Expression and purification of an NSD3-GB1 fusion protein as a way to study the structure of complex formation with Brd4 ET.

Research Thesis

Presented in partial fulfillment of the requirements for graduation *with research distinction* in Biochemistry in the undergraduate colleges of The Ohio State University

by

Jennifer Tuokkola

The Ohio State University

April 2020

Project Advisor: Professor Mark Foster, Department of Chemistry and Biochemistry

Introduction

Brd4

The Brd4 protein is a eukaryotic transcription factor belonging to the bromodomain and extraterminal domain (BET) family of proteins[1]. Brd4 is involved in both transcriptional and epigenetic regulation and its overexpression is heavily implicated in both carcinogenesis and tumor metastasis[2].

As is characteristic of the BET family, Brd4 has two bromodomains which are responsible for binding acetylated lysine residues. Brd4 has a preference to bind histones in hyperacetylated regions; this directs it to promoter and enhancer regions where it promotes transcription at both initiation and elongation steps[2]. During transcription initiation, Brd4 acts as a kinase and phosphorylates Serine 2 on the RNA Polymerase II C-terminal domain (CTD)[3]. Brd4 also recruits PTEFb (positive transcription elongation factor) to promoters and activates its kinase subunit. PTEFb then phosphorylates CTD Serine 2 during productive elongation[4].

Studies have identified Brd4 target genes associated with multiple myeloma, colon cancer, lung cancer, prostate cancer, breast cancer, and more. This has made it a popular target in the field of cancer research. All current inhibitors of the BET family target the bromodomains by mimicking acetylated lysines and preventing Brd4 from binding to chromatin[2]. Bromodomains are not specific to the BET family, meaning these inhibitors do not target only the BET family.

The extraterminal (ET) domain of Brd4 is specific to the BET family but remains unexplored as a therapeutic target. The ET domain is responsible for recruiting chromatin-modifying factors that

result in transcriptional regulation and co-activation and also serves as a binding site for several viral factors[5]. Several structures have been solved of the ET domain in complex with peptides derived from these chromatin-modifying and viral factors, but the complete interactions have not been characterized.

Previous work from the Foster Lab characterizes the structure of interaction between Brd4 ET and a peptide derived from murine leukemia virus (MLV) integrase using nuclear magnetic resonance (NMR)[6]. The structure shows the formation of a three-stranded antiparallel β-sheet motif. This study goes on to indicate that MLV integrase can outcompete nuclear set receptor SET domain protein 3 (NSD3) for binding by the ET domain. This suggested a conserved mechanism of interaction between the ET domain and its binding factors involved in transcriptional activation.[7]

NSD3

NSD3 is a histone methyltransferase responsible for the methylation of lysine residue 36 on histone H3. This methylation results in looser binding of DNA by the histone, making the DNA accessible to replication and transcriptional machinery and thereby promoting cell cycle progression. Due to this role, NSD3 has been shown to be imperative in the maintenance of acute

Figure 1. Structure of Brd4 ET domain (green and red) bound to a peptide 152-EIKLKITKTIQN-163 derived from NSD3 (orange)[4]

myeloid leukemia (AML)[8]. The NSD3-short isoform is 645 amino acid residues and contains only the PWWP1 domain which is a chromatin reader domain. While this isoform lacks the
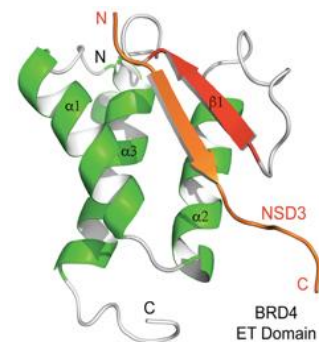
methyltransferase activity of the long isoform, it effectively serves as an adaptor protein which couples Brd4 to the chromatin remodeler protein CHD8[9].

Due to its relevance in transcriptional regulation and AML maintenance, the interaction between Brd4 ET and NSD3 has been studied. Previous studies showed that NSD3 residues 100-263 were sufficient to maintain binding with the ET domain with an affinity of 2.1 μM, measured by surface plasmon resonance (SPR)[6]. Subsequent studies using nuclear magnetic resonance indicated that residues 152-163 of NSD3 form an intermolecular β-sheet with the ET domain (Figure 1) with a reported K_d of 140 μM[7], measured by isothermal titration calorimetry (ITC). NSD3 residues 152-163 show sequence similarity with the ET binding region of MLV IN, yet the K_d of the peptide is almost 10 times higher than the longer construct, indicating that more than just residues 152-163 are involved in this interaction. Thus, the overall goal of my research is to determine the structure of the interaction between Brd4 ET and NSD3. This research will provide insight into an important biological interaction and will provide a valuable information on possible development of novel anti-cancer therapeutics which inhibit the BET family ET domain.

