Crystallization of GSK3β with an inhibiting peptide of psychiatric risk factor DISC1

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master's
In the Department of Veterinary Biomedical Sciences
University of Saskatchewan
Saskatoon

By
NARSIMHA PUJARI

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Veterinary Biomedical Sciences
52 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5B4
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

Disrupted in Schizophrenia 1 (DISC1) is a candidate risk gene in several major mental illnesses, e.g. depression, bipolar disorder, and schizophrenia. The full-length DISC1 protein comprises of 854 amino acids. It is a scaffold protein that interacts with a very large number of other proteins, forming a sizeable protein-protein-interaction network that coordinates various stages of brain development. One of those important interactors is the enzyme, glycogen synthase kinase 3β (GSK3β). As a target for lithium, GSK3β itself is implicated in bipolar disorder. The interaction of DISC1 and GSK3β was discovered at the cross-section of the canonical Wnt/β-catenin signalling which controls the proliferation of neural progenitors. DISC1 specifically inhibits GSK3β's function in this pathway via a direct physical interaction. GSK3β is involved with the Axin-APC complex in this pathway, and its main role is to phosphorylate and regulate the levels of β-catenin in the cell which in turn is a regulator of gene expression levels. The most potent GSK3β inhibitory region has been mapped to a small region in the N-terminus (residue 195-238) of DISC1. This 44-amino acid region (hD1) inhibits GSK3β in an ATP non-competitive mechanism and its binding site partially overlap with that of a peptide from another GSK3β binding protein, FRATide. Knowledge regarding the molecular network of the interactions between hD1, FRATide, and GSK3β remains elusive. This thesis focusses on the effort towards obtaining crystals for the protein complexes GSK3β-hD1 and GSK-hD1-FRATide. Crystals of the putative binary and ternary protein complexes were successfully obtained.

Acknowledgements

I extend my gratitude towards Dr. Adelaine Leung for guiding me through my M.Sc. degree. I thank her for providing me the opportunity to join USask, her faith in me as a student scientist, and the constant push to develop critical thinking. She also helped me build scientific skillsets, and provided guidance for academic ambitions.

I appreciate the suggestions regarding experiments & growth as a graduate student from my committee members Drs. David Sanders and Daniel MacPhee. I would like to thank Drs Stanley Moore, Michel Fodge, and Michal Boniecki for the help with crystallography in PCCF and Synchrotron (CLS).

I thank my former and current lab members Stephanie, Robert, Sean, and Anand, for their significant contribution towards the project. I acknowledge Dara McMunn and Kane Simmans for proofreading this thesis.

Good science is impossible without friends and family. I am grateful to my parents, Aditya Pujari and Ratna Mund, for pushing me toward my dreams. I would like to acknowledge the support I received from my girlfriend, Jade, to become a better person and develop my English proficiency. I appreciate Cassidy for keeping me focussed on thesis writing.

A hearty thanks to Drs. Shreyas and Arinjay for teaching me to how to remain calm through grad school. Thanks to Poonam, Kevin, & Mamata for being the best office friends possible, and for guiding me through the experiments of lab and life. I appreciate the help of our lab manager and friend, Angela. It takes a village to raise a graduate student.

Dedication

I dedicate my thesis to all the graduate student scientists around the world, because their undying curiosity keeps the future of research alive.

“The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them”

- Sir William Lawrence Bragg, recipient of Nobel prize for Physics, 1915 for devising Bragg’s law of X-ray diffraction as a graduate student.)

Table of Contents

List of tables

List of figures

1 Chapter 1: Introduction

1.1 Psychiatric illnesses: DISC1 and its importance in etiology

More than 16% of the world population experienced psychiatric illnesses in 2019[1]. Schizophrenia, autism spectrum disorder, and depression are some of the mental disorders which can severely impact quality of life and pose a significant economic burden. Complex genetic factors, variation in the etiology across different groups of the population, and environmental effects raise significant challenges in understanding the molecular basis of psychiatric diseases. Defects in neurodevelopmental processes have been found to be at the central stage of etiology for the aforementioned psychiatric illnesses[2].

Genetic inheritance of schizophrenia and related illnesses was established via twin and adoption studies as well as genetic analysis of large pedigrees[3]. The need to determine genetic factors for psychiatric diseases led to an extensive search by the scientists for the different genes involved, their mutations, and the study of altered functions. A large Scottish family with a high prevalence of a variety of psychiatric illnesses, such as schizophrenia, major depressive disorder, and autism, was investigated to uncover the critical genetic factor. Two genes, *Disrupted in Schizophrenia 1 & 2 (DISC1/2)* were discovered to have a balanced translocation (1; 11) (q42; q14.3) that co-segregated with major psychiatric illnesses in the family[4,5]. Multiple SNPs and haplotypes of DISC1 have also been linked to segregation of autism spectrum disorders and asperger syndrome in Finnish families[4]. Association of the DISC1 protein with a high incidence of multiple neuropsychiatric disorders garnered the interest of researchers to understand its structure, function, and interaction.

Further research using tissue cell culture, animal models, and human brain samples of schizophrenic patients provided evidence of DISC1's biological role in different stages of neurodevelopment. "DISCopathies" is a term used to categorize mental disorders resulting from the malfunctioned DISC1 protein[6]. The protein has been found at the convergence of multiple critical neuro-developmental pathways. The pathways involved are responsible for a diverse array

of functions in the brain such as neural progenitor proliferation, neural plasticity, memory, neural migration, and maturation[7]. The multifunctional nature of DISC1 can be attributed to the structural ability of the protein to interact with many different protein partners[8].

1.2 DISC1: Structure, function, and its protein interactome

A DISC1 interactome was established using a yeast two-hybrid system (Y2H), and it identified potential cellular pathways based on the function of the interactors[9,10]. DISC1 interactome analysis determined an array of proteins from the human fetus and adult brain libraries to potentially interact with DISC1. The research implied a strong involvement of the protein in neuronal migration, progenitor proliferation, and NMDA receptor activation (glutamate receptors)[9]. A multi-faceted functional capacity of the protein makes it a distinct and ideal target for novel therapeutics aimed at psychiatric illnesses. Novel therapy designing would succeed when DISC1 and a specific interacting partner are targeted from a particular pathophysiology. This process, however, has a significant hurdle. The three-dimensional atomic model of DISC1 is unknown. The availability of a high-resolution molecular structure of DISC1 will give crucial insight into its function.

The lack of structural data for DISC1 impedes our understanding of how mutations and SNPs give rise to pathophysiology. Sequence analysis has been used by scientists to understand the nature of different regions on DISC1 protein. These findings may provide leads to purify specific regions of the protein along with the concerned interacting partner. The *DISC1* gene contains 13 exons. Many splice variants of the protein exist. The longest version, DISC1-L-isoform, contains 854 amino acids with a molecular weight of ~100 kDa. Sequence analysis reveals two major distinct regions of human DISC1: an N-terminal region (1-350) and a C-terminal region (350-854). Sequence alignment and comprehensive bioinformatics studies show more significant conservation in the C-terminal region when compared with the N-terminus[5,11–13].

1.2.1 N-terminal disordered region

Only a few notable conserved regions are recorded in the N-terminus of the protein: nuclear localization signal (NLS) and serine-phenylalanine-rich (SF-rich) motifs (Fig 1-1)[12,13]. The N-terminal region (1-350 amino acid) of the protein was previously recognized as a globular domain, which means it is natively folded[5,12]. However, an improvement of the sequence analysis algorithm that considers both secondary structure and disorder sequence signature concurrently suggests that

the N-terminus of DISC1 may be unstructured[14,15]. Intrinsically disordered proteins carry a distinct sequence identity, where the majority of the amino acids are highly charged and hydrophilic. These amino acids (e.g. Glu, Asp, Lys, Ser, Pro, Arg, etc.) constitute 60% of the N-terminus in DISC1[7]. DISC1 has a ~15% serine and ~23% alanine and glycine content in the N-terminal region[12]. This analysis strongly suggested an intrinsically disordered nature of the N-terminus with very low complexity[14]. Updated sequence analysis revealed a ~65% disordered content in the N-terminus. Five intrinsically disordered portions were predicted: (i) 1-35 (ii) 65-95 (iii) 146-205 (iv) 218-260 (v) 278-322 amino acid residues[7] (Fig 1-1). This result gave the first insight into the intrinsic disorder of the N-terminus of DISC1.

A significant percentage (~33%) of the eukaryotic proteome is partially or entirely disordered, most of the disordered protein being at the intersection of cellular regulations and signalling[16]. One theory suggests the ability of disordered regions to be involved in a plethora of signalling pathways. The disordered regions can transition into more folded structures whenever they bind to a protein target. Multiple structural transitions allow the protein to be able to interact with many signalling cascades. This could be the reason why DISC1 interactome involves more than 200 diverse proteins[9]. Another theory suggests the availability of a substantial intermolecular interaction surface when a disordered region is involved[15].

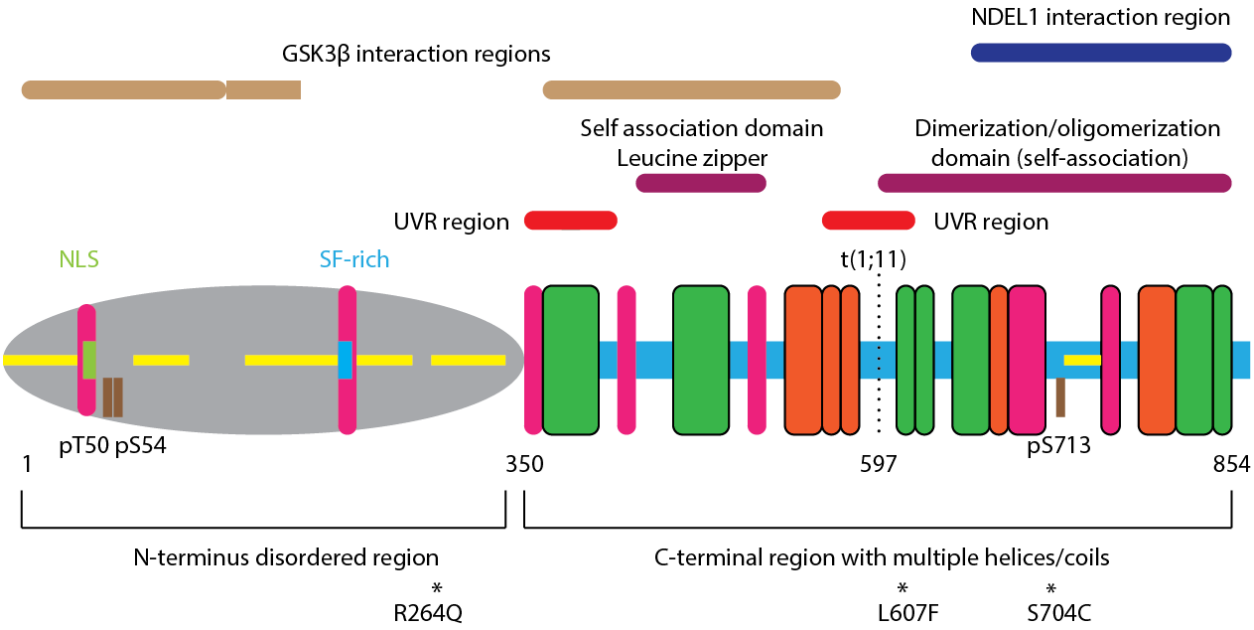


Figure 1-1. Biophysical characterization of DISC1. Sequence analysis and experimental data suggest an N-terminal disordered region (grey, 1-350 residues). Yellow horizontal bars represent five major disordered sites. Nuclear localization signal (NLS) and Serine-phenylalanine (SF) rich domains are shown in green and blue small vertical bars inside N-terminus. C-terminus (blue horizontal, 350-854) is comprised of many α-helices and coiled-coil regions. Pink vertical bars represent regular α-helices, green vertical bars show coiled-coil helices, orange bars represent regular and coiled-coil helices[7]. The typical translocation break from the Scottish pedigree is seen at position 597[4]. Mutations marked with stars are common SNPs found in human DISC1. Anti-parallel α-helices such as UVR-like regions are predicted with red horizontal bars (in C-terminus)[17]. Multiple self-association sites, dimerization, and oligomerization sites are highlighted with purple bars[18]. The light brown bars indicate interaction sites for GSK3β, the rectangular section (contrary to smooth end bars of the same colour) containing the most potent interacting and inhibiting region (hD1)[19]. Horizontal dark blue colour represents the minimum necessary interaction site for NDEL1[20].

1.2.2 C-terminal region with self-association and oligomerization sites

The C-terminus of DISC1 is defined by multiple heptad repeats with a pattern of "abcdefg" amino acids where the 4th position is often a non-polar residue. Few of these repeats (helices) get intertwined to provide stability via supercoil formation[21]. DISC1 contains at least four regions with the structural potential of a coiled-coil region: residues (i) 347-393, (ii) 452-499, (iii) 603-680, and (iv) 805-828[7]. Coiled-coil regions have been associated with protein-protein interaction surface and oligomerization potential which could be the case for DISC1. Besides these regions, analyses by multiple programs suggest at least five α-helical regions in the C-terminus.

Analysis of the purified full-length DISC1 protein by dynamic light scattering (DLS) and size-exclusion chromatography initially suggested the possible existence of a dimer or a higher-order structure in mammalian cell lines[22]. Brandon *et al.* observed DISC1 with a molecular weight of ~250 kDa, suggesting at least a dimerized state combined with other possible oligomeric states. Oligomerization of proteins is often determined via self-interacting domains. A self-interacting domain (residue 403-504) containing a leucine zipper (455-495) was discovered via deletion studies[23]. Oligomerization is often aided by the presence of a leucine zipper (Fig 1-1).

The self-association domain described above is flanked by two UVR domains. The UVR domains (found across many species) are capable of self-association while often being part of the protein-protein interaction interfaces. They are characterized by anti-parallelly packed α-helices connected via a left-handed twist[24]. The UVR-like regions in DISC1 are predicted to be between residues 343-394 and 574-625[17].

Other sites necessary for the oligomerization states of DISC1 were discovered while the NDEL1 (Nuclear distribution element 1)-DISC1 interaction sites were being investigated. A DISC1 fragment with residues from 598-854 was observed with the simultaneous presence of dimeric, octameric, and oligomeric states, and only the species with the octameric state was able to interact with NDEL1[20]. These findings suggest a possible adaptive mechanism of oligomerization for specific interactions in the cell. Further analysis of DISC1 segments with multiple lengths, provided proof of another coiled-coil region (765-854), while a stretch of residues from 668 to 747 was responsible for multimerization of DISC1 (Fig 1-1)[18].

1.2.3 Studying DISC1-protein interaction framework

The large size of DISC1 and the intrinsic disorder in the N-terminus region are possibly the key reasons that impede expression/purification and crystallization of the full-length protein. The sequence analysis of disordered regions can be used to design constructs that are more ordered[25]. Another option is to co-express the protein with an interaction partner. Out of more than 200 interaction partners, the focus of our group is the interaction between the kinase GSK3β and DISC1.

1.3 Discovery of DISC1's interaction with the protein kinase GSK3β

Since the discovery of DISC1's implication in many neurodevelopmental defects, researchers have investigated the structure and function of the protein. Mao *et al*.[19] used DISC1 transgenic mouse lines to characterize its expression and function. High level of DISC1 expression was observed in nestin- and Sox2-positive neural progenitors present in the ventricular and subventricular zone of the cerebral cortex in the embryo. These zones are actively involved in neurogenesis. In the adult mouse brain, DISC1 was found to be highly expressed in the dentate gyrus and the olfactory bulbs, both regions that can still undergo neurogenesis.

Therefore, neural progenitor proliferation became a mechanism of interest where DISC1 might be actively engaged[19]. This hypothesis was tested both *in vitro* (AHP cells) and *in utero* (embryonic mouse brains- E13/14/15). Cell proliferation rates were significantly affected by a knockdown of DISC1. Progenitor proliferation halted, and early cell cycle exit was observed in the absence of DISC1. This set of experiments strongly suggested that DISC1 plays a decisive role in progenitor proliferation and that knockdown of DISC1 potentially leads to premature cell cycle exit and differentiation.

Wnt/β-catenin canonical signalling plays a significant role in regulating progenitor proliferation in the nervous system. An apparent correlation was investigated between canonical Wnt signalling and the function of DISC1 in cell proliferation. Mao *et al*.[19] observed an increase in proliferation rate in AHP culture when they added Wnt3a, and DISC1 knockdown neutralized this effect. Conclusions from these experiments were further strengthened by the investigation of the effect of DISC1 knockdown on transcription factors and target genes involved in Wnt/β-Catenin signalling. Both gain-of-function and loss-of-function outcomes of DISC1 were tested, and they were found to have a positive and negative effect of neural progenitor proliferation, respectively[19].

DISC1 knockdown caused defects in progenitor proliferation. However, β-catenin overexpression was able to correct the defects, suggesting DISC1's modulating effect on β-catenin. β-catenin levels are known to be regulated by phosphorylation which is orchestrated by the kinase glycogen synthase kinase 3β (GSK3β). GSK3β phosphorylates the Ser33/37 and Thr41 residues of β-catenin. Phosphorylated β-catenin is degraded via the ubiquitin-proteasome pathway[26,27].

Observation of a decrease in the level of the aforementioned phosphorylation after DISC1 knockdown raised the possibility of GSK3β involvement.

Understanding protein-protein interaction was the next necessary step to solve the puzzle of DISC1's role in progenitor proliferation. Direct binding of DISC1 and GSK3β was established for a physiological interaction. Incubation with DISC1 fragment spanning peptide residues 1-220 inhibited the phosphorylation of β-Catenin. This DISC1 fragment was further scrutinized for potentially independent conserved regions. Out of all of the shorter peptides tested, DISCtide-1 (spanning 195-238 on mDISC1) was able to inhibit GSK3β. Determining the 44 amino acid long DISCtide-1 was a crucial discovery in understanding DISC1's specific inhibition of kinase towards the phosphorylation of β-catenin[19].

Investigation performed by Mao *et al.*[19] provides strong evidence of direct physical interaction between DISC1 and GSK3β in regulating neural progenitor proliferation. They determined the most potent inhibitory region on mDISC1 to be amino acids spanning from 195-238 (and 193-236 in hDISC1, referred to as hD1) (Fig 1-2). The exact nature of interaction and the details of the GSK3β inhibition mechanism, however, remain elusive. DISC1's direct and specific inhibition of GSK3β holds the potential for novel therapeutic research targeted for pathological conditions due to genetic mutations disrupting DISC1- GSK3β interaction. An essential criterion for any therapeutic research is to solve the detailed molecular framework involved in the protein-protein interaction. Our lab aims to understand the molecular framework required between DISC1 and GSK3β interaction using crystallography.

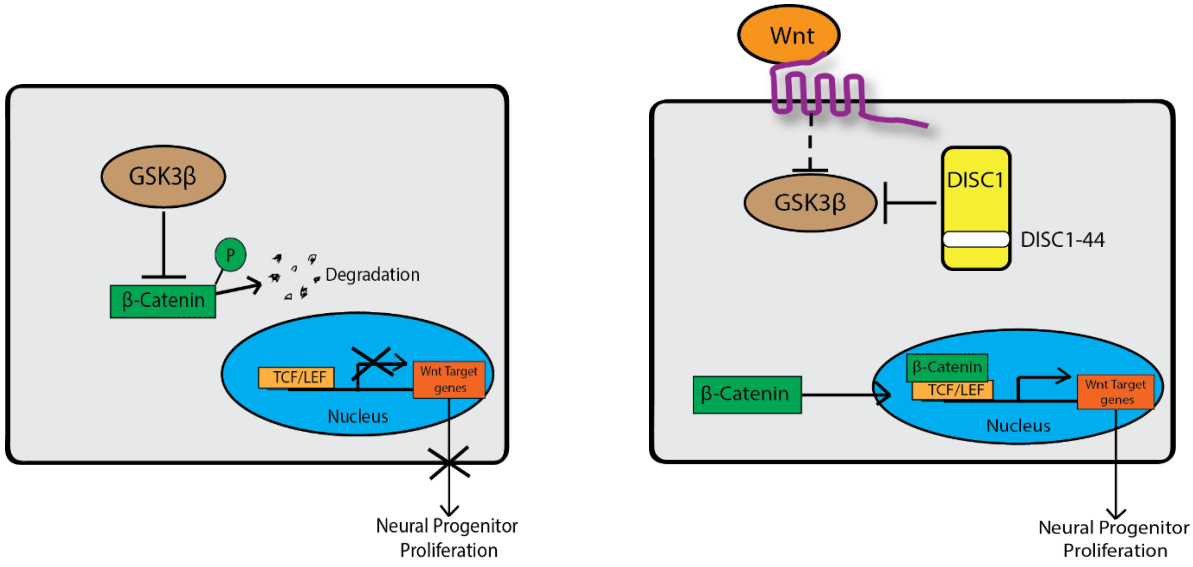


Figure 1-2. Neural progenitor proliferation and involvement of DISC1 in canonical Wnt/β-catenin signalling. β-catenin is responsible for activating genes for neural progenitor proliferation after nuclear translocation. The picture on the left shows phosphorylation of β-catenin by GSK3β and degradation, limiting the rate of proliferation. The image on the right shows an active Wnt/β-catenin signalling pathway where DISC1 acts as one of the regulators of GSK3β inhibition. A 44 amino acid extended region in the N-terminal of DISC1 (hD1) is shown as the most potent inhibitor of GSK3β by Mao *et al*[19].

1.4 GSK3β: Structural basis of function, regulation, and relevance to Wnt-signalling

Glycogen synthase kinase 3β (GSK3β) was initially discovered to be one of the critical regulators for glycogen metabolism in the liver. Insulin triggers the glycogen deposition via initiation of dephosphorylation and activation of glycogen synthase[28]. The deactivation of glycogen synthase via phosphorylation is carried out by the GSK3β[29]. Once thought to have an important but narrow role in insulin mechanism, GSK3β was eventually established as a kinase involved with perhaps the most diverse cellular pathways. In addition to β-catenin, almost 100 other proteins have been experimentally implicated as GSK3β substrates to date, and theoretical kinase analysis predicted that the kinase could hypothetically have as many as 500 substrates[30,31].

The *GSK3β* gene is conserved across species ranging from fungi and flies to mice and humans[32,33]. Extensive research into the role of this kinase revealed its involvement in an array of pathways other than the regulation of blood glucose. These pathways are directly linked to neuronal progenitor proliferation, cell proliferation, migration, etc., and have been implicated in a

multitude of diseases such as psychiatric illnesses, glucose metabolism disorders, and cancers[34]. Established as a regulator in many cell signalling pathways, GSK3β has been regarded as a classic therapeutic target. The kinase's ubiquitous nature, however, poses a challenge for targeted therapeutics design. This problem can be circumvented by investigating regulation of GSK3β in a specific pathway.

1.4.1 Fundamental regulatory mechanisms of GSK3β

Three main regulation mechanisms are known to control the GSK3β kinase's activity, each type being used by specific pathways. They are: (i) substrate preference for primed substrate, (ii) different phosphorylation states of the protein, and (iii) subcellular compartmentalization of the protein to different sites in the cell. Each of these regulatory functions is discussed below.

1.4.1.1 The kinase activity of GSK3β is often specific towards a primed phosphorylated substrate

GSK3β shows a significant preference specifically towards primed phosphorylated substrates[35–37]. The priming phosphorylation on GSK3β substrates happens on $(P+4)^{th}$ residue where P is the serine/threonine residue to be phosphorylated by GSK3β, and $(P+4)^{th}$ residue is a serine already phosphorylated by another enzyme prior to the GSK3β phosphorylation (Fig 1-3). The sequence of the primed phosphorylation site can be written as Ser/Thr-XXX-Ser(p), where the N-terminal Ser/Thr site is GSK3β target, and Ser(p) is primed phosphorylated site by another enzyme. This sequence specificity and preference is probably an evolutionary advantage that puts GSK3β at the intersection of major regulatory pathways. In Wnt/β-catenin signalling, priming phosphorylation of β-catenin is accomplished by casein kinase Iα (CKIα) that phosphorylates β-catenin at Ser45. Then, GSK3β further phosphorylates the primed β-catenin at 41, 37, and 33 residues, marking it for ubiquitin-mediated proteolysis[38–40]. The sequence specificity towards a primed substrate is somewhat flexible. Some substrates with acidic/anionic residue instead of the primed phosphorylated group can also be processed as GSK3β substrates. The absence of priming phosphorylation from most of the GSK3β substrates has been shown to decrease the enzyme efficacy by ~100-1000 times[41].

1.4.1.2 Regulation by different phosphorylation states of GSK3β

GSK3β activity is modulated by two phosphorylation sites. Tyr216P aids to activate while Ser9P inhibits the enzyme. Tyr216P was found to be auto-phosphorylated under active conditions[42,43]. Recent research has suggested that Y216 phosphorylation may not be absolutely essential for activity as unphosphorylated GSK3β structure in an active conformation has been obtained by crystallography[44]. Nonetheless, the phosphorylated tyrosine residue plays a significant part in the stabilization of the activation loop[45]. The decrease in GSK3β activity by 5-10 fold has been noted by mutating Tyr216 to a non-phosphorylatable residue[46]. This suggests only a modest stimulation provided by Tyr216P, contrary to a ~1000-fold stimulation seen in other kinases.

A crucial negative regulation of the kinase is brought about by phosphorylation of Ser9 residue in the N-terminus. This regulation was discovered to be used by the insulin mechanism via PKB/Akt[47,48]. Prephosphorylated primed substrates were found to interact with Arg96 in the substrate binding cleft of the kinase[44]. Site specific mutations and crystallographic models proved that the phosphorylated Ser9 of the N-terminal segment binds to the Arg96 residue of the substrate-binding cleft in GSK3β, posing as a pseudo-substrate (Fig 1-3)[49,44]. Binding of the pseudo-substrate blocks the access of primed substrates to the substrate binding cleft, auto-inhibiting the kinase.

Research published by Mao *et al.*[19] showed DISC1 knockdown resulted in a significant increase in Tyr216P in GSK3β. Ser9P levels in the kinase were not affected. The overexpression of DISC1 led to a substantial decrease in Tyr216 phosphorylation in AHP.

