known as the selenocysteine incorporation element (SECIS). The SECIS, found in the 3' untranslated region (UTR) of messenger ribonucleic acid (mRNA), is a stem-loop secondary structure that directs the translation machinery to incorporate a Sec residue upon recognition of a UGA codon. Typically, the UGA codon signals the translation machinery to stop the translation process. The base pairing configuration that comprises the SECIS which gives rise to its unique secondary structure is used as a template in bioinformatics as a tool to elucidate novel selenoproteins from a given genomic sequence. [12] The incorporation of SeM is instead a stochastic process, occurring via random incorporation. In these cases, atomic selenium replaces sulfur in the amino acid molecule and is incorporated using the same mechanism undergone by Cys. [13] There are a total of 25 proteins that have been identified in humans that possess Sec, and they are referred to as selenoproteins. [14] The function of several of these proteins remain unknown; however, it has been shown that knocking out the genes encoding some of these proteins is lethal [15-25].

One of selenium's core functions in biology is to act as a redox cofactor in antioxidant enzymes such as glutathione peroxidase (GPx) and thioredoxin reductase (TrxR). These enzymes, as well as others in the same functional family, are classified as oxidoreductases, which catalyze the reduction and/or oxidation of their respective coenzymes via a Sec residue which undergoes a transition from the selenol (Se-H) form to the selenenic acid (Se-OH), or vice-versa [26]. In the case of the GPx family of enzymes, their main function in the physiological context is the catalytic transformation of reactive oxygen species (ROS) into non-toxic metabolites, protecting the organism from oxidative stress and downstream effects of their presence in the body. It is thought that these enzymes play a role in certain neurodegenerative diseases, including Alzheimer's and Parkinson's diseases, through their prevention of disease state progression via their ability to protect from neuronal damage by scavenging various ROS species.

Additionally, a technique in X-ray crystallography known as multi-wavelength anomalous dispersion (MAD) is extensively utilized and involves the substitution of Met residues with SeM. This allows a simplified approach to determining the three dimensional structure of biological macromolecules such as proteins, but these selenium-incorporated proteins are not identified canonically as selenoproteins [27].

## 1.4 The $^{77}$Se Atom

In nature, selenium is found as a mixture of seven isotopes: $^{74}$Se, $^{76}$Se, $^{77}$Se, $^{78}$Se, $^{79}$Se, $^{80}$Se, and $^{82}$Se [28]. All of these are stable isotopes, with the exception of $^{79}$Se,

which is only found in trace amounts as a fission product in various uranium ores. In addition to the naturally-occurring isotopes, there are several synthetic isotopes of selenium, ranging from $^{64}$Se to $^{95}$Se. Of the seven naturally-occurring isotopes, $^{80}$Se is the most abundant in nature at approximately 49.6%.

It is the $^{77}$Se isotope that is of particular use in unlocking important information about intramolecular interactions within biological systems. This can be done through NMR spectroscopy of the $^{77}$Se atom incorporated into biological samples. $^{77}$Se is found in nature at 7.6% natural abundance, and most importantly, has a nuclear spin of +1/2. This property lends the $^{77}$Se atom a unique role in this area of chemistry, as the overwhelming majority of NMR spectroscopic experiments conducted in research leverage atoms of spin +1/2. This is due to the lesser technical and computational complexity that one encounters when conducting NMR spectroscopy of atoms whose spins are not +1/2.

**Table 1.** Physicochemical properties of sulfur and selenium [29]. (Courtesy of Dr. Qingqing Chen)

| **Physical Property** | **Sulfur** | **Selenium** |
|---|---|---|
| Atomic Radius (Å) | 1.80 | 1.90 |
| Covalent Radius (Å) | 1.04 | 1.18 |
| Electron Affinity (kJ/mol) | 200.41 | 194.97 |
| Electronegativity (Pauling Scale) | 2.58 | 2.55 |
| Oxidation States | -2, 0, +4, +6 | -2, 0, +4, +6 |

| Electron Configuration | [Ne] $3s^2 3p^4$ | [Ar] $3d^{10} 4s^2 4p^4$ |
|---|---|---|
| Free Amino Acid pKa | 8.25 | 5.24 |

## REFERENCES


1. Haynes, W. ed. CRC Handbook of Chemistry and Physics. CRC Press/ Taylor and Francis, 95th Edition, (Internet Version 2015).
2. Coursey, J., Schwab, D., Tsai, J., Dragoset, R. Atomic Weights and Isotopic Compositions (Version 4.1). (2015). National Institute of Standards and Technology, Gaithersburg, MD.
3. Cottrell, T. The Strengths of Chemical Bonds. Butterworth, London. (1954).
4. Jackson, M. Periodic Table Advanced. Bar Charts Inc. (2002). ISBN 978-1-57222-542-8.
5. Anthony, J., Bideuax, R., Bladh, K., Nichols, M. C. Handbook of Mineralogy. Mineralogical Society of America. Chantilly, VA. (2003).
6. Birrigen, M., Pilawa, S., Flohé, L. Trends in selenium biochemistry. (2002) Nat Prod Rep 19: 693-718. doi: 10.1039/b205802m
7. Labunskyy V, Hatfield DL and Gladyshev VN. Selenoproteins: molecular pathways and physiological roles. (2014). Physiol Rev 94 (3): 739-777. doi:10.1152/physrev.00039.2013.
8. Barage, S., Deobagkar, D., Baladhye, V. Characterization of structural and functional role of selenocysteine in selenoprotein H and its impact on DNA binding. (2018). Amino Acids 50 (5): 593-607. doi: 10.1007/s00726-018-2543-5.
9. Cremlyn, R. An Introduction to Organosulfur Chemistry. (1996). Chichester: John Wiley and Sons. ISBN 0-471-95512-4.
10. Valentina, A., Shchedrina, S., Novoselov, M., Gladyshev, M. V. Identification and characterization of a selenoprotein family containing a diselenide bond in a

redox motif Proceedings of the National Academy of Sciences. (2007). 104 (35) 13919-13924; https://doi.org/10.1073/pnas.0703448104

11. Lee, S., Bar-Noy, S., Kwon, J., Levine, R., Stadtman, T., Rhee, S. Mammalian thioredoxin reductase: Oxidation of the C-terminal cysteine/selenocysteine active site forms a thioselenide and replacement of selenium with sulfur markedly reduces catalytic activity. Proceedings of the National Academy of Sciences. (2000). 97 (6) 2521-2526. doi: 10. 1073/pnas.050579797.
12. Lambert, A., Lescure, A., Gautheret, D. A survey of metazoan selenocysteine insertion sequences. (2002). *Biochimie*. 84 (9): 953–959. doi:10.1016/S0300-9084(02)01441-4. PMID 12458087.
13. Block, E., Birringer, M., Jiang, W., Nakahodo, T., Thompson, H., Toscano, P., Uzar, H., Zhang, X., Zhu, Z. Allium chemistry: synthesis, natural occurrence, biological activity, and chemistry of Se-alk(en)ylselenocysteines and their γ-glutamyl derivatives and oxidation products. (2001). Journal of Agricultural and Food Chemistry. 49 (1): 458–70. doi:10.1021/jf001097b. PMID 11305255.
14. Avery, J. & Hoffmann, P. Selenium, selenoproteins, and immunity. Nutrients. (2018). 10 (9): 1203. doi:10.3390/nu10091203. PMC 6163284. PMID 30200430.
15. Birrigen M, Pilawa S and Flohé L. Trends in selenium biochemistry. Nat Prod Rep. (2002). 19: 693-718. doi: 10.1039/b205802m
16. Labunskyy V, Hatfield DL and Gladyshev VN. (2014) Selenoproteins: molecular pathways and physiological roles. Physiol Rev 94 (3): 739-777. doi: 10.1152/physrev.00039.2013

17. Barage SH, Deobagkar DD and Baladhye VB. (2018) Characterization of structural and functional role of selenocysteine in selenoprotein H and its impact on DNA binding. Amino Acids 50 (5): 593-607. doi: 10.1007/s00726-018-2543-5
18. Liu J, Zhang Z and Rozovsky S. (2014) Selenoprotein K form an intermolecular diselenide bond with unusually high redox potential. FEBS Lett 588 (18): 3311-3321. doi: 10.1016/j.febslet.2014.07.037
19. Ji S, Xia J and Xu H. (2016) Dynamic Chemistry of Selenium: Se–N and Se–Se Dynamic Covalent Bonds in Polymeric Systems. ACS Macro Lett 5 (1): 78-82. doi: 10.1021/acsmacrolett.5b00849
20. Fairweather-Tait SJ, Bao Y, Broadley MR, Collings R, Ford D, Hesketh JE and Hurst R. (2011) Selenium in human health and disease. Antioxid Redox Sign 14 (7): 1337-1383. doi: 10.1089/ars.2010.3275
21. Cardoso BR, Roberts BR, Bush AI and Hare DJ. (2015) Selenium, selenoproteins and neurodegenerative diseases. Metallomics 7: 1213-1228. doi: 10.1039/c5mt00075k
22. Kurokawa S and Berry MJ. (2013) Selenium. Role of the essential metalloid in health. Met Ions Life Sci 13: 499-534. doi: 10.1007/978-94-007-7500-8_16
23. Klein EA, Thompson IM, Tangen CM, Crowley JJ, Lucia MS, Goodman PJ, Minasian L, Ford LG, Parnes HL, Gaziano JM, Karp DD, Lieber MM, Walther PJ, Klotz L, Parsons JK, Chin JL, Darke AK, Lippman SM, Goodman GE, Meyskens FL and Baker LH. (2011) Vitamin E and the Risk of Prostate Cancer: Updated Results of The Selenium and Vitamin E Cancer Prevention Trial (SELECT). JAMA 306 (14): 1549-1556. doi: 10.1001/jama.2011.1437

24. Hatfield DL and Gladyshev VN. (2009) The Outcome of Selenium and Vitamin E Cancer Prevention Trial (SELECT) Reveals the Need for Better Understanding of Selenium Biology. Mol Interv 9 (1): 18-21. doi: 10.1124/mi.9.1.6
25. Norton R and Hoffmann PR. (2012) Selenium and asthma. Mol Aspects Med 33: 98-106. doi: 10.1016/j.mam.2011.10.003
26. Bhabak, K., Mugesh, G. Functional mimics of glutathione peroxidase: bioinspired synthetic antioxidants". (2010). Accounts of Chemical Research. 43 (11): 1408–19. doi:10.1021/ar100059g. PMID 20690615.
27. Hendrickson, W. Analysis of Protein Structure from Diffraction Measurement at Multiple Wavelengths. (1985). Transactions of the ACA. 21.
28. Audi, G., Kondey, F., Wang, M., Huang, W., Naimi, S. The NUBASE2016 evaluation of nuclear properties. (2017). Chinese Physics C. 41 (3): 030001. doi 10.1088/1674-1137/41/3/030001.
29. Johansson, L., Gafvelin, G., Arnér, ESJ. Selenocysteine in proteins—properties and biotechnological use. (2005). Biochim Biophys Act. 1726: 1-13. doi: 10.1016/j.bbagen.2005.05.010.

# Chapter 2

# SOLUTION $^{77}$Se NMR OF GB1 SeM VARIANTS

## 2.1 Background

In an effort to better understand the nature of how sulfur interacts within the tertiary structure of proteins, six variants of the model protein GB1 were investigated due to their unique and important interactions both individually and with other residues within the protein structure. The locations and expected intermolecular interactions of the six GB1 variants are summarized in Table 2. The selenium-substituted variants of GB1 were designed to explore a range of interactions, including the differences in local environment (solvent exposed vs. oriented into the hydrophobic core of the protein), position in the secondary structure (α-helix, β-sheet, loop), and intermolecular interactions (sulfur-π interaction, hydrophobic repulsion, dipole-dipole interaction). Because the natural abundance of $^{77}$Se is 7.63%, it was a viable approach to collect the solution NMR spectra of these protein variants without the need to isotopically enrich the samples with higher percentages of $^{77}$Se relative to the natural abundance. Knowing the intramolecular position of each variant within the GB1 molecule, acquisition of 1D $^{77}$Se solution NMR enables the observation of any trends that may be seen that relate to distinct intramolecular environments. Isotropic chemical shifts ($\delta_{iso}$), temperature dependence (277 K, 293 K, and 303 K), and T1 relaxation times were collected at a

magnetic field strength of 11.74 T. The results of the 1-D solution $^{77}$Se NMR of the six GB1 variants are summarized in sections 2.3-2.8.

Table 2. Local environment and intramolecular interactions of GB1 SeM variants. (Courtesy of Dr. Qingqing Chen)

| Mutation | Location | Interaction |
|---|---|---|
| L5M | Hydrophobic; close to Phe30 Trp43 | Sulfur – Aromatic |
| I6M | Solvent exposed near Lys4 and Glu15 | London Dispersion |
| V29M | Surface exposed on helix | London Dispersion |
| A34M | Hydrophobic; close to Phe30 Trp43 | Sulfur – Aromatic |
| V39M | Loop, oriented into core | Flexible |
| V54M | Hydrophobic; close to Trp43 | Sulfur – Aromatic |

## 2.2 Methods

The following sub-sections within this section (2.2) are reprinted from Chen et al. (2020) with permission from the author.

### *2.2.1 Materials*

Elemental $^{77}$Se (99.2% purity) was purchased from lsoflex, USA. Enzymes used for molecular biology were acquired from New England Biolabs. The pMHTDelta238 plasmid expressing Tobacco Etch Virus (TEV) protease fused to the cytoplasmic maltose binding protein (MBP) was purchased from the Protein Structure Initiative: Biology Materials Repository. Chromatography media was supplied by GE Healthcare Bio-

Sciences Corporation. All other chemicals and reagents were supplied by Sigma-Aldrich, Acros Organics, and GoldBio.

### *2.2.2 Gene, Plasmid, Cloning, & Site-Directed Mutagenesis*

The gene of the immunoglobulin-binding B 1 domain of streptococcal protein G (GB1) was codon optimized for expression in E. coli and synthesized by DNA2.0. This synthetic GB1 gene, which included an N-terminal hexahistidine-tag, was cloned into the bacterial expression vector pJexpress414 (DNA2.0). The hexahistidine-tag was followed by the TEV protease cleavage sequence ENLYFQG to allow for affinity tag removal of the hexahistidine-tag. Following cleavage, the tag-free GB 1 has Gly at position l instead of Met so that a sole Met residue can be inserted at other locations of the protein (abbreviated as SeM-free GBl). In addition, the engineered GBl sequence following cleavage with TEV protease also contained a substitution of Gln for Thr at position 2 since the majority of studies on GBl utilize this mutation to improve sample homogeneity. This vector was subsequently used as a template to introduce the Met at the desired position using PCR-based mutagenesis. Bacterial expression plasmids for SeM-free GB 1 and the six variants were made available through Add gene (plasmids 128685-128691).

### *2.2.3 Protein Expression: Natural Abundance $^{77}Se$*

Protein samples for solution NMR were prepared by utilizing the direct incorporation of the SeM amino acid into the protein chain, which results in a final sample containing natural abundance $^{77}$Se at 7.63%. To begin, the appropriate Met variant was transformed into E. coli BL21 (DE3) cells. To incorporate the expressed

protein with SeM, a defined growth medium was used that supplied no Cys or Met sources by the time protein expression was induced. In brief, an overnight starter culture in low salt LB broth (10 g of tryptone, 5 g of yeast extract, 5 g NaCl per L), supplemented with 100 µg/ ml ampicillin was inoculated and grown at 37 °C with good aeration. Next day, 10 ml of starter culture was used to inoculate 1 L of LB broth, supplemented with 100 µg/mL ampicillin. When optical density (OD) at 600 nm reaches 0.7, the cells were pelleted by centrifugation for 10 min at 4 °C and 5000 g. The supernatant was discarded, and the cell pellet was washed with ice cold sterilized water. The cells were pelleted again and resuspended in modified MDAG media from Studier as listed in Table 2 [1]. The temperature was lowered to 18 °C, and the cells were allowed to recover for 30 min. SeM was supplied at 0.1 g/L with no additional Met supplementation, and after an additional 10 min, protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The cells were grown for 20 h at 18 °C.

### *2.2.4 Protein Expression: $^{77}$Se Enriched*

Proteins were prepared as reported by Schaefer et. al. to incorporate elemental $^{77}$Se into proteins [1]. To begin, the appropriate Met variant was transformed into E. coli BL21(DE3) cells. The growth medium for the GB1 SeM variants are summarized in Table 3. A 10 mL starter culture supplemented with 5 mM $Na_2SO_4$ and 100 µg/mL of ampicillin was grown at 37 °C for 9 h. The starter culture contained 50 mM $Na_2HPO_4$, 50 mM $KH_2PO_4$, 10 mM NaCl, 50 mM NH 4 Cl, 2 mM MgCl 2, 0.2x metals, 1x vitamins, 0.4% glucose and 200 µM $CaCl_2$. The 1x trace metal solution contained 50 µM $FeCl_3$, 20 µM $CaCl_2$, 10 µM $MnCl_2$, 10 µM $ZnCl_2$, 2 µM $CoCl_2$, 2 µM $CuCl_2$, 2 µM $NiCl_2$, 2 µM $Na_2MoO_4$ and 2 µM $H_3BO_3$. The recipe for the 1000x vitamin solution followed the

details described by Studier [2]. It is worth noting that the components used in this growth medium, including the antibiotic selection, were kept sulfur-free except for $Na_2SO_4$, which was the sole source of sulfur. When the OD of the starter culture reached approximately 1.2, 1 mL of the starter culture was used to inoculate 1 L of growth media supplemented with 50 μM $Na_2SO_4$ and 100 μM/mL of ampicillin in 2.8 L baffled flasks. Cells were grown at 37 °C with good aeration until sulfur depletion, which was indicated by a plateau in OD at 600 nm at about 0.95, a value typically reached after 15 h. At this point, the cells were provided with a mixture of 20 μM $Na_2SO_4$ and 70 μM $Na_2SeO_3$. Protein expression was induced with 0.8 mM IPTG. Cells were grown for an additional 12 h at 37 °C then harvested by centrifugation as previously described. When $^{77}Se$ isotope was to be used, an appropriate amount of elemental $^{77}Se$ powder was carefully oxidized to $^{77}$selenite using nitric acid. No more than 100 μM of nitric acid can be present in the cell culture to minimize cellular toxicity.

### *2.2.5 Protein Purification*

The cells were pelleted as described and resuspended in 50 mM sodium phosphate, 200 mM NaCl, 10 mM imidazole, pH 8.0, supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1 mM benzamidine. Cells were disrupted by homogenization, and the cell debris was removed by centrifugation at 17,000 g for 1 h. The supernatant was pooled and loaded onto an immobilized metal affinity chromatography (IMAC) column that was pre-equilibrated with 5 column volumes (CV) of IMAC equilibration buffer (50 mM sodium phosphate, 500 mM NaCl, and 10 mM imidazole, pH 8.0). The column was washed with 15 CV of buffer supplemented with 30 mM imidazole, and His6-GB1 was eluted with buffer supplemented with 500 mM

imidazole. The yield of GB1 SeM is on average 50 mg per L of defined growth medium. The polyhistidine tag was cleaved by TEV protease at 1:3 molar ratio TEV:GB1 in 50 mM sodium phosphate, 200 mM NaCl, pH 8.0, supplemented with 5 mM of β-mercaptoethanol (BME). The progress of the cleavage was monitored by 16% Tricine-SDS-PAGE. After TEV cleavage, the protein mixture was loaded onto a second IMAC column in order to remove the TEV protease. The flow-through containing the tag-free GB1 was pooled and concentrated using an Amicon Ultra-15 centrifugal filter unit with MWCO of 3 kDa. The concentrated protein was further purified with an appropriate size exclusion column suitable for small proteins such as Superdex™ 75 or Sephacryl™ S-100. Fractions containing GB1 were collected and visualized on 16% Tricine-SDS-PAGE. Protein purity was higher than 95%. Protein concentration was determined by UV-Vis spectroscopy with an extinction coefficient of 9,970 $M^{-1}$ $cm^{-1}$, or 1.617 g $L^{-1}$. The ratio of SeM incorporation into GB1 was assessed by mass spectrometry, as summarized in Table 4. Using this method, SeM incorporation typically reached more than 95% with the Met variant being undetectable in some preps.

Table 3. Defined growth medium for natural-abundance SeM incorporation (Courtesy of Dr. Qingqing Chen).

| **Components** | **Volume (mL)** |
|---|---|
| Sterilized water | 940 |
| 1 M $Mg_2SO_4$ | 2 |
| 1000x metals* | 0.2 |
| 1000x vitamins* | 1 |
| 40% w/v glucose | 10 |
| 50 XM* | 20 |

| | |
|---|---|
| 25% aspartate | 10 |
| 17 amino acids (10 mg/mL)* | 20 |
| L-SeM | 0.1 g/L |

* The recipes for the preparation of 1000x metals, 1000x vitamins, 50x M salts and 17 amino acids are as described by Studier [2].

Table 4. Calculated and observed molecular masses of the GB1 variants (Courtesy of Dr. Qingqing Chen).

| **Protein** | **Theoretical (Da)** | **Observed (Da)** |
|---|---|---|
| L5SeM | 6213.6 | 6215.4 |
| I6SeM | 6213.6 | 6214.0 |
| V29SeM | 6227.7 | 6228.5 |
| A34SeM | 6255.7 | 6256.4 |
| V39SeM | 6227.7 | 6227.9 |
| V54SeM | 6227.7 | 6228.4 |

### *2.2.6 Sample Preparation for Solution NMR spectroscopy*

Solution NMR samples included proteins both with and without isotopic enrichment. This is a viable strategy since the natural abundance of $^{77}$Se is 7.63%. Solution-state NMR samples of $^{77}$Se-enriched GB1 contained 1.5 to 3 mM GB1 in 50 mM phosphate (pH 5.5), 1 mM EDTA, and 10% $D_2O$. $^{77}$Se incorporation into SeM was higher than 85% (the remaining fraction was sulfur in Met). Protein concentrations in the

natural abundance GB1 NMR sample ranged from 7 to 10 mM protein. The NMR tubes were either from Shigemi or, as in the case of air-tight tubes, from Wilmad-LabGlass.

### *2.2.7 Electrospray Ionization Mass Spectrometry*

Mass spectra of intact proteins were obtained using a Xevo G2-S QTOF on a Waters ACQUITY UPLC Protein BEH C4 reverse-phase column (300 Å, 1.7 μm, 2.1 mm x 150 mm). An acetonitrile gradient from 5%-95% was used with 0.1% formic acid, over a run time of 5 min and constant flow rate of 0.5 mL/min at room temperature. Spectra were acquired from m/z 350 to 2000, at a rate of 1 sec/scan. Each analysis was 25 min under constant flow rate of 0.2 mL/min at room temperature. Data were acquired from m/z 350 to 2500, at a rate of 1 sec/scan. The spectra were deconvoluted in MassLynx using a maximum entropy algorithm. All data are reported within 30 ppm.