The first problem encountered in this project is that NSD3 100-263 did not remain soluble at

the concentrations needed to characterize its structure using NMR spectroscopy. Our

recombinant NSD3-expression construct was originally fused to a glutathione S-transferase

(GST) tag to facilitate purification and maintain solubility during expression. However, the large

size of GST (26.9 kDa) is incompatible with NMR

structural studies, and GST must be removed

prior to these experiments. Removal of the tag

by treatment with the sequence-specific tobacco

etch virus protease (TEV protease) resulted in

immediate precipitation of NSD3 even when in

complex with Brd4 ET. To determine if NSD3

100-263 was the precipitate, an SDS-page gel

was run showing GST-NSD3 alone and in

Figure 2. SDS-PAGE gel verifying precipitate formation by NSD3 when cleaved from GST tag. Lane 1: Size marker, Lane 2: GST-NSD3, Lane 3: TEV Protease, Lane 4: GST, Lane 5: Brd4 ET, Lane 6: GST-NSD3 with Brd4 ET, before TEV cleavage, Lane 7: GST-NSD3 with Brd4 ET, after TEV cleavage.

complex with Brd4 ET (Figure 2).   The gel shows the disappearance of the product band

corresponding to GST-NSD3 following TEV cleavage, indicating that it is no longer in solution.

To combat this, GB1 was used as an alternative solubility tag. GB1 is a 56 amino acid residue

protein that originates from the B1 domain of *Streptococcal* protein G[10]. It has also been shown

to increase the solubility of the proteins to which it is fused. The main benefit of using GB1 is

that it has a well-defined NMR spectrum. This means that once fused, there is no need to
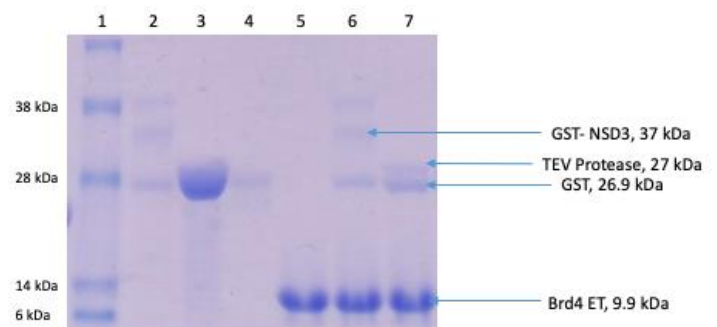
cleave the GB1 tag as long as its signals don't overlap or convolute the spectrum of the NSD3-Brd4 complex.

NMR to Study Structure

A 2D {^{1}H, ^{15}N} HSQC spectrum provides valuable structural information regarding protein structure and dynamics. Each signal in a ^{1}H-^{15}N-HSQC spectrum arises from correlation between nitrogen and hydrogen, such as in the amide bond in the protein backbone. This means that each amino acid residue in a protein, with the exception of proline, will produce a peak on the spectrum. The position of each peak on the spectrum is characteristic of the chemical environment surrounding that amide Changes in peak position, referred to as chemical shift perturbations (CSPs), can be tracked upon the addition of a binding partner to determine which residues are interacting.

To measure the interaction between Brd4 ET and NSD3, several spectra will be recorded. First, apo spectra of ET, NSD3-GB1, and GB1 alone will each be recorded. This provides reference spectra and allows for verification that the peaks arising from GB1 can be distinguished from the peaks arising from NSD3. Next, a spectrum will be recorded of the ET:NSD3 complex. The ET domain will be uniformly ^{15}N-labeled, and NSD3-GB1 will be unlabeled. As ^{14}N is NMR invisible, only peaks from the ET domain will be observed. CSPs upon the addition of NSD3-GB1 will be tracked. As signals from the ET domain have previously been assigned, these CSPs will indicate which residues of the ET domain are involved in the interaction. If residues beyond those reported to interact with NSD3 152-163 undergo significant CSPs, this will indicate an extended or additional interaction interface. These experiments could then be repeated with the ET

domain unlabeled and NSD3-GB1 labeled to determine which residues of NSD3 are interacting

with the ET domain. As all signals from NSD3 100-263 have not previously been assigned,

additional triple resonance experiments will be performed to assign the peaks and map the

CSPs.

