

https://dspace.org/

Assessing the potential interaction between CBD and TBR1 CBD and T-box domain

Blignaut, Chanel

2024
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/42940
Blignaut, Chanel. (2024). Assessing the potential interaction between CBD and TBR1 CBD and T-box domain [Master's dissertation, University of the Witwatersrand, Johannesburg]. WireDSpace.

Downloaded from a DSpace repository. https://dspace.org/



WITS
UNIVERSITY



ASSESSING THE POTENTIAL INTERACTION BETWEEN CBD AND TBR1 T-BOX DOMAIN

Chanel Blignaut (1831350)



A Dissertation submitted to the Faculty of Health Science, University of Witwatersrand, in fulfilment of the requirements of the degree of Masters of Medicine

Supervisor:

Dr. Sylvia Fanucchi

Dr. Samson Adeyemi

DECLARATION

I Chanel Blignaut (1831350), declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Medicine (MRA03) at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



_____ (Signature of candidate)

19th day of January 2024 at Johannesburg, Gauteng

ABSTRACT

In 2020, Cannabidiol (CBD) emerged as the most commonly used recreational substance among pregnant women and was perceived as a natural and safer option for alleviating pregnancy-related symptoms, yet its potential effects on foetal neurodevelopment remain uncertain. With varied results from existing literature on the association between prenatal cannabis use and Autism Spectrum Disorder (ASD) development, this study focuses on filling these knowledge gaps. It investigates the potential interaction of CBD with the T-box domain of TBR1, a transcription factor implicated in ASD. CBD may cross the placenta and distribute throughout the developing foetal organs, including the brain, where it may interact with TBR1. This study's objective is to lay the groundwork for future research into the impact of CBD binding on TBR1 functionality, whose dysregulation is implicated in ASD. The study aims to use *in vitro* and *in silico* methods to identify and characterise the interaction between CBD and TBR1 T-box Domain. Initially, predictive models were utilised to determine the structure of the TBR1 T-box domain and its binding domains. Subsequently, the ADMET properties of CBD are assessed to determine its potential interaction with TBR1 T-box domain within the body. Through the optimisation of the TBR1 T-box domain and CBD structures, induced fit docking and molecular dynamics simulations, the study aims to predict the potential interaction sites, dynamics and stability of this interaction. The study confirms the computational results using *in vitro* methodologies. After expressing and purifying the TBR1 T-box domain, a pull-down assay (PDA), thermal shift assay (TSA) and Time-resolved Fourier-transformed infrared spectroscopy (TR-FTIR) are used to evaluate the potential binding, stability and physiochemical properties of the interaction. Computational analysis, using Maestro Schrödinger Induced Fit Protocol, predicts that CBD binds stably within a hydrophobic pocket of TBR1 T-box domain, away from its DNA-interacting residues. Pose 2 and 3 from molecular docking shows the highest binding affinity and molecular dynamics simulations, using Maestro Schrödinger Desmond Molecular Dynamics System, reveal that the TBR1 T-box domain stabilises upon interaction with CBD. Specifically, the interaction is facilitated by hydrophobic interactions and hydrogen bond formation with residues Ser238, Pro335, Thr360, Glu363 and Asn240. Experimental validation through PDA and TSA provided inconclusive results, but TR-FTIR confirmed the dynamic nature of the CBD-TBR1 interaction, characterised by time-dependent spectral changes. While the results do not directly indicate an impact of CBD on TBR1 functionality, further DNA binding studies are necessary for confirmation. This study suggests caution in using CBD during pregnancy due to its complex and largely

unexplored interaction with TBR1, underscoring the need for more comprehensive research to conclusively understand its influence on neurodevelopmental disorders and its therapeutic potential.

ACKNOWLEDGEMENTS

I extend my sincere gratitude to the University of the Witwatersrand for providing me with an exceptional academic environment and the necessary resources to pursue my research endeavours. Your commitment to excellence has been instrumental in shaping my educational journey.

I would like to express my heartfelt appreciation to the Protein Structure Function Research Unit (PSFRU) for their invaluable assistance in facilitating and completing my experiments. Your support has been pivotal in the successful execution of my research.

I am profoundly thankful to Dr. Sylvia Fanucchi and Dr. Samson Adeyemi for their unwavering guidance, mentorship and scholarly insights. Your dedication to my academic and personal growth has left an indelible mark on my journey. I would like to extend my sincere gratitude to Dr. Ikechukwu Achilonu for his invaluable assistance with the computational aspects of this research, utilising Maestro Schrödinger. Additionally, I am grateful to the Centre for High Performance Computing (CHPC) for providing access to Maestro Schrödinger, which was crucial in facilitating the computational analysis and significantly contributed to the quality and depth of the research findings.

Furthermore, I extend my deep gratitude to the National Research Foundation (NRF) for their generous bursary support. Your financial assistance has lightened my academic burden and allowed me to focus on my studies and research.

This journey would not have been possible without the collective support, encouragement, and contributions from these institutions and individuals. I am truly privileged to have had such a robust network of support throughout my academic pursuit.

Please note that AI has been utilised in the preparation of this document. The AI was employed solely for the purposes of grammar and spelling checks. No content or writing in this document has been generated by the AI. All substantive content has been created independently by the author.

