Dissecting the Mechanism of ETV6 Polymerization

by

Helen Huang-Hobbs

B.Sc., McGill University 2009

A Thesis Submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

The Faculty of Graduate and Postdoctoral Studies

(Chemistry)

The University of British Columbia

(Vancouver)

December 2013

© Helen Huang-Hobbs, 2013

Abstract

ETV6 (or TEL), a member of the ETS family of eukaryotic transcription factors, normally functions as a transcriptional repressor and putative tumor suppressor. ETV6 is modular, containing a SAM (or PNT) domain and a DNA-binding ETS domain joined by a flexible linker sequence. The ETV6 SAM domain self-associates in a head-to-tail fashion, forming helical polymers proposed to generate extended repressive complexes at target DNA sites. ETV6 is also frequently involved in chromosomal translocations yielding unregulated chimeric oncoproteins with the SAM domain fused to the catalytic domain of a tyrosine receptor kinase such as NTRK3. Cellular transformation likely results from SAM domain-mediated polymerization and constitutive activation of the kinase domain. In the case of the ETV6-NTRK3 fusion (EN), this transformation is linked to congenital fibrosarcomas. Our goal is to investigate via mutations within its SAM domain, the thermodynamic and dynamic mechanisms underlying the altered transformation properties of ETV6-NTRK3. These studies have been carried out using monomeric variants of the isolated SAM domains with "head" or "tail" point mutations that prevent self-association, yet allow for formation of a mixed dimer with a native binding interface. Specifically, we used a combination of NMR spectroscopy and isothermal titration calorimetry to study the effects of additional mutations on their dimerization. Consistent with its involvement in a crystallographically-observed interdomain salt bridge, mutation of Lys99 was found to weaken the association of ETV-SAM monomers in solution, and to disrupt cellular transformation by EN. This supports the role of the SAM domain self-association in the activation of ETV6-NTRK3, and helps define the mechanisms underlying cellular transformation by similar chimeric oncoproteins.

Preface

Chapter 1 and 4. Versions of these data (Table 2, Figure 6, Figure 22, Figure 23, Figure 24) have been published within the manuscript: Naniye Cetinbas, Helen Huang-Hobbs, Cristina Tognon, Gabriel Leprivier, Jianghong An‖, Steven McKinney, Mary Bowden, Connie Chow, Martin Gleave, Lawrence P. McIntosh and Poul H. Sorensen. Mutation of the Salt Bridge-forming Residues in the ETV6-SAM Domain Interface Blocks ETV6-NTRK3-induced Cellular Transformation. *The Journal of Biological Chemistry*. 2013; 288:27940-27950. © the American Society for Biochemistry and Molecular Biology.

Mass Spectroscopy (Figure 13, Figure 26) was run by Jason Rogalski of the CHiBi Proteomics Core Facility. For the NMR studies, I prepared the samples myself and the experiments were set up with the assistance of Mark Okon. Preliminary results from Eric Escobar were consulted in the design of this project.

Figure 6 was produced by our collaborators (Naniye Cetinbas, Cristina Togon, and Poul Sorenson)

Table of Contents

Abstract i

Preface ii

Table of Contents iii

List of tables v

List of figures vi

Abbreviations viii

Acknowledgements x

Chapter 1 Introduction 1

1.1 ETS PROTEINS AND THEIR SIGNIFICANCE 1
1.2 THE SAM DOMAIN 2
1.3 THE ETV6 GENE AND PROTEIN 6
1.4 RECEPTOR TYROSINE KINASES AND THEIR ROLE IN CELLULAR REGULATION 9
1.5 ETV6-NTRK3 FUSION GENE 11
1.6 SUMOYLATION OF ETV6 13
1.7 THE ROLE OF K99 AT THE BINDING INTERFACE 14
1.8 GOALS AND HYPOTHESIS 17

Chapter 2 Materials and methods 18

2.1 DESIGNING CONSTRUCTS OF ETV6 PNT DOMAIN 18
2.1.1 Monomeric N90SA93D-ETV6$^{1-125}$ 18
2.1.2 ETV6$^{43-125}$ constructs 21
2.2 PROTEIN EXPRESSION AND PURIFICATION 22
2.3 NMR SPECTROSCOPY 24
2.4 ISOTHERMAL TITRATION CALORIMETRY 25

Chapter 3 NMR studies of the ETV6 PNT domain 26

3.1 PROTEIN NMR: A GENERAL INTRODUCTION 26
3.2 AN NMR ASSIGNMENT OF ETV6 PNT DOMAIN 26
3.2.1 An assignment of the backbone residue signals of a full-length construct 28
3.2.2 A shortened construct and doubled peaks 31
3.2.3 Secondary structure analysis 41
3.3 COMPARING THE VARIOUS MONOMERIC POINT MUTANTS BY NMR 45
3.4 AN NMR TITRATION OF ETV6 PNT DOMAIN MONOMERS 49

Chapter 4 Binding studies of the SAM-ETV6 domain using ITC 52

4.1 INTRODUCTION TO ITC 52
4.2 K99 AFFECTS BINDING OF MONOMERS 53

Chapter 5 Concluding remarks and future directions 59

5.1 GOALS OF THIS THESIS 59
5.2 FUTURE WORK 62

Bibliography 63

Appendices 68

Appendix A: Chemical shift assignments 68

Appendix B: SSP Scores from NMR assignments 74

Appendix C: ITC titration results 78

List of tables

Table 1. The primers used for PCR 19

Table 2. Binding thermodynamic parameters determination from ITC measurements..... 57

Table 3. A summary of the chemical shift residue assignments for N90SA93D-ETV6$^{1-125}$ 68

Table 4. A summary of assignment of chemical shift residues for A93D-ETV6$^{43-125}$. 71

Table 5. SSP score values for A93D-ETV6$^{1-125}$ 74

Table 6. SSP score values for A93D-ETV6$^{43-125}$ 76

Table 7. ITC results for individual experiments 78

List of figures

Figure 1. Structures of several SAM domains 4

Figure 2. The structure of the ETV6 SAM domain with a native dimer interface 5

Figure 3. Head-to-tail SAM domain polymerization 8

Figure 4. Structure of the ligand binding domain of TrkC and the kinase domain of TrkB 10

Figure 5. Modular structures of wild type ETV6, ETV6-NTRK3 chimera, and wild type NTRK3 12

Figure 6. A summary of results from our collaborators in the Sorenson lab 16

Figure 7. Schematics of the initial ETV6 $^{1-125}$ and subsequent ETV6 $^{43-125}$ constructs 20

Figure 8. Strategy for assignment of protein NMR signals. 27

Figure 9. An assignment of the ^{1}H-^{15}N HSQC spectrum of N90SA93DETV6$^{1-125}$ 30

Figure 10. An assigned ^{1}H-^{15}N HSQC spectra for A93D-ETV6$^{43-125}$ 32

Figure 11. A ^{1}H-^{15}N HSQC monitored pH titration of A93D-ETV6$^{43-125}$ 35

Figure 12. Overlaid ^{1}H-^{15}N HSQC spectra of A93D-ETV6$^{43-125}$recorded over 4 days 37

Figure 13. Reconstructed ESI mass spectrum of A93D-ETV6$^{43-125}$ 38

Figure 14. ^{1}H-^{15}N HSQC spectrum of N90SA93D-ETV6$^{43-125}$ 40

Figure 15. SSP scores for N90SA93D-ETV6$^{1-125}$ 43

Figure 16. Secondary structure SSP score for the A93D-ETV6$^{43-125}$ 44

Figure 17. A93D-ETV6$^{43-125}$ construct overlaid with A93DK99R-ETV6$^{43-125}$ construct 46

Figure 18. ^{1}H-^{15}N HSQC overlay of A93D-ETV6$^{43-125}$ with V112E-ETV6$^{43-125}$ 47

Figure 19. ^{1}H-^{15}N HSQC overlay of A93DK99R-ETV6$^{43-125}$ with V112E-ETV6$^{43-125}$ 48

Figure 20. A ^{1}H-^{15}N HSQC monitored titration of unlabeled A93D-ETV6$^{43-125}$ into ^{15}N-labeled V112E-ETV6$^{43-125}$ 50

Figure 21. A ^{1}H-^{15}N HSQC monitored titration of unlabeled A93DK99R-ETV6$^{43-125}$ into ^{15}N-labeled V112E-ETV6$^{43-125}$ 51

Figure 22. An ITC titration of A93D-ETV6$^{43-125}$ into V112E-ETV6$^{43-125}$ 54

Figure 23. An ITC titration of A93DK99R-ETV6$^{43-125}$ into V112E-ETV6$^{43-125}$ 55

Figure 24. An ITC titration of A93D-ETV6$^{43-125}$ into V112EK99R-ETV6$^{43-125}$ 56

Figure 25. A cartoon summary of the native binding interface of the ETV6 SAM domain 61

Figure 26. EMS m/z measurement of A93D-ETV6$^{43-125}$ 79

Abbreviations

2D	two-dimensional
3D	three-dimensional
Da	Dalton
D_2O	deuterium oxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EMSA	electrophoretic mobility shift assay
ETS	E-twenty-six transformation specific
ETV6	ets-variant gene six
EN	ETV6-NTRK3
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl-β-D-thiogalactopyranoside
ITC	isothermal titration calorimetry
MAE	modulator of the activity of Ets
MOPS	3-(N-morpholino)propanesulfonic acid
NMR	nuclear magnetic resonance
NTRK	neurotrophic tyrosine kinase
PCR	polymerase chain reaction
PNT	pointed
PTK	protein tyrosine kinase
ppm	parts per million
RNA	ribonucleic acid

SAM	sterile alpha motif
SDS	sodium dodecyl sulfate
SUMO	small ubiquitin-like modifier
TEL	translocation Ets leukemia
Tris	tris(hydroxymethyl)aminomethane
WT	wild type

Acknowledgements

This work presented in this thesis would not have been possible without guidance from and support of numerous people. Firstly and foremost I would like to thank my supervisor Dr. Lawrence McIntosh who gave me the opportunity to work on this project.

I would also like to thank past and current members of the McIntosh lab for their advice, support, and time. In particular I would like to thank Dr. H. Jerome Coyne for his implementation of protein purification protocols which were the basis for the techniques used in the majority of this thesis, Desmond Lau for his help with learning new techniques and troubleshooting protein precipitation problems, and Dr. Mark Okon for his endless NMR knowledge and patience in teaching me.

Genevieve Desjardins, thank you for the helpful scientific discussions, editing expertise, and ongoing support and friendship over the years. I would also like to thank Dr. Soumya De for his patience and time in teaching me how to run ITC and discuss my results, and Dr. Laura Packer for generosity with her time, helpful editing, and pep-talks.

I would like to thank my family and friends for their support in all of my endeavors. Particularly, Pippa Payne, your outpouring of encouragement and motivating work sessions have been invaluable in the completion of this thesis. My thanks and gratitude to Lyndsey Earl, for supplying meals, pictures of puppies and helpful chats. Lastly my mother, father, and brothers Maxwell and Emmet have been irreplaceable and their support of me through the years this is something I am forever thankful for.

Chapter 1 Introduction

1.1 Ets proteins and their significance

ETS (E26 transformation specific) proteins play a significant role in cellular differentiation and development. These proteins form a family of eukaryotic DNA-binding transcription factors that direct gene expression by binding to promoters and enhancers and facilitating assembly of transcriptional machinery [23]. ETS members include, among others, PU.1, Ets-1, Fli1, and ETV6 [2]. The biological roles of ETS members vary from organ cell fate determination to neuronal development, oogenesis, and hematopoiesis [2, 3, 5, 15, 23, 37, 38, 47]. Because of the wide variety of functionality of ETS proteins, errors in regulatory pathways of ETS factors can result in a multitude of cellular problems [18]. Understanding the regulation and role of these proteins is of particular interest because of their involvement in oncogenesis.

ETS proteins all have a conserved winged helix-turn-helix ETS domain that binds to the major groove of DNA over a region of ~ 9 base pairs with a common $^{5'}$ GGA(A/T) $^{3'}$ core [23, 40]. The structural mechanisms for DNA-binding by several ETS domains have been determined from X-ray crystallography and NMR spectroscopy [23]. Although the DNA binding domain of ETS proteins is highly conserved among the 28 members of the ETS family, they have a wide array of cellular functions. This diversity amongst ETS family members arises from numerous routes including additional regulatory domains located elsewhere on the proteins, post-translational modifications, and protein-protein interactions. One such regulatory domain that is found in approximately one third of ETS members is the protein-protein interacting SAM (or PNT) domain.

1.2 The SAM domain

The SAM (sterile alpha motif) domain is a commonly found helical bundle with diverse functions that generally involve serving as a protein-interaction module in homo-SAM, hetero-SAM, and heterotypic interactions with non-SAM domain containing proteins. Some SAM domains bind even to RNA [18, 25]. Proteins containing SAM domains can dimerize or oligomerize with one another depending on cellular conditions and the structure of the particular SAM domain. The SAM domain is found in over two hundred and fifty regulatory proteins including receptor tyrosine kinases, serine and threonine kinases, transcription factors and adapter proteins [24]. A subset of about one-third of ETS transcription factors, including Ets-1, Ets-2, Erg, Fli-1, GABPα, and ETV6 contain a SAM domain [33].

Because of its role as a protein-interaction module, SAM domains are diverse in both function and regulation. SAM domains are important for the cellular regulation of transcription. One well-documented example of the SAM domain's role in transcriptional regulation is exemplified by the Sevenless EGF-receptor signaling pathway leading to *Drosophila* eye development. The pathway involves two ETS factors; Pointed-P2 is an activator of developmental genes in *Drosophila*, whereas Yan is a repressor [23]. Transcriptional repression by Yan requires higher order structure formation, which is mediated by Yan-SAM polymerization [44]. Mae (modulator of the activity of Ets) has a SAM domain, but not an ETS domain, and aids in Yan depolymerization by binding to its SAM domain [23]. When Mae binds to Yan it exposes a phosphoacceptor site on Yan allowing it to be phosphorylated, by the Rolled kinase (activated by EGF signaling). The resulting phosphorylated Yan is subject to nuclear export and cytoplasmic degradation [30]. Meanwhile, the SAM domain of Pointed-P2 is a docking site for Rolled and its phosphorylation leads to transcriptional activation of many genes

including that encoding Mae. As a feedback mechanism, Pointed-P2 is also regulated by Mae. In this case, Mae can heterodimerize with the Pointed-P2 SAM domain, blocking Rolled docking and reducing phosphorylation-dependent transcriptional activation. This example shows that protein partnership and post translational modification of SAM domains are essential for both the activation and repression of a transcriptional pathway.

High resolution structures of many SAM domains have been determined from both X-ray crystallographic and NMR spectroscopic studies (Figure 1) [23]. The core structure of the SAM domain is a four α-helical bundle (denoted as helices H2-H5 for consistency with literature publications) with a short 3_{10} or α-helix (H2') [23]. The SAM domain structures of ETS family members, Ets1, Pnt-P2, Erg, GABPα, Yan, and ETV6 all show the common α-helical core of the SAM domain with slight variations [11, 47-50, 53]. The interaction surfaces of the SAM domain have been termed the mid-loop (ML) and end-helix (EH). In the crystal structure of the ETV6-SAM domain, the interactions between these two surfaces results in an extended superhelical head-to-tail polymer (Figure 2) [24]. ETV6, Yan, and likely ETV7 are the only ETS proteins with SAM (PNT) domains having exposed ML and EH interfaces necessary for self-association in this head-to-tail fashion. All others, including Ets1 and Pnt-P2, have one or both surfaces blocked by amino acid substitutions or appended helices and are therefore monomeric in solution [23].

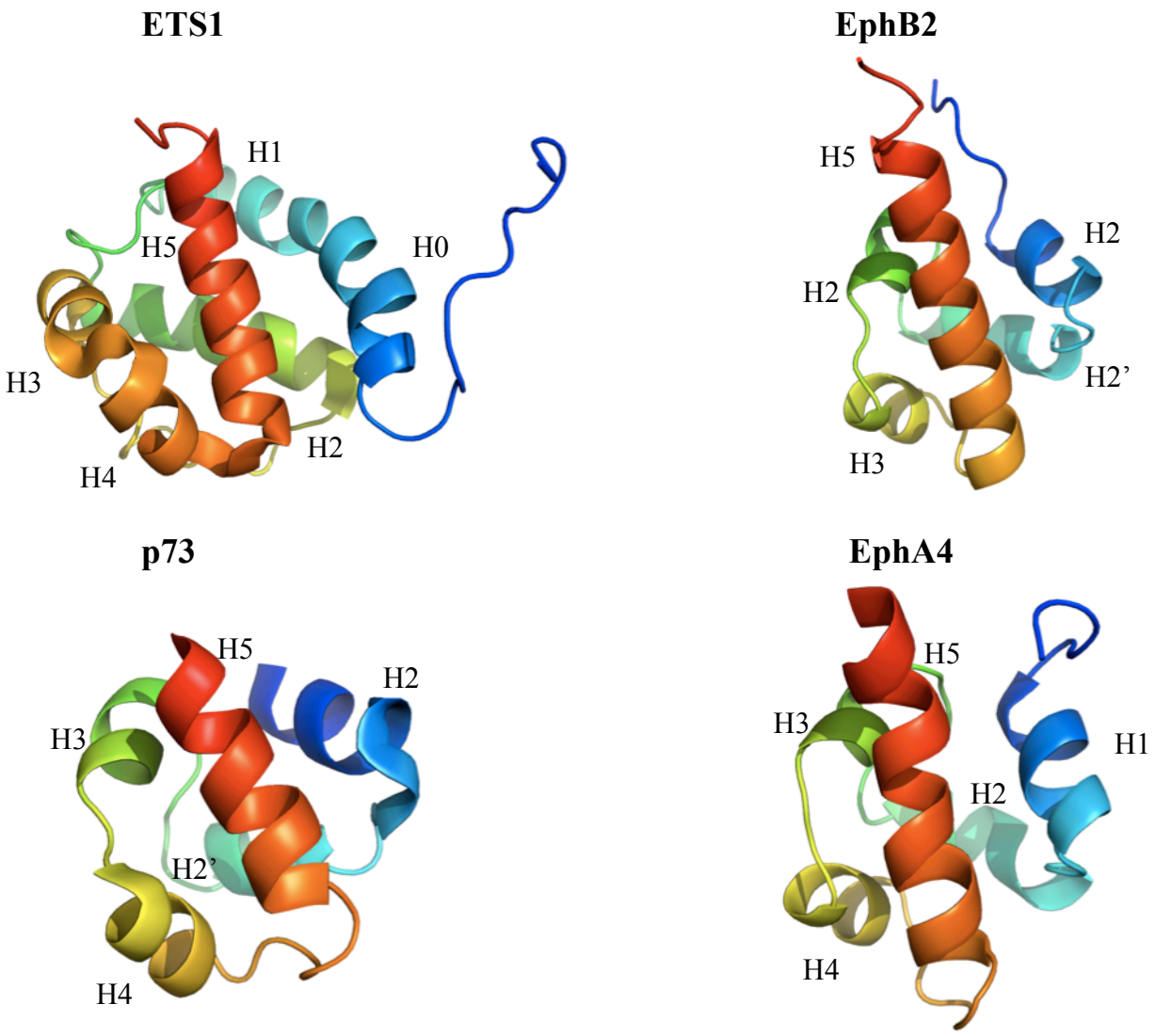


Figure 1. Structures of several SAM domains. Clockwise from top left: ETS1 (2JV3.pdb), EphB2 (1SGG.pdb), EphA4 (1B0X.pdb), and p73 (1DXS.pdb). Helices are labeled and are colored from N-terminus to C-terminus in blue to red. The common SAM domain core helical structure can be readily seen.

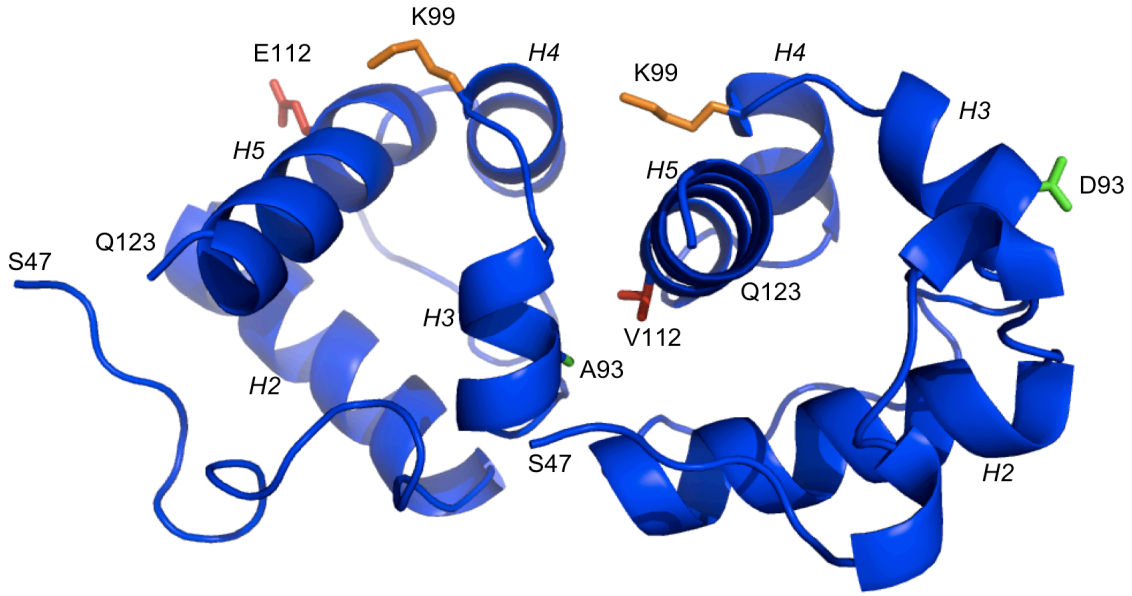


Figure 2. The structure of the ETV6 SAM domain with a native dimer interface (1LKY.pdb). Mutation of either Ala93 to Asp (A93D; green) or Val112 to Glu (V112E; red) prevents self-association. However, the A93D and V112E mutants can still form a "heterodimer" with a native interface when added together. Terminal residues (S47 and Q123 respectively) and helices H2-H5 are also labeled.

1.3 The ETV6 gene and protein

ETV6 (also known as TEL or TEL1 for translocation Ets leukemia) is an ETS family member that plays key roles in embryonic development as well as hematopoietic regulation and is essential for yolk sac angiogenesis [29]. The ETV6 gene consists of 8 exons, with alternative start codons at positions 1 and 43 [12]. Full length ETV6 protein contains 452 amino acids with an N-terminal SAM domain and C-terminal ETS domain. Unlike most ETS factors that activate gene expression, ETV6 is a transcriptional repressor and reported tumor suppressor [32]. Critical to its role as a transcriptional repressor is its ability to self-associate via the SAM domain [12].

SAM domain self-association is hypothesized to allow ETV6 to cooperatively bind adjacent DNA recognition sites as an extended polymer [19]. In contrast, the isolated, monomeric ETV6-ETS domain binds consensus sequence with relatively low affinity (2.8×10^{-8} M) [19]. This low affinity is caused by a dynamic appended inhibitory helix that sterically blocks the DNA-binding interface of the ETS domain and thereby attenuates DNA binding ~50-fold [19]. Previous work by our lab in collaboration with the Graves group (University of Utah), supports a model where ETV6 binds to DNA as a cooperative polymer, thereby compensating for the otherwise low affinity of the inhibited ETS domain [19]. This result was obtained using electrophoretic mobility shift assays (EMSAs) to measure the dissociation kinetics of monomeric versus dimeric forms of ETV6 from DNA. These data showed that ETV6 dimers were more stably bound than monomers on tandem ETS binding sites, supporting the theory that the SAM domain mediated oligomerization of ETV6 would allow overall higher affinity binding as a cooperative polymer [19].