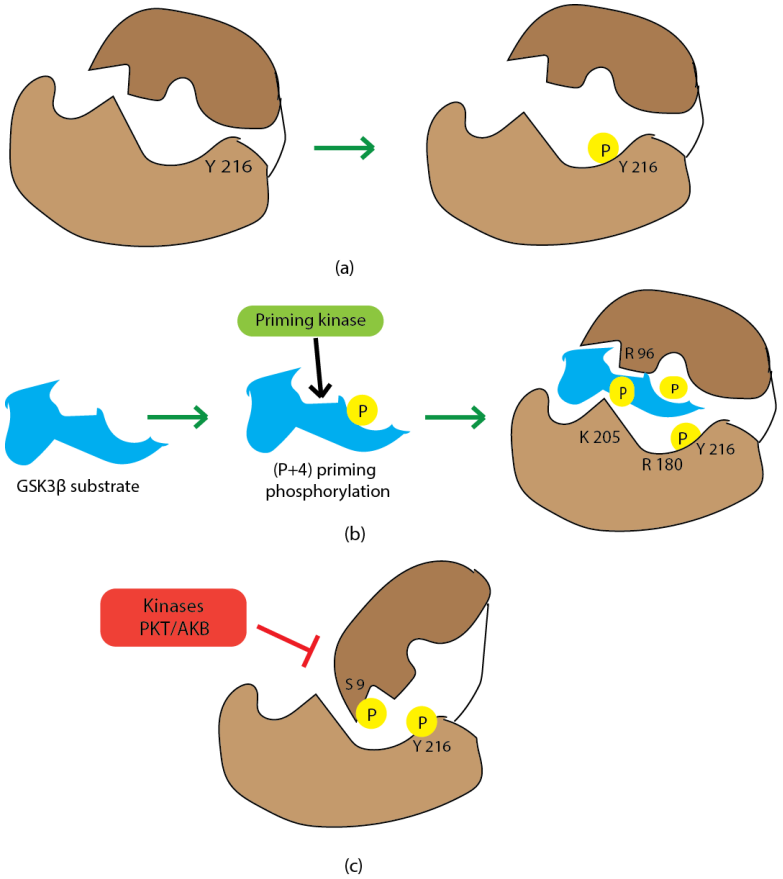


Figure 1-3. Substrate specificity and regulatory mechanisms of GSK3β. (a) Schematic representation shows the possible conformational change in the substrate-binding pocket after Tyr 216 residue is phosphorylated. This could explain some level of increase in the binding efficiency of primed substrates to the kinase with the Tyr216P. (b) Serine/threonine kinase GSK3β mostly shows specificity towards the primed substrates where priming phosphorylation by another kinase has taken place at (P+4) serine residue, where P is the site of phosphorylation by GSK3β. Three basic residues, Arg96, Arg180, and Lys205, show interaction towards the primed phosphorylated site and help stabilize the substrate in substrate binding pocket. (c) While Tyr216P has been associated with active conformation of the kinase, a negative regulation via phosphorylation of Ser 9 residue also exists. Ser9P leads to the serine residue acting as a pseudo-substrate towards the substrate-binding pocket, blocking the access of other substrates. The image design is based on a model presented by Patel et. al[50].

1.4.1.3 Regulation by subcellular compartmentalization

Subcellular compartmentalization is another important regulatory mechanism of GSK3β[51]. By associating with different complexes, the activity of GSK3β is restricted to substrates residing in these complexes. As will be discussed further below, subcellular compartmentalization is a critical regulatory mechanism of GSK3β in the Wnt signalling pathway.

1.4.2 GSK3β at the molecular intersection of Wnt/β-catenin signalling

A *GSK3β* homolog in *Drosophila* (known as Zeste-white3 or Shaggy) was implicated as a critical component of Wingless Wg/Wnt signal transduction[52]. Canonical Wnt-signalling pathway in humans is a major pathway involved in cell growth, differentiation, and migration, especially in the brain[53]. Humans have genes responsible for 18 different Wnt molecules, all of which contain conserved cysteine residues and glycosylation as part of the post-translational modification (molecular weight range between 39 to 46 kDa)[53]. This pathway involves activation of a signalling cascade by Wnt that leads to stabilization and nuclear localization of β-catenin, a co-activator of TCF/LEF family of transcription factors. Activation of T-cell factor/lymphoid enhancer-binding factor (TCF/LEF)-bound genes initiate neural progenitor proliferation[54].

In the absence of Wnt signalling molecules, ubiquitous GSK3β overphosphorylates the N-terminal of β-catenin. β-catenin phosphorylation leads to ubiquitin-mediated proteasomal degradation, thus stopping the nuclear localization of the molecule and halting the neuronal progenitor proliferation[26]. Destruction of β-catenin via phosphorylation is a very complicated process. It is carried out by subcellular compartmentalization of a complex known as "β-catenin destruction complex," and it comprises scaffold proteins such as Adenomatous polyposis coli gene product (APC) and Axin[55,56]. In the absence of Wnt signalling molecules, these two scaffold proteins orchestrate a large protein complex *via* direct interaction with Casein Kinase 1α (CK1α), GSK3β, and β-catenin. The two kinases of this Axin-APC-orchestrated β-catenin phosphorylation and destruction processes are CKIα and GSK3β[40,57,58]. As described earlier, CK1α performs priming phosphorylation on β-catenin at Serine 45 residue, prompting GSK3β phosphorylation at residues 41, 37, and 33. E3 ubiquitin ligase β-TrCP recognizes phosphorylated β-catenin and marks it for ubiquitination and proteasomal degradation[26,59,60]. The stage of β-catenin destruction complex set by APC-Axin could be a reason why GSK3β is such a widely expressed enzyme with little or no obstruction of one pathway with the other (Fig 1-4).

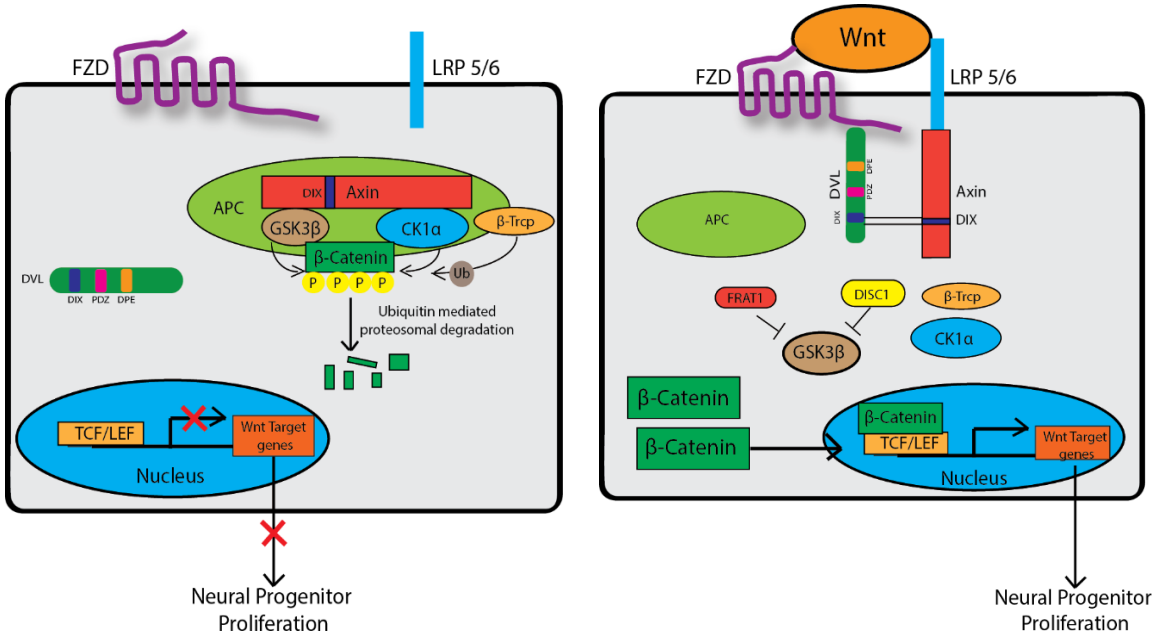
In the presence of Wnt signalling molecules bound to their receptors, the β-catenin destruction complex is disassembled. As a result, CKIα and GSK3β no longer phosphorylate β-catenin. Unphosphorylated β-catenin is stable, and it accumulates in the cytoplasm. Excess β-catenin in the cytoplasm translocates to the nucleus and acts as a co-activator of genes responsible for cell proliferation. The canonical Wnt/β-catenin signalling pathway is a very intricately designed complex pathway. The exact cascade involved in disassembling the Axin-APC complex is still obscure, although there are some established vital players (Fig 1-4).

Wnt molecules interact with two sets of co-receptors during active Wnt-signalling: (a) the frizzled receptor (FZD) that spans the plasma membrane seven times and (b) the low-density lipoprotein receptor-related protein 5 or 6 (LRP5/6) that spans the plasma membrane only once[61–66]. Binding of Wnt molecule to these two receptors leads to activation and sequestration of downstream effectors. A critical effector of the pathway is Dishevelled (DVL 1-3 in humans). Its mutation in animal models was lethal as their loss severely hampered Wnt signalling[67]. DVL (~700 amino acids) serves as a hub via interaction of its different domains with different proteins. Three important domains (~80-90 amino acids each) of DVL are DIX (Dishevelled, Axin), PDZ (Postsynaptic density 95, discs large, zona occludens-1), and DPE (Dishevelled, Egl-10, Pleckstrin). The carboxy-cytosolic region of FZD receptor sequesters the PDZ domain of DVL when Wnt signalling is active[68,69]. Recent developments show the affinity of DPE domains for the phospholipids of the plasma membrane as well as for some protein interactions. The DIX domain of DVL is capable of forming homo-dimers/oligomers as well as hetero-dimers with Axin, which also has a DIX domain for interaction[70–72]. (Fig 1-4).

The hetero-dimers of DVL-Axin *via* interaction of their DIX (DAX in Axin) domains could be one of the many ways how Wnt-signalling disrupts the β-catenin destruction complex. Axin acts as one of the critical scaffolds in the absence of Wnt and provides a stage for GSK3β and CKIα for over phosphorylation of β-catenin. Sequestration and isolation of Axin to the plasma membrane by DVL may be interrupting the physiological proximity required by the destruction of complex partner proteins.

The PPPSxS motif on LRP6 is also known to be phosphorylated by GSK3β and CKIα in the presence of Wnt molecules. The phosphorylated LRP6 motif in the cytoplasm is a docking site for Axin[73–76]. Binding of Axin to the phosphorylated LRP 5/6 cytosolic PPPSxS motif and DVL

mediated Axin dimerization could be happening simultaneously, according to a classic biochemical model[77]. The destruction complex clogged model suggests transport of the whole Axin-APC destruction complex to the plasma membrane where β-catenin is still loaded but trapped in the complex. This could lead to stabilization of new β-catenin in the cytosol despite undisturbed integrity of the β-catenin destruction complex in the cell in the presence of Wnt[78]. The Axin auto-inhibition model focusses on the dephosphorylation of Axin in the wake of active Wnt/β-catenin signalling by Protein Phosphatase I (PP1). Axin is phosphorylated by GSK3β in canonical Wnt/β-catenin signalling. However, a correlation has been established between β-catenin stabilization, Axin dephosphorylation by PP1 in the presence of Wnt, and Axin phosphorylation in the absence of Wnt. The Axin auto-inhibition model suggests the inability of dephosphorylated Axin to sustain the β-catenin destruction complex, thus stabilizing the protein when Wnt/β-catenin signalling is active[40,79–82].



(a) Wnt OFF (b) Wnt ON

Figure 1-4. Canonical Wnt/β-catenin signalling pathway. (a) The picture on the left describes the overall mechanism of neural progenitor proliferation in the absence of Wnt/β-catenin signalling molecules. A β-catenin destruction complex via scaffold proteins such as APC and Axin is arranged. The destruction complex uses the kinase function of GSK3β to over phosphorylate β-catenin, once CK1α primes it. The phosphorylated β-catenin gets degraded via ubiquitin-mediated proteasome using βTrcp. (b) Schematics on the right describe the positive regulation of progenitor proliferation once Wnt binds to the co-receptors Frizzled (FZD) and LRP 5/6. Multiple theories of disassembly of β-catenin destruction complex have been suggested. Possibly, the interaction of Dishevelled (DVL) with FZD, Axin heterodimerization with DVL via DIX domain, and Axin sequestration via LRP 5/6 could be disrupting the destruction complex. This leads to cytosolic accumulation and nuclear accumulation of β-catenin, thus switching on genes for proliferation.

1.4.3 Emergence of FRAT1 as a potential regulator of GSK3β in Wnt signalling

Cytoplasmic complexity of the canonical Wnt/β-catenin signalling has been long studied. Involvement of multitude of factors makes it difficult to announce a complete model of GSK3β inhibition and β-catenin stabilization during the active state of Wnt-signalling. Discovery of a new causative gene *Frequently rearranged in T-cell lymphomas 1* (FRAT1) from mouse T-cell lymphomas led to a series of investigations that established the Frat1 protein to be involved in the dissociation of Axin-APC β-catenin destruction complex[83]. The Frat1 homologous gene/protein in Xenopus called GSK-3 binding protein (GBP) was identified as one of the key positive regulators in Wnt-dependent regulation of β-catenin stability[84]. A detailed investigation of the physiological placement of Frat1 in the dissociation process of Wnt-dependent β-catenin destruction complex

revealed the presence of a quaternary complex involving GSK3β, Axin, DVL as well as Frat1[85]. This quaternary complex was comparatively less stable than the Axin-APC destruction complex. Li *et al*.[85] established a more robust and stable complex formed by just Frat1 and GSK3β. Later studies have strongly suggested a mutually exclusive complex formation of Axin-GSK3β and Frat1-GSK3β, which may be counteracting each other's role in β-catenin stabilization[86].

Interaction of Frat1 and GSK3β garnered a lot of attention regarding its peculiar inhibition of the kinase. *In vitro* studies revealed that an N-terminal peptide spanning residues 188-226 on Frat1 (referred to as FRATide) was able to inhibit GSK3β interaction and phosphorylation with Axin and β-catenin. They also identified this interaction to be very specific towards the canonical β-catenin stabilization substrates. FRATide was unable to inhibit GSK3β kinase action towards common substrates such as Glycogen Synthase (GS)[41]. These findings repeatedly establish the presence of a specific GSK3β inhibition mechanism in place for canonical Wnt/β-catenin signalling.

1.4.4 Literature review of relevant crystal structures of GSK3β

1.4.4.1 Overall structure

The serine-threonine kinase GSK3β maintains a usual protein kinase structure, an N-terminal β-sheet domain and a C- terminal α helical core[87]. The *GSK3β* gene is present in human chromosome 3 and has 11 exons that translate to a protein with molecular weight ~47 kDa. The N-terminal β-strand domain, also known as β-barrel due to its unique arrangement, spans from residues 25-139[45]. The C-terminal α-helical core domain lies between residues 139-343. The C-terminal residues from 344 to 382 contain small α-helices and lies outside the core kinase domain. Amino acids from 1-25 and 382-420 are thought to be disordered for structures obtained from crystallographic analysis[88].

The N-terminal segment of the kinase is mainly comprised of seven anti-parallel β-strands, where the strands 2-6 form a β-barrel[44]. A short α-helix (residue 94-104) (C-loop) connects β-strands 5 and 6 in the N-terminus. Arg96 of the C-loop is an essential part of the catalytic domain that helps position the primed substrate. The α-helical core domain retains structural integrity similar to other MAP kinases[44]. Residues following 343 until 384 contain secondary structures that include a β-turn from 350-353 and short α-helices between 363-368 and 373-378. These structures are packed against the large α-helix formed by 155-175 residues of the core domain[44,45,88].

1.4.4.2 Structural details of GSK3β involved in kinase activity and substrate specificity

The crystal structure of Tyr216P-containing GSK3β revealed an interaction of the tyrosine residue with Arg220 and Arg223. Tyr216P in GSK3β has been compared to Tyr185P in ERK2, another kinase of the same class. However, the tyrosine residue in GSK3β shows a broader opening of substrate binding groove which could be responsible for a larger number of substrates of the kinase and the stabilization of the substrate-binding pocket (Fig 1-5)[45].

The kinase activity of GSK3β would solely rely on careful positioning of catalytic groups and substrate so that there are energy and stereo-favourable transfers of ϒ-phosphate from ATP to Ser/Thr of the substrate side chain[89]. Most of the protein kinases become active due to interaction between the basic residues of the N-terminal and the catalytic groups with the phosphate ion of the activation segment. Three basic residues, Arg96, Arg180, and Lys205, were determined to bind to the primed phosphorylated site of the substrate[44,45,88]. This was further proven by the inability of the kinase to phosphorylate primed substrates when Arg96 was mutated to an alanine[49]. Thus, these three residues coordinate the phosphorylated primed serine of the substrate and stabilize the substrate binding. These sites would also interact with Ser9P during inhibitory regulation and block the entry of primed substrates. GSK3β aligns both of its domains using primed phosphorylated serine of the substrate and its interaction with the basic residues (Arg96/180, Lys205) (Fig 1-5). ATP binding region of the kinase is bordered with the glycine-rich loop and the hinge[88]. GSK3β structures also revealed an interaction of the anionic group with -NH backbone group of Val214 along with the Tyr216 in the activation loop. This loop (residues 200-226) outlines the activation loop. Once phosphorylated, Tyr216 residue undergoes a sizeable conformational change so that the side chain is exposed to other amino acids. The Tyr216P position also interacts with catalytic residues Asp181, Arg220, and Arg 223[45].

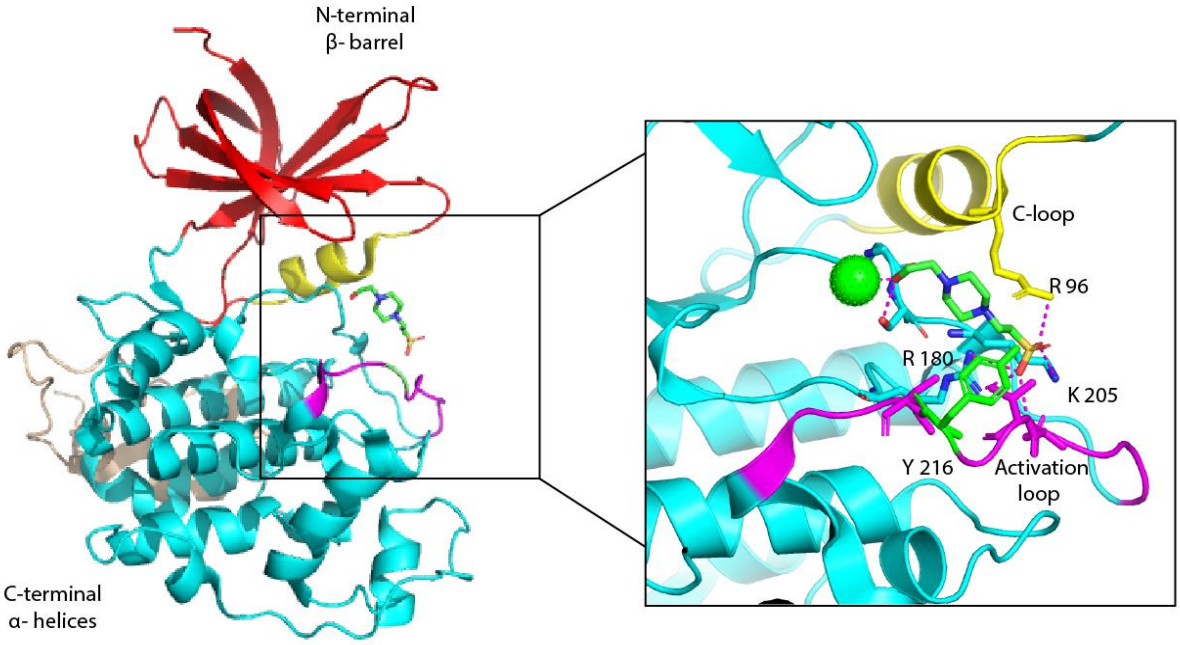


Figure 1-5. Three-dimensional structure of GSK3β. The cartoon model of the kinase is derived from the PDB model 1H8F. The overall structure of the protein shows a clear distinction of N-terminal β-barrel (coloured red), C-terminal α-helical domain (coloured cyan), and a terminal chain of small helices in the back (coloured wheat). The box shows the enlarged structure of the substrate-binding domain and activation loop (magenta, residues 210-220) that is sandwiched between N- and C-terminus. Arg96 from C-loop (yellow) from N-terminal and two basic residues from C-terminus (Arg180 and Lys205, one of them from the activation loop) show direct interaction with the sulfonate anion present (in the substrate-binding pocket) in the structure above. These residues interact with the primed phosphorylation site on the substrate. A polar interaction with a water molecule (green ball) is also seen in the pocket. Tyrosine at 216 position is unphosphorylated and is marked as the green residue in the middle of the magenta activation loop**.**

1.4.4.3 The structural framework involved in Axin GID-GSK3β

GSK3β directly interacts with Axin and is involved in the formation of Axin-APC complex. Binding of the GSK3β to Axin-APC scaffold has shown a ~20,000-fold increase in β-catenin phosphorylation. Detailed interaction between a minimal GSK3β binding region on Axin (residues 383-401, known as Axin-GID) and GSK3β is discussed in this section. Genetic and biochemical studies suggested overlapping binding sites for FRAtide and Axin-GID on GSK3β. A comparison of the binding sites and types of interaction for both FRATide- GSK3β and AxinGID- GSK3β complexes is discussed in the next section.

The Axin-GID forms a single α-helix that comprises a mixture of both hydrophobic and hydrophilic residues (Fig 1-6)[46]. Unlike FRATide, Axin-GID forms a single helix without a break. The interaction framework involves a hydrophobic surface channel that involves two main regions on GSK3β, (i) an α-helix in C-terminal between 262-273 and (ii) a loop between 285-299. One side of the hydrophobic groove comprises of hydrophobic interaction between Axin residues Phe388, Leu392, Leu396, and Val399 and GSK3β residues Val263, Leu266, Val267, Ile270 (the α-helix). Interactions on the opposite wall involve Axin residues Pro385, Ala389, Ile393, and Leu396 packed with GSK3β residues Tyr288, Phe291, Pro294, and Ile296. Although most of the interactions between Axin-GID and GSK3β are hydrophobic, some polar interactions play a key role as well. Hydrogen bond and ionic interaction are observed between Arg395 of Axin-GID and Asp264 of GSK3β. A hydrogen bond is formed between Gln400 of Axin-GID and Gln295 of GSK3β. Axin binding to the C-terminal of GSK3β also prevents dimerization of the kinase under micro-molar concentration[46].

The discussion above shows a common interaction region on GSK3β by both FRATide and Axin-GID, the α-helix (262-273) and the loop (285-299) in the C-terminal. One of the major differences between FRATide and Axin-GID is the presence of a sharp turn with a break in the former. FRATide shows a sharp turn between the residues Gly210 and Asn211 (Fig 1-7). This allows a hydrogen bond between FRATide and Tyr288 and Glu290 of GSK3β. Absence of this break in Axin-GID results in a hydrophobic interaction instead. This interaction is observed between phenol of Tyr288 of GSK3β and Pro385, Phe388 of Axin peptide (Fig 1-6).

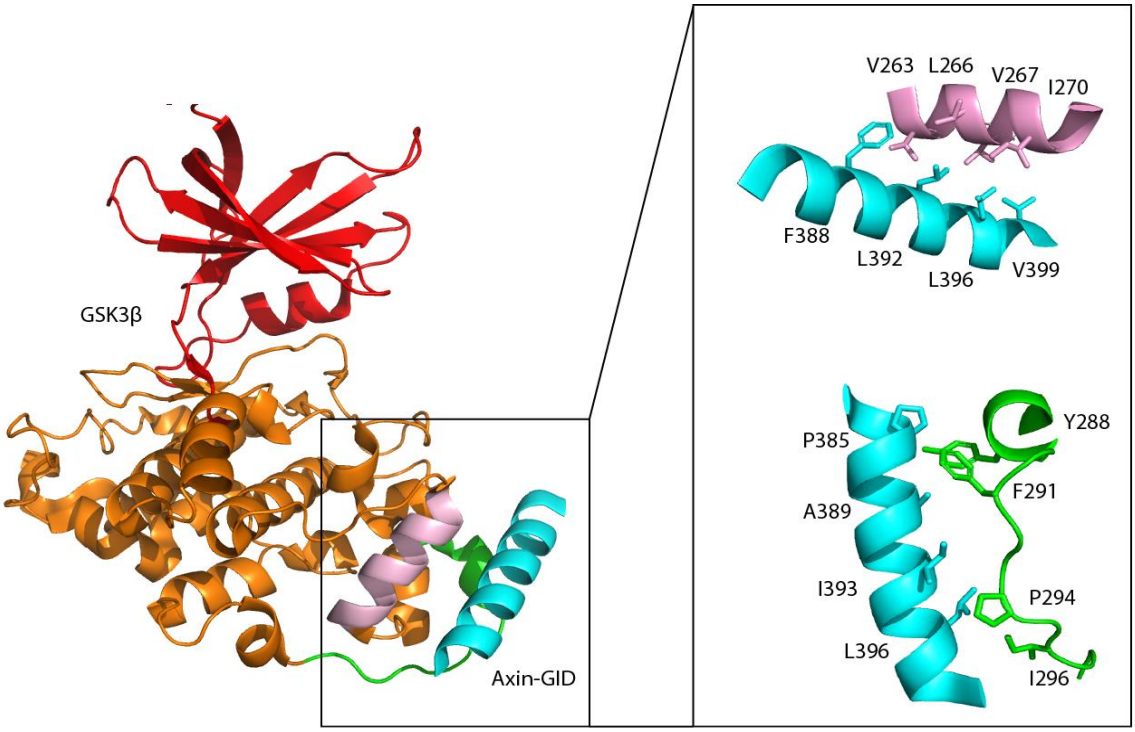


Figure 1-6. GSK3β-Axin GID complex. Crystal structure of the binary complex between kinase GSK3β and minimal GSK3β binding region on Axin (Axin-GID peptide) was recorded at a resolution of 2.4 Å. The cartoon model with detailed interaction framework is based on the PDB model 1O9U (published by Dajani *et al.* (2003)). The Axin-GID forms a single amphipathic α-helix (shown in cyan) and interacts with 2 major regions of C-terminal of the kinase, (i) an α-helix formed between GSK3β residues 262 to 273 (shown in light pink) and (ii) a loop between residues 285-299 (shown in green). The boxed portion on the right highlights the major interaction sites on both protein and peptide. One of the two major hydrophobic interaction sites is formed between Phe388, Leu392, Leu396, Val399 of Axin-GID on one side of the peptide with Val263, Leu266, Val267, Ile270 of GSK3β. The other site is between the other side of the Axin-GID peptide (residues Pro385, Ala389, Ile393, Leu396) and the GSK3β loop residues Tyr288, Phe291, Pro294, Ile296).