### *2.2.8 Data Acquisition & Analysis*

Solution-state NMR spectra at 11.74 T were acquired using a Bruker AVANCE III HD spectrometer equipped with a Bruker 5 mm broadband inner coil probe. The number of scans for protein samples for isotopically labeled samples was 256 to 500 scans and for natural abundance samples between 5000 to 13,350 scans for all variants beside GB1 A34SeM, this more challenging mutant required 55,500 scans. The recycle delay was 1.0 s for all variants but 0.5 s was used for GB1 L5SeM. At 11.74 T, proton-decoupled 77 Se spectra were recorded at a basic transmitter frequency of 95.3824 MHz and an offset of 75 ppm, with a spectral width of 19.13 kHz, a 90 º pulse width of 12.13 μs, and WALTZ64 decoupling. The acquisition time was 1.0 s for all variants beside GB1 A34SeM and V54SeM for which 0.2 s was employed in some experiments. All

samples contained $^{77}Se$ at natural abundance, and the number of scans depending on sample concentration and linewidth, ranged from 6,300 to 10,000, beside GB1 V54SeM, which required up to 550,000 scans and GB1 A34SeM, which required up to 750,000 scans. All data were processed with an exponential apodization of about a 1/2 of the natural line width. The FWHM was obtained either without exponential apodization or alternatively with low line broadening (LB). 77 Se chemical shift was referenced by setting 1 M diphenyl diselenide (in $CDCl_3$) signal at 463.0 ppm. The error reported for measurement of the T1 relaxation time is the standard deviation of the curve fitting.

## 2.3 GB1 L5SeM

The leucine residue at residue 5 of WT GB1 was selected as a SeM incorporation site due to its orientation into the hydrophobic core of the protein, as well as its proximity to the aromatic phenylalanine in position 30 and tryptophan in position 43. Specifically, it can be inferred that the selenium atom will interact with the indole ring of the neighboring tryptophan residue. This in theory would elucidate the behavior of $^{77}Se$ in the hydrophobic core of a protein and the potential influence of the better-understood sulfur-π interaction.

The solution NMR data for GB1 L5SeM are summarized in Table 5. GB1 L5SeM follows the trend of the majority of GB1 SeM variants in that as temperature decreases, chemical shift also decreases (Figure 1). The shorter T1 relaxation time of 166 ms, relative to the other five GB1 SeM variants, indicates a less mobile arrangement, which makes sense due to residue 5 being oriented towards the core of GB1.

Table 5. Solution isotropic $^{77}$Se NMR properties of GB1 L5SeM

| **277 K** | **293 K** | **303 K** | **$\Delta_{ppm}$** | **$\Delta_{ppm}$/K** | **T1 (ms)** |
|---|---|---|---|---|---|
| 100.05 | 101.22 | 101.50 | 1.45 | 0.056 | 166 ± 5 |

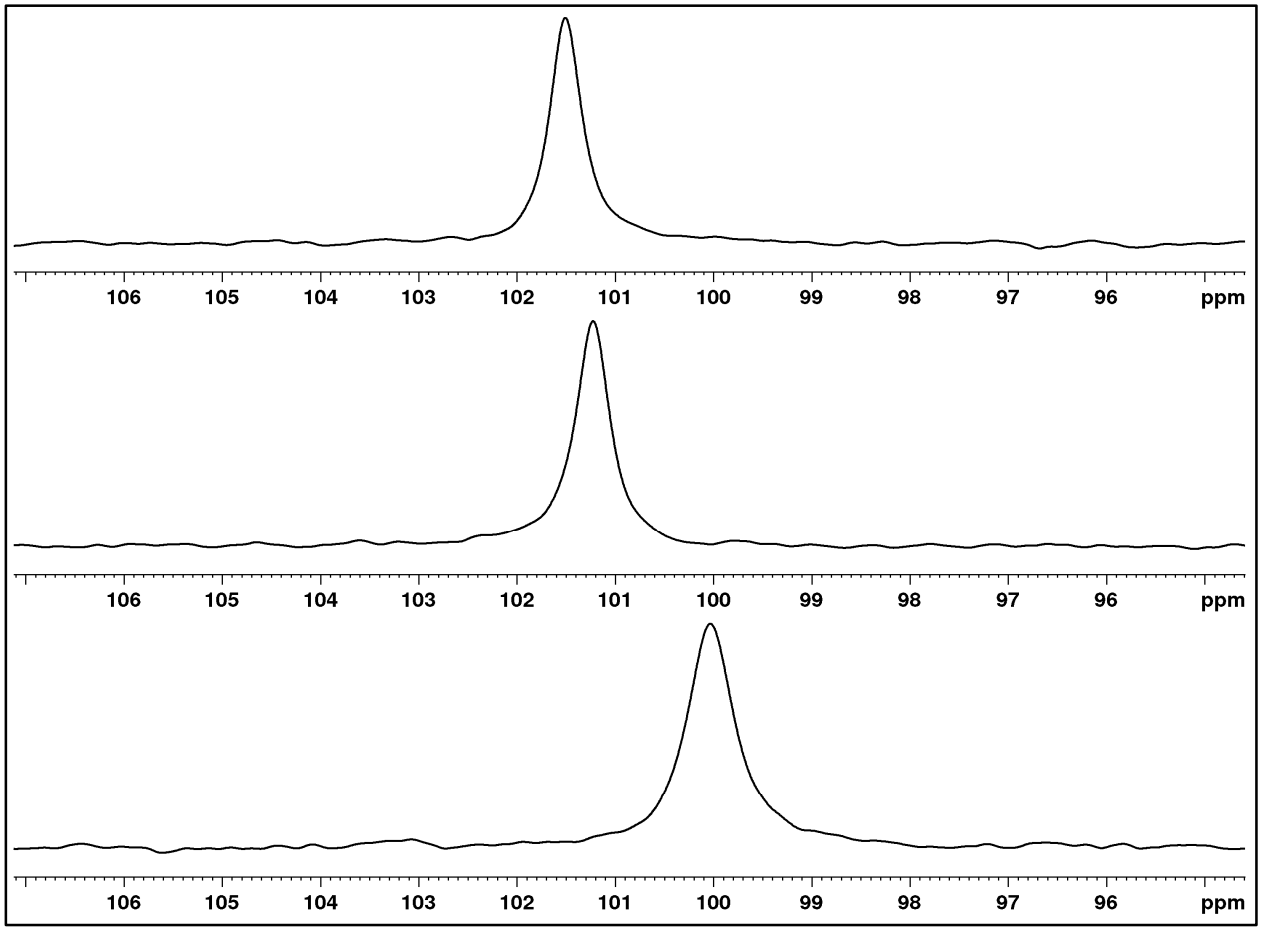


Figure 1. $^{77}$Se NMR spectrum of GB1 L5SeM at 303 K (top), 293 K (middle) and 277 K (bottom).

## 2.4 GB1 I6SeM

In order to observe the behavior of a SeM residue in a fully solvent exposed environment, the isoleucine residue at position 6 was selected for mutagenesis and SeM incorporation. I6 is located on the first β-strand and is not believed to be involved in any intramolecular hydrophobic interactions. This lends itself well to observe the behavior of an unconstrained SeM residue interacting primarily with the solvent matrix.

The solution NMR data for GB1 I6SeM are summarized in Table 6. Once again, GB1 I6SeM follow the trend seen in other GB1 SeM variants, displaying a decrease in chemical shift at decreasing temperatures; however, the change in chemical shift over the observed range of temperatures was by far the largest for GB1 I6SeM, with a total change in chemical shift of 4.94 ppm (0.19 ppm/K) (Figure 2). Additionally, the T1 relaxation time of GB1 I6SeM is the longest of all the observed GB1 SeM variants. This dynamic behavior of the SeM is representative of a residue with unconstrained mobility, unbound by the various intramolecular forces experienced by the other GB1 SeM variants in this study.

Table 6. Solution $^{77}$Se NMR properties of GB1 I6SeM

| **277 K** | **293 K** | **303 K** | **$\Delta_{ppm}$** | **$\Delta_{ppm}$/K** | **T1 (ms)** |
|---|---|---|---|---|---|
| 50.05 | 52.92 | 54.99 | 4.94 | 0.19 | 264 ± 3 |

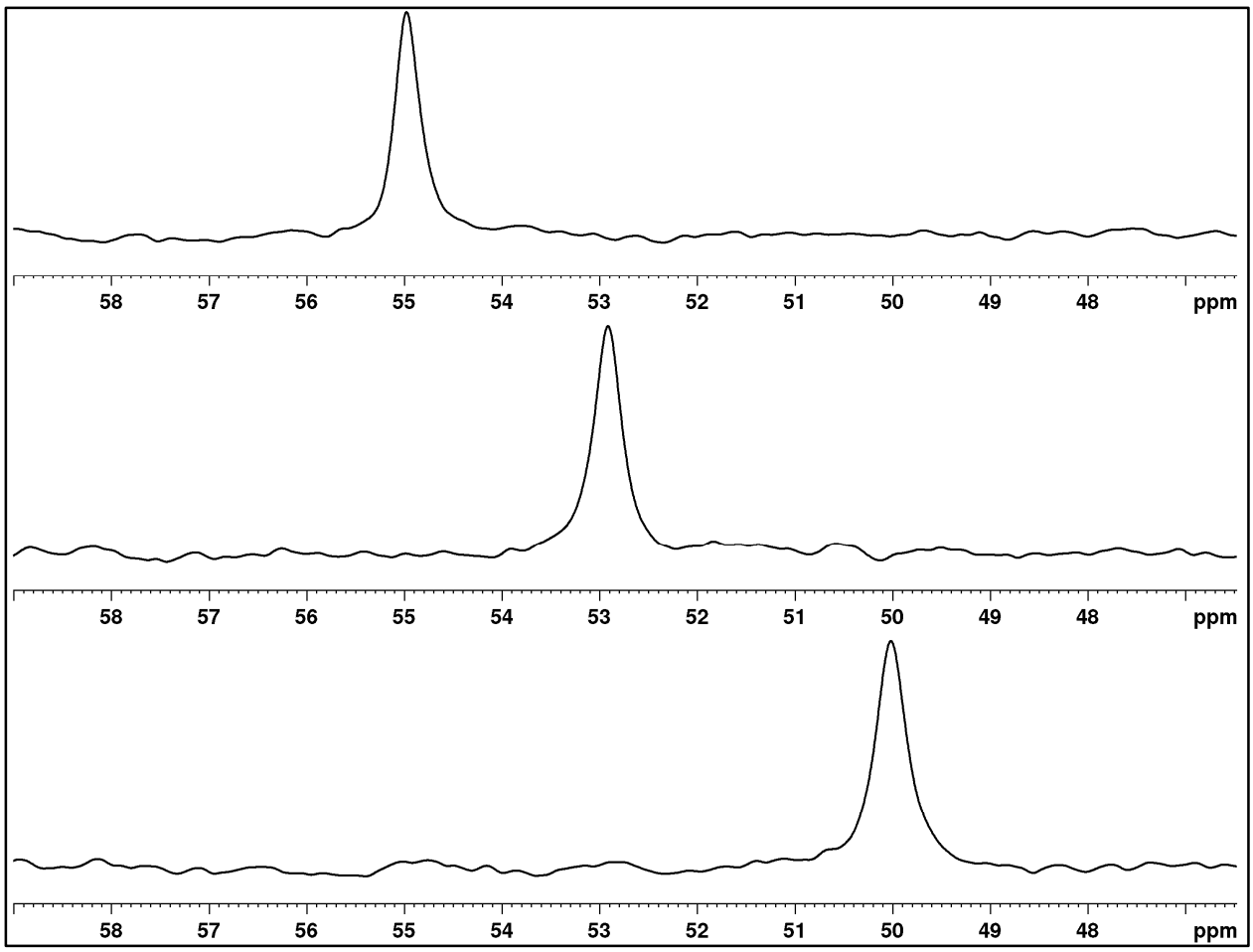


Figure 2. $^{77}$Se NMR spectrum of GB1 I6SeM at 303 K (top), 293 K (middle) and 277 K (bottom).

## 2.5 V29SeM

Similar to GB1 I6SeM, the valine at position 29 is GB1 was selected for mutagenesis due to its position in a solvent-exposed environment. In contrast to GB1 I6SeM, V29SeM is located on an α-helix and should have a high degree of mobility compared to other SeM variants in this study.

The solution NMR data for GB1 V29SeM are summarized in Table 7. The magnitude of chemical shift change as a function of temperature in GB1 V29SeM is only 20% of what is observed in GB1 I6SeM, the other variant situated in a solvent-exposed environment. Indeed, the chemical shift follows the common trend of decreasing as the temperature decreases; however, the magnitude of change is lower than that of some variants situated in the hydrophobic core of the protein (Figure 3). This suggests the possibility that in the absence of contributing intramolecular forces, conformational flexibility alone may not be the sole force affecting changes in chemical shift as a function of temperature. The T1 relaxation time of GB1 V29SeM was found to be intermediate relative to the other variants examined in this study.

Table 7. Solution $^{77}$Se NMR properties of GB1 V29SeM

| **277 K** | **293 K** | **303 K** | **$\Delta_{ppm}$** | **$\Delta_{ppm}$/K** | **T1 (ms)** |
|---|---|---|---|---|---|
| 73.62 | 74.13 | 74.53 | 0.91 | 0.035 | 189 ± 2 |

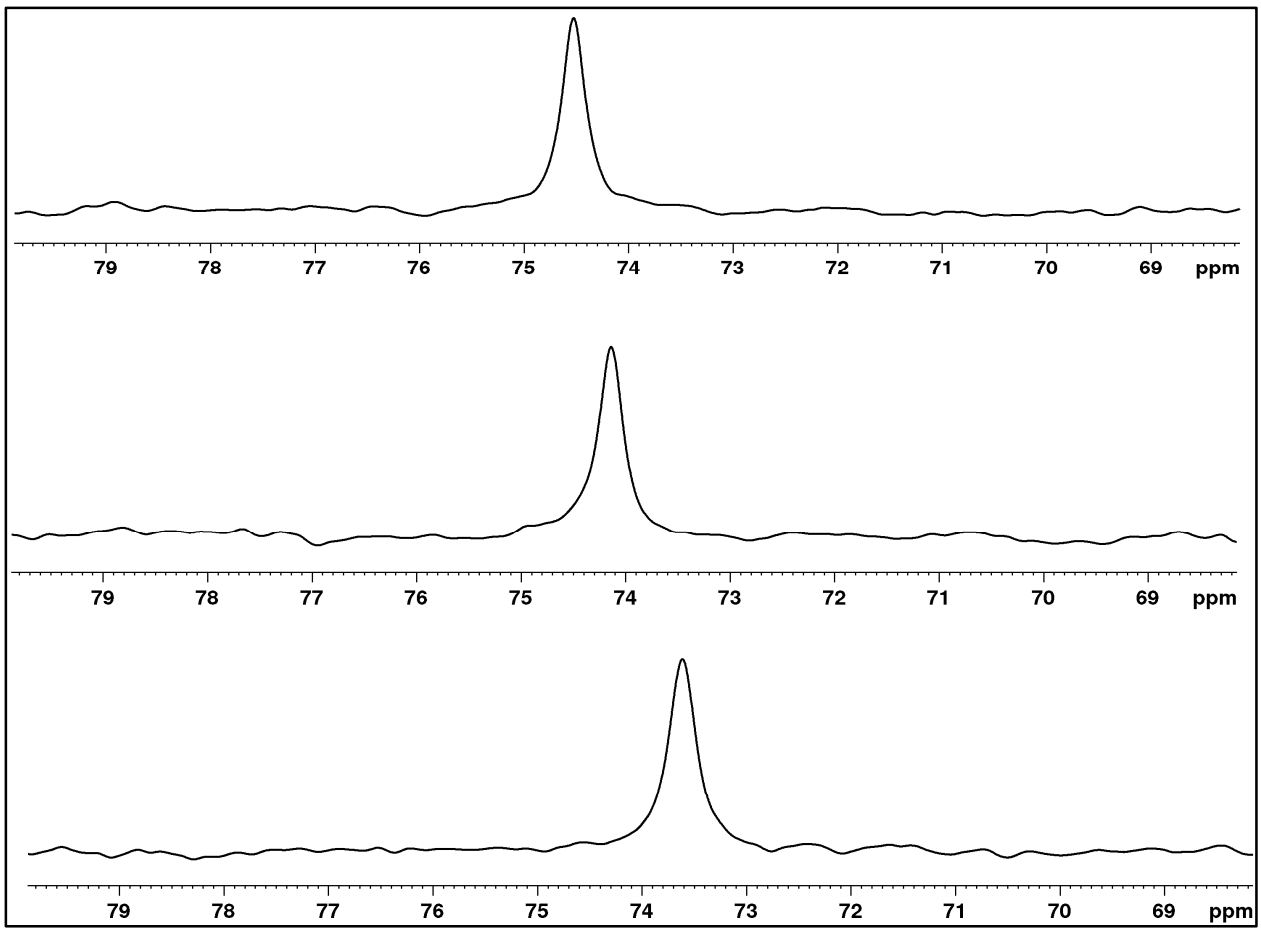


Figure 3. $^{77}$Se NMR spectrum of GB1 V29SeM at 303 K (top), 293 K (middle) and 277 K (bottom).

## 2.6 A34SeM

The alanine at position 34 in GB1, located on an α-helix, was selected for mutagenesis as another representative residue that will pack into the hydrophobic core of the protein. Of all six GB1 SeM variants produced for this study, the alanine to selenomethionine substitution was the most structurally disruptive due to the large discrepancy in size between the two amino acids. Incorporation of a SeM residue into

this position caused a perturbation in the tertiary structure of the GB1 molecule, displacing the nearby Trp43 and creating a solvent pocket. The formation of this solvent pocket allowed the incorporation of an external imidazole molecule to fit between the selenium atom and the ring within Trp43. Additional studies by Chen et al. utilizing x-ray crystallography determined that the predominant conformation of GB1 A34SeM included a nearby imidazole molecule [3].

The solution NMR data for GB1 A34SeM are summarized in Table 8. Due to the broad signal and prohibitively long acquisition times, much of the analogous data collected from other GB1 SeM could not be collected for GB1 A34SeM. The isotropic shift at 293 K was acquired, allowing incorporation of this variant into the overall understanding of chemical shift as a function of local environment.

Table 8. Solution $^{77}$Se NMR properties of GB1 A34SeM

| **277 K** | **293 K** | **303 K** | **$\Delta_{ppm}$** | **$\Delta_{ppm}$/K** | **T1 (ms)** |
|---|---|---|---|---|---|
| N/A | 74.8 | N/A | N/A | N/A | N/A |

## 2.7 V39SeM

Val39 was selected for SeM incorporation due to its relatively unique local environment on a loop structure that faces into the hydrophobic core of the protein. This should allow the residue to experience some degree of mobility due its placement on the low-order loop structure, but also some rigidity owing to its orientation into the core of the protein. In addition, the majority of residues surrounding the SeM at this position are

hydrophobic, so it is predicted that the primary interaction of this residue will be van der Waals forces.

The solution NMR data for GB1 V39SeM are summarized in Table 9. Once again, GB1 V39SeM behaves in a similar fashion to the other GB1 SeM variants, with a decrease in chemical shift as temperature decreases (Figure 4). The magnitude of change over the range of temperatures sampled is relatively intermediate, as is the T1 relaxation time.

Table 9. Solution $^{77}$Se NMR properties of GB1 V39SeM

| **277 K** | **293 K** | **303 K** | **$\Delta_{ppm}$** | **$\Delta_{ppm}$/K** | **T1 (ms)** |
|---|---|---|---|---|---|
| 80.93 | 82.09 | 82.61 | 1.68 | 0.065 | 198 ± 3 |

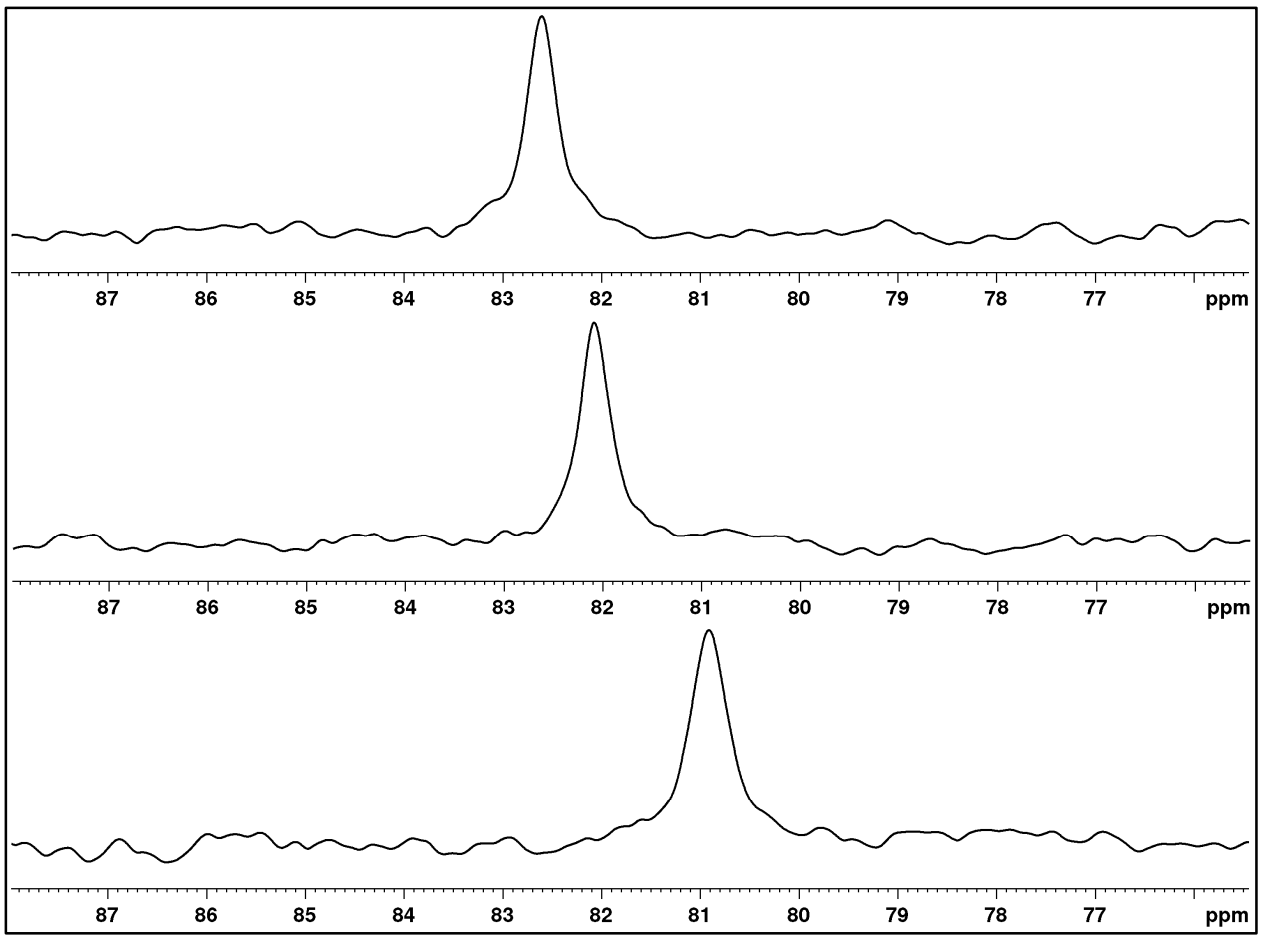


Figure 4. $^{77}$Se NMR spectrum of GB1 V39SeM at 303 K (top), 293 K (middle) and 277 K (bottom).

## 2.8 V54SeM

Val54 was selected for mutagenesis as yet another residue facing into the hydrophobic core of the protein. Located on an α-helix, Val54 is also likely in contact and interacting with Trp43, allowing a substitution to SeM to possibly participate in sulfur-aromatic interactions.