Materials and Methods

Molecular Cloning

NSD3 100-263 in pop5BT vector

The gene sequence of NSD3 100-263 was amplified from bacterial vector pGEX-6P-1 (Sigma Aldrich GE28-9546-48) using the following primers: forward, 5'-CTTGTCGACTTTTGGCGCGGTGC GTAACTTCAG-3' and reverse, 5'GCACGGCTGGTGGCTTCAATCGATTCTTAAGGGTT-3'. The PCR amplified product was purified and digested with SalI and EcoRI and cloned into these restriction sites of the pOP5BT vector (Addgene 112609) with an N-terminal GB1 protein and His tag (6x-His). There is a 7-residue linker (TGSGTSG) between the final residue of GB1 and the start of NSD3 100-263.
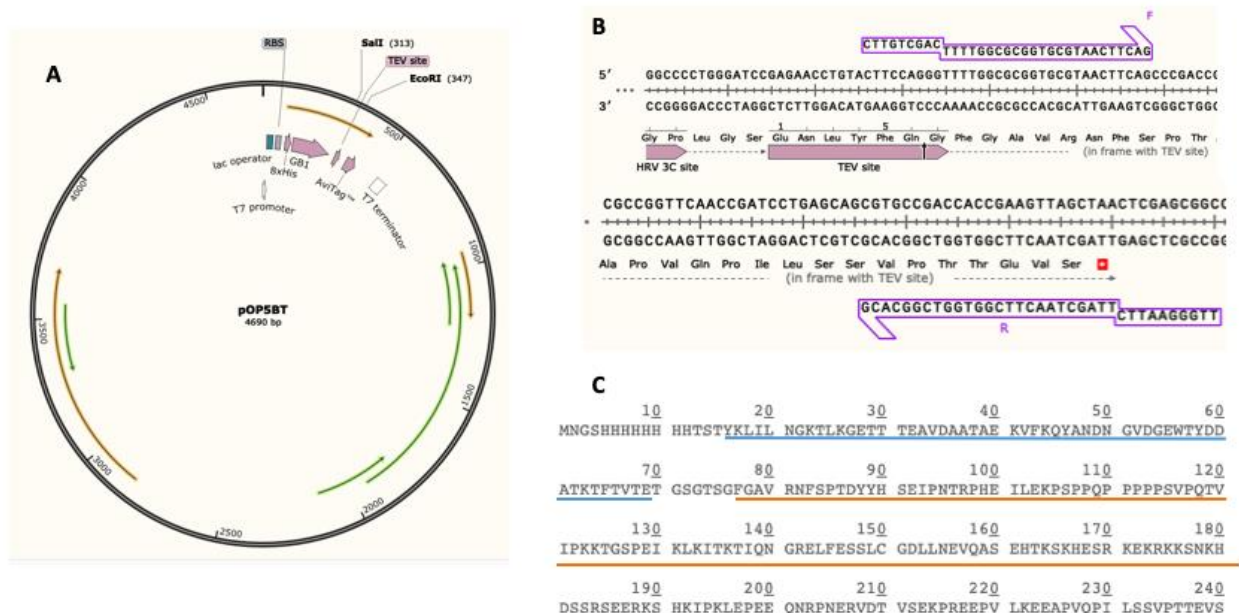

Figure 3. *Design of NSD3-GB1 in the pOP5BT vector. (A) The pOP5BT vector with restriction sites for the insertion of NSD3 100-263 highlighted. (B) Primers designed to amplify NSD3 100-263 from the pGEX-6P-1 vector. (C) The*

expected protein sequence of NSD3-GB1 in pOP5BT, GB1 is underlined in blue and NSD3 100-263 is underlined in orange.

NSD3 100-263 in GEV 2 vector

NSD3 100-263 was amplified from bacterial vector Puc57 (GenScript SD1176) using the following primers: forward, 5'-TTTGGATCCTTCGGTGCGGT-3' and reverse, 5'-GGCGCCCTGACTCGAG-3'. The insert was amplified using TD-PCR, purified, and digested with BamHI and EcoRI. It was cloned into these restriction sites of the GEV 2 vector (Addgene 12616) with an N-terminal GB1 protein and a C-terminal His tag.
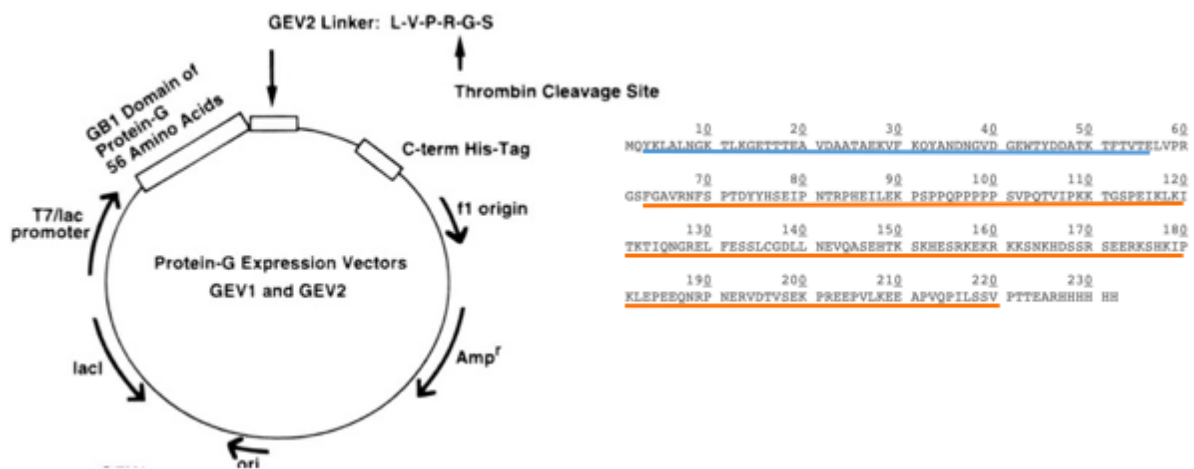

Figure 4. *An overview of the GEV2 vector and the expected sequence of NSD3-GB1 in GEV2 with GB1 underlined in blue and NSD3 100-263 underlined in orange.*

Protein Expression and Purification

Plasmids were transformed into electrocompetent *E. coli* BL21(DE3) cells by electroporation.

Both plasmids are under control of an IPTG (isopropyl-β-thiogalactopyranoside) inducible

promoter. To begin expression, a colony from a fresh transformation was picked and used to inoculate a 100 mL LB starter culture with 100 µg/mL of carbenicillin, this culture was grown overnight at 37°C with shaking. To express unlabeled protein, 20 mL of the starter culture was added to 1 L of LB media with 100 µg/mL of carbenicillin. This culture was grown at 37°C with shaking while the OD_{600} was measured. To express uniformly [U-^{15}N]-labeled protein, 20 mL of the starter culture was added to 1 L of M9 minimal media containing 1 g/L ^{15}N-ammonium chloride (Cambridge Isotope Laboratories, Inc.) as the sole nitrogen source.

