IN SILICO AND *IN VITRO* CHARACTERISATION OF PEPTIDE BINDING SEQUENCE OF TRIM25 CC DOMAIN TOWARDS 14-3-3 SIGMA PROTEIN

CHIANG DE CHEN

UNIVERSITI SAINS MALAYSIA

2023

IN SILICO AND *IN VITRO* CHARACTERISATION OF PEPTIDE BINDING SEQUENCE OF TRIM25 CC DOMAIN TOWARDS 14-3-3 SIGMA PROTEIN

by

CHIANG DE CHEN

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

September 2023

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to the following people who have supported me throughout my research journey. First and foremost, I would like to thank my supervisor, Dr Yap Beow Keat, for his guidance, encouragement and constructive feedback. He has been a great mentor and a source of inspiration for me. I would also like to thank my co-supervisor, Dr Teh Aik Hong, for his valuable insights and suggestions. He has helped me to refine my research questions and methodology, and the resource support. I would like to acknowledge the scholarship support GRA-ASSIST 2021 from Universiti Sains Malaysia, which enabled me to pursue my interest and education without worrying on my finance. I would also like to thank my colleagues and friends at School of Pharmaceutical Sciences, especially Ghazi Ahmad Aljabal and Choong Fei Her, for their help and support. They have made my research experience enjoyable and memorable. Last but not least, I would like to thank my family, especially my parents, for their unconditional love and support. They have always believed in me and encouraged me to pursue my dreams.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

α	alpha
β	beta
γ	gamma
δ	delta
ζ	zeta
η	eta
ε	epsilon
θ	theta
τ	tau
μ	micro
°C	degree Celsius
K_D	equilibrium dissociation constant

LIST OF ABBREVIATIONS

3D	Three-dimensional
ACN	Acetonitrile
AI	Artificial Intelligence
APS	Ammonium persulfate
BME	beta-mercaptoethanol
BPB	Bromophenol blue
CC	Coiled-coil
CPMG	Carr-Purcell-Meiboom-Gill
DEE	Diethyl ether
DIPEA	N,N-Diisopropylethylamine
DMB	1,3-dimethoxybenzene
DMSO	Dimethylsulfoxide
FT	Flow-through
GST	Glutathione S-transferase
HCTU	O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
IPTG	isopropyl ß-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LC	Liquid chromatography
MBP	Maltose binding protein
MD	Molecular dynamics
MeOH	Methanol
MM/PBSA	Molecular Mechanics Poisson-Boltzmann Surface Area

MS	Mass spectrometry
MW	Molecular weight
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PPI	Protein-protein interaction
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
RP-HPLC	Reverse-phase high performance liquid chromatography
SDS	Sodium dodecyl sulphate
Semi-prep LC	Semi-preparative liquid chromatography
SPPS	Solid phase peptide synthesis
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic acid
TIPS	Triisopropyl silane
TRIM	Tripartite motif

PENCIRIAN IN SILICO DAN IN VITRO IKATAN TURUTAN PEPTIDA DOMAIN CC TRIM25 TERHADAP PROTEIN 14-3-3 SIGMA

ABSTRAK

Protein 14-3-3σ ialah isoform yang paling bercapah antara tujuh keluarga protein 14-3-3. Tidak seperti isoform 14-3-3 yang lain, penurunan ekspresinya ditunjukkan dalam pelbagai perkembangan kanser dan tumor, yang selaras dengan fungsinya sebagai protein penyesuai yang mengawal kitaran dan perkembangan sel. TRIM25 ialah ligase ubiquitin E3 yang fungsi utamanya bertanggungjawab terhadap degradasi protein yang dimediasi ubiquitin. Menariknya, interaksi 14-3-3σ-TRIM25 membawa kepada ubiquitinasi 14-3-3σ dan degradasi proteasomal seterusnya. Walau bagaimanapun, urutan peptida tepat yang terlibat dalam interaksi dua protein ini masih belum dikenal pasti. Kajian ini bertujuan untuk mengenal pasti ikatan urutan peptida protein TRIM25 terhadap protein 14-3-3σ. Sebanyak 5 peptida yang berasal daripada TRIM25 telah dikenal pasti daripada analisis bioinformatik *in-silico* (carian motif dan penjajaran jujukan berbilang) untuk ujian pengikatan seterusnya. Peptida sama ada dibeli secara komersil, atau disintesis secara dalaman menggunakan kaedah Fmoc SPPS, diikuti dengan penulenan menggunakan LC separa persediaan dan disahkan dengan LC-MS. Protein rekombinan 14-3-3σ berjaya diekspresikan menggunakan sistem ekspresi *E. coli* dan ditulenkan dengan kromatografi afiniti. Interaksi antara 6 peptida yang dikenal pasti sebelum ini dan protein 14-3-3σ yang dinyatakan telah diuji dengan ujian ^{1}H CPMG NMR, yang mana Peptida 1 dengan urutan ^{402}KLPp(T)FG^{407} menunjukkan pertalian pengikatan tertinggi kepada protein 14-3-3σ. Ujian CPMG NMR kompetitif bagi Peptida 1 dengan Peptida 6 (dikenali sebagai pengikat 14-3-3σ)

mendedahkan persaingan yang besar, mencadangkan bahawa Peptide 1 berkongsi tapak pengikatan yang serupa dengan Peptida 6 pada poket amfipati. Perkaitan mengikat, K_D Peptida 1 dianggarkan pada 116.4 μM dengan ujian ITC. Simulasi dinamik molekul menunjukkan pembentukan ikatan hidrogen dan titi garam yang konsisten antara pThr405 Peptida 1 dengan Lys49, Arg56, Arg129 dan Tyr130 daripada 14-3-3σ, selaras dengan mod pengikatan yang ditunjukkan oleh pasangan protein berfosforilasi bagi keluarga protein 14-3-3 yang lain. Selain itu, Phe406 daripada Peptida 1 membentuk ikatan hidrogen dengan Lys122 dan Asn175 daripada 14-3-3σ secara konsisten, yang tidak pernah dilaporkan sebelum ini. Pro404 Peptide 1 juga berinteraksi dengan Val178 daripada 14-3-3σ melalui interaksi van der Waals, yang selaras dengan interaksi 14-3-3η-TRIM25 yang dilaporkan sebelum ini. Di samping itu, kajian ini juga mencadangkan bahawa Ile219 daripada 14-3-3σ juga mungkin penting untuk mengikat TRIM25, mungkin melalui interaksi bukan polar dengan Phe406. Penemuan kajian ini berguna sebagai titik permulaan untuk menyiasat potensi perencat interaksi TRIM25-14-3-3σ dalam menangani penurunan regulasi 14-3-3σ dan akhirnya menyekat pertumbuhan tumor dalam pelbagai jenis kanser.