As mentioned previously, the native interface of the ETV6-SAM domain has been determined by crystallographic methods (Figure 2) [24, 55]. The interfaces have a non-polar core

with residues Met79, Ala83, Leu86, and Leu87, on the ML surface and Phe67, Leu69, Val112, and Leu106 on the EH surface [55]. Several interdomain salt bridges surround this hydrophobic core including Glu66 to Lys82', Arg95 to Asp101', Asp91 to Arg93', Glu90 to Lys89', and Asp91 to Lys89' (with ‘ denoting the partner protein) [55]. Salt bridges contribute to overall protein structure and stability. These salt-bridge interactions are not necessarily conserved and some mutations can be accommodated at the dimer interface [55].

The polymerization of wild-type ETV6-SAM domain makes *in vitro* biophysical studies difficult. Fortunately point mutations for the ETV6-SAM domain have been found which substitute in a hydrophobic to charged residue on either the ML or EH surface [23]. These mutations (A93D and V112E) disrupt polymerization of the SAM domain and are also monomerizing. Mixing a ML-mutant with EH-mutant yields a "hetero-dimer" that is tightly associated (K_d=1.7 nM by surface plasmon resonance [24]) via their respective wild-type interfaces (Figure 3). This dimer is used as a model polymer because it retains one wild-type interaction between the ML and EH surfaces and allows for investigation of the native binding surface of the SAM-ETV6 domain.

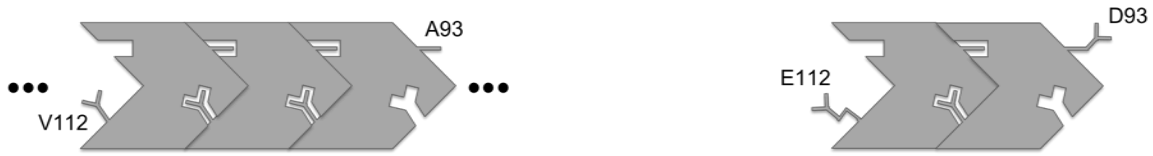


Figure 3. Head-to-tail SAM domain polymerization of native ETV6 (left). Solubilizing mutations A93D and V112E are highlighted. A mixture of the two monomerized mutants results in a heterodimer with a native binding interface (right).

Based on SAM-ETV6 domain crystal structure, Bowie and co-workers proposed a model whereby monomers self-associate in a head-to-tail fashion with a 6_5-screw symmetry to form an extended, repressive polymer along DNA [24]. Evidence for this model was obtained by adding soluble monomerized mutants to wild type protein in order to cap the otherwise insoluble polymers [24]. Subsequent electron microscopy showed soluble filaments with a width dimension similar to those of the crystallized helical polymer [24]. This suggested that ETV6 can form a polymeric structure that links many DNA binding elements together and spread repression over a large segment of chromatin [24]. Furthermore, the proposed model shows the C-terminus of the SAM domain pointing outwards away from the polymer axis [24]. The C-terminal ETS domain is therefore also pointed outwards and can help recruit chromatin to toroidally wrap around the polymer [24]. The model is also supported by a 53 Å repeat of the ETV6 polymer, which is the same dimension as a nucleosome core particle [24]. The authors proposed that once nucleated on DNA, only weak-binding affinity for DNA is required for further ETV6 protein association [24]. Thus, the entropic cost of uniting protein and DNA is paid by the strong SAM-SAM interactions [24]. This is consistent with the previously mentioned studies of Dr. Graves showing that co-operative binding to tandem DNA sites by ETV6 overcomes the auto-inhibition of its ETS domain [19].

1.4 Receptor tyrosine kinases and their role in cellular regulation

Mis-regulation of Neurotrophin-tyrosine kinase receptors (NTRKs) contributes significantly to neuroblastoma development [31]. TrkC (also called NTRK3) in particular has been shown to be involved in soft tissue cancers [21]. Furthermore, it is of interest to us because of its involvement in an oncogenic chimeric protein, resulting from a chromosomal translocation that fuses the genes encoding a functional SAM domain of ETV6 with the receptor tyrosine kinase domain of NTRK3.

Tyrosine kinases are important regulatory proteins that transfer a phosphate group from ATP to a tyrosine residue of a protein. Receptor tyrosine kinases (RTKs) are a subset of these proteins that also have a transmembrane domain linking an extracellular receptor domain to the intercellular catalytic domain. As such, RTKs are central for signal transduction and cellular regulation. The oncogenic potential of aberrant tyrosine kinase receptors is well documented and thus they are attractive drug targets [28].

Neurotrophin-tyrosine kinase receptors are RTKs which are essential for development and maintenance of the nervous system [57]. These proteins generally consist of a ligand binding domain, as well as a kinase domain with a linker region. The crystal structure of the TrkC kinase domain has been solved in complex with an inhibitor (3V5Q.pdb) [1] and related protein kinase domains TrkA, and TrkB have also been solved by X-ray crystallography (4F0I.pdb and 1HCF.pdb respectively) [4]. The crystal structure of the TrkC ligand binding domain has also been previously determined (Figure 4, 1WWC.pdb) [57].

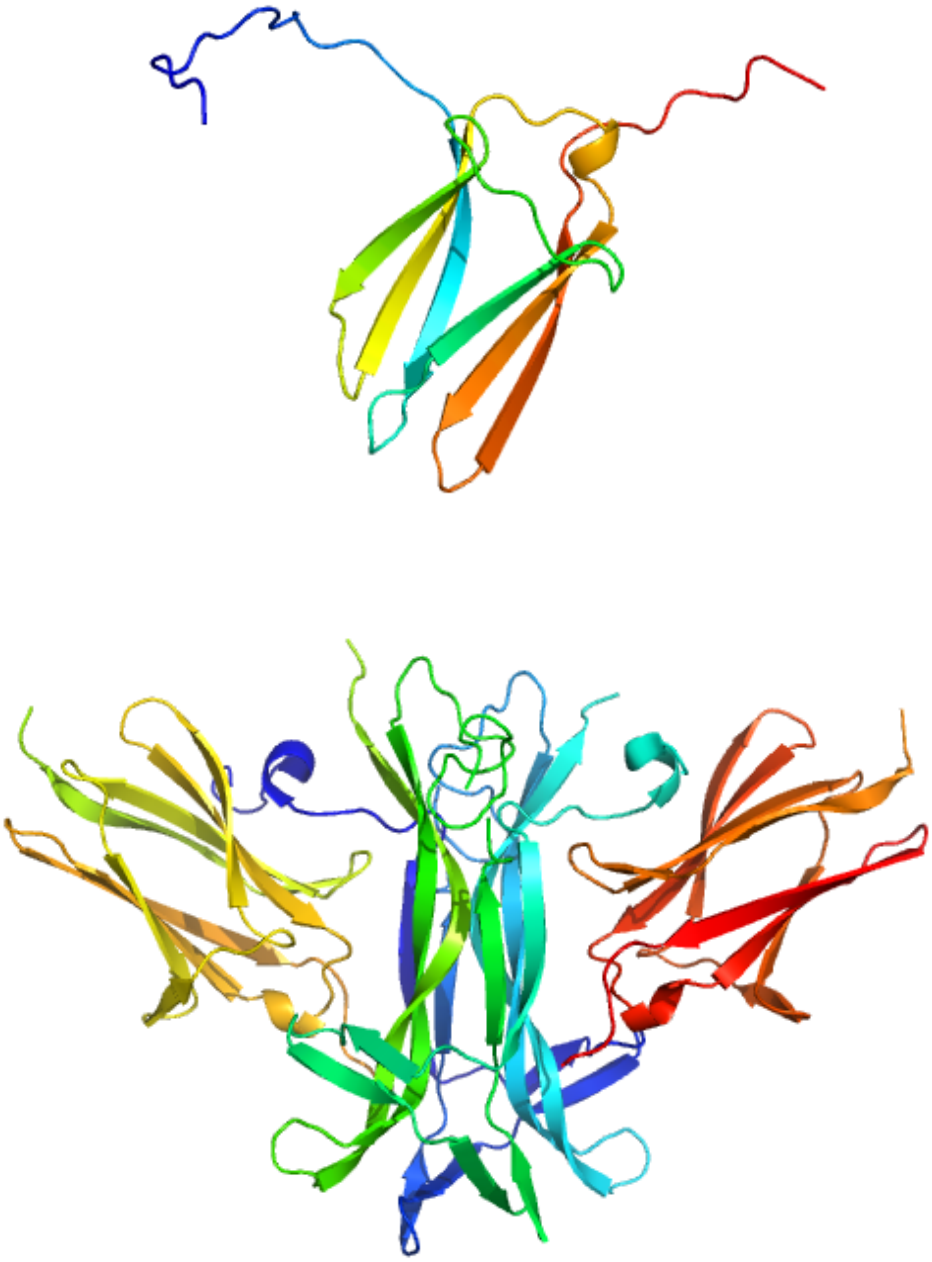


Figure 4. Structure of the ligand binding domain of TrkC (top, 1WWC.pdb) and the kinase domain of TrkB (bottom, 1HCF.pdb). The kinase domain of TrkC is similar to that of TrkB.

1.5 ETV6-NTRK3 fusion gene

ETV6 is involved in numerous chromosomal translocations, several of which result in fusion genes encoding chimeric oncoproteins that contribute to the development of leukemia or leukemogenesis [12]. In these chimeric oncoproteins the SAM domain of ETV6 mediates polymerization that is critical for transformation activity. Frequently, the translocations fuse the SAM domain to either a DNA-binding transcription factor such as AML or a protein tyrosine kinase (PTK) such as PDGFRβ, ABL, JAK2, ARG, or FGFR3 [12, 29]. These resultant fusion proteins have oncogenic potential and are often associated with myeloid and lymphoid malignancies [60].

Transformation of cells by ETV6 chimeras with PTK's have been shown to require both the active tyrosine kinase domain and an intact SAM oligomerization domain [29]. As previously mentioned, one of these translocation products fuses the N-terminal SAM domain of ETV6 to the C-terminal tyrosine kinase domain of NTRK3 (Figure 5). This chimeric fusion (EN) was first characterized by Dr. Poul Sorensen's group at the BC Cancer Agency [26]. EN polymerizes via the SAM domain, which results in auto-phosphorylation and constitutive activation of the tyrosine kinase domain [55]. This fusion shows *in vivo* transformation activity in soft agar colony [29], a technique which will be discussed in detail later. The introduction of the previously mentioned monomerizing mutations A93D and V112E into the SAM domain prevents polymerization and abrogates transformation of cells in soft agar assays [54]. However, replacement of the SAM domain with an inducible dimerization system was shown to not transform cells [54]. Thus, oligomerization, rather than simple dimerization appears to be necessary for cellular transformation and oncogenic activity.

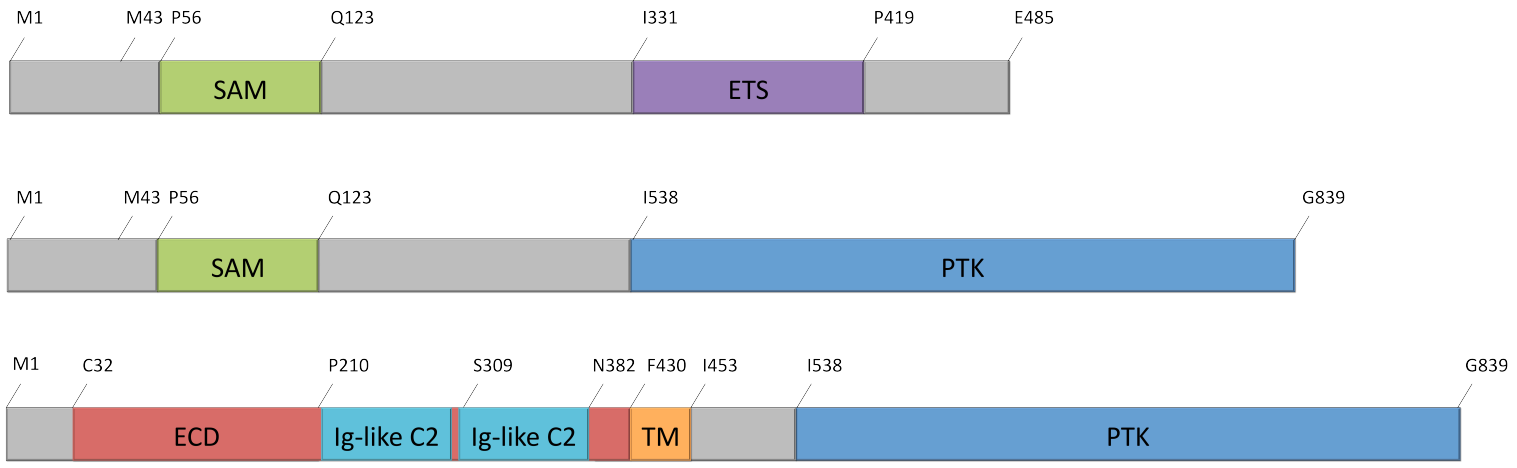


Figure 5. Modular structures of wild type ETV6 (top), ETV6-NTRK3 chimera (middle), and wild type NTRK3 (bottom). The EN chimera results from fusion of the ETV6 PNT domain with the NTRK3 protein tyrosine kinase (PTK) domain. Protein domain boundaries are noted for the ligand binding Ig-like C2 domain (cyan), extracellular domain (red), transmembrane domain (orange), SAM domain (green), ETS domain (purple) and protein tyrosine kinase domain (blue).

1.6 Sumoylation of ETV6

As mentioned previously, post-translational modification of proteins is an important way of regulating function. Sumoylation of ETV6 was initially reported to occur at a non-consensus Lys99 within the ordered SAM domain [9, 10, 61]. Small ubiquitin-like modifier (SUMO) is a protein that becomes covalently linked to other proteins in order to mediate their function [20]. Sumoylation of proteins is an important post-translational cellular regulatory mechanism and results from several enzymatic steps similar to the ubiquitin E1-E2-E3 cascade in order to occur [20]. SUMO has been shown to help regulate transcription factors [16]. In most cases, sumoylation inhibits transcription, although the mechanism is not fully understood and likely involves the recruitment of additional co-repressors [16]. Sumoylation frequently occurs in unstructured regions of proteins at a consensus modification site ΨKXE where Ψ is a large hydrophobic residue, E is glutamate, X is any residue, and K is the lysine sumoylation site [16].

Several studies have shown that mutation of the non-consensus Lys99 site to an arginine residue affects localization and potentially the ability of the SAM domain to polymerize [45]. Initially this effect was attributed to the inability of the arginine residue to act as a sumoylation site. However a more recent study has shown that Lys11 is the primary sumoylation site in the SAM-ETV6 domain and this post-translational modification was found to inhibit ETV6 repression of gene expression [45]. This study also showed that both monomers and oligomers of ETV6 were sumoylated in vitro, whereas in vivo only oligomers were found to be sumoylated [45]. Although Lys99 is not the primary sumoylation site for the SAM domain of ETV6 is it of interest due to its proximity to the ML/EH interface and the non-transforming phenotypic change that occurs in soft agar assays when it is mutated to an arginine residue [8].

1.7 The role of Lys99 at the binding interface

In a previous collaborative effort between the Sorensen and McIntosh group, the monomerizing mutations A93D and V112E were found to abrogate cellular transformation by EN [54]. More recently, Naniye Cetinbas, a graduate student in the Sorenson group, observed similar effects while investigating the role of Lys99 [8]. In particular she found that the K99R and K99D mutations reduced high molecular mass complex formation of EN and abrogated its transformation activity. Additionally, soft agar assays were conducted on various constructs of EN. NIH3T3 fibroblasts were used in these assays with an empty vector used as a negative control, and WT-EN was used as a positive control. When the cells are grown in the soft agar, negative controls undergo detachment-induced apoptosis, whereas those cells transformed by WT-EN have been previously shown to be resistant to detachment-induced apoptosis [35, 39]. Several EN constructs were tested, one lacking the SAM domain of the ETV6 portion of EN (ΔSAM-EN), a kinase dead EN (K380N-EN) and K99R-EN. The ΔSAM-EN, and K380N-EN were both found to be non-transforming and K99R-EN was shown to have high levels of PARP, a marker for detachment-induced apoptosis. Additionally, these three cell lines showed smaller sized spheroids in the soft agar analysis when compared to cells expressing WT-EN. These results are summarized in Figure 6. She also showed that mutation of Asp101, the intermolecular salt bridge partner of Lys99 to alanine or lysine similarly blocked transformation of NIH3T3 cells by EN, reduced EN tyrosine phosphorylation, inhibited Akt and Mek1/2 signaling downstream of EN, and abolished tumor formation in nude mice. In contrast, mutations of Glu100 and Arg103 residues in the vicinity of the interdomain Lys99-Asp101 salt bridge, had little or no effect of these oncogenic characteristics of EN. Her results underscored the importance of Lys99 and Asp101 for SAM polymerization and EN transformation. Her results

led to the hypothesis that these residues contribute to a salt bridge important for SAM domain polymerization and therefore important for transformation. Using several biophysical techniques I sought to test this hypothesis regarding the non-transforming phenotype present in the K99R-EN cell line.

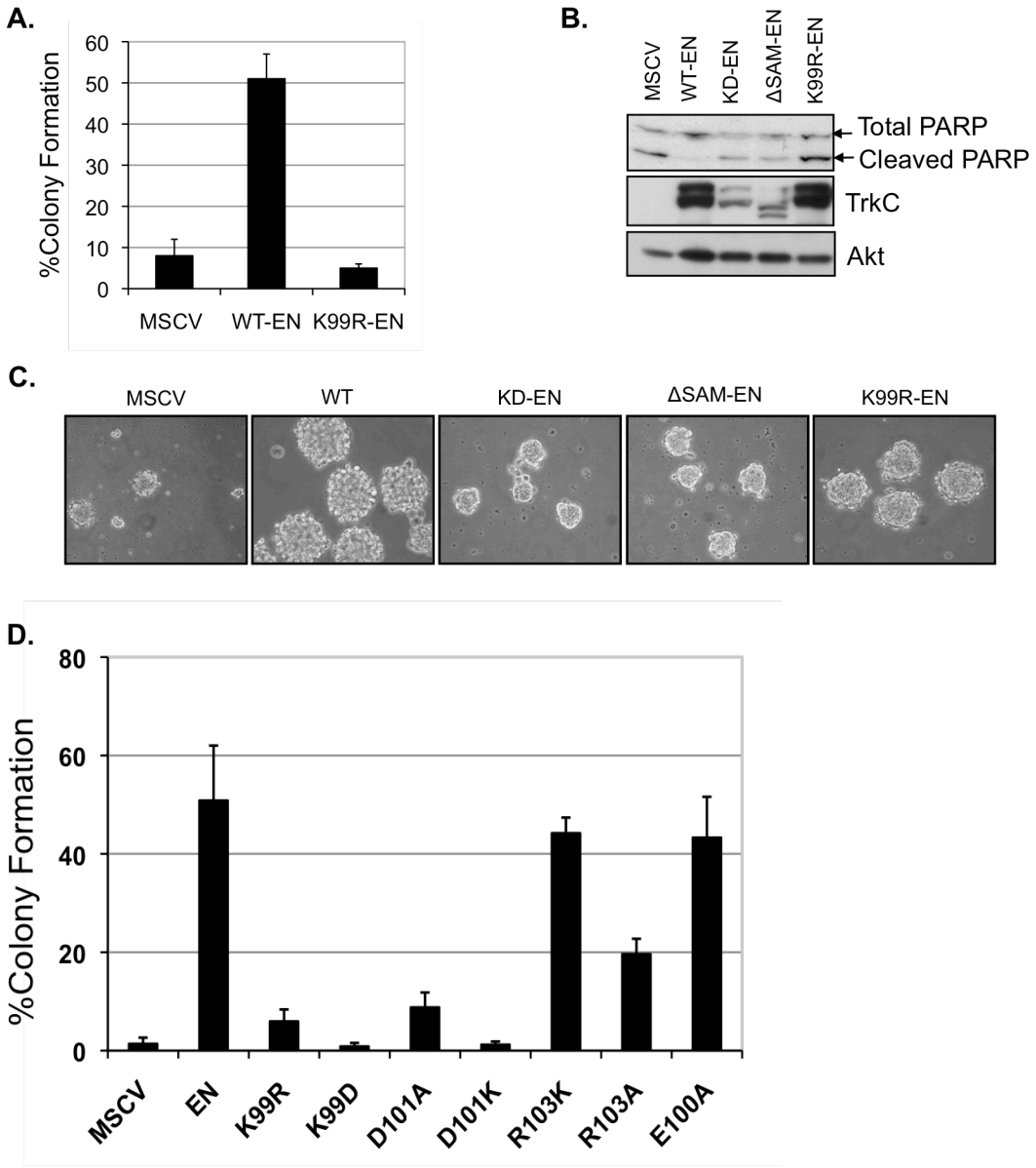


Figure 6. A summary of results from collaborators in Sorenson lab; (A) Soft agar colony formation assay shows K99R mutation gives lower colony formation at (B) Western blot analysis of cleaved PARP to probe detachment-induced cell death. (C) Pictures depicting the spheroid sizes formed by the NIH3T3 cells expressing the indicated constructs. Empty MSCV vector was used as a negative control, KD-EN indicated a kinase-dead version of EN ΔSAM-EN indicates a version of EN without a SAM domain, and K99R-EN has the single point mutation K99R. (D) NIH3T3 cells stably expressing the indicated constructs were seeded in soft agar. Colony formation was shown to depend on the residue selected for point mutation. Versions of this data have been previously published [8].

1.8 Goals and hypothesis

The goal of this project was to investigate the role that Lys99 plays in the binding interface of the ETV6 SAM domain. Additionally we sought to determine how a mutation of Lys99 could alter the phenotypic transformation of cells by EN in soft agar assays. We hypothesized that the effect of Lys99 and Asp101 mutations on cell transformation is due to the attenuation of a SAM domain polymerization via disruption of an intermolecular salt bridge. The specific aims of this thesis were therefore to investigate the effect of point mutations on the structure of the isolated ETV6-SAM domain in solution as well as on the thermodynamic parameters underlying polymerization.

Using NMR spectroscopy and isothermal titration calorimetry (ITC), I found that mutation of Lys99 to arginine does not significantly perturb the structure of the isolated ETV6-SAM domain with the monomerizing A93D substitution. However mixed dimer formation was weakened 400-fold when the K99R mutation was present in the wild-type interface. This quantitatively demonstrated the importance of intermolecular salt bridges in SAM domain polymerization. Furthermore these results support the hypothesis that cellular transformation by EN is critically dependent upon SAM domain-mediated self-association, and thus is particularly relevant in terms of oncogenesis.