1.4.4.4 Structural framework involved in the binary complex of FRATide-GSK3β

Crystal structure of FRATide (197-222)-GSK3β complex was obtained by Bax *et al.* The molecular framework involved in the FRATide-GSK3β complex is mainly sustained *via* hydrophobic interactions with significant support from hydrogen bonds and van der Waals interactions[45]. FRATide binds at the C-terminal domain of GSK3β. FRATide bound GSK3β shows a substantial displacement between the loop spanning residues 288-294 and the α-helix spanning 262-273, leading to the formation of a hydrophobic groove (Fig 1-7). FRATide is comprised of two helices where the helix-turn-helix contains α1 and α2 (the break is between Gly210 and Asn211). The α1 helix (Histidine at 200 position of FRAT1) remains proximal to the Tyr216 on GSK3β but is less buried in the hydrophobic groove. The α2 helix interacts primarily with the hydrophobic groove on GSK3β. Lys271 of GSK3β forms van der waals interaction with FRATide via its hydrophobic side chain regions. Leucine and isoleucine at the N-terminal of the α2 helix (residue 212 and 213) are crucial to the interaction process where they form hydrogen bonds with side chains of Tyr288 and Glu290 on GSK3β (Fig 1-7)[84]. The leucine and isoleucine from FRATide are the only conserved residues in GID from Axin. Initially, it was hypothesized that they compete against each other to bind GSK3β as they share a partial binding site (based solely on the Leu-Ile sequence as they share no homology for any different amino acid sequence). The kinetic analysis showed a competition of FRATide and GID for GSK3β with less affinity between GID and GSK3β, which could explain why FRATide is able to dissociate Axin-APC complex from GSK3β.

A comparison of GSK3β structures in complex with peptides (Axin-GID and FRATide) from canonical Wnt/β-catenin pathway may give us insight into the binding mechanism between GSK3β and hD1 (minimal GSK3β binding and inhibiting peptide of DISC1, 44 amino acid long), if hD1 interacts at a similar location.

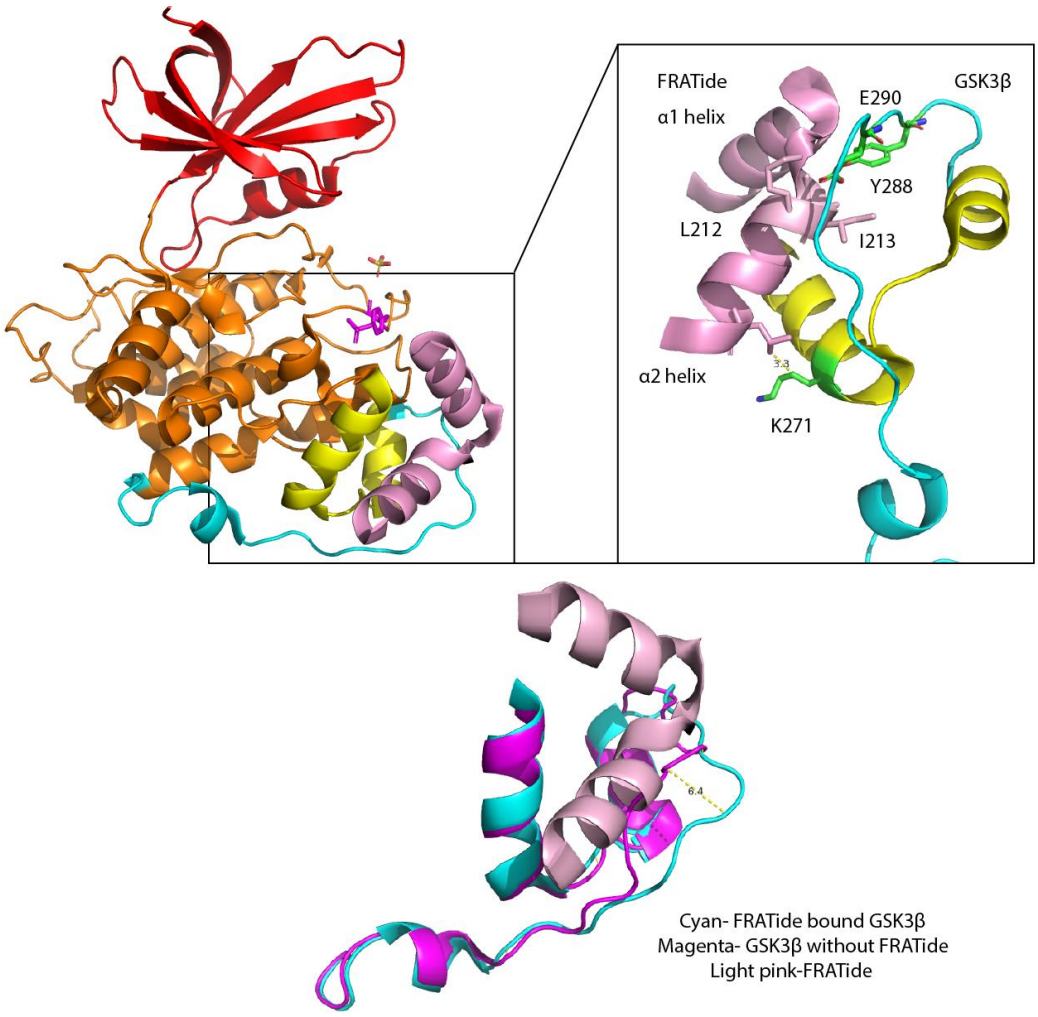


Figure 1-7. FRATide-GSK3β complex. Schematics of the GSK3β-FRATide complex shown in the cartoon model, based on the PDB model 1GNG. FRATide (shown in light pink) binds to GSK3β in the C-terminus region (shown in orange, yellow, and cyan). The interaction does not affect the substrate-binding pocket or the activation loop. A phosphorylated Tyr216 can be seen in magenta (in the activation loop). FRATide interacts mainly with residues 285-310 on GSK3β (shown in cyan). A helix-turn-helix is seen with FRATide (α1 and α2). Mostly hydrophobic interactions are observed between Leu212, Ile213 (FRATide) and Tyr288, Glu290 (GSK3β). A van der waals interaction is seen between α2 helix of FRATide and Lys271 of GSK3β. Overlay of GSK3β and GSK3β+FRATide structure shows the creation of a hydrophobic groove (shown in the bottom picture). This groove in GSK3β is created when the loop 288-294 moves around 6.4 Å from helix 262-273, α2 helix mainly interacts with this groove.

1.5 *In vitro* analysis of hD1 binding to GSK3β

Our lab has investigated the mechanism and molecular framework behind the interaction between synthetic hD1 and GSK3β using a variety of biochemical and biophysical assays[90]. SPR (Surface plasmon resonance) studies showed that FRATide can still bind GSK3β that has been saturated with hD1. The SPR assays suggested different binding sites for these peptides on GSK3β.

Further investigation of this finding was done using an enzymatic assay. hD1 can inhibit GSK3β at an IC50 of ~3 μM. Surprisingly in the presence of FRATide, the inhibitory effect of hD1 on GSK3β is suppressed. This finding hints towards a possible overlap of interaction sites for both peptides on the kinase, contradictory to the SPR findings.

These opposing inferences could be explained by findings shown by Mao *et al.*[19] They discovered a minimal GSK3β binding sequence (spanning 15 amino acids, 211-225 of mouse DISC1 protein) via SPR experiments. This minimal GSK3β binding sequence did not show any inhibitory effect on the kinase. Further experiments by our lab confirmed that the conserved minimal GSK3β region on hD1 of the human DISC1 protein binds but does not inhibit GSK3β (referred to as GSK3β binding region, GB, 209-225). Deletion experiment showed that the kinase inhibitory region resides at the C-terminal end of hD1 (referred to as GSK3β inhibitory region, GI, spanning 226-236)

Based on the SPR, enzymatic assays, and the findings mentioned in the previous paragraph, two possible models were proposed for the interaction framework of FRATide, hD1, and GSK3β. The first one hypothesizes that the FRATide physically blocks the "GI" binding site on GSK3β but not the "GB" binding site. The other model predicts a conformational change of the "GI" binding site but only once FRATide binds to GSK3β. These two models are the most likely scenarios given that FRATide and hD1 can both bind the kinase simultaneously, but FRATide prevents hD1 inhibition of GSK3β. FRATide is known to change conformation on GSK3β to create a hydrophobic groove (Fig 1-8).

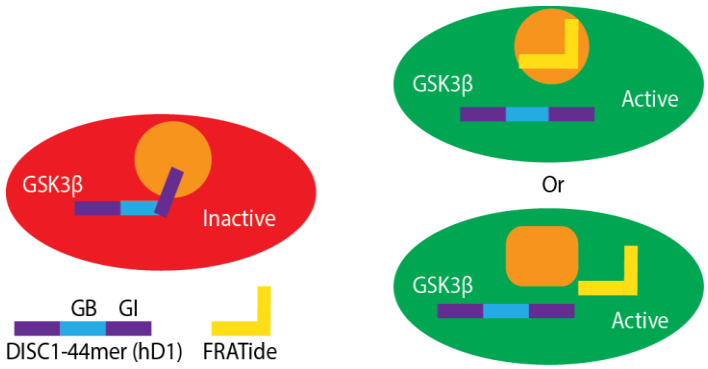


Figure 1-8. Predicted model FRATide and hD1 interaction with GSK3β. *In vitro* experiments have shown proofs of two regions on hD1, a central GSK binding region "GB" and a C-terminal inhibitory interacting domain "GI". FRATide prevents hD1 mediated GSK3β inhibition by either physically blocking the GI binding site on GSK3β (while allowing the GB to bind) or changing the conformation of GI binding region by binding elsewhere.

However, these models only predict the possible scenarios and offer essential cues for future investigation to determine the exact binding site. Detailed crystallography based structural analysis will be able to provide a final answer. This calls for structural investigation using crystallization of GSK3β-hD1 complex.

1.6 Hypothesis

Based on the models described above (Fig 1-8), I hypothesize that GB and GI regions of hD1 directly bind to GSK3β but binding of FRATide to hD1-GSK3β complex results in impairment of hD1-GI binding to GSK3β, preventing the inhibitory action of hD1.

1.7 Objective

The crystallographic analysis of the hD1/GSK3β complex may reveal the hD1 interacting region on GSK3β. The objective of my project is to crystallize the GSK3β & hD1 protein complex with and without FRATide. X-ray crystallography will be used for structural investigation. I have tried to achieve the objective using the following strategies, described and discussed in detail in chapters 3, 4 & 5.

Strategy 1: Complex formation of HisMBP- hD1with GSK3β

Strategy 2: Complex formation using GSK3β- hD1 co-expression construct

Strategy 3: Co-crystallization of GSK/hD1 protein complex, with and without FRATide

2 Chapter 2: Materials and Methods

2.1 Overview

The current section describes the cloning, overexpression, purification and crystallization techniques to achieve the target protein complex. This chapter is divided based on techniques with sub-sections titled with individual recombinant proteins or peptides used for the project. Two methods of cloning were used e.g., traditional restriction enzyme based cloning and ligase independent cloning to clone the human sequences of proteins of interest into plasmids for bacterial expression systems. Overexpression and purification protocols mentioned for each protein in this section were optimized *via* a series of experiments to yield pure and high quantities of proteins. Finally, complex formation for each strategy and the corresponding crystallization trials are described.

2.2 Rationale

Molecular understanding of any protein complex structure requires structural biology methods such as X-ray crystallography. Critical criteria for this method are the high quantity and purity of the purified protein (complex). Among all protein expression systems available, bacterial systems provide more flexibility, higher expression levels, and ease of purification. Different cloning methods use different strategies to derive a protein complex; for example, mono-cistronic plasmids allow expression of a single protein from one system. In contrast, a polycistronic plasmid allows expression of multiple proteins (possibly a complex) from a single system. Many chromatography techniques, such as affinity, ion exchange, size exclusion, etc., use different sets of principles to separate the target protein of interest from most of the contaminants. Many contaminants resistant to one type of chromatography could be removed by another. We have used various combinations of chromatography techniques for purification of different complexes. The quality of the protein complexes is evaluated by gel filtration, western blot, and dynamic light scattering (DLS) before crystallization. All catalogue numbers are listed in Appendix A.

2.3 Peptide: hD1

The synthetic peptide (residue 193 to 236) hD1, derived from the full-length human DISC1 (gene ID- 27185) (Uniprot- Q9NRI5.3), and it was purchased from NEO Biolabs (Cat. P12485D, 96.5% purity). The 44 amino acid long peptide has a molecular weight of 4.53 kDa and an isoelectric point of 4.75. The sequence of the peptide is $^{(193)}$ PEVPPTPPGS HSAFTSSFSF IRLSLGSAGE RGEAEGCPPS REAE $^{(236)}$ of the hDISC1 protein.

2.4 Peptide: FRATide

Synthetic FRATide (residues 174-196) peptide sequence was derived from the full-length human FRAT1 gene (Gene ID- 10023) (Unirpot-Q92837). The 25 amino acid-long peptide has a molecular weight of 2.89 kDa and an isoelectric point of 10.74. The sequence of the peptide is $^{(198)}$DPHRLLQQLVLSGNLIKEAVRRLHS$^{(222)}$.

2.5 Cloning

2.5.1 HisMBP-hD1

The sequence coding for hD1 (44 amino acids, residues 193-236) from hDISC1 (Gene ID: 27185) (854 amino acids) was cloned into a T7-promoter based protein expression system (pST44) with ampicillin resistance gene as a selective marker. The vector contains a 6x histidine tag (6xHis), a maltose-binding protein tag (MBP), and a tobacco etch virus (TEV) protease cleavage site (amino acid sequence ENLYFQ-S). Ligation independent cloning was used to clone the hD1 sequence into the vector C-terminal to the 6x His, MBP, and TEV protease site.

The vector was linearized using the restriction enzyme SspI-HF (NEB: R3132S) in the presence of cut smart buffer (NEB: B7204S). The hD1 coding sequence on hDISC1 was amplified using the protocol from table 2-1. Linearized vector and PCR products were confirmed by a 1.5% agarose gel.

Table 2-1. The PCR protocol for the amplification of hD1 sequence from the *hDISC1* gene.

Step	Temperature (℃)	Time
Initial denaturation	98℃	30 s
Primer annealing	98℃	10 s
	72℃	30 s
	72℃	6 s (for 172 nucleotides)
Final extension	72℃	2 mins
Hold	4℃	

The primers used for PCR are as mentioned below

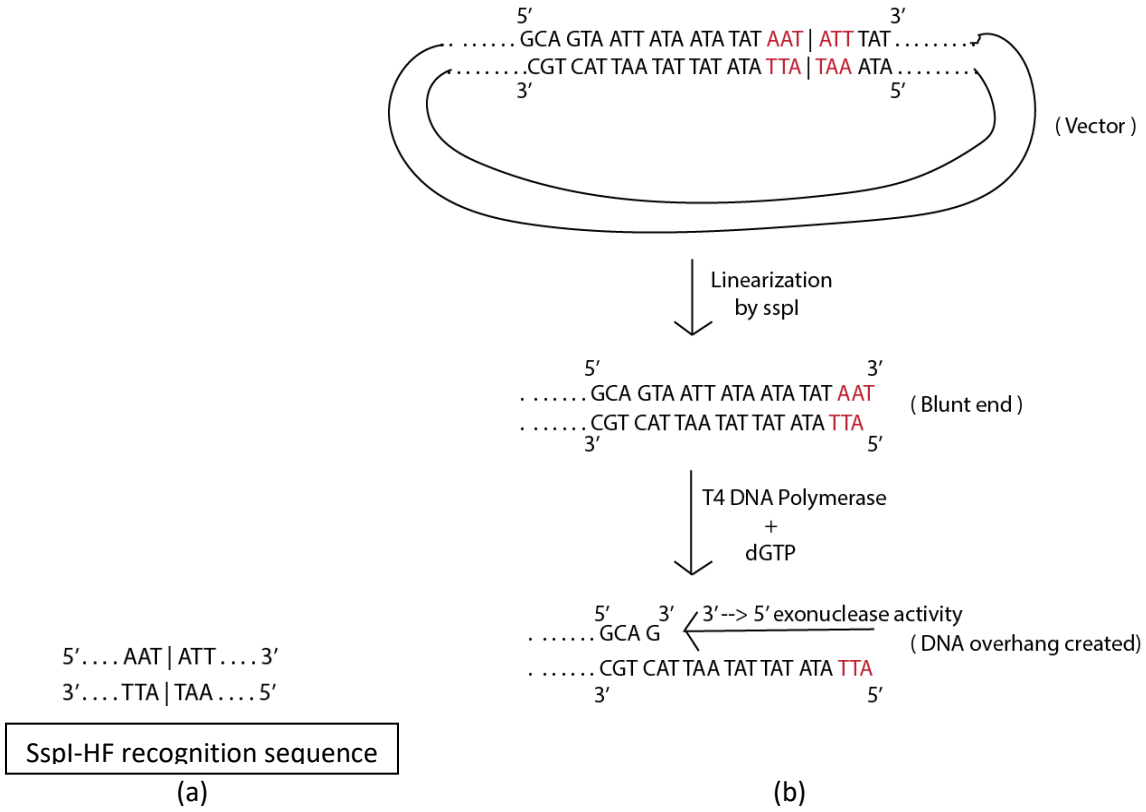
Forward primer- 5' TAC TTC CAA TCC AAT GCA CCG GAA GTG CCG CCGAC 3'

Reverse primer- 5' TTA TCC ACT TCC AAT GTT ATT ATT CAG CTT CGC GAC TCG 3'

Ligation independent cloning (LIC) method was used to yield the final construct (Fig 2-1)[91]. This cloning method uses T4 DNA polymerase and specially designed constructs. T4 DNA polymerase is capable of both exonuclease and endonuclease activities. The exonuclease activity of the polymerase is triggered in the absence of dNTPs, and it is used to create cohesive ends. After incorporating a restriction site in the construct, the overhang is designed in a way that it will miss a specific nucleotide for at least 15 bases. In this construct, the nucleotide bases are guanine and cytosine in the positive and negative strands, respectively. These initially missing nucleotides create the boundary of the cohesive ends. First, the construct was linearized by the restriction enzyme, creating blunt ends. Next, a T4 DNA polymerase reaction was performed with only one dNTP that was complementary to the boundary nucleotide. In the absence of complementary dNTPs, the 3'→5' exonuclease activity of T4 DNA polymerase would dominate until it slided to the boundary location where the one dNTP was present for the polymerase activity to take over. As a result, a cohesive end of 15 bases was generated.

An insertion PCR overhang with a pairing sequence for the vector overhang is also created using the same principle as previously described. The annealing of the T4-treated PCR product and the vector was done at a ratio of 3:1 and transformed into competent bacteria. The nicks present at the overhang joining sites were ligated by bacterial ligases *in vivo*. This process used neither the

restriction endonucleases to create overhangs nor the externally provided ligases for ligation. The detailed protocols are pictured below.


5′
3′
. . . . . GCA GTA ATT ATA ATA TAT AAT | ATT TAT . . . . . .
. . . . . CGT CAT TAA TAT TAT ATA TTA | TAA ATA . . . . . .
3′
5′
(Vector)
Linearization
by sspI
5′
3′
. . . . . GCA GTA ATT ATA ATA TAT AAT
. . . . . CGT CAT TAA TAT TAT ATA TTA
(Blunt end)
3′
5′
T4 DNA Polymerase
+
dGTP
5′
3′
3′--> 5′ exonuclease activity
. . . . . GCA G
(DNA overhang created)
. . . . . CGT CAT TAA TAT TAT ATA TTA
3′
5′
5′. . . . AAT | ATT . . . . 3′
3′. . . . TTA | TAA . . . . 5′
SspI-HF recognition sequence
(a)
(b)
5′
3′
TA ATT ATA ATA TAT AAT GCT TAC AGC . . . . . .
AT TAA TAT TAT ATA TTA CGA ATG TCG . . . . . .
(PCR product)
3′
5′
T4 DNA Polymerase
+
dCTP
5′
3′
TA ATT ATA ATA TAT AAT GCT TAC AGC . . . . . .
CGA ATG TCG . . . . . .
(DNA overhang created)
3′
5′
(c)

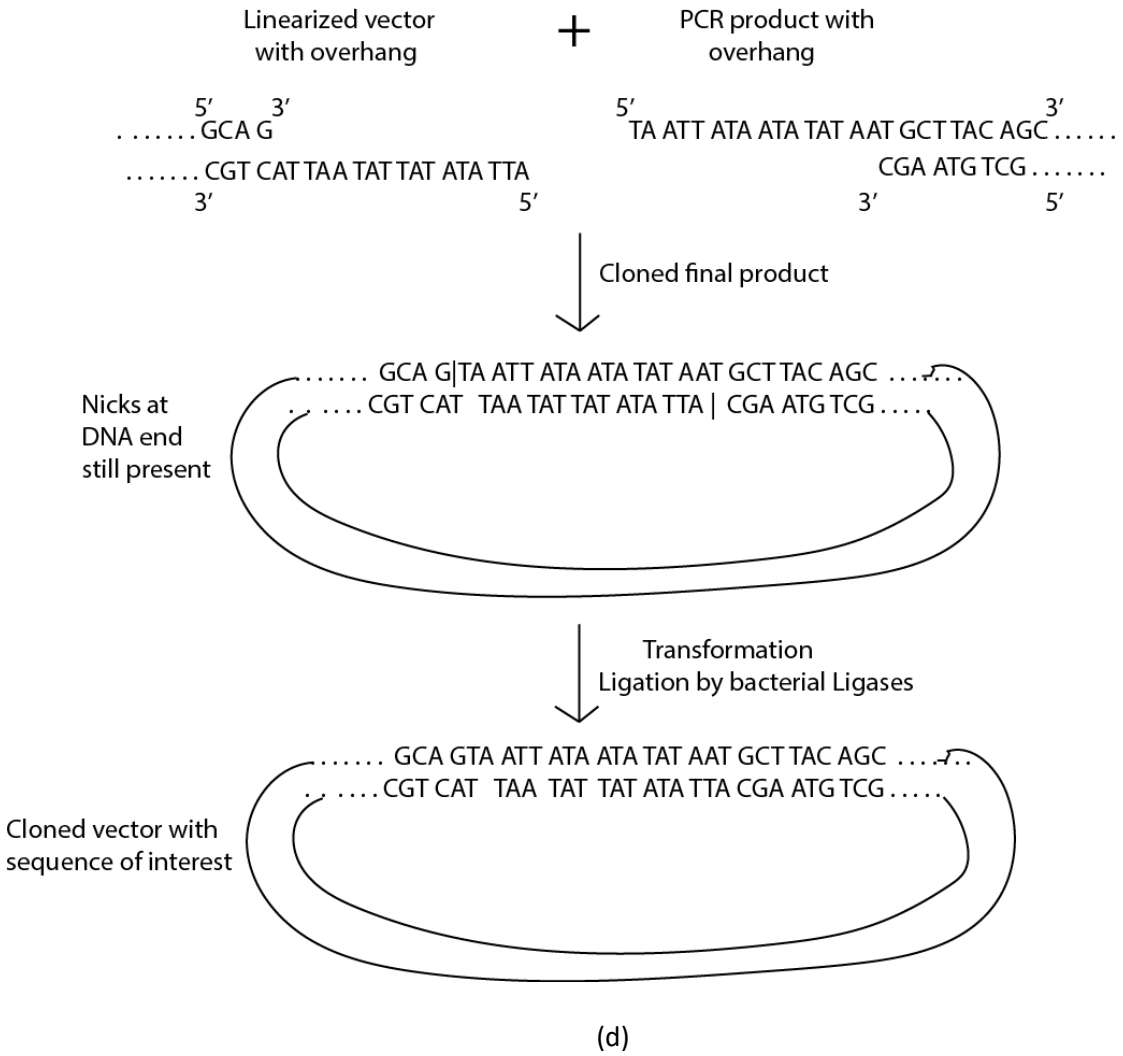


Figure 2-1 Ligation independent cloning. (a) The cleavage site for restriction endonuclease sspI-HF. SspI recognizes this specific sequence on the pST44 vector. SspI site is located after the sequence for 6xHis and TEV cleavage site. (b) The flowchart shows cleavage of a vector at the sspI site. The restriction digestion creates a linearized sequence with blunt ends. T4 DNA Polymerase creates an overhang using exonuclease activity. The exonuclease activity is limited by the supply of dGTP, which switches on the polymerase activity at the first cytosine on the negative strand. (c) PCR products are treated with T4 DNA polymerase and dCTP. An overhang is created at the first Guanine on the positive strand. (d) The last figure shows the annealing of the matching sequence of the PCR product and the vector. The final ligation happens inside the bacterial cells after the transformation.

Transformed colonies were tested for successful cloning by colony PCR using the following protocol (Table 2-2)

Table 2-2. Contents and protocol for colony PCR.

Content	Volume
10X ThermoPol Reaction Buffer	2.5 µL
10mM dNTPs	0.5 µL
10 µM Forward primer	0.5 µL
10 µM Reverse primer	0.5 µL
Colony	Pipet tip full
Control plasmid	
H_2O	20.875 µL
Taq Polymerase	0.125 µL

Step	Temperature (°C)	Time	Number of cycles
Initial Denaturation	95	5 mins	1
Denature	95	15 s	30
Annealing (primer Tm)	62	15 s	30
Extension (1 min/kb)	68	20 s	30
Final extension	68	5 mins	1
Hold	4	Infinite	1

The final protein sequence of the construct is as follows, with the names of the tags or genes bolded and bracketed right before the corresponding amino acid sequence (Fig 2-2).