The solution NMR data for GB1 V54SeM are summarized in Table 10. In contrast to all the other GB1 SeM variants examined in this study, GB1 V54SeM is the only variant that displayed an inverse relationship between temperature and chemical shift (Figure 5). In this variant, as the temperature increased, the isotropic chemical shift decreased. Furthermore, GB1 V54SeM exhibited the shortest T1 relaxation time of all the variants examined in this study.

Table 10. Solution $^{77}$Se NMR properties of GB1 V39SeM

| **277 K** | **293 K** | **303 K** | **$\Delta_{ppm}$** | **$\Delta_{ppm}$/K** | **T1 (ms)** |
|---|---|---|---|---|---|
| 121.82 | 119.85 | 118.62 | -3.20 | -0.123 | 107 ± 2 |

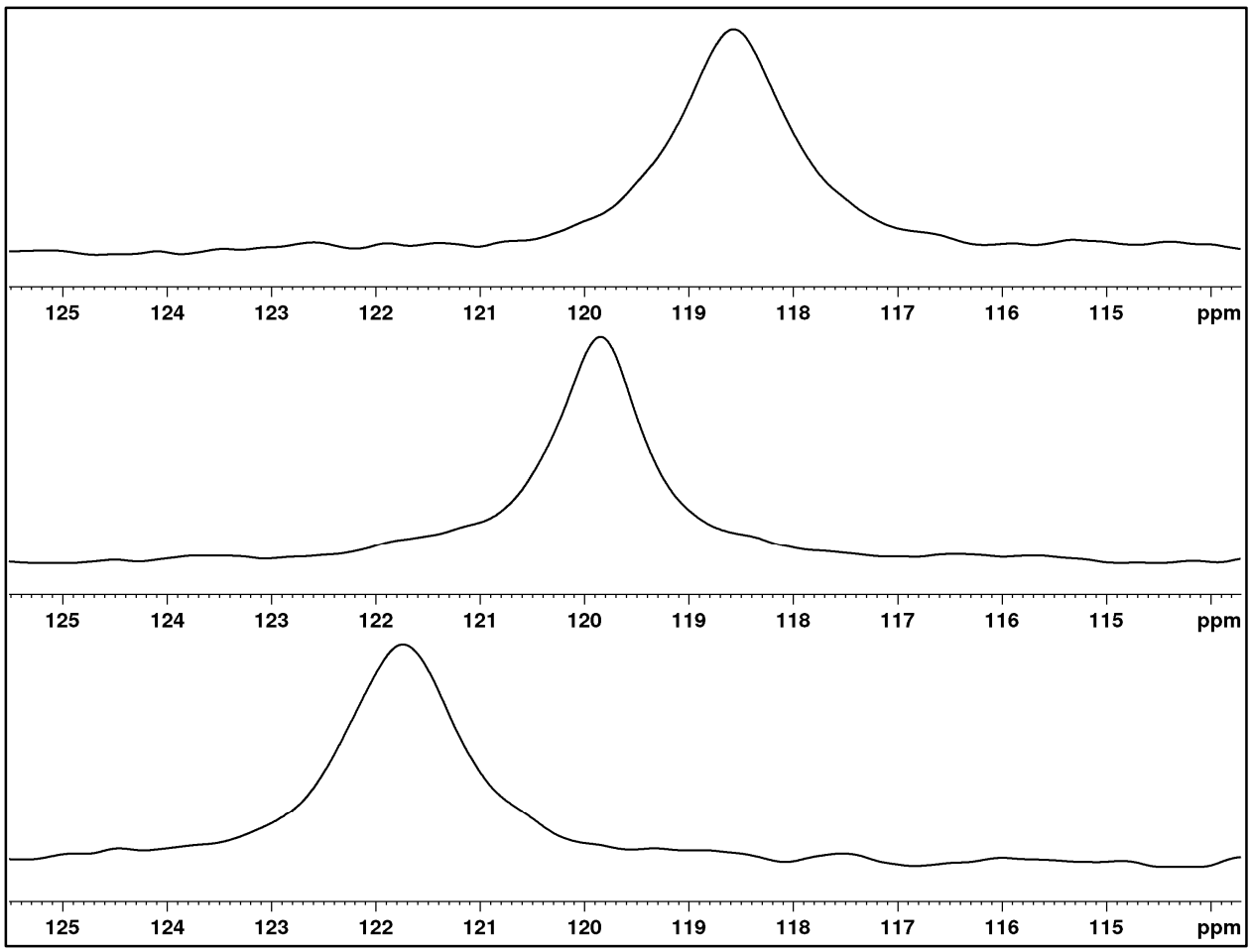


Figure 5. $^{77}$Se NMR spectrum of GB1 V54SeM at 303 K (top), 293 K (middle) and 277 K (bottom).

## 2.9 Conclusions

Isotropic chemical shifts were collected at 277 K, 293 K, and 303 K for five of six of the GB1 SeM variants produced for this study, as well as T1 relaxation times. The isotropic chemical shift for GB1 A34SeM was only acquired at 293 K due to signal to noise limitations, and it was not possible experimentally to obtain the T1 relaxation time.

The peaks of the isotropic chemical shifts for each of the six GB1 SeM variants at 293 K are shown in Figure 6.

While there are many contributing factors that can influence chemical shift of a heavy atom such as selenium (pH, solvent composition, temperature), one of the primary contributing factors is conformational mobility. The GB1 SeM variant that displayed the largest change in chemical shift over the temperature range examined was I6SeM, with an observed change of 0.19 ppm/K. If mobility is indeed the dominant force governing chemical shift range for a given atomic position, that would agree with observations for GB1 I6SeM due to its solvent-exposed local environment and lack of intramolecular forces acting upon it. Conversely, the $^{77}$Se chemical shifts of SeM variants oriented into the core of the protein were shown to respond to changes in temperature to a lesser extent.

In addition to the magnitude of change in chemical shift as a function of temperature, the trend of each variant's chemical shift either upfield or downfield as a function of temperature was observed. In all variants except GB1 V54SeM, chemical shift decreased as the temperature of the sample decreased. The reason for the inverse relationship observed in GB1 V54SeM may be attributable to changes in the aromatic stacking in the core of the protein as the temperature changes.

Longitudinal relaxation times were also examined, and once again, clear trends can be seen that correlate conformational flexibility with T1 relaxation times. Of the five variants for which T1 relaxation times were measured, GB1 I6SeM was found to have the longest T1, while GB1 V54SeM was found to have the shortest. It can be inferred that the long time scale of GB1 I6SeM's T1 relaxation is attributable to its conformational

flexibility, low steric hindrance, and high degree of freedom. Conversely, because GB1 V54SeM is oriented into the hydrophobic core of the protein and is conformationally bound by other intramolecular interactions, the relaxation time is made shorter. The T1 relaxation times acquired in this study once again point to conformational flexibility as a major factor in the NMR spectroscopic properties of selenium within a biological macromolecule, as well as in general.

Taken together, the 1D solution $^{77}$Se NMR data acquired from these six SeM-containing protein variants provides further studies utilizing intramolecular selenium NMR with benchmark chemical shift reference values, as well as informed interpretations of what the data tells the researcher in terms of the contributions of the local environment to the observed data. Further studies, correlating SeM properties with the protein local environment, will aid calculations of factors that control the NMR parameters.

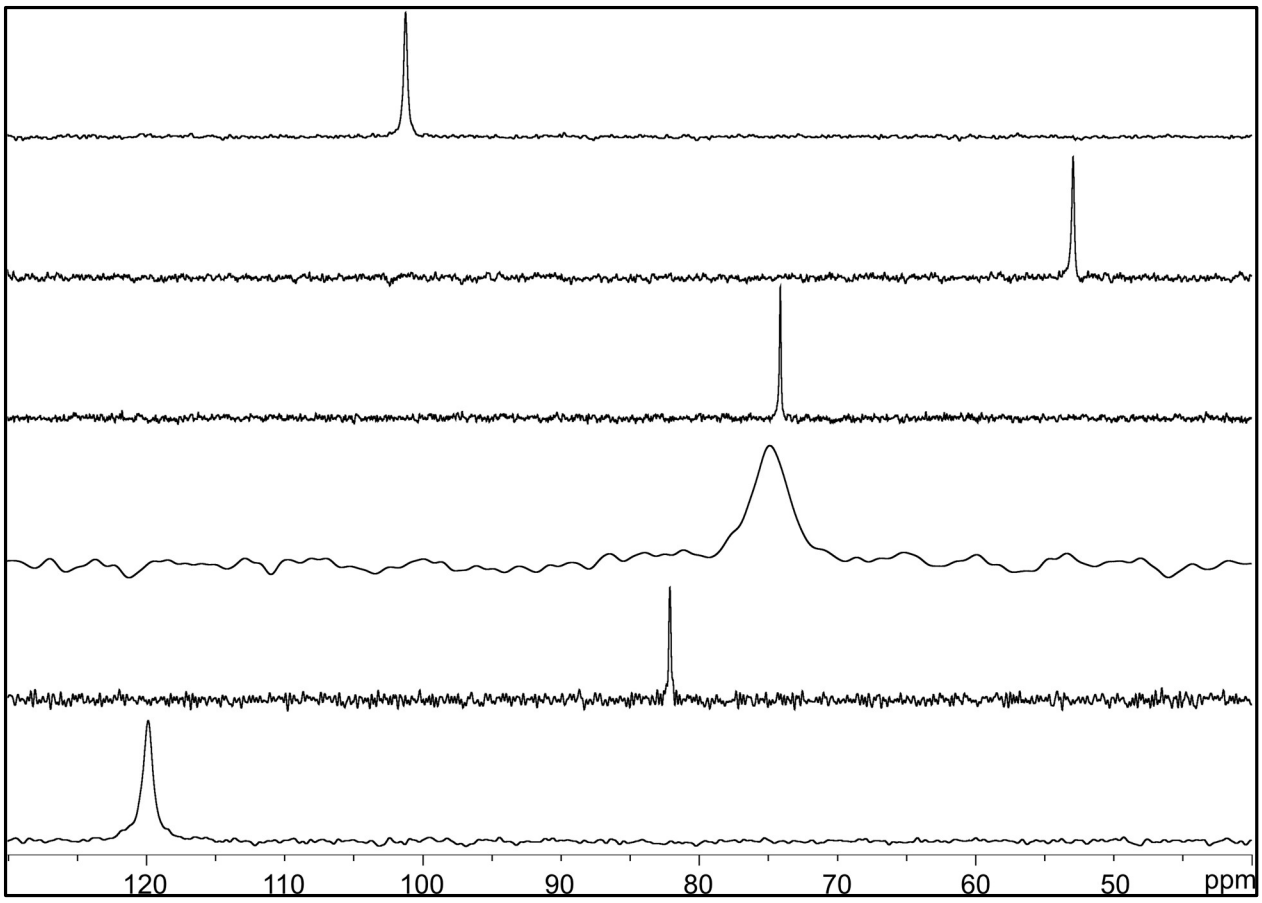


Figure 6. $^{77}$Se solution NMR acquired at 293 K. From the top to bottom: L5SeM, I6SeM, V29SeM, A34SeM, V39SeM and V54SeM.

## REFERENCES


1. Schaefer, S., Dong, M., Rubenstein, R., Wilkie, W., Bahnson, B., Thorpe, C., Rozovsky, S. $^{77}$Se enrichment of proteins expands the biological NMR toolbox. (2013). Journal of Molecular Biology. 425, 222-231.
2. Studier, F. W., Protein production by auto-induction in high-density shaking cultures. (2005). Protein Expression and Purification. 41 (1), 207-234.
3. Chen, Q., Xu, S., Lu, X., Boeri, M., Pepelyayeva, Y., Diaz, E., Soni, SD., Allaire, M., Forstner, M., Bahnson, B., Rozovsky, S. $^{77}$Se probes the protein environment of selenomethionine. (2020). Journal of Physical Chemistry B. 124, 601-616.

# Chapter 3

# 2-D NMR OF GB1 VARIANTS

## 3.1 Background

To ensure that GB1 variants are all well folded and to examine their mobility, 2D NMR correlation spectroscopy was conducted at natural abundance with carbon, nitrogen, and selenium. All six GB1 SeM variants were studied by $^{1}$H-$^{13}$C and $^{1}$H-$^{15}$N HSQC, as well as $^{1}$H-$^{77}$Se HMBC, while the $^{1}$H-$^{15}$N HSQC of WT GB1 was acquired as a reference to monitor perturbations of specific residues in each GB1 SeM variant. The information provided by these experiments includes insight into the nature of the perturbations that are introduced when substituting a WT residue with a SeM residue. Additionally, the $^{1}$H-$^{77}$Se HMBC spectra provide yet another point of reference in the literature that provides insight into the NMR spectroscopic properties of the selenium atom within a biological macromolecule.

## 3.2 Methods

2D $^{1}$H- $^{13}$C heteronuclear single-quantum coherence (HSQC) spectra at 295 K were recorded employing natural abundance GB1 samples at 11.74 T with frequencies of 500.1324 MHz and 125.7691 MHz for $^{1}$H and $^{13}$C, respectively. In the direct dimension ($^{1}$H), 32 scans with 1024 data points for each scan were accumulated. In the indirect

dimension ($^{13}$C), 256 data points were collected. The recovery delay between consecutive scans was set to 2.0 s. The $^{13}$C and $^{1}$H carrier frequencies were set to 90 ppm and 4.70 ppm, respectively, with a spectral width of 200 ppm and 16 ppm.

Natural abundance ($^{1}$H, $^{15}$N)-HSQC were recorded at 500.1324 MHz and 50.6838 MHz for $^{1}$H and $^{15}$N, respectively. In the direct dimension ($^{1}$H), 400-800 scans with 2,048 data points for each scan were accumulated. In the indirect dimension ($^{15}$N), 256 data points were collected. The recovery delay between consecutive scans was set to 2.0 s. The carrier frequencies for $^{15}$N and $^{1}$H were set to 120.00 ppm and 4.70 ppm, respectively, with the exception of GB1 I6SeM, where the $^{15}$N carrier frequency was set to 100 ppm. The spectral width for $^{1}$H was set to 16.02 ppm, while the spectral width for $^{15}$N ranged between 40 to100 ppm.

Natural abundance 2D $^{1}$H- $^{77}$Se heteronuclear multiple-bond correlation (HMBC) spectra at 298 K were recorded at 500.1324 MHz and 95.39 MHz for $^{1}$H and $^{77}$Se, respectively. In the direct dimension ($^{1}$H), 3000 - 4096 scans with 2048 data points for each scan were accumulated. In the indirect dimension ($^{77}$Se), 64 data points were collected. The recovery delay between consecutive scans was set to 2.0 s. The mixing time was set to 62.5 ms, which corresponds to J-coupling of 8 Hz. Recycle delay time was 1.5 sec. Spectral width for $^{1}$H was set to 16.02 ppm for all variants, while the spectral width for the $^{77}$Se dimension was set to 100 ppm for all variants except GB1 V54SeM, which was set to 150 ppm. The $^{1}$H offset frequency was 4.69 ppm for all variants, and 48 ppm for $^{77}$Se for all variants except GB1 L5SeM (99 ppm) and V54SeM (75 ppm). In all of the HSQC and HMBC experiments, $^{1}$H chemical shifts were referenced by setting the strongest DSS $^{1}$H signal to 0 ppm; $^{13}$C chemical shift are referenced by setting the

strongest DSS $^{13}C$ signal to 0 ppm; $^{15}N$ chemical are referenced by setting the liquid ammonia signal to 0 ppm; $^{77}Se$ chemical shift is referenced by setting 1 M diphenyl diselenide (in $CDCl_3$) signal to 463.0 ppm.

### 3.3 $^{1}H$-$^{13}C$ HSQC Spectra of GB1 SeM Variants

We initiated the studies by recording 2D NMR data acquired with the six GB1 SeM variants were $^{1}H$-$^{13}C$ spectra. This experiment serves as a method of validating that the tertiary structure of the protein being measured is intact, and in the case of this study in which a large amino acid residue is being substituted, it serves as a quality control method to ensure the amino acid substitution has not compromised the protein. The $^{1}H$-$^{13}C$ spectra of all six GB1 SeM variants are shown in Figures 7-12.

$^{1}H$-$^{13}C$ spectra can be generally divided into three representative categories of atoms within the protein structure based on the chemical shift distribution of the resonances. The resonances that are found far downfield in both dimensions (~ 6.1 – 7.7 ppm in the direct dimension and ~ 110 – 130 ppm in the indirect dimension) represent the atoms found in aromatic structures within the side chains of the aromatic residues. The second general population of resonances found in the intermediate range of the spectrum (~ 3.0 – 5.7 ppm in the direct dimension and ~ 45 – 70 ppm in the indirect dimension) are those pertaining to the atoms in the backbone of the polypeptide chain. Lastly, the resonances found far upfield in both dimensions (~ -0.5 – 3.0 ppm in the direct dimension and ~ 5 – 45 ppm in the indirect dimension) correspond to the many aliphatic atoms distributed throughout the protein structure.

As can be seen in each of the six GB1 SeM variant $^{1}H$-$^{13}C$ spectra, the tertiary structure of the protein does not appear to be compromised by the substitution of a SeM residue into the protein structure. Small perturbations can be seen in each of the three distinct subpopulations of resonances within each spectrum, but these are both minor and are expected. No major changes in chemical shift are seen; however, the information provided by $^{1}H$-$^{13}C$ spectra are generally qualitative in nature and do not provide residue-specific information in the case of this study. Other 2D techniques, including $^{1}H$-$^{15}N$ HSQC spectroscopy can provide more detail about residue-specific alterations.

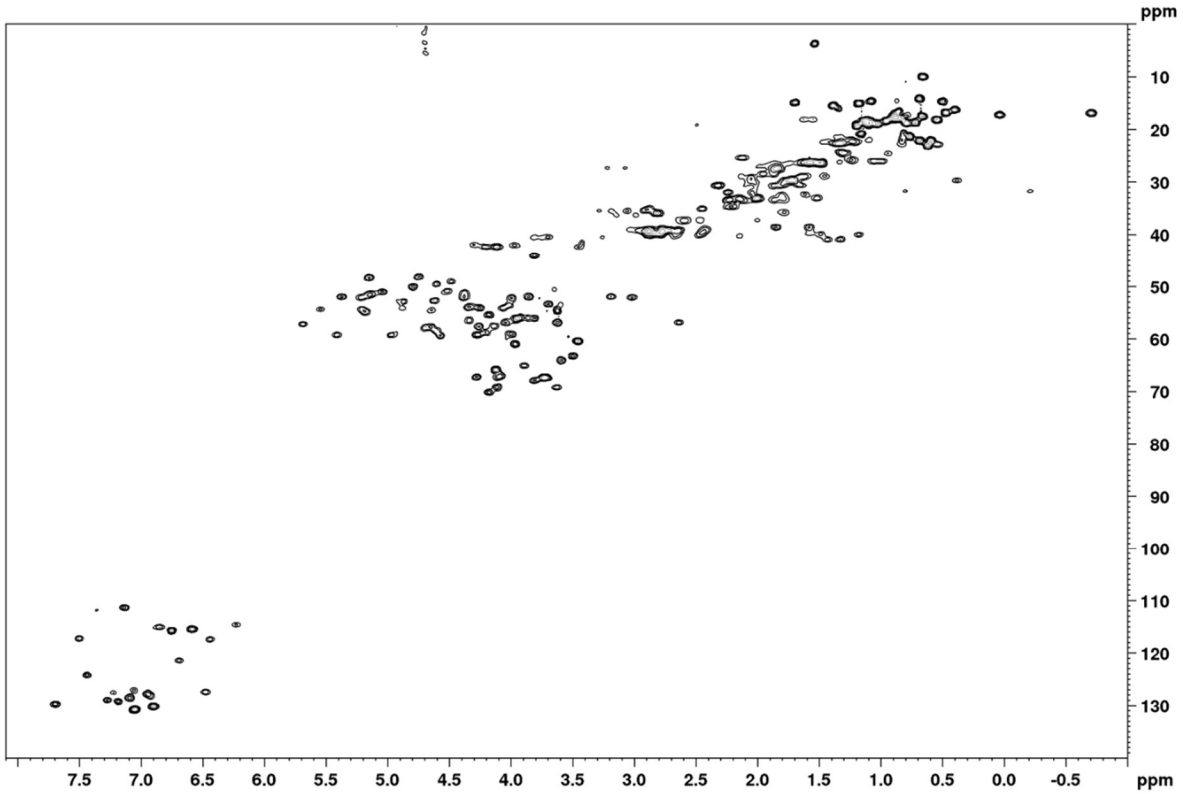


Figure 7. $^{1}H$-$^{13}C$ HSQC NMR spectrum of GB1 L5SeM.

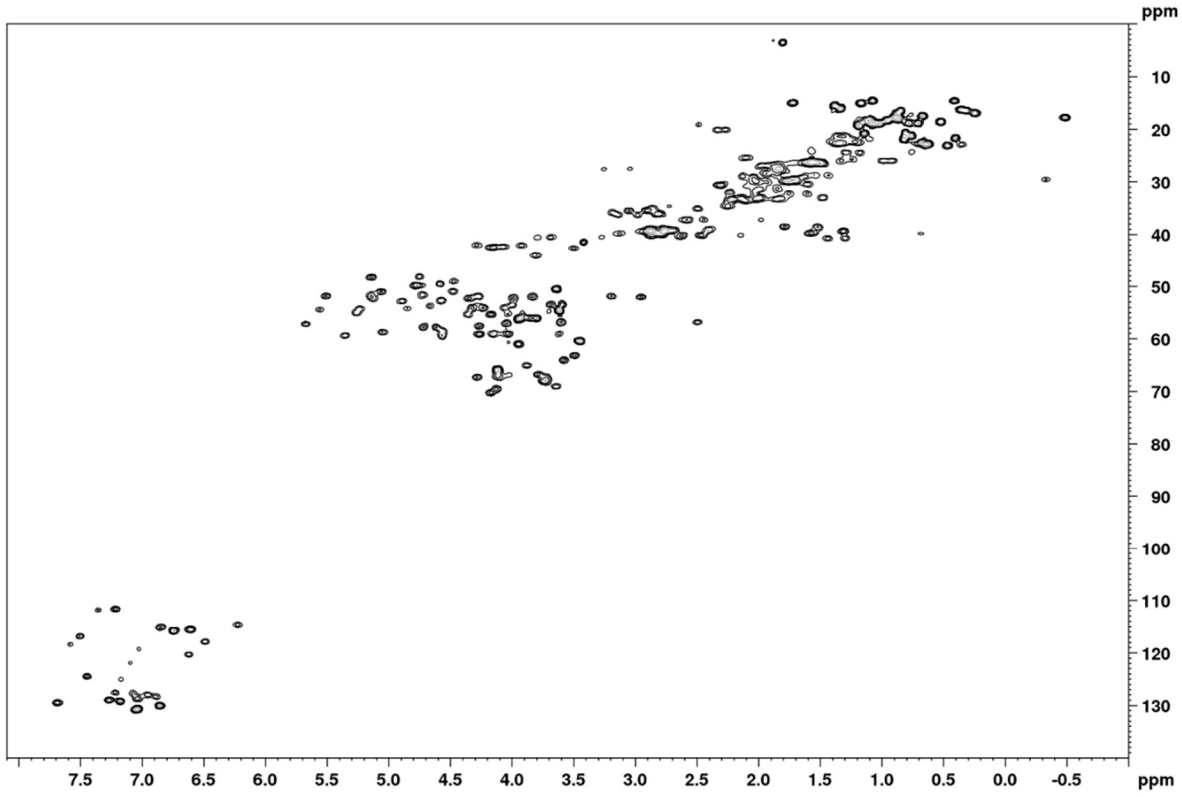


Figure 8. $^{1}H$-$^{13}C$ HSQC NMR spectrum of GB1 I6SeM.