With profound appreciation,

Chanel Blignaut

CONTENTS

DECLARATION I

ABSTRACT II

ACKNOWLEDGEMENTS IV

LIST OF FIGURES IX

LIST OF TABLES XI

LIST OF SYMBOLS XII

LIST OF ABBREVIATIONS XIII

1. CHAPTER 1: INTRODUCTION 1

1.1 Classification of *Cannabis sativa. L* 1

1.2 Chemical Entities 2

1.2.1 CANNABIDIOL (CBD) 3

1.3 Endocannabinoid System 4

1.4 Prenatal Cannabis Use 6

1.4.1 Prenatal Cannabis Use and Autism Spectrum Disorder (ASD) 7

1.5 Transcription Factors 9

T-box Transcription Factors 10

1.5.1 TBR1 T-box Proteins 12

Study Rationale and Aims and Objectives 16

Study Rationale 16

Aim: 18

Objectives: 18

2 CHAPTER 2: LITERATURE REVIEW 20

2.1 History of Cannabis 20

2.1.1 Cannabis in South Africa 20

2.2 Cannabis in the Human Body 21

2.3 Therapeutic Potential of Cannabis 22
2.3.1 Cannabis in the Treatment of Diseases 23
2.3.1.1 In Silico Methods and Their Benefits in Biomedical Research 24
3 CHAPTER 3: METHODOLOGY 26
Objective 1: Assess the Potential Interaction between TBR1 T-box Domain and CBD using Bioinformatics 26
3.1 Structural Analysis 27
3.2 Predicting Binding Domains 28
Schrödinger Maestro User Interface 30
3.3 CBD ADMET Properties 31
3.4 Structure Preparation 32
3.4.1 Protein Preparation: TBR1 T-box domain 32
3.4.2 Ligand Preparation: CBD (Cannabidiol) 33
3.5 Molecular Docking 34
3.6 Molecular Dynamics 37
3.6.1 Post Simulation Analysis 38
Objective 2: Assessing the Potential Interaction between TBR1 T-box domain and CBD using Experimental Approaches 42
3.7 Transformation of pET-11a Plasmid into *E.coli* cells 43
3.8 Expression of TBR1 T-box domain 46
3.9 Purification and Protein Purity 48
3.9.1 Immobilised Metal Affinity Chromatography 48
3.9.2 Size Exclusion Chromatography 51
3.9.3 SDS-PAGE 53
3.9.4 Protein Concentration and Purity 55
3.10 Structural Analysis 55
3.10.1 Far-UV Circular Dichroism 55

3.11 Pull-Down Assay 57
3.12 Thermal Shift Assay 60
3.13 Time-resolved Fourier Transformed Infrared Spectroscopy 62
4 CHAPTER 4: RESULTS 66
Objective 1: Assessing the Potential Interaction between TBR1 T-box Domain and CBD using Bioinformatics 67
4.1 Structure Analysis 67
4.2 Predicting Binding Domains 69
4.3 CBD Structure and ADMET Properties 72
4.4 Molecular Docking 76
4.5 Molecular Dynamics: Post Simulation Analysis 83
4.5.1 Simulation Quality Analysis 83
4.5.2 Molecular Dynamics Analysis 85
Objective 2: Determine the Potential Interaction between TBR1 T-box Domain and CBD using Experimental Approaches 96
4.6 Transformation and Expression 96
4.7 Purification and Protein Purity 98
4.7.1 Immobilised Metal Affinity Chromatography (IMAC) 98
4.7.2 Size Exclusion Chromatography (SEC) 100
4.8 Structural Analysis of TBR1 T-box Domain 101
4.8.1 Far-UV Circular Dichroism (CD) 101
4.9 Pull-Down Assay 104
4.10 Thermal Shift Assay 107
4.11 Time-resolved Fourier Transform Infrared Spectra 108
5 CHAPTER 5: DISCUSSION 116
TBR1 T-box Domain Structure 117
DNA Binding Residues 117

The Potential for CBD to reach TBR1 T-box domain 118
CBD Meets TBR1 T-box domain: A Computational Choreography 119
CBD Predicted Binding Pocket 119
The stabilising effect of CBD 120
Affinity Matters: CBD's Interaction with TBR1 T-box domain 122
Biologically relevant CBD-docked pose 123
CBD and TBR1 T-Box Domain: Time to Dance in the Lab 124
The Pull-Down Assay 124
Thermal shift assay 126
Time-Resolved FTIR 127
CONCLUSION 130
Limitations 132
Future Recommendations 132
References: 133
Appendix A: Turn-it-in Report 159

LIST OF FIGURES

Figure 1: Taxonomy for Cannabis.. 2

Figure 2: The Chemical Structure of the Main Phytocannabinoids found in Cannabis sativa. L.. 3

Figure 3: Chemical Structure of CBD and THC 4

Figure 4: Prenatal Cannabis Exposure and ASD Risk Mechanisms. 8

Figure 5: T-box Family Tree and Identity Matrix. 11

Figure 6: TBR1 T-box Domain Structure Comparison with TBX5 14

Figure 7: Comparing Rigid and Induced Fit Docking Models. 34

Figure 8: pET-11a Plasmid Utilized in Transformation.....44

Figure 9: pET11-a Plasmid Expression System in T7 E.coli Cells 46

Figure 10: IMAC Resin Structure and Protein Purification Process. 49

Figure 11: Principle for Size Exclusion Chromatography (SEC) 51

Figure 12: Principle of Far-UV Circular Dichroism 56

Figure 13: Epoxy Activated Agarose Pull-Down Assay.....58

Figure 14: Thermal Shift Assay Principle.....61

Figure 15: The fundamental principle underpinning Fourier Transform Infrared (FT-IR) spectroscopy..…..63

Figure 16: Amide I, Amide II and Amide III Vibrations in FTIR..…..64

Figure 17: AlphaFold Predicted Structure of TBR1 T-box Domain.....67

Figure 18:: Sequence and Structure Alignment between the T-box domains of TBR1 and TBX5. 68

Figure 19: TBR1 T-box Domain Binding Pocket Prediction. 70

Figure 20: Various Poses Generated from the Induced Fit Docking between TBR1 T-box Domain and CBD. 79

Figure 21: Protein Cα and ligand Fit-to-protein RMSD Plots for TBR1 T-box domain and CBD for the APO and CBD-docked TBR1 T-box domain systems.. 87

Figure 22: 2D Ligand Interaction Diagrams and Protein-Residue Bonds Distribution Across all Four Poses in 250 ns Simulation. 91

Figure 23: Cα-Root Mean Square Fluctuation (RMSF) for each residue in the TBR1 T-box domain, emphasising residues involved in hydrogen bond formation with the ligand across the entire 250 ns simulation.. 92

Figure 24: Radius of Gyration difference between CBD-Bound and Unbound TBR1 T-box Domain over the 250 ns simulation. 94