(6xHis)MKSSHHHHHH**(MBP)**GSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKV
TVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTW
DAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEP
YFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSI
AEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASP
NKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEI
MPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIE
(TEVsite)ENLYFQSNA**(hD1)**PEVPPTPPGSHSAFTSSFSFIRLSLGSAGERGEAEGCPPSRE
AE

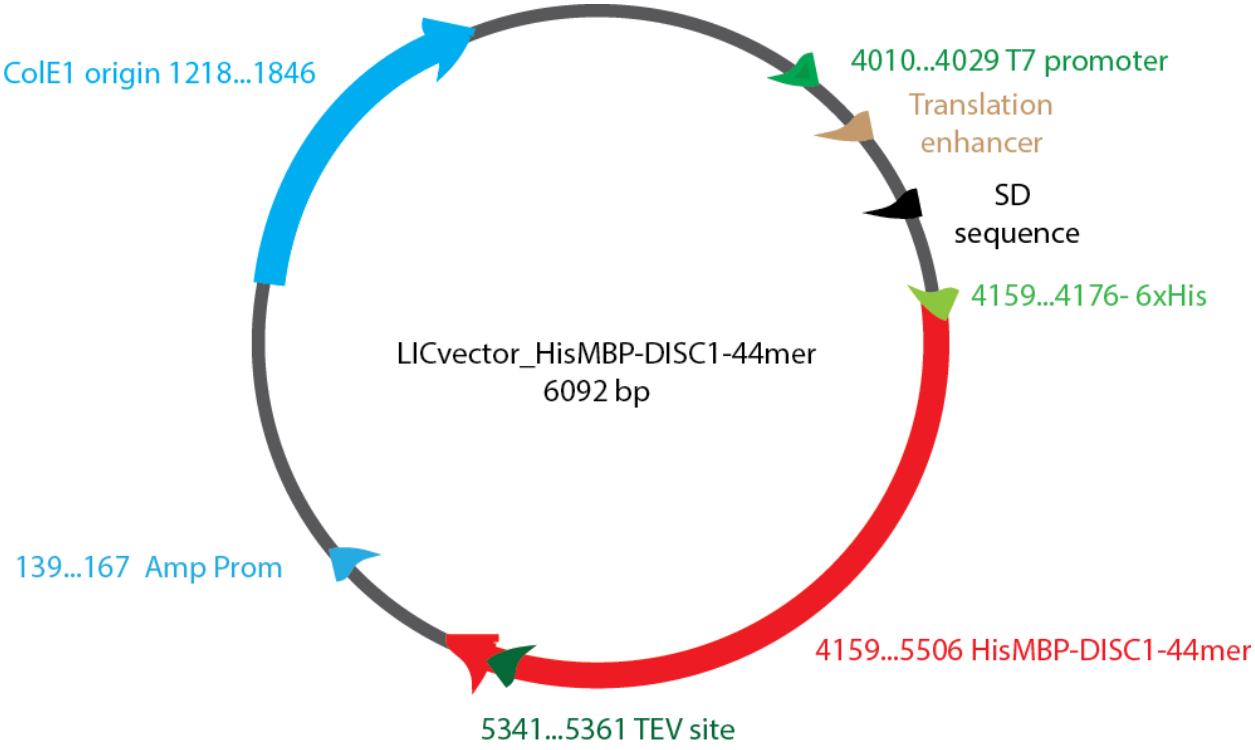


Figure 2-2. Graphic map of HisMBP- hD1.

2.5.2 His-GSK3β

The sequence for the human GSK3β (Gene ID: 2932, NCBI) was cloned into a T7 promoter-based pST50Trc1 vector backbone with ampicillin resistance as the selective marker[92,93]. A 6x histidine (6xHis) and TEV protease cleavage site sequences (TEV) are present N-terminal to the GSK3β sequence.

His-GSK3β contains 447 amino acids with a molecular weight of 49.40 kDa and an isoelectric point of 8.85. The protein sequence of the protein is as follows, with tags bolded and in brackets before the amino acid sequence (Fig 2-3).

6xHis)MGSSHHHHHHSSGSGGGGG**(TEV)**ENLYFQGS**(GSK3β)**MSGRPRTTSFAESCKPV
QQPSAFGSMKVSRDKDGSKVTTVVATPGQGPDRPQEVSYTDTKVIGNGSFGVVYQAKL
CDSGELVAIKKVLQDKRFKNRELQIMRKLDHCNIVRLRYFFYSSGEKKDEVYLNLVLDY
VPETVYRVARHYSRAKQTLPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLLDPDTA
VLKLCDFGSAKQLVRGEPNVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLLGQ
PIFPGDSGVDQLVEIIKVLGTPTREQIREMNPNYTEFKFPQIKAHPWTKVFRPRTPPEAIAL
CSRLLEYTPTARLTPLEACAHSFFDELRDPNVKLPNGRDTPALFNFTTQELSSNPPLATILI
PPHARIQAAASTPTNATAASDANTGDRGQTNNAASASASNST

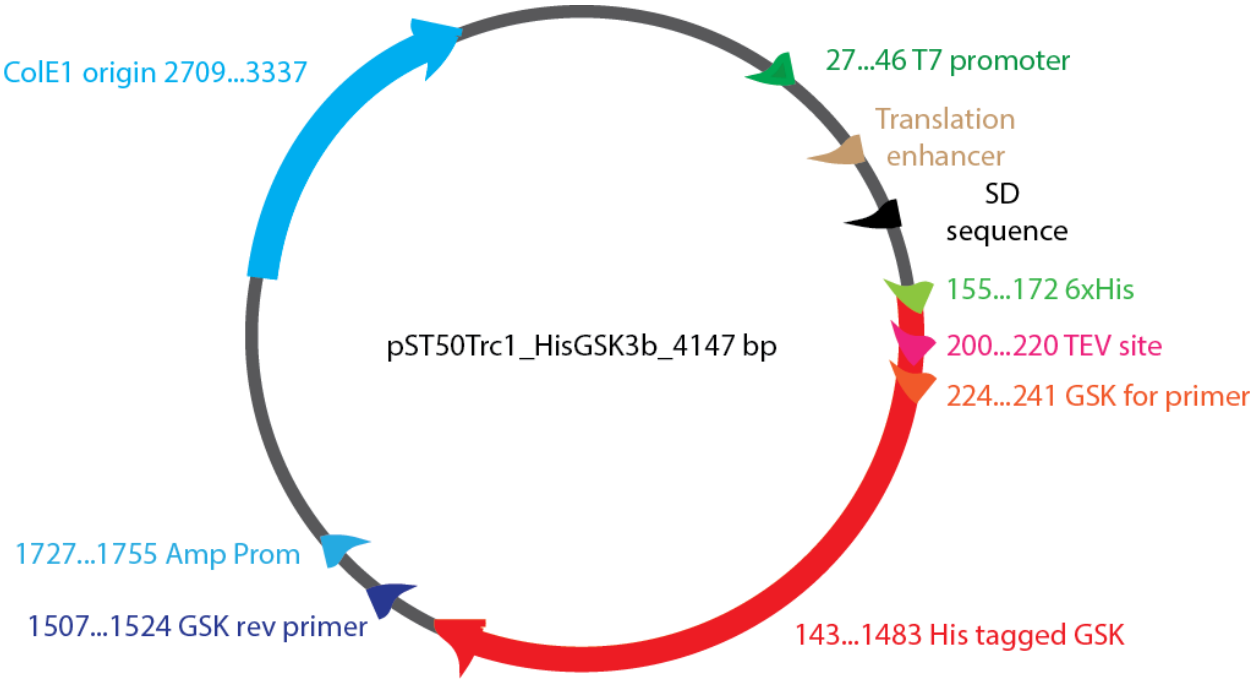


Figure 2-3. Graphic map of genetic sequence for His-GSK3β.

2.5.3 Co-expression construct HisGSK3β-StrnhD1

Single T7 promoter based polycistronic pST44 vector backbone was used to clone sequences for both proteins of interest, hD1 and GSK3β[93]. The 6xHis and streptavidin (Strn) tags were present upstream of GSK3β and hD1 coding genes, respectively. The ampicillin resistance gene is used for selective growth marker. The polycistronic vector backbones were developed by Tan *et al.* (2004) to provide modularity for expression of multiple recombinant proteins in one vector. This system uses conventional digestion and ligation of multiple genes from pST50Trc1/2 vectors into four serial gene cassettes available in the pST44 vector. Each cassette also contains sequence coding for cleavable or uncleavable affinity tags on either side of the genes of interest. A total of up to 4 different genes can be cloned into the final backbone for protein expression under a T7 promoter system. Every gene is flanked by various restriction enzyme digestion sites on both 5' and 3' ends. A schematic of the polycistronic protein expression system is depicted in figure 2-4.

The hD1 coding sequence was PCR amplified from the full-length DISC1 plasmid by using the following primers and Q5 Hot start high-fidelity DNA polymerase (NEB: M0493L)

Forward primer- 5' GCT AAA GGA TCC CCG GAA GTG CCG CCG A 3'

Reverse primer- 5' GGG CCC TGT ACA TCA TTC AGC TTC GCG ACT CG 3'

The PCR product was cloned into the pST50Trc2 backbone vector using NdeI (NEB: R0111S) and BsrGI (NEB: R0575S) in the presence of NEBuffer 2.1 (NEB: B7202S), downstream to a sequence coding for Streptavidin tag (Strn). The human GSK3β clone in the pST50Trc1 backbone was already made (Fig 2-3)[93]. The HisGSK3β sequence and the Strn-hD1 sequence were digested out from their respective vectors and cloned sequentially into the pST44 backbone using the XbaI (NEB: R0145S) &BglII (NEB: R0144S) sites for HisGSK3β and the EcoRI (NEB: R0101S) & HindIII (NEB: R0104S) sites for Strn-hD1 in the presence of NEBuffer 2.1 (Fig 2-4). The two genes of interest were cloned into the cassette 1 and 2 of the pST44 vector one at a time (Fig 2-4). T4 DNA ligase (NEB: M0202S) in the presence of T4 DNA ligase buffer (NEB: B0202S) was used to ligate 3:1 molar ratio of the inserts and vectors. The cloned construct was transformed into DH5-α competent cells (NEB: C2988J). Cloning was confirmed by sequencing.

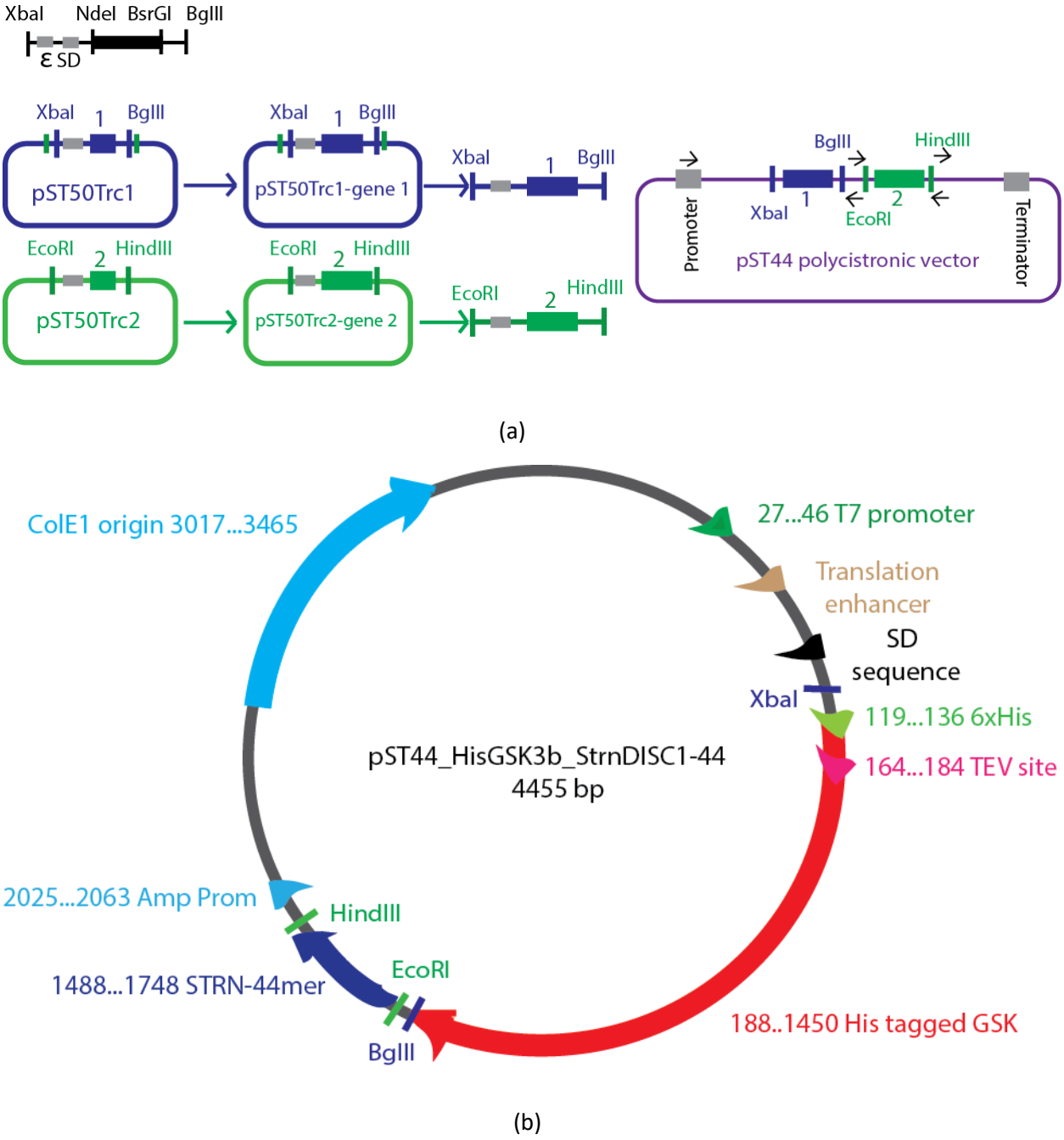


Figure 2-4. Polycistronic system for co-expression construct. (a) Schematics of pST44 polycistronic protein expression system. This system uses four transfer plasmids named as pST50Trc1/2 to transfer four different genes into the pST44 expression system. All transfer plasmids contain target clone region flanked by an internal set of restriction enzymes, NdeI and BsrGI. Outside of this region are the restriction sites that help to transfer the whole gene. All restriction sites are unique and highlighted in different colours for a different gene. Ɛ- Translation enhancer, SD- Shine-Dalgarno sequence, Arrows in pST44- primer sites, Blue, green, orange, red- for genes 1/2 respectively. (b) The graphic map for the genetic sequence of HisGSK3β-StrnhD1 co-expression construct.

2.6 Overexpression

2.6.1 HisMBP-hD1

The clone with the confirmed sequence was used for protein expression using BL21 Star (DE3) competent cells (source: Dmitriev lab, CoM, USask). Optimized growth parameters are listed below:

Table 2-3. Overexpression parameters variation for His-MBP-hD1 expression.

Flask (for both clones)	IPTG concentration	Temperature	Duration of growth Post induction
1	1 mM	37˚C	3 hrs
2	0.5 mM	25˚C	ON
3	0.2 mM	15˚C	ON

Cells were grown using 50 mL of 2xYT media (16 g tryptone, 10 g yeast extract, 5 g NaCl per 1 L of media, pH 7.0, Ampicillin at 100 µg/mL) in 250 mL Erlenmeyer flasks (Corning: 4980-250) and shaken at 220 RPM. The optimized purification protocol required growth at 37˚C for 3 hours after induction with 1 mM IPTG at OD600 range of 0.4-0.6. The cells were spun down at 5000 RPM, at 4˚C in 50 mL conical tubes (Fisher: 430044), using an Allegra 25R centrifuge (TS 5.1-500 rotor). Cell pellets were flash-frozen with liquid nitrogen and stored at -20˚C.

2.6.2 His-GSK3β

Protein was overexpressed using BL21 Star (DE3) pLysS cells. Seed culture was grown with 15 mL of Lysogeny broth (LB) media (10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract, pH 7, 100 µg/mL ampicillin, 25 µg/mL chloramphenicol) in a 50 mL conical tube at 30˚C. A 1:1000 ratio was used to inoculate seed culture to main culture media. TB media (24 g/L yeast extract, 20 g/L tryptone, 4 mL/L glycerol, 100 mL/L phosphate buffer (0.17 M KH2PO4, 0.72 M K2HPO4)) in the presence of ampicillin (100 µg/mL) was used for protein overexpression in the main culture. A total of 12 L media would be grouped in 8 Ultra yield Thomson flasks (1.5 L media in each 3 L flask) for one batch of overexpression. The culture was induced for protein expression using 0.1 mM IPTG when the OD 600 value was 0.2 and the temperature was 18˚C. Cells were grown for 48 hours with ampicillin replenished after the first 24 hours. Cells were harvested by spinning the

culture at 4700 RPM using 2 L PPCO Nalgene centrifuge bottles in a Sorvall RC12BP low-speed centrifuge (H-12000 swinging bucket rotor) (4℃, 30 minutes). Cell pellets were flash-frozen with liquid nitrogen.

2.6.3 HisGSK3β-StrnhD1 (Co-expression construct)

The clone with the confirmed sequence was first optimized for the most suitable cell strain based on the final yield of the protein. The following three expression systems were tested, BL21 Star (DE3), BL21 star (DE3) pTf16, and Rosetta (DE3) pLysS. The protein yield was further optimized using BL21 Star (DE3) and BL21 Star (DE3) pTf16 using the following parameters for optimization (Table 2-4).

Table 2-4. Overexpression optimization for co-expression construct.

Cell strain	IPTG concentration	Temperature	Duration of growth Post induction
BL21 Star (DE3)	1 mM	37℃	3 hrs
BL21 Star (DE3)	0.15 mM	16℃	ON
BL21 DE3 Star pTf16	1 mM	37℃	3 hrs
BL21 DE3 Star pTf16	0.15 mM	16℃	ON

Optimized overexpression of the protein was done with BL21 Star (DE3) pTf16. Seed culture was grown using 10 mL LB broth. Seed culture (1:1000) was inoculated into 1 L of 2xYT media supplemented with ampicillin (100 µg/mL) and L-arabinose (0.5 mg/mL) in a 3 L Ultra yield Thomson flask. L-arabinose was added in the beginning to induce the expression of chaperone pTf16. The cells were induced at an OD600 value of 0.6 with 0.15 mM IPTG and grown at 16℃ overnight. Cells were harvested by spinning the cells in 1 L polycarbonate centrifuge bottles with a JLA 8.1 rotor at 5000 RPM (4℃, 30 mins). Cell pellets from a total of 3 L culture were flash-frozen in liquid nitrogen and stored at -20℃.

2.7 Purification

2.7.1 HisMBP-hD1

2.7.1.1 Small scale purification

0.5 g of frozen cell pellet (the equivalent of 50 mL culture) was resuspended in 1.5 mL (cell pellet to lysis buffer ratio, 1:3) of lysis buffer (LB 1) (50 mM Hepes pH 7.4, 200 mM NaCl, 5% glycerol, protease inhibitor, lysozyme at 1 mg/ml). The resuspended lysate was incubated at 37℃ for 30 minutes. The supernatant was recovered by spinning the lysate at 15000 RPM (TA 15-1.5 rotor, Allegra 25R centrifuge) for 30 minutes using 2 mL microcentrifuge tubes. The supernatant was further incubated and nutated with 50 µL loose NiNTA resin for 1 hour at 4°C. The resin was previously equilibrated with equilibration buffer 1 (EB 1) (50 mM Tris pH 7.5, 300 mM NaCl, 20 mM Imidazole). Resins were washed with ~40 column volume (CV) of EB 1 to remove contaminants, and then the protein was eluted using 10 CV of elution buffer 1 (ELB 1) (50 mM Tris pH 7.5, 300 mM NaCl, 250 mM Imidazole). The washes and elution fractions were collected by spinning the mix at 14000 RPM in a table-top centrifuge.

2.7.1.2 Large scale purification

In order to yield more protein, the supernatant fractions from 8 small scale cultures (50 mL) were combined and purified using a 1 mL HisTrap FF column. 40 CV of washes and 10 CV of protein elution were performed manually by passing EB 1 and ELB 1 respectively. A solubility screening of the pure protein was performed where 50 µL protein at 1 mg/mL was mixed with the final concentration of the following solvents. The solvents screened were 20% ethylene glycol, 20% glycerol, 50% glycerol, 20% glycerol with 500 mM NaCl, and 20% PEG 4000. Dynamic Light Scattering (DLS) ((Wyatt DynaPro Plate Reader II) was used to screen the best storage buffer for the protein.

2.7.2 HisGSK3β

Cell pellets from 12 L culture (~90 g) were resuspended in ~300 mL of lysis buffer 2 (LB 2) (50 mM Hepes pH 7.2, 5% glycerol, 50 mM NaCl, 1 mM phenyl-methyl-sulfonyl fluoride (PMSF)) at a pellet weight to lysis buffer volume ratio of 1:3. The resuspended cell pellet was lysed by cell disruption (Constant systems LTD 0.75KW, two runs with 25K Psi). The soluble

fraction of the lysate was collected after spinning at 40000 RPM for 45 minutes (Rotor type 70Ti, Beckman Coulter Optima XPN-100).

Over the course of optimization for purification, we tried two protocols which differed based on the sequence of ion exchange and affinity chromatography.

Protocol 1: Affinity -> Cation-exchange chromatography

The supernatant was loaded onto a 5mL His-Prep FF column equilibrated with equilibration buffer 2 (EB 2) (50 mM Tris pH 7.5, 300 mM NaCl, 12.5 mM imidazole). The column was washed with 40 CV of EB2 to remove contaminants. His-GSK3β was eluted with 10 CV of elution buffer, ELB 1. A flow rate of 5 mL/min. was maintained throughout the affinity chromatography step.

The affinity-purified protein was diluted with dilution buffer (DLB, 50 mM Hepes pH 7.2, 5% glycerol, 50 mM NaCl) in a 1:7 volume ratio. The diluted pool was passed through a 1 mL SP sepharose column at a flow rate of 1 mL/min. The column was washed with 10 CV SP buffer A (SPA 1, 50 mM HEPES pH 7.2), and the proteins were eluted in a 20CV linear gradient from 0% to 100 % of SP Buffer B (SPB 1, 50 mM Hepes pH 7.2, 1 M NaCl). HisGSK3β containing fractions were pooled and treated with TEV protease at 1:10 molar ratio.

The dialyzed and TEV-treated protein pool was purified using the same NiNTA column with similar method parameters. Cleaved GSK3β was collected from the flow-through and concentrated using a 3K MWCO 4 mL Amicon ultra-4 concentrator to a volume of 500 µL or not more than 1-2 mg/mL. The concentrated protein was loaded on to the size exclusion chromatography column, Superdex 200 increase 10/300 GL at 0.5 ml/min in gel filtration buffer (50 mM Hepes pH 7.2, 500 mM NaCl, 1 mM TCEP, 2 mM $MgCl_2$). Gel filtration elution fractions were analyzed by SDS-gel electrophoresis to check for purity of the protein.

Protocol 2: Cation-exchange -> affinity chromatography

The supernatant was diluted with 400 mL of DLB (in 1:7 ratio of pellet weight to DLB volume) to reduce the salt concentration for cation exchange. The diluted supernatant was applied to a 50 ml self-packed SP column equilibrated with 5 CV of SP Buffer A at a flow rate of 10

mL/min. The column was washed with 40 CV of SPA and elution was performed with a 10 CV linear gradient of SPB 1.

The SP elution fractions containing His-GSK3β were pooled and applied to a 10 mL HisTrap FF column (two 5 mL HisTrap FF columns connected in tandem and equilibrated with 5 CV with EB 2), and His-GSK3β was eluted with 10 CV of ELB 1.

TEV protease was added to the NiNTA elution pool in a molar ratio of 1:10. Tag cleavage and dialysis of the protein with dialysis buffer (DB, 50 mM Hepes pH 7.2, 500 mM NaCl, 5 mM BME, protein to buffer volume ratio of 1:100) were carried out simultaneously at room temperature for 17 hours. A clean and cleaved GSK3β pool was separated from the rest of the proteins in a flow-through fraction using the same 10 mL NiNTA column and purification protocol. Buffer exchange of GSK3β was done with the final buffer (50 mM Hepes pH 7.2, 500 mM NaCl, 1 mM TCEP, 2 mM $MgCl_2$). The pure crystallizable protein was further concentrated up to 4-8 mg/mL using a 3K MWCO 4 mL Amicon ultra-4 concentrator.

2.7.3 HisGSK3β-StrnhD1

2.7.3.1 Small scale purification test

Small scale purification experiments for the co-expression construct were performed with 0.5 g of cell pellet. Cells were lysed using lysis buffer 3 (LB 3, 50 mM Tris pH 7.5, 300 mM NaCl, 5% glycerol, 20 mM imidazole, protease inhibitor tablet) at a pellet weight to buffer volume ratio of 1:3. The soluble fraction was collected by spinning the resuspension at 15000 RPM for 30 minutes at 4℃ (TA 15-1.5 rotor, Allegra 25R centrifuge). The possible protein complex was eluted using 10 CV of ELB1 from 50 µL loose NiNTA resin that is equilibrated with 5 CV of EB 1.

2.7.3.2 Large scale purification

For the larger scale of purification, 2.5 g of cell pellets were resuspended in 7.5 mL of LB 3 (ratio 1:3), and the supernatant was collected the same way as described for small-scale purification. A multiple step purification was performed for the larger batch.

Protein was eluted with 10 resin volume of ELB 1 (50 mM Tris pH 7.5, 300 mM NaCl, 250 mM Imidazole) from 500 µL loose NiNTA resin that is equilibrated by 10 resin volume EB 1 (50 mM Tris pH 7.5, 300 mM NaCl, 20 mM Imidazole).

The His-GSK3β of the complex mixture was simultaneously dialyzed and cleaved with TEV protease in a 1:10 molar ratio in dialysis buffer (50 mM Hepes pH 7.2, 500 mM NaCl, 5 mM BME). The purification of the dialyzed protein/TEV mixture was performed the same way as described above, and the cleaved protein complex was recovered from the flow-through. Flow-through protein was concentrated using an Amicon ultra 3K MWCO Ultra-4 mL concentrator. The concentrated protein sample was further purified with size exclusion chromatography with a Superdex 200 increase 10/300 GL (24 ml bed volume) column in gel filtration buffer (GF, 50 mM Hepes pH 7.2, 500 mM NaCl, 1 mM TCEP, 2 mM $MgCl_2$) at a flow rate of 0.5 mL/min.

2.8 Complex formation and biochemical analysis

2.8.1 Complex formation with HisMBP-hD1

Purified GSK3β and HisMBP-hD1 proteins were mixed at 200 nM concentration in Reaction buffer (RB 1) (50 mM Tris pH 7.5, 5 mM $MgCl_2$, 0.01% Tween-20, 5 mM TCEP). The possible complex mixture was purified using 100 µL of loose NiNTA resin equilibrated with EB 1 (50 mM Tris pH 7.5, 300 mM NaCl, 20 mM Imidazole). The resins were washed with 20 CV EB 1, and the complex was eluted with 10 CV of ELB 1 (50 mM Tris pH 7.5, 300 mM NaCl, 250 mM Imidazole). A negative control experiment with just HisMBP- hD1 was included in the experiment as well. The complex formation was tested with western blot using antibodies for GSK and HisMBP- hD1.

2.8.2 Complex formation with hD1 and FRATide

Synthetic hD1 and GSK3β were mixed at 8 µM or 200 nM, at a 1:1 ratio. Running buffer (RB 2, 50 mM Tris pH 7.5, 5 mM MgCl2, 150 mM NaCl, 0.005% T-20, 1 mM TCEP) was used to mix both the proteins for complex formation. Protein mixture was incubated for 30 minutes in room temperature to help the complex formation process. The complex was further concentrated to ~50 µL or at least 7 mg/mL using a 3K MWCO 500 µL concentrator.