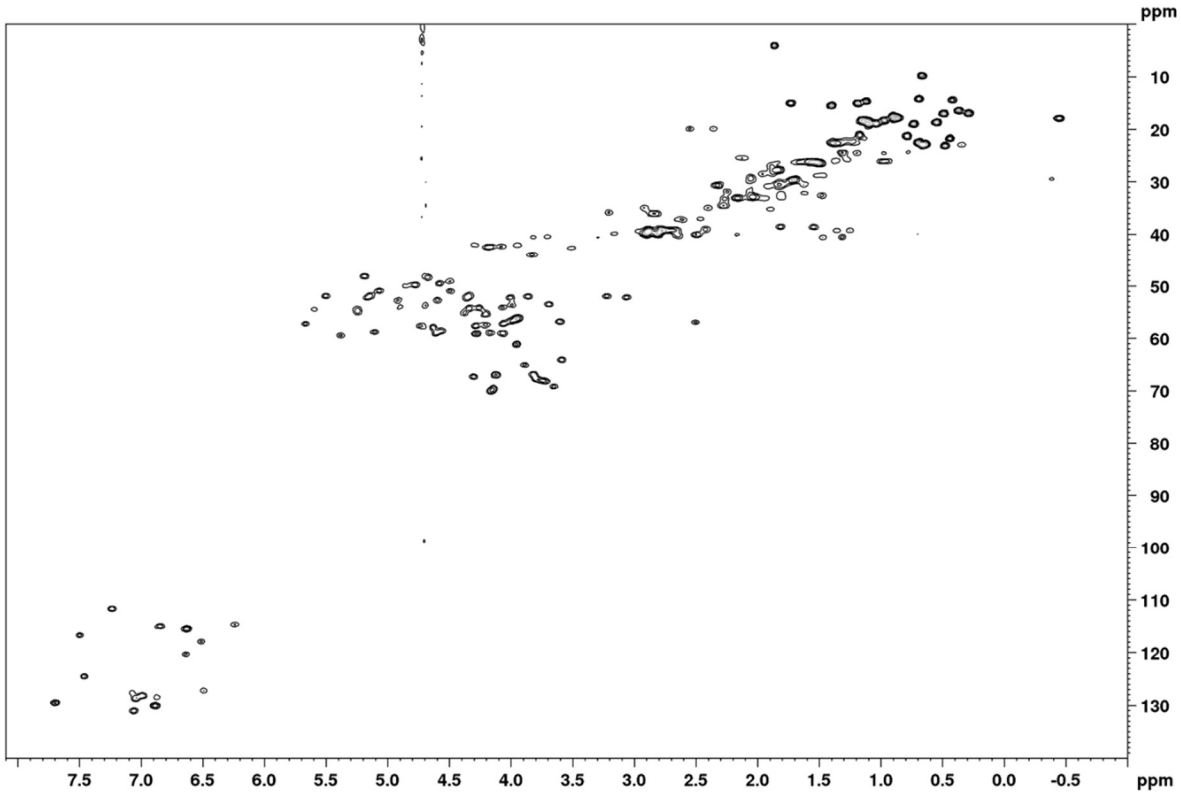


Figure 9. $^{1}H$-$^{13}C$ HSQC NMR spectrum of GB1 V29SeM.

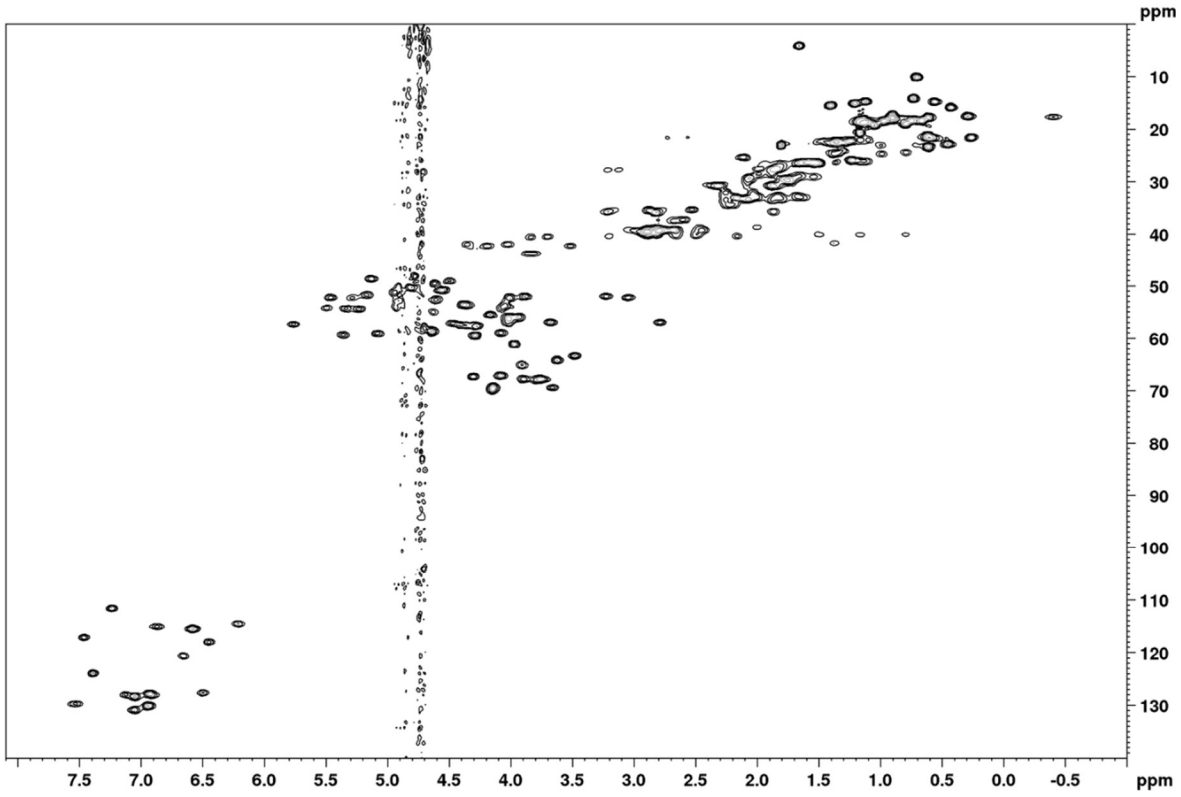


Figure 10. $^{1}H$-$^{13}C$ HSQC NMR spectrum of GB1 A34SeM.

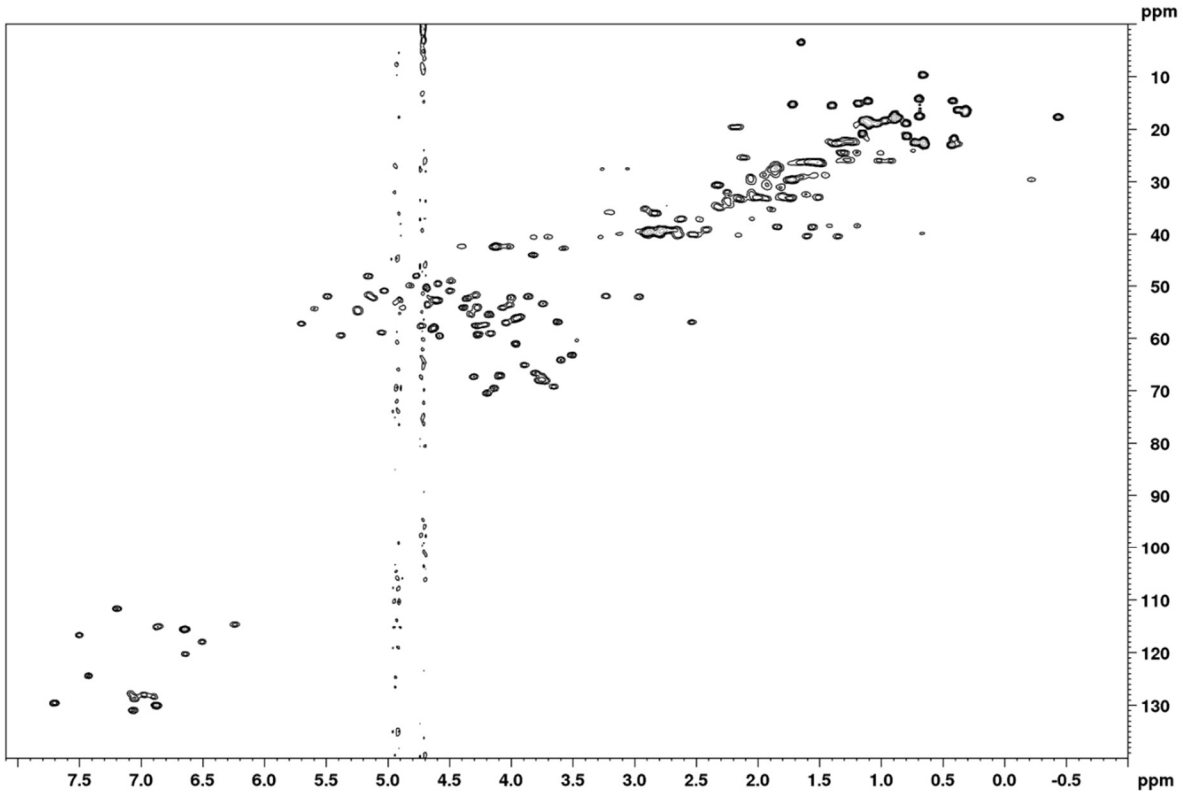


Figure 11. $^{1}H$-$^{13}C$ HSQC NMR spectrum of GB1 V39SeM.

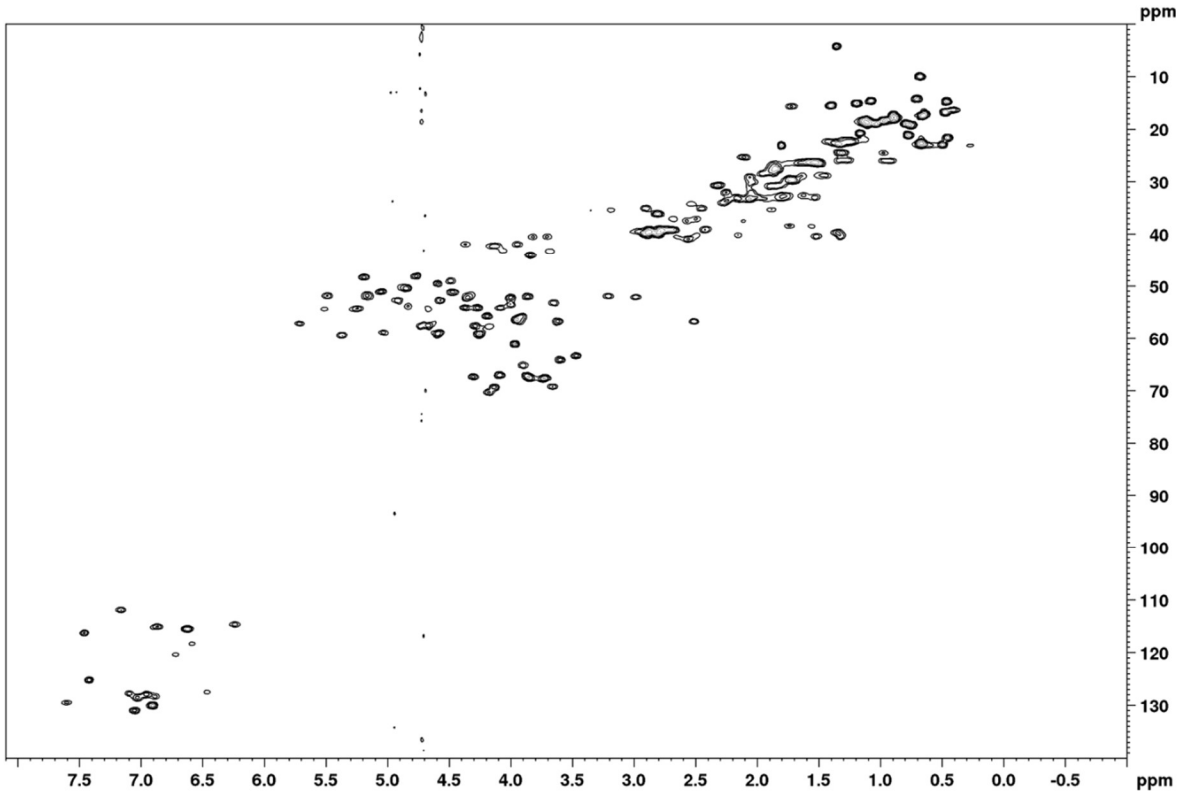


Figure 12. $^{1}$H-$^{13}$C HSQC NMR spectrum of GB1 V54SeM.

## 3.4 $^{1}$H-$^{15}$N HSQC Spectra of GB1 SeM Variants

In an attempt to elucidate the nature of the perturbations induced by SeM substitution in the GB1 variants, $^{1}$H-$^{15}$N HSQC spectra of all six variants were collected. In addition, in order to provide a reference for residue assignment in the variant spectra, the $^{1}$H-$^{15}$N HSQC spectrum of WT GB1 was acquired and each resonance within its spectra was assigned according to Wilton et al. (PDB: 2J52; BMRB: 7280) [1-2]. The acquisition of $^{1}$H-$^{15}$N HSQC spectra of the GB1 variants serves as a way to confirm the integrity and proper folding of the expressed proteins, in addition to the precise structural

changes that arise as a result of their introduction. The $^{1}H$-$^{15}N$ HSQC spectra of WT GB1 and all six GB1 SeM variants are shown in Figures 13-19.