IN SILICO AND *IN VITRO* CHARACTERISATION OF PEPTIDE BINDING SEQUENCE OF TRIM25 CC DOMAIN TOWARDS 14-3-3 SIGMA PROTEIN

ABSTRACT

14-3-3σ protein is the most diverged isoform among the seven isoforms of the 14-3-3 protein family. Unlike other 14-3-3 isoforms, its downregulation was demonstrated in various cancer and tumour development, which is in line with its function as an adaptor protein controlling cell cycle and development. TRIM25 is an E3 ubiquitin ligase that largely responsible for ubiquitin-mediated protein degradation. Interestingly, 14-3-3σ-TRIM25 interaction leads to 14-3-3σ ubiquitination and subsequent proteasomal degradation. Their exact interacting peptide sequence, however, has yet to be identified. The present study aimed to identify the peptide binding sequence of TRIM25 protein towards 14-3-3σ protein. A total of 5 peptides were identified from *in silico* bioinformatic analysis (motif search and multiple sequence alignment) for subsequent binding assay. The peptides were either purchased commercially, or synthesised in-house using Fmoc SPPS method, followed by purification using semi-preparative LC and mass-validated with LC-MS. 14-3-3σ protein was successfully expressed using *E. coli* expression system and purified with affinity chromatography. Interactions between the 6 peptides identified previously and the 14-3-3σ protein expressed were tested with ^{1}H CPMG NMR assay, of which Peptide 1 with sequence ^{402}KLPp(T)FG^{407} demonstrates the highest binding affinity to 14-3-3σ protein. Competitive CPMG NMR assay of Peptide 1 with Peptide 6 (a known 14-3-3σ binder) revealed substantial binding competition, suggesting that Peptide 1 shares similar binding site with Peptide 6 at the amphipathic pocket. The binding

affinity, K_D of Peptide 1 was estimated at 116.4 μM with ITC assay. Molecular dynamics simulation demonstrated consistent formation of hydrogen bonds and salt bridges between pThr405 of Peptide 1 with Lys49, Arg56, Arg129 and Tyr130 of 14-3-3σ, which resembles the binding mode of other phosphorylated protein partner of 14-3-3 family protein. Additionally, Phe406 of Peptide 1 formed hydrogen bonds with Lys122 and Asn175 of 14-3-3σ consistently, interactions which have not been reported previously. Pro404 of Peptide 1 also interacted with Val178 of 14-3-3σ via van der Waals interaction, which is in agreement with the previous work on 14-3-3η-TRIM25 interaction. In addition, this study also proposed that Ile219 of 14-3-3σ may also be important for binding to TRIM25, presumably through its non-polar interaction with the Phe406. The finding of this study is useful as a starting point to probe the potential of designing an inhibitor of TRIM25-14-3-3σ in addressing the downregulation of 14-3-3σ and ultimately suppressing tumour growth in various cancers.

CHAPTER 1

INTRODUCTION

1.1 Background of the study

14-3-3σ protein is one of the seven isoforms from the highly conserved eukaryotic 14-3-3 protein family beside α/β, γ, δ/ζ, η, ε, and θ/τ isoforms. In most cases, they function as adaptor proteins mediating protein-protein interaction for cell cycle regulation, apoptosis, protein trafficking and signal transduction [1-3]. Intriguingly, only 14-3-3σ is being downregulated in various cancers in contrast to other 14-3-3 isoforms [2]. While all other 14-3-3 isoforms can interact among themselves to form heterodimers, 14-3-3σ protein only forms homodimers [1]. Sequence and structural analysis suggest that 14-3-3σ protein is the most diverged isoform among the 7 isoforms, which further highlights its distinctive properties among 14-3-3 family protein [1, 2].

Downregulation of 14-3-3σ expression in various tumours is causally linked to hypermethylation at the CpG-rich region of the 14-3-3σ promoter, and ubiquitin(Ub)-mediated proteasomal degradation [4]. Though the loss of 14-3-3σ expression is frequently linked to epigenetic silencing of CpG methylation, inconsistency in 14-3-3σ expression level with its promoter methylation status in some cancers suggests the role of post-transcriptional regulation through proteasome-mediated degradation [2, 5, 6].

Tripartite motif (TRIM) protein family consists of more than 80 identified genes in humans, of which most of them are defined as E3 Ub ligases due to their ubiquitination ability [7]. Ub-mediated proteasomal degradation effectively degrades proteins through a cascade process, in which Ub is conjugated to protein followed by

proteasome-mediated degradation [8]. Similar to 14-3-3 protein, TRIM proteins are also involved in modulating cell cycle, apoptosis, cellular development and carcinogenesis [7, 9-11]. Particularly, interferon-inducible Tripartite motif-containing 25 (TRIM25), also known as estrogen-responsive finger protein (Efp), is found to be responsible for proteolytic degradation of 14-3-3σ in breast cancer and endometrial cancer, in which abrogation of TRIM25 suppressed tumour growth through 14-3-3σ upregulation [12-14]. Chimeric small interfering RNA (siRNA) targeting TRIM25 effectively inhibits breast cancer cell proliferation [15]. Similarly, upregulation of TRIM25 correlated with prognosis and proliferation of prostate, gastric, lung, liver, colon and many other cancers [16-20]. On top of that, TRIM25 expression also correlates with cisplatin resistance via the MDM2-p53 pathway, suggesting the essential role of TRIM25 in addressing the chemoresistance of cancers [21].