Chapter 2 Materials and methods

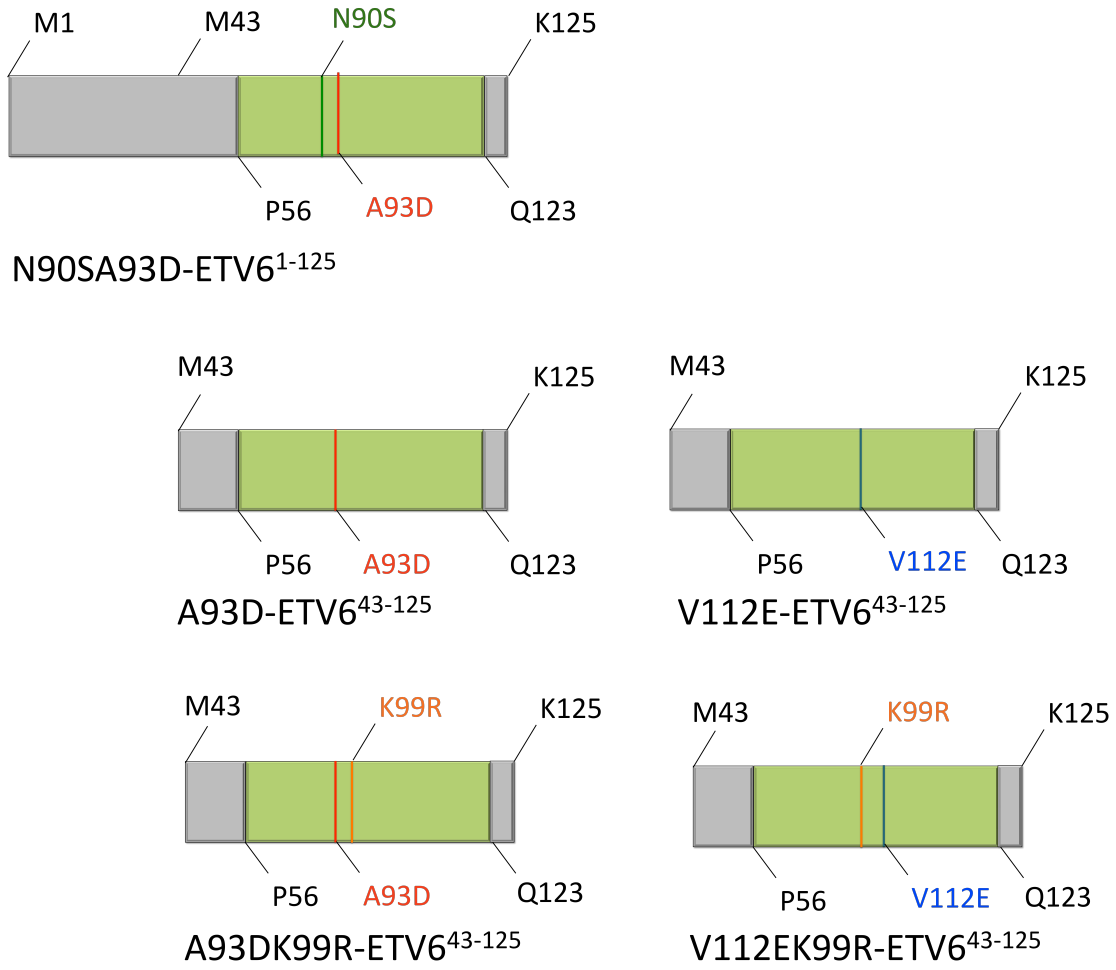
2.1 Designing constructs of ETV6 PNT domain

2.1.1 Monomeric N90SA93D-ETV6$^{1-125}$

Initially, the gene encoding a fragment of ETV6 spanning residues 1-125 (Figure 5) was generated by standard PCR methods [27] from the full length EN gene provided by the Sorenson laboratory. Primers used for this PCR reaction can be found in Table 1. This first construct, denoted N90SA93D-ETV6$^{1-125}$, encodes the SAM domain (residues Pro56-Gln123) with the monomizering mutation A93D, preceded by the full N-terminal segment of ETV6. Using introduced *EcoR*I site and *Nde*I restriction sites, the PCR product ligated into a pET28a plasmid (Invitrogen). The resulting gene also included an N-terminal His_6-tag and a thrombin cleavage site. The ligated plasmid was transformed into *Escherichia coli* DH5α cells and grown on LB agar plates containing kanamycin (35 mg/L). Plasmids from the resulting colonies were extracted using a GeneJET Plasmid Miniprep Kit (Fermentas) and analyzed by commercial DNA sequencing (GENEWIZ). It was at this time that the unwanted mutation N90S was also discovered. This initial full-length construct is shown in Figure 7 along with the amino acid sequence for the SAM domain of ETV6.

Primer	Direction	Sequence (5' to 3')
EcoRI		CTGAAGCAGAGGAAATGAGAATTCTATA 3'
*Nde*I(1-125)		CTGAAGCAGAGGAAATGAAAGCTTTATA3'
*Hind*III		TATAAAGCTTTCATTTCCTCTGCTTCAG 3'
*Nde*I(43-125)		TATACATATGGAGGAAGACTCGATCCG 3'
S90N	Forward	CAGCAACACGTTTGAAATGAATGGCAAAGATCTCCTGCTGC 3'
	Reverse	GCAGCAGGAGATCTTTGCCATTCATTTCAAACGTGTTGCTG 3'
D101K	Forward	GAATGAGGAGATCGATAGCGAAATTTCTCTTTGGTCAGCAGCAGGAG
	Reverse	CTCCTGCTGCTGACCAAAGAGAAATTTCGCTATCGATCTCCTCATTC
N90S into A93D	Forward	GCAGCAGGAGATCTTTGCCACTCATTTCAAACGTGTTGCTG
	Reverse	CAGCAACACGTTTGAAATGAGTGGCAAAGATCTCCTGCTGC
K99R into V112E	Forward	GGCAAAGCTCTCCTGCTGCTGACCCGTGAGGACTTTCGCTATCGATCTCC
	Reverse	GGAGATCGATAGCGAAAGTCCTCACGGGTCAGCAGCAGGAGAGCTTTGCC
K99R into A93D	Forward	GGCAAAGATCTCCTGCTGCTGACCCGTGAGGACTTTCGCTATCGATCTCC
	Reverse	GGAGATCGATAGCGAAAGTCCTCACGGGTCAGCAGCAGGAGATCTTTGCC
V112E	Forward	GATCTCCTCATTCAGGCGACGAGCTCTATGAACTCCTTCAGC
	Reverse	GCTGAAGGAGTTCATAGAGCTCGTCGCCTGAATGAGGAGATC
D93A	Forward	GTTCGAAATGAATGGCAAGGCCCTCCTGCTGCTGACCAAAG
	Reverse	CTTTGGTCAGCAGCAGGAGGGCCTTGCCATTCATTTCGAAC

Table 1. The primers used for PCR are shown. From a full length EN construct, primers *EcoR*1, and *Nde*I(1-125) were used to create an N90SA93D-ETV6$^{1-125}$ construct in a pET28a plasmid. Primers S90N (forward and reverse) were used for site directed mutagenesis to remove the N90S point mutation found in this construct. The primer set *Hind*III and *NdeI*(43-125) were used to amplify the A93D-ETV6$^{43-125}$ and then ligated into pET28a or pET28-MHL vector.



MSETPAQCSIKQERISYTPPESPVPSYASSTPLHVPVPRALR**M**EEDSIRLPAHLRLQPIY

WSRDDVAQWLKWAENEFSLRPIDSNTFEM**N**GK**A**LLLLT**K**EDFRYRSPHSGD**V**LYEL

LQHILKQRK

Figure 7. Schematics of the initial ETV6$^{1-125}$ and subsequent ETV6$^{43-125}$ constructs. The SAM domain is highlighted in green. The solubilizing mutations A93D and V112E, as well as additional point mutations N90S and K99R, are indicated. Below The wild-type amino acid sequence of ETV6$^{1-125}$ with the alternative start site M43 in bold and sites of interest N90, A93, K99, D101, V112, colored (green, red, orange, blue respectively)

2.1.2 ETV6$^{43-125}$ constructs

In addition to containing an unwanted N90S mutation, N90SA93D-ETV6$^{1-125}$ also contained a thrombin cleavage site at residues V37-P38-R39||A40. Due to problems with producing stable soluble protein without a His_6-tag, and given that the residues preceding the SAM domain appeared unstructured in preliminary NMR spectra, a set of shorter ETV6$^{43-125}$ fragments were generated. Residue Met43 is an alternative start site for ETV6 [42] and thus was chosen as the new N-terminus. Similar constructs of this length have been previously studied [55] and thus we felt justified in using a shortened construct. We also explored the introduction of a TEV cleavage system into the clones. In order to efficiently obtain constructs, several cloning strategies were attempted concurrently. The N90S$^{1-125}$ construct was first corrected back to Asn90 using site directed mutagenesis, and then using *Nde*I(43-125) and *Hind*III primers (Table 1), a shortened version of the SAM domain was PCR amplified from this plasmid. This DNA was then cleaved using the appropriate enzymes and ligated into a similarly cleaved pETMHL plasmid. Site directed mutagenesis was then used to introduce the K99R mutation into the A93D construct. Additionally shortened versions of the V112E point mutant were created using an A93D-ETV6$^{43-125}$-pET28a construct. First the A93D mutation was corrected back to Ala93, and then the V112E mutation was introduced. From this shortened V112E-ETV6$^{43-125}$ pET28a plasmid, a K99R mutation was added using site directed mutagenesis. Additionally a D101K mutation was added into the V112E-ETV6$^{43-125}$ construct. Site directed mutagenesis was also used to introduce N90S mutation back into the A93D-ETV6$^{43-125}$ shortened construct to confirm that a side deamidation reaction was not occurring. This will be discussed later in detail.

2.2 Protein expression and purification

The sequence-verified plasmids were introduced into *E. coli* BL21 (λDE3) cells by either electroporation or heat shock. SOC media (1 mL) was added to freshly transformed cells and they were incubated at 37°C for 1 hr. These cells were then plated on LB agar plates with kanamycin (35 mg/L) and grown overnight at 37°C. Single colonies from these plates were selected and placed in 25 mL LB overnight. These cultures were spun down at 5000 rpm (GSA rotor Sorvall) for 15 minutes, resuspended in 1 mL media and used to inoculate larger volume expression media.

The ETV6 constructs were expressed in *E. coli* BL21 (λDE3) cells grown at 37°C in Luria broth (LB) or minimal M9 media containing 1g/L $^{15}HN_4Cl$ alone or 1 g/L$^{15}HN_4Cl$ and 3 g/L$^{13}C_6$-glucose (Sigma Aldrich) [36]. Kanamycin (35mg/L) was included in all media. The cells were induced at OD_{600} = 0.6 with a final concentration of 1 mM IPTG, grown overnight, and then harvested by centrifugation (GSA rotor, Sorvall) at 5000 rpm for 30 minutes. The cell pellet was resuspended in 50 mL of binding buffer (0.22μm filtered 20 mM imidazole, 50 mM Na_2HPO_4, 500 mM NaCl, pH 7.5) and either frozen for storage at -80°C or processed immediately. The cell pellet was lysed by passage through a French press twice at 10,000 psi followed by sonication (Branson Sonifier 250, VWR Scientific) at 60% duty cycle for 3x10 minutes. The lysate was spun (SS34 rotor Sorvall) at 15,000 rpm for 1 hour and the resulting supernatent was passed through a 0.8 μm filter before being loaded on a Ni-NTA column (GE Healthcare) pre-equilibrated with binding buffer. The column was washed with several column volumes of binding buffer and then the ETV6 constructs were eluted with elution buffer (0.22μm filtered 500 mM imidizole, 50 mM Na_2HPO_4, 400 mM NaCl, pH 7.5). The fractions containing the desired protein were identified by SDS-PAGE gel analysis, pooled and placed in dialysis

tubing. For samples studied using ITC, these samples were dialysed into a buffer containing 20 mM MOPS, 50 mM NaCl, 0.5 mM EDTA pH 8.0 overnight. For NMR protein samples the appropriate protease to remove the His_6-tag was also added to the dialysis tubing. Constructs with a TEV cleavage site were dialyzed overnight with TEV protease (made in house) at room temperature in 50mM Tris, 50 mM NaCl, pH 8.0, 1mM DTT. Constructs with a thrombin cleavage site were dialyzed overnight in 20 mM Tris pH 8.4, 0.15 M NaCl, 2.5 mM $CaCl_2$ at 4 ^{o}C with 1 unit of thrombin (Novagen). Cleavage was stopped the next day with p-aminobenzamidine beads.

During preparation of the $^{13}C/^{15}N$-labeled N90SA93D-ETV6$^{1-125}$-pET28a construct formation of precipitate during the overnight cleavage of the His_6-tag indicated sample instability. In order to collect as much data as possible, this sample was transferred to a 25 mM KCl, 25 mM Na_2HPO_4, pH 7.0 buffer, concentrated, and 3D NMR data was collected over the course of 4 days. Further purification of this sample was not able to be done in a timely fashion and thus, the His_6-tag cleaved and uncleaved portions of the protein were not separated from one another. Additionally this sample was not purified by size exclusion, and likely contained a mixture of species.

For the shortened ETV6$^{43-125}$ constructs, His_6-tag cleaved proteins were separated from uncleaved protein by passage through a Ni-NTA column (GE Healthcare) pre-equilibrated with binding buffer. The flow-through was collected and concentrated by ultrafiltration with an Amicon (Ultra-15; 3K) to a volume of 2-5 mL. This concentrated sample was spun to remove any precipitate, and then injected onto a pre-equilibrated Sepharose S75 gel filtration column and eluted with NMR buffer (20 mM MOPS, 50 mM NaCl, 0.5 mM EDTA pH 8.0). Fractions containing desired protein were identified using SDS-PAGE, pooled and concentrated for further

analysis. Sample concentration was evaluated by absorbance values at A_{280} using Beer's Law based on the predicted molar absorptivity (20950 $M^{-1}cm^{-1}$) given by the ProtParam program (http://web.expasy.org/protparam/). Mass and purity were confirmed by SDS-PAGE gel. After cleavage with either thrombin or TEV cleavage enzymes, some residual amino acids remain before the wild-type M43 or M1 start sites. Leader sequences of either MGSSHHHHHHSSGLVPRGSHIH (for thrombin cleavage – pET28a vector) or MHHHHHHSSGRENLYFQGHIH (for TEV cleavage – pET28MHL) were incorporated into the V112E and A93D variants, respectively.

2.3 NMR spectroscopy

ETV6 samples were at 0.25-1.0 mM in the previously defined NMR buffer with an additional 10% v/v D_2O added for signal lock. Spectra were recorded at 25°C using 500 MHz Varian Unity, 600 MHz Varian Inova, 500 MHz Bruker Avance III, or 600 MHz Bruker Avance III NMR spectrometers equipped with triple resonance cryogenic probes. Resonance assignments were obtained using standard main-chain correlation experiments, including HNCO, HN(CA)CO, HNCACB, and CBCA(CO)NH [46]. Spectra were analyzed using NMRpipe [13], nmrDraw [13], and Sparky [17].

^{1}H-^{15}N-HSQC spectra were used to monitor titrations of ^{15}N labeled A93D$^{43-125}$ conducted over a pH range from 5.5-8.5 using increments of approximately 0.5 units. From this range it was determined that the protein was relatively stable and soluble over a variety of pH conditions. NMR spectra were collected at pH 7.5 in order to minimize signal loss due to hydrogen exchange and to yield good signal-to-noise. Titrations of monomers were observed using ^{1}H-^{15}N-HSQC spectra of ^{15}N-labeled V112E-ETV6$^{43-125}$; these spectra were collected as

aliquots of unlabeled A93D-ETV6$^{43-125}$ or A93DK99R-ETV6$^{43-125}$ were added to a final 1:1 equivalency.

2.4 Isothermal titration calorimetry

Fractions of His_6-tag purified protein were pooled and dialyzed overnight in 4 L of buffer (20 mM MOPS, 50 mM NaCl, 0.5 mM EDTA, pH 8.0). The His6-tag was left on the constructs to expedite experiments. Since previously reported surface plasmon resonance experiments have been done with an intact N-terminal His_6-tag [24] we conclude that leaving the tag attached would produce comparable binding data. Proteins were concentrated using an Amicon Ultra-15 3K filter. Due to limited solubility, V112E-ETV6$^{43-125}$ and V112EK99R-ETV6$^{43-125}$ mutants were at 0.019-0.038 mM and used within the ITC cell. A93D-ETV6$^{43-125}$ and A93DK99R-ETV6$^{43-125}$ were concentrated further (0.16-0.4 mM) for use in the ITC syringe.

ITC measurements were run using an ITC200 instrument (GE Healthcare). Experiments consisted of 20-25 injections (1.5-2.0 μL) of A93D-ETV6$^{43-125}$ or A93DK99R-ETV6$^{43-125}$ (0.16-0.4 mM) into the stirred (1000 rpm) cell. The resulting data were processed using Origin 7.0. Buffer blank titrations were performed in order to determine that there was no thermodynamic contribution due to dilution.

Chapter 3 NMR studies of the ETV6 PNT domain

3.1 Protein NMR: a general introduction

The use of NMR spectroscopy is a well-established tool for characterizing the structure and dynamics of a protein. The development of heteronuclear NMR experiments, paired with techniques that allowed for uniform ^{15}N and ^{13}C labeling, have led to the ability to fully assign the signals from most nuclei in a protein. The chemical shift information obtained during these experiments allows for structural analysis and the observation of dynamics or conformational changes caused by differing environment or addition of binding substrates.

3.2 An NMR assignment of ETV6 PNT domain

The ^{1}H-^{15}N HSQC spectrum is a 2D NMR spectrum that shows the correlated signals between directly bonded ^{1}H and ^{15}N nuclei. The spectrum typically contains the same number of amide signals as the number of non-proline residues in a protein. Additionally, depending upon sample conditions, the ^{1}H-^{15}N signals from Trp, Gln, Asn, and Arg, Lys and His side chains may appear in the ^{1}H-^{15}N HSQC spectrum. Assignment of the ^{1}H-^{15}N HSQC spectrum of a uniformly $^{13}C/^{15}N$-labeled protein is typically accomplished using complementary CBCA(CO)NNH and HNCACB spectra (Figure 8).

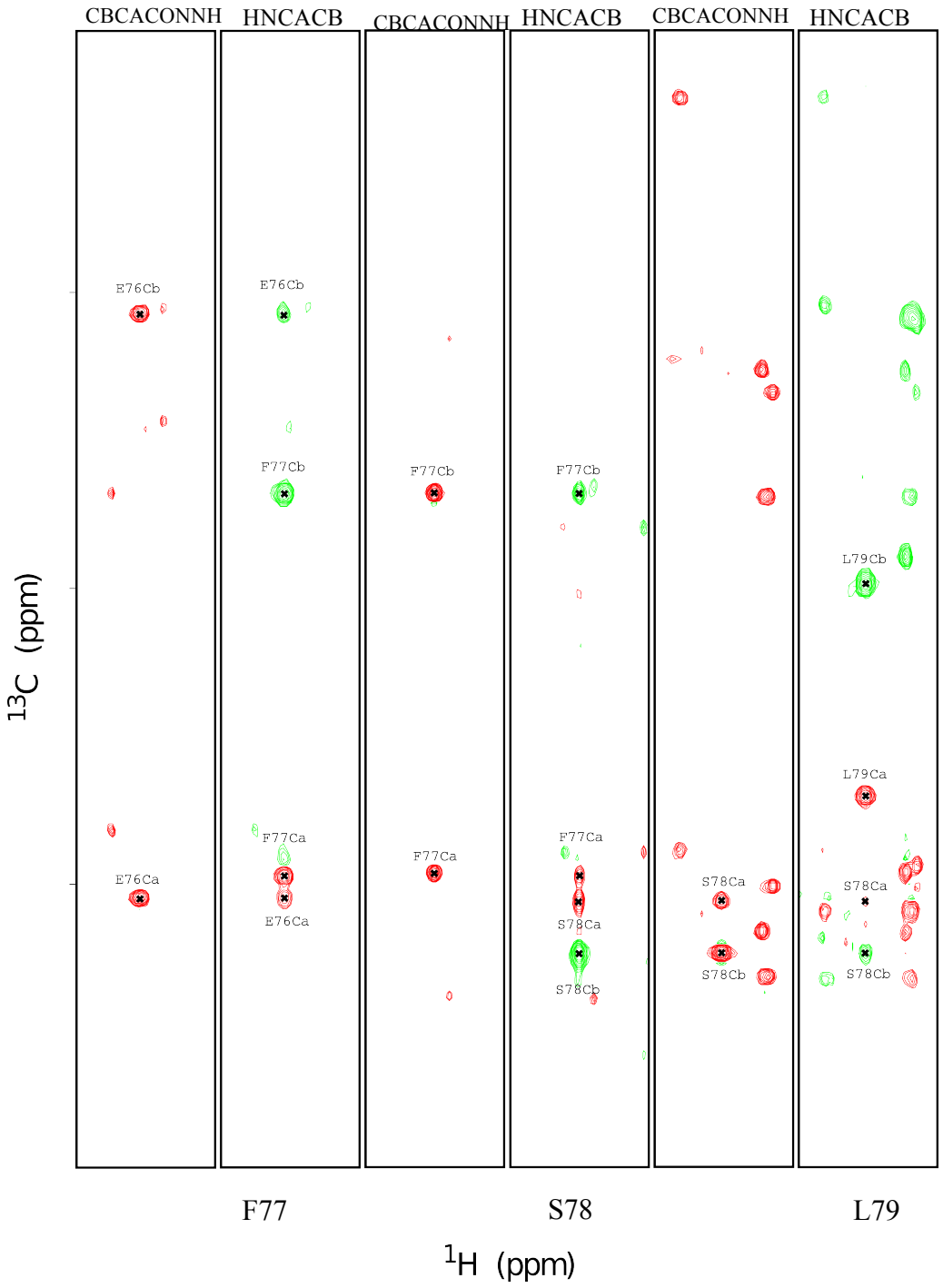


Figure 8. Strategy for assignment of protein NMR signals. Strips plots from HNCACB and CBCACONNH spectra are shown at the amide ^{15}N planes of the indicated adjacent residues (A93D-ETV6$^{43-125}$). The $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ are seen for each of the experiments. The HNCACB experiment has $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ shift information for both the i and i-1 residues, whereas the CBCACONNH shows only the inter-residue correlations. In addition, the $^{13}C_{\beta}$ (green) have opposite phases to the $^{13}C_{\alpha}$ (red) in the HNCACB spectra,

The CBCA(CO)NNH correlate the amide signals of residue i to the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ of the previous residue i-1. The HNCACB correlates the amide to both the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ of residues i and i-1. Additionally, in the latter spectrum $^{13}C_{\beta}$ signals are of opposite phase to those of the $^{13}C_{\alpha}$ residues. Combination of these two experiments coupled with knowledge of typical ^{13}C shifts for each amino acid type, the peaks of N90SA93D-ETV6$^{1-125}$ and A93D-ETV6$^{43-125}$ constructs were assigned. The HNCO and HNCACO experiments allow for assignment of carbonyl carbons on the protein backbone and thus helped confirm assignments obtained with the HNCACB and CBCA(CO)NNH spectra and resolve any ambiguities. Using these complementary experiments coupled with knowledge of typical ^{13}C shifts for each amino acid type the amide signals of N90SA93D-ETV6$^{1-125}$ and A93D$^{43-125}$ constructs were assigned.