Figure 25: Codon Optimised Sequence and Experimental Sequence of TBR1 T-box Domain Sequence Alignment. 97

Figure 26: Purification and Purity Profile of TBR1 T-box Domain using IMAC 99

Figure 27: Purification and Purity of TBR1 T-box domain using SEC. 100

Figure 28: Far UV Circular Dichroism Spectrum for TBR1 T-box Domain. 102

Figure 29: Absorbance Spectrum of CBD and Results of the Pull-Down Assay with TBR1 T-box Domain at 1 Hour and 24 Hours.....104

Figure 30: Thermal Melting Curves for TBR1 T-box Domain upon CBD Binding. 107

Figure 31: FTIR Spectra of Equilibration Buffer, CBD and TBR1 T-box Domain Controls.. 109

Figure 32: Time-resolved FTIR Spectrum of TBR1 T-box domain and CBD over 70 minutes..112

Figure 33: Time-resolved FTIR spectrum of the Interaction between CBD and TBR1 T-box Domain in the Amide region. 114

LIST OF TABLES

Table 1: TBR1 T-box Domain Predicted Binding Pocket Analysis by SiteMap 71

Table 2: Table Describing the ADMET Properties of CBD.....73

Table 3: Cα-RMSD Comparison Between Native and Docked Ligands, and Native and Docked T-Box Domain Poses 82

Table 4: Molecular Dynamics Simulation Quality Analysis After Equilibration.....83

Table 5: Secondary Structure Prediction after MD Simulations 95

LIST OF SYMBOLS

Degree Celsius - °C

Energy - kcal/mol

Gravitational force - Xg

Kelvin - K

Kilogram - Kg

Kilopascals - kPa

Microlitre - μl

Micromolar - µM

Milligram per millilitre - mg/mL

Millimolar - mM

Nanometres - wavelength (nm)

Percentage - %

Area in Angstroms - $Å^2$

Percentage - %

Volts - V

Volume in Angstroms - $Å^3$

Micromolar per millilitre - µM/mL

Per centimetre per mole - M^{-1} cm^{-1}

Absorbance - Au

Millilitre - mL

LIST OF ABBREVIATIONS

∆9-THC - . ∆9-tetrahydrocannabinol

2D – 2 dimensional

2xYT media - Double-strength Yeast Extract Tryptone media

3D - 3 dimensional

ADMET properties - Absorption, Distribution, Metabolism, Excretion, and Toxicity properties

AKTA - Automated Continuous Flow Chromatography System

APS - Ammonium Persulfate

ASD – Autism Spectrum Disorder

ATR - Attenuated Total Reflectance

ATR-TR-FTIR - Attenuated Total Reflectance - Fourier Transform Infrared Spectroscopy

Au – Absorbance

BBB - Blood-Brain Barrier

CAGR - Compound Annual Growth Rate

CASK: Calcium/Calmodulin-Dependent Serine Protein Kinase

CASP14 - Critical Assessment of Protein Structure Prediction 14

CASTp3.0 - Computed Atlas of Surface Topography of proteins 3.0

CB1 - Cannabinoid Receptor Type 1

CB2 - Cannabinoid Receptor Type 2

CBC – cannabichromene

CBD - Cannabidiol

CBDA - Cannabidiolic acid

CBG – cannabigerol

CBN – cannabinol

CD – Circular Dichroism

CHPC - Centre for High Performance Computing

Cl^- - Chloride ion

cm – centimeter

CNS - Central Nervous System

CO_2 - Carbon dioxide

CPL - circularly polarised light

CYP3A - Cytochrome P450 3A

Cα-RMSD – Alpha carbon Root Mean Square Deviation

Cα-RMSF - Cα-Root Mean Square Fluctuations

DD/ID - Developmental Delays/Intellectual Disabilities

DNA - Deoxyribonucleic Acid

Dscore - Druggability Score

DTT - Dithiothreitol (Cleland's reagent)

E. coli - Escherichia coli

ECS - Endocannabinoid system

FDA - U.S. Food and Drug Administration

Fourier Transform Infrared

FOXP2: Forkhead Box Protein P2

Glide - Grid-Based Ligand Docking with Energetics

HSA - human serum albumin

HEPES - N-2-Hydroxyethylpiperazine-N′-2-ethanesulfonic acid

His-tag – Histidine Tag

HTH - helix-turn-helix

IFD - Induced Fit Docking

IFDScore - Induced Fit Docking Score

IMAC - Immobilised Metal Affinity Chromatography

IPTG - Isopropyl β-D-1-thiogalactopyranoside

kDa – kilodalton

LacO gene – Lac Operator Gene

LP - Ligand-Fit-to Protein

LP-RMSD – Ligand fit to protein root mean square deviation

LZ - leucine zipper

MD - Molecular Dynamics

mg – milligram

mL – mililitre

mM – Milimolar

mRNA -messenger RNA

MS-Excel - Microsoft Excel

Na^+ - Sodium ion

NASEM - National Academy of Sciences, Engineering, and Medicine

nm – nanometer

NPT - Number of particles, Pressure, and Temperature

NRF - National Research Foundation

ns – Nanosecond

OD600 – Optical Density of 0.6

OH - Hydroxyl Group

OPLS - Optimised Potentials for Liquid Simulations

PAE - Predicted aligned error

PDA - Pull-Down Assay

PDB - Protein Data Bank

PEM - Photo Elastic Modulator

pET11a - plasmid Expression Vector-11a

pLDDT - per-residue model confidence score

PMSF - Phenylmethylsulfonyl Fluoride

PMT - Photo Multiplier Tube

PSFRU - Protein Structure Function Research Unit

qPCR - quantitative polymerase chain reaction

QPlogBB - Predicted logarithm of the blood-brain barrier partition coefficient.

QPlogKhsa - Predicted logarithm of the binding constant to human serum albumin.

QPlogPC16 - Predicted logarithm of the octanol-water partition coefficient for compounds at pH 7.4.

QPlogPo/w - Predicted logarithm of the octanol-water partition coefficient.