For the NSD3-GB1 in GEV2 vector, the best induction was observed when IPTG was added to a final concentration of 1 mM when the OD_{600} was 0.6. Following induction, the protein continued to grow at 37°C for 6-8 hours before being harvested by centrifugation at 5000 rpm using Sorvall SLA-3000 rotor for 20 minutes at 4°C.

For the NSD3-GB1 in pOP5BT vector, several conditions were tested for induction of the protein. IPTG concentrations ranging from 0.4-1 mM were tested and induction temperatures of 37° C, 30° C, and 20° C were tested, but induction was not observed with any condition tested. Despite not observing induction, the protein was still harvested in the same way as listed above and purified.

To fulfill the goal of obtaining an NMR spectrum of GB1 by itself, the pHLIC GB1 vector was used. This vector was a gift from the Magliery laboratory at The Ohio State University. GB1 was expressed in the same way as described above and induced to a final concentration of 1 mM IPTG when the OD_{600} was between 0.6-0.7.

To uniformly label these with ^{15}N, they were grown in defined minimal media. To make the minimal media, 9.465 g Na_2HPO_4, 0.18 g KH_2PO_4, and 1 g $^{15}NH_4Cl$ are dissolved and autoclaved in 1 L of ddH_2O. Before inoculating the culture, you must also add: 100 μL 1M $CaCl_2$, 2 mL 1M $MgSO_4$, 10 mL Gibco MEM Vitamin Solution (100x), 4 g glucose, 1 mL 1000x trace metal mix (50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnCl_2$, 10 mM $ZnSO_4$, 2 mM $CoCl_2$, 2 mM $CuCl_2$, 2 mM $NiCl_2$), and carbenicillin to a final concentration of 100 μg/mL. Following inoculation, these cultures are grown, induced, and harvested in the same manner as listed above.

To purify these proteins, the cell pellet is removed from the -80°C freezer and thawed on ice. After it is thawed, the pellet is dissolved in 35 mL Lysis Buffer (20 mM $NaPO_4$, pH 7.0, 500 mM NaCl, 15 mM imidazole) and 500 μL 10 mM phenylmethylsulfonyl fluoride (PMSF). A complete mini protease tablet is also added to the cells. Cells are lysed by sonication following a cycle time of 5 minutes with 5 seconds on and 2 seconds off. The sonication step is generally repeated two times. Following sonication, the cells are centrifuged at 15,000 rpm using Sorvall SS-34 rotor for 45 minutes at 4°C. The supernatant is then filtered through a 0.45-micron filter and then loaded onto the appropriate column.

Both constructs mentioned above contain a His-tag, so a Ni column is used for purification. A 5 mL HisTrap FF column from GE healthcare was used for all purification. The column was stripped of metal by washing with 25 mL 0.5 M EDTA, washed with 10 mL H_2O, and charged with 25 mL of 100 mM $NiSO_4$. The column was then equilibrated with 25 mL of lysis buffer. All samples were loaded onto the HisTrap FF column using a peristaltic pump at a flow rate of 5 mL/min.

In the case of NSD3-GB1 in pOP5BT vector, all purifications were done using a peristaltic pump with an isocratic elution. After loading the sample, the column was washed with 35 mL of lysis buffer and then eluted with 25 mL of elution buffer (20 mM NaPO$_4$, pH 7.0, 500 mM NaCl, 500 mM imidazole). This elution was concentrated to 1 mL using an Amicon centrifugal filter with MWCO of 10 kDa by centrifuging at 5000 rpm using Sorvall SS-34 rotor at 4°C. The sample was further purified on a HiLoad 16/60 Superdex 75 (GE Healthcare) in 20 mM NaPO$_4$, pH 7.0, and 500 mM NaCl. Fractions containing the protein, as assayed by SDS-PAGE were collected and prepared for NMR.

Free GB1 was also purified on a HisTrap FF column as described above. The elution from this column was concentrated to 10 mL and dialyzed overnight with 200 µL of 200 mM TEV protease in buffer containing 20 mM NaPO$_4$, pH 7.0, and 500 mM NaCl The TEV protease removed the His-tag from GB1. The TEV protease used contained a His-tag so following dialysis, the dialysate was run on a second HisTrap FF column. This time, GB1 came through in the wash step and TEV protease came off in the elution. To ensure its purity, GB1 was also run on an S75p size exclusion column and the fractions from this column were collected and prepared for NMR.

In the case of the NSD3-GEV2 construct, several difficulties were encountered during its purification. After loading onto the HisTrap FF column, a gradient elution was set up to reach a final concentration of 100% elution buffer over 10 column volumes. The complex elutes off the column at extremely low concentrations of imidazole with no separation of protein components. To combat this, the lysis buffer was remade with 0 mM imidazole and the elution

buffer was remade with only 100 mM imidazole. There was still no separation of protein components seen during this purification.