3.2.1 An assignment of the backbone residue signals of a full-length construct

Initial constructs were made with a portion of ETV6 (residues 1-125) starting at its natural N-terminus. Although the SAM domain was expected to span residues Pro56-Gln123 we wanted to include the N-terminal sumoylation site, K11. Also, some ETS PNT domains contain a dynamic helix preceding the core helical bundle [30]. Thus we first wished to determine if the ETV6 SAM domain also had this additional helix and to characterize the N-terminal tail containing the sumoylation site Lys11.

Three-dimensional NMR experiments were run on a $^{13}C/^{15}N$ labeled N90SA93D-ETV6$^{1-125}$ sample. Crowding of peaks in the center of the spectra indicated that many residues in this construct had random coil conformations. Unfortunately this also made a full spectral assignment difficult. During the course of this assignment it was discovered that an unwanted point mutation was located at residue N90S. Furthermore the protein was prone to degradation and thus it was

difficult to obtain reproducible high quality samples and spectra. Despite issues with sample integrity, the spectra were assigned with certainty to ~2/3 completeness (Figure 9, Table 3).

During the assignment, the unfortunate N90S mutation and a cryptic thrombin site were discovered in the amino acid sequence. The thrombin cleavage site was found at residues V37-P38-R39-A40. Therefore, the N-terminal region was likely partially proteolyzed during the thrombin His_6-tag removal. However, some of the protein was believed to be intact as residues in this unstructured tail region were confidentially assigned. These assignments did allow for chemical shift-based secondary structure prediction to be run. Based on this analysis, the SAM domain helical boundaries were confirmed to be similar to those of the X-ray crystallographic structure. No additional N-terminal helices were discovered, and the residues preceding the SAM domain were shown to be disordered. This structure prediction analysis will be discussed later.

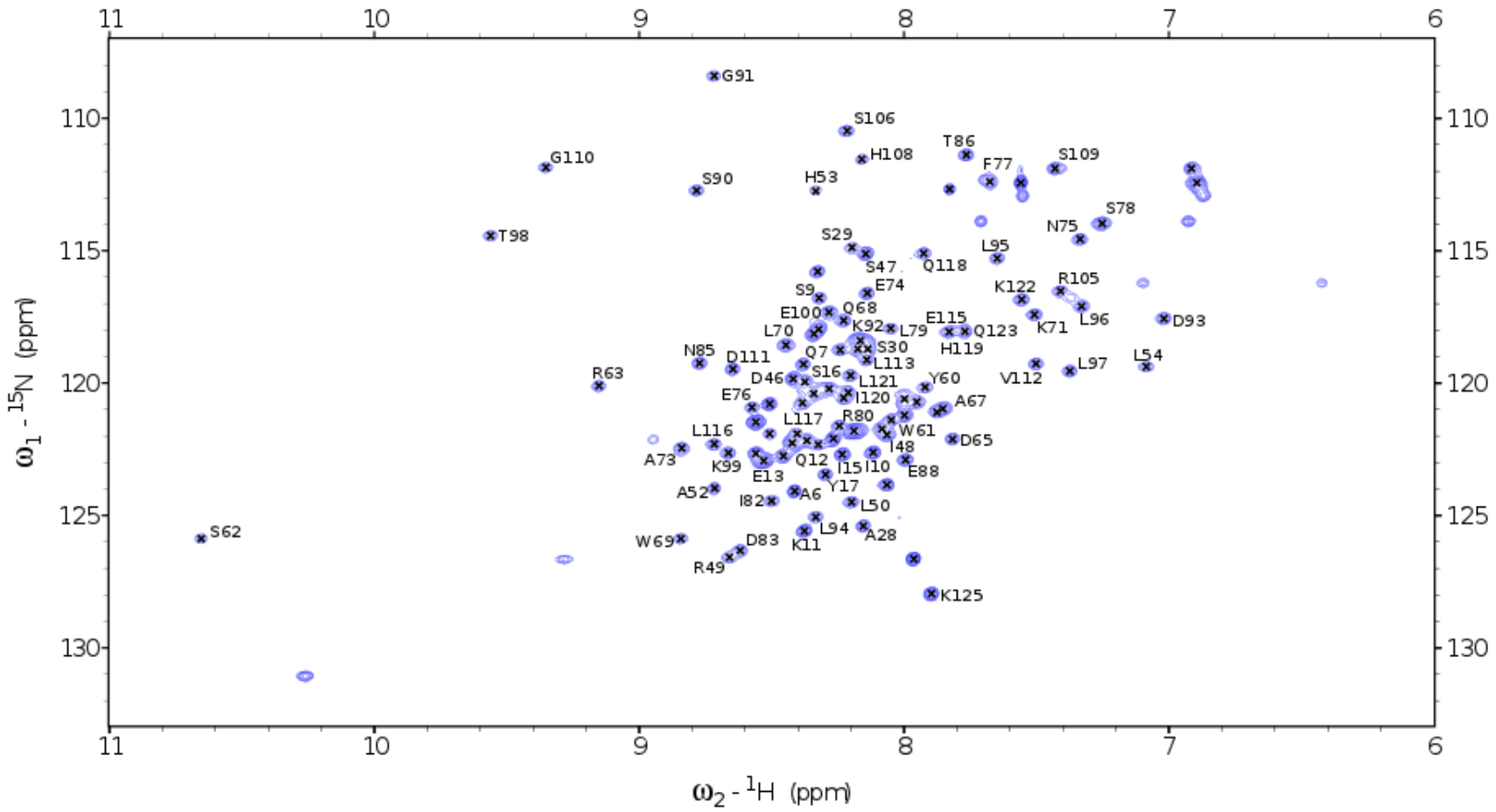


Figure 9. An assignment of the ^{1}H-^{15}N HSQC spectrum of the N90SA93DETV6$^{1-125}$ construct is shown. Full chemical shift value information can be found in Table 3. Well dispersed signals indicated folded protein.

3.2.2 A shortened construct with conformational heterogeneity

Ongoing issues with protein stability and sample integrity caused us to a resort to a more stable construct. We decided to remove the unstructured N-terminal residues and continue binding studies with a shortened construct, predicting that the removal of the unstructured tail would result in a more stable sample. The shortened construct (residues 43-125) with the N90S mutation corrected back to wild-type Asn90, was expressed and NMR experiments were run on the resulting $^{13}C/^{15}N$ labeled A93D-ETV6$^{43-125}$. Unexpectedly, spectra of the shorter constructs (residues 43-125) now showed a doubling of many amide peaks in the ^{1}H-^{15}N HSQC, indicating more than one form of the folded protein in solution. From assignment of the A93D-ETV6$^{43-125}$ (Figure 10) we are able to see that two isoforms of the protein exist in solution. Similar peak intensities suggested that their populations in solution were approximately equal.

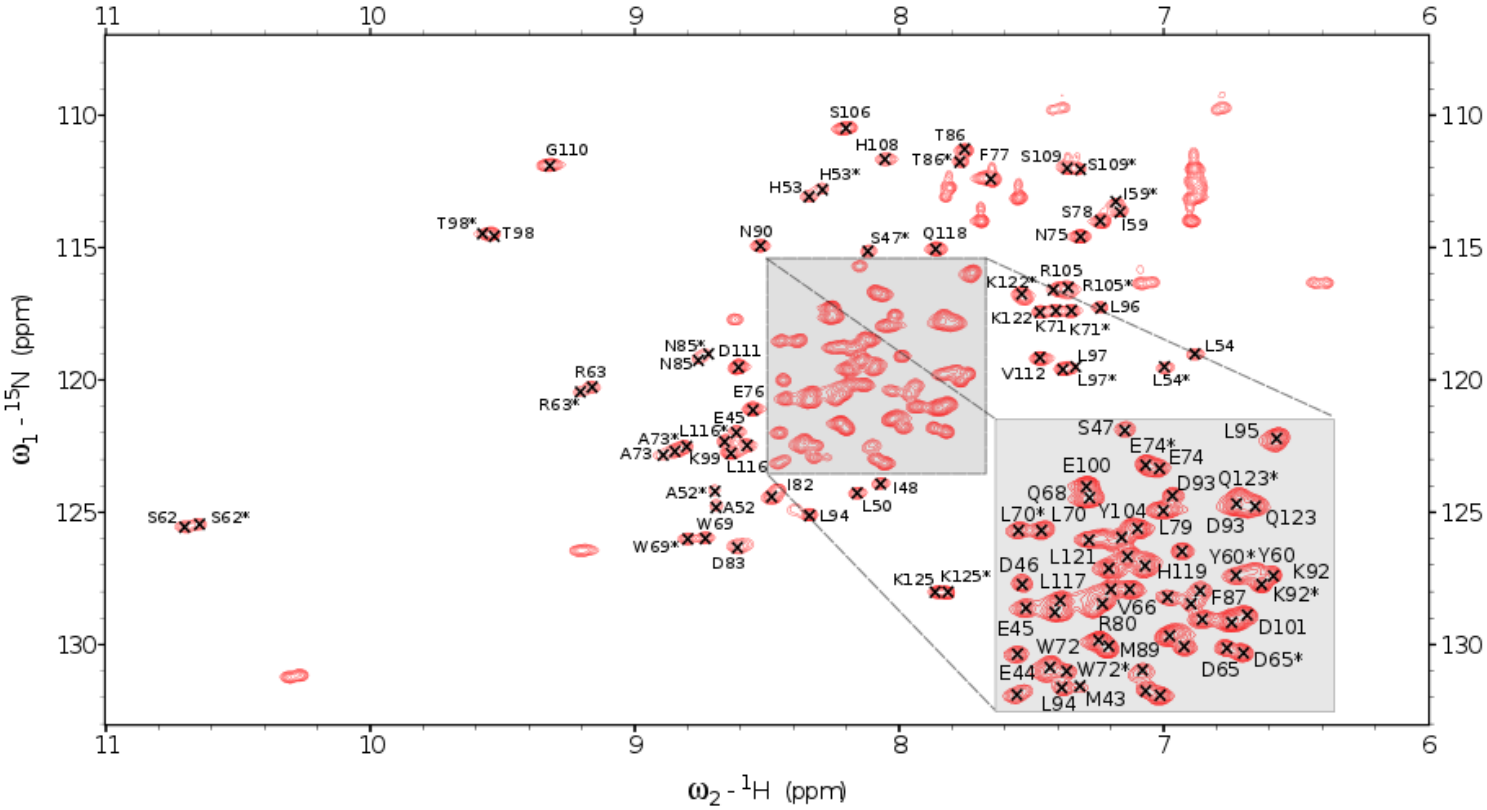


Figure 10. An assigned ^{1}H-^{15}N HSQC spectra for the shortened A93D-ETV6$^{43-125}$ construct. The crowded center of the spectrum is expanded in the grey inset. Note the spectral overlap and the doubling of many peaks.

Conformational exchange can yield distinct NMR behavior depending on its timescale and the chemical shift differences between the resulting conformations. In a simple 2-state equilibrium, fast exchange gives one peak with population-weighted chemical shift of the two forms of a protein. In contrast, two distinct signals arise with slow exchange. These two signals show the separate conformers of a protein, with information about structure from chemical shift values and population from the relative signal intensities.

One common source of multiple amide signals in ^{1}H-^{15}N- HSQC spectra results from the fact that, X-Pro groups can isomerize between cis and trans conformation of their peptide bond. This isomerization can provide a rate limiting kinetic barrier in protein folding [7, 14, 59] and because of its relatively high activation energy (~20 kcal/mol) [51] the interconversion between these two forms in a folded protein usually occurs on the seconds timescale or longer. However, because we observed numerous peaks of approximately equal intensity, this helps rule out the idea that the second set of peaks is caused by proline isomerization. That is, in a disordered polypeptide the cis isomer is usually populated to only ~10%. If either the cis or trans form offer a preferred folding arrangement for the protein, exchange may not occur. Also, if the isomerization has an effect on local but not the global, structure of the protein then only residues near the proline residue would have doubled peaks with similar ppm shifts, whereas those distal to the proline should yield single signals. This effect was not seen thus we concluded the doubling of peaks was not due to proline cis/trans isomerization.

Given the ability of the SAM domain to self-associate it is possible that A93D-ETV6$^{43-125}$ is polymerizing or dimerizing in solution despite the monomerizing mutations introduced at the binding interface. If slow exchange between self-association states were occurring, then the monomeric forms would have sharper lineshapes than the lineshapes of the oligomeric form.

However, based on the similar lineshape and signal intensity of the two sets of peaks, such self-association does not appear to be occurring. Furthermore doubled peaks were observed between pH 5.8-8.6 (Figure 11). Along with the fact that N90SA93D-ETV6$^{1-125}$ showed only single peaks, this further argues against self-association occurring.

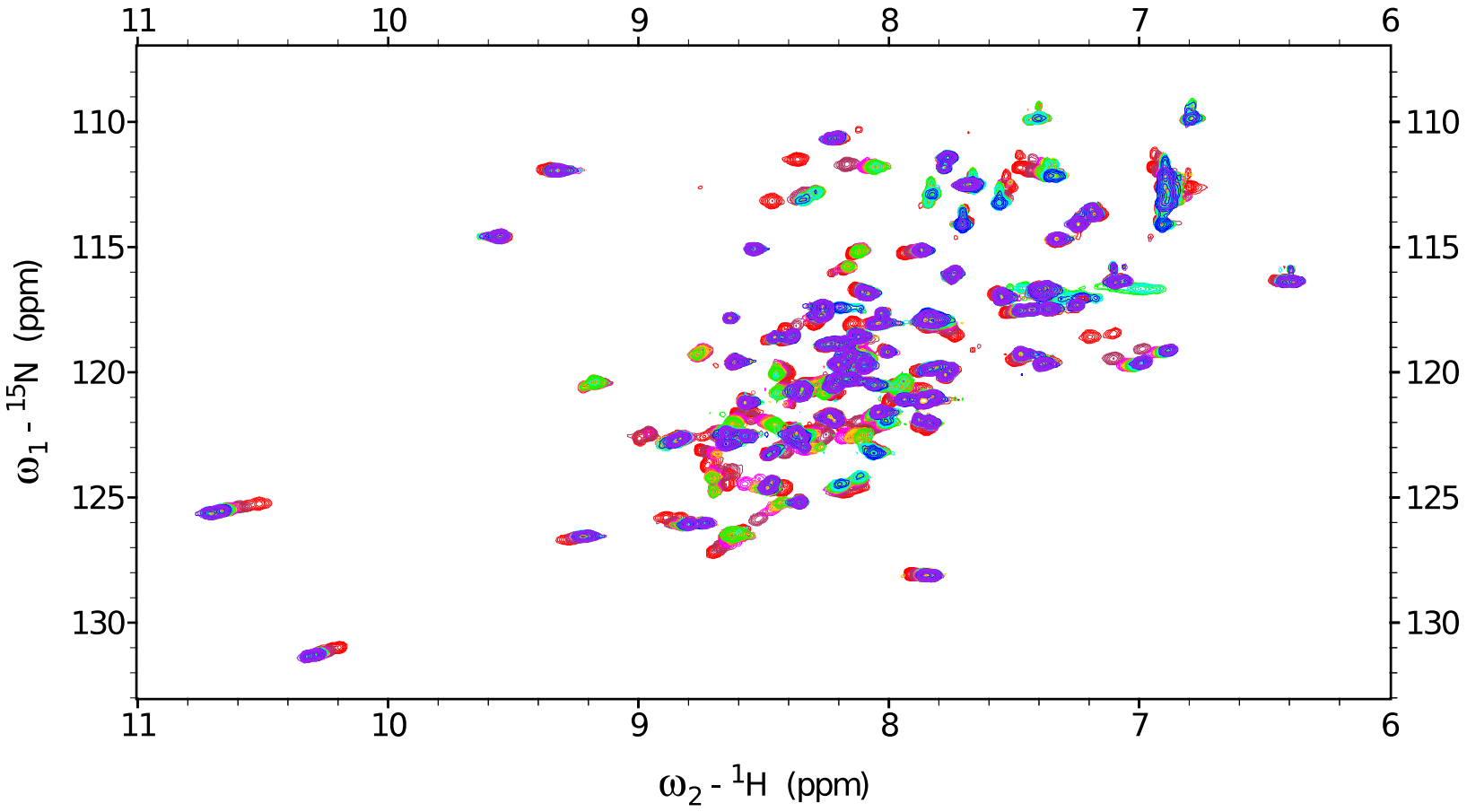


Figure 11. A ^{1}H-^{15}N HSQC monitored pH titration of A93D-ETV6$^{43-125}$. The pH of the sample was raised (pH 8.6, purple) and lowered (pH 5.8, red). Whereas some signals disappeared at higher pH due to rapid exchange with solvent, overall no large rearrangement of signals indicates that the protein structure remains constant over the pH range examined.

Another source of the doubled peaks could be issues with purification or degradation. However, over the course of 4 days (Figure 12) the spectra are indicating that it is unlikely that the samples are degrading. Mass spectrometry was performed on the A93D-ETV6$^{43-125}$ sample and only one species was present (Figure 13). This result supports the theory that these doubled peaks are not due to degradation of the protein. The doubling of peaks however could be due to deamination, which causes a mass change of only 1 Da. Deamidation is known to occur in proteins, particularly at asparagine-glycine (N-G) dipeptide to yield an aspartate-glycine (D-G) [43, 52]. The same deamidation reaction can also yield an isoaspartyl residue with its neighboring residue linked via a peptide bond formed with its γ-carboxyl. Indeed the Asn90 residue is followed by Gly91, making the site vulnerable to deamidation. Because the longer construct had the unfortunate point N90S mutation we hypothesized that deamidation led to peak doubling in the spectra of A93D-ETV6$^{1-125}$ but not N90SA93D-ETV6$^{43-125}$. If the deamidation led to an isoaspartyl group, characteristic patterns in HNCACB-type spectra would result [56]. Although such patterns were not observed in our spectra, the difficulty in full assignment and the presence of extra unassigned peaks did not allow us to confidently exclude deamidation via this evidence alone.

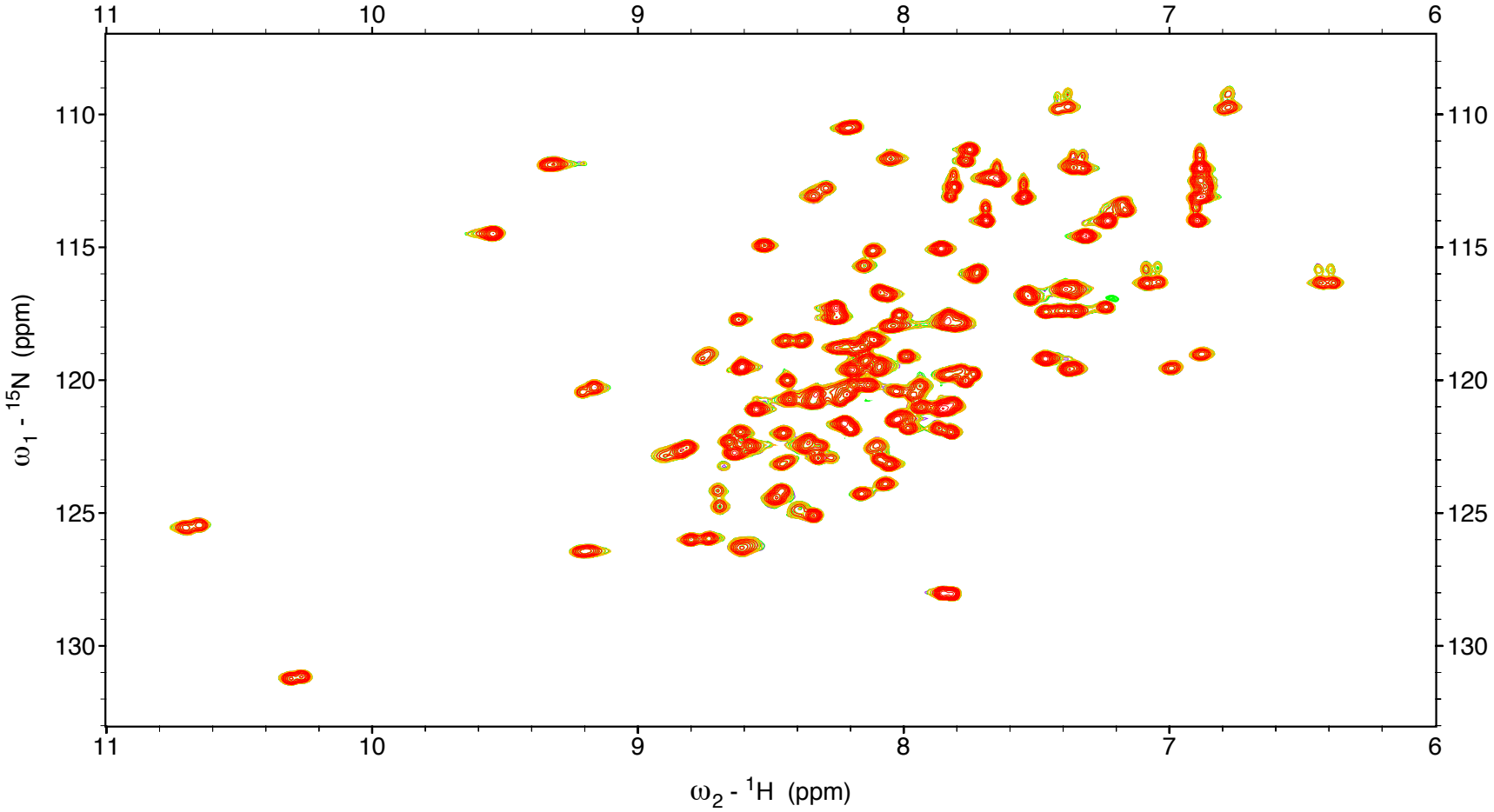


Figure 12. Overlaid ^{1}H-^{15}N HSQC spectra of A93D-ETV6$^{43-125}$ recorded as a function of time (purple to red). The significant overlap over the course of time showed that the sample did not degrade over a period of 4 days while 3D data were being collected.

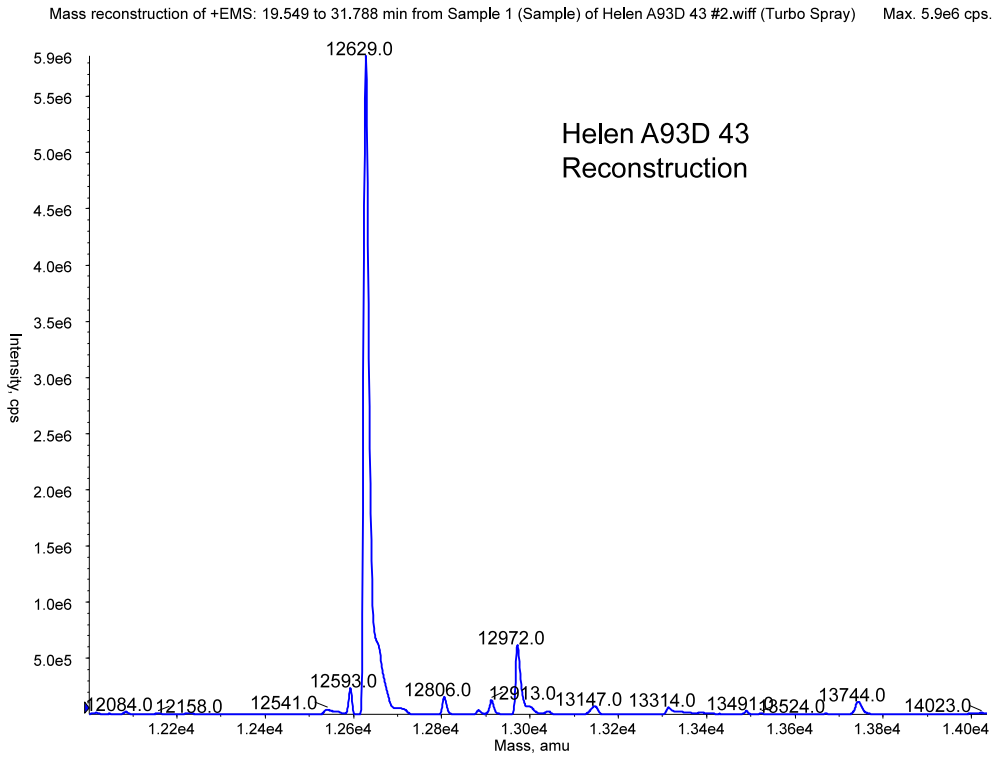


Figure 13. Reconstructed ESI mass spectrum of A93D-ETV6$^{43\text{-}125}$. The observed mass of 12629 Da closely matches the predicted mass of the construct, eliminating degradation as a source of doubled peaks in its NMR spectra.