The mechanisms of proteolytic inactivation of 14-3-3σ by TRIM25 were found to be through ubiquitination and ISGylation, which essentially regulates 14-3-3σ at the post-transcriptional level [4, 14]. The transfer of Ub to 14-3-3σ subsequently promotes proteasomal degradation of 14-3-3σ, while ISG15 transfer leads to structural modification of unknown function to date, presumably through stabilising or interfering its functionality as demonstrated in other ISGylated protein in cancers [22]. Together with their role in cell growth regulation and cancers, direct interaction among TRIM25 and its target 14-3-3σ protein is undoubtedly an interesting subject to be studied.

1.2 Problem statement

Structural analysis revealed a conserved protein architecture of TRIM25 characterised by a RING catalytic domain, one or two B boxes (B1 and B2), a coiled-coil region (CC) and a B30.2/SPRY domain [23, 24]. The N-terminal RING domain is essential for E2 ligase binding, which gives rise to TRIM25's ubiquitinating ability, while the C-terminal SPRY domain facilitates various substrate interactions in immune signalling, including pathogen recognition and protein-protein interactions [12, 25-27]. However, there are very few studies illustrating the role of the B boxes or CC domain of TRIM25, except that the CC domain may be involved in dimerisation and higher-order assembly of TRIM25 [24]. Surprisingly, a TRIM25 truncate with only B boxes and the CC domain was found to be crucial for interaction and subsequent downregulation of 14-3-3σ [12]. Another independent study further supported the finding, in which the direct binding of TRIM25 to 14-3-3η is solely dependent on the 270-residues-containing CC domain truncate [28]. As existing studies have yet to pinpoint the key interacting region between TRIM25 and 14-3-3σ protein, this study aimed to identify of the key residues among the CC domain responsible for TRIM25-14-3-3σ binding. This study may serve as a starting point in addressing the downregulation of 14-3-3σ protein in various cancer and thus, designing a new therapeutic agent to regulate TRIM25-14-3-3σ interaction.

1.3 Research aim and objectives

To identify the peptide binding sequence of TRIM25 CC domain towards 14-3-3σ protein.

1.3.1 Specific objectives

(a) To predict and synthesise potential 14-3-3σ-interacting peptide fragments of TRIM25's CC domain.

(b) To recombinantly express the 14-3-3σ protein.

(c) To investigate the interaction and binding conformation of 14-3-3σ-interacting peptide fragment of TRIM25's CC domain to 14-3-3σ protein.

CHAPTER 2

LITERATURE REVIEW

2.1 Protein-protein interactions (PPIs)

Proteins govern most of the cellular functions in a living system: they catalyse chemical reactions, regulate gene expression, transport tiny molecules, and send signals across membranes. Even viral infections are frequently triggered by virus-host protein interactions. In simple words, protein-protein interactions (PPIs) are processes of two or more proteins binding together to perform a biological function [29]. Multiprotein complexes control a number of crucial physiological processes, such as transmembrane signal transmission, transcriptional activation, translation, and cell development [30]. The PPIs that exist between the different complex subunits often regulate the function, activity, and specificity of these complexes.

As PPIs are involved in many biological processes and pathways that are essential for normal cellular function and homeostasis, they are often dysregulated or altered in various diseases, such as cancer, infectious diseases, neurodegenerative diseases, autoimmune diseases and metabolic disorders [31-33]. Therefore, PPIs can also serve as biomarkers or indicators of disease status, progression, prognosis or response to treatment. Besides, PPIs could also influence the pharmacokinetics, pharmacodynamics and toxicity of drugs by affecting their absorption, distribution, metabolism or excretion [31-33]. For instance, some drugs can interact with transporters, receptors or enzymes that are involved in drug metabolism or clearance.

With the significant role of PPIs in cellular activities, the study on PPI networks is useful to analyse complicated pathways and uncover the roles of undiscovered proteins. These research on protein networks has shed light on molecular evolution, how cells react to perturbations, and how proteins perform their tasks. Particularly in

the case of complicated diseases like cancer, an understanding of how proteins interact and communicate with one another has allowed the molecular basis of a variety of different diseases to be clarified. As a result, the number of possible therapeutic targets is increasing rapidly as our understanding of cell signalling mechanisms continues to develop.

However, current understanding on the function and binding mechanism of PPIs is limited due to the diversity and complexity of the highly dynamic PPIs network [30]. As protein-protein binding often being dynamic and possibly involves multi targets or multiple very large interaction surfaces, therapeutic approach targeting PPIs have long been regarded as high-risk drug targets in the past decade [34]. Furthermore, PPIs are challenging targets for drug discovery because they often have large and flat binding surfaces that are difficult for small molecules to bind to [35]. Some PPI interfaces may also have cryptic binding sites that only emerge upon ligand binding [36]. Another challenge is to achieve high selectivity and specificity for the target PPI, as many proteins share similar interacting regions or motifs that may result in off-target effects [37].

With the advancement in sensitivity, accuracy and precision of molecular technology, an increasing amount of research has been conducted in various academies and industries on the discovery and modulation of PPIs. Advances in technology have helped to identify PPIs in drug discovery in various ways, for example, mapping PPI network and the interactome using high-throughput methods such as yeast two-hybrid, affinity purification-mass spectrometry, protein microarrays and bioinformatics tools [31]. Techniques such as X-ray crystallography, nuclear magnetic resonance, surface plasmon resonance and isothermal titration calorimetry allow further investigation on the structure and dynamics of PPI interfaces [33]. On drug discovery, there are diverse

approaches to aid in discovering and designing small molecules, peptides and antibodies that can modulate PPIs, including structure-based drug design, ligand-based drug design, fragment-based drug discovery, phage display and computational screening [38]. These technologies have enabled the identification of novel PPI targets, the characterisation of their binding modes and mechanisms, and the development of potent and selective PPI modulators for various diseases.