NMR

To prepare samples for NMR, their concentrations were first tested. Both free GB1 and NSD3 in pOP5BT vector contain sufficient tryptophan and tyrosine residues to check their concentrations by absorbance at 280 nm. All samples used for NMR had a concentration around 100 μM. The samples were then buffer exchanged into NMR buffer (10 mM NaCl, 5 mM DTT, 20 mM d_{11}-Tris, pH 7.0) using a PD-10 desalting column. The samples were then concentrated to approximately 1 mL. In the NMR sample tube, 475 μL of sample were combined with 50 μL D_2O and 1 μL DSS. All NMR spectra were collected on a Bruker 600 MHz instrument. Raw data was processed using NMRFx and peak assignments were competed using NMRViewJ[11].

Results and Discussion

Cloning, Expression, Purification, and NMR Spectroscopy of NSD3-GB1 in pOP5BT Vector

Amplification of NSD3 100-263 from pGEX-6P-1 vector was verified via agarose gel. Controls of both restriction enzymes were done by showing linearization of plasmid DNA when digested by only one enzyme. Despite this, all ligation attempts of NSD3 100-263 into the pOP5BT vector were unsuccessful. Two ligase enzymes were tested with multiple incubation times and ratios of insert to vector. After many failed ligation attempts, cloning of NSD3 100-263 into the pop5BT vector was ultimately performed by GenScript.

NSD3-GB1 in pOP5BT vector was expressed and purified as described in Materials and

Methods. Figure 5 shows the expression and purification of NSD3-GB1. Throughout the

induction process, there is no change in the band intensity corresponding to the protein despite

being under the control of an IPTG-inducible promoter. NSD3-GB1 in pOP5BT has an expected

molecular weight of 27 kDa but runs much higher on SDS-PAGE. This may be caused by a high

number of aspartate and glutamate residues (15% overall composition) or proline residues

(10% overall composition) in the protein. There may also be a mutation in the construct that
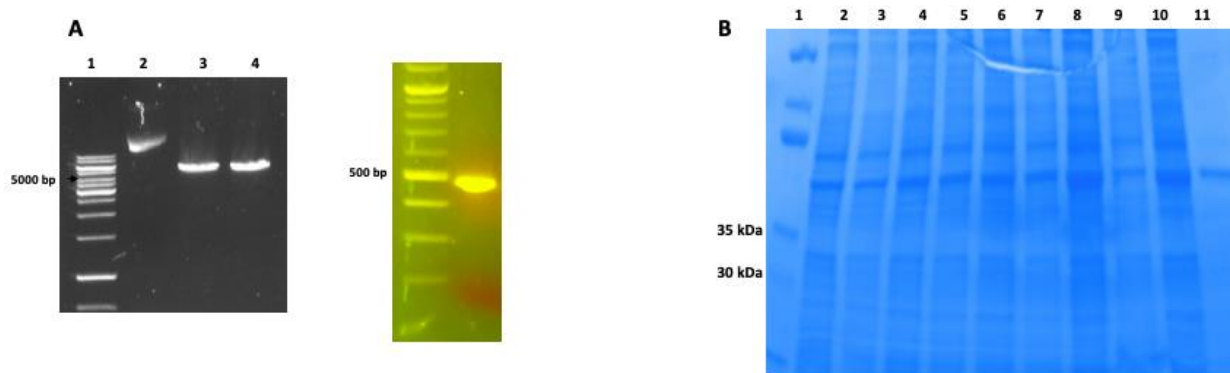
results in the molecular weight being higher.

Figure 5. *Cloning, expression and purification of NSD3-GB1 in pOP5BT monitored by agarose gel and SDS-PAGE.*

(A) The gel on the left shows the restriction digest control done of the pOP5BT vector. Lane 1: Size marker, Lane 2: Undigested pOP5BT vector, Lane 3: SalI digested vector, Lane 4: EcoRI digested vector. The gel on the right shows the amplification of NSD3 100-263. Lane 5: Size marker, Lanes 6-9: PCR-Amplified NSD3 100-263.

(B) Expression of GB1-NSD3 in pOP5BT in E. coli BL21(DE3) and purification using the HisTrap FF column. Lane 1: MW marker, Lanes 2-3: Cells before induction with 0.5 mM IPTG, Lanes 4-5: Cells 1 hour post-induction, Lanes 6-7: Cells 2 hours post-induction, Lane 8: Soluble fraction following sonication, Lane 9: Insoluble fraction following sonication, Lane 10: Ni column flow-through and wash, Lane 11: Ni column elution.

Purification of NSD3-GB1 in pOP5BT yielded 500 μL protein with concentration of 91 μM,

measured by the A_{280}. This sample was prepared for NMR as described in materials and

methods and a ^{1}H-^{15}N-HSQC was recorded. Upon overlay with the GB1 spectrum, several peaks

arising from GB1 could be assigned, however the spectrum shows that NSD3 100-263 is not

folding properly. The overlapping peaks in the middle of the spectrum indicate lack of chemical

shielding caused by secondary structure interactions. It is possible that the location of GB1 in

the construct was preventing NSD3 100-263 from folding properly and induction of the