Deamidation has been seen as a result of protein aging and degradation [43, 52] and is affected strongly by the C-flanking residue to the Asn/Asp-Gly pair in addition to the local flexibility of the region [43]. Deamidation results in both a small change in mass (1 Da) and a change in charge. While the mass difference is difficult to observe, in principle an isoelectric focusing gel could be used to detect if the charge of the protein has changed due to deamidation. However, we did not have ready access to such electrophoresis. Additionally this reaction is accelerated at higher pH. During a titration of the protein over a range of pH values, it was seen that the ratio of the two populations did not significantly change upon raising or lowering the pH. (Figure 11). Also deamidation is effectively irreversible and should proceed to completion rather

than approximately equal population of protein species. From Figure 12 we saw unchanging ^{1}H-^{15}N HSQC spectra which indicate that deamidation was not occurring to the sample over the course of 4 days.

Using site directed mutagenesis, the point mutation, N90S, was introduced in the shortened construct to eliminate the site of the potential deamidation reaction. A ^{1}H-^{15}N HSQC spectrum of the sample was collected (Figure 14), and the double peaks were also observed for the N90SA93D-ETV6$^{43-125}$ sample. Along with the above results, this allows us to conclude that deamidation of this sample is not occurring at the N90-G91 residues, and thus deamidation is not a source of the doubled peaks in the ^{1}H-^{15}N HSQC spectra of the shortened ETV6$^{43-125}$ constructs. Because these doubled peaks were not observed in the ^{1}H-^{15}N HSQC spectrum of the N90SA93D-ETV6$^{1-125}$ sample it is possible that the unstructured tail region (residues 1-42) play a role in selecting one conformer of the SAM-ETV6 domain over another. Determining the exact cause of the multiple forms of the SAM-ETV6$^{43-125}$ in solution will require additional studies.

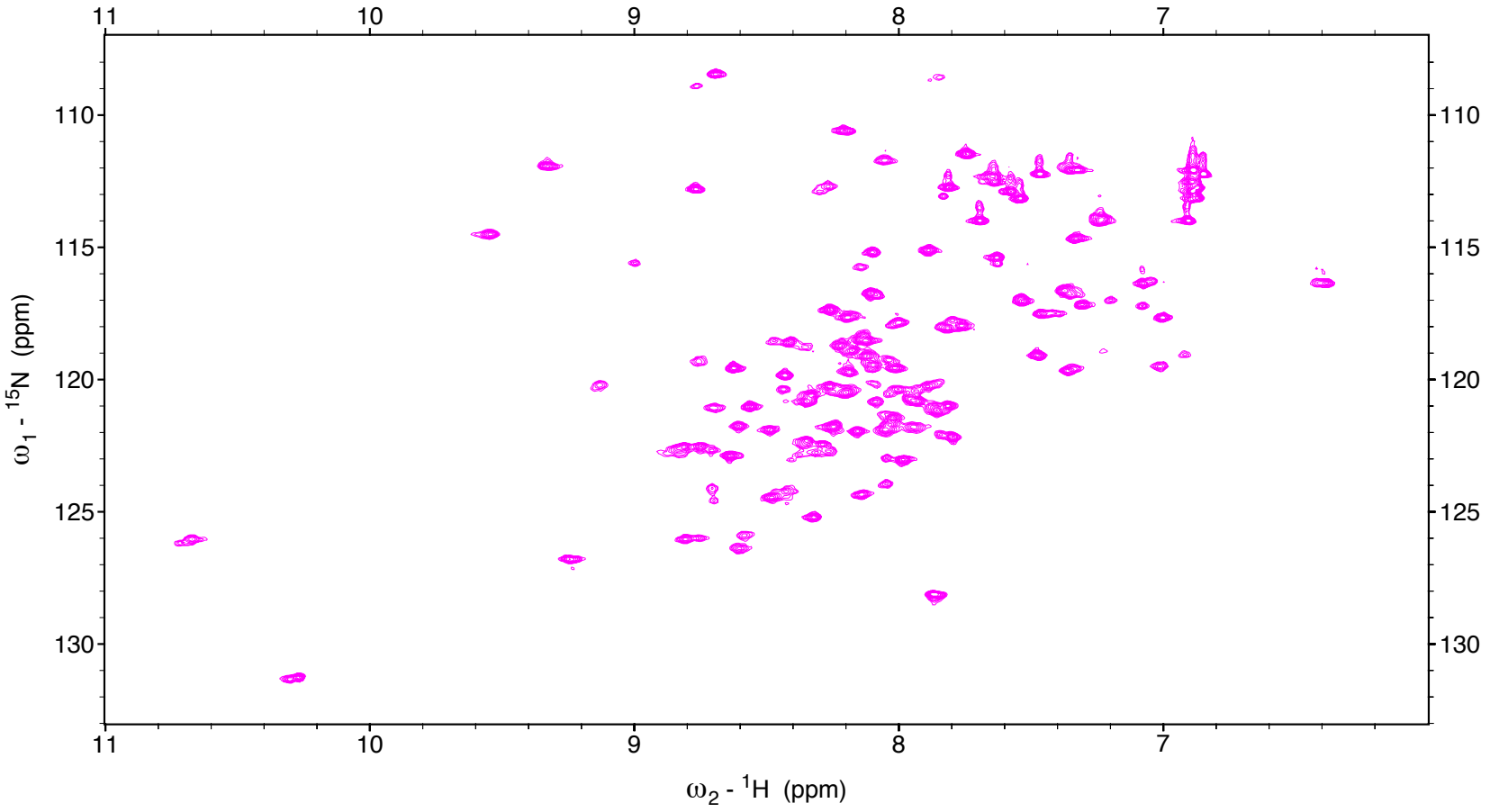


Figure 14. ^{1}H-^{15}N HSQC spectrum of N90SA93D-ETV6$^{43-125}$ construct. Again a similar pattern to the rest of the SAM-ETV6 constructs can be seen. This shortened construct (43-125) has a set of doubled peaks, eliminating deamidation of the N90-G91 site as the source of peak doubling in shortened constructs.

While the peak-doubling observed in the various ETV6$^{43-125}$ ^{1}H-^{15}N HSQC spectra made assignment and interpretation of spectra initially difficult it provided insight into the differences between the short and long constructs and we could eliminate several potential sources of the peak-doubling. From the mass-spectroscopy, collection of ^{1}H-^{15}N HSQC over several time-points, and pH titrations we could see that this doubling is not due to degradation of the protein over time. Additionally, the introduction of an N90S point mutation in the shortened construct coupled with the HNCACB spectra of the A93D-SAM-ETV6$^{43-125}$ allow us to conclude that the doubling of peaks is not due to the deamidation of residues N90-G91. Because the doubled peaks were not observed in the ^{1}H-^{15}N HSQC of the N90SA93D-SAM-ETV6$^{1-125}$ it is possible that the unstructured tail region (residues 1-42) play a role in selecting one confomer of the SAM-ETV6 domain over another. We can also conclude that mutiple forms of the SAM-ETV6$^{43-125}$ exist in solution and warrents future study in order to beter understand the SAM domain of ETV6.

3.2.3 Secondary structure analysis

Despite the challenges in interpreting and assigning the NMR spectra of the ETV6 constructs, relatively complete assignments were obtained. Using the SSP algorithm [34], secondary structural elements were predicted for the N90SA93D-ETV6$^{1-125}$ and A93D-ETV6$^{43-125}$ constructs based on the measured chemical shift values of their backbone nuclei (Figure 15, Figure 16). The observed $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shift differences were compared to those of a random coil polypeptide. Positive values correspond to alpha helices and negative values to beta strands. Despite the difficulty of assigning the double-peaked shortened constructs we were able to run secondary structure propensity on the samples using one set of the assigned amino acid residues.

From this approach it is clear that the characteristic SAM domain alpha helix bundle previously found in the ETV6-SAM domain crystal structure [24] is present in solution in both the full and shortened constructs of ETV6. Furthermore, no additional helices (H0 and H1) as found for the SAM domain of Ets1 are present. Rather, residues N-terminal to the SAM domain are disordered with random coil chemical shifts. While SSP scores were only obtained for the A93D variants of the long (1-125) and short (43-125) constructs, when comparing the ^{1}H-^{15}N HSQC's of the various shortened constructs, we see only small shift perturbations located in areas around the point mutants. This overlap leads us to conclude that these various point mutants of the SAM-ETV6 domain retain the same structure as one another and thus have similar helical boundaries.

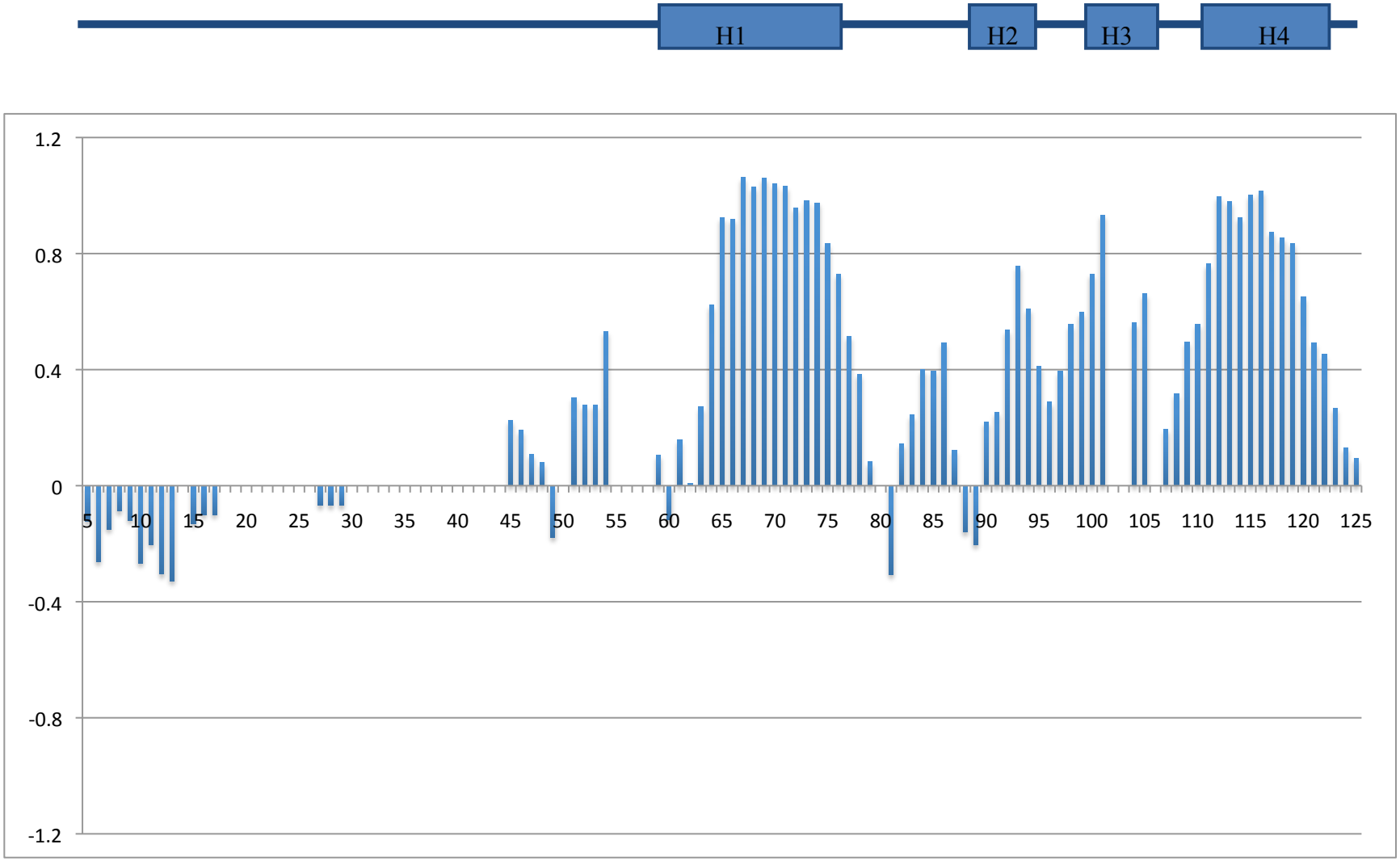


Figure 15. SSP scores for the N90SA93D-ETV6$^{1-125}$ construct. These values were calculated using chemical shifts from $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ assignments. SSP score (y-axis) is plotted against the residue number (x-axis). Helical boundaries from the crystal structure (1LKY.pdb) are shown for comparison Positive values correspond to α-helices and negative values to β-strands.

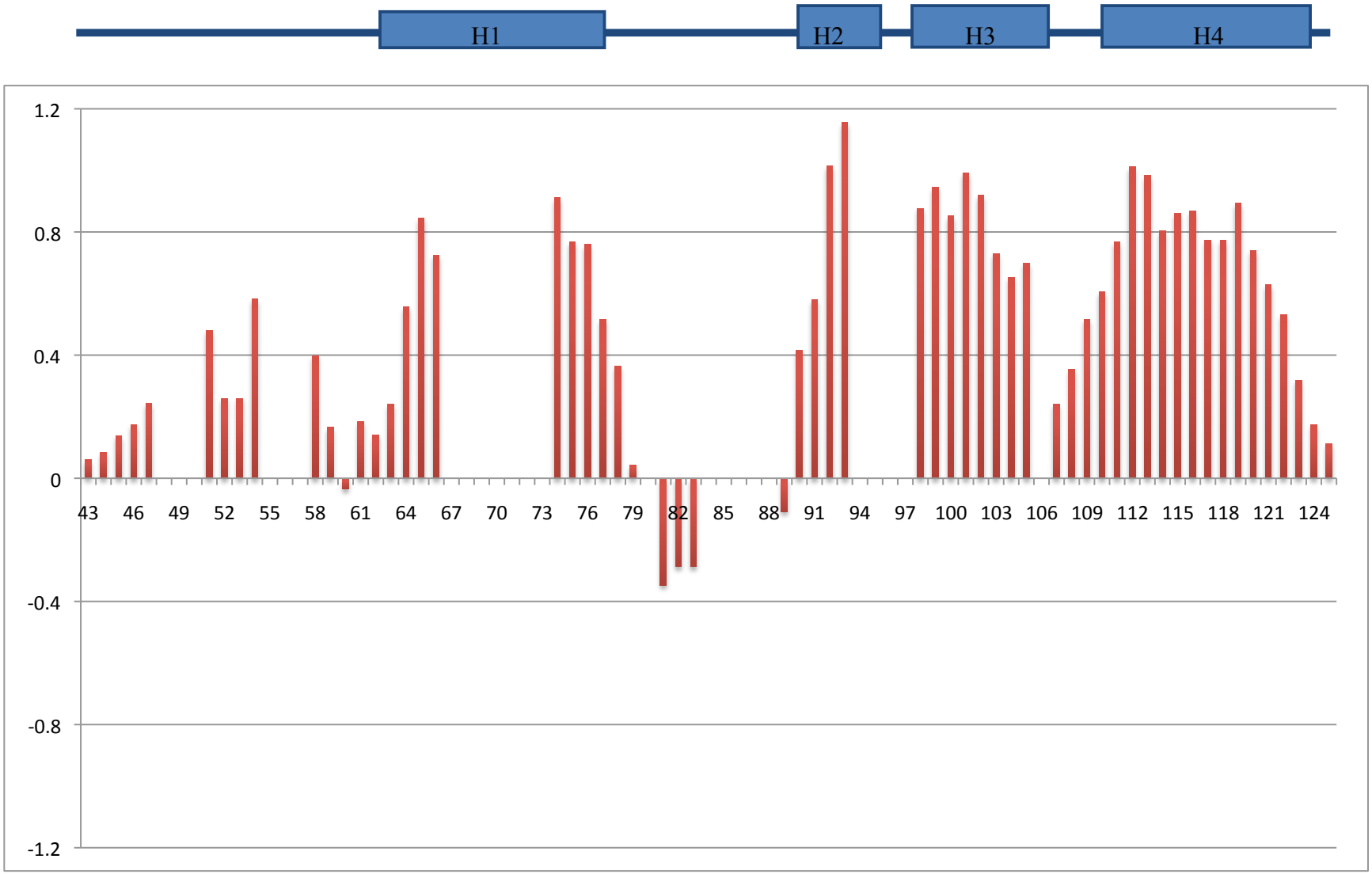


Figure 16. Secondary structure SSP score for the A93D-ETV6$^{43-125}$ construct. SSP score (y-axis) is plotted against the residue number (x-axis). Helical boundaries from the ETV6 SAM domain crystal structure (1LKY.pdb) are shown for comparison. Positive values correspond to α-helices and negative values to β-strands.

3.3 Comparing the various monomeric point mutants by NMR

^{1}H-^{15}N HSQC spectra were obtained for ^{15}N-labeled ETV6$^{43-125}$ constructs with the mutations, A93D, A93DK99R, V112E, and N90SA93D. Comparison of these spectra by overlaying the data, allows us to see that while there are subtle changes to the chemical shifts of certain residues, overall there are no major peak shifts (Figure 17, Figure 18, Figure 19). This overlap indicated that the structural core of the ETV6-PNT domain remains unchanged. The subtle changes seen for some residues could be due to small local changes from the addition of point mutations. Because a full assignment requires a ^{13}C/^{15}N-double-labeled sample, these were was not done on each point mutation and the doubled peaks crowd the center region of the spectra, the chemical shift perturbations of specific residues are difficult to accurately determine.

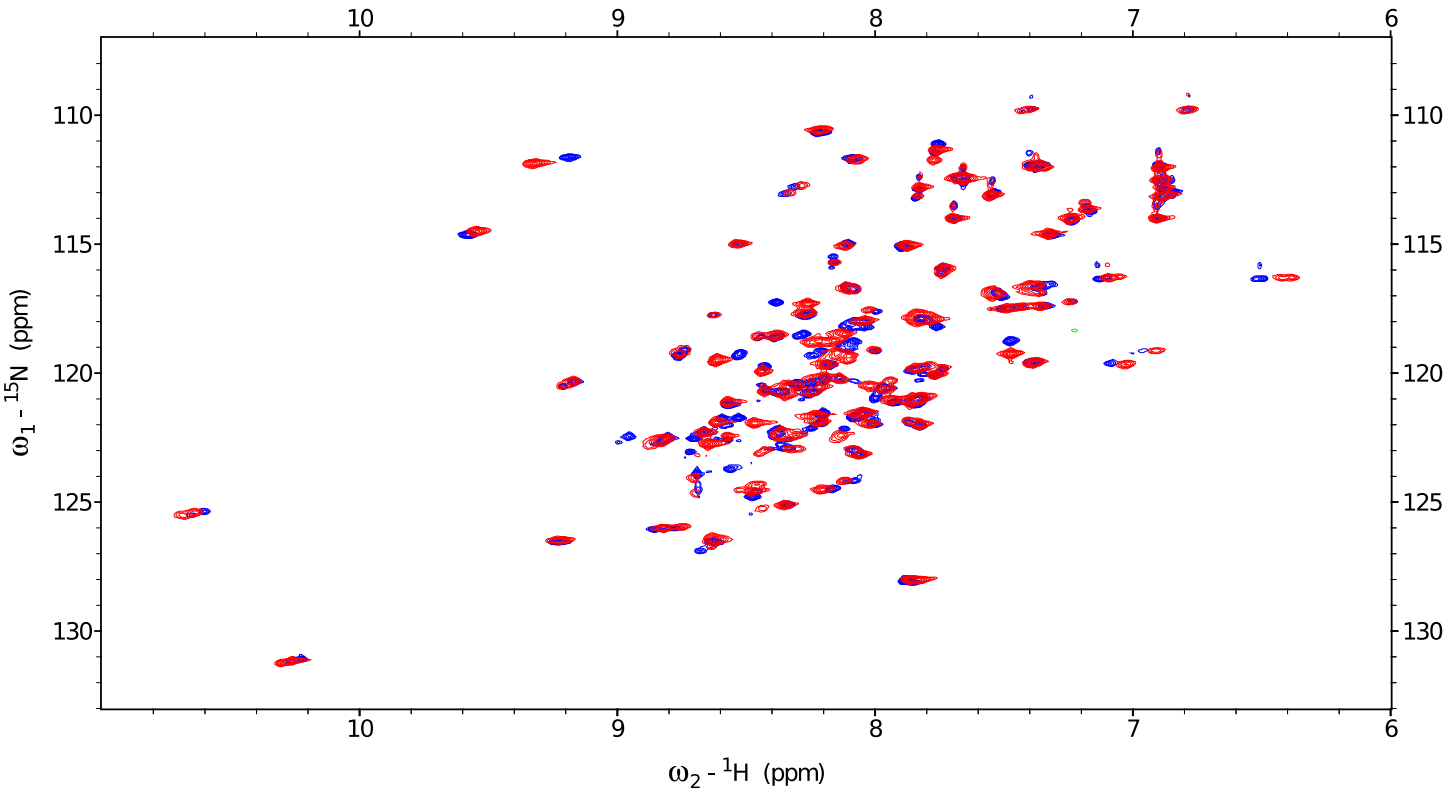


Figure 17. A93D-ETV6$^{43\text{-}125}$ construct (red) overlaid with A93DK99R-ETV6$^{43\text{-}125}$ construct (blue). Significant spectral overlap is seen between the two constructs, indicating minimal structural perturbation due to the K99R mutation.