Scientists have been utilising PPIs to discover the biological underpinnings of some illnesses as well as possible therapeutic targets. Some examples of successful PPI inhibitors include milademetan (MDM2/p53), venetoclax (BCL-2/BAX), navitoclax (BCL-2/BCL-XL), ibrutinib (BTK/SH2), idelalisib (PI3Kδ/SH2), lenalidomide (CRBN/IKZF1) and palbociclib (CDK4/6/cyclin D) [32, 34, 38-41].

2.2 Techniques for detection and analysis of PPIs

2.2.1 *In silico* prediction of PPIs

In silico methods are computational methods used in drug discovery that are critical in the cost-effective identification of promising drug candidates. These methods are relevant in limiting the use of animal models [42]. Especially with the introduction of Next-Generation Sequencing (NGS) technologies in the past decade, there has been tremendous progress in terms of speed, read length, and throughput [43]. These advances have democratised NGS and paved the way for new applications in genomics and molecular biology [44]. For example, NGS has enabled researchers to sequence entire genomes at a fraction of the cost and time required by previous technologies. GenBank database recorded sequences from more than 260000 organisms, which projected to double in size every 18 months [45]. Submission of quality protein sequences to UniProt was also observed to increase [46]. By leveraging

these large databases of biological information, researchers can identify the underpinning mechanism of PPI and possibly new drug targets.

Proteins bind to each other through a combination of hydrophobic bonding, van der Waals forces, and salt bridges at specific binding domains on each protein. These domains can be small binding clefts or large surfaces and can be just a few peptides long or span hundreds of amino acids. Typically, protein-protein interactions are mediated by short conserved sequences of amino acid, which is commonly known as protein binding motif [47]. The presence of a conserved motif between proteins is also often associated with a distinct structural site performing a particular function, such as active sites of an enzyme. A typical protein motif with associated function, such as Zinc Finger motif, is 10 to 20 amino acids long [48].

Binding motif search is a computational method used to identify short conserved amino acid sequences that are involved in PPIs. Researchers can use binding motif search to identify the specific amino acid sequences that are involved in PPIs between two proteins. By identifying these binding motifs, researchers can better understand how proteins interact with each other and how these interactions can be disrupted or enhanced. This information will be useful in identifying the underlying mechanism of diseases and to develop new drugs that target the specific PPIs. Some examples of motif search tool include FIMO and ScanProSite [49, 50]. FIMO employs position-specific scoring algorithms and keep track of false discovery rate, while ScanProSite matches protein sequences with pre-calculated match database based on motif libraries that are manually curated, aligned, and annotated, along with motif-specific information.

Another way to discover PPI is by using protein sequence alignments, which are methods to align and compare the amino acid sequences of different proteins and

identify regions of similarity or difference [51]. Protein sequence alignment can help to predict protein function by inferring evolutionary relationships, structural features or functional motifs from the sequence conservation or variation [52, 53]. For example, proteins that share a high degree of sequence similarity are likely to have similar functions or belong to the same family or superfamily. Proteins that have conserved regions or motifs are likely to have specific functions or interactions associated with those regions. Proteins that have variable regions are likely to have different functions or adaptations to different environments or conditions. Therefore, protein sequence alignments can help to identify potential PPI interfaces, hot spots, or motifs, which in turn serves as the starting point to design peptides or small molecules that can mimic or disrupt the natural PPIs by targeting these regions.

In addition to sequence-dependent prediction, experimental protein structure information deposited on online databases can be leveraged for PPI predictions. Structural information can help predict protein interaction by providing information about the shape and orientation of proteins and identify regions of proteins that are likely to interact with other proteins [54]. The structure-based method predicts protein-protein interaction if two proteins have a similar structure, such if the known protein partner having similar structure (usually secondary structure) with other protein, they might share similar binding partner.

2.2.2 Nuclear Magnetic Resonance (NMR) spectroscopy

NMR spectroscopy is a well-established method for analysing protein structure, interaction, and dynamics at atomic resolution [55]. NMR can detect changes in the local electronic environment caused by binding events, allowing researchers to identify the areas of a protein that are engaged in a binding interface [56]. Analysis of

data from NMR spectroscopy can also reveal detailed information about protein-protein interaction, such as binding affinity and thermodynamic.

In order to screen ligands for PPIs, one can apply NMR spectroscopy with the use of a technique called chemical shift perturbation (CSP) analysis. Chemical shift perturbation (CSP) analysis is a technique used to study PPIs by comparing the NMR spectra of a protein in its free form to that of the protein in complex with its binding partner [57]. The difference between these two spectra can be used to identify residues that are involved in the binding interface [56]. These residues will experience changes in their chemical environment upon binding and will therefore exhibit changes in their NMR chemical shifts [57]. Changes in NMR chemical shifts suggests binding events between the protein and the binding partner. In addition to detecting binding site, CSP analysis can also quantitatively determine the binding affinity of a protein-protein complex [56]. By titrating one protein into the other and monitoring changes in chemical shifts, it is possible to deduce the dissociation constant (Kd) of the complex, which is useful to determine the binding affinity between the two proteins.

CPMG stands for Carr-Purcell-Meiboom-Gill and is a pulse sequence used in NMR spectroscopy to measure relaxation times of nuclei [58]. It is a modification of the Carr-Purcell sequence [59] that uses a train of π 180-degree pulses to refocus the transverse magnetization and reduce the effects of static field inhomogeneities [60, 61]. The T_2 relaxation times are then analysed to study chemical exchange processes in biomolecules. This is especially useful for screening of PPIs involving large proteins, as it enhances signal-to-noise ratio and reduce experimental time. Protein-bound ligand generally has a short T2 relaxation time similar to the protein, in which the resonances are being filtered out during CPMG measurement [62]. Therefore, when a ligand is bound to a protein, the ligand's signal intensities will drop as the

population of the unbound ligand at any one time decreases. Ligand-observed ^{1}H CPMG NMR screening method is proved to be as effective as 1D protein-observed fluorine NMR (PrOF NMR), which the former does not require protein labelling [63]. This technique was used to screen ligands against multiple proteins such as BPTF, PfGCN5, BSA [64, 65].