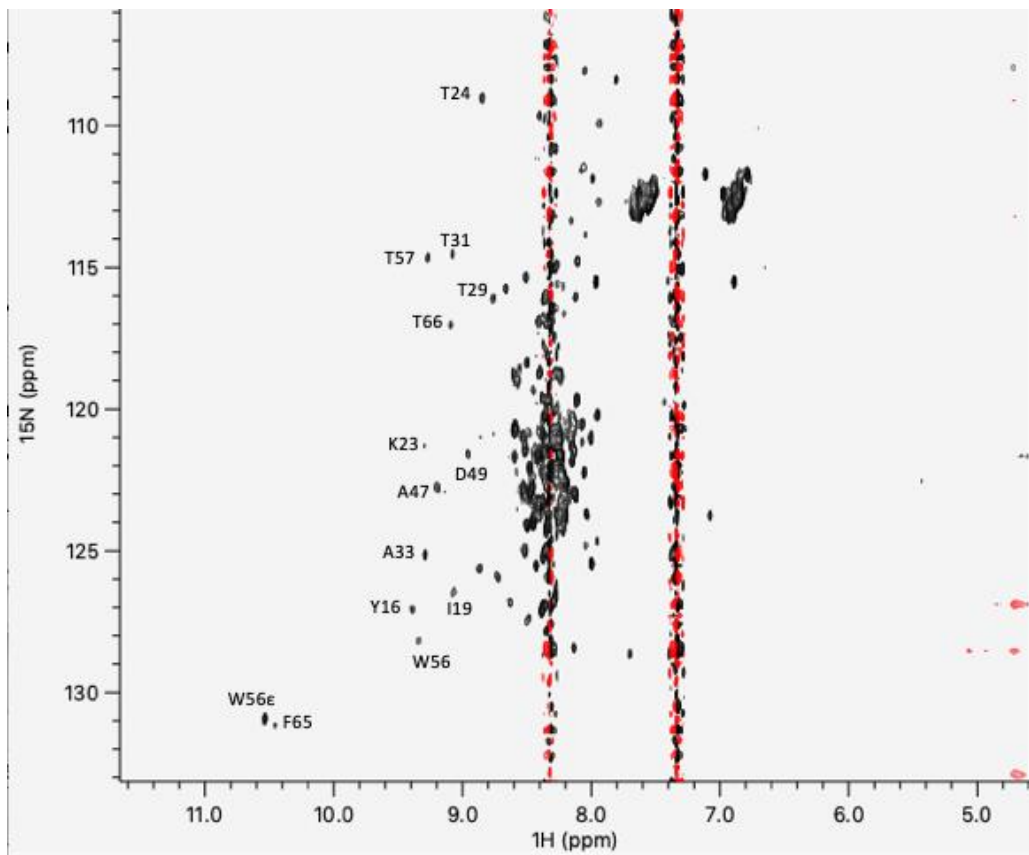construct was never observed, so NSD3 100-263 was cloned into an alternate GB1-encoding

vector.

Figure 6. Two-dimensional {^{1}H, ^{15}N}- HSQC spectrum of [U-^{15}N-NSD3-GB1 from pOP5BT recorded at 600 MHz. GB1 peak assignments made from pHLIC GB1 spectrum (Figure 10).

Cloning, Expression, and Purification of NSD3-GB1 in GEV2 Vector

In an attempt to achieve better protein induction a different GB1-encoding vector was tested, GEV2. NSD3 100-263 was amplified from the Puc57 vector using touchdown (TD) PCR. This amplification was verified with an agarose gel. The amplified and double digested NSD3 100-263 was ligated with the double digested GEV2 vector to create the recombinant NSD3-GB1 in GEV2 vector. The results of the ligation vector were verified via Sanger Sequencing (Figure 7).

Protein expression was optimized as described in Materials and Methods and induction of the product band is observed on an SDS-page gel. The crude lysate was loaded onto the HisTrap FF column, but the target protein eluted at very low imidazole concentrations (< 100 mM imidazole). As a consequence, there is no separation seen between proteins in the crude lysate when purification is done via Ni affinity chromatography, as shown in figure 8. While hexa-histidine tagged constructs can have varying affinity to the Ni resin, it is unexpected to have such low affinity that no separation from crude lysate is observed. A possible explanation for this low affinity is that this tag is not completely exposed and to interact with the Ni resin.
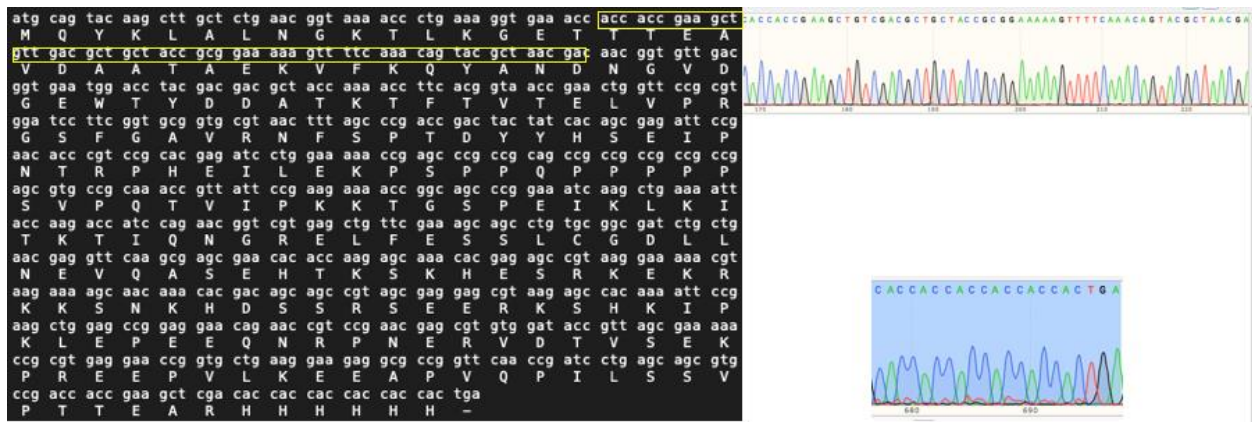