Co-crystallization of hD1 and FRAtide with GSK3β was performed at 200 nM, at a 1:1:1 ratio. The proteins were mixed in RB 2 and were concentrated up to ~8 mg/mL using a 3K MWCO 500 µL concentrator.

2.9 Western blot

Confirmation of protein identity for all constructs was done by western blot. Proteins of interest from the SDS gel were transferred to nitrocellulose membrane of pore size 0.45 μm. The transfer was done for 60 minutes at 350 mA. Two types of blocking solutions were used depending on the type of antibody: (i) Blocking solution 1 (BS 1, dry milk as the blocking agent) was made with 1xTBST (20 mM Tris-HCl of pH 7.4, 150 mM NaCl, 0.1% Tween 20) and 5% (w/v) dry milk powder and (ii) Blocking solution 2 (BS 2, Bovine serum albumin as the blocking agent) was made with 1xTBST and 5% (w/v) BSA powder. Blots were incubated with blocking solution for 1 hour, primary antibodies and conjugated antibodies for 1 hour, and secondary antibodies overnight.

Complex confirmation in the interaction assay using HisMBP- hD1 and GSK3β was done with primary antibodies mouse anti-6xHis (to detect the His tagged protein) and mouse anti-GSK3β (1-160) (for the detection of GSK3β). BS 1 was used to prepare the antibody solutions at the ratio of 1:2000 for the anti-His and 1:6000 for the anti-GSK antibody. Anti-mouse DyLight 680 was used as the secondary antibody for this test, diluted with BS 1 at a ratio of 1:10,000. Fluorescent signals for the respective bands were recorded at 680 nm wavelength using G-box (Syngene).

Expression confirmation for His-GSK3β from the co-expression construct (HisGSK3β-StrnhD1) was done using mouse penta-His alexa flour 488 antibody. The antibody solution was prepared with BS 1 at a ratio of 1:10,000. The fluorescent signal at a wavelength of 488 nm was recorded using G-box.

Strep-tagged hD1 (Strn-hD1) of the co-expression construct was not visible in denaturing gels due to its low molecular weight. Expression of this protein was detected using StrepMAB-classic mouse anti-strep antibody. BS 2 was used to dilute the antibody at a ratio of 1:2000. Chemiluminescence was used to detect signals in this case. The blots were incubated with the ECL solution (10 mM Tris pH 8.5, 0.19 mM p-Coumaric acid, 1.25 mM Luminol, 0.01% H_2O_2) for an hour before signal detection using the chemiluminescence setting of G-Box.

2.10 Crystallization

2.10.1 Crystallization of protein complexes

A sparse matrix commercial widescreen JCSG+ suite (MD1-40, Molecular dimensions) was used to screen potential crystallization conditions for the concentrated protein complex. The screening was done in a 96-2 well intelliplate (Art Robbins) using the Crystal Gryphon robot (Art Robbins). A 0.15 µl drop per well setting was used for the screen. The plates were stored inside the Crystal Farm imaging system. Promising conditions of crystallizations were further optimized with a fine screen to improve the quality of the crystals. Fine screen optimizations were done using a 24-well sitting drop Cryschem M plate (Hampton research: HR1-002). The freezing parameters tested for different crystals are listed below (Table 2-5)

Table 2-5. Crystal freezing parameters. Harvest buffers used for each crystal conditions are described along with the buffer composition of the protein, interaction concentration, and crystallizing condition, etc.

Complex type	Sample buffer	Mother liquor (ML) (Crystallization)	Harvest buffer (HB) (Freezing)
GSK3β+ hD1 (8 μM)	(Running buffer) 50 mM Tris pH 7.5 5 mM MgCl2 150 mM NaCl 0.005% Tween-20 1 mM TCEP	10 mM Hepes pH 7.5 10%/14% PEG 8000 8% Ethylene glycol	10 mM Hepes pH 7.5 14% PEG 8000 10%/20%/30% Ethylene glycol
GSK3β+ hD1 (8 μM)	Running buffer	0.1 M Bicine pH 9 10% PEG 6000	0.1 M Bicine pH 9 10% PEG 6000 20% glycerol
GSK3β+ hD1 (8 μM)	Running buffer	0.1M Tris pH 8 12% PEG 8000	0.1 M Tris pH 8 12% PEG 8000 20% glycerol
GSK3β+ hD1 (200 nM)	Running buffer	10 mM Hepes pH 7.5 8%/10% PEG 8000 8% Ethylene glycol	10 mM Hepes pH 7.5 14% PEG 8000 20% ethylene glycol

2.10.2 Diffraction studies

Diffraction data for the frozen crystals were collected at 08ID-1 beamline (CMCF-ID) at the Canadian Light Source (CLS). The CMCF-ID beamline is equipped with a double crystal monochromator (DCM) and a Pilatus3 s 6M X-ray detector. Data were collected at 100 K.

3 Chapter 3: Complex formation with HisMBP-hD1 and GSK3β

3.1 Rationale

The hD1 region of DISC1 protein is the most potent inhibitory fragment against GSK3β. Understanding the molecular framework involved in this inhibition is the primary objective of our project. The first strategy of yielding a hD1-GSK3β complex is to express and purify each protein independently. Complex formation was attempted by mixing the purified protein in equal molar concentration. In this strategy, the hD1 fragment was fused with a large tandem affinity tag that contains 6xHis and MBP. The 6xHis tag helps with targeted affinity purification, and the MBP tag increases the solubility of the small hD1 peptide.

3.2 Overview

This chapter describes the individual expression and purification protocols, including major optimizations for each interaction partners. Consistent overexpression of HisMBP-hD1 was achieved after selecting a clone with the highest expression level. Overexpression of GSK3β was optimized by other lab members. This chapter, however, discusses a few major challenges that I have overcome. By combining several chromatography steps in different sequences, highly pure GSK3β that is crystallizable was obtained. This chapter describes my attempt to obtain the binary complex by mixing together the purified HisMBP-hD1 and His-GSK3β.

3.3 Purification of HisMBP-hD1

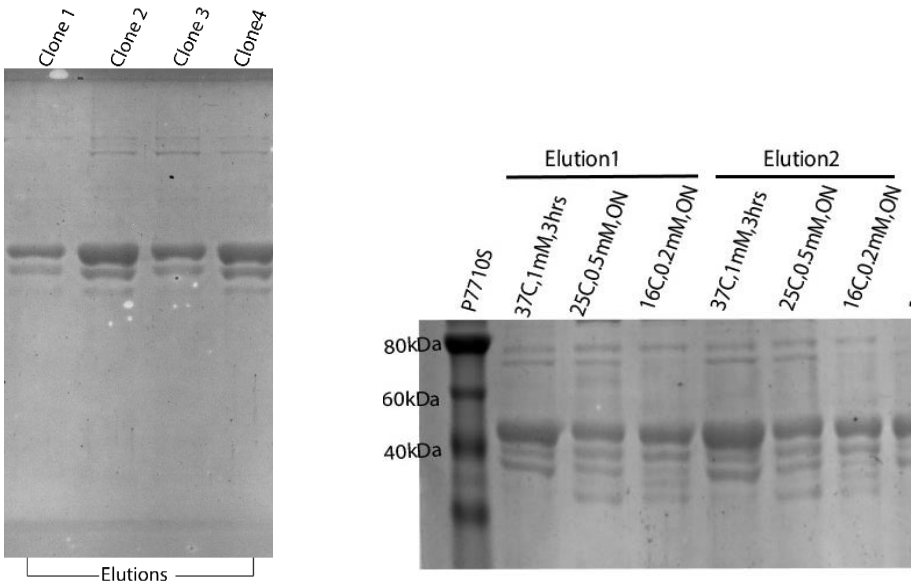
Expression levels of the protein were not clear when lysate samples were analyzed on a denaturing gel. Due to the lack of intense bands in lysate samples, small scale purification experiments were done to check the expression level. Purification of small-scale cultures (50 mL) after successful cloning revealed fluctuation in protein yield from one experiment to another. A clonal selection test was performed between four random colonies selected from the transformed plate. The small-scale purification resulted in a visible difference in the final protein yield (Fig 3-1 (a)). The clone with the highest protein yield was selected for further use. A 1ml aliquot of an

overnight culture grown from this clone was frozen with 20% glycerol. Subsequent cultures were grown from this frozen stock to yield a consistently higher level of HisMBP-hD1.

Maiden expression of the protein of interest required optimization of overexpression experiments. Overexpression experiments were done to evaluate and determine possible conditions to yield a very high amount of the protein. This test was done to compare several overexpression parameters, such as growth duration, temperature, and IPTG concentration. Small scale purification experiments were necessary for comparable protein yield so that the best expression condition could be determined. The conditions are listed below (same as Table 2-3),

Flask (for both clones)	IPTG concentration	Temperature (℃)	Duration of growth Post induction
1	1 mM	37℃	3 hrs
2	0.5 mM	25℃	ON
3	0.2 mM	15℃	ON

The overexpression optimization experiment of HisMBP-hD1 determined expression of the protein induced with 1 mM IPTG at 37˚C for 3 hours to be the optimized set of conditions (Fig 3-1 (b)). The protein elution fractions collected from 50 µL NiNTA resins contained a negligible level of contaminants and were pure enough for interaction experiments. To complete the requirement of enough protein for interaction, a larger scale culture was purified (500 mL). Approximately 1 mg of protein with more than 80% purity was recovered from the larger-scale culture using 500 µL NiNTA resin (Fig 3-2).



(a) (b)

Figure 3-1. HisMBP-hD1 protein expression optimization. (a) A 10% SDS gel shows differential protein expression strength amongst 4 different clones. Samples are NiNTA elution from a small-scale purification. Clone 2 was selected and stored because of its higher expression level. (b) The denaturing SDS gel shows the test of varying over-expression parameters to achieve optimal protein yield. Elutions 1 and 2 represent protein eluted from the NiNTA resin.

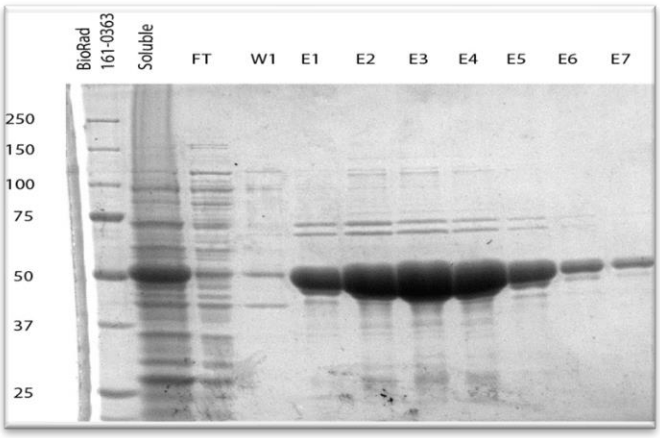


Figure 3-2. Large scale purification of HisMBP-hD1. An SDS gel is showing a high yield of protein from 500 ml culture using the optimized O/E parameter(37℃, 3 hours, 1mM IPTG), Soluble- a soluble fraction of the lysate (supernatant), FT- Flow-through, W1- Wash, E1 to E7- Protein eluted using 250 mM imidazole.

Purified protein was tested for size distribution and monodispersity using DLS. Molecular weight of the most common molecular species in the solution was 52 kDa (Table 3-1), which is very close to the theoretical molecular weight of HisMBP-hD1. To determine the optimal long-term storage condition for HisMBP-hD1, DLS analysis were done with samples stored in different cryoprotectants. The solutions tested were 20% ethylene glycol, 20% glycerol, 50% glycerol, 20% glycerol with 500 mM NaCl, and 20% PEG 4000.

Out of all the solvents tested, 20% ethylene glycol maintained the quality of the protein much better than others. DLS suggested the presence of a major molecular species with a molecular weight of ~77 kDa, which is also close to the theoretical molecular weight of HisMBP-hD1 (~48 kDa). The protein species was also mostly monodispersed (Table 3-2). All HisMBP-hD1 pools were frozen with 20% ethylene glycol until GSK3β was purified for interaction and complex formation experiment.

Table 3-1. DLS analysis of HisMBP-hD1. A dynamic light scattering analysis shows HisMBP-hD1 maintains a monomeric state as a monodisperse (%Pd) species when expressed and purified using our protocol.

Intensity distribution	Radius (nm)	% Pd	Mw-R (kDa)	% Intensity	% Mass	% Number
Peak 1	3.232	11.7	52	7.9	91.2	99.9
Peak 2	14.066	11.7	1636	47.2	7.0	0.1
Peak 3	69.443	11.0	68583	42.4	0.3	0.0
Peak 4	2925.840	4.2	434021000	25	1.5	0.0

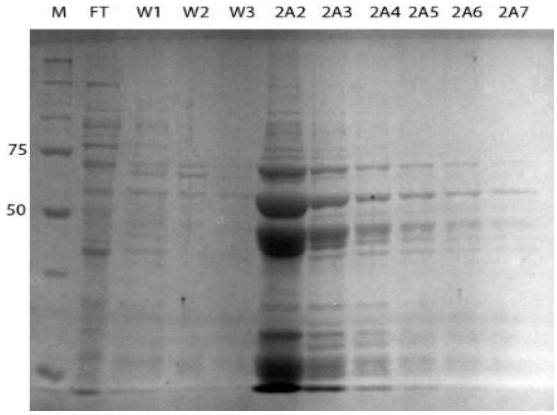
Table 3-2. DLS analysis of HisMBP-hD1 with cryoprotectant. A dynamic light scattering analysis shows HisMBP-hD1 maintains a monomeric state and monodisperse (%Pd) species when stored in 20% ethylene glycol.

Intensity distribution	Radius (nm)	% Pd	Mw-R (kDa)	% Intensity	% Mass	% Number
Peak 1	3.800	19.7	77	11.6	91.9	99.8
Peak 2	22.473	44.8	4896	83.4	7.9	0.2
Peak 3	261.880	38.9	1531280	5.0	0.3	0.0

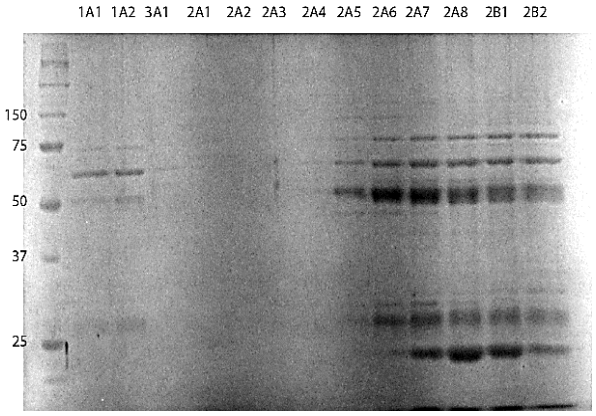
3.4 Purification of GSK3β

Overexpression of GSK3β construct was optimized by Stephanie Saundh (Thesis titled, "Binding kinetics investigation of a peptide fragment of the psychiatric risk protein, DISC1, to the kinase GSK3β". (2019)) and Anand Krishnan (Thesis titled, "Structural and functional characterization of the psychiatric risk protein, DISC1 and its interaction partner GSK3β). Overexpression of the protein was done using BL21 Star (DE3) pLysS, grown for 48 hours at 18℃, and induced with 0.15 mM IPTG at an OD600 value of 0.2.

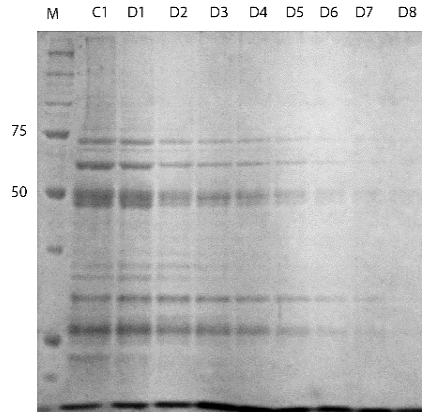
According to the affinity -> cation-exchange chromatography protocol, HisGSK3β was collected from NiNTA resin as the major eluent in the first step. The elution fraction 2A2 (Fig 3-3 (a)) contains significant contaminants such as ~75 kDa, ~60 kDa, and~25 kDa. The pooled protein fraction was then subjected to the cation exchange SP sepharose column, and HisGSK3β was eluted along with the contaminants between 29-77% salt gradient. The separation of contaminants from the protein of interest was inefficient (Fig 3-3 (b,c)). The separations of the ~60 kDa, ~75 kDa, and ~25 kDa contaminants were inefficient even after the size exclusion chromatography step (Fig 3-3(d)). We tried the cation-exchange -> affinity chromatography protocol next (where the two chromatography steps were reversed) as the protein was impure for crystallization or interaction process.


M FT W1 W2 W3 2A2 2A3 2A4 2A5 2A6 2A7
75
50

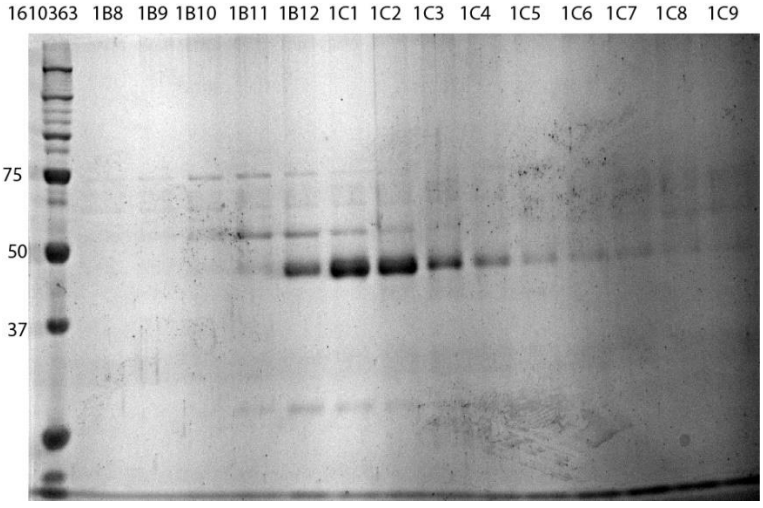
(a)


1A1 1A2 3A1 2A1 2A2 2A3 2A4 2A5 2A6 2A7 2A8 2B1 2B2
150
75
50
37
25

(b)


M C1 D1 D2 D3 D4 D5 D6 D7 D8
75
50

(c)

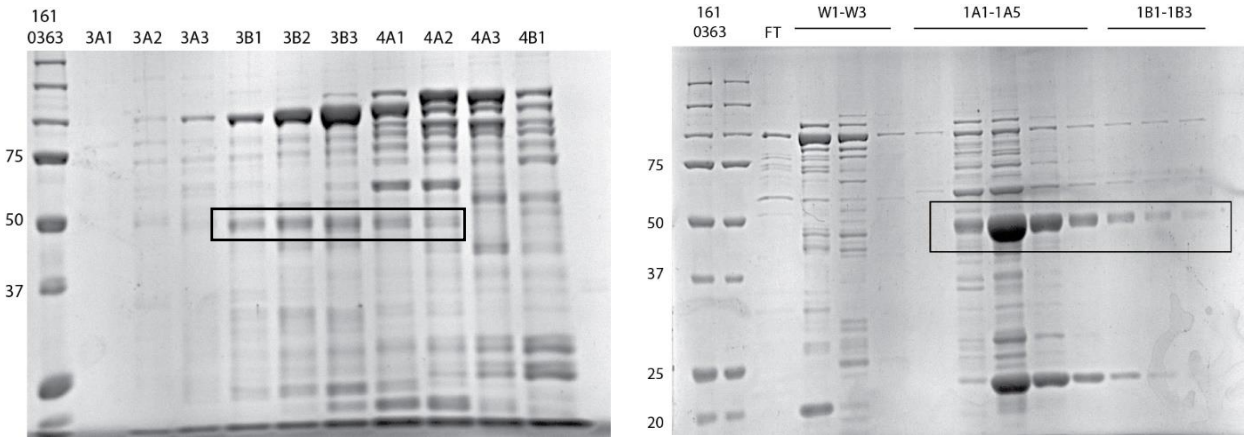

1610363 1B8 1B9 1B10 1B11 1B12 1C1 1C2 1C3 1C4 1C5 1C6 1C7 1C8 1C9
75
50
37

(d)

Figure 3-3. Purification of GSK3β using NiNTA->Ion exchange method. (a) The NiNTA purification result shows the presence of GSK3β (~50 kDa) as a major band on the SDS gel. Presence of other contaminants (~75, 60, & 25 kDa) in higher amount can also be observed. FT-Flowthrough, W1/2/3- Washes, 2A2 to 2A7- Protein elutions. (b) & (c) Ion exchange chromatography results on the gel show almost no removal of any major contaminant from GSK3β. The pools of fractions between 29-77% gradient have GSK3β along with most of the major contaminants. (d) SDS denaturing gel shows the size exclusion chromatography results. The protein pool with GSK3β (~48 kDa) still showed the presence of the higher molecular weight contaminants (~60, ~75 kDa) and one lower molecular weight contaminant (~25 kDa).

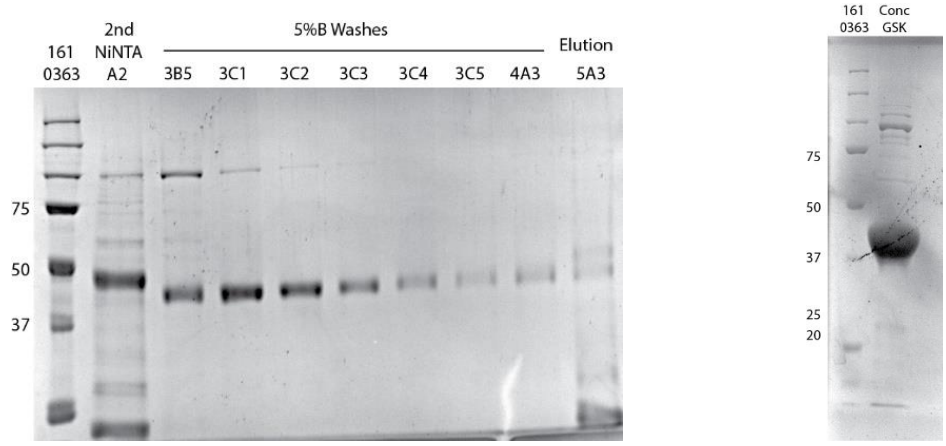
Using the new protocol, cell lysate was first subjected to cation chromatography under a low salt condition (50 mM). Most of the HisGSK3β protein was recovered from the column between 7-34% salt gradient. This step was able to separate some major lower molecular weight contaminants from HisGSK3β. The SP elution pool with the highest amount of GSK3β (3B1-4A1) (Fig 3-4 (a)) was loaded on a 10 mL NiNTA column. HisGSK3β was the major protein in the eluted fractions from this affinity chromatography step (1A1-1A5) (Fig 3-4(b)).

The 6xHis tags were removed from GSK3β in the NiNTA elution pool using TEV protease. A complete and efficient cleavage was achieved after 16 hours of dialysis and cleavage. The TEV/GSK3β mixture was purified using the same 10 mL NiNTA purification protocol as before. Majority of the cleaved GSK3β protein was recovered with a low imidazole (12.5 mM) wash step (Fig 3-5(a), fractions 3B5-3C4). These fractions were pooled and concentrated up to 4 mg/mL using a 3K MWCO concentrator (Fig 3-5(b)). The resulting GSK3β was >90% pure, and it was crystallizable (Chapter 5).



(a) (b)

Figure 3-4 Purification of GSK3β using ion exchange-> NiNTA method. (a) Ion exchange chromatography in the first step elutes most of the His-GSK3β between 7-35% of NaCl (3B1-4A1) (a gradient of 0 M to 1 M salt) 3A1-4B1- Ion exchange chromatography elutions (b) The SDS denaturing gel showed elution of His-GSK3β (1A1-1A5) with 250 mM imidazole when the ion exchange pool of His-GSK was applied to the NiNTA column. FT- Flowthrough, 1A1-1B3- Elutions



(a) (b)

Figure 3-5. Final affinity purification and concentration of GSK3β. (a) A denaturing 12% SDS gel shows fractions with pure post cleavage GSK3β pool in the wash elution of NiNTA (with 12 mM imidazole). 161-0363- Marker, 3B5-4A3- Washes, 5A3- Elution (b) The gel shows a sample of a concentrated and very pure (>90%) GSK3β. Presence of other contaminants is negligible.

3.5 Interaction of HisMBP-hD1 and GSK3β

An *in vitro* complex formation assay was performed by mixing HisMBP-hD1 and GSK3β at a 1:1 ratio at 200 nM in the reaction buffer (RB 1, 50 mM Tris pH 7.5, 5 mM $MgCl_2$, 0.01% Tween-20, 5 mM TCEP). The reaction mix was incubated with 100 µL of NiNTA resin equilibrated with EB 1 (50 mM Tris pH 7.5, 300 mM NaCl, 20 mM imidazole). The resins were washed with 20 CV EB 1, and the complex was eluted with 10 CV of ELB 1 (50 mM Tris pH 7.5, 300 mM NaCl, 250 mM imidazole). The negative control experiment with just HisMBP-hD1 was mixed with the reaction buffer and purified in the same way. Detection of the binary complex formation was done by assessing whether the tagless GSK3β can be pulled down by NiNTA resin that binds to the 6xHis tagged HisMBP-hD1. The negative control contains the same mixture without the HisMBP-hD1 protein. Western blot was done to detect the presence of both proteins in different mixtures.

GSK3β was detected in the wash and elution fractions in the NiNTA pull-down experiment (Fig 3-6). This result indicated possible interaction between both proteins, as GSK3β without His-tag remained bound to His-MBP-hD1 when the latter was being eluted. Unfortunately, GSK3β was also detected in elution fractions of the negative control. This result was classified as a false positive result as GSK3β bound to the NiNTA resin non-specifically.

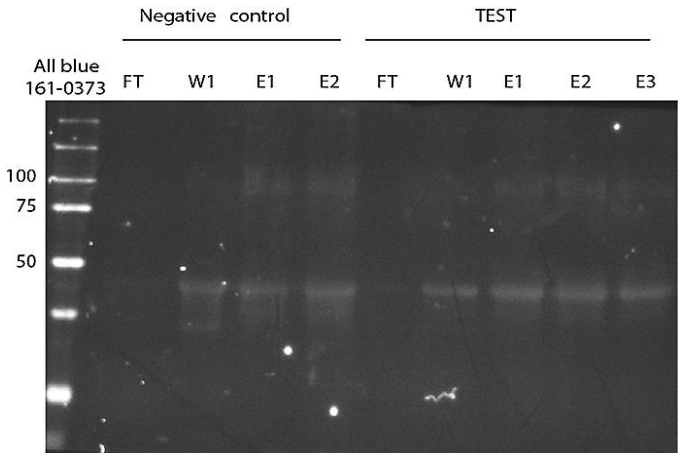


Figure 3-6 Confirmation of interaction experiment. An Anti-His western blot compares NiNTA elutions of the negative control (only GSK3β) and test (HisMBP-hD1 and GSK3β interaction). The presence of GSK3β in the elutions of the negative control is an indication of a false-positive result. FT- Flow-through, W2- Wash, E1/E2- Elution fraction from loose NiNTA resin.