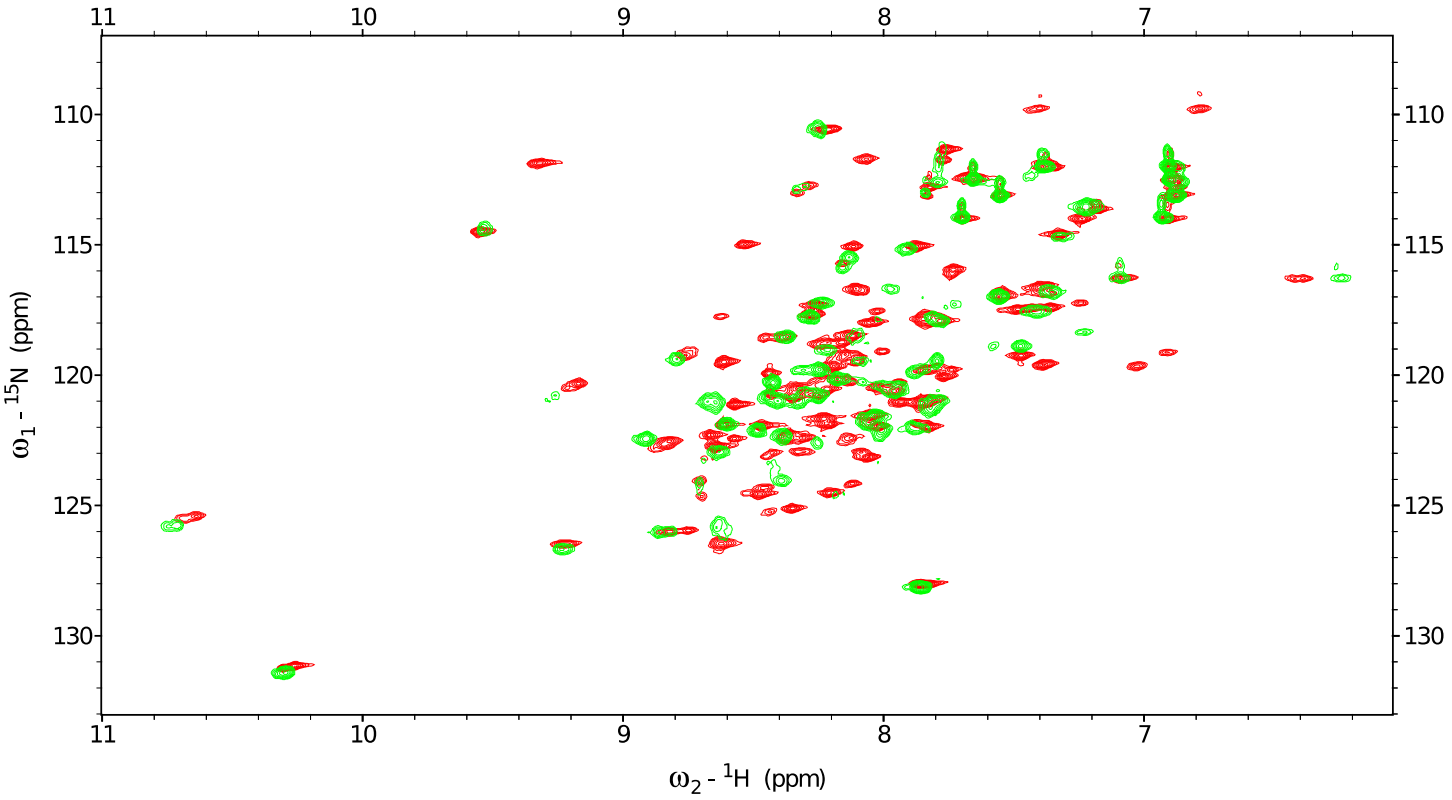


Figure 18. A93D-ETV6$^{43\text{-}125}$ construct (red) overlaid with V112E-ETV6$^{43\text{-}125}$ construct (green). Significant overlap is also seen between the two constructs. However several peaks are not seen for the V112E-ETV6$^{43\text{-}125}$ construct as its solubility was lower than that of any of the A93D-ETV6$^{43\text{-}125}$ constructs

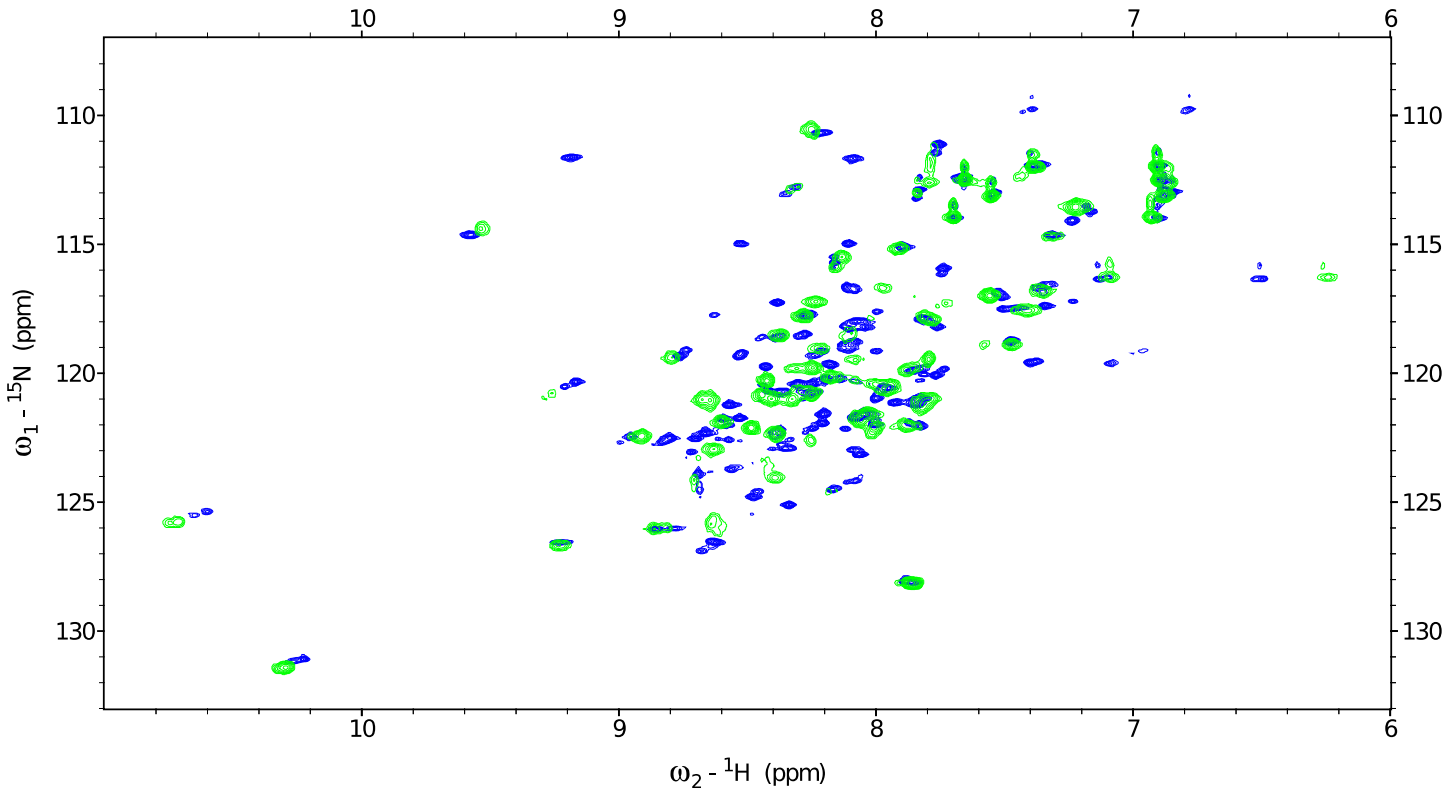


Figure 19. A93DK99R-ETV6$^{43-125}$ construct (blue) overlaid with V112E-ETV6$^{43-125}$ construct (green). Again overlap is seen between the two constructs. However several peaks are not seen for the V112E-ETV6$^{43-125}$ construct as its solubility was significantly lower than that of any of the A93D-ETV6$^{43-125}$ constructs

3.4 An NMR titration of ETV6 SAM domain monomers

^{1}H-^{15}N HSQC spectra were obtained for the ^{15}N-labeled construct V112E-ETV6$^{43-125}$. To this sample either concentrated unlabeled A93D-ETV6$^{43-125}$ or A93DK99R-ETV6$^{43-125}$ was added in aliquots to a final molar ratio of 1:1 (Figure 20, Figure 21). Prior to the titration samples were dialyzed overnight into the identical buffer, and each titration point was monitored with the collection of a ^{1}H-^{15}N HSQC spectra. During the course of these titrations, many amides in ^{15}N-labeled V112E-ETV6$^{43-125}$ showed chemical shifts perturbations. Furthermore, over the course of the titration, these shifts demonstrated that there was a binding interaction between the two SAM domains. Additionally binding occurred in the slow exchange regime as several peaks disappeared and new peaks appeared over the course of the titration. This slow exchange behavior indicated that the two SAM domains bound tightly. However, we were unable to determine the binding constants governing the interaction by NMR methods and sought to characterize the binding event using isothermal titration calorimetry.

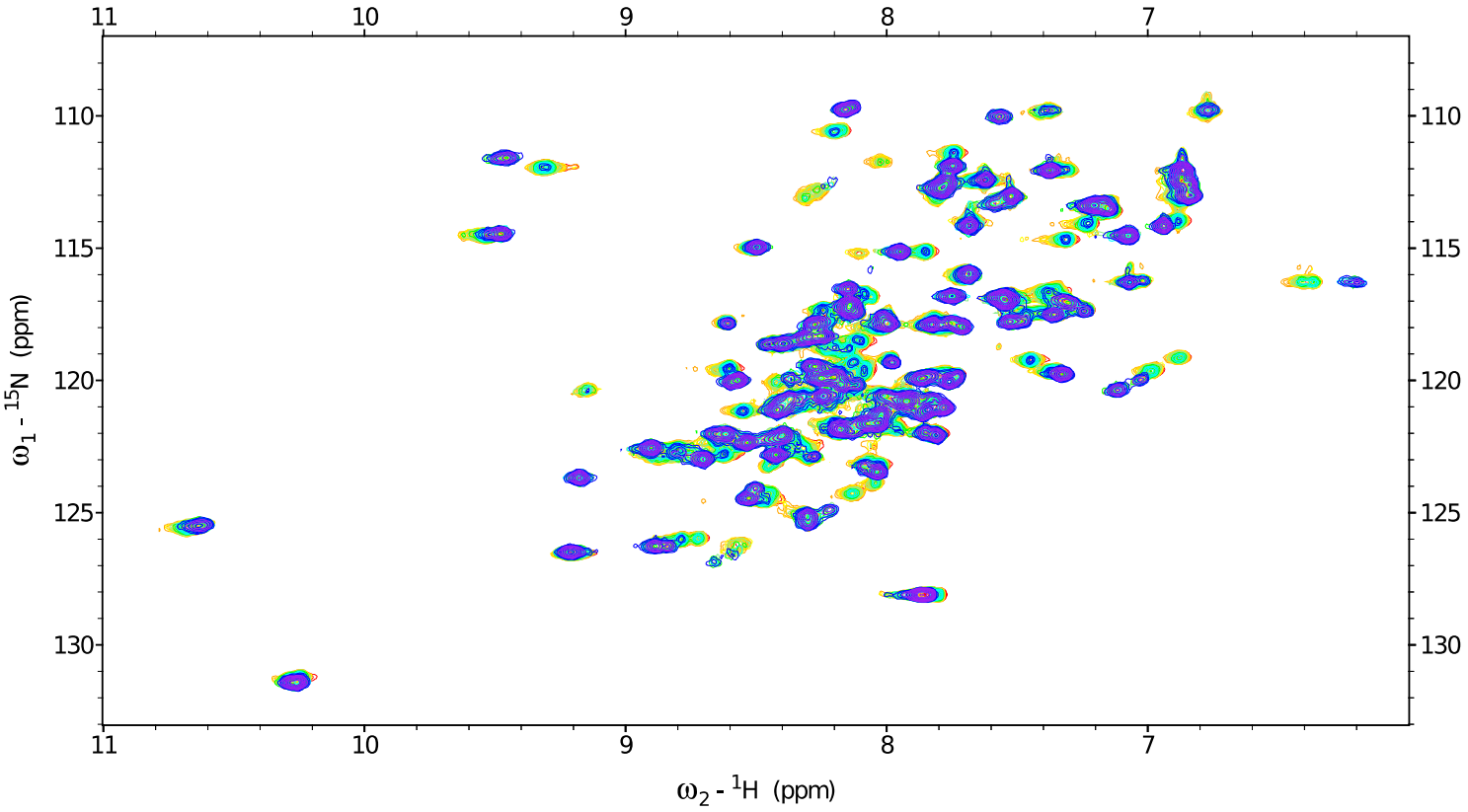


Figure 20. A titration of unlabeled A93D-ETV6$^{43-125}$ into ^{15}N-labeled V112E-ETV6$^{43-125}$. The superimposed spectra were recorded during addition from red (molar ratio 0:1) to purple (molar ratio 1:1) with peaks visibly shifting and several disappearing with complex formation. The spectral perturbations demonstrate that the two proteins bind in the slow exchange regime. Note that the doubling of peaks is not affected by the binding of the two monomers.

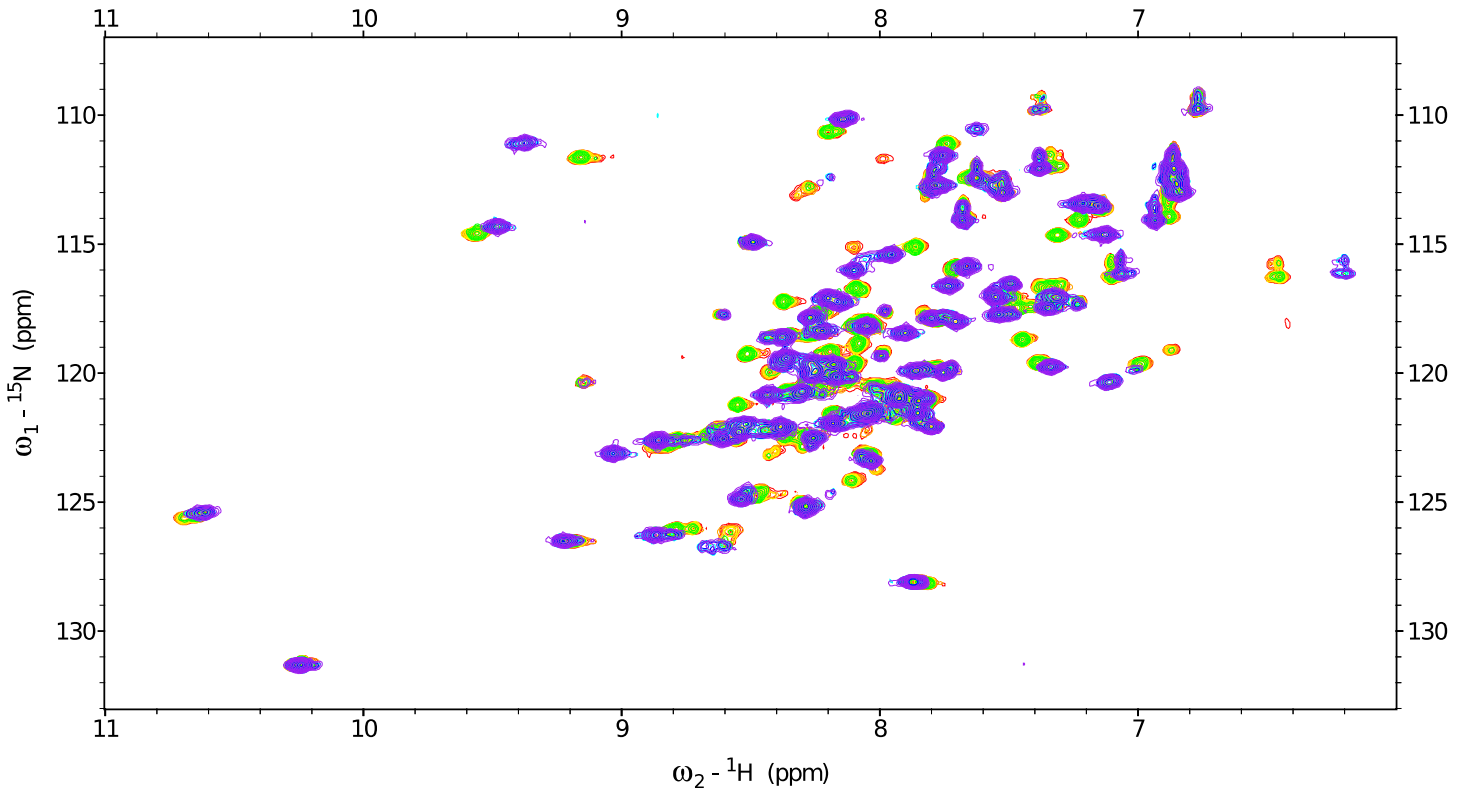


Figure 21. A titration of unlabeled A93DK99R-ETV6$^{43-125}$ into ^{15}N-labeled V112E-ETV6$^{43-125}$. Similar to the titration of the A93D-ETV6$^{43-125}$ mutant into the V112E-ETV6$^{43-125}$, the titration goes from red (0:1 molar ratio) to purple (1:1 molar ratio) with increasing additions of the A93DK99R construct. Again some peaks visibly shift over the titration while others disappear, and presumably reappear elsewhere. These spectral perturbations demonstrate that the interaction between the two domains occurs in the slow exchange regime. Note that the doubling of peaks does not appear to be affected by the titration and binding of the two monomers with one another.

Chapter 4 Binding studies of the SAM-ETV6 domain using ITC

4.1 Introduction to ITC

Isothermal titration calorimetry (ITC) provides details of the binding energetics governing the interaction between a titrant and an analyte. This technique is widely used to measure biological interactions involving proteins, nuclei acids, lipids, carbohydrates and small molecule metabolites or drugs [22, 41, 58]. The ITC instrument is composed of small coin-shaped sample and reference cells that are enclosed in an adiabatic shield and maintained at the same temperature [58]. A feedback system measures any difference in temperature between the two cells and keeps this difference as close to zero throughout the course of a titration experiment.

For an ITC an experiment, the titrant is placed in the syringe and the analyte within the sample cell. The motorized syringe injects small aliquots of the titrant into the sample cell, which is stirred rapidly to ensure fast mixing. If the two reactants interact, such as forming a non-covalent complex, there is most likely an accompanying release or absorption of heat (exothermic or endothermic reaction, respectively). This release or absorption of heat causes the temperature of the sample cell to transiently change from that of the reference cell. The feedback system either raises or lowers the thermal power required to return the cell to the same temperature as the reference. Integrating the resulting peaks in a power versus time plot gives the total amount of heat associated with each injection and subsequent binding interaction [22]. As the titration continues, this amount of heat progressively decreases in magnitude with the saturation of the reactant in the cell and eventually plateaus at a typically small baseline value associated with dilution effects. The latter can be corrected via control experiments. Commonly, the raw data is re-plotted in terms of heat change per mole of titrant injected versus the molar

ratio of titrant to analyte, and fit to an appropriate binding model. This fitting yields directly the enthalpy change (ΔH) upon binding, as well as the equilibrium association constant (K_a) and stoichiometry (n) for the given interaction. [22] From the standard thermodynamic relationships $\Delta G = \Delta H - T\Delta S$ and $\Delta G^o = -RT\ln K_a$, the accompanying entropy change (ΔS) can be calculated.

4.2 Mutation of Lys99 affects binding of monomers

Monomeric samples of ETV6 were titrated by ITC and their thermodynamic constants were measured along with the stoichiometry and binding constant of the interaction (Figure 22, Figure 23, Figure 24, Table 2). From these experiments we can reach several conclusions regarding the ETV6 SAM domain interactions. From the thermodynamic information, we see that each of the three binding reactions are exothermic. Furthermore, the favorable ΔS values for these reactions are indicative of hydrophobic interactions between the two monomers. A major contributor to stability of a protein is the hydrophobic effect which brings together apolar residues and allows for the entropically favorable release of caged water [6]. This conclusion is supported by the crystal structure, which has shown shared hydrophobic surfaces at the binding interface. The stoichiometries (n) close to 1 allow us to conclude that upon mixing, the monomeric SAM domains form heterodimers.

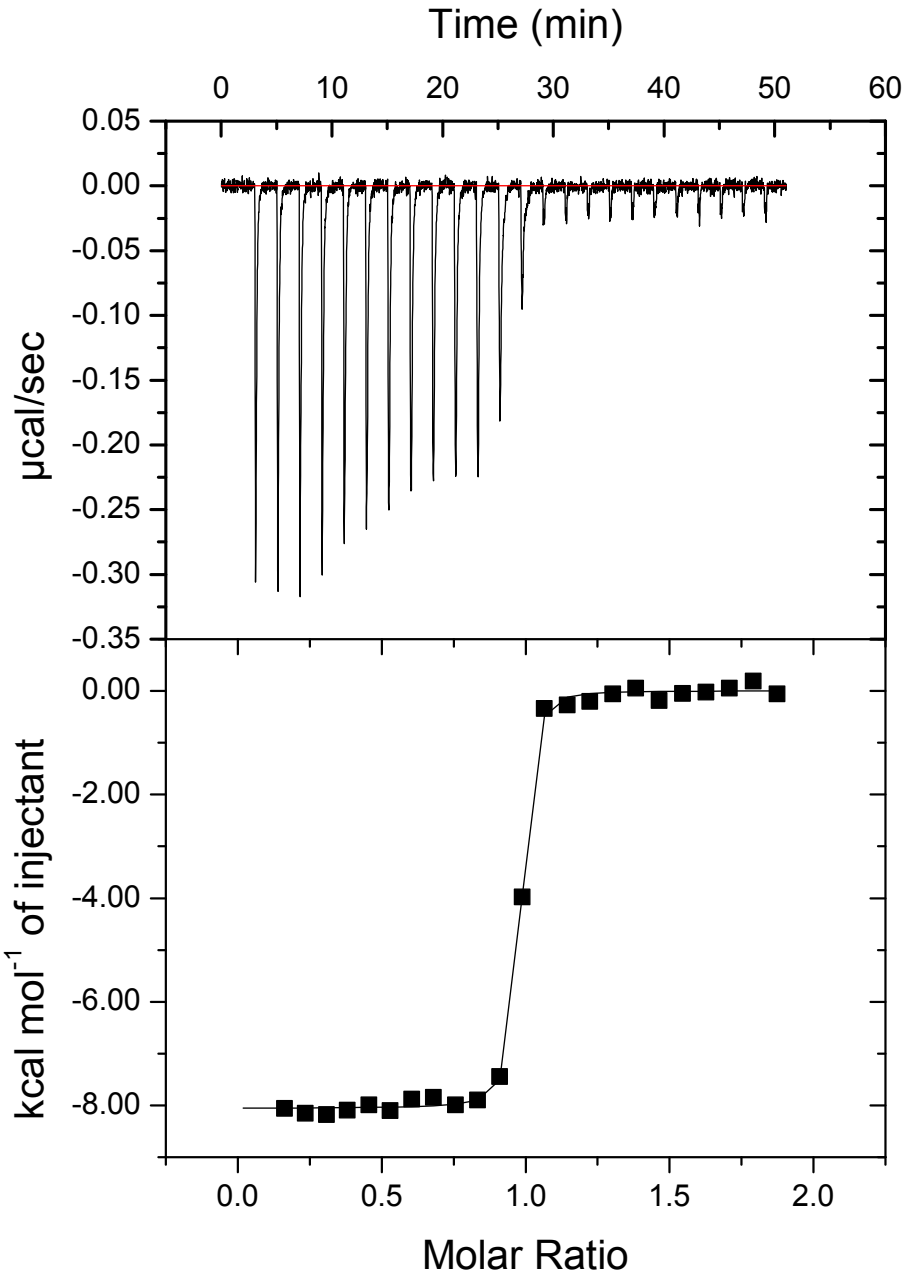


Figure 22. An ITC titration of A93D-ETV6$^{43-125}$ into V112E-ETV6$^{43-125}$. The power required to maintain a constant temperature in the sample cell is plotted against time in the top panel. The heat evolved per mole of A93D-ETV6$^{43-125}$ added is plotted in the lower panel. Fitting yielded the data summarized in Table 2.