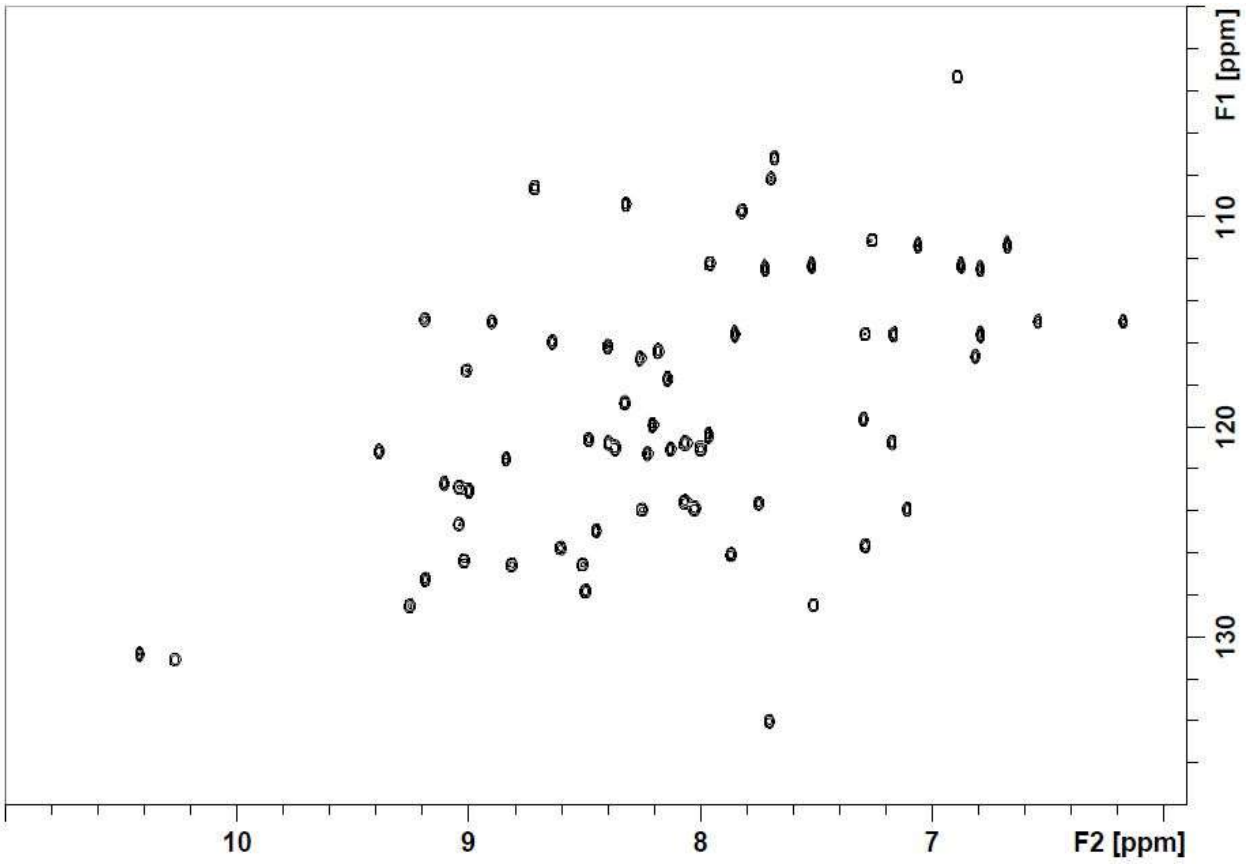


Figure 13. $^{1}H$-$^{15}N$ HSQC spectrum of WT GB1

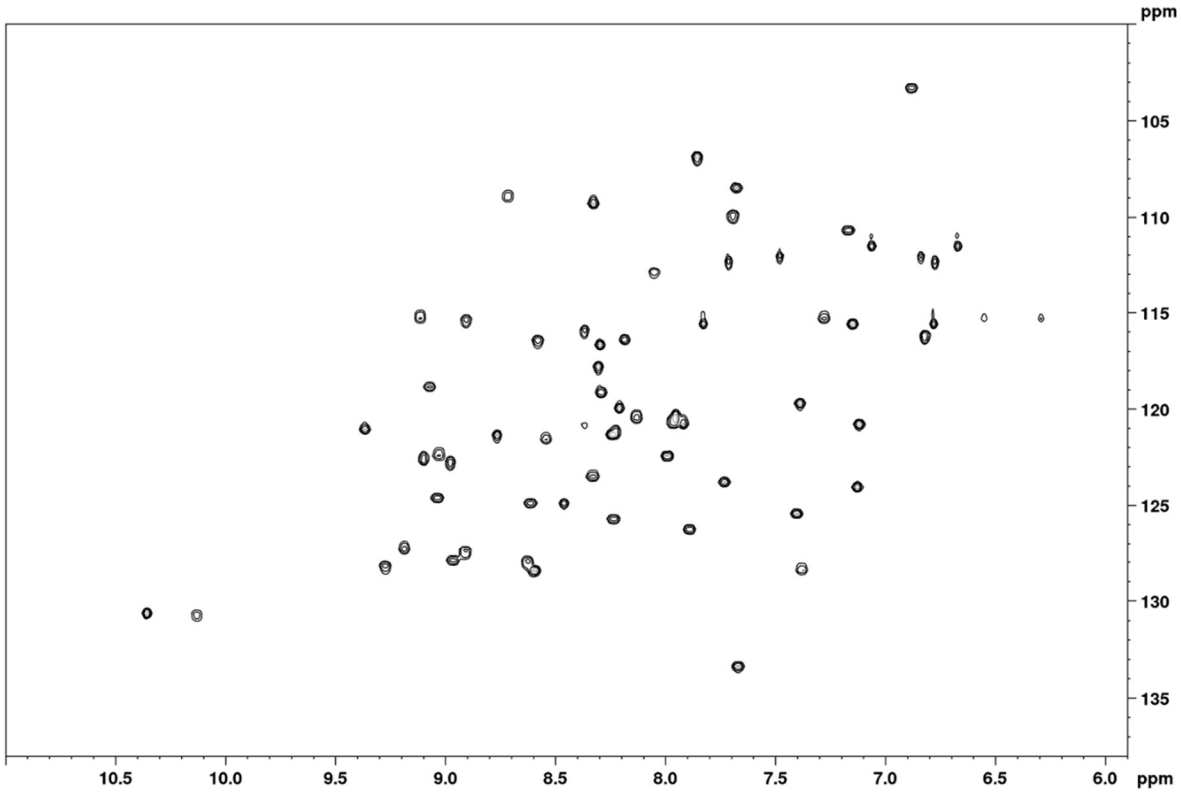


Figure 14. $^{1}H$-$^{15}N$ HSQC spectrum of GB1 L5SeM

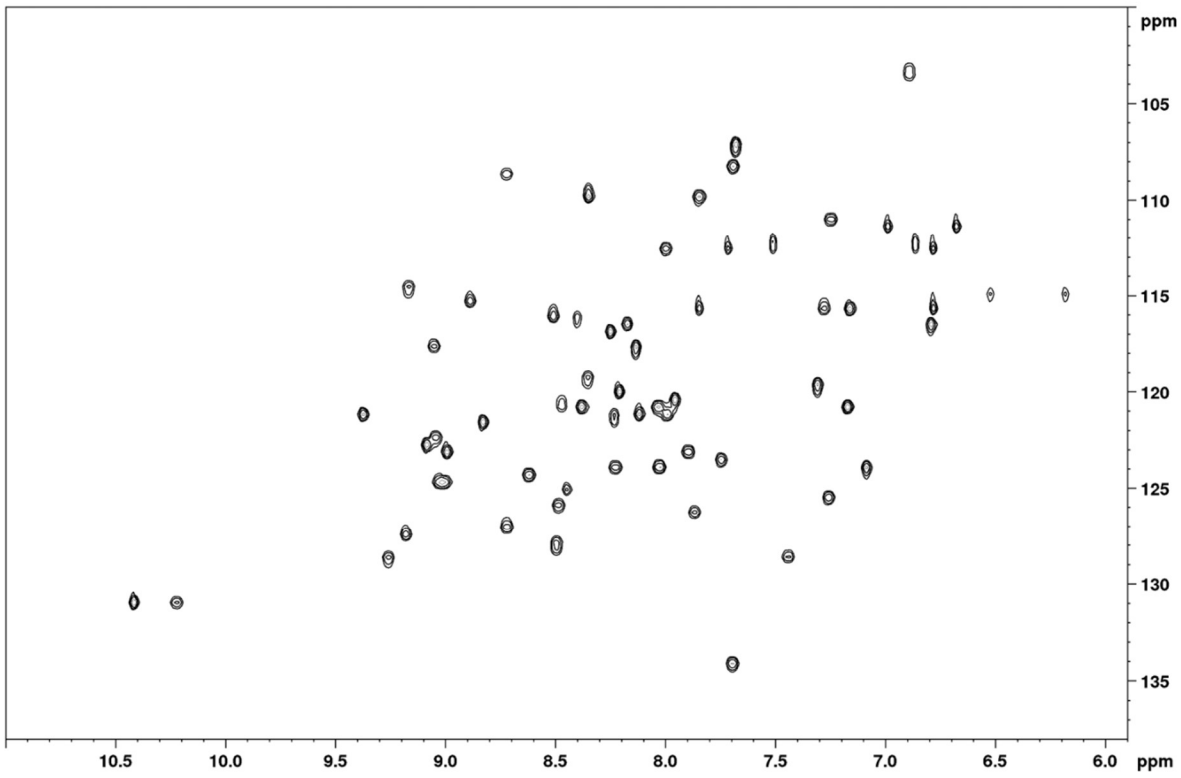


Figure 15. $^{1}H$-$^{15}N$ HSQC spectrum of GB1 I6SeM

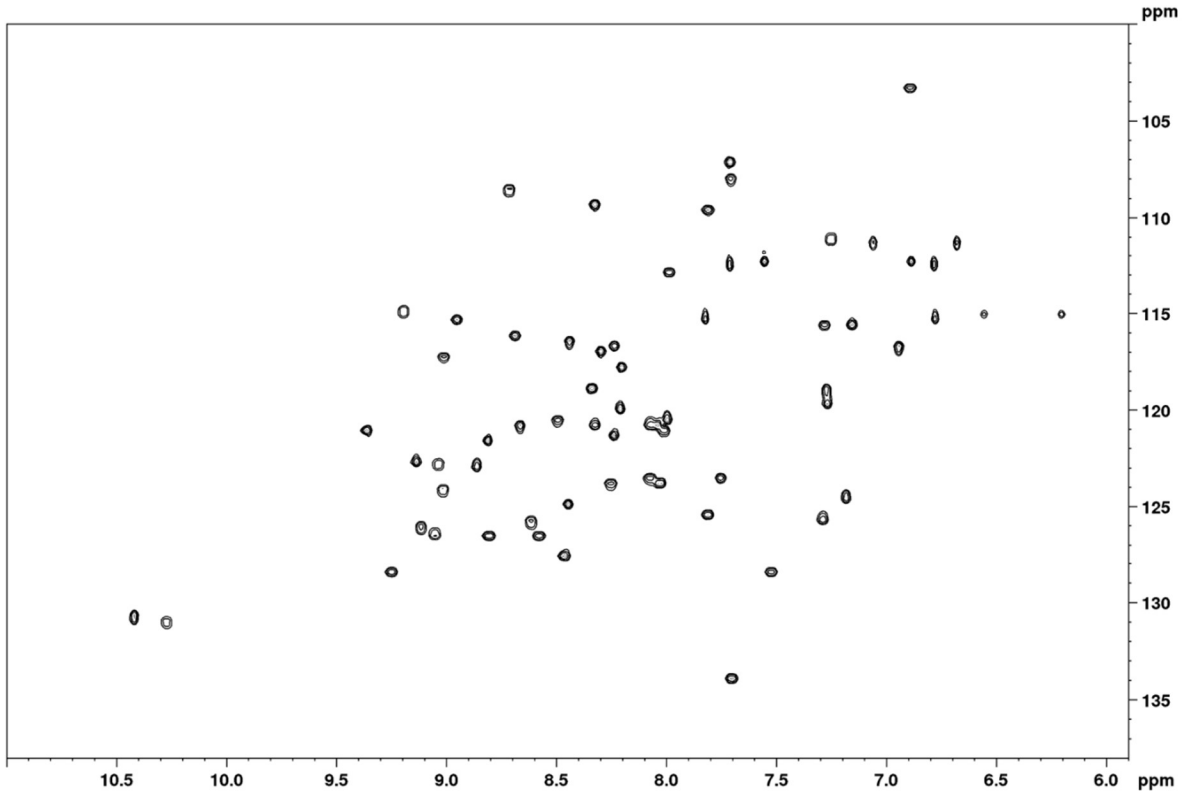


Figure 16. $^{1}H$-$^{15}N$ HSQC spectrum of GB1 V29SeM

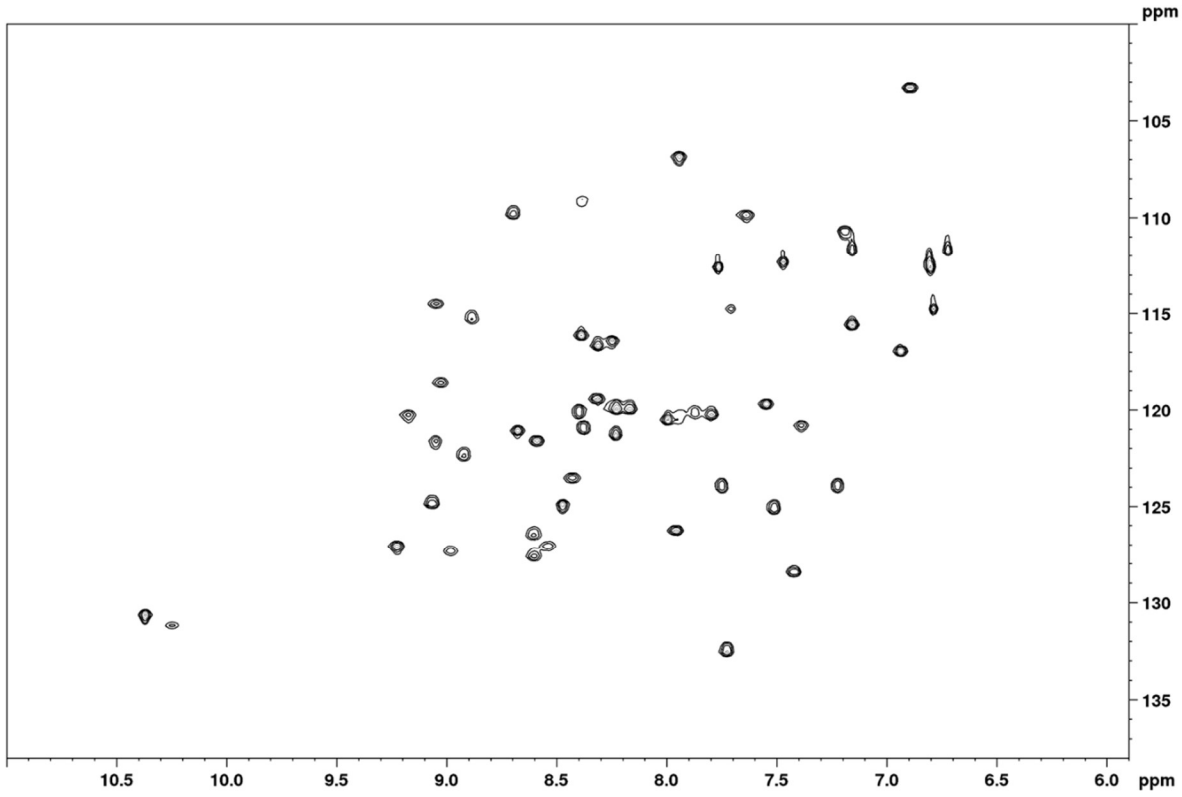


Figure 17. $^{1}H$-$^{15}N$ HSQC spectrum of GB1 A34SeM

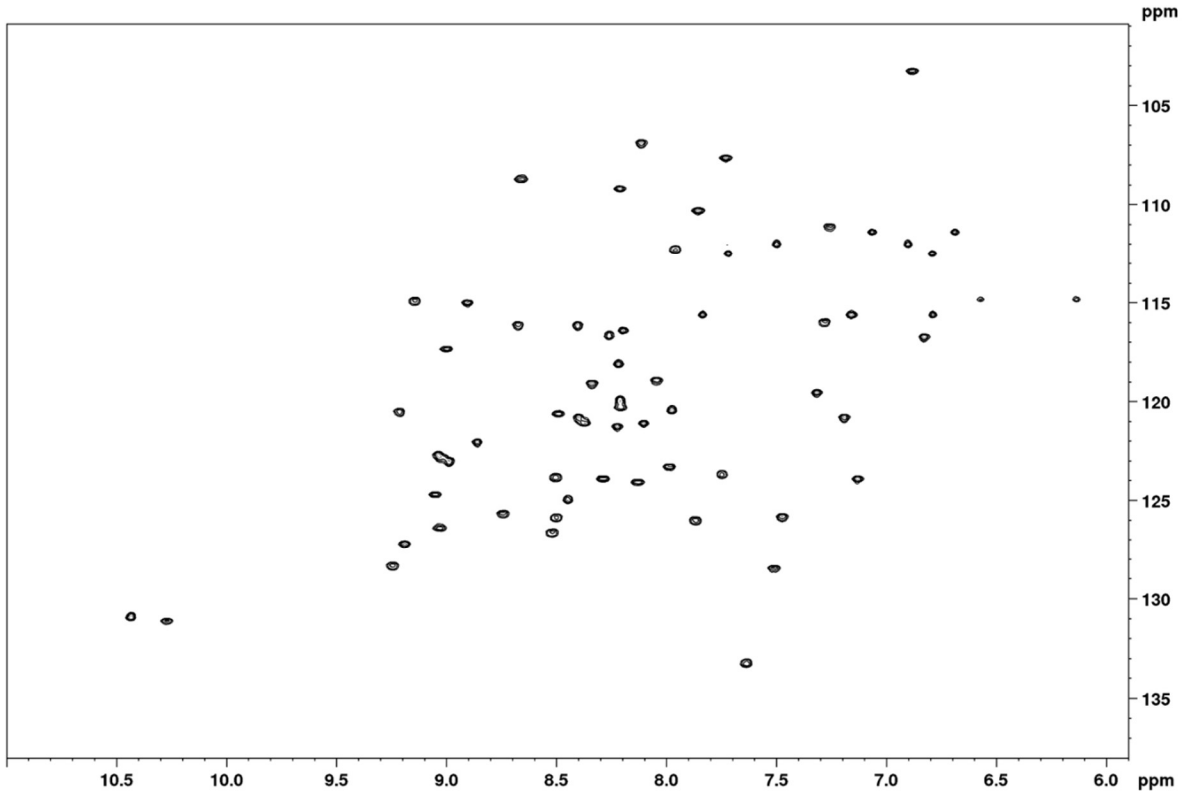


Figure 18. $^{1}H$-$^{15}N$ HSQC spectrum of GB1 V39SeM

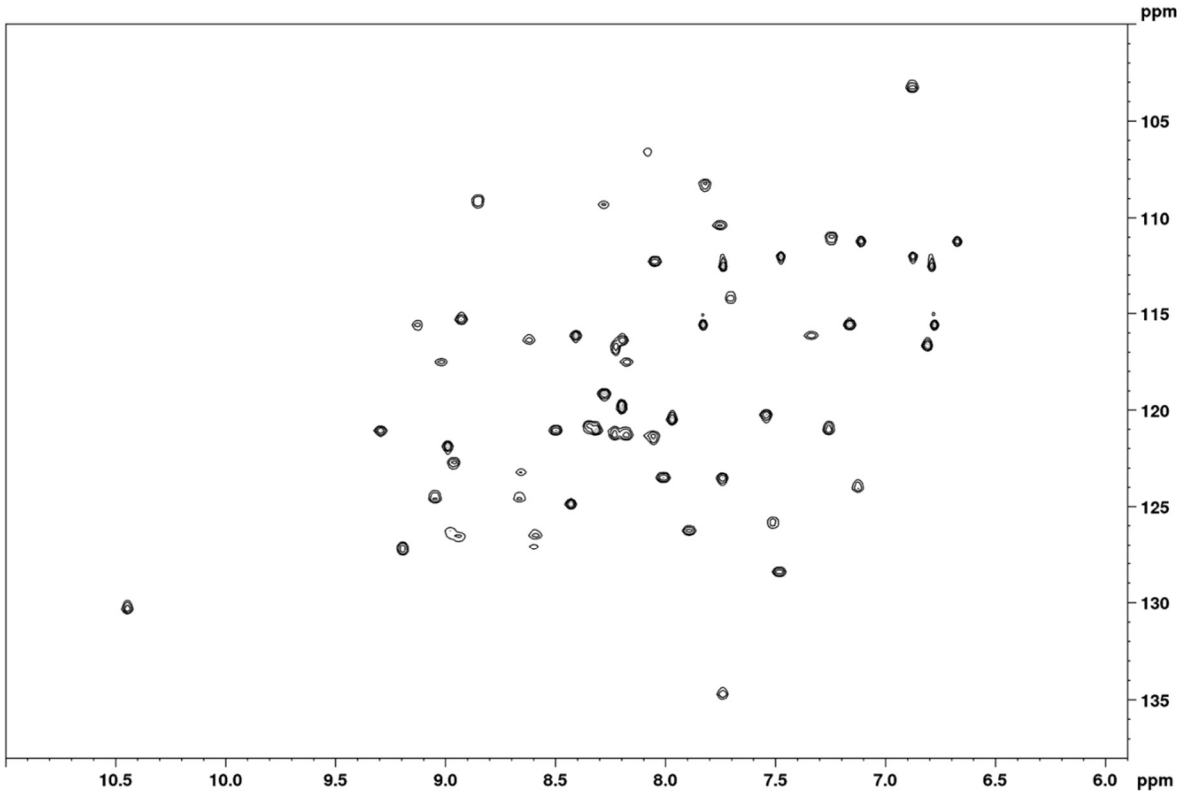


Figure 19. $^{1}H$-$^{15}N$ HSQC spectrum of GB1 V54SeM

## 3.5 Structural Perturbations in GB1 SeM Variants

For this particular study, the ability to assign residues of the WT GB1 based on published chemical shifts from a $^{1}H$-$^{15}N$ HSQC spectrum is of great importance. When superimposed with the $^{1}H$-$^{15}N$ HSQC spectrum of a GB1 SeM variant, it permits the identification of residues that experience a change in chemical shift, and therefore, a change in their local environment. This approach was particularly effective in this study, as the interpretation of data was straightforward based on deviations from the spectrum of the WT GB1.

When comparing the $^{1}H$-$^{15}N$ HSQC spectra of the six GB1 SeM variants to the WT GB1, the overall distribution of resonances remains largely similar; however, equipped with the resonance assignments provided by the WT GB1 spectrum, it becomes possible to identify specifically which residues within each variant have been perturbed due to SeM incorporation. This can be done by identifying which resonances in the variant spectra have altered chemical shifts relative to the WT spectrum. In order to visualize the proximity of perturbed residues to the site of mutagenesis, PDB structures of WT GB1 with annotation of the mutated residue and those that experience changes in chemical shift are provided for each variant [3]. The observed perturbations are summarized in the following sections.

### *3.5.1 L5SeM Perturbations*

The incorporation of SeM in the 5$^{th}$ position, occupied by a leucine in the WT GB1, caused significant disruption in chemical shifts in the $^{1}H$-$^{15}N$ HSQC spectrum of GB1 L5SeM. This is likely due to the fact that this residue is oriented into the core of the protein and the difference in size of the amino acid side chain. When superimposed with the spectrum of the WT GB1, a majority of the resonances in the $^{1}H$-$^{15}N$ HSQC spectrum of GB1 L5SeM deviate from the WT, with some of the significant residues highlighted in Figure 20. Furthermore, these perturbed residues, as well as leucine-5, are displayed in the structure of WT GB1 in Figure 21 (PDB: 2J52). These data show that the substitution

of a core-facing amino acid significantly perturbs tertiary structure of a protein, but importantly does not compromise the overall structure.

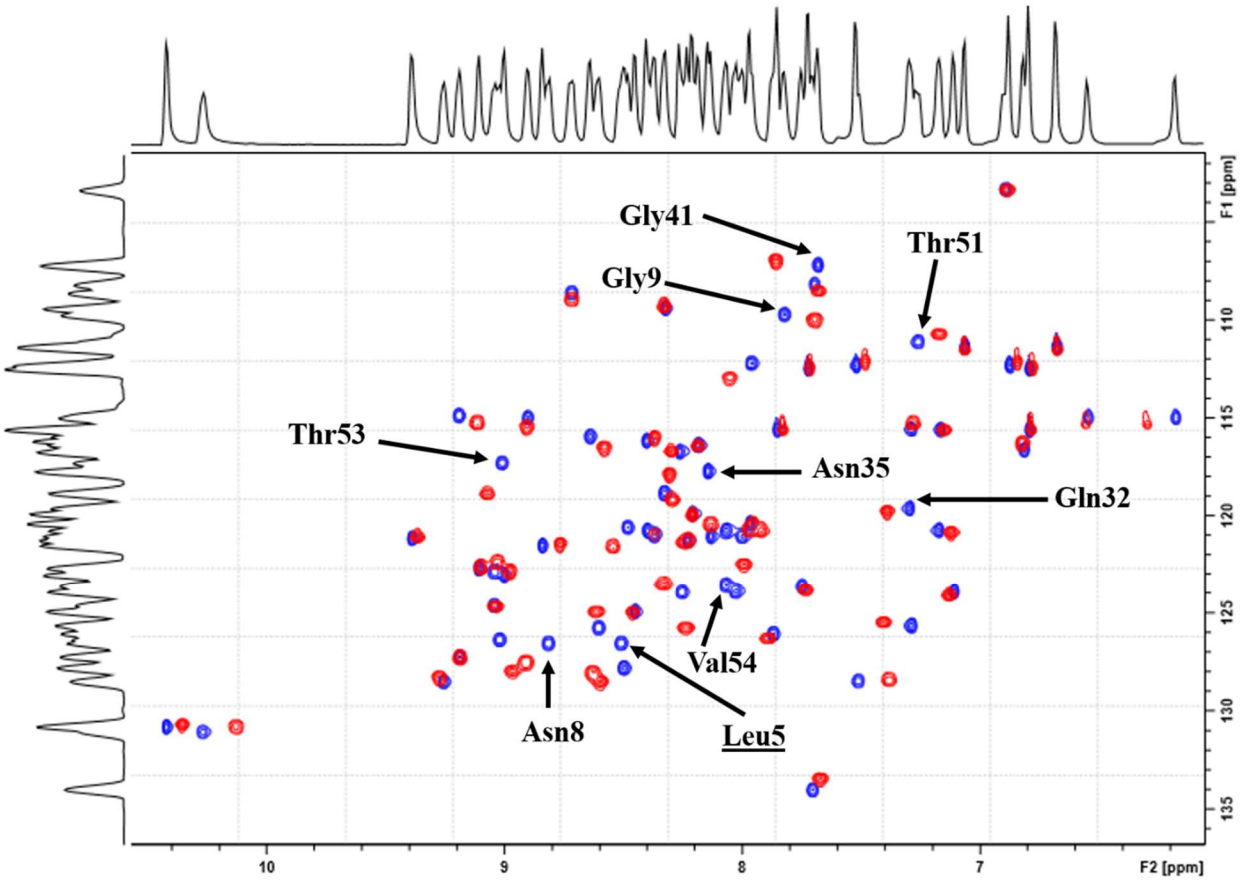


Figure 20: Overlay of $^{15}N$ HSQC spectra of WT (blue; PDB: 2J52) and L5Sem (red) GB1. The substituted residue (underlined) and those displaying significant deviations in chemical shift between spectra are noted.

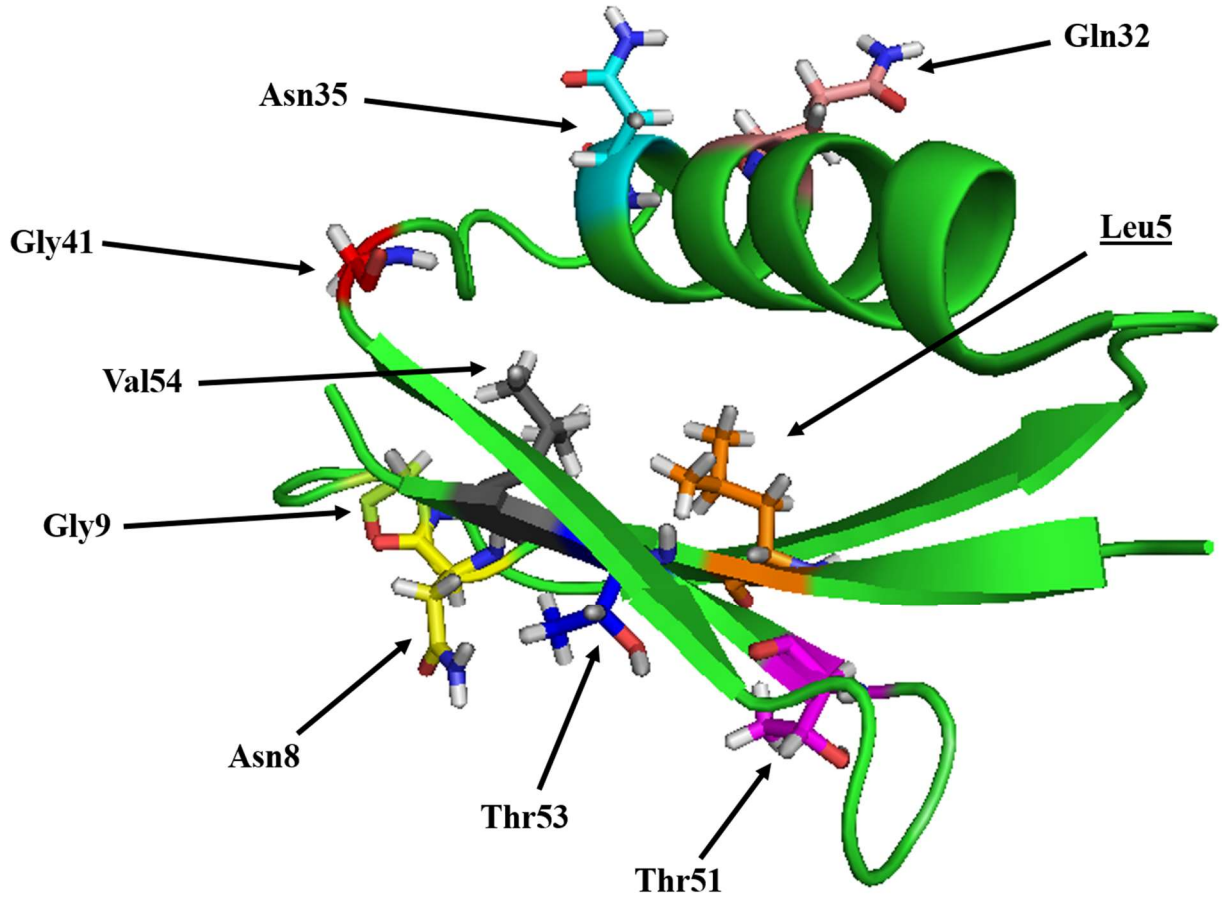


Figure 21: Physical proximity of leucine 5 to perturbed residues in WT GB1 (PDB: 2GB1).

### *3.5.2 I6SeM Perturbations*

The addition of SeM in place of an isoleucine residue in the 6$^{th}$ position was markedly different in comparison to GB1 L5SeM, as seen in Figure 22. In contrast, replacement of this solution-facing residue caused far less disruption in the $^{1}$H-$^{15}$N HSQC spectrum of GB1 I6SeM relative to WT GB1, which is to be expected for a residue that is surface exposed and does not interact with many neighboring residues; however,

significant perturbation is still observed (Figure 23). Upon comparison with the WT GB1 spectrum, the residues that experience changes in chemical shift are exclusively in close proximity with the SeM residue. The correlation is not exclusive to residues that neighbor in terms of amino acid sequence, but instead are seen in those which are in close proximity, regardless of sequence. In all, this data suggests that the substitution of a solution-facing amino acid has little effect on the overall structure of the protein, but does still affect the local environment of residues that are in close proximity.