2.2.3 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) is a powerful physical technique that can be used to determine the thermodynamic parameters of interactions in solution. It is most often used to study the binding of small molecules (such as medicinal compounds) to larger macromolecules (proteins, DNA etc.) in a label-free environment. The principles of isothermal titration calorimetry (ITC) are based on the measurement of heat changes that occur when two or more molecules interact [66]. ITC measures the heat released or absorbed during a binding event and uses this information to determine the thermodynamic parameters of the interaction.

The ITC instrument consists of two cells which are enclosed in an adiabatic jacket, one containing a reference solution and the other containing a solution of the macromolecule of interest [67]. In ITC, a small amount of one molecule (the ligand) is injected into a solution containing a larger amount of another molecule (the macromolecule). As the ligand binds to the macromolecule, heat is either released or absorbed. This heat change is measured by a highly sensitive calorimeter and used to calculate the thermodynamic parameters of the interaction [66]. Analysis of ITC experiment data enable identification of the binding enthalpy of two proteins, i.e. exothermic or endothermic. Quantification of the reaction heat is also possible, providing a complete thermodynamic description of the protein interaction such as

binding affinity, stoichiometry, enthalpy change, entropy change and free energy change [68]. In addition, if structural information is available, the contributions of specific amino acids mediating the binding event can be identified and their thermodynamic contributions can be quantified [69].

2.2.4 Structural prediction of PPIs

Protein-peptide docking is a computational method for predicting complex structures by sampling potential peptide binding conformations and evaluating the probable protein-peptide complexes using an energy scoring function. [70]. Generally, there are three categories of protein-peptide docking methods, namely template-based docking, local docking, and global docking. Different approaches offer different levels of prediction accuracy, often determined by the amount of interaction information provided as input [71]. Template-based docking uses a known protein-peptide complex as a template to predict the structure of a new complex [72]. It is useful only when the sequence of the peptide is similar to that of a peptide with known structure. Local docking predicts the structure of a protein-peptide complex by searching for the best binding site within the user-defined search box on the protein surface [70], while global docking searches for all possible binding sites on the protein surface [73].

AutoDock Vina is an open-source program to perform molecular docking, with twice the speed of its predecessor, AutoDock4, while providing improved accuracy of prediction [74]. It is one of the docking tools provided under AutoDock suite. AutoDock Vina uses a scoring function that combines terms for intermolecular energy, desolvation energy, and entropic cost of binding. The program performs a global search of the conformational space of the ligand by iteratively sampling conformations and optimising their binding energy using an efficient gradient-based algorithm. The

peptide docking using AutoDock Vina involves several steps. First, the protein structure and peptide structure are prepared using software such as AutoDockTools. The peptide is then docked into the protein using AutoDock Vina. A set of docked poses will be generated, which are ranked based on their predicted binding affinity. The best pose can be selected for further analysis.

A study compared the success rate of AutoDock Vina with AutoDock4 for a large and diverse set of protein-ligand complexes [75]. The study found that AutoDock Vina was more accurate than AutoDock4 in predicting binding poses but losing its edge in binding free energy estimation. AutoDock Vina has been used in several studies to investigate PPIs or protein-ligand interactions. One study had modelled interaction of candidate inhibitors with SARS-CoV-2 main protease using AutoDock Vina, which are structurally homologous with other coronavirus proteases [76]. AutoDock Vina was used to dock antimicrobial peptides with their target protein to predict their residue interactions [77].

On the other hand, molecular dynamics (MD) simulation is another *in silico* technique found to be useful to predict the structure of PPIs. MD simulation is a computational method that mimics the physical motions of atoms and molecules in a system over time [78]. They were first developed in the 1950s and applied to large molecules that lasted less than 10 ps in 1975 [79, 80]. The simulations give detailed information about atomic motions over time. Molecular Mechanics force fields are mathematical models used in MD simulation to describe the interactions and energies of atoms and molecules in biological systems. Molecular Mechanics force fields are usually divided into bonded and non-bonded terms [81]. The bonded terms describe the interactions between atoms that are connected by covalent bonds, such as bond lengths, bond angles, and torsion angles. The non-bonded terms describe the

interactions between atoms that are not covalently bonded, such as electrostatics and van der Waals forces.

MD simulations are used to simulate how biomolecules behave under different conditions, such as temperature, pressure, and solvent, which provide insights into the structural and energetic features of the interface between two protein molecules [78]. The solvent is the liquid environment that surrounds the biomolecules, and it can affect their shape, stability, and function. Proteins may undergo a "glass" transition, which is a change from a flexible state to a rigid state, when the temperature or the solvent properties are changed. This can affect their biological activity and interactions with other molecules. With MD simulation, the positions and velocities of atoms and molecules can also be calculated with respect to time, which helps us to understand how the structure and energetic features of biomolecules are related to their function and regulation [82]. On a higher level, it reveals the dynamics and flexibility of protein interface, which may affect the binding affinity and specificity of the interaction. For instance, simulations can reveal how proteins change their conformation to bind to other molecules, or how they respond to external stimuli or signals. There are many variations of MD methods that use different techniques and models, in which user can adjust the parameters of the simulations to study different properties [83]. Some of the most popular force fields are CHARMM [84], GROMOS [85] and AMBER[86].

MD simulation serves as an alternative *in silico* tool to explore how proteins change their shape and orientation over time, allowing us to visualise and simulate protein-protein interaction [87]. They can also be compared with NMR experiments to validate and improve the simulations [78]. MD simulation can be used to study various aspects of protein structure and function, such as conformational changes, ligand binding, folding, stability, and interactions with other molecules [82, 88]. Together

with molecular docking techniques, MD simulations can help to validate peptide docking by providing mechanistic information and assessing the stability and flexibility of the docked protein-peptide complex [89]. The conformation and energy profiles generated from MD simulation can also complement the docking results by evaluating the binding affinity and atomic properties of the peptide. Researcher had used MD simulation to validate the docking of natural compounds against the SARS-CoV-2 main protease [90]. They performed 50 ns MD simulations for the top-ranked compounds from AutoDock4 to assess the conformational stability and fluctuations of the protein-ligand complexes.