QPPCaco - Predicted apparent Caco-2 cell permeability.

QPPMDCK - Predicted apparent MDCK cell permeability.

QSAR - Quantitative Structure-Activity Relationships

rGyr - Radius of Gyration

RMSD - Root Mean Square Deviation

RMSF - Root Mean Square Fluctuation

RNA - Ribonucleic Acid

RuleofFive - Rule of Five (a rule for drug-likeness)

SDS - Sodium Dodecyl Sulphate

SDS-PAGE - Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

SEC - Size Exclusion Chromatography

SOC media - Super Optimal Broth with Catabolite repression

std. dv – Standard Deviation

T-BET - T-Box Transcription Factor (TBX21)

TBR1 - T-Brain 1 Transcription Factor

TBR2 - T-Brain 2 Transcription Factor (EOMES)

TBX5 - T-box transcription factor 5

TEMED - Tetramethyl ethylenediamine

TF - Transcription Factor

THC - Tetrahydrocannabinol

THCA - Tetrahydrocannabinolic Acid

THCV – Tetrahydrocannabivarin

TIP3P - Transferable Intermolecular Potential with 3 Points

Tm - Melting Temperature

TR-FTIR – Time Resolved Fourier Transformed Infrared Spectroscopy

TSA - Thermal Shift Assay

UV – Ultraviolet

w/v - Weight/Volume (concentration)

WHO – World Health organization

W-HTH - winged helix-turn-helix

ZF - zinc finger

ΔRg – Change in radius of gyration

1. CHAPTER 1: INTRODUCTION

Cannabis stands out as the most widely used, cultivated, and trafficked illicit drug globally (Smart and Pacula., 2019; WHO., 2023). It holds the top position as the most commonly chosen recreational substance among pregnant women (Badowski and Smith., 2020). Within the realm of maternal health, the issue of cannabis use during pregnancy has become a complex and contentious landscape. More expectant mothers are opting for this popular recreational drug, perceiving it as a natural and safer alternative to conventional western medicine (Badowski and Smith., 2020). Over the last decade, there has been nearly a twofold increase in cannabis use among expectant mothers, particularly during the critical first trimester (Brown *et al.,* 2017). This increase is linked to the desire to alleviate pregnancy-related symptoms such as morning sickness and pain (Beatty *et al.,* 2012).

Cannabis, commonly known as marijuana, is a plant sought after for inducing altered states of consciousness, yet concurrently valued for its therapeutic properties (Young-Wolff *et al.,* 2022). This dual impact of cannabis on the human body is mediated through its interaction with the endocannabinoid system (ECS). It's important to note that the ECS is abundant in the developing foetal brain (Fride *et al.,* 2009). As the chemical constituents of cannabis readily traverse the placenta, concerns arise about their direct influence on foetal neurodevelopment. Notably, using cannabis during pregnancy has been linked to neurodevelopmental disorders, particularly Autism Spectrum Disorder (ASD) (Shorey-Kendrick *et al.,* 2023). This connection emphasises the critical need to understand the implications of maternal cannabis use on the developing foetus.

1.1 Classification of *Cannabis sativa. L*

Cannabis has been cultivated worldwide for thousands of years, serving recreational, spiritual, and medical purposes (Vučković *et al*., 2018). This herbaceous plant, part of the Cannabaceae family, annually grows up to 5 m in warm and temperate zones. The taxonomic classification of cannabis has been disputed for centuries. The botanist Karl Linnaeus recognised the only species of cannabis to be *Cannabis sativa.L*., with the "L." denoting Linnaeus. However, other botanists like Small and Cronquist (1976) adopted a biphasic approach, considering both morphology and chemical content. This led to the further classification of the species *Cannabis sativa. L* into four cannabis taxa, as shown in Figure 1 (Small and Cronquist., 1976).

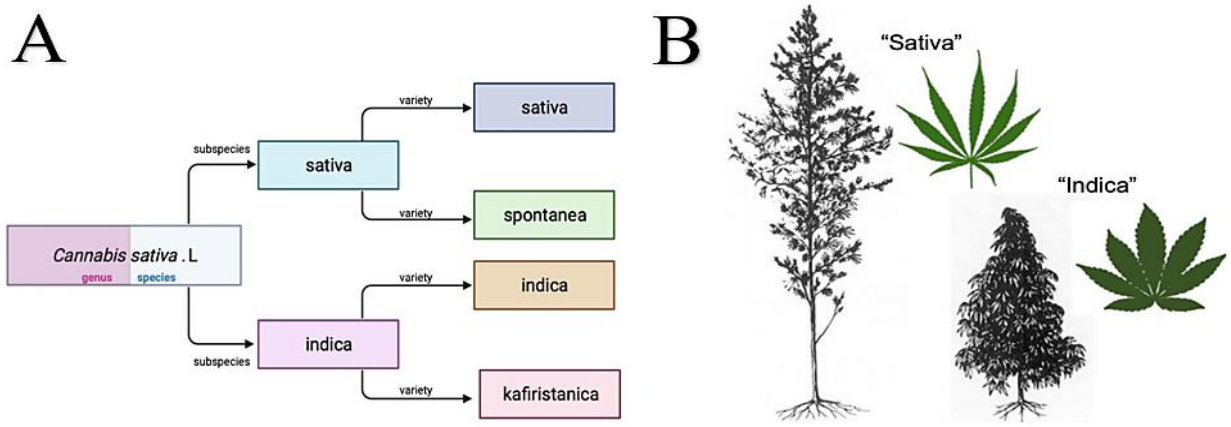


Figure 1: Taxonomy for Cannabis. (A): The most widely accepted taxonomy of Cannabis. Classifies Cannabis sativa.L as the only species of Cannabis however, it can be further classified into subspecies, namely Cannabis sativa and Cannabis indica, depending on their major cannabinoid content (CBD and THC). Cannabis sativa, higher in THC, are further classified into the varieties sativa and spontanea (both with low CBD content). Cannabis indica, higher in CBD, is further classified into the varieties indica and kafiristanica (both low in THC content) (Small and Cronquist., 1976). (B): *C. sativa* is classified as having narrow, slender and tall leaves and the trees are from Indian, African, Colombian, and Mexican descent. *C. Indica* is classified as having broad and compact leaves and is from Afghanistan descent. The image is adapted from Anderson (1980).