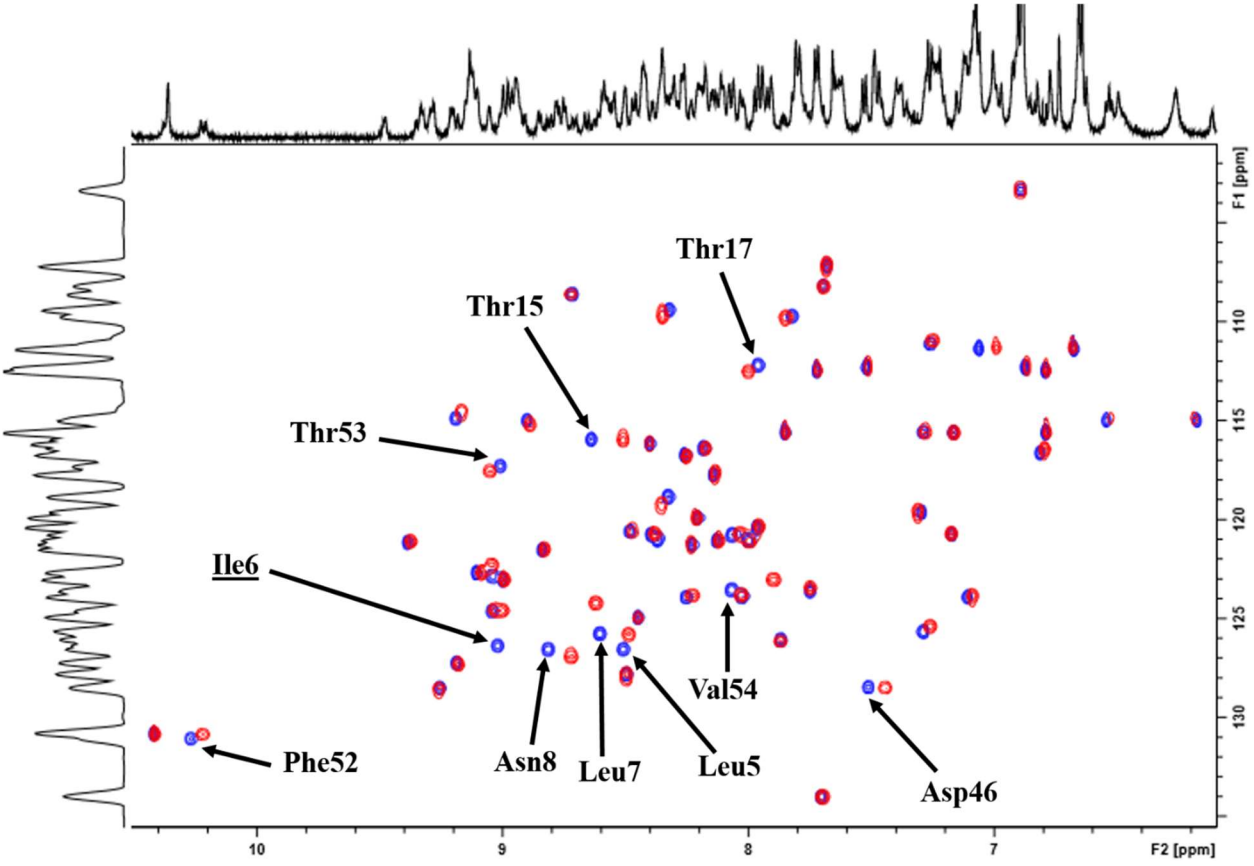

Figure 22: Overlay of $^{15}$N HSQC spectra of WT (blue; PDB: 2J52) and I6Sem (red) GB1. The substituted residue (underlined) and those displaying significant deviations in chemical shift between spectra are noted.

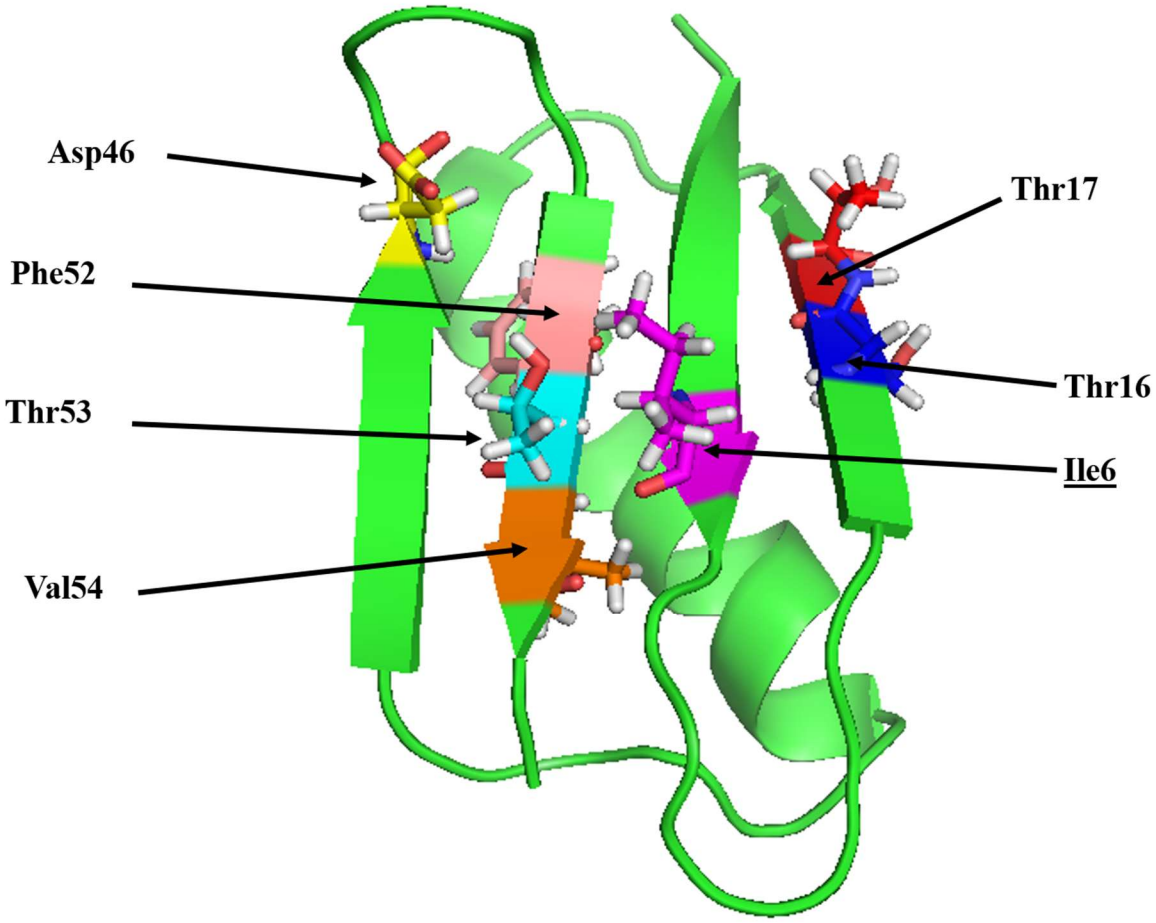


Figure 23: Physical proximity of isoleucine 6 to perturbed residues in WT GB1 (PDB:

### *3.5.3 V29SeM Perturbations*

Substitution of a SeM residue into the 29$^{th}$ position, occupied by a valine residue in WT GB1, is another case of substitution of a surface-exposed residue. As was

observed in the substitution of isoleucine 6 with a SeM residue, the overall perturbation to the chemical shifts in the $^{1}H$-$^{15}N$ HSQC spectrum of GB1 V29SeM relative to WT, seen in Figure 24, are minor compared to the spectrum of a core-facing substitution such as GB1 L5SeM. Similar to GB1 I6SeM, only the residues in close physical proximity are perturbed, and in general, the magnitude of perturbations observed in the $^{1}H$-$^{15}N$ HSQC spectrum are minor (Figure 25).

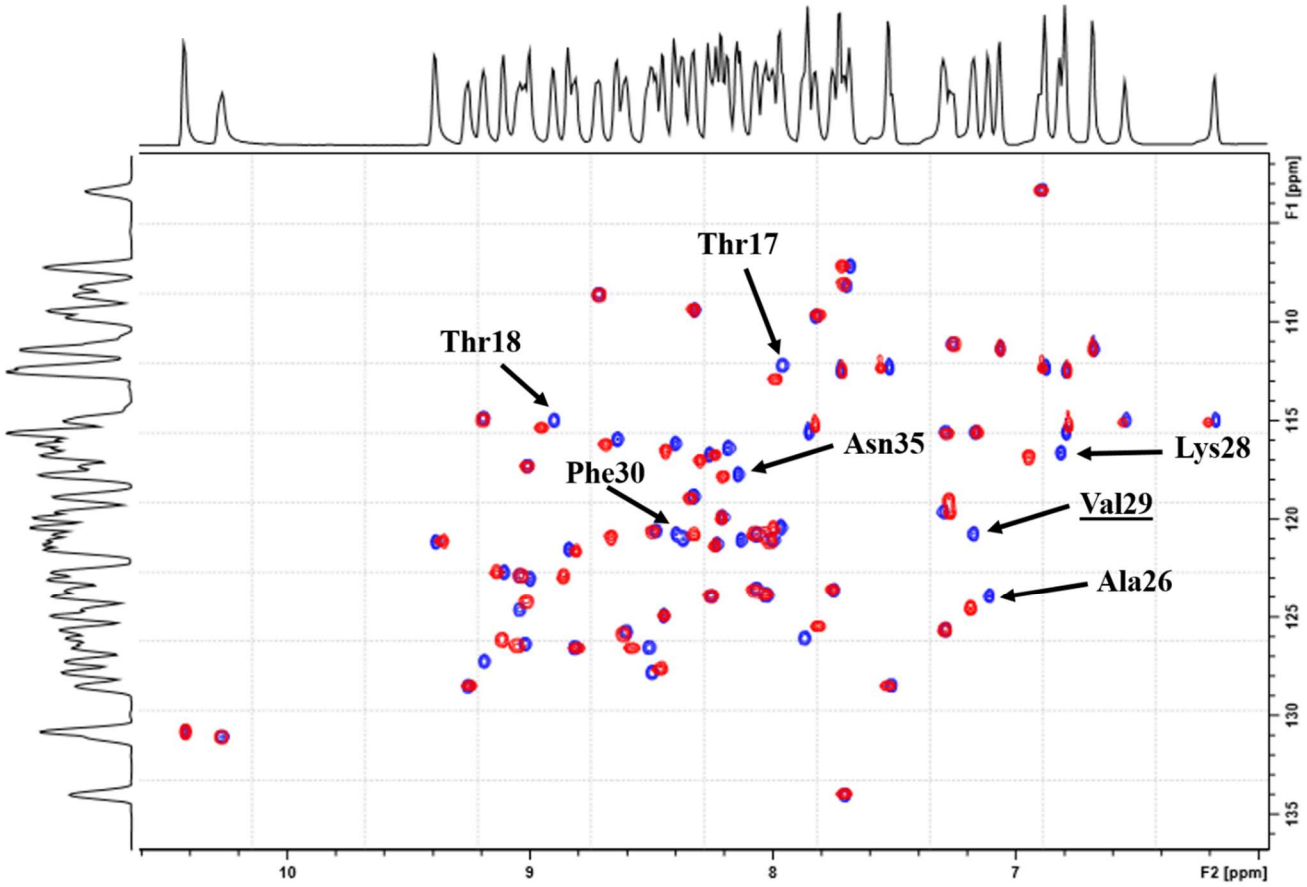


Figure 24: Overlay of $^{15}N$ HSQC spectra of WT (blue; PDB: 2J52) and V29Sem (red) GB1. The substituted residue (underlined) and those displaying significant deviations in chemical shift between spectra are noted.

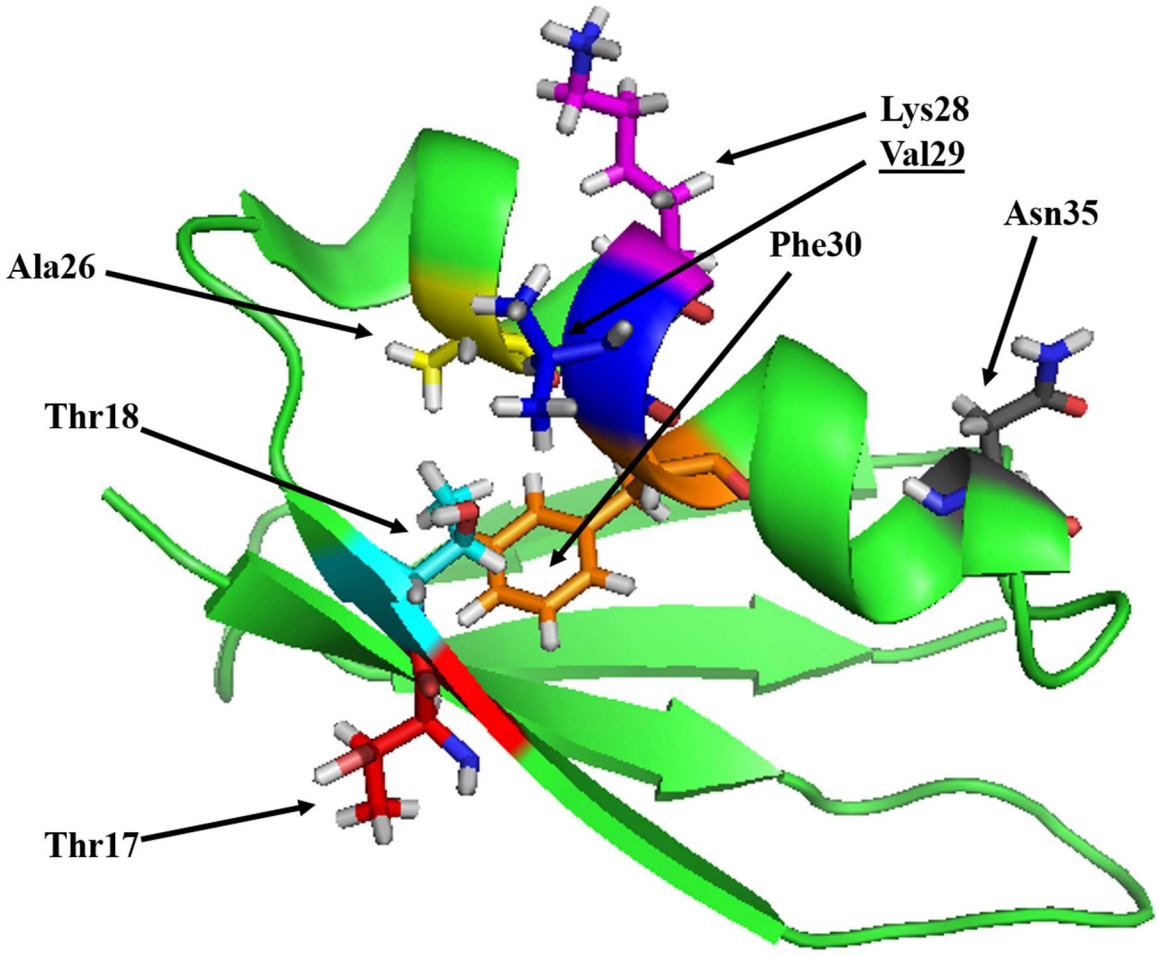


Figure 25: Physical proximity of valine 29 to perturbed residues in WT GB1 (PDB: 2GB1).

### *3.5.4 A34SeM Perturbations*

The substitution of alanine 34 for a SeM residue, in terms of position within the tertiary structure and size difference between the amino acid side chains, is the most significant deviation from WT GB1 that is seen in the six GB1 SeM variants examined in this study. Not only is this a substitution of a core-facing residue, it also introduces a significantly larger side chain. In contrast to the simple methyl group side chain of

alanine, the side chain of SeM contains three carbons and a large selenium atom. This is thought to introduce significant steric hindrance within the core of the protein.

Comparison of the $^{1}H$-$^{15}N$ HSQC spectrum of GB1 A34SeM to the WT GB1 shows a systematic disruption to the chemical shift of nearly every residue within the protein, as can be seen in Figure 26. This suggests an overall change in structure; however, the overall tertiary structure remains intact. The structural change introduced by SeM incorporation at the 34$^{th}$ position in GB1 can be summarized as minor increase in the size of the protein, affecting nearly all residues to a measurable degree. Some of the residues that are significantly affected by the substitution can be seen in Figure 27.

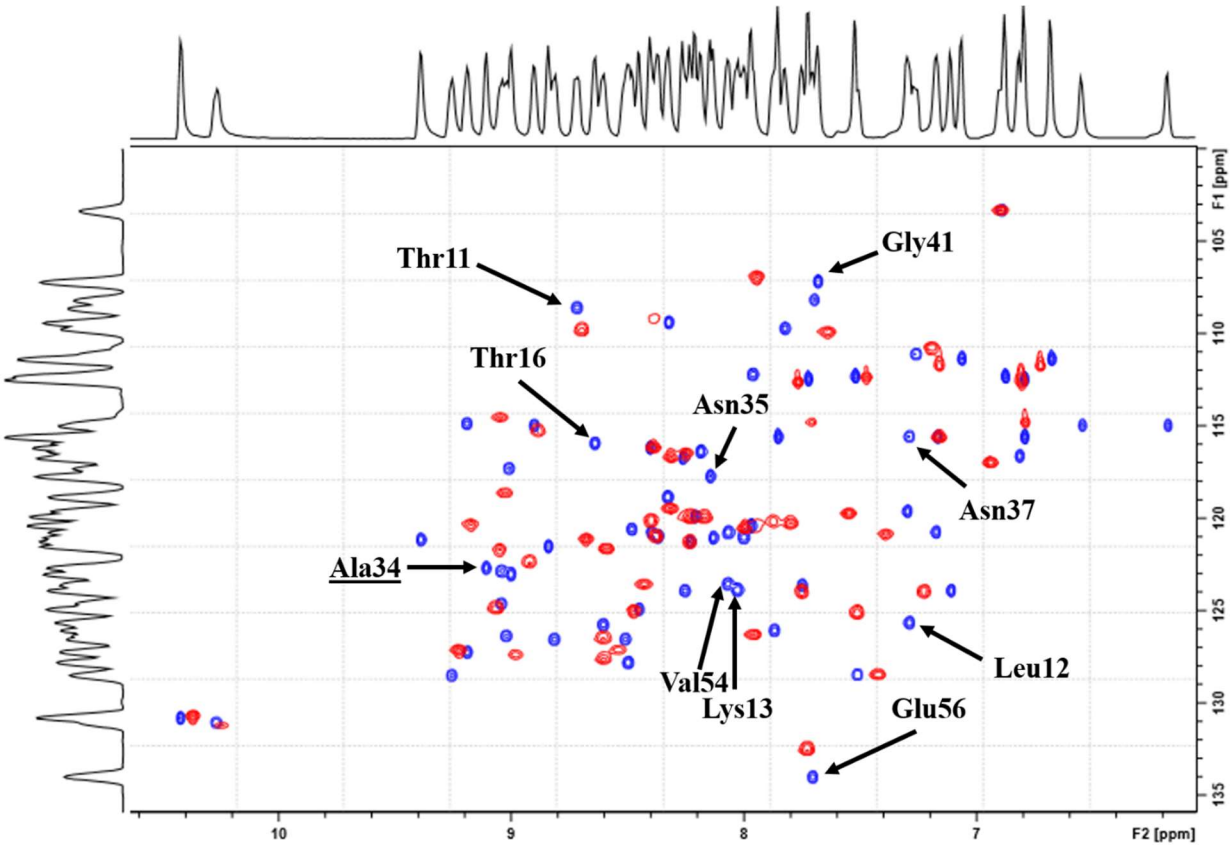

Figure 26: Overlay of $^{15}$N HSQC spectra of WT (blue; PDB: 2J52) and A34Sem (red) GB1. The substituted residue (underlined) and those displaying significant deviations in chemical shift between spectra are noted.

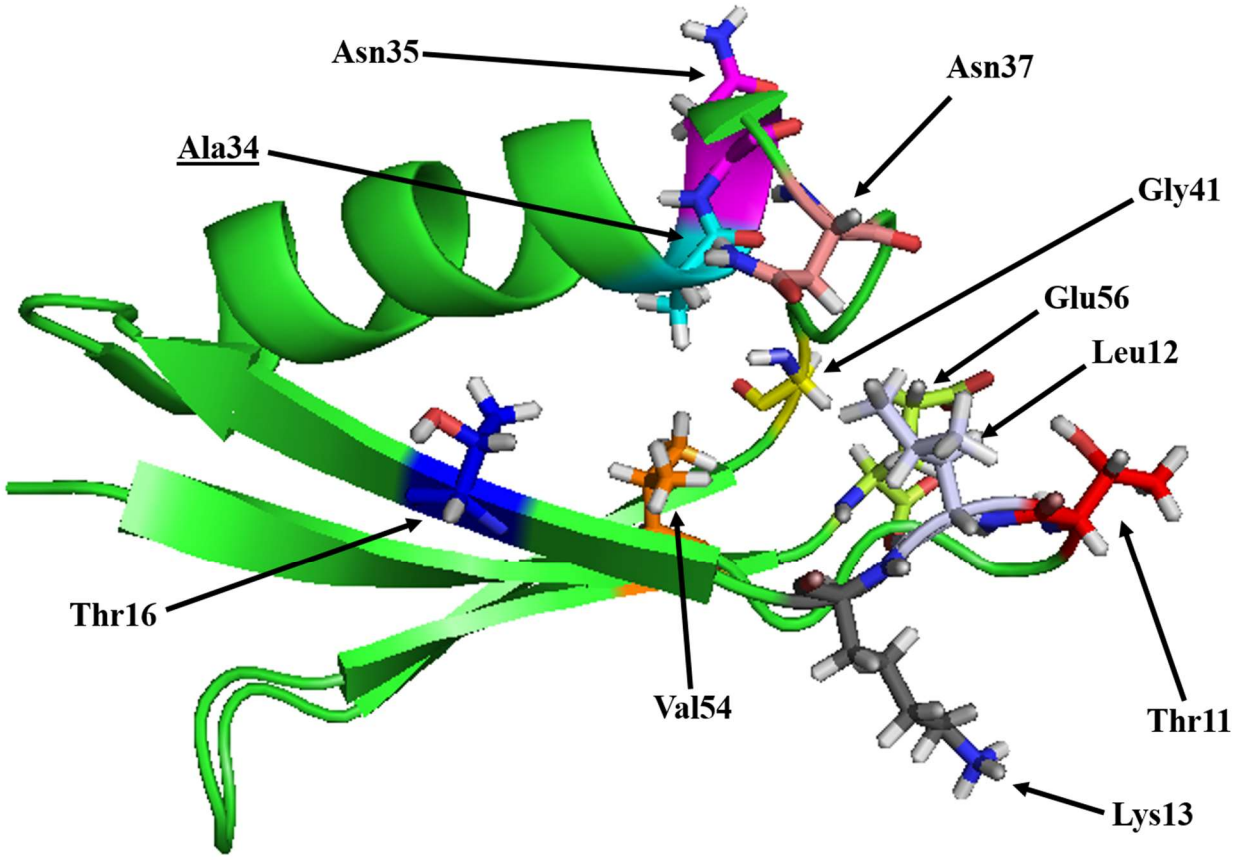


Figure 27: Physical proximity of alanine 34 to perturbed residues in WT GB1 (PDB: 2GB1).