2.3 14-3-3σ

The pivotal role of 14-3-3 proteins in intracellular signalling and cell cycle control originates from their interactions with vast variety of signalling proteins. Most of the 14-3-3 protein partners contain consensus phosphoserine/phosphothreonine-containing peptide motifs, which selectively bind to the major amphipathic groove at the inner concave face of 14-3-3 protein [1] (**Figure 2.1**). Generally, the canonical peptide-binding motifs are characterised as RSX(pS/pT)XP (Mode I), RXXX(pS/pT)XP (Mode II) and (pS/pT)$X_{1–2}$–COOH (Mode III) [91, 92]. Another canonical binding motif similar to Mode I and Mode II was also suggested as (R/K)X_{2-3}(pS/pT)X(P/G) [91, 93]. Nonetheless, unphosphorylated peptides were also reported to interact with residues at the 14-3-3 binding groove [94-96]. Other than the canonical binding groove, peptide fragment screening of 14-3-3 protein identified hypothetical secondary binding sites where fragments bind at the top helices at the C-terminal "hedge" or near the N-terminal dimer interface of 14-3-3 protein [97, 98] (**Figure 2.1**). Consistent with the findings, the binding of several larger proteins to 14-3-3 protein

were discovered at the putative secondary binding sites, for example, serotonin N-acetyltransferase (AANAT), small heat shock protein (HSPB6), plant proton pump PMA2 and plant transcription factor FT [93, 99-101].

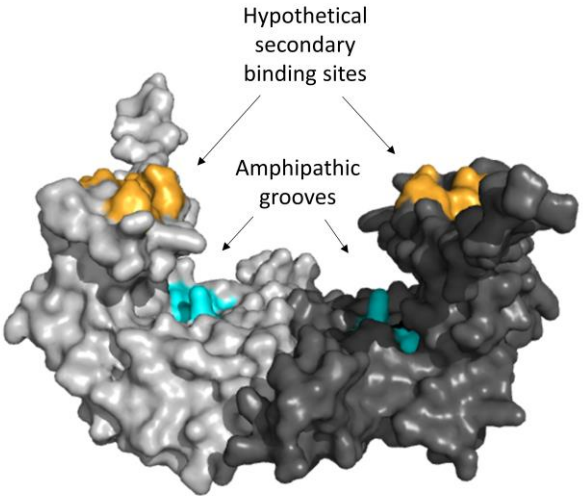

Figure 2.1 AlphaFold model of 14-3-3σ dimer with its canonical amphipathic binding grooves (blue) and hypothetical secondary binding sites (yellow). Monomer model was obtained from AlphaFold and aligned to 14-3-3σ dimer crystal structure (PDB ID: 1YZ5).

Phylogenetic analysis implied that 14-3-3σ is the most diverged isoform in the 14-3-3 family [1]. The X-ray crystal structure of 14-3-3σ suggests its horseshoe-shaped dimer consisting of 9 antiparallel α-helices [1]. Structurally, 14-3-3σ only exists as a homodimer, thanks to its most diverged sequence compared to other 14-3-3 isoforms [2]. 14-3-3σ is also one-turn shorter at its N-terminus, giving it a slightly closer homodimer conformation [1]. Mutation in Phe25 and Gln55 residues slightly promoted heterodimerisation and reduced homodimerisation of 14-3-3σ, while the mutation in Ser5, Glu20 and Glu80 residues lead to heterodimerisation [102]. 14-3-3σ carrying the mutations of all five residues lose its growth-suppressiveness, proposing the importance of 14-3-3σ homodimerisation in tumour suppression [102]. Other 14-3-3σ-specific residue interactions are Phe25-Tyr84 ring-ring interaction and Lys9-

Glu83 salt bridge [1, 2]. Notably, all these residues are located at the homodimer interface, supporting their importance in controlling dimerisation and possibly the prerequisite for 14-3-3σ functionality. On the other hand, pull-down assay showed that ISGylated 14-3-3σ was able to form a dimer with unmodified 14-3-3σ [14], suggesting that ISGylation of 14-3-3σ monomer is insufficient to block its dimerisation. This also suggests that ISG15 transfer (mediated by TRIM25 upon IFN treatment) regulates 14-3-3σ function through a mechanism other than dimerisation blocking. As TRIM25 regulates 14-3-3σ protein level through ubiquitination and ISGylation, whether ubiquitination affects dimerisation of 14-3-3σ or not remains an interesting question to be studied.

2.4 TRIM25

In contrast to the smaller 14-3-3σ, TRIM25 is a 630-residues protein harbouring a RING domain, two B-boxes, coiled-coil (CC) domain and a PRY-SPRY domain (**Figure 2.2A**). N-terminal RING domain is critical for TRIM25's ubiquitination activity, where it binds to E2 ubiquitin ligases for ubiquitin transfer to its target protein [103]. The C-terminus, on the other hand, equipped with the substrate binding PRY-SPRY domain that recognises and interacts with various signalling proteins such as RIG-I (retinoic acid-inducible gene I), Respiratory syncytial virus (RSV) non-structural 1 (NS1) protein and JC polyomavirus (JCV) small t antigen (tAg) [104-106]. Although there is limited information on the structural functions of the two B-boxes and CC domain, these domains were indispensable for TRIM25's interaction with several proteins. Tumour suppressive AT motif-binding factor 1 (ATBF1) demonstrated binding at residue 200-300 of TRIM25's CC domain truncates, and also the intact B-boxes, which leads to proteasomal degradation of ATBF1 [107]. X-ray

crystallography revealed the structure of TRIM25-Influenza A NS1 binding complex, of which NS1-ED (effector domain) interacts at the region spanning residues 212–229 of one TRIM25-CC monomer and residues 274-277 of the other monomer [108]. The NS1 interaction interferes the positioning of PRY-SPRY domain on TRIM25-CC domain, which possibly explains the inhibition of NS1 on RIG-I ubiquitination and subsequent signalling activation [108].