Figure 7. *The expected sequence of NSD3-GB1 in GEV2 shown on the left with chromatograms from Sanger Sequencing on the right, verifying the sequence of the ligated construct.*
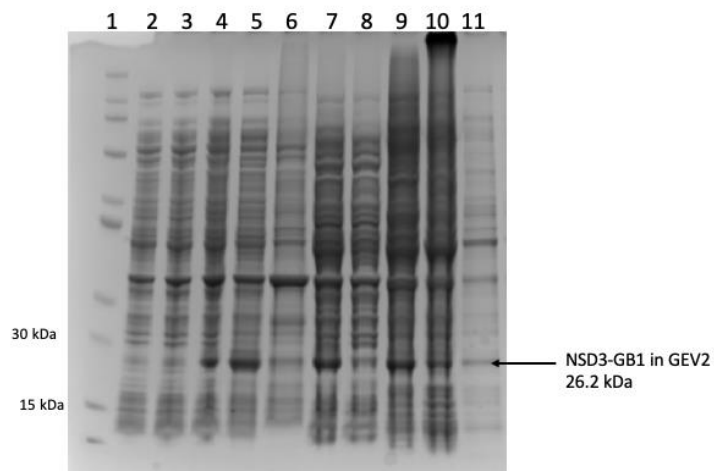
Figure 8. *Expression of NSD3-GB1 in GEV2 monitored by SDS-PAGE. Lane 1: MW Marker, Lanes 2-3: Cells before induction with 1 mM IPTG, Lanes 4-5: Cells after induction, Lane 6: Insoluble fraction following sonication, Lane 7: Soluble fraction following sonication, Lane 8: Ni column flow-through, Lanes 9-11: Ni column elution fractions.*

Expression, Purification, and NMR Spectroscopy of pHLIC GB1

Free GB1 was expressed and purified as described in Materials and Methods. Figure 9 outlines the expression and purification process of GB1.
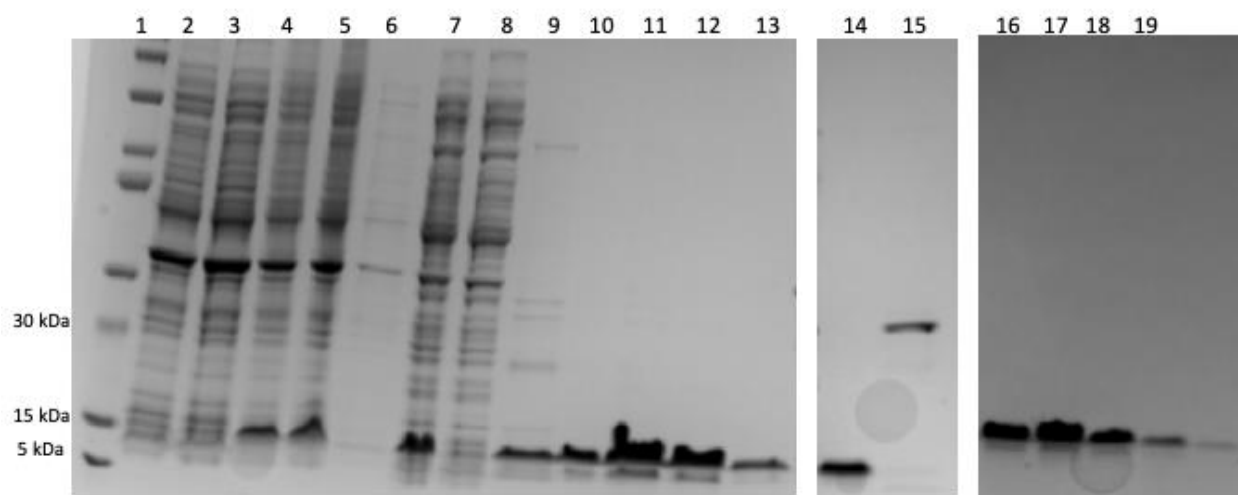

Figure 9. *SDS-PAGE gels outlining the expression and purification process of GB1. Lane 1: Protein Marker, Lanes 2-3: pre-induction samples, Lanes 4-5: post-induction samples, Lane 6: insoluble fraction, Lane 7: insoluble fraction,*

Lane 8: Ni column flow-through, Lanes 9-13: Ni column elution fractions, Lane 14: Ni column flow-through following TEV cleavage, Lane 15: Ni column elution following TEV cleavage, Lanes 16-19: S75p size exclusion fractions.

The fractions from the size exclusion were concentrated to 500 µL and their concentration was determined to be 97 µM. The sample was prepared for NMR and an HSQC was recorded. Peak assignments were made based on WT GB1 assignments published in the Biological Magnetic Resonance Data Bank (BMRB 25909). pHLIC GB1 was expected to have 60 amino acids following TEV cleavage with 3 asparagine residues and 1 glutamine residues. 63 total peaks were identified and 57 of them were assigned. Figure 10 shows the labeled HSQC with the assigned peaks.