### *3.5.5 V39SeM Perturbations*

Substitution of SeM at the 39$^{th}$ position, occupied by a valine residue in WT GB1, introduces a change to a residue located on a flexible loop section of the protein's secondary structure, oriented somewhat into the core of the protein from a more

peripheral position on the outside. The introduction of SeM at this position causes unique changes in the $^{1}H$-$^{15}N$ HSQC spectrum of GB1 V39SeM relative to WT GB1 in that there are not a large number of perturbation; however, in the residues that are affected, the magnitude of change is relatively significant (Figure 28). At a glance, there appears to be a dichotomy of behaviors, with the chemical shifts either being unchanged, or if they are changed, the magnitude is significant compared to the $^{1}H$-$^{15}N$ HSQC spectra of other variants. This is further elucidated when visualizing the position of the substitution relative to the residues that have been perturbed, as is observed in Figure 29 and Figure 30. These figures highlight the observation that changes in chemical shift following an amino acid substitution are highly localized and have little to no effect on residues that are spatially distant.

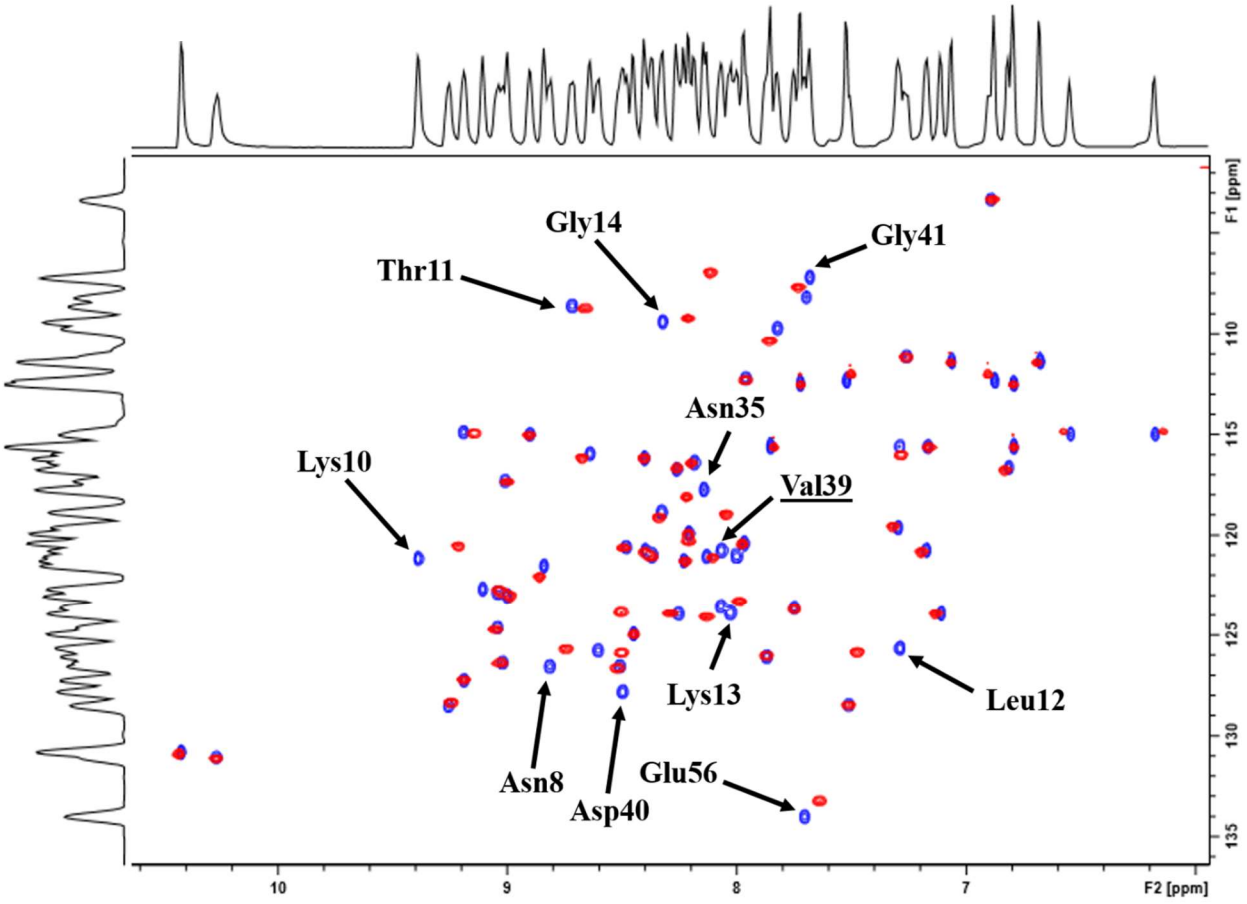


Figure 28: Overlay of $^{15}$N HSQC spectra of WT (blue; PDB: 2J52) and V39Sem (red) GB1. The substituted residue (underlined) and those displaying significant deviations in chemical shift between spectra are noted.

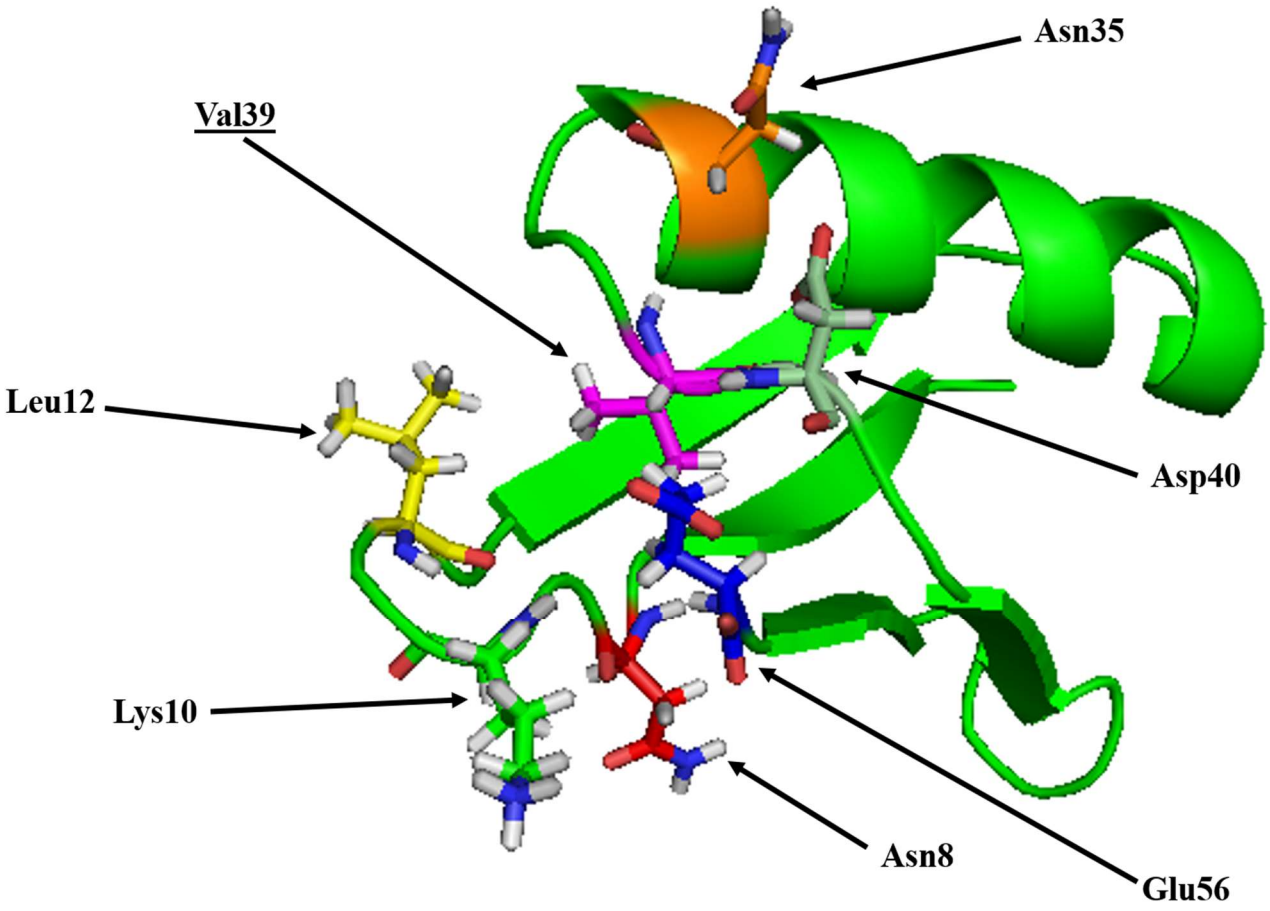


Figure 29: Physical proximity of valine 39 to perturbed residues in WT GB1 (PDB: 2GB1).

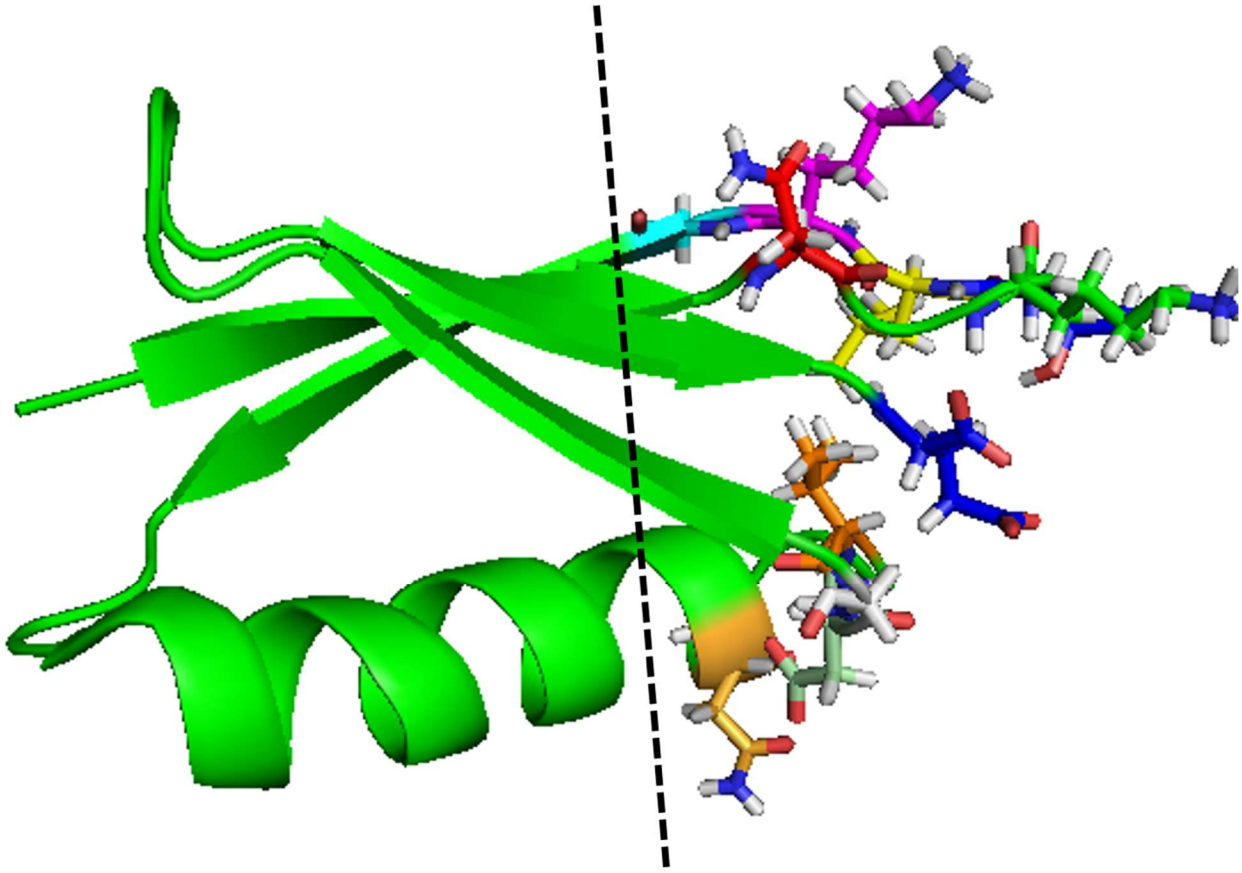

Figure 30: Alternate view of the physical proximity of valine 39 to perturbed residues in WT GB1 (PDB: 2GB1). The dashed line is added to emphasize the proximity of the substituted residue to those which display significant perturbation in chemical shift in the HSQC

### *3.5.6 V54SeM Perturbations*

The introduction of SeM at the 54$^{th}$ position, which is occupied by a valine residue in WT GB1, represents another change to a core-facing residue. As is observed in other core-facing GB1 SeM variants, perturbations to chemical shifts in the $^{1}$H-$^{15}$N HSQC spectrum of GB1 V54SeM are systematic, affecting the majority of residues in the protein (Figure 31). This follows the trend that is observed in other similar variants, and is to be expected due to the centralized proximity of the residue within the protein's tertiary structure (Figure 32)

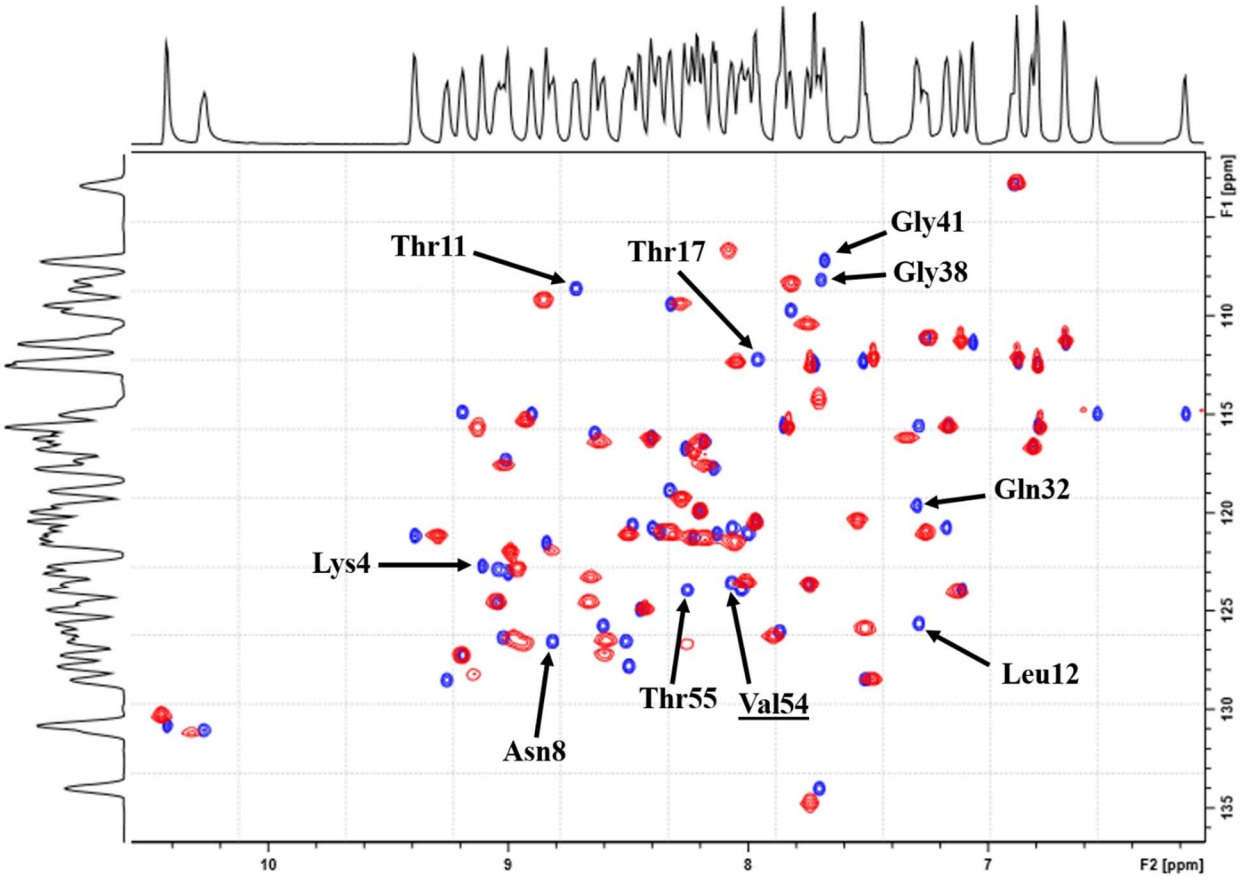


Figure 31: Overlay of $^{15}$N HSQC spectra of WT (blue; PDB: 2J52) and V54Sem (red) GB1. The substituted residue (underlined) and those displaying significant deviations in chemical shift between spectra are noted.

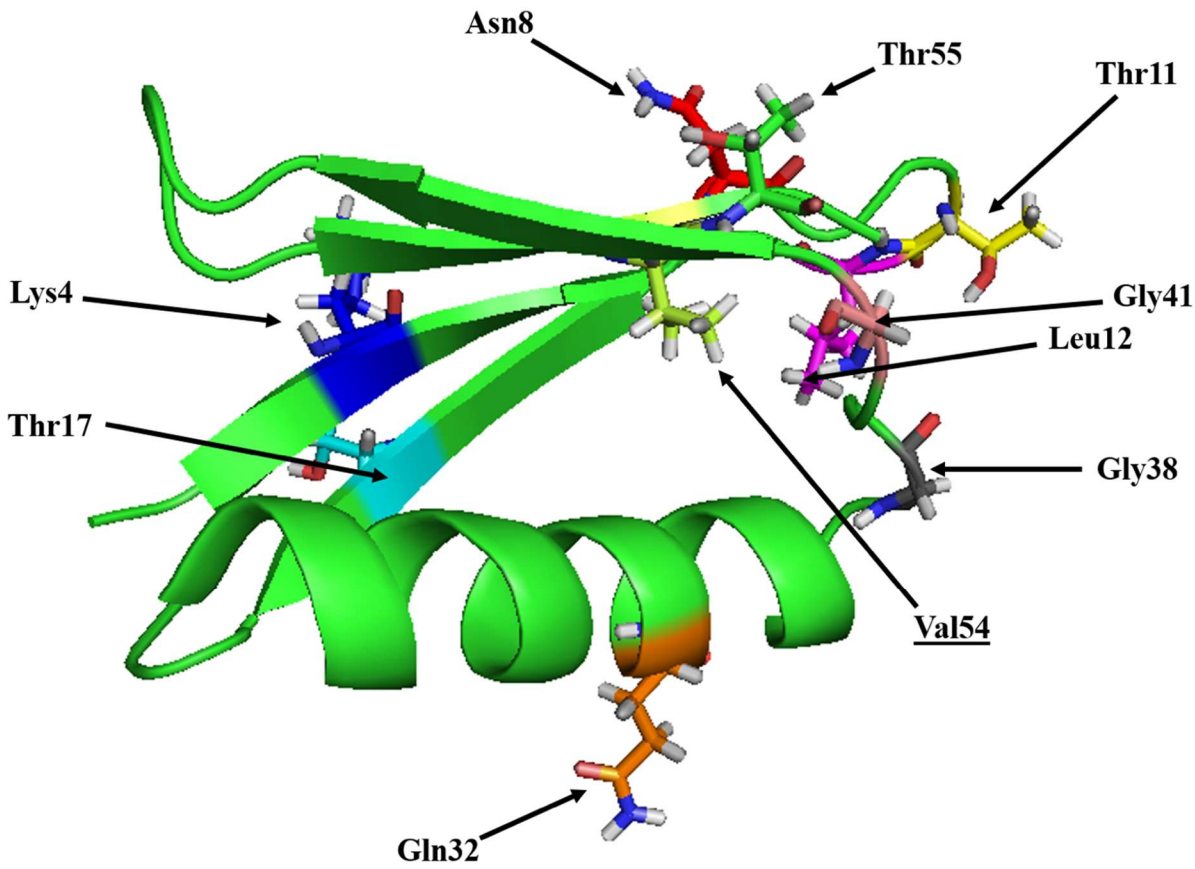


Figure 32: Physical proximity of valine 54 to perturbed residues in WT GB1 (PDB: 2GB1).

### 3.6 $^{1}H$-$^{77}Se$ HMBC Spectra of GB1 SeM Variants

The final 2D NMR experiment in this study was the acquisition of $^{1}H$-$^{77}Se$ HMBC spectra of the six GB1 SeM variants. The $^{1}H$-$^{77}Se$ HMBC spectra of other SeM-containing proteins in solution have been reported, including a study by Zhang et al. that observed the changes in chemical shift of the selenium atoms in calmodulin between the *apo* and $Ca^{2+}$-bound forms in solution [4]. Therefore, by providing yet another data set of 2D $^{77}Se$ NMR, the volume of reference information is further expanded by providing supplemental information on such studies.

Quality $^{1}H$-$^{77}Se$ HMBC spectra were acquired for all GB1 SeM variants except for GB1 A34SeM, likely due to reasons described in 2.6, and can be seen in Figures 33-38. Each spectrum (except A34SeM) shows a single, well-defined resonance that corresponds to the single selenium atom of the SeM side chain within the protein. The distribution of both $^{1}H$ and $^{77}Se$ chemical shifts fall within a range comparable to previous studies using $^{1}H$-$^{77}Se$ HMBC in solution-state proteins, with $^{1}H$ chemical shifts ranging from 1.38 – 1.88 ppm and $^{77}Se$ chemical shifts ranging from 53.97 – 119.23 ppm. The large range of chemical shift distribution in the $^{77}Se$ dimension further highlights the sensitivity of this particular nucleus, as well as its benefit as a surrogate for sulfur, thus enabling its use as a probe within biological molecules.

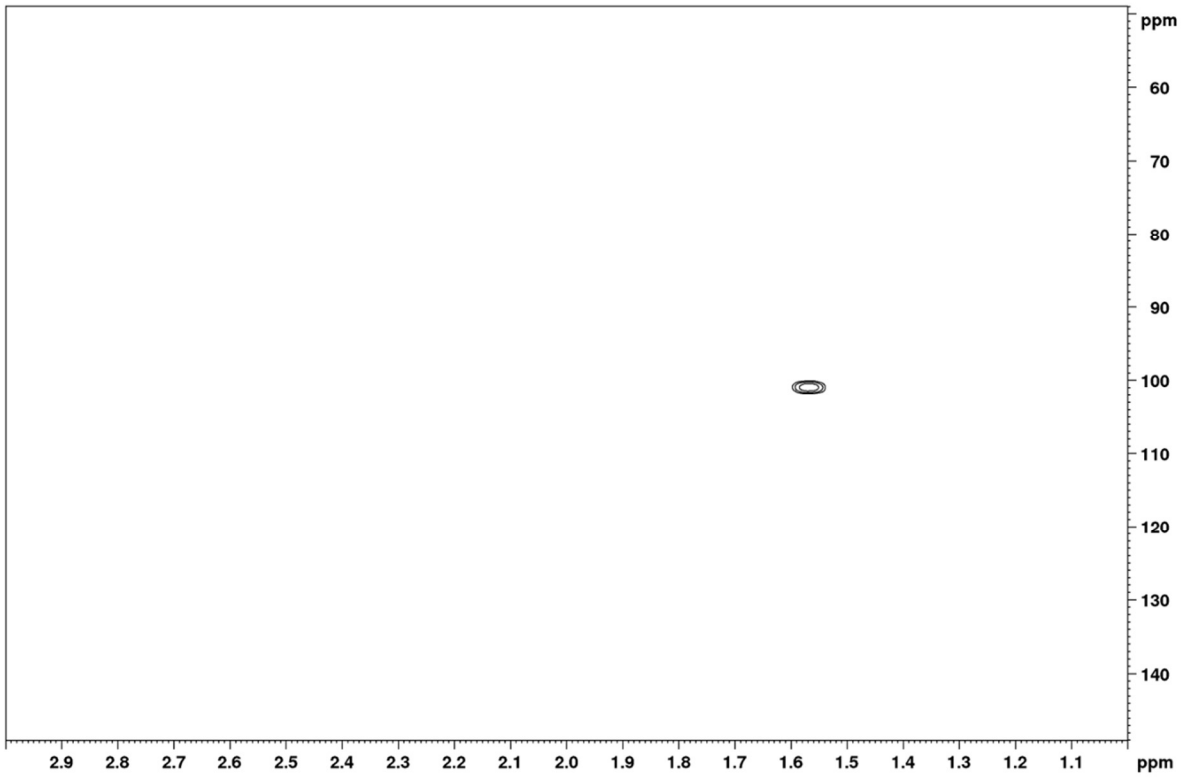


Figure 33: $^{1}H$-$^{77}Se$ HMBC NMR spectrum of GB1 L5Sem.