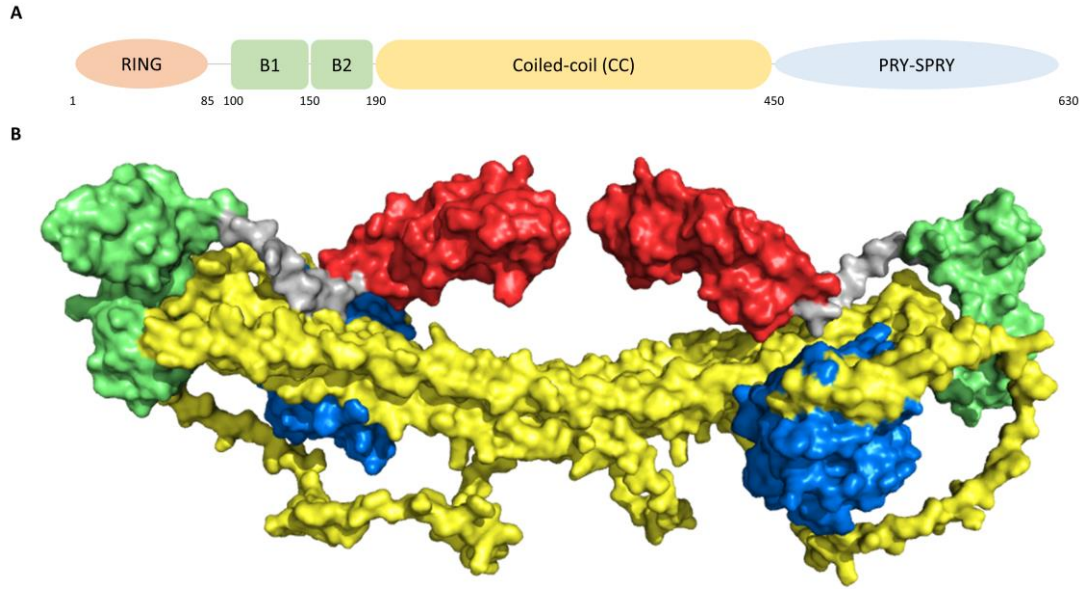

Figure 2.2 (A) Domain architecture of TRIM25 (residue numbering based on existing crystal structure). (B) Full-length TRIM25 dimer model with the RING domain (red), B-boxes (green), CC domain (yellow) and PRY/SPRY domain (blue). Monomer model was obtained from AlphaFold and aligned to TRIM25-CC dimer crystal structure (PDB ID: 4LTB).

In respect of its dimerisation, TRIM25 exists as a stable obligate dimer with a conformation of interdigitating antiparallel helical hairpins, positioning each RING domain at opposite ends and the SPRY domain at the centre [24, 109] (**Figure 2.2B**). A structural study suggests that monomeric TRIM25 is catalytically active, while dimerisation of TRIM25 enhanced its ubiquitinating activity [110]. The phylogenetic analysis illustrates that almost all human TRIM proteins harbour a unique central sequence in their CC domain, which may contribute to the general preference of TRIM

proteins to form a homodimer, and only form heterodimer with very similar family members [24]. Ser195, Ala216, Lys285, Ala234, Glu267 and Leu308 residues were essential for TRIM25 dimer formation through the first helix segment of CC domain [24]. Though TRIM25's dimer architecture effectively improves its ubiquitinating ability, TRIM25 is still functional as a monomer in ubiquitinating substrates, and therefore, the need of TRIM25 to dimerise for 14-3-3σ ubiquitination is yet to be uncovered.

2.5 14-3-3σ-TRIM25 interacting region

RING-domain-deleted TRIM25 failed to downregulate 14-3-3σ level, indicating the importance of the RING catalytic domain for UbcH8 E2 ligase recruitment and subsequently 14-3-3σ ubiquitination [12]. Similarly, the RING domain of TRIM25 is also required for the ISGylation of 14-3-3σ [14]. Co-immunoprecipitation suggests a direct interaction between TRIM25 and 14-3-3σ using mouse xenograft model, in which TRIM25 truncate harbouring B boxes and CC domain is required for the binding [12]. A recent study which showed binding of TRIM25 with another isoform of 14-3-3, 14-3-3η, suggesting that only the CC domain is required for their binding [28]. Another E3 Ub ligase COP1 demonstrated its binding with 14-3-3σ, mainly dependent on its CC domain and slightly on its RING domain [111]. Although these interactions were with a different isoform of 14-3-3 protein or different E3 ligase, all studies suggest that the CC domain of TRIM25 possibly harbours the interacting peptide sequence with 14-3-3σ. Nevertheless, no study showing where or how the CC domain of TRIM25 binds to the 14-3-3σ has been reported to date, though the recent evidence with its η isoform suggested that the 14-3-3η-TRIM25 binding requires interaction at V181 residue but not K50 at the

amphipathic groove of 14-3-3η [28]. While their interaction is independent of phosphorylation, the exact peptide regions of the TRIM25's CC domain that interacts directly with the 14-3-3σ protein are yet to be identified.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Chapter 3 aims to clarify the methodology of the experiments conducted, along with all materials and apparatus used. The methodology was written in succession of the experiments carried out. To summarise, Section 3.2 lists the materials, chemicals, apparatus and software used throughout the study. Section 3.3 describes how *in silico* tools were used to predict potential 14-3-3σ-interacting TRIM25-CC peptide fragments. Section 3.4 explains the process of Fmoc SPPS and procedures to conduct RP-HPLC, Semi-prep LC and LC-MS for peptide validation. Section 3.5 describes the steps to conduct protein expression, SDS-PAGE, ^{1}H NMR and *de novo* peptide sequencing. Section 3.6 illustrates the methodology to perform ^{1}H NMR CPMG assay, competitive CPMG assay and ITC experiments in examining peptide-protein interactions. Section 3.7 describes the procedure in conducting *in silico* molecular docking using AutoDock Vina, and MD simulation using GROMACS software.