3.6 Discussion

3.6.1 Improved purification results due to change in chromatography sequence

The gel filtration results of the protein pool recovered from NiNTA -> cation exchange protocol show elution of GSK3β with two major higher and lower molecular weight contaminants (~75 kDa and ~25 kDa). Presence of a contaminant in comparable quantity to the protein of interest is not suitable for interaction or crystallization experiments. The second affinity chromatography step in cation exchange -> NiNTA protocol, however, was able to generate pure crystallizable GSK3β. The advantage of cation-exchange -> NiNTA process over the NiNTA -> cation exchange sequence could be due to concentration-dependent interaction between the major contaminant proteins and GSK3β.

It is possible that the interaction between these two proteins is much stronger and maybe irreversible at high concentrations such as NiNTA elution in the first step. Affinity chromatography of cell lysate concentrates the protein elution at the first step. This interaction could be hard to disrupt in the subsequent steps when NiNTA elution fractions are treated with 1mL SP column. In order to deal with this possible problem, we also tried diluting the NiNTA elution fraction before loading them on to the SP cation exchange column. The results were not positive as there was a significant loss in GSK3β when it was diluted in a comparatively purer state. Cation exchange -> NiNTA protocol was in an attempt to counteract this problem where the cell lysate was diluted before being loaded on to the SP column in the first step.

The cell lysate was diluted up to 7 times before it was passed through a 50 mL cation exchange column in the first step. The diluted state of supernatant (10X) maintained a lower density of contaminants and GSK3β around each other, leading to the weak interaction between the 25 kDa contaminant and GSK3β. It may have helped in an efficient separation of the lower molecular contaminant and protein of interest using this protocol.

3.6.2 False-positive interaction result could be due to GSK3β

The interaction experiment between GSK3β and HisMBP-hD1 had false-positive results. It failed as GSK3β bound to NiNTA resin despite having no 6xHis tags. If the experiment had worked, there would have been no signal recorded from elution fractions of the negative control. The suspicion that GSK3β could be the reason for the false-positive test came to light after we took a detailed and closer look at the purification results of the kinase.

In the cation exchange -> NiNTA purification protocol, GSK3β was eluted in the wash fractions (Fig 3-5(a)) of the second NiNTA chromatography after the 6xHis tags were cleaved. Without any histidine tag, GSK3β should ideally be eluted in the flow-through fraction. Elution of the protein in wash fraction suggests GSK3β binds non-specifically to the NiNTA resin. Increasing the imidazole concentration for washing and equilibration can eliminate non-specific binding and could help in obtaining a more conclusive result from the interaction experiment.

3.6.3 The low salt condition may disfavour a physiological interaction

The reaction buffer (RB 1, 50 mM Tris pH 7.5, 5 mM $MgCl_2$, 0.01% Tween-20, 5 mM TCEP) used for the interaction process had a meager amount of salt. Synthetic hD1 is more soluble in the low salt buffer. It is also possible that the low salt buffer could hamper the solubility of bigger proteins like HisMBP-hD1 and GSK3β in the reaction solution. A decrease in solubility could have caused partial or complete precipitation of the proteins on the NiNTA resin. This hypothesis could be tested by using a reaction buffer with a higher salt content (e.g., 300-500 mM) for the interaction experiment that could promote better solubilization of the proteins[94].

3.6.4 Size-exclusion chromatography as an alternative method to test the interaction

Size-exclusion chromatography is often used to monitor the particle size of different protein solutions. This method could also be utilized to facilitate the detection of a possible physiological complex. The interacting proteins would form a molecular species of higher mass and size upon the achievement of a stable complex. GSK3β (~47 kDa) and HisMBP-hD1 (~49 kDa) would elute under their specific peaks when loaded individually. Upon forming a complex, they would create a species of average molecular weight around 96 kDa. The binary protein complex molecule of larger size would elute under a different peak that correlates to its size. This shift in the molecular weight could be used as a tool to detect complex formation. Deciding the nature of the interaction could become a challenge if the interaction leads to a heterogenous mixture of different oligomers.

4 Chapter 4: HisGSK3β-StrnhD1 (co-expression construct)

4.1 Rationale

Co-expression of interacting proteins is one method to yield physiological protein complexes. A T7-promoter-based poly-cistronic vector was used to express GSK3β and hD1 in the same vector. The rationale of co-expression strategy relies on the availability of a cellular milieu for better interaction between both proteins of interest. If their interaction *in vivo* requires simultaneous expression or folding, this method may be deemed helpful for complex formation. As described in the method section, the presence of different tags in different proteins provides modularity for downstream affinity purification processes. In case they form a stable complex, affinity purification of one protein should ideally result in the co-elution of the other.

4.2 Overview

The current chapter discusses the results derived from co-expression strategy. As the co-expression strategy was attempted for the first time, overexpression and purification experiments needed optimization for optimal yield of the proteins. The overexpression of the protein was quantified across multiple bacterial strains and overexpression parameters. Purification tests were performed along with respective western blots to confirm each protein. The crystallization trial of the purified GSK-StrnhD1 is also covered in this chapter.

4.3 Purification of the HisGSK3β-StrnhD1 construct

Protein expression levels needed to be tested for the novel HisGSK3β-StrnhD1 construct. This co-expression construct contained both the interacting proteins with their respective N-terminal affinity tags. Overexpression tests were done with three cell strains: BL21 Star (DE3), BL21 Star (DE3) pTf16, and Rosetta 2 (DE3) pLysS. BL21 Star (DE3) pTf16 is an expression system with a pTf16 plasmid. This plasmid codes for pTf16 chaperone which could help in better folding and expression of our proteins of interest[95]. Protein expression levels were compared between pre- and post-induction samples from cultures of all three systems. All cultures in this

experiment were grown under the same conditions. A protein band correlating to the molecular weight of HisGSK3β (~47 kDa) was visible in the post-induction sample (lysate) on the denaturing gel. This optimization test revealed a visibly higher expression level of GSK3β in BL21 Star (DE3) and BL21 Star (DE3) pTf16 systems (Fig 4-1(a)).

Overexpression of the HisGSK3-StrnhD1 construct was done for the first time, hence confirmation of individual protein expression was necessary. As the band correlating to the HisGSK3β molecular weight had a visibly higher band density, and western blot with mouse anti-Penta His Alexa 488 confirmed the identity. The test showed that the overexpressed protein was, in fact, HisGSK3β (Fig 4-1(b)).

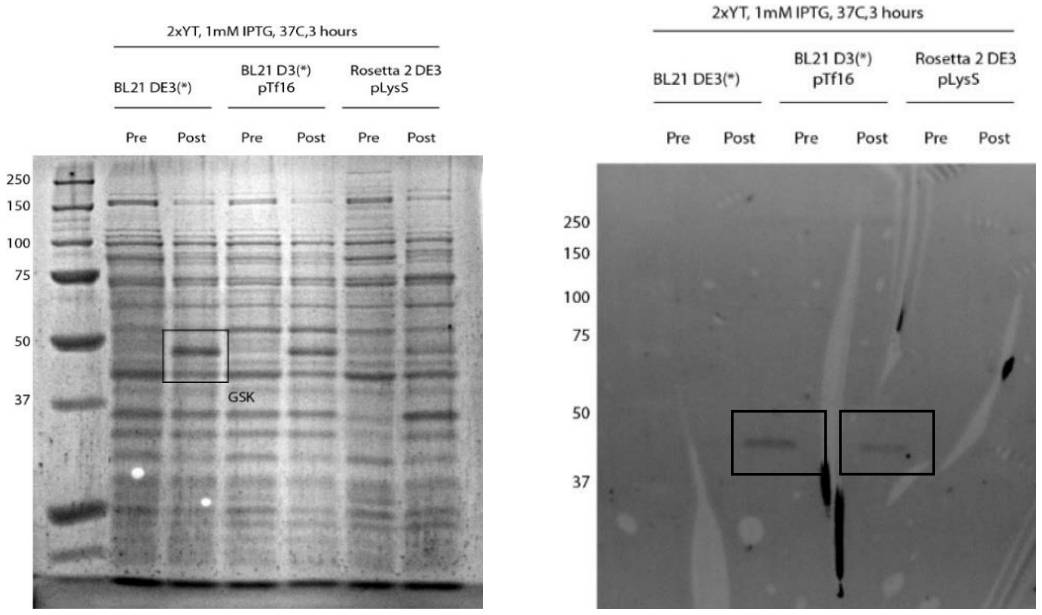


Figure 4-1. Overexpression confirmation of HisGSK3β from co-expression construct. (a) A 12% denaturing gel depicts a comparison of His-GSK3β (~49 kDa) expression between pre- and post-induction samples from 3 different cell strains. A putative GSK3β band is visible in the post-induction lanes of BL21 DE3 (*) and BL21 DE3 (*) pTf16 cell strains. GSK bands are highlighted with rectangular boxes. An overexpression in 2xYT media for 3 hours at 37˚C using 1 mM IPTG was followed for all cell strains. (b) The western blot with anti-PentaHis confirms His-GSK3β in the post-induction lanes of the two cell lines.

Expression confirmation of HisGSK3β from the HisGSK-StrnhD1 construct was done with two systems, BL21 Star (DE3) and BL21 Star (DE3) pTf16. This preliminary test was done at 37ºC where expression was induced at 1 mM IPTG for 3 hours. More optimization tests were necessary to explore higher expression levels under different conditions. Lower temperature (16ºC) induction and growth over a more extended period (overnight) were tested for both the expression systems. A much higher protein expression level was observed in BL21 Star (DE3) pTf16 system. The optimized overexpression parameter is as follows, induction with 0.15 mM IPTG at 16ºC and total growth duration of overnight (Fig 4-2).

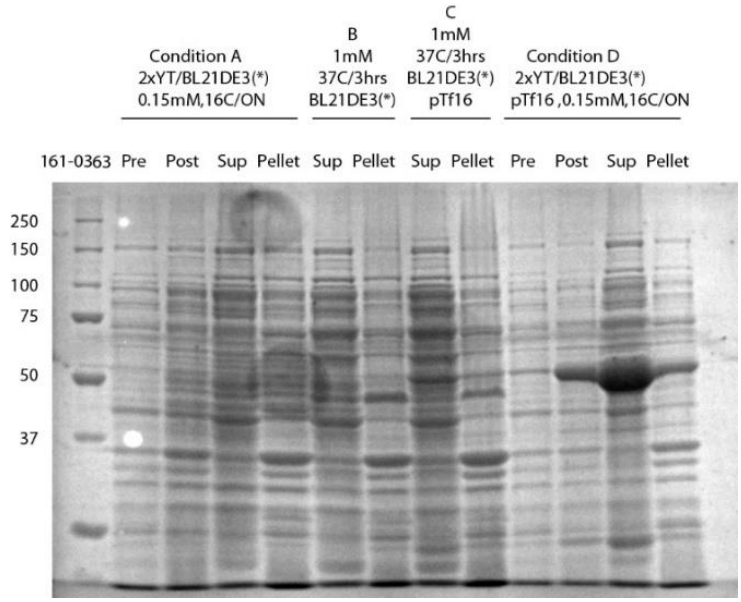
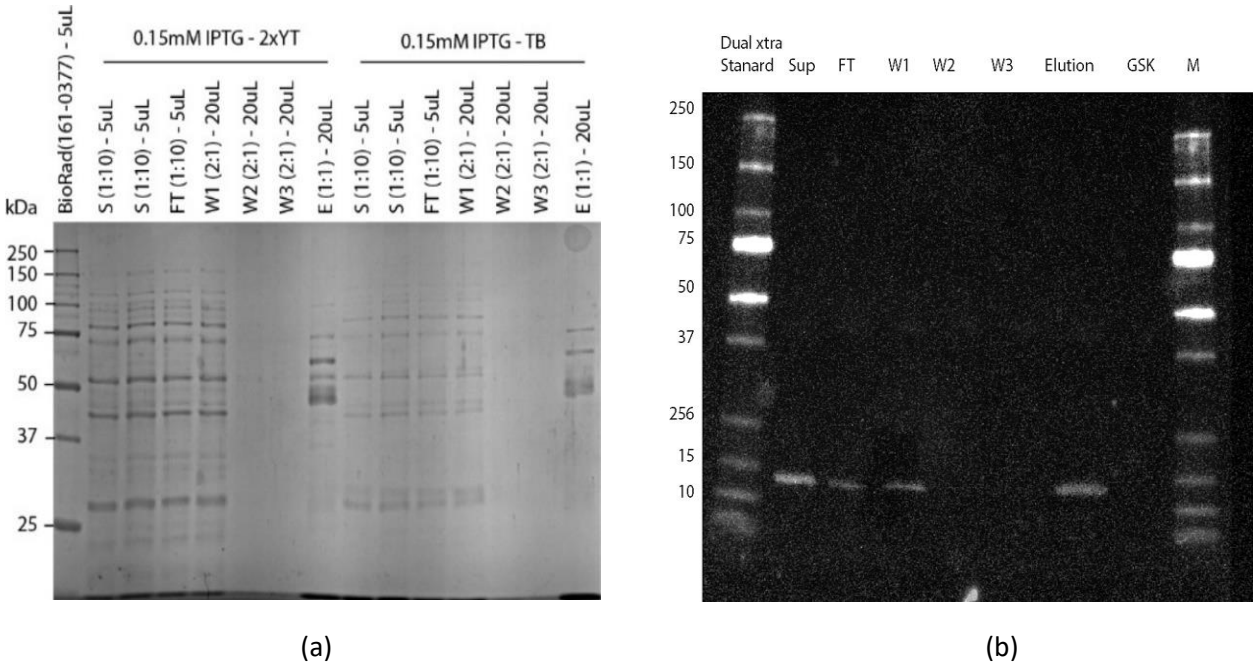


Figure 4-2. Overexpression optimization of the co-expression construct. Overexpression parameters were optimized (Conditions A/B/C/D) for the BL21 DE3 (*) and BL21 DE3 (*) pTf16 cell strain. A very high level of His-GSK3β (~49 kDa) expression was observed in BL21 DE3 (*) pTf16 cell strain when protein expression was induced with 0.15 mM IPTG at 16˚C, and it was grown for overnight. Most of the expressed protein was also present in the soluble fraction. Sup- supernatant. Pre and Post- Pre and post-induction samples, Sup- Soluble supernatant fraction, Pellet- insoluble fraction.

A longer period of growth can lead to a change in the pH of the culture media, and a destabilized pH can potentially hamper proper expression and folding of proteins. Therefore, a final optimization test was done to compare the final yield of protein from 2xYT media and a pH-controlled TB media. No significant difference was observed when proteins were eluted from a small-scale purification (0.5g cell weight) using 50 µL NiNTA resin (Fig 4-3(a)). This experiment showed the successful purification of the HisGSK3β from HisGSK-StrnhD1 co-expression construct.

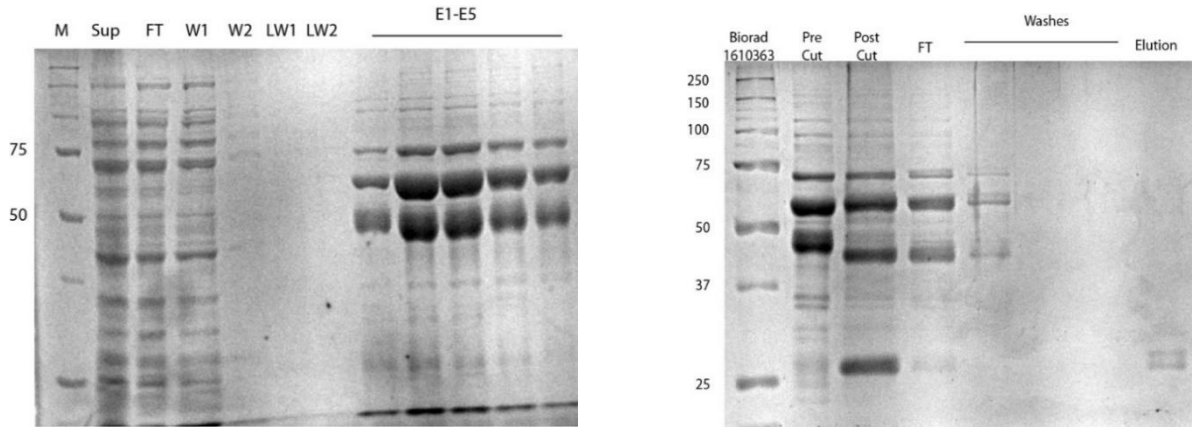


(a) (b)

Figure 4-3. Small scale purification and detection of the possible complex. (a) A denaturing gel shows similar levels of overexpression between 2xYT and TB media (Lane marked with E(1:1) under both 2xYT and TB media. A small-scale purification using NiNTA (50µL) method yields HisGSK3β as the main protein in the elution. Some endogenous chaperones and proteins are also present along with GSK. (S- Supernatant, FT- Flowthrough, W1/2/3- Washes, E1- Elution) Volume written beside the labels is the amount of denatured protein loaded on the gel. (b) A western blot with anti-strep antibody detects the presence of StrnhD1 in the elution fraction as well as supernatant, flow-through and wash from NiNTA step. Presence of Strep signal provides strong evidence for complex formation. StrnhD1 not involved in the interaction is seen in sup/FT/wash. Dual Xtra standard/M- protein marker, Sup-Supernatant, FT- Flowthrough, W1/2/3- Washes, GSK- Control pure GSK

The goal of co-expression construct was to express both proteins in the same system, possibly yielding a physiological complex. Both proteins were cloned with different tags for different affinity purifications. If they form a stable complex, then purifying one protein should ideally elute the other. His-GSK3β had a distinctive thick band in the denaturing gel of Fig 4-3 (a). However, the peptide Strn-hD1 band was not distinctively visible in the elution fractions of the small-scale purification results (Fig 4-3(a)). Hence, western blot using anti-strep antibody was used to detect the peptide and the putative complex. Signal for Strn-hD1 peptide (~6.8 kDa) was observed at ~12 kDa molecular weight range in the elution fractions containing HisGSK3β (Fig 4-3(b)). Presence of the peptide along with HisGSK3β strongly suggested the presence of a binary protein complex.

Crystallization trials require a pure and concentrated protein sample. A larger-scale purification (2.5 g cell) was done to yield enough protein sample for the crystallization experiment. Because of the higher expression level, HisGSK3β was tracked through subsequent purification experiments. Approximately 300 µg of HisGSK3β was recovered using 500 µL loose NiNTA resin. The elution fractions, however, contained equal proportions of endogenous contaminant proteins (~60 kDa, & ~75 kDa) (Fig 4-4(a)). The histidine tag was removed from GSK3β using TEV protease, and cleaved protein was again purified using NiNTA affinity chromatography. The cleaved protein remained bound to the contaminants and was recovered from the flow-through fraction (Fig 4-4(b)).

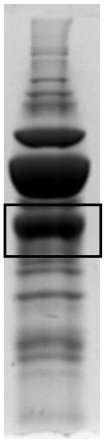


(a) (b)

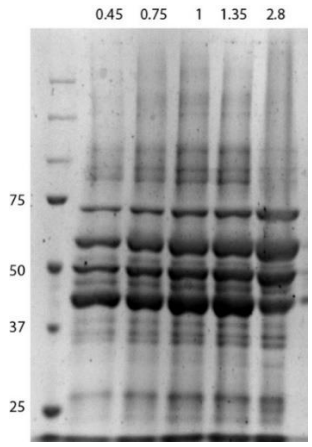
Figure 4-4. Affinity purification and His-tag cleavage. (a) NiNTA purification of a 2.5 g cell batch from the co-expression construct provides an elution fraction on a denaturing 12% SDS gel with HisGSK3β as one of the major proteins. Other significant bands present could be chaperone pTf16 (~56 kDa) and an endogenous chaperone (~75 kDa). M- Marker, Sup- supernatant, FT- Flow-through, W1/2/LW1/LW2- Washes, E1-E5- Elution fractions. (b) 6xHis tag cleavage using TEV protease was able to achieve a very efficient cleavage of GSK3β as seen from the SDS gel. A comparison of pre-cut and post-cut sample showcases lower band size of GSK3β (~47 kDa) compared to HisGSK3β (~49 kDa). Most of the cleaved protein was collected from flow-through and wash fractions as expected. Major contaminant bands, however, were still eluted with GSK, requiring a size exclusion chromatography. Pre-cut- Sample before adding TEV protease, Post-cut- Sample was taken when cleavage was complete. FT- Flow through.

Size exclusion chromatography efficiently separates proteins based on the difference in their molecular weight. A maximum resolution of the proteins is achieved when the load volume is between 0.5-2% of the total column volume. A maximum sample size of 500 µL is required for the Superdex 200 15/300GL column with 24 mL bed volume. The concentration step of the TEV treated GSK3β-StrnhD1 sample before the size exclusion step revealed a gradual loss of GSK3β (Fig 4-5(a)). Protein samples from different stages of the concentration steps were tested to determine an optimal concentration value. Running these samples on a denaturing gel revealed a significant loss of GSK3β once the concentration value crossed 1.35 mg/mL. A maximum concentration of 1.3 mg/mL was targeted for gel filtration load sample in further experiments (Fig 4-5(b), (c)). The possible binary protein complex was eluted in the void fraction of the repeated gel filtration experiment. A possible higher-order oligomerization or aggregation was suspected because of this result. A significant loss of contaminants was observed in the void volume fraction with the protein complex. Concentrating this protein pool for crystallization, however, revealed

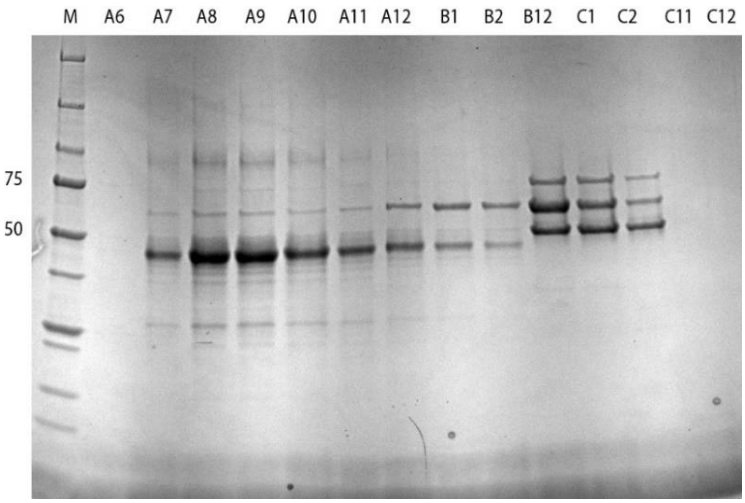
the presence of contaminants that were not visible in the diluted pool. The void volume protein was concentrated up to 6 mg/mL (Fig 4-5(d)).



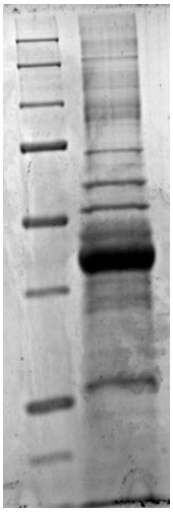
(a)

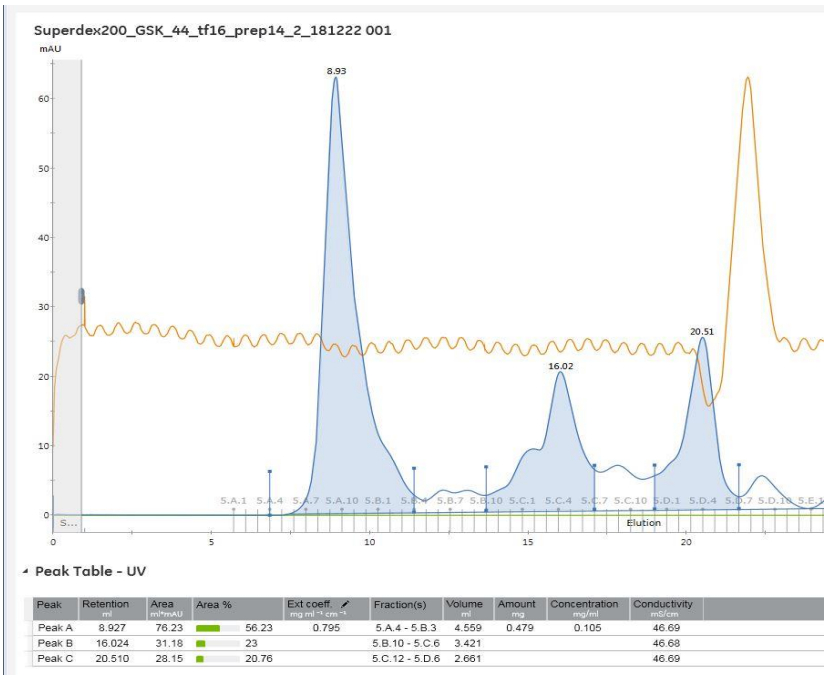


(b)



(c)





Peak	Retention ml	Area mAU*ml	Area %	Ext coeff. mg ml⁻¹ cm⁻¹	Fraction(s)	Volume ml	Amount mg	Concentration mg/ml	Conductivity mS/cm
Peak A	8.927	76.23	56.23	0.795	5.A.4 - 5.B.3	4.559	0.479	0.105	46.69
Peak B	16.024	31.18	23		5.B.10 - 5.C.6	3.421			46.68
Peak C	20.510	28.15	20.76		5.C.12 - 5.D.6	2.661			46.69

(d) (e)

Figure 4-5. Size exclusion chromatography of His-tag cleaved GSK3β. (a) A denaturing gel shows bands of concentrated TEV cleaved GSK3β pool at 3 mg/mL. GSK3β is displayed inside a rectangular box. The ratio of the contaminants to GSK3β is alarmingly high compared to pre-concentrated fractions. (b) A concentration test reveals loss of GSK3β from the cleaved protein pool when the concentration reaches more than ~2 mg/mL. A sample concentrated to ~1.5 mg/mL was loaded on a size exclusion chromatography column. (c) Fractions from size exclusion chromatography revealed the GSK3β elution in the void volume (A7-A11) of the gel filtration column Superdex 200 15/300GL. Other larger contaminants were eluted at later fraction sizes correspondent to their molecular sizes. (d) GSK3β sample collected from the void fraction and concentrated to 6 mg/mL at ~70% purity. (e) A chromatogram shows three main peaks of the gel filtration chromatogram, where the first prominent peak represents elution of the complex, and subsequent peaks represents other major contaminants.