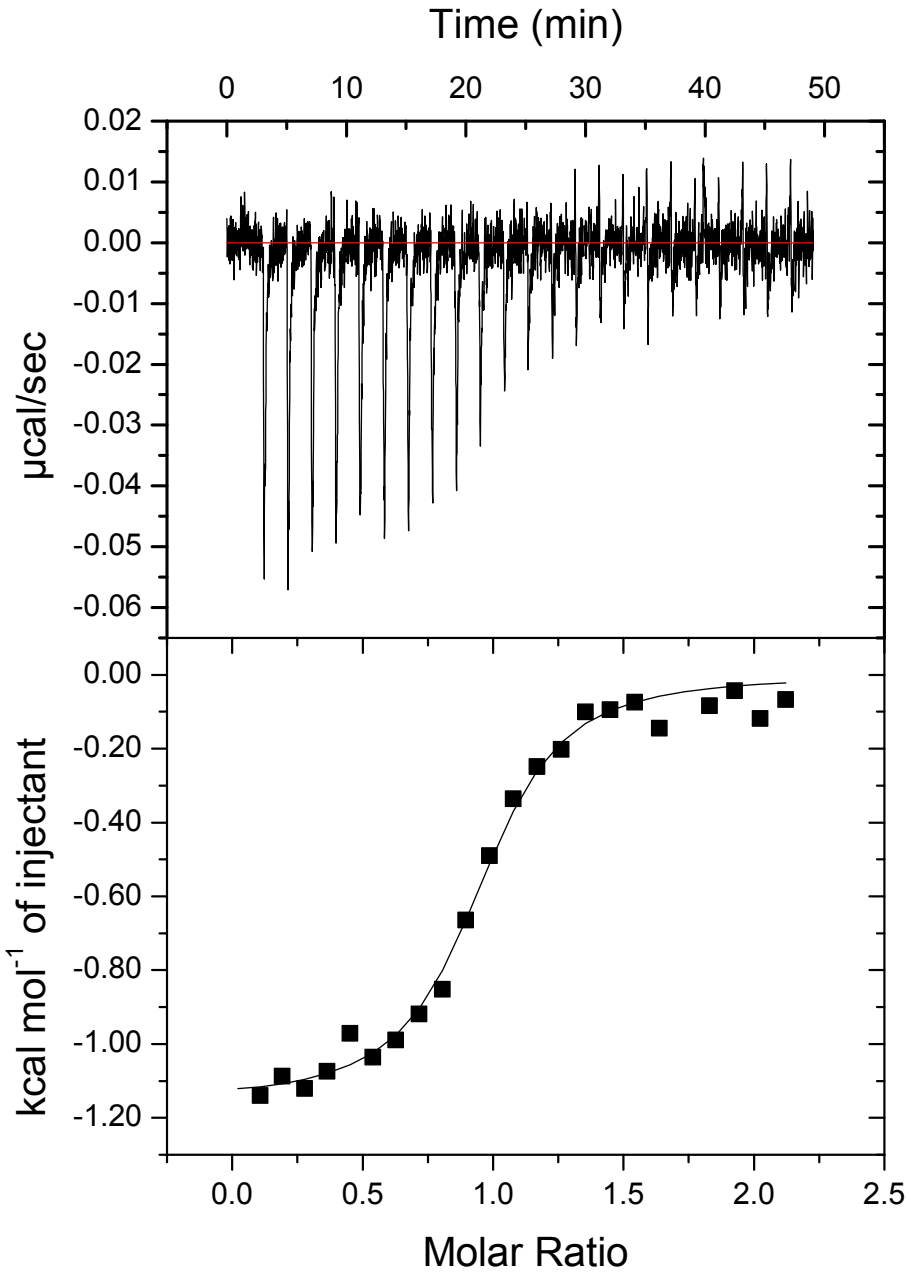


Figure 23. An ITC titration of A93DK99R-ETV6$^{43-125}$ into V112E-ETV6$^{43-125}$. The power required to maintain a constant temperature in the sample cell is plotted against time in the top panel. The heat evolved per mole of A93DK99R-ETV6$^{43-125}$ added is plotted in the lower panel. Fitting yielded the data summarized in Table 2.

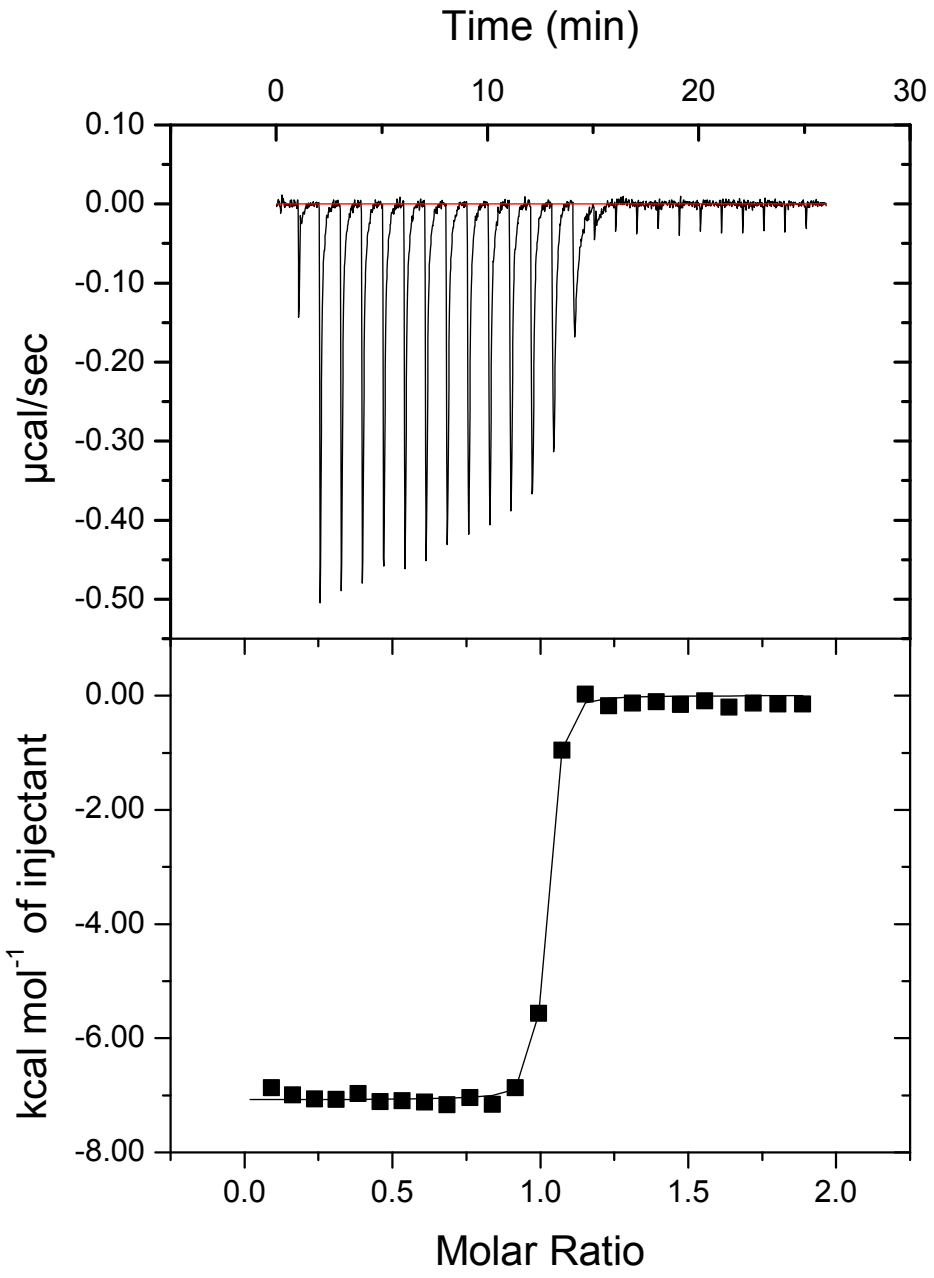


Figure 24. An ITC titration of A93D-ETV6$^{43-125}$ into V112EK99R-ETV6$^{43-125}$.The power required to maintain a constant temperature in the sample cell is plotted against time in the top panel. The heat evolved per mole of A93D-ETV6$^{43-125}$ added is plotted in the lower panel. Fitting yielded the data summarized in Table 2.

Table 2. Binding thermodynamic parameters determination from ITC measurements

Protein in Cell	Protein in Syringe	n	K_d (M^{-1})	ΔH (kcal/mol)	ΔS (cal/mol-K)
A93D	V112E	0.90 ± 0.03	(4.4 ± 2.2) x10^{-9}	-8.4 ± 0.2	10 ± 1.0
A93D	V112E K99R	0.94 ± 0.05	(6.4 ± 1.6) x10^{-9}	-7.3 ± 0.3	13 ± 0.8
A93D K99R	V112E	0.98 ± 0.19	(1.9 ± 1.2) x10^{-6}	-2.0 ± 1.0	19 ± 5

Most importantly the ITC experiments yielded a low dissociation constant K_d ($1/K_a$) value of 4.4 nM for the dimerization of A93D-ETV6$^{43-125}$ and V112E-ETV6$^{43-125}$. These results are entirely consistent with previously reported Surface Plasmon Resonance (SPR) data [24] yielding a dissociation constant of 1.7 ± 0.5 nM for the dimerization of A93D-ETV6 and V112E-ETV6. Our result shows a similarly low K_d of 4.4 nM for this pairing. Also, the dimerization of A93D-ETV6$^{43-125}$ and V112EK99R-ETV6$^{43-125}$ has a similar K_d of 6.6 nM. This similarity of this interaction, for which the mutated Lys99 is not at the resulting dimer interface, supports the conclusion that K99R point mutation does not disrupt the structure of the SAM-ETV6 domain.

In striking contrast, the K_d of the interaction of the A93DK99R-ETV6$^{43-125}$ with V112E-ETV6$^{43-125}$ was 1900 nM. Also, ΔH was substantially smaller in magnitude than observed for the wild-type pairings. Therefore, we conclude that the K99R point mutation dramatically affects the binding thermodynamics of the SAM domain interaction when present at the dimer interface. The results of these ITC experiments show that the K99R point mutation substantially weakens the interaction of the ETV6 SAM domains when present at the binding interface. This diminished binding provides an explanation for the loss of EN-mediated cellular transformation observed by our collaborators (Figure 6).

The participation of residue Lys99 in a salt bridge could be the reason for this large enthalpic change. Salt bridges are bonds that form between oppositely charged residues which are close enough together to experience electrostatic attraction [6]. Salt bridges can contribute to

protein structure and specificity in interactions with biomolecules [6]. Electrostatic effects can have a highly variable effect on protein folding and interaction attraction or repulsion of charged residues requires ordering and charges should be desolvated to interact [6].

Chapter 5 Concluding remarks and future directions

5.1 Goals of this thesis

The overall goal of this project was to better understand the mechanism by which a K99R mutation affects the self-association of the N-terminal SAM domain of ETV6. Mutation of residue Lys99 in the chimeric EN protein has been shown to interrupt the Ras-dependent transformation that results in aberrant growth (Figure 6) [8] and our hypothesis was that this resulted from disrupting the binding interface of the SAM domain. Consistent with this hypothesis, the K99R point mutant did not disrupt the structure of the SAM domain, and weakened the K_d value for dimerization by two orders of magnitude.

In my work I faced several obstacles in producing samples of soluble concentrated monomeric protein. I first cloned a long construct, N90SA93D-ETV6$^{1-125}$ but difficulties in solubility and longevity of the sample and the discovery of an unfortunate N90S residue made further investigation using constructs of this length problematic. Nevertheless, I was able to assign ~2/3 of the amino acid residues with certainty. Using these chemical shift assignments we established that this construct in solution has a helical secondary structure with helical boundaries that match those of the published crystal structure [55]. Also, the N-terminal sequence of this construct is intrinsically disordered with no additional helices such as those found in other ETS-family SAM domains. This disordered tail region also shows that residue Lys11 is conformationally dynamic and accessible for sumoylation.

Truncated constructs lacking residues 1-42 were generated and used for the remainder of my studies. ^{15}N-labeled versions of these constructs (A93D-ETV6$^{43-125}$; A93DK99R-ETV6$^{43-125}$; V112E-ETV6$^{43-125}$) of similar length to the previously published crystal structure were successfully purified. The ^{1}H-^{15}N HSQC of these constructs showed similar chemical shifts to

one another, as well as N90SA93D-ETV6^{1-125-}, indicating similar structures. Unfortunately these proteins also showed approximately twice as many peaks as were expected in their ^{1}H-^{15}N HSQC spectra. This was not due to degradation, proline isomerization, or deamidation and remains currently unexplained. The spectra of double-labeled ^{13}C/^{15}N-A93D-ETV6$^{43-125}$ were partially assigned. Based on the main chain chemical shift assignments, secondary structure propensity scores were calculated and found to have similar helical boundaries to the crystallized ETV6-SAM domain.

Titrations of unlabeled A93D-ETV6$^{43-125}$ or A93DK99R-ETV6$^{43-125}$ into ^{15}N-labeled V112E-ETV6$^{43-125}$ were conducted and interaction was observed by collection of ^{1}H-^{15}N HSQC spectra. The binding observed during these titrations was in the slow exchange regime, indicative of a strong interaction between monomers. This interaction was quantified using isothermal titration calorimetry. The binding constants of a pairs of monomers with native interface were determined to be on the same order of magnitude as previously published studies (K_d = 4.4 nM versus 1.7 nM). However, when present at the binding interface, the K99R point mutation was seen to affect the dissociation constant of the SAM domain by over two orders of magnitude. In contrast, when the mutation is introduced but not at the dimerization interface, tight binding still occurs. Thus the K99R mutation does not disrupt the structure of the SAM domain but does disrupt its intermolecular interactions. A summary of the binding interactions observed during these titrations is shown in Figure 25.

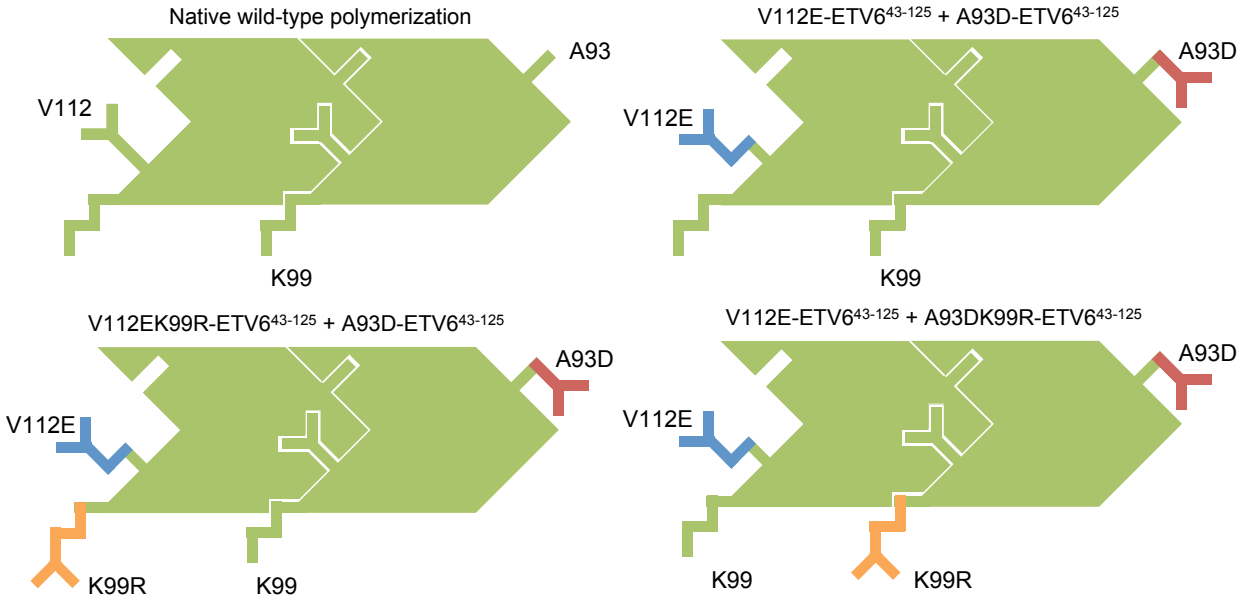


Figure 25. A cartoon summary of the native binding interface of the ETV6 SAM domain. A93D (red), V112E (blue), and K99R (orange) are highlighted. Clockwise from top left are: the native polymerization occurring in wild-type protein; a combination of A93D-ETV6$^{43-125}$ and V112E-ETV6$^{43-125}$ monomers with a native interface; a dimer interface with a K99R point mutation perturbation in the pairing A93DK99R-ETV6$^{43-125}$ and V112E-ETV6$^{43-125}$; and a native interface in the A93D-ETV6$^{43-125}$ and V112EK99R-ETV6$^{43-125}$ pairing.

5.2 Future work

The self-association of the SAM domain of ETV6 was studied in the scope of this thesis. The importance of residue Lys99 was confirmed by our binding studies and our results help explain the mechanisms underlying cellular transformation by the EN oncoprotein. However these results have generated many additional questions. From the long (1-125) and short (43-125) constructs of ETV6 we saw a doubling of peaks in the ^{1}H-^{15}N HSQC spectra. We showed that this doubling was not due to degradation, deamidation, or proline cis/trans isomerization. Ultimately the source of the doubled peaks has yet to be determined. Perhaps aliquots of the unstructured peptide tail (1-42) could be titrated into the 43-125 constructs and the effect on the number of peaks observed. Additionally, the D101KV112E-ETV6$^{43-125}$ construct was cloned but was not found to be soluble in the same NMR buffer as the other constructs. Determination of soluble conditions for this construct would be of great interest as it forms a salt bridge with the Lys99 residue at the binding interface and has also been shown to affect phenotype of colony formation in soft agar assays [8].

Ideally the ITC titrations should also be performed in reverse. If a better buffer solution for the D101K construct could be obtained this would also allow for further concentration of the other V112E mutants. Furthermore studying the effect of the other stabilizing salt bridges is of interest. Soft agar studies have been conducted by the Sorenson lab [8] for point mutations of Arg103 which is seen to form a salt bridge with E100 in the crystal structure. In the colony formation it appears that this salt bridge is less important for EN transformation activity. Thus determination of binding constants for these point mutants would further our understanding of this binding event.

Bibliography

1. Albaugh, P., et al., *Discovery of GNF-5837, a Selective TRK Inhibitor with Efficacy in Rodent Cancer Tumor Models.* ACS Medicinal Chemistry Letters, 2012. **3**(2): p. 140-145.

2. Bartel, F.O., T. Higuchi, and D.D. Spyropoulos, *Mouse models in the study of the Ets family of transcription factors.* Oncogene, 2000. **19**(55): p. 6443-54.

3. Barton, K., et al., *The Ets-1 transcription factor is required for the development of natural killer cells in mice.* Immunity, 1998. **9**(4): p. 555-63.

4. Bertrand, T., et al., *The crystal structures of TrkA and TrkB suggest key regions for achieving selective inhibition.* J Mol Biol, 2012. **423**(3): p. 439-53.

5. Bories, J.C., et al., *Increased T-cell apoptosis and terminal B-cell differentiation induced by inactivation of the Ets-1 proto-oncogene.* Nature, 1995. **377**(6550): p. 635-8.

6. Bosshard, H.R., D.N. Marti, and I. Jelesarov, *Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings.* J Mol Recognit, 2004. **17**(1): p. 1-16.

7. Brandts, J.F., H.R. Halvorson, and M. Brennan, *Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues.* Biochemistry, 1975. **14**(22): p. 4953-63.

8. Cetinbas, N., et al., *Mutation of the salt bridge-forming residues in the ETV6 SAM domain interface blocks ETV6-NTRK3 induced cellular transformation.* J Biol Chem, 2013.

9. Chakrabarti, S.R., et al., *Modulation of TEL transcription activity by interaction with the ubiquitin-conjugating enzyme UBC9.* Proc Natl Acad Sci U S A, 1999. **96**(13): p. 7467-72.

10. Chakrabarti, S.R., et al., *Posttranslational modification of TEL and TEL/AML1 by SUMO-1 and cell-cycle-dependent assembly into nuclear bodies.* Proc Natl Acad Sci U S A, 2000. **97**(24): p. 13281-5.

11. Chi, S.W., A. Ayed, and C.H. Arrowsmith, *Solution structure of a conserved C-terminal domain of p73 with structural homology to the SAM domain.* EMBO J, 1999. **18**(16): p. 4438-45.

12. De Braekeleer, E., et al., *ETV6 fusion genes in hematological malignancies: A review.* Leuk Res, 2012. **36**(8): p. 945-61.

13. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.

14. Fischer, G. and F.X. Schmid, *The mechanism of protein folding. Implications of in vitro refolding models for de novo protein folding and translocation in the cell.* Biochemistry, 1990. **29**(9): p. 2205-12.

15. Garrett-Sinha, L.A., et al., *PU.1 and Spi-B are required for normal B cell receptor-mediated signal transduction.* Immunity, 1999. **10**(4): p. 399-408.

16. Gill, G., *Something about SUMO inhibits transcription.* Curr Opin Genet Dev, 2005. **15**(5): p. 536-41.

17. Goddard, T.D., Kneller, D. G. , *Sparky 3*, 2009, University of California: San Francisco.

18. Green, J.B., et al., *RNA Recognition via the SAM Domain of Smaug.* Molecular Cell, 2003. **11**(6): p. 1537-1548.

19. Green, S.M., et al., *DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association.* J Biol Chem, 2010. **285**(24): p. 18496-504.

20. Hay, R.T., *SUMO: a history of modification.* Mol Cell, 2005. **18**(1): p. 1-12.

21. Hisaoka, M., et al., *Gene expression of TrkC (NTRK3) in human soft tissue tumours.* J Pathol, 2002. **197**(5): p. 661-7.

22. Holdgate, G.A. and W.H. Ward, *Measurements of binding thermodynamics in drug discovery.* Drug Discov Today, 2005. **10**(22): p. 1543-50.

23. Hollenhorst, P.C., L.P. McIntosh, and B.J. Graves, *Genomic and biochemical insights into the specificity of ETS transcription factors.* Annu Rev Biochem, 2011. **80**: p. 437-71.

24. Kim, C.A., et al., *Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression.* EMBO J, 2001. **20**(15): p. 4173-82.

25. Kim, C.A. and J.U. Bowie, *SAM domains: uniform structure, diversity of function.* Trends Biochem Sci, 2003. **28**(12): p. 625-8.

26. Knezevich, S.R., et al., *A novel ETV6-NTRK3 gene fusion in congenital fibrosarcoma.* Nat Genet, 1998. **18**(2): p. 184-7.

27. Kramer, M.F. and D.M. Coen, *Enzymatic Amplification of DNA by PCR: Standard Procedures and Optimization*, in *Current Protocols in Molecular Biology*2001, John Wiley & Sons, Inc.

28. Krause, D.S. and R.A. Van Etten, *Tyrosine kinases as targets for cancer therapy.* N Engl J Med, 2005. **353**(2): p. 172-87.

29. Lannon, C.L. and P.H. Sorensen, *ETV6-NTRK3: a chimeric protein tyrosine kinase with transformation activity in multiple cell lineages.* Semin Cancer Biol, 2005. **15**(3): p. 215-23.

30. Lau, D.K., M. Okon, and L.P. McIntosh, *The PNT domain from Drosophila pointed-P2 contains a dynamic N-terminal helix preceded by a disordered phosphoacceptor sequence.* Protein Sci, 2012. **21**(11): p. 1716-25.