1.2 Chemical Entities

Cannabis exerts its effects on the human body through the various chemical compounds found in the plant. Cannabinoids, the primary chemical entities produced by the cannabis plant, fall into three classes: Phytocannabinoids (from the cannabis plant), endocannabinoids (naturally produced by the human body), and synthetic cannabinoids (man-made) (Schurman *et al.,* 2020). Phytocannabinoids, including tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA), are synthesised in the stalked glandular trichomes of *Cannabis sativa L.* (Fairbairn, 1972). Trichome concentrations are lower in male plants compared to their female counterparts (Dayanandan and Kaufman., 1976). These compounds are stored within the small secretory cavities developed by trichomes, also holding terpenes and flavonoids (ElSohly and Slade., 2005). The nature of cannabinoids, marked by their non-polar and highly lipophilic characteristics, results in low solubility in polar solvents like water (Hanuš and Hod., 2020). Among the diverse range of cannabinoids, THC (tetrahydrocannabinol) and CBD (cannabidiol) emerge as standout entities due to their distinct properties (depicted in Figure 2, along with other less prevalent cannabinoids).

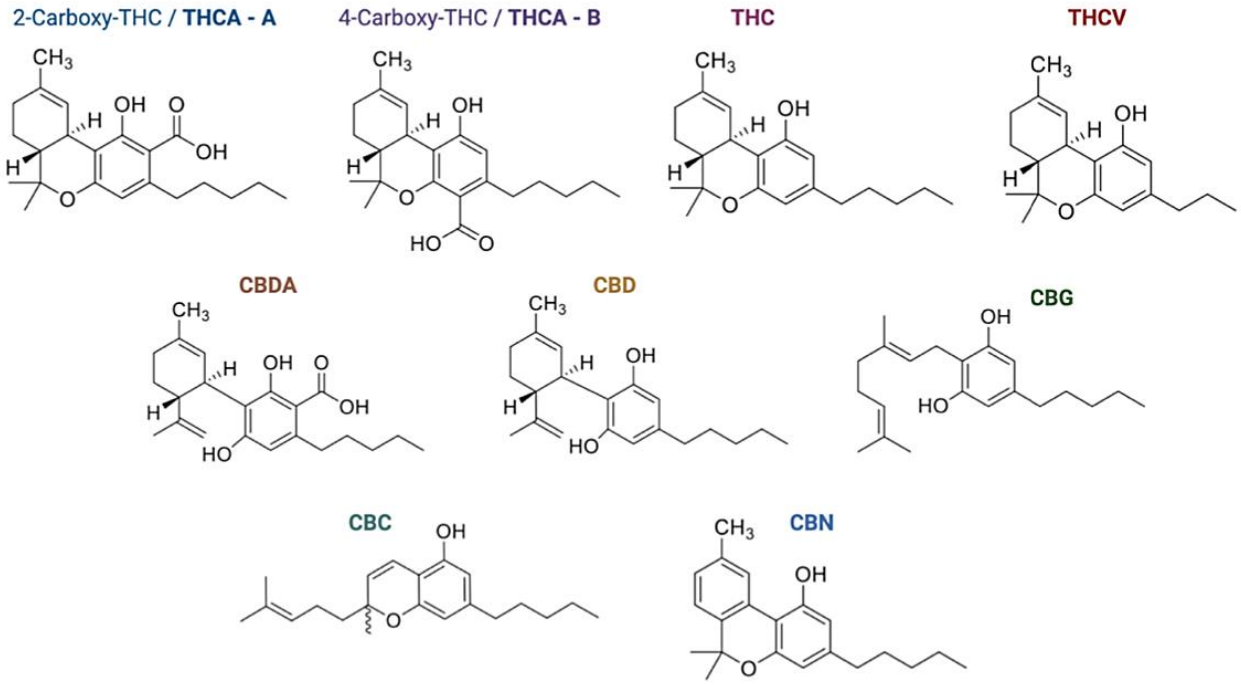


Figure 2: The Chemical Structure of the Main Phytocannabinoids found in Cannabis sativa. L. Δ^9-tetrahydrocannabinol (Δ^9-THC), more commonly known as THC, has two acidic counterparts – THCA-A and THCA-B is shown, alongside Tetrahydrocannabivarin (THCV). Cannabidiol (CBD) can be seen alongside its acidic counterpart CBD acid (CBDA), together with cannabigerol (CBG), cannabichromene (CBC) and cannabinol (CBN). Adapted from (Tahir *et al*., 2021).

1.2.1 CANNABIDIOL (CBD)

THC and CBD, key cannabinoids in cannabis, have unique properties that attract considerable attention. THC, renowned for its psychoactive effects, sharply differs from CBD, which lacks psychoactivity but is intriguing for its potential therapeutic uses. CBD has been explored for its neuroprotective qualities (Hampson *et al.,* 1998), analgesic properties (Russo and Guy, 2006), sedative effects (Russo and Guy., 2006), anti-inflammatory attributes (Malfait *et al.,* 2000), and anti-spasmodic potential (Baker *et al.,* 2000). Due to the array of health benefits, CBD usage has become a popular topic of interest. Despite the structural similarity between CBD and THC (Figure 3), CBD does not induce the same euphoric effect as THC (Russo., 2017).

CBD ($C_{21}H_{30}O_2$), a terpene phenolic compound comprising 21 carbon atoms, naturally occurs as (-)-CBD, featuring a cyclohexene ring, phenolic ring and a pentyl side chain (Elmes *et al.,* 2015). The -OH groups and pentyl chain located at C-1/5 and C-3 positions of the phenol ring contribute

to a significant hydrophobic area in the structure, influencing its hydrophobicity and lipophilicity (Baban *et al.,* 2021). The structural composition of CBD is a key determinant of its physiochemical characteristics, exerting a profound influence on its therapeutic potential within the human body through the ECS.