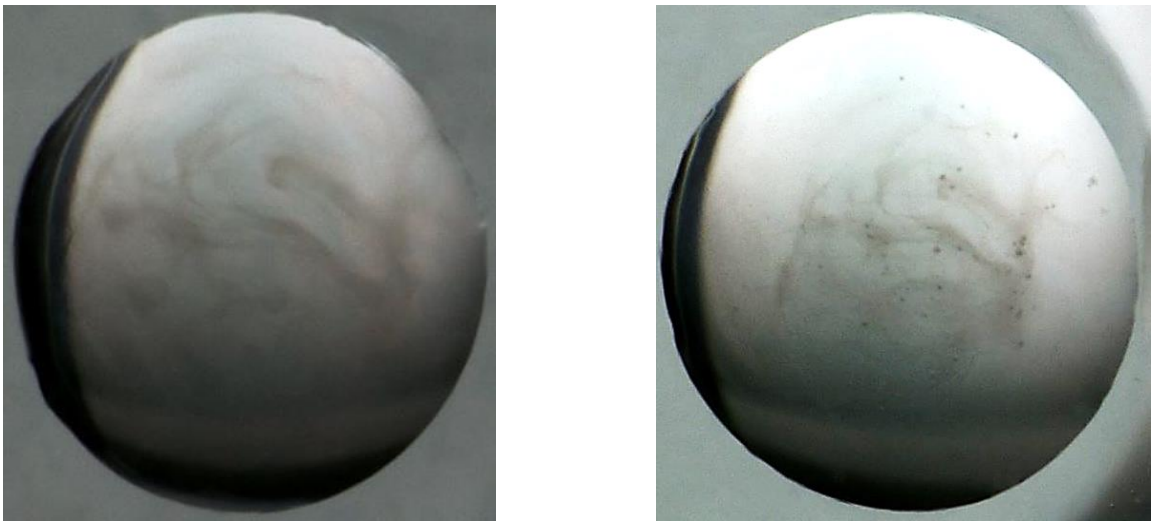
4.4 Biochemical characterization and Crystallization

DLS and crystallization trials were performed with the GSK3β-StrnhD1 sample purified as described above. The theoretical molecular weight of the binary protein complex would be ~55 kDa (assuming the interaction is 1:1). Still, it eluted in the void volume of the size exclusion chromatography step (molecular weight of the complex being more than 440 kDa, according to the manual). DLS sheds light on an approximate molecular size of the most common molecular species present in a protein solution, so we decided to use DLS before using the sample for a widescreen crystallization trial. The most common species in the solution had a molecular weight of 483 kDa (Table 4-1). A high molecular weight could be due to a nonameric complex formation or aggregation due to the presence of other contaminants with the complex.

Table 4-1. DLS data for the HisGSK3β-StrnhD1 co-expression protein

Intensity distribution	Radius (nm)	% Pd	Mw-R (kDa)	% Intensity	% Mass	% Number
Peak 1	8.353	11.3	483	3.8	66.7	99.0
Peak 2	36.229	35.6	14965	96.2	33.3	1.0

The purified GSK3β-StrnhD1 was subjected to crystallization using the commercially available sparse matrix JCSG+ screen. The development of dense precipitate like possible crystalline structures, was observed in the A9 condition, which contains 0.2M ammonium chloride, 20% w/v PEG 3350 as the mother liquor (Fig 4-6). This strategy was not further pursued as the strategies described in the next chapter gave us more promising results.



(a) (b)

Figure 4-6. Crystallization of co-expression complex. (a) JCSG+ sparse matrix widescreen was used to find a crystallizing condition for the co-expression complex. Picture shown is for condition A9 from day 1. (b) The same well from day 3 can be seen to have developed many densely precipitated structures throughout the crystallization drop.

4.5 Discussion

4.5.1 Dimer-like behavior of the Strn-hD1 peptide in SDS gel

The anti-strep blot in Fig 4-3(b) was performed to identify strnhD1 as a part of the physiological complex with GSK3β. The signal for Strep-tag was, however, detected at ~12 kDa, which is twice the theoretical molecular weight of Strn-hD1 (~6.8 kDa). This signal suggested a SDS-stable dimerization of the peptide. The hD1 peptide sequence had never been reported to dimerize under denaturing conditions. Thus, this finding requires literature and experimental research to determine the actual cause.

A study published by Walkenhorst (2009) on decreased electrophoretic mobility of small peptides in SDS-PAGE suggests a specific interaction between Asparagine (Asn) and SDS detergent micelles as a driving factor. The slower movement of the peptide could be the reason for the misleading dimerization results[96]. The protein sequence of StrnhD1 also contains an asparagine residue at the 19th position as part of the TEV cleavage site. The polar asparagine residue may lead to dimerization-like results when present with detergents like SDS.

If Asn19 on StrnhD1 was responsible for the slower electrophoretic mobility and cause the StrnhD1 to migrate like a dimer on the SDS gel, then TEV cleavage of the peptide, removing Asn19, should lead to an SDS band that reflects the molecular weight of a monomeric StrnhD1. Since the western blot results were done with the samples from the first affinity chromatography test before it was treated with TEV protease. A future blot with a TEV-treated protein sample could support this hypothesis.

4.5.2 Improvement of the purification protocol

An essential prerequisite of the crystallization process is the purity of the protein. While a high quantity of the possible protein complex has been achieved, we only managed to get a final purity of ~70%. A combination of many chromatography techniques could be applied to improve the purity level of the final yield in future experiments.

Talon resin (CoNTA) could be used instead of NiNTA resin in the first step for more specific binding. Talon resin uses Cobalt ion (Co^{2+}) instead of Nickel ion (Ni^{2+}) as the immobilized metal ion. Talon resin has also shown a more specific binding of poly-histidine tags under milder elution conditions. This could potentially separate some of the contaminants in the first step[97].

The GSK3β-StrnhD1 protein complex elutes in the void volume fraction of Superdex 200 increase 10/300 GL column. Resolution of the protein separation for this column lies between 10 kDa to 500 kDa. One of the general causes for elution of proteins in the void volume fraction is either higher oligomerization or aggregation of the protein. The higher oligomerization in this case seems unlikely as the upper limit of resolution is ~10 times the theoretical molecular weight of the complex (~55 kDa). This hypothesis is based on the assumption that the stoichiometry of GSK3β and hD1 interaction is 1:1. A solubility screening experiment using DLS could determine the proper buffer environment to prevent aggregation of the complex. An oligomeric and properly folded complex would possibly elute in gel filtration fraction corresponding to its size.

However, the protein size and molecular weight estimation from size exclusion chromatography is usually based on the assumption that the protein is globular in nature. But, if the GSK3β-StrnhD1 complex adapts an elongated fibrous shape instead of globular shape, it would elute in the void volume fraction. The size estimate from the size exclusion chromatography will not be accurate despite the molecular weight of the protein complex being within the fractionation limit of the column. Indeed, preliminary negative-stained EM images of GSK3β/hD1 (discussed in the next chapter) suggests the complex may be fibrous and elongated in shape.

4.5.3 Crystallization hit requires further optimization

The crystallization screening using JCSG+ sparse matrix gave a preliminary hit. Most of the structures were clusters of spherulites with a rounded morphology. It is necessary to try a vast range of parameters to improve the quality of the crystals. A fine screen optimization of the protein complex is essential. In future trials, the PEG percentage and type can be used as variables over a wider range. Individual spherulites were found with 20% PEG 3350. So a range of 10-30% of PEG 3350 could be a start. In addition, the concentration and the type of monovalent salt could be varied to investigate the effect of ionic environment on crystal growth.

5 Chapter 5: GSK/hD1Peptide

5.1 Rationale

Crystallization of a protein complex relies on its purity, quantity, and interaction parameters. Our lab has investigated the biophysical and the biochemical nature of the interaction between GSK3β and a synthetic hD1 peptide. So, my priority shifted to obtaining co-crystals of GSK3β/hD1 using conditions we learned from the *in vitro* assays.

5.2 Overview

This chapter describes the co-crystallization methods used for GSK3β, hD1, and FRATide. Crystallization of GSK3β was tested first to check the quality and purity of the protein. The GSK3β crystals were soaked in various concentrations of hD1 to yield a binary complex crystal but soaking caused damage to the crystals. The failed soaking experiment led us to try co-crystallization of GSK3β and hD1 after they interacted *in vitro*. Optimization of sample preparation to ensure protein complex formation was done before co-crystallization.

5.3 Crystallization of GSK3β with synthetic hD1

The successful purification of GSK3β in high purity and quantity was followed by the crystallization of the protein. Based on the published conditions by Aoki *et al*. (2000), purified GSK3β at 6 mg/mL was crystallized using the sitting drop method[98]. Multiple thin square like crystals of GSK3β with a size ~ 0.3 x 0.5 mm were observed in the condition containing 0.1 M Tris pH 8, 20% PEG 8000 (Fig 5-1). Crystals were harvested and tested for diffraction at the home source X-ray diffractometer. These diffracting crystals were soaked for 5 minutes in 2.5 μM hD1 solution made with the harvest buffer (10 mM Hepes pH 7.5, 14% PEG 8000, 10% ethylene glycol). The diffraction spots from the hD1-soaked GSK3β crystals was streaky (Fig 5-2(c), Fig 5-3), suggesting damage to the crystal. The streaky pattern often arises due to structural defects in the crystal packing due to extra soaking.

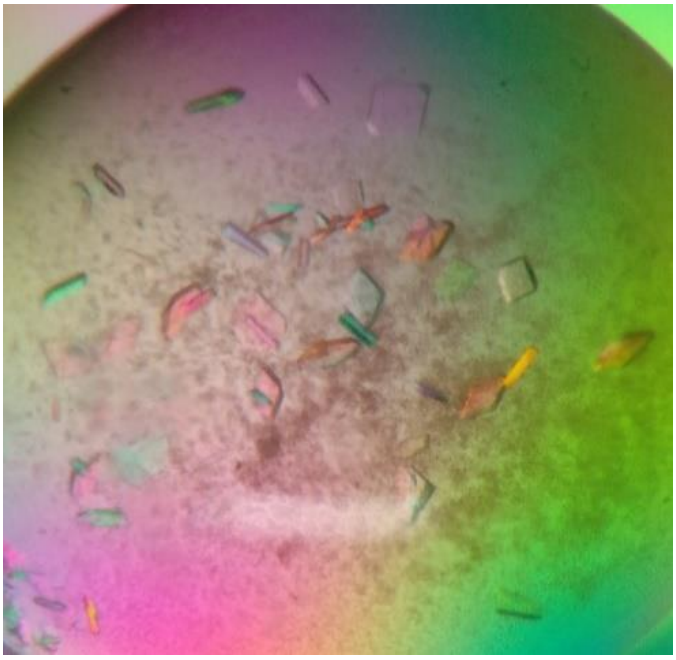
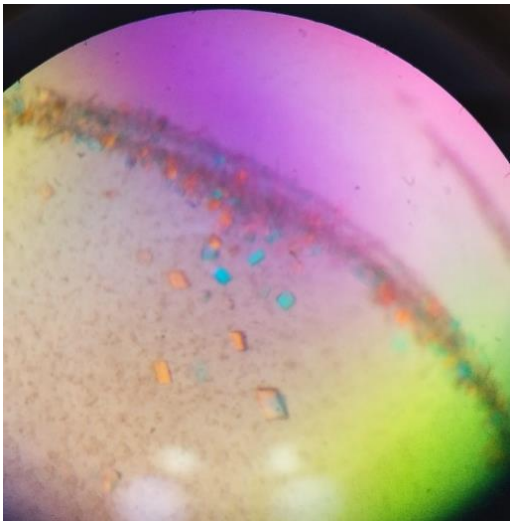


Figure 5-1. GSK3β crystals. GSK3β crystals seen through a polarizer lens in a sitting drop plate. Crystallizing condition was 0.1 M Tris pH 8, 20% PEG 8000. The square plate-like crystals grew mostly in 2 dimensions leaving the crystals very thin. Most of the crystals were singular and diffracted well.

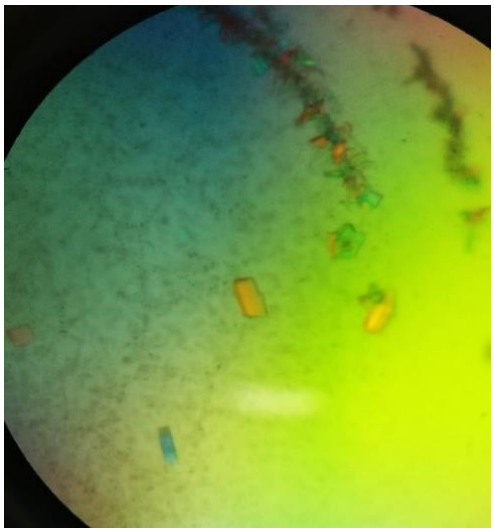
GSK3β from the same batch was simultaneously used for the interaction and co-crystallization experiments. GSK3β protein and hD1 peptide were mixed to interact at 8 µM concentration, at a molar ratio of 1:1. The lyophilized powder of synthetic hD1 peptide was dissolved in the buffer with no monovalent salt (50 mM Tris pH 7.5, 5 mM $MgCl_2$, 0.005% T-20, 1 mM TCEP) to make a 500 µM stock solution. Then, the peptide stock was diluted with an equal volume of the high salt buffer (50 mM Tris pH 7.5, 5 mM $MgCl_2$, 300 mM NaCl, 0.005% T-20, 1 mM TCEP), resulting in a working stock solution of the peptide with 150 mM NaCl in the buffer, which is the same as the buffer for the interaction assay. 8 µM of the working peptide solution was mixed with GSK3β in the same buffer and allowed to incubate for 30 minutes at room temperature. This interaction condition was based on the published SPR protocol (Thesis by Saundh, Stephanie. Binding kinetics investigation of a peptide fragment of the psychiatric risk protein, DISC1, to the kinase GSK3β (2019)). The binary protein complex was concentrated up to 8.7 mg/mL. The complex was subjected to crystallization using the sparse matrix JCSG+ screen.

At least three conditions from the widescreen test provided promising crystallizing conditions for the possible protein complex. These three conditions were used as a reference for a fine screen

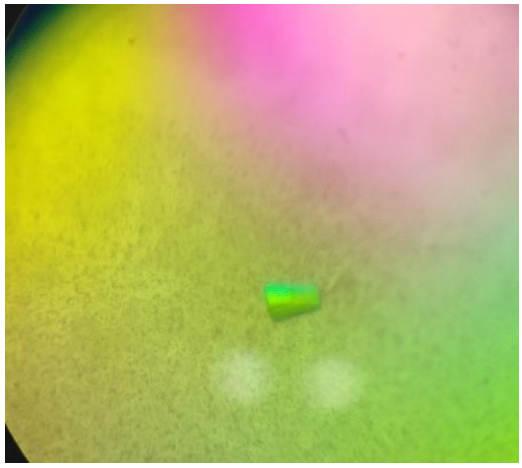
matrix used to fine-tune the quality of crystals (Fig 5-2(a), (b), (c)). Three major conditions with good crystal hits were observed within a week. The list of conditions is mentioned in Table 5-1 following the crystal figures.



(a)



(b)



(c)

Figure 5-2. GSK-hD1 crystals. Set of GSK-hD1 crystals achieved using fine screening, based on three separate hits from JCSG+ (B4, E10, E12). The following table contains the details of crystallizing conditions pictured above.

A series of freezing/diffraction test revealed that 20% ethylene glycol was the optimal cryoprotectant for these crystals. The co-crystals (Fig 5-2) were harvested and cryoprotected in the presence of 2.5 μM hD1 and incubated for 5 minutes before freezing. The diffraction spots were observed at 2.8 Å resolution, at CMCF-ID beamline of CLS, USask. However, the spots were streaked, indicating that the harvesting procedure needed further optimization.

Table 5-1. Crystallizing conditions for GSK-hD1 crystal hits. The table lists the crystallization conditions and quality of crystals depicted in Figure 5-2

Figures	**Crystallization condition**	**Approximate size and shape** (for the one used for diffraction)
5-2 (a)	0.1 M Hepes pH 7.5 8% Ethylene glycol 8% PEG 8000	0.1 mm x 0.05 mm A group of the square plate-like thin crystals as well as many tiny joint crystals
5-2 (b)	0.1 M Hepes pH 7.5 8% Ethylene glycol 10% PEG 8000	0.1 mm x 0.05 mm Most of the crystals were tiny and clustered, very few of them were single
5-2 (c)	0.1 M Bicine pH 9 10% PEG 6000	0.2 mm x 0.1 mm X 0.5 mm One big crystal with an optimum thickness unlike other conditions

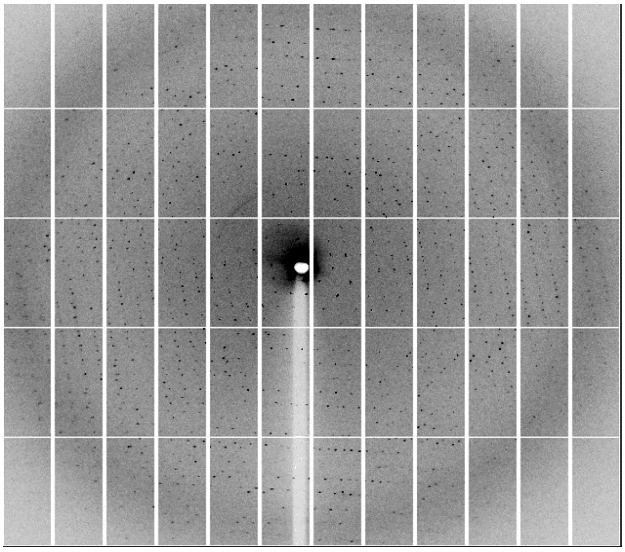


Figure 5-3. Diffraction data collected from crystals of Fig 5-2(c). The diffraction pattern shown in the picture was streaky, suggesting crystal damage due to extra peptide soaking, and the data was not of sufficient quality to solve the structure**.**

To avoid crystal damage, we attempted to harvest the co-crystals without soaking with additional peptide, from which a better data set was obtained. Diffraction data from the new batch of co-crystals were recorded up to 2.23 Å. Molecular replacement was successfully performed. Two GSK3β molecules are present in each a.s.u., but no density for the hD1 peptide was observed (Fig 5-4).

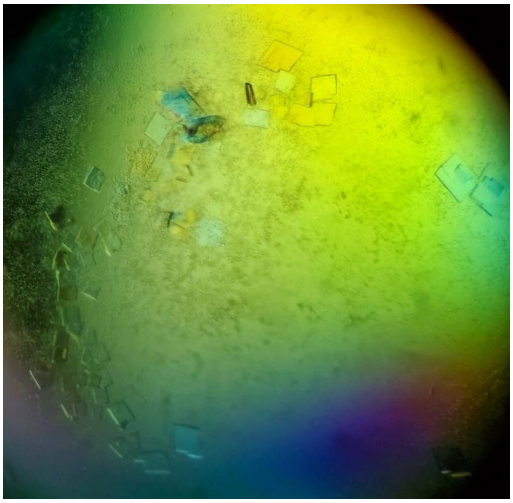


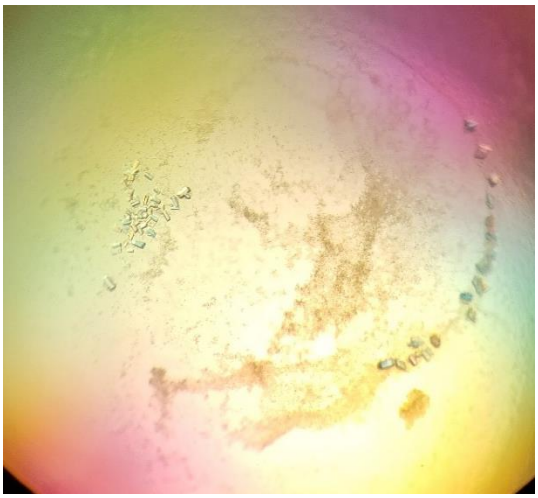
Figure 5-4. Crystals of GSK3β-hD1 at 8 µM. Single and square plate-like crystals with an average size of 0.3 mm x 0.5 mm were obtained with protein interaction at 8 µM.

Table 5-2. Structural data analysis for the crystal from Fig 5-4. Values in the brackets are for the highest resolution shell (2.30-2.23 Å).

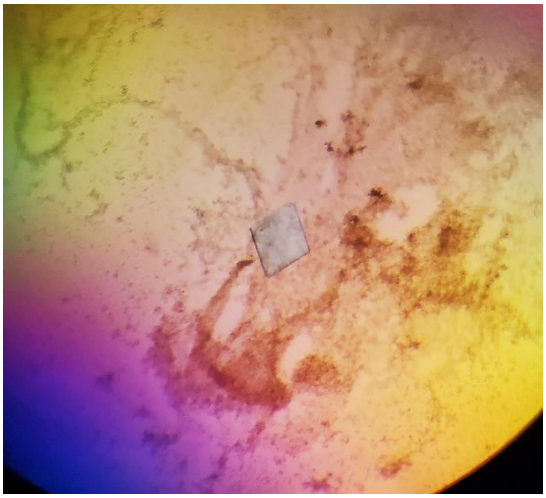
Observed parameters	Values
Wavelength (Å)	0.97949
Space group	P1
Unit cell (Å) and angles	63.75, 67.27, 67.40, 78.1, 76.8, 89.4
All reflections	170448
Unique reflections	49770
Multiplicity	3.4
Resolution (Å)	46-2.23
Completeness *	95.6% (81%)
Multiplicity #	3.4 (3.1)
I/Sigma (I)	6.9 (0.3)
$CC_{1/2}$	0.994
R_{merge} $	0.116 (3.195)

* The completeness represents the percentage of reflections recorded compared to theoretically possible measurements. # Multiplicity shows the quantity of redundant measurements. $ R_{merge} stands for internal R factor.

A careful review of GSK3β structure article published by Dajani et al. (2001) suggested that GSK3β forms a dimer at micromolar concentration. Since we prepared our previous complex by mixing GSK3β and hD1 at 8 μM concentration, we reasoned that GSK3β would exist as a dimer before the addition of hD1. If the binding surface of hD1 is at the GSK3β dimer interface, then a GSK3β/hD1 would not be possible in the condition that we used for complex formation; therefore, we repeated sample preparation by mixing the proteins at a concentration of 200 nM and repeated crystallization screen with the widescreen sparse matrix JCSG+. Indeed, we discovered multiple new hits with the new sample. In contrast to only observing crystals with square, plate-like morphologies, we obtained crystals with diamond and needle shapes in different conditions. This change strongly suggested a possible structural change at the molecular level. Crystals obtained from the fine screen based on the widescreen hits are listed in the following figure (Fig 5-5(a), (b), (c)).



(a)



(b)

(c)

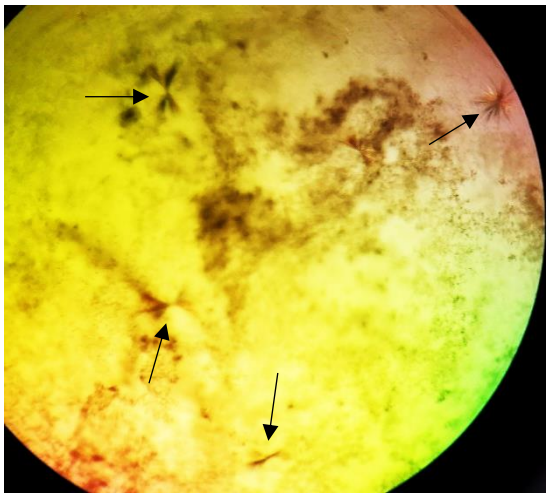


Figure 5-5. GSK-hD1 crystals at 200 nM. Crystals obtained from GSK3β-hD1 interaction experiment done at 200 nM. Arrows in fig (c) point towards crystalline needle structures. The crystallizing conditions are listed in the following table.

Table 5-3. Crystallization conditions for crystals obtained from the interaction at 200 nM.

Figures	**Crystallization condition**	**Approximate size and shape** (for the one used for diffraction)
5-5 (a)	0.1 M Hepes pH 7.5 8% Ethylene glycol 8% PEG 8000	0.03 mm x 0.03 mm A group of the small square plate-like thick and joined crystals as
5-5 (b)	0.1 M Hepes pH 7.5 8% Ethylene glycol 10% PEG 8000	0.5 mm x 0.5 mm One big square plate like crystal
5-5 (c)	0.1 M Bicine pH 9 8% PEG 6000	Multiple needle flowers, size less than 0.05 mm

Structural data were obtained from crystals of Fig 5-5 (b) and (c). The data statistics for those two crystals are listed in tables 5-4 and 5-5, respectively.

Table 5-4. Structural data statistics of crystals from Fig 5-5(b). Values in the brackets are for the highest resolution shell (3.08-2.90 Å).

Observed parameters	Values
Wavelength (Å)	0.97949
Space group	P121
Unit cell (Å) and angles	67.41, 113.48, 67.51, 90.0, 102.63, 90.0
All reflections	100914
Unique reflections	22003
Multiplicity #	4.6 (4.7)
Resolution (Å)	113.59-2.90
Completeness *	99.9% (100%)
I/Sigma (I)	9.53 (1.26)
$CC_{1/2}$	0.95
R_{merge} $	0.119 (0.424)

* The completeness represents the percentage of reflections recorded compared to theoretically possible measurements. # Multiplicity shows the quantity of redundant measurements. $ R_{merge} stands for internal R factor.

Table 5-5. Structural data statistics of crystals from Fig 5-5(c). Values in the brackets are for the highest resolution shell (3.08-2.90 Å).

Observed parameters	Values
Wavelength (Å)	0.97949
Space group	P1 21 1
Unit cell (Å) and angles	106.78, 85.64, 116.50, 90.00, 95.3, 90.00
All reflections	113827
Unique reflections	35500
Multiplicity #	3.2 (3.3)
Resolution (Å)	116.07-3.15
Completeness *	99.9% (100%)
I/Sigma (I)	1.5 (0.1)
$CC_{1/2}$	0.972
R_{merge} $	0.244 (4.485)

* The completeness represents the percentage of reflections recorded compared to theoretically possible measurements. # Multiplicity shows the quantity of redundant measurements. $ R_{merge} stands for internal R factor.