Figure 10. *2D {^{1}H, ^{15}N}- HSQC spectrum of ^{15}N labeled GB1 protein recorded at 600 MHz.*

Conclusion and Next Steps

Although NSD3 100-263 did not appear folded in the ^{1}H-^{15}N-HSQC, the GB1 tag was successful at keeping NSD3 100-263 soluble for purification. The design of a construct in the pOP5BT vector appeared to inhibit the folding of NSD3 100-263, making its spectrum indistinguishable. The GEV2 vector holds promise in allowing proper folding of NSD3 100-263 but purification of this construct needs to be optimized.

It is also a possibility that NSD3 100-263 does not have secondary structure. There is no solved structure of this peptide, and it is predicted that even NSD3 152-163 is unstructured before binding ET. Although NSD3-GB1 in pOP5BT does not show a well-defined secondary structure on its HSQC, it there is a possibility that it may adopt a secondary structure when in complex with Brd4 ET. This could be monitored by taking the HSQC of NSD3-GB1 in pOP5BT while it is in complex with ET. This would give better insight into if the GB1 tag is interfering with any possible secondary structure formation.

Due to the problems faced involving expression and purification of NSD3-GB1, there is an option to clone NSD3 100-263 into the pHLIC GB1 vector. Free GB1 in this vector showed good induction and simple purification by Ni affinity chromatography. The pHLIC vector offers cloning sites so NSD3 100-263 could be inserted with the GB1 tag on the N-terminus.

Overall, characterizing the complete interaction between ET:NSD3 holds promise for directing future development of cancer therapeutics. ET demonstrating a conserved mechanism of interaction would open up many possibilities in terms of specific targeting of the BET family.

Acknowledgements

Work Cited

(1)   Sanchez, R.; Zhou, M.-M. The Role of Human Bromodomains in Chromatin Biology and Gene Transcription. *Curr. Opin. Drug Discov. Devel.* **2009**, *12* (5), 659–665.

(2)   Donati, B.; Lorenzini, E.; Ciarrocchi, A. BRD4 and Cancer: Going beyond Transcriptional Regulation. *Mol. Cancer* **2018**, *17* (1), 164. https://doi.org/10.1186/s12943-018-0915-9.

(3)   Devaiah, B. N.; Singer, D. S. Two Faces of BRD4. *Transcription* **2013**, *4* (1), 13–17. https://doi.org/10.4161/trns.22542.

(4)   Wu, S.-Y.; Chiang, C.-M. The Double Bromodomain-Containing Chromatin Adaptor Brd4 and Transcriptional Regulation. *J. Biol. Chem.* **2007**, *282* (18), 13141–13145. https://doi.org/10.1074/jbc.R700001200.

(5)   Rahman, S.; Sowa, M. E.; Ottinger, M.; Smith, J. A.; Shi, Y.; Harper, J. W.; Howley, P. M. The Brd4 Extraterminal Domain Confers Transcription Activation Independent of PTEFb by Recruiting Multiple Proteins, Including NSD3. *Mol. Cell. Biol.* **2011**, *31* (13), 2641–2652. https://doi.org/10.1128/MCB.01341-10.

(6)   Crowe, B. L.; Larue, R. C.; Yuan, C.; Hess, S.; Kvaratskhelia, M.; Foster, M. P. Structure of the Brd4 ET Domain Bound to a C-Terminal Motif from γ-Retroviral Integrases Reveals a Conserved Mechanism of Interaction. *Proc. Natl. Acad. Sci.* **2016**, *113* (8), 2086–2091. https://doi.org/10.1073/pnas.1516813113.

(7)   Zhang, Q.; Zeng, L.; Shen, C.; Ju, Y.; Konuma, T.; Zhao, C.; Vakoc, C. R.; Zhou, M.-M. Structural Mechanism of Transcriptional Regulator NSD3 Recognition by the ET Domain of BRD4. *Structure* **2016**, *24* (7), 1201–1208. https://doi.org/10.1016/j.str.2016.04.019.

(8)   Vougiouklakis, T.; Hamamoto, R.; Nakamura, Y.; Saloura, V. The NSD Family of Protein Methyltransferases in Human Cancer. *Epigenomics* **2015**, *7* (5), 863–874. https://doi.org/10.2217/epi.15.32.

(9)   Shen, C.; Ipsaro, J. J.; Shi, J.; Milazzo, J. A.; Wang, E.; Roe, J.-S.; Suzuki, Y.; Pappin, D. J.; Joshua-Tor, L.; Vakoc, C. R. NSD3-Short Is an Adaptor Protein That Couples BRD4 to the CHD8 Chromatin Remodeler. *Mol. Cell* **2015**, *60* (6), 847–859. https://doi.org/10.1016/j.molcel.2015.10.033.

(10)   Cheng, Y.; Patel, D. J. An Efficient System for Small Protein Expression and Refolding. *Biochem. Biophys. Res. Commun.* **2004**, *317* (2), 401–405. https://doi.org/10.1016/j.bbrc.2004.03.068.

(11)   Johnson, B. A. From Raw Data to Protein Backbone Chemical Shifts Using NMRFx Processing and NMRViewJ Analysis. In *Protein NMR: Methods and Protocols*; Ghose, R.,

Ed.; Methods in Molecular Biology; Springer: New York, NY, 2018; pp 257–310. https://doi.org/10.1007/978-1-4939-7386-6_13.