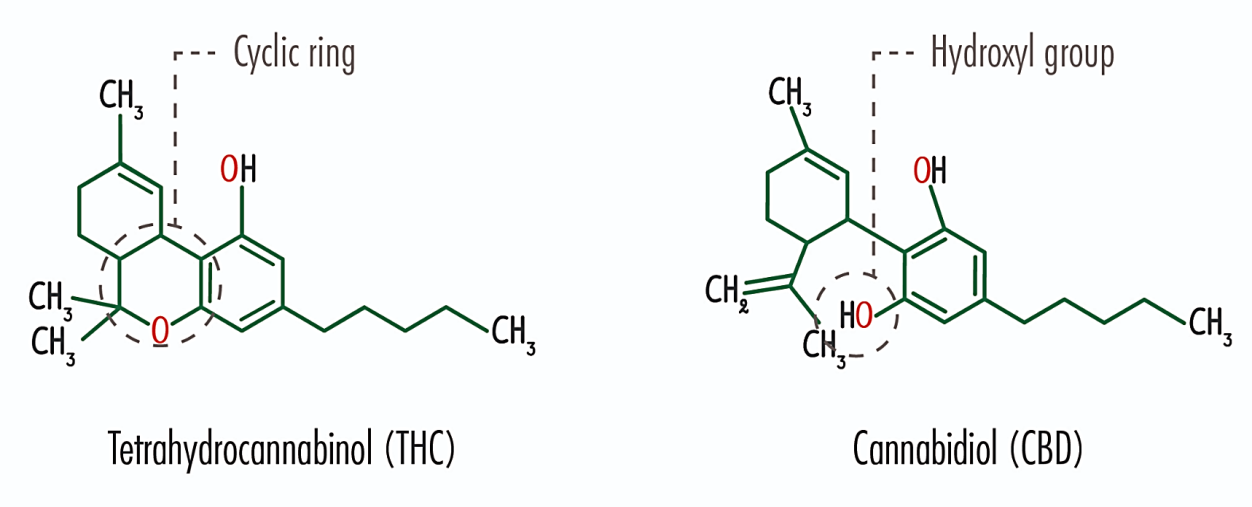


Figure 3: Chemical Structure of CBD and THC: (A): Structure of Δ-9-THC. (B): Structure of CBD. CBD and THC share identical molecular formulas, denoted as $C_{21}H_{30}O_2$. Despite having the same molecular formula, THC contains a cyclic ring which is key to its psychoactive effects. Additionally, their molecular weights are nearly indistinguishable, with THC possessing a mass of approximately 314.469 g/mol and CBD having a slightly lesser mass of around 314.464 g/mol. Image sourced from Hanuš *et al.,* (2005).

1.3 Endocannabinoid System

In recent decades, the once niche field of cannabis research has witnessed a surge in interest, particularly in understanding the intricate mechanisms of the ECS (Lawrence and Gill., 1975). The ECS, a complex network involving cannabinoid receptors (CB1 and CB2), natural cannabinoids and associated enzymes, plays a pivotal role in regulating various bodily functions, ranging from memory and mood to pain modulation and cardiovascular processes (Vučković *et al.,* 2018).

Exploration into the physiological impact of CBD and its synthetic counterparts has illuminated clear evidence of drug-receptor interactions, marked by stringent structural and stereo-selectivity (Hollister., 1974; Jones *et al.,* 1974). CB1 receptors (Cannabinoid Receptor 1) are predominantly found in the central nervous system (CNS), including the brain and spinal cord, while a second cannabinoid receptor, the CB2 receptor, is primarily expressed in immune cells and some peripheral tissues such as the spleen, tonsils, and bone marrow (Pacher *et al.,* 2006; Cabral and

Griffin-Thomas., 2009). Beyond neural territories, CB1 receptors extend their reach into non-neuronal tissues such as the liver, adipose tissue, and skeletal muscle (Batkai *et al.,* 2007; Pagano *et al.,* 2007). CB2 receptors, too, make their presence known in the gut and liver, hinting at their potential roles in governing gastrointestinal and hepatic functions (Izzo *et al*., 2009). Intriguingly, THC functions as a partial agonist at both cannabinoid receptors, favouring CB1 receptors (Cooper and Haney., 2008). Conversely, CBD has a low affinity for both CB1 and CB2 receptors and assumes an antagonistic role, reducing the effects and potency of THC (Babalonis *et al.,* 2017). The strategic distribution of CB1 receptors primarily within the CNS and CB2 receptors in immune-related cells, combined with the distinct binding affinities of CBD and THC, provide insights into their diverse effects. This explains why THC is linked to psychoactive properties, while CBD, with its heightened therapeutic affinity, exerts a more substantial medicinal influence on the body (Vučković *et al.,* 2018).

Exploring how cannabinoids interact with the human body unveils complexity in bioavailability. When inhaled, THC has a 20% bioavailability, dropping to 6% orally. Less than 1% of these amounts reaches the brain, underscoring THC's potent psychoactive effects (Cabral and Jamerson., 2014). In comparison, CBD has a 31% bioavailability when inhaled and 6% orally, with less than 1% reaching the brain, similar to THC. The heightened lipophilicity of cannabinoids contributes to its rapid distribution throughout adipose tissue and various organs, including the brain. This swift journey is attributed to its ability to traverse the blood-brain barrier (BBB) and interact with receptors within the brain, particularly those embedded within the intricate ECS. The BBB historically acts as a formidable barrier, impeding the entry of large molecules and toxins into the brain. Nevertheless, cannabinoids, armed with their lipophilic properties, efficiently diffuse through the cell membranes, establishing interactions with transporter proteins that actively facilitate their passage across the BBB (Calapai *et al*., 2020; Pacher *et al.,* 2006).