5.4 Crystallization of GSK3β with synthetic hD1 and FRATide

The hypothesis of my project predicts that the binding of FRATide to the GSK3β disrupts the binding of the GSK3β inhibiting region of hD1 GI to the kinase. We tried to crystallize the GSK3β with hD1 peptide using multiple strategies. However, to prove the hypothesis I suggested, a FRATide and hD1 bound GSK3β crystal structure is necessary. The comparison of structural framework changes between GSK3β-hD1 and GSK3β-hD1-FRATide crystals is required to prove the hypothesis.

To examine the effect of FRATide binding, an interaction and co-crystallization experiment was performed. FRATide, hD1, and GSK3β were mixed at 200 nM in reaction buffer (50 mM Tris pH 7.5, 5 mM MgCl2, 150 mM NaCl, 0.005% T-20, 1 mM TCEP) and allowed to interact for 30 minutes. The interaction pool was concentrated up to 7.8 mg/mL and subjected to the sparse matrix widescreen, JCSG+ screen. Fine screens based on preliminary hits were performed. Needle-shaped crystals were obtained using a 1:1 mixture of the protein complex and mother liquor (0.1 M bis-

tris pH 5.5, 0.3 M magnesium formate, based on the H1 condition of JCSG+). Crystals were harvested and frozen with the mother liquor and 20% ethylene glycol as the cryoprotectant. Structural data will be collected from these crystals once the synchrotron reopens.

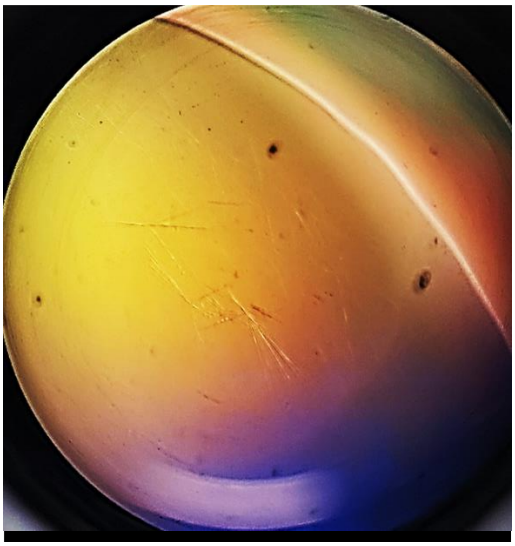


Figure 5-6. Crystals of GSK3β-hD1-FRATide. Crystalline needles are observed in the crystallization drop of GSK3β co-crystallized with hD1 and FRATide peptides at 200 nM. Crystallizing condition- 0.1 M Bis-Tris pH 5.5, 0.3 M magnesium formate

5.5 Discussion

Three strategies were used to generate the GSK3β-hD1 complex. The strategy described in this chapter was able to provide the most promising results. The crystallization screens often require optimization and availability of the protein complexes in higher quantities. We were able to express ~600 µg of GSK3β from a 12 L batch cell culture using our optimized overexpression and purification protocol. The current yield of GSK3β provides a sufficient amount to generate complexes to set up one to two 96-wide screen conditions only. In the future, we could explore using a fermenter to increase the size of the bacterial culture or using a eukaryotic expression system, such as yeast or insect cell lines that may provide a higher protein yield. Nonetheless, we are able to routinely obtain crystallizable GSK3β with our current expression and purification protocol.

Our first attempt to freeze co-crystals of GSK3β/hD1 with additional soaking of hD1 was unsuccessful. Crystals were soaked for 5 minutes in 5 µM peptide solution made with cryoprotectant harvest buffer. Diffraction spots observed for this batch were streaky. Streaky spots

arise when crystal packing is disrupted. Comparison of the first and second co-crystallization trials revealed that soaking the crystals with an excess of peptides may have caused distortion to the crystal lattices, thus affecting the integrity[99]. The detrimental effect is not surprising since hD1 is quite a large molecule for soaking experiments.

Although we are hopeful that the new crystals with different morphologies grown from new crystallization conditions contain the peptide(s) bound to GSK3β, we will not know until we determine the crystal structures. The possibility that the new crystallization conditions disrupt the complex remains and we may end up with GSK3β alone in these new crystal forms. Crosslinking the complex is a possible solution if we failed to find a crystallization condition where the peptide(s) remain bound to GSK3β. Chemical crosslinking molecules such as glutaraldehyde can be used to stabilize the protein complexes. Polymerization of glutaraldehyde and its interaction with lysine residues of the protein of interest helps preserve the structure of the enzymes. We could use chemical crosslinking reagents to covalently link hD1 to GSK3β. This method will ensure that we do not waste time pursuing conditions that favour the crystallization of GSK3β alone. To make sure the peptide and the kinase remain bound, it may also be beneficial to introduce a flexible glycine rich linker in the construct. Linkers often promote the proximity of the interaction partners without interefering with the protein-protein interaction framework.

Alternative to crystallography, electron microscopy may provide structural information of GSK3β/hD1. I have performed negative staining of the GSK3β/hD1 and GSK3β/hD1/FRATide complexes in their orginal buffers. We were able to identify some recurrent shapes of the protein particles. In addition, we can utilize this method to screen for buffers that disrupt and stabilize the complex to complement our crystallization effort.

Finally, hD1 is a small region of the 100 kDa full-length DISC1 protein. The structure of GSK3β/hD1 will partially reflect how the full-length proteins interact. Extending the length of the peptide to various lengths (possibly targeting for a nearby predicted α-helix sequence) in the co-expression complex could generate additional interacting regions that may stabilize the complex more. Ultimately, we may need to understand how the full-length proteins interact to understand the complete inhibition mechanism of DISC1 against GSK3β. Crystallization of such a large complex is a challenging task and only provides a static image of the complex. Biochemical and biophysical analyses will be able to provide other insights on the interaction.

6 Conclusion and future work

The primary objective of my thesis was to yield crystals of GSK3β, hD1, and FRATide. I used three distinct strategies to develop the crystals of the binary/ternary protein complex. The objective was achieved as we derived crystals for the complexes under different crystallization conditions. However, we need to solve the crystallographic data collected to determine the atomic structure of the complex. Future work in extension of this thesis will focus on solving the data.

Three strategies were attempted to obtain crystallizable samples. Strategy 1 resulted in a false positive interaction result between GSK3β and HisMBP-hD1, which will require a thorough investigation. This strategy was preliminary, and we never reached up to the crystallization step. This approach will not have high priority in the future due to the extensive experiments needed.

In the future, I would like to further explore strategies 2 and 3. Strategy 2 was based on the co-expression of the interacting protein partners in a single protein expression system. Expression systems with GSK3β alone were inefficient as it has been predicted that the expression of the kinase may not be favourable for the growth of bacterial cells. The cells cope with the negative effect by limiting the expression of the kinase. The expression of GSK3β using strategy 2, however, resulted in very high yield (~10 times) compared to HisGSK3β expression using strategy 1. The high yield may have been possible due to the expression of the kinase inhibiting peptide hD1 in the same system. This could be a firm indication of intracellular interaction. We observed the growth of many clusters in the crystallization drop. Optimization of the purification protocol for strategy 2 would be worth the resources to yield enough protein for multiple crystallization trials.

We derived crystals from strategy 3, but the data needs to be analyzed. Using the suggestions mentioned in the discussion section of chapter 5, we could fast-track the development of binary/ternary crystal structures.

Appendix A

Reagent list-

Acrylamide (Fisher # 01065-500)

Ampicillin (VWR # CA97061-442)

APS (Sigma #A3678-100G)

Bacto-Agar (Anachemia # 02116-300)

Bacto-Tryptone (Fisher # BP1421-2)

EDTA (Fisher # PR-V4231)

Glycerol (Fisher # BP2291)

HEPES (Fisher # BP3101)

Imidazole (Fisher # BP30550)

KH2PO4 (Fluka # 60220)

K2HPO4 (Fisher BP363-1)

Magnesium Chloride (VWR # W1917603)

Pierce Protease Inibitor (Thermo Fisher# 88666)

Precision Plus Protein Unstained Standard (Biorad # 161-0363)

Sodium Chloride (VWR # CA71008-564)

Sodium Hydroxide (Alfa Aesar # AAA1603736)

TCEP HCl (Fisher-Pierce # PI20491)

TEMED (Sigma # T9281-100ML)

Tris Base (Fisher # BP1521)

Tryptone (Fisher # BP14212)

Tween 20 (VWR # CA97062-332)

Yeast Extract (Fisher #BP14222)

References

1. Rehm, J. & Shield, K. D. Global Burden of Disease and the Impact of Mental and Addictive Disorders. *Curr. Psychiatry Rep.* **21**, (2019).

2. Ross, C. A., Margolis, R. L., Reading, S. A. J., Pletnikov, M. & Coyle, J. T. Neurobiology of Schizophrenia. *Neuron* **52**, 139–153 (2006).

3. McGuffin, P., Owen, M. & Farmer, A. Genetic basis of schizophrenia. *Lancet* **346**, 678–682 (1995).

4. Blackwood, D. H. R. *et al.* Schizophrenia and affective disorders - Cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: Clinical and P300 findings in a family. *Am. J. Hum. Genet.* **69**, 428–433 (2001).

5. Millar, J. K. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum. Mol. Genet.* **9**, 1415–1423 (2000).

6. Korth, C. DISCopathies: Brain Disorders Related to DISC1 Dysfunction. *Rev. Neurosci.* **20**, (2009).

7. Soares, D. C., Carlyle, B. C., Bradshaw, N. J. & Porteous, D. J. DISC1: Structure, function, and therapeutic potential for major mental illness. *ACS Chem. Neurosci.* **2**, 609–632 (2011).

8. Brandon, N. J. *et al.* Understanding the role of DISC1 in psychiatric disease and during normal development. *J. Neurosci.* **29**, 12768–12775 (2009).

9. Millar, J. K., Christie, S. & Porteous, D. J. Yeast two-hybrid screens implicate DISC1 in brain development and function. *Biochem. Biophys. Res. Commun.* **311**, 1019–1025 (2003).

10. Camargo, L. M. *et al.* Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol.*

Psychiatry **12**, 74–86 (2007).

11. Chubb, J. E., Bradshaw, N. J., Soares, D. C., Porteous, D. J. & Millar, J. K. The DISC locus in psychiatric illness. *Mol. Psychiatry* **13**, 36–64 (2008).

12. Taylor, M. S., Devon, R. S., Millar, J. K. & Porteous, D. J. Evolutionary constraints on the Disrupted in Schizophrenia locus. *Genomics* **81**, 67–77 (2003).

13. Ma, L. *et al.* Cloning and characterization of Disc1, the mouse ortholog of DISC1 (Disrupted-in-Schizophrenia 1). *Genomics* **80**, 662–672 (2002).

14. Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet.* **42**, 38–48 (2001).

15. Fink, A. L. Natively unfolded proteins. *Curr. Opin. Struct. Biol.* **15**, 35–41 (2005).

16. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **337**, 635–645 (2004).

17. Sanchez-Pulido, L. & Ponting, C. P. Structure and evolutionary history of DISC1. *Hum. Mol. Genet.* **20**, 175–181 (2011).

18. Leliveld, S. R. *et al.* Oligomer assembly of the C-terminal DISC1 domain (640-854) is controlled by self-association motifs and disease-associated polymorphism S704C. *Biochemistry* **48**, 7746–7755 (2009).

19. Mao, Y. *et al.* Disrupted in Schizophrenia 1 Regulates Neuronal Progenitor Proliferation via Modulation of GSK3β/β-Catenin Signaling. *Cell* **136**, 1017–1031 (2009).

20. Leliveld, S. R. *et al.* Insolubility of disrupted-in-schizophrenia 1 disrupts oligomer-dependent interactions with nuclear distribution element 1 and is associated with sporadic mental disease. *J. Neurosci.* **28**, 3839–3845 (2008).

21. Crick, F. H. C. The packing of α-helices: simple coiled-coils. *Acta Crystallogr.* **6**, 689–697 (1953).

22. Brandon, N. J. *et al.* Disrupted in Schizophrenia 1 and Nudel form a neurodevelopmentally regulated protein complex: Implications for schizophrenia and

other major neurological disorders. *Mol. Cell. Neurosci.* **25**, 42–55 (2004).

23. Kamiya, A. *et al.* A schizophrenia-associated mutation of DISC1 perturbs cerebral cortex development. *Nat. Cell Biol.* **7**, 1067–1078 (2005).

24. Sohi, M. *et al.* Crystal structure of Escherichia coli UvrB C-terminal domain, and a model for UvrB-UvrC interaction. *FEBS Lett.* **465**, 161–164 (2000).

25. Ishizuka, K. *et al.* DISC1-dependent switch from progenitor proliferation to migration in the developing cortex. *Nature* **473**, 92–96 (2011).

26. Aberle, H., Bauer, A., Stappert, J., Kispert, A. & Kemler, R. B-Catenin Is a Target for the Ubiquitin-Proteasome Pathway. *EMBO J.* **16**, 3797–3804 (1997).

27. Orford, K., Crockett, C., Jensen, J. P., Weissman, A. M. & Byers, S. W. regulated Ubiquitination and Degradation of -Catenin *. *October* 24735–24738 (1997).

28. Larner, J. *et al.* Hormonal and non-hormonal control of glycogen synthesis — Control of transferase phosphatase and transferase I kinase. *Adv. Enzyme Regul.* **6**, 409–423 (1968).

29. EMBI, N., RYLATT, D. B. & COHEN, P. Glycogen Synthase Kinase-3 from Rabbit Skeletal Muscle. *Eur. J. Biochem.* **107**, 519–527 (2005).

30. Sutherland, C. What are the bona fide GSK3 substrates? *Int. J. Alzheimers. Dis.* **2011**, (2011).

31. Linding, R. *et al.* Systematic Discovery of In Vivo Phosphorylation Networks. *Cell* **129**, 1415–1426 (2007).

32. Woodgett, J. R. A common denominator linking glycogen metabolism, nuclear oncogenes and development. *Trends Biochem. Sci.* **16**, 177–181 (1991).

33. Doble, B. W. & Woodgett, J. R. GSK-3: Tricks of the trade for a multi-tasking kinase. *J. Cell Sci.* **116**, 1175–1186 (2003).

34. Krishnankutty, A. *et al.* In vivo regulation of glycogen synthase kinase 3β activity in neurons and brains. *Sci. Rep.* **7**, 1–15 (2017).

35. Fiol, C. J. *et al.* Phosphoserine as a recognition determinant for glycogen synthase kinase-

3: Phosphorylation of a synthetic peptide based on the G-component of protein phosphatase-1. *Arch. Biochem. Biophys.* **267**, 797–802 (1988).

36. Fiol CJ, Wang, A. & Roach, P. J. Ordered Multisite Protein Phosphorylation. *J. Biol. Chem.* **265**, 6061–6065 (1990).

37. Wang, Y. R. & Roach, P. J. Inactivation of rabbit muscle glycogen synthase by glycogen synthase kinase-3. Dominant role of the phosphorylation of Ser-640. *J. Biol. Chem.* **268**, 23876–80 (1993).

38. KURET, J., WOODGETT, J. R. & COHEN, P. Multisite phosphorylation of glycogen synthase from rabbit skeletal muscle: Identification of the sites phosphorylated by casein kinase-I. *Eur. J. Biochem.* **151**, 39–48 (1985).

39. Woodgett, J. R. & Cohen, P. Multisite phosphorylation of glycogen synthase. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **788**, 339–347 (1984).

40. Liu, C. *et al.* Control of β-Catenin Phosphorylation/Degradation by a Dual-Kinase Mechanism. *Cell* **108**, 837–847 (2002).

41. Thomas, G. M. *et al.* A GSK3-binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of Axin and β-catenin. *FEBS Lett.* **458**, 247–251 (1999).

42. Robbins, D. J. *et al.* Regulation and properties of extracellular signal-regulated protein kinases 1 and 2 *in vitro*. *J. Biol. Chem.* **268**, 5097–106 (1993).

43. Hughes, K., Nikolakaki, E., Plyte, S. E., Totty, N. F. & Woodgett, J. R. Modulation of the glycogen synthase kinase-3 family by tyrosine phosphorylation. *EMBO J.* **12**, 803–8 (1993).

44. Dajani, R. *et al.* Crystal structure of glycogen synthase kinase 3β: Structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell* **105**, 721–732 (2001).

45. Bax, B. *et al.* The structure of phosphorylated GSK-3β complexed with a peptide, FRATtide, that inhibits β-catenin phosphorylation. *Structure* **9**, 1143–1152 (2001).

46. Dajani, R. *et al.* Structural basis for recruitment of glycogen synthase kinase 3β to the axin-APC scaffold complex. *EMBO J.* **22**, 494–501 (2003).

47. Cross, D. A. E., Alessi, D. R., Cohen, P., Andjelkovich, M. & Hemmings, B. A. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* **378**, 785–789 (1995).

48. Eldar-Finkelman, H., Seger, R., Vandenheede, J. R. & Krebs, E. G. Inactivation of Glycogen Synthase Kinase-3 by Epidermal Growth Factor Is Mediated by Mitogen-activated Protein Kinase/p90 Ribosomal Protein S6 Kinase Signaling Pathway in NIH/3T3 Cells. *J. Biol. Chem.* **270**, 987–990 (1995).

49. Frame, S., Cohen, P. & Biondi, R. M. A common phosphate binding site explains the unique substrate specificity of GSK3 and its inactivation by phosphorylation. *Mol. Cell* **7**, 1321–1327 (2001).

50. Patel, P. & Woodgett, J. R. Glycogen Synthase Kinase 3: A Kinase for All Pathways? *Curr. Top. Dev. Biol.* **123**, 277–302 (2017).

51. Shin, S. H. *et al.* The nuclear localization of glycogen synthase kinase 3β is required its putative PY-nuclear localization sequences. *Mol. Cells* **34**, 375–382 (2012).

52. Dierick, H. & Bejsovec, A. Cellular Mechanisms of Wingless/Wnt Signal Transduction. in *Current Topics in Developmental Biology* **43**, 153–190 (1998).

53. Miller, J. R. Protein family review The Wnts. *Gene* 1–15 (2001). doi:10.1186/gb-2001-3-1-reviews3001

54. Barker, N., Huls, G., Korinek, V. & Clevers, H. Restricted High Level Expression of Tcf-4 Protein in Intestinal and Mammary Gland Epithelium. *Am. J. Pathol.* **154**, 29–35 (1999).

55. Hart, M. J., De Los Santos, R., Albert, I. N., Rubinfeld, B. & Polakis, P. Downregulation of β-catenin by human Axin and its association with the APC tumor suppressor, β-catenin and GSK3β. *Curr. Biol.* **8**, 573–581 (1998).

56. Kishida, S. *et al.* Axin, a negative regulator of the Wnt signaling pathway, directly interacts with adenomatous polyposis coli and regulates the stabilization of β-catenin. *J. Biol. Chem.* **273**, 10823–10826 (1998).

57. Amit, S. Axin-mediated CKI phosphorylation of beta -catenin at Ser 45: a molecular

switch for the Wnt pathway. *Genes Dev.* **16**, 1066–1076 (2002).

58. Yost, C. *et al.* The axis-inducing activity, stability, and subcellular distribution of β-catenin is regulated in Xenopus embryos by glycogen synthase kinase 3. *Genes Dev.* **10**, 1443–1454 (1996).

59. Liu, C. *et al.* β-Trcp couples β-catenin phosphorylation-degradation and regulates Xenopus axis formation. *Proc. Natl. Acad. Sci.* **96**, 6273–6278 (1999).

60. Latres, E., Chiaur, D. S. & Pagano, M. The human F box protein β-Trcp associates with the Cul1/Skp1 complex and regulates the stability of β-catenin. *Oncogene* **18**, 849–854 (1999).

61. Wodarz, A. & Nusse, R. Mechanisms of Wnt Signaling in Development. *Annu. Rev. Cell Dev. Biol.* **14**, 59–88 (1998).

62. Sawa, H., Lobel, L. & Horvitz, H. R. The Caenorhabditis elegans gene lin-17, which is required for certain asymmetric cell divisions, encodes a putative seven-transmembrane protein similar to the Drosophila Frizzled protein. *Genes Dev.* **10**, 2189–2197 (1996).

63. He, X. *et al.* A member of the frizzled protein family mediating axis induction by Wnt-5A. *Science (80-.).* **275**, 1652–1654 (1997).

64. Pinson, K. I., Brennan, J., Monkley, S., Avery, B. J. & Skarnes, W. C. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**, 535–538 (2000).

65. Tamai, K. *et al.* LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530–535 (2000).

66. Wehrli, M. *et al.* Arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407**, 527–530 (2000).

67. Perrimon, N. & Mahowald, A. P. Multiple functions of segment polarity genes in Drosophila. *Dev. Biol.* **119**, 587–600 (1987).

68. Wong, H.-C. *et al.* Direct Binding of the PDZ Domain of Dishevelled to a Conserved Internal Sequence in the C-Terminal Region of Frizzled. *Mol. Cell* **12**, 1251–1260 (2003).

69. Umbhauer, M. The C-terminal cytoplasmic Lys-Thr-X-X-X-Trp motif in frizzled

receptors mediates Wnt/beta-catenin signalling. *EMBO J.* **19**, 4944–4954 (2000).

70. Schwarz-Romond, T. *et al.* The DIX domain of Dishevelled confers Wnt signaling by dynamic polymerization. *Nat. Struct. Mol. Biol.* **14**, 484–492 (2007).

71. Schwarz-Romond, T., Merrifield, C., Nichols, B. J. & Bienz, M. The Wnt signalling effector Dishevelled forms dynamic protein assemblies rather than stable associations with cytoplasmic vesicles. *J. Cell Sci.* **118**, 5269–5277 (2005).

72. Schwarz-Romond, T., Metcalfe, C. & Bienz, M. Dynamic recruitment of axin by Dishevelled protein assemblies. *J. Cell Sci.* **120**, 2402–2412 (2007).

73. Tamai, K. *et al.* A Mechanism for Wnt Coreceptor Activation. *Mol. Cell* **13**, 149–156 (2004).

74. Mao, J. *et al.* Low-density lipoprotein receptor-related protein-5 binds to Axin and regulates the canonical Wnt signaling pathway. *Mol. Cell* **7**, 801–809 (2001).

75. Davidson, G. *et al.* Casein kinase 1γ couples Wnt receptor activation to cytoplasmic signal transduction. *Nature* **438**, 867–872 (2005).

76. Zeng, X. *et al.* A dual-kinase mechanism for Wnt co-receptor phosphorylation and activation. *Nature* **438**, 873–877 (2005).

77. Tortelote, G. G., Reis, R. R., de Almeida Mendes, F. & Abreu, J. G. Complexity of the Wnt/β-catenin pathway: Searching for an activation model. *Cell. Signal.* **40**, 30–43 (2017).

78. Li, V. S. W. *et al.* Wnt Signaling through Inhibition of β-Catenin Degradation in an Intact Axin1 Complex. *Cell* **149**, 1245–1256 (2012).

79. Willert, K., Shibamoto, S. & Nusse, R. Wnt-induced dephosphorylation of Axin releases β-catenin from the Axin complex. *Genes Dev.* **13**, 1768–1773 (1999).

80. Hernandez, A. R., Klein, A. M. & Kirschner, M. W. Kinetic Responses of -Catenin Specify the Sites of Wnt Control. *Science (80-.).* **338**, 1337–1340 (2012).

81. Kim, S. E. *et al.* Wnt stabilization of β-catenin reveals principles for morphogen receptor-scaffold assemblies. *Science (80-.).* **340**, 867–870 (2013).

82. Luo, W. *et al.* Protein phosphatase 1 regulates assembly and function of the β-catenin degradation complex. *EMBO J.* **26**, 1511–1521 (2007).

83. Jonkers, J., Korswagen, H. C., Acton, D., Breuer, M. & Berns, A. Activation of a novel proto-oncogene, Frat1, contributes to progression of mouse T-cell lymphomas. *EMBO J.* **16**, 441–450 (1997).

84. Yost, C. *et al.* GBP, an inhibitor of GSK-3, is implicated in Xenopus development and oncogenesis. *Cell* **93**, 1031–1041 (1998).

85. Li, L. *et al.* Axin and Frat1 interact with Dvl and GSK, bridging Dvl to GSK in Wnt-mediated regulation of LEF-1. *EMBO J.* **18**, 4233–4240 (1999).

86. Fraser, E. *et al.* Identification of the Axin and Frat binding region of glycogen synthase kinase-3. *J. Biol. Chem.* **277**, 2176–2185 (2002).

87. Hanks, S. K., Quinn, A. M. & Hunter, T. The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science (80-.).* **241**, 42–52 (1988).

88. Ter Haar, E. *et al.* Structure of GSK3β reveals a primed phosphorylation mechanism. *Nat. Struct. Biol.* **8**, 593–596 (2001).

89. Johnson, L. N., Noble, M. E. M. & Owen, D. J. Active and inactive protein kinases: Structural basis for regulation. *Cell* **85**, 149–158 (1996).

90. Saundh, S. *et al.* Identification and Mechanistic Characterization of a Peptide Inhibitor of Glycogen Synthase Kinase (GSK3β) Derived from the Disrupted in Schizophrenia 1 (DISC1) Protein. **1**, (2020).

91. Aslanidis, C. & de Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18**, 6069–6074 (1990).

92. Tan, S. A modular polycistronic expression system for overexpressing protein complexes in Escherichia coli. *Protein Expr. Purif.* **21**, 224–234 (2001).

93. Tan, S., Kern, R. C. & Selleck, W. The pST44 polycistronic expression system for producing protein complexes in Escherichia coli. *Protein Expr. Purif.* **40**, 385–395 (2005).

94. Kramer, R. M., Shende, V. R., Motl, N., Pace, C. N. & Scholtz, J. M. Toward a molecular

understanding of protein solubility: Increased negative surface charge correlates with increased solubility. *Biophys. J.* **102**, 1907–1915 (2012).

95. Nishihara, K., Kanemori, M., Yanagi, H. & Yura, T. Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. *Appl. Environ. Microbiol.* **66**, 884–889 (2000).

96. Walkenhorst, W. F., Merzlyakov, M., Hristova, K. & Wimley, W. C. Polar residues in transmembrane helices can decrease electrophoretic mobility in polyacrylamide gels without causing helix dimerization. *Biochim. Biophys. Acta - Biomembr.* **1788**, 1321–1331 (2009).

97. Bornhorst, J. A. & Falke, J. J. [16] Purification of proteins using polyhistidine affinity tags. in *Journal of Investigative Dermatology* **135**, 245–254 (2000).

98. Aoki, M. *et al.* Structural insight into nucleotide recognition in tau-protein kinase I/glycogen synthase kinase 3β. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 439–446 (2004).

99. Collins, P. M. *et al.* Gentle, fast and effective crystal soaking by acoustic dispensing. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 246–255 (2017).