31. Light, J.E., et al., *Clinical significance of NTRK family gene expression in neuroblastomas.* Pediatr Blood Cancer, 2012. **59**(2): p. 226-32.

32. Luo, Y., et al., *NTRK3 Is a Potential Tumor Suppressor Gene Commonly Inactivated by Epigenetic Mechanisms in Colorectal Cancer.* PLoS Genet, 2013. **9**(7): p. e1003552.

33. Mackereth, C.D., et al., *Diversity in structure and function of the Ets family PNT domains.* J Mol Biol, 2004. **342**(4): p. 1249-64.

34. Marsh, J.A., et al., *Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation.* Protein Sci, 2006. **15**(12): p. 2795-804.

35. Martin, M.J., et al., *The insulin-like growth factor I receptor is required for Akt activation and suppression of anoikis in cells transformed by the ETV6-NTRK3 chimeric tyrosine kinase.* Mol Cell Biol, 2006. **26**(5): p. 1754-69.

36. McIntosh, L.P. and F.W. Dahlquist, *Biosynthetic incorporation of 15N and 13C for assignment and interpretation of nuclear magnetic resonance spectra of proteins.* Q Rev Biophys, 1990. **23**(1): p. 1-38.

37. McKercher, S.R., et al., *Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities.* EMBO J, 1996. **15**(20): p. 5647-58.

38. Muthusamy, N., K. Barton, and J.M. Leiden, *Defective activation and survival of T cells lacking the Ets-1 transcription factor.* Nature, 1995. **377**(6550): p. 639-42.

39. Ng, T.L., et al., *The AMPK stress response pathway mediates anoikis resistance through inhibition of mTOR and suppression of protein synthesis.* Cell Death Differ, 2012. **19**(3): p. 501-10.

40. Obika, S., S.Y. Reddy, and T.C. Bruice, *Sequence specific DNA binding of Ets-1 transcription factor: molecular dynamics study on the Ets domain--DNA complexes.* J Mol Biol, 2003. **331**(2): p. 345-59.

41. Pierce, M.M., C.S. Raman, and B.T. Nall, *Isothermal titration calorimetry of protein-protein interactions.* Methods, 1999. **19**(2): p. 213-21.

42. Poirel, H., et al., *The TEL gene products: nuclear phosphoproteins with DNA binding properties.* Oncogene, 1997. **14**(3): p. 349-57.

43. Potter, S.M., W.J. Henzel, and D.W. Aswad, *In vitro aging of calmodulin generates isoaspartate at multiple Asn-Gly and Asp-Gly sites in calcium-binding domains II, III, and IV.* Protein Sci, 1993. **2**(10): p. 1648-63.

44. Qiao, F., et al., *Derepression by depolymerization; structural insights into the regulation of Yan by Mae.* Cell, 2004. **118**(2): p. 163-73.

45. Roukens, M.G., et al., *Identification of a new site of sumoylation on Tel (ETV6) uncovers a PIAS-dependent mode of regulating Tel function.* Mol Cell Biol, 2008. **28**(7): p. 2342-57.

46. Sattler, M., J.r. Schleucher, and C. Griesinger, *Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients.* Progress in Nuclear Magnetic Resonance Spectroscopy, 1999. **34**(2): p. 93-158.

47. Scott, E.W., et al., *Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages.* Science, 1994. **265**(5178): p. 1573-7.

48. Slupsky, C.M., et al., *Structure of the Ets-1 pointed domain and mitogen-activated protein kinase phosphorylation site.* Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12129-34.

49. Smalla, M., et al., *Solution structure of the receptor tyrosine kinase EphB2 SAM domain and identification of two distinct homotypic interaction sites.* Protein Sci, 1999. **8**(10): p. 1954-61.

50. Stapleton, D., et al., *The crystal structure of an Eph receptor SAM domain reveals a mechanism for modular dimerization.* Nat Struct Biol, 1999. **6**(1): p. 44-9.

51. Steinberg, I.Z., et al., *The Configurational Changes of Poly-L-proline in Solution.* Journal of the American Chemical Society, 1960. **82**(20): p. 5263-5279.

52. Stephenson, R.C. and S. Clarke, *Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins.* J Biol Chem, 1989. **264**(11): p. 6164-70.

53. Thanos, C.D., et al., *Monomeric structure of the human EphB2 sterile alpha motif domain.* J Biol Chem, 1999. **274**(52): p. 37301-6.

54. Tognon, C.E., et al., *Mutations in the SAM domain of the ETV6-NTRK3 chimeric tyrosine kinase block polymerization and transformation activity.* Mol Cell Biol, 2004. **24**(11): p. 4636-50.

55. Tran, H.H., et al., *Native interface of the SAM domain polymer of TEL.* BMC Struct Biol, 2002. **2**: p. 5.

56. Tugarinov, V., et al., *Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g.* J Am Chem Soc, 2002. **124**(34): p. 10025-35.

57. Ultsch, M.H., et al., *Crystal structures of the neurotrophin-binding domain of TrkA, TrkB and TrkC.* J Mol Biol, 1999. **290**(1): p. 149-59.

58. Velazquez-Campoy, A., et al., *Isothermal titration calorimetry.* Curr Protoc Cell Biol, 2004. **Chapter 17**: p. Unit 17 8.

59. Wedemeyer, W.J., E. Welker, and H.A. Scheraga, *Proline cis-trans isomerization and protein folding.* Biochemistry, 2002. **41**(50): p. 14637-44.

60. Wlodarska, I., et al., *ETV6 gene rearrangements in hematopoietic malignant disorders.* Leuk Lymphoma, 1996. **23**(3-4): p. 287-95.

61. Wood, L.D., et al., *Small ubiquitin-like modifier conjugation regulates nuclear export of TEL, a putative tumor suppressor.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3257-62.

Appendices

Appendix A Chemical shift assignments

Table 3. A summary of the chemical shift residue assignments for N90SA93D-ETV6$^{1-125}$ in 25 mM KCl, 25 mM Na_2HPO_4, pH 7.0 buffer and recorded at 25°C

Residue	^{13}CO	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	$^{1}H^{N}$	^{15}N
P5	176.8	63.39	32.21	-	-
A6	178.1	52.81	19.23	8.414	124.1
Q7	176	56.1	29.52	8.381	119.3
C8	176.1	55.65	32.99	-	-
S9	174.3	58.49	63.8	8.319	116.8
I10	176.1	61.31	38.84	8.117	122.6
K11	176.3	56.34	33.06	8.375	125.6
Q12	175.8	55.64	42.73	8.418	122.3
E13	176.3	-	30.78	8.532	122.9
R14	172.8	-	-	8.323	118
I15	176	61.09	38.94	8.233	122.7
S16	173.7	57.92	63.9	8.375	119.9
Y17	175.1	57.93	39.39	8.296	123.4
T18	181.6	59.13	70.06	7.998	121.2
Y27	175.5	57.67	38.73	-	-
A28	177.4	52.48	19.49	8.154	125.4
S29	174.6	58.29	63.83	8.196	114.9
S30	-	-	-	8.174	118.7
E45	176.6	57.58	30.34	8.421	119.8
D46	176.4	54.88	38.46	8.42	119.8
S47	174.5	58.73	63.93	8.144	115.1
I48	175.7	61.3	38.46	8.067	122
R49	175.5	55.22	30.63	8.615	126.4
L50	174.8	52.97	36.38	8.198	124.5
P51	177.2	62.35	32.52	-	-
A52	179.4	55.8	18.57	8.716	124
H53	175.5	57.62	29.03	8.335	112.7
L54	175.8	53.97	42.66	7.09	119.4
P59	-	63.35	37.53	-	-
Y60	176.7	57.47	39.03	7.922	120.2
W61	178.5	55.62	30.35	8.05	121.4

Residue	^{13}CO	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	$^{1}H^{N}$	^{15}N
S62	175.1	57.12	66.11	10.65	125.9
D63	178.2	-	-	8.213	120.4
R63	179.4	59.6	30.01	9.153	120.1
D64	178.2	58.2	45.17	8.219	120.4
D65	178	57.75	39.6	7.817	122.1
V66	177.7	67.63	32.38	8.284	120.2
A67	181.2	55.86	18	7.85	121
Q68	179	58.87	28.55	8.229	117.6
W69	174.4	61.57	27.12	8.844	125.9
Y69	176.9	-	-	8.84	125.9
L70	176.9	58.26	42.14	8.444	118.6
K71	179	58.17	32.08	7.509	117.4
W72	177.6	61.53	27.92	8.367	122.2
A73	179.1	54.83	17.68	8.84	122.4
E74	178.6	59.41	30.08	8.141	116.6
N75	177.4	55.85	39.66	7.338	114.6
E76	177.8	58.3	28.86	8.575	120.9
F77	174	57.09	37.85	7.69	112.4
S78	174.5	59.12	60.99	7.254	114
L79	177.3	53.12	42.5	8.051	117.9
R80	174	54.72	28.92	8.246	121.6
P81	177.1	63.71	34.9	-	-
I82	176.1	60.41	40.19	8.5	124.5
D83	177.7	54.2	41.09	8.617	126.3
S84	175.8	61.09	63.06	-	-
N85	176.4	54.95	37.95	8.772	119.2
T86	173.9	63.76	69.23	7.767	111.4
F87	174.4	57.37	40.25	7.953	120.7
E88	174.1	55.98	28.1	7.994	122.9
M89	172.3	54.58	34.62	8.176	121.8
S90	175.5	57.44	65.27	8.784	112.7
G91	174.6	47.72	-	8.718	108.4
K92	177.8	59.69	32.02	8.163	118.4
D93	179.1	56.72	41.7	7.021	117.6
L94	178.3	57.89	43.76	8.332	125
L95	178.3	56.57	41.88	7.648	115.3
L96	178.3	54.79	42.45	7.331	117.1
L97	178.5	55.43	42.21	7.375	119.5

Residue	^{13}CO	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	$^{1}H^{N}$	^{15}N
T98	175.6	60.43	72	9.561	114.4
K99	178.4	61.2	31.91	8.663	122.6
E100	178.9	60.23	28.84	8.283	117.3
D101	179.9	57.73	42.36	7.876	121.1
F102	178.4	-	-	8.335	120.4
Y104	177.8	60.48	38.51	-	-
R105	177	58.24	31.2	7.412	116.5
S106	181.7	54.06	62.26	8.217	110.5
P107	179.3	65.33	31.94	8.16	-
H108	177.2	57.46	39.78	8.159	111.5
S109	174.5	58.26	65.24	7.42	111.9
G110	174.1	48.92	-	9.353	111.9
D111	178.1	58.39	40.28	8.648	119.5
V112	178.1	65.82	31.83	7.503	119.3
L113	177.8	58.02	42.64	8.141	119.1
Y114	177.6	62.25	38.08	8.242	118.7
E115	179.6	58.88	29.15	7.833	118.1
L116	178.3	57.93	42.59	8.72	122.3
L117	177.9	57.94	40.8	8.406	120.8
Q118	178.8	58.02	27.69	7.927	115.1
H119	178.1	59.74	32.2	8.137	118.7
I120	178.4	65.32	38.42	8.225	120.6
L121	178.8	57.37	42.29	8.138	119.7
K122	177.2	57.57	32.87	7.557	116.9
Q123	176.3	56.75	29.25	7.792	118.3
R124	175	56.03	30.42	8.002	120.6
K125	181.2	57.76	33.74	7.897	128

Table 4. A summary of assignment of chemical shift residues for A93D-ETV6$^{43-125}$. Signals arising from doubled peaks are marked with (*). Recorded in 20 mM MOPS, 50 mM NaCl, 0.5 mM EDTA pH 8.0 buffer at 25°C

Residue	^{13}CO	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	$^{1}H^{N}$	^{15}N
M43	175.6	55.29	33.09	8.275	122.9
E44*	177	-	-	-	-
E44	176.8	56.54	30.26	8.458	122
E45*	176.6	-	-	8.434	120.7
E45	176.6	57.65	30.14	8.619	122
D46	-	54.86	40.96	8.445	120
D46*	-	-	-	8.24	120.7
S47	-	-	-	8.153	115.7
S47*	-	-	-	8.116	115.1
P51	177	62.34	32.35	-	-
P51*	177.1	62.27	32.37	-	-
A52*	179.4	55.75	18.32	8.706	124.2
A52	179.6	55.85	18.25	8.698	124.7
H53*	175.6	57.67	28.9	8.299	112.8
H53	175.6	57.8	28.77	8.345	113.1
L54*	175.9	54.05	42.66	6.999	119.6
L54	176.2	54.29	42.76	6.885	119
P58*	-	63.17	29.99	-	-
P58	-	63.19	30.01	-	-
I59	-	63.2	37.39	7.171	113.6
I59*	-	63.2	37.39	7.175	113.4
Y60	-	-	-	7.794	119.7
W61	-	-	-	8.037	121.5
S62*	-	-	-	10.65	125.5
S62	-	-	-	10.7	125.5
R63	-	-	-	9.165	120.3
R63*	-	-	-	9.208	120.5
D65	-	-	-	7.868	121.8
D65*	-	-	-	7.822	122
V66*	-	-	-	8.134	120.2
V66	-	-	-	8.188	120.2
Q68	-	-	-	8.252	117.6
W69*	-	-	-	8.8	126
W69	-	-	-	8.734	126
L70*	-	-	-	8.446	118.5
L70	-	-	-	8.383	118.5
K71	-	-	-	7.465	117.4

K71*	-	-	-	7.407	117.4
W72	-	-	-	8.357	122.3
W72*	-	-	-	8.377	122.5
A73*	-	-	-	8.816	122.5
A73	-	-	-	8.837	122.7
E74	178.6	59.37	29.88	8.056	116.8
E74*	-	-	-	8.087	116.7
N75	177.4	55.86	39.52	7.319	114.6
E76	177.8	58.21	28.74	8.562	121.1
F77	174.1	57.01	37.79	7.672	112.4
S78	174.6	58.37	60.99	7.236	114
L79	177.3	53.03	42.35	8.046	118
R80	174	54.67	28.71	8.229	121.6
P81	-	63.68	32.18	-	-
I82	-	60.31	40.01	8.466	124.2
D83	-	54.11	40.99	8.617	126.3
N85	-	-	-	8.756	119.2
N85*	-	-	-	8.737	119
T86	-	-	-	7.751	111.3
T86*	-	-	-	7.766	111.7
F87	-	-	-	7.934	121
M89	-	54.44	34.47	8.2	121.8
N90	-	51.02	38.78	8.531	114.9
G91	-	47.96	-	8.424	105.1
G91*	-	47.38	-	8.489	105.5
K92	178.8	59.99	31.75	7.741	119.8
K92*	178.6	59.92	31.57	7.774	120
D93*	179.6	56.96	40.99	7.848	117.8
D93	179.8	56.86	40.47	8.022	117.6
L94	-	-	-	8.339	125.1
L97	-	-	-	7.351	117.4
T98	-	60.42	71.99	9.546	114.5
T98*	-	-	-	9.591	114.5
K99	-	61.26	31.78	8.642	122.7
K99?	-	-	-	8.677	123.2
D100	-	60.16	-	-	-
E100	-	60.09	28.68	8.262	117.3
D101	-	60.09	-	7.864	121.1
F102	-	63.58	39.77	-	-
R103	-	59.08	30.62	8.161	118.7
Y104	-	60.44	38.35	8.127	118.4
R105	-	58.19	31	7.394	116.6

R105*	-	-	-	7.366	116.6
S106	-	54.03	62.24	8.213	110.5
P107	-	65.24	31.76	-	-
H108	-	57.52	31.22	8.053	111.7
P108	-	-	31.7	-	-
S109	-	58.15	65.34	7.367	112
S109*	-	-	-	7.325	112
G110	-	48.85	-	9.323	111.9
D111	178.1	58.34	40.14	8.614	119.5
V112	178.3	65.78	31.65	7.472	119.2
L113	177.7	57.96	42.64	8.147	119.3
Y114	-	62.29	38.03	8.221	118.7
Y114*	-	-	-	8.251	118.8
E115	-	58.74	28.85	-	-
L116*	-	57.81	-	8.655	122.3
L116	-	52.66	42.37	8.584	122.5
L117	178.4	58.01	40.67	8.344	120.8
L117*	-	-	-	8.334	120.5
Q118	178.7	57.87	27.46	7.864	115.1
H119	178.2	60	30.57	8.097	119.5
I120	178.4	65.39	38.23	8.219	120.5
L121	178.8	57.25	42.22	8.198	119.6
L121*	-	57.42	42.21	-	-
K122*	177.5	57.67	32.75	48.46	75.74
K122	177.4	57.62	32.72	7.537	116.9
Q123	176.4	56.58	29.14	7.789	117.8
Q123*	176.5	56.61	29.16	7.839	117.6
R124*	174.9	55.95	30.35	7.944	120.2
R124	175	55.98	30.42	7.962	120.6
K125*	181.2	57.76	33.52	7.818	128
K125	181.3	57.77	33.54	7.846	128

Appendix B SSP Scores from NMR assignments

Table 5. SSP score values of A93D-ETV6$^{1-125}$

Residue Number	SSP Value
5	-0.123
6	-0.261
7	-0.15
8	-0.088
9	-0.119
10	-0.268
11	-0.204
12	-0.305
13	-0.328
14	
15	-0.13
16	-0.1
17	-0.1
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	-0.068
28	-0.068
29	-0.068
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	

Residue Number	SSP Value
44	
45	0.225
46	0.191
47	0.107
48	0.079
49	-0.179
50	
51	0.303
52	0.277
53	0.277
54	0.531
55	
56	
57	
58	
59	0.106
60	-0.118
61	0.157
62	0.007
63	0.271
64	0.622
65	0.924
66	0.919
67	1.062
68	1.029
69	1.059
70	1.042
71	1.031
72	0.958
73	0.983
74	0.975
75	0.835
76	0.729
77	0.515
78	0.385
79	0.082
80	
81	-0.306
82	0.143

Residue Number	SSP Value
83	0.244
84	0.399
85	0.396
86	0.493
87	0.122
88	-0.16
89	-0.204
90	0.219
91	0.253
92	0.536
93	0.756
94	0.608
95	0.411
96	0.288
97	0.394
98	0.555
99	0.597
100	0.73
101	0.932
102	
103	
104	0.561
105	0.661
106	
107	0.193
108	0.316
109	0.495
110	0.557
111	0.764
112	0.997
113	0.979
114	0.925
115	1.002
116	1.016
117	0.875
118	0.854
119	0.836
120	0.65
121	0.492
122	0.452
123	0.268
124	0.129
125	0.095

Table 6. SSP values for construct A93D-ETV6$^{43-125}$

Residue Number	SSP Value
43	0.06
44	0.084
45	0.138
46	0.175
47	0.244
48	
49	
50	
51	0.481
52	0.26
53	0.26
54	0.583
55	
56	
57	
58	0.398
59	0.166
60	-0.035
61	0.185
62	0.141
63	0.242
64	0.556
65	0.846
66	0.725
67	
68	
69	
70	
71	
72	
73	
74	0.911
75	0.769
76	0.759
77	0.515
78	0.365
79	0.043
80	
81	-0.348
82	-0.287

Residue Number	SSP Value
83	-0.287
84	
85	
86	
87	
88	
89	-0.108
90	0.415
91	0.581
92	1.014
93	1.155
94	
95	
96	
97	
98	0.877
99	0.945
100	0.852
101	0.992
102	0.92
103	0.73
104	0.653
105	0.699
106	
107	0.241
108	0.354
109	0.516
110	0.606
111	0.767
112	1.011
113	0.983
114	0.803
115	0.861
116	0.868
117	0.773
118	0.773
119	0.893
120	0.74
121	0.628
122	0.532

123	0.319
124	0.174
125	0.113

Appendix C ITC titration results

Table 7. ITC results for individual experiments

Protein in Cell	Protein in Syringe	n	K_a (M^{-1})	ΔH (kcal/mol)	ΔS (cal/mol-K)
V112E	A93D	0.948	$1.15x10^8$	-8055	9.9
V112E	A93D	0.888	$3.19x10^8$	-8395	10.8
V112E	A93D	0.879	$1.44x10^8$	-8559	8.6
V112E	A93D	0.874	$3.32x10^8$	-8440	10.7
V112E	A93DK99R	0.937	$9.35x10^5$	-1155	23.4
V112E	A93DK99R	0.728	$6.22x10^5$	-1326	22.1
V112E	A93DK99R	1.140	$3.39x10^5$	-2332	17.5
V112E	A93DK99R	1.100	$2.04x10^5$	-3271	13.3
V112EK99R	A93D	0.897	$1.15x10^8$	-7078	13.1
V112EK99R	A93D	0.938	$1.91x10^8$	-7326	13.3
V112EK99R	A93D	0.989	$1.61x10^8$	-7638	11.9

Appendix D EMS results

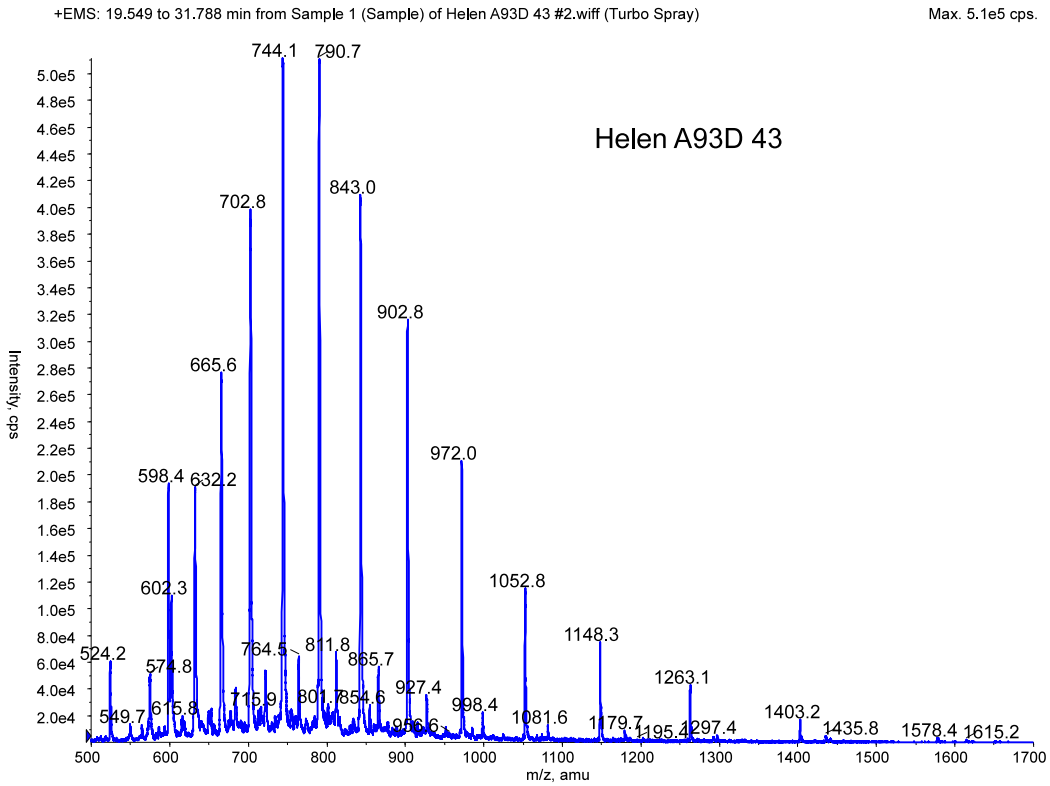


Figure 26. EMS m/z measurement of A93D-ETV6$^{43\text{-}125}$