The early presence of the ECS in various brain structures during embryonic development, underscores its potential influence on shaping neurodevelopmental processes in the foetus (Fride *et al.,* 2009; Harkany *et al.,* 2008). The ECS plays a role in the movement of neurons and the extension of axons (Fride *et al.,* 2009), the formation of glial cells, the proliferation and differentiation of neural stem cells (Galve-Roperh *et al.,* 2013), as well as the coordination of axonal migration, connectivity, and the formation of synapses (Berghuis *et al.,* 2007). However, this finely tuned system is not immune to external factors. Research, spanning both animals and humans, has uncovered compelling evidence suggesting that exposure to external cannabinoids

during foetal development may lead to persistent neurobehavioral abnormalities in the offspring (Fried and Smith., 2001). This intricate web of interactions sets the stage for a deeper exploration into the broader implications of such influences, particularly in the context of cannabis use during pregnancy.

1.4 Prenatal Cannabis Use

Extensive marketing campaigns have led to CBD gaining popularity as a safe, non-psychoactive self-treatment option (Wieckiewicz *et al.,* 2022). Pregnant women perceive CBD as a natural alternative to traditional medications for pregnancy-related symptoms such as nausea and pain, despite uncertainties surrounding its safety on foetal development (Vanstone *et al.,* 2022). The rising perception that cannabis consumption in pregnancy carries no risks has contributed to an increase in usage among expectant mothers (Vanstone *et al.,* 2022). However, both the U.S. Surgeon General and the U.S. Food and Drug Administration (FDA) express concerns about potential adverse effects of cannabis use during pregnancy. Although specific consumption rates for pregnant individuals are currently unavailable, a study detected 22% THC metabolites in cord blood samples from this demographic (Metz *et al.,* 2019), implying potential CBD consumption within the same cohort (Bidwell *et al.,* 2018). This trend aligns with the increased accessibility of CBD since its reclassification to a schedule 0 drug in 2018 under The Medicines and Related Substances Act (No 101 of 1965) of South Africa (The Medicines Act).

Upon administration, CBD enters the blood–stream and by its lipophilic nature, transverses through the placenta, into foetal circulation and distributes throughout the foetal organs (NIDA., 2023). CBD is known to bind and activate receptors in the brain required for proper foetal neurodevelopment (Grand *et al.,* 2010). This prenatal exposure may result in persistent developmental consequences for offspring, involving cognitive impairments and behavioural issues (Campolongo *et al.,* 2011). Supporting these concerns, the Oregon Health and Science University conducted a study revealing that cannabis exposure during pregnancy influences the placental and foetal epigenome. This, in turn, results in changes consistent with those observed in prevalent neurobehavioral conditions, including ASD, intellectual disability, and other neuropsychiatric disorders (Shorey-Kendrick *et al*., 2023). Although there is a lack of extensive clinical evidence on the impact of cannabis exposure during pregnancy, the increasing incidence of prenatal cannabis use is a major worry. This is because any immediate medicinal advantages

for alleviating nausea and pain may be overshadowed by the potential for negative long-term effects on the offspring (Young-Wolff *et al*., 2019).

1.4.1 Prenatal Cannabis Use and Autism Spectrum Disorder (ASD)

A limited number of empirical studies have investigated the relationship between prenatal cannabis use and ASD. In a study conducted by Corsi *et al.,* (2020), ASD was the primary focus when exploring the potential link between maternal cannabis use during pregnancy and the likelihood of the offspring developing neurodevelopmental disorders. Among the group exposed to prenatal cannabis use, the incidence of ASD was 4.0 per 1,000 person-years, while the unexposed group reported an incidence of 2.4 (Corsi *et al.,* 2020). This research stands out as one of the limited investigations that explores the enduring effects associated with prenatal cannabis exposure, utilising a substantial and representative sample. Few studies have ventured into this aspect of prenatal cannabis exposure research, making such research crucial and in demand. Additionally, prenatal cannabis use has been linked to an increased risk of lower birth weight in the offspring (National Academies of Sciences., 2017), which has been identified as a risk factor for ASD (Autism National Institute of Environmental Health Sciences., 2020). However, in contrast to these findings, a more recent study found no correlation between ASD in offspring and self-administered prenatal cannabis use (DiGuiseppi *et al.,* 2021). Consequently, there is a pressing need for additional research to address existing gaps in the current literature on this subject. Figure 4 presented below illustrates the distinct stages of prenatal cannabis exposure and their potential implications in the context of ASD susceptibility.