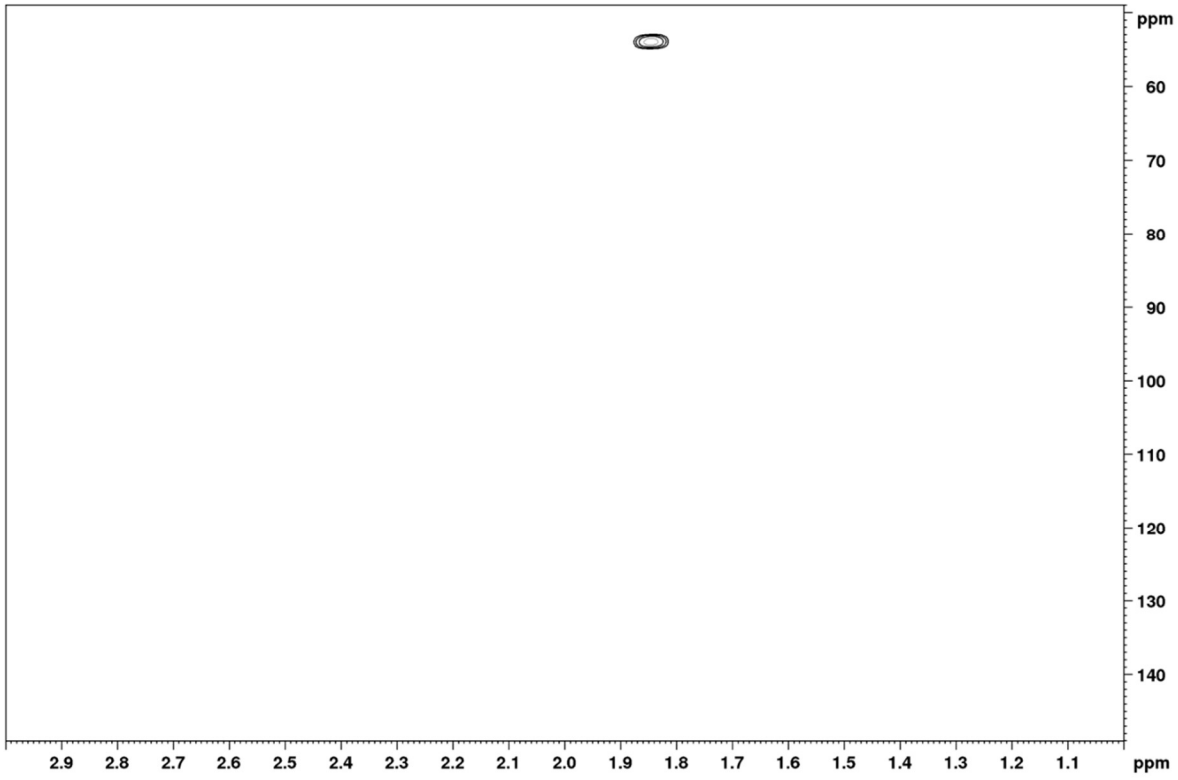


Figure 34: $^{1}H$-$^{77}Se$ HMBC NMR spectrum of GB1 I6Sem.

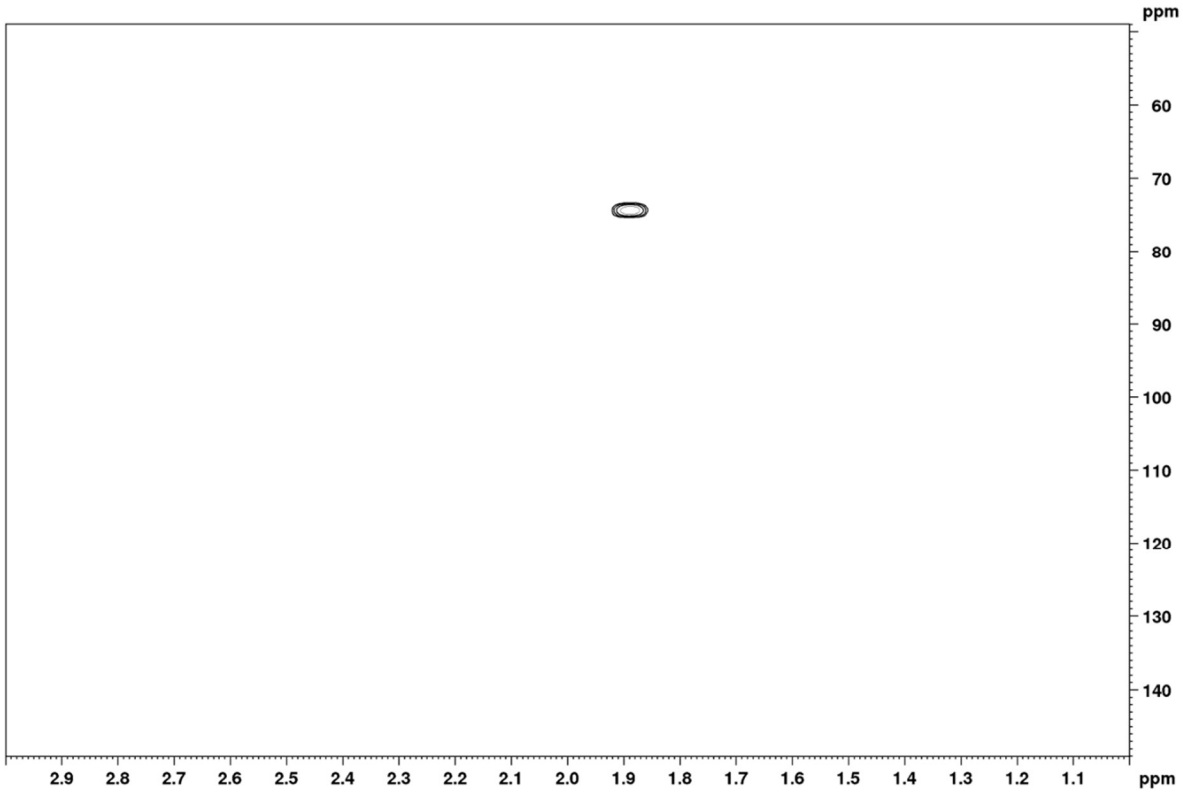


Figure 35: $^{1}H$-$^{77}Se$ HMBC NMR spectrum of GB1 V29Sem.

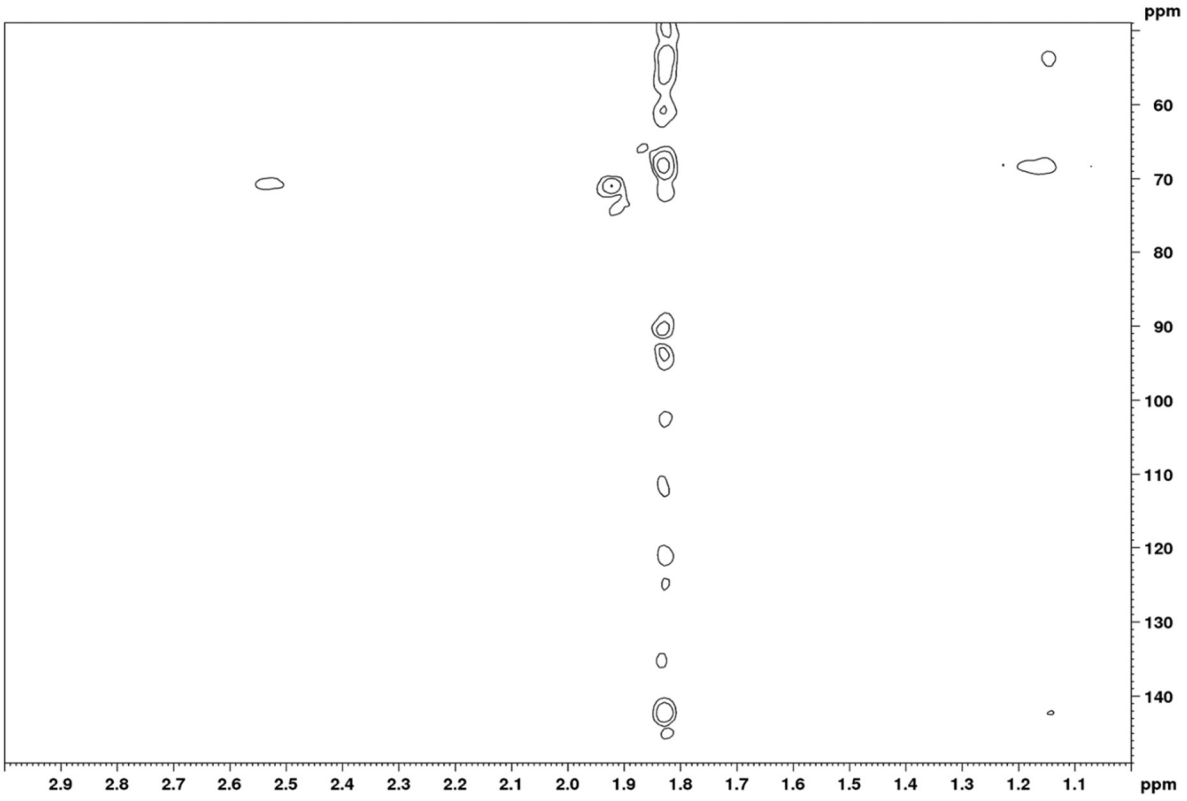


Figure 36: $^{1}H$-$^{77}Se$ HMBC NMR spectrum of GB1 A34Sem.

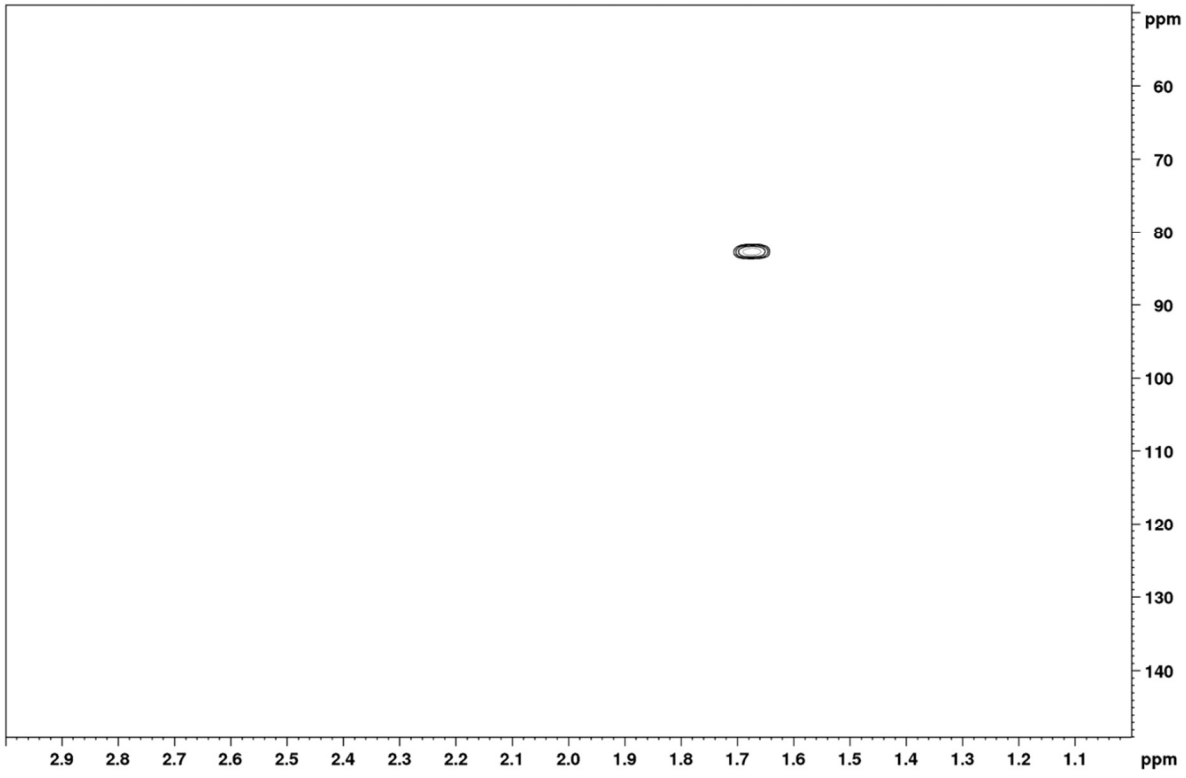


Figure 37: $^{1}H$-$^{77}Se$ HMBC NMR spectrum of GB1 V39Sem.

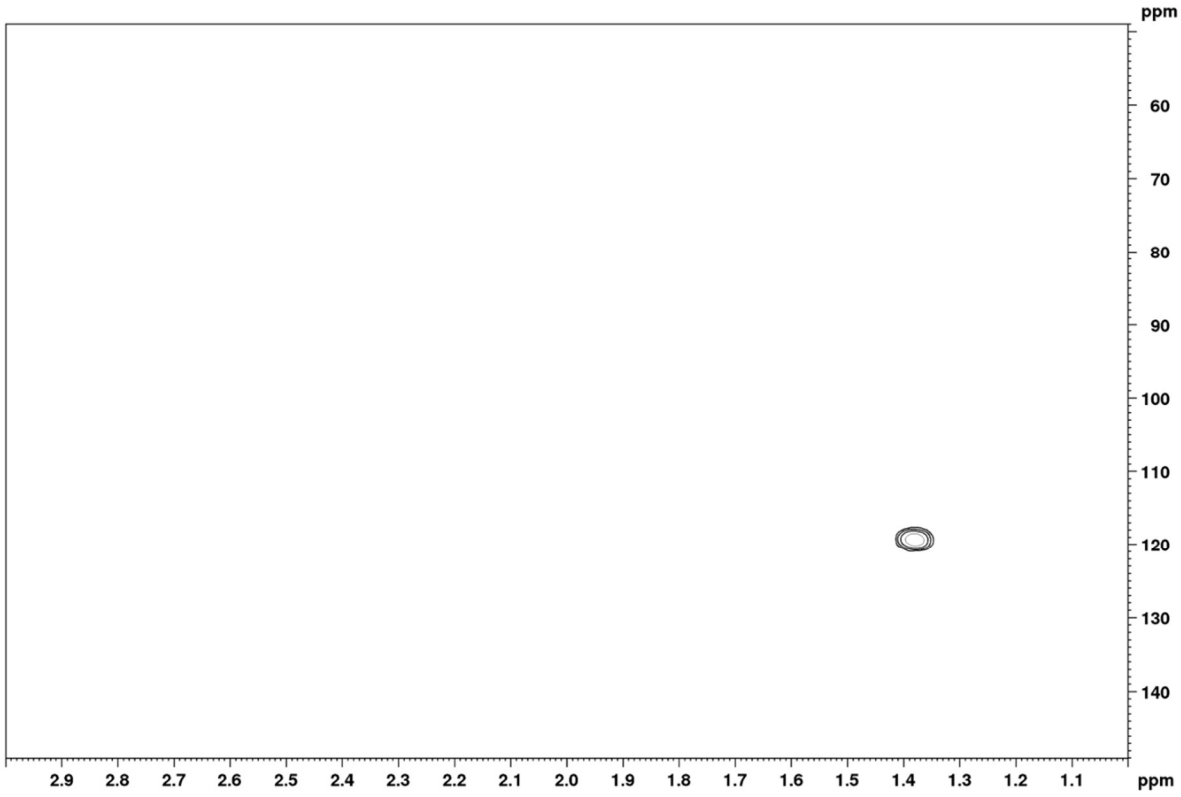


Figure 38: $^{1}H$-$^{77}Se$ HMBC NMR spectrum of GB1 V54Sem.

## REFERENCES


1. Wilton, D., Tunnicliffe, R., Kamatari, Y., Akasaka, K., Williamson, M. Pressure-induced changes in the solution structure of the GB1 domain of protein G. (2008). Proteins. 71 (3), 1432-1440.
2. Tunnicliffe, R. & Williamson, M. $^{1}H$, $^{13}C$, and $^{15}N$ chemical shift assignments for non-aromatic parts of GB1 domain. (2008). BMRB Entry: 7280
3. Gronenborn, A., Filpula, D., Essig, N., Achari, A., Whitlow, M., Wingfield, P., Clore, G. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. (1991) Science. 253 (5020), 657-661. PMID:
4. Zhang, M., Vogel, H. 2-dimensional NMR studies of selenomethionyl calmodulin. (1994). Journal of Molecular Biology. 239 (4), 545-554.

# Chapter 4

# CONCLUSIONS

## 4.1 Conclusions and Future Work

This 1D and 2D solution $^{77}$Se NMR spectroscopy survey of the six GB1 SeM variants was a part of a larger, systematic evaluation of the behavior of selenium in proteins, as examined by $^{77}$Se NMR spectroscopy, within different chemical environments inside a biological macromolecule. Another aspect of the larger work was the acquisition of solid-state NMR (ssNMR) spectra, which allowed the measurement of chemical shift anisotropy (CSA) and subsequent calculation of the chemical shift tensors of the six GB1 SeM variants. Furthermore, the broader study was augmented by the acquisition of high-resolution structures of the GB1 SeM variants through the use of x-ray crystallography. Structures of five of the six GB1 SeM variants were determined with resolution that met the requirements for deposition into the Protein Data Bank (PDB). The synergy created by the acquisition of these diverse data sets yielded a novel, previously-unestablished understanding of the behavior of the $^{77}$Se atom in a biological context. The specific goal of the study described herein was to establish a foundational understanding of how the selenium atom responds to various local environments and create a point of reference for further studies that explore solution $^{77}$Se NMR spectroscopy of selenium in biological systems.

The results of the 1D solution NMR of the six GB1 SeM variants showed that the chemical shift of $^{77}$Se is very sensitive, with selenium atoms incorporated into the same SeM framework and the same protein structure exhibiting a range of chemical shifts spanning nearly 70 ppm. These observations were further probed by examining the influence of solution temperature on both chemical shift and line width of the signals measured in each of the six GB1 SeM variants. In general, lowering the temperature of the solution will result in a broadening of the line, as well as a decrease in the $^{77}$Se chemical shift. The lone exception to this observation is seen in GB1 V54SeM, in which the chemical shift increased as the temperature of the measurement solution decreased, a dynamic that may be explained by aromatic stacking in the protein core as a result of temperature change. Aside from direct measurables such as temperature and line width, the influence of local environment on the chemical shift and its dynamism were observed as a function of location within the protein that the selenium atom occupied. As the hydrophobicity, in this case the proximity to the core of the protein, increased, so too did the chemical shift value. Conversely, as hydrophobicity of the selenium atom's environment increased, the change in chemical shift vs. temperature decreased, suggesting a reduced conformational mobility that is imparted to those residues facing the solvent.

The results of the 2D solution NMR experiments on the six GB1 SeM variants provided additional information on the macroscale properties and dynamics brought about as a result of SeM incorporation. This was aided substantially by the acquisition of the $^{1}$H-$^{15}$N HSQC spectrum of WT GB1, which serves as a point of reference for comparison with the spectra of the six GB1 SeM variants. Furthermore, the ability to

assign specific amino acid residues to the resonances within the WT spectrum allowed precise identification of the perturbed residues in the GB1 SeM variants.

The $^{1}H$-$^{13}C$ HSQC spectra of all six GB1 SeM variants were acquired and displayed the characteristic distribution of resonances that is seen in a well-folded, intact protein. While there were local variations in the resonances, which are expected in a protein sample containing substitutions in different positions, the overall pattern was similar. The acquisition of these spectra, for the purposes of this study, confirmed sample integrity and informed about overall protein dynamics. This once again highlights the need to establish a library of data that incorporates $^{77}Se$, either directly, or in the case of these $^{1}H$-$^{13}C$ HSQC spectra, indirectly.

Acquisition of the $^{1}H$-$^{15}N$ HSQC spectra of both the WT GB1 and the SeM variants provided a substantial amount of data that sheds light on the nature of how SeM substitution affects the macroscale structure of a model protein. Upon superimposition of variant spectra with that of WT GB1, individual residues whose identity are known from published values can be identified to have varied from the WT. On the whole, substitution of core-facing residues had the largest disruptive effect to the global distribution of chemical shifts, while solvent exposed residues were significantly less disruptive. In general, the governing factor that influenced perturbation was direct proximity to the local environment of the substitution. This result is to be expected, but it is beneficial to once again establish component datasets that can contribute to a larger library of data that will enable the expansion of the use of selenium NMR in the study of biological macromolecules.

Finally, $^{1}H$-$^{77}Se$ HMBC spectra of all six GB1 SeM variants were acquired, five of which provided quality data ($^{1}H$-$^{77}Se$ HMBC experiments with A34SeM did not produce a useable spectrum). The five spectra for which meaningful data was acquired each show a single, high signal-to-noise ratio (SNR) peak. While the span of $^{77}Se$ chemical shifts were acquired through the 1D solution experiments, this technique provides the same information in the $^{1}H$ dimension, further expanding the information that is available to those who in the future may wish to utilize these same techniques in their research. Similarly to the $^{1}H$-$^{13}C$ HSQC spectra of the GB1 SeM variants, these HMBC spectra serve more as an affirmation of the applicability of the technique, opening the door to further research studies. It is quite likely that this technique will further be used to monitor protein dynamics and binding, as we seen in the work by Zhang et al.

The data and results of this study serve several purposes, including an expansion of the understanding of $^{77}Se$ NMR and the different ways it can be applied. As a surrogate to the NMR-insensitive sulfur atom, selenium unlocks the application of the powerful tool of NMR spectroscopy to sites that normally include sulfur. Moreover, the wide scope of this study included a matrix of six GB1 SeM variants, examined 4 separate atomic nuclei via both 1D and 2D NMR experiments, and examined several important experimental variables such as temperature and local chemical environment. When taken together, the data and results from the solution NMR, ssNMR, and structure acquisition build a robust foundation for future work that seeks to utilize similar experimental approaches to study biological systems.

## REFERENCES

1. Chen, Q., Xu, S., Lu, X., Boeri, M., Pepelyayeva, Y., Diaz, E., Soni, SD., Allaire, M., Forstner, M., Bahnson, B., Rozovsky, S. $^{77}$Se probes the protein environment of selenomethionine. (2020). Journal of Physical Chemistry B. 124, 601-616.