3.2 Materials and apparatus

3.2.1 Chemical reagents

The reagents used in this study were analytical grade or higher, with > 95% purity. All chemicals and salts were purchased from Sigma-Aldrich (St Louis, Missouri. USA), Merck Millipore (Darmstadt, Germany) and Gibco (Thermo Fisher Scientific. Inc. Waltham, MA, USA), unless stated otherwise in text.

3.2.2 Fmoc-protected amino acids and resin

Fmoc-protected amino acids (Fmoc-Phe-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Asp(OtBu)-OH and Fmoc-Arg(PBF)-OH) and rink amide AM resin for peptide synthesis were purchased from GL Biochem (Shanghai) Ltd (Shanghai, China).

3.2.3 Bacterial strain

Escherichia coli strain BL21 (DE3) was used for transformation and protein expression of GST-tagged 14-3-3σ protein in this study, with the following genotype: F^- *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm* (DE3) (Novagen, USA).

3.2.4 Recombinant plasmid

Full-length human 14-3-3σ expression construct used in this study was purchased from Eurofins Genomics Blue Heron with the following characteristic: the 14-3-3σ construct, 759 base pairs, vector name: pEX-N-GST (PS100029, ORIGENE) (**Figure 3.1**). Full-length human 14-3-3σ gene was cloned at restriction sites Sgf I and Mlu I, with sequence as described in Appendix A. The recombinant plasmid was expressed as a glutathione-S-transferase (GST) fusion protein in *E. coli* BL21 (DE3).

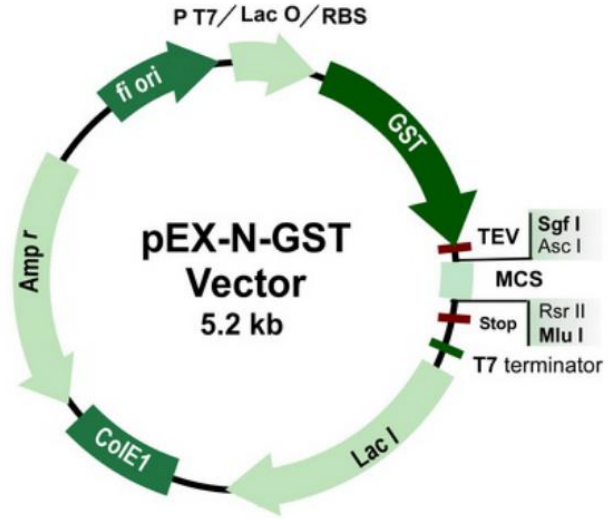

Figure 3.1 pEX-N-GST vector.

3.2.5 Bacterial media

Lysogeny Broth or Luria Broth (LB) and LB Agar were used as bacterial media in this study. The LB media were prepared from LB Broth (Miller) powder (10 g/L NaCl, 10 g/L Tryptone, and 5 g/L Yeast Extract), at 25 g in 1 L of distilled water. To prepare LB agar, 15 g/L of bacteriological grade agar powder (1st BASE, Malaysia) was added to LB to reach a final concentration of 1.5% (w/v).

3.2.6 Buffers and solutions

The buffers and solutions used in this study are summarised in **Table 3.1**.

Table 3.1 List of buffers and solutions with their composition.

Buffer/ solutions	**Component**	**Amount**
Buffer A	Tris HCl	4.44 g
	Tris base	2.65 g
	250 mM NaCl	14.61 g
	Ultrapure water	Top up to 1 L
100 mM Phosphate buffer (PBS), pH 7.0	NaH_2PO_4	6.0 g
	Na_2HPO_4	8.0 g
	50 mM NaCl	2.9 g
	Ultrapure water	Top up to 1 L
NMR buffer	NaH_2PO_4	0.12 g
	Na_2HPO_4	0.16 g
	50 mM NaCl	0.058 g
	Deuterated water (D_2O)	2 mL
	Ultrapure water	Top up to 20 mL
ITC buffer	HEPES	5.95 g
	100 mM NaCl	5.84 g
	Ultrapure water	Top up to 1 L
10x SDS-PAGE running buffer	1.9 M glycine	144.4 g
	250 mM Tris base	30.3 g
	0.1% (w/v) SDS	10.0 g
	Ultrapure water	Top up to 1 L
SDS staining solution	0.04% (w/v) Coomassie Brilliant Blue (R-250)	0.4 g
	25% (v/v) isopropanol	250 mL
	10% (v/v) acetic acid	100 mL
	Ultrapure water	650 mL
SDS de-staining solution	10% (v/v) isopropanol	100 mL
	10% (v/v) acetic acid	100 mL
	Ultrapure water	800 mL

Table 3.1 Continued

8x SDS-PAGE loading dye	1 M Tris, pH 6.8	6 mL
	SDS	2.4 g
	Glycerol	12 mL
	β-mercaptoethanol (BME)	6 mL
	Bromophenol blue (G-250)	0.06 g
	Ultrapure water	Top up to 30 mL

3.2.7 Instruments

Various instruments utilised for the experiments in this study were summarised in **Table 3.2**. Other common lab equipment were not included in the table.

Table 3.2 Instruments and equipment with their manufacturer and application.

Instrument	**Manufacturer**	**Application**
UV-Vis spectrophotometer	PerkinElmer	Peptide synthesis, purification and characterisation
RP-HPLC	Agilent Technologies	
Semi-prep LC	Shimadzu	
LC-MS Accela™	Thermo Scientific	
GSTrap HP column	Cytiva	Recombinant protein expression and purification
MBPTrap HP column	Cytiva	
NMR spectrometer (500 MHz)	Bruker	Protein folding validation, protein-ligand binding assay
MicroCal iTC200 microcalorimeter	Malvern Instruments	Protein-ligand binding assay

3.2.8 Software and online databases/servers

The software and online databases/servers used in this study were summarised in **Table 3.3**.