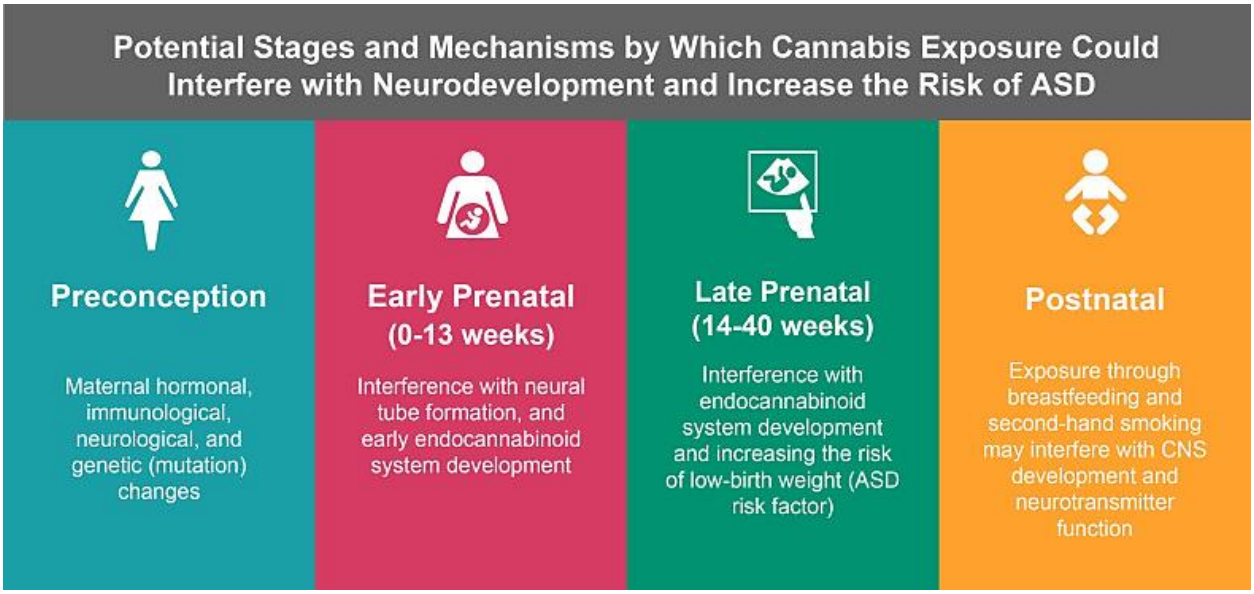


Figure 4: Prenatal Cannabis Exposure and ASD Risk Mechanisms. The figure illustrates the potential phases and mechanisms of cannabis-induced neurodevelopmental disruption and enhanced ASD susceptibility. These stages consist of preconception, early prenatal (0-13 weeks), late prenatal (14-40 weeks) and postnatal stages. The blocks describe how cannabis use during each of these stages may lead to neurodevelopmental disorders in the offspring. The image was adapted from Sajdey *et al.,* (2021).

Autism is a disorder that affects neural development, leading to chronic deficiencies in social interaction and communication – often associated with the prevalence of repetitive and limited behavioural patterns, interests and activities (American Psychiatric Association., 2013). Autism does not have a singular identifiable cause but is rather believed to result from the interplay of various genetic, biological, and environmental factors (Hyman *et al.,* 2020). These factors, which encompass hereditary predispositions, sibling history of ASD, specific genetic or chromosomal conditions, prenatal exposure to certain substances, environmental factors like air pollution and maternal health conditions such as obesity, diabetes, or immune disorders, along with other circumstances like extreme prematurity and oxygen deprivation during birth, collectively contribute to potential abnormalities in brain structure and function (Hyman *et al.,* 2020).

Exome sequencing studies on ASD have identified high-confidence ASD genes through the analysis of mutations in affected individuals. Initial research by Willsey *et al.*, (2013) pinpointed nine genes with recurrent de novo loss-of-function mutations, termed the “high-confidence Willsey subset.” Subsequent large-scale studies by Iossifov *et al.,* (2014) and De Rubeis *et al.,* (2014) expanded these findings, collectively identifying 35 high-confidence ASD genes. These genes are associated with embryonic expression and play roles in synapse formation, transcriptional

regulation, and chromatin remodelling. Functional characterisation reveals connections to cortical development, particularly deep-layer projection neurons (Iossifov *et al.,* 2014). Concluding this investigation into high-confidence ASD genes, TBR1, a T-box transcription factor (TF), seamlessly emerges as a recurring player in all three studies. Its intricate involvement in guiding the differentiation and identity of deep-layer projection neurons during neocortical development lays the groundwork for a more in-depth exploration of TBR1's functional dynamics in the context of ASD. (Hevner *et al.,* 2001; Leone *et al.,* 2008; Bedogni *et al.,* 2010; Han *et al.,* 2011)

1.5 Transcription Factors

At the core of molecular biology lies the concept of transcription factors, crucial players in the intricate regulation of DNA transcription and protein expression. The central dogma, a foundational principle in molecular biology (Crick., 1970), outlines the process by which genetic information, stored in the form of DNA or RNA, is converted into functional proteins through the intermediary molecule known as messenger RNA (mRNA). While the DNA in all cells of a multicellular organism is inherently identical, the specificity of tissue function is achieved through the orchestration of tissue-specific proteins, referred to as transcription factors. These transcription factors play a pivotal role in the intricate regulation of DNA transcription and, consequently, protein expression (Sonawane *et al.,* 2017). TFs enable gene expression by selectively binding to specific DNA sequences, known as promoter regions (Yang., 1998). This initial interaction represents the first step in a series of precise molecular contacts involving general TFs, RNA polymerase II, and the promoter region. Specifically, RNA polymerase II binds to a distinct segment within the promoter sequence, known as the TATA box, in conjunction with various general transcription factors, including TFIID, TFIIB, TFIIE, TFIIF, TFIIA, and TFIIH. Together, they assemble into a complex known as the pre-initiation complex (PIC) (Yang., 1998). The actual transcription process commences approximately 25 to 35 bases downstream from the TATA box, precisely at a site termed the transcription start site. This region also houses the initiator sequence, which not only dictates the location but also governs the rate of transcription (Clancy., 2008). General TFs are responsible for the expression of genes crucial for maintaining fundamental cellular functions (Clancy., 2008). In contrast, specific TFs exert control over the transcriptional activity of particular genes. These specific TFs respond to tissue-specific requirements and external stimuli by binding to enhancer or repressor regions within the gene. This finely tuned regulatory mechanism is known as differential gene expression, or facultative

gene expression, and it plays a crucial role in shaping the unique characteristics of different tissues and cellular responses (Yang., 1998).

T-box Transcription Factors

T-box proteins are specialised TFs that govern the differential expression of genes crucial for various developmental processes. The T-box class of TFs contain approximately 50 members, with 18 identified in mammals (Wilson and Conlon., 2002). T-box proteins, typically ranging from 50 to 70 kDa, feature a transactivation/repression domain and a structurally conserved DNA binding domain known as the T-box domain (Herrmann., 1994). T-box proteins are classified according to sequence similarity within the T-box domain, expression patterns and function between various vertebrates, including fish, frogs, dogs, and mice (Herrman *et al*., 1990; Pflugfelder *et al.,* 1992; Papaioannou., 2001; Smith., 1999; Papaioannou and Silver, 1998; Haworth *et al.,* 2001; Herrmann and Kispert, 1994). This has given rise to 5 distinct subfamilies, shown in Figure 5A (Wingender*.,* 2013). The TBR1 T-box domain is highly conserved amongst the T-brain 1 subfamily, with TBR2 and TBX21 demonstrating sequence similarities of 85.64% and 73.48%, respectively (Figure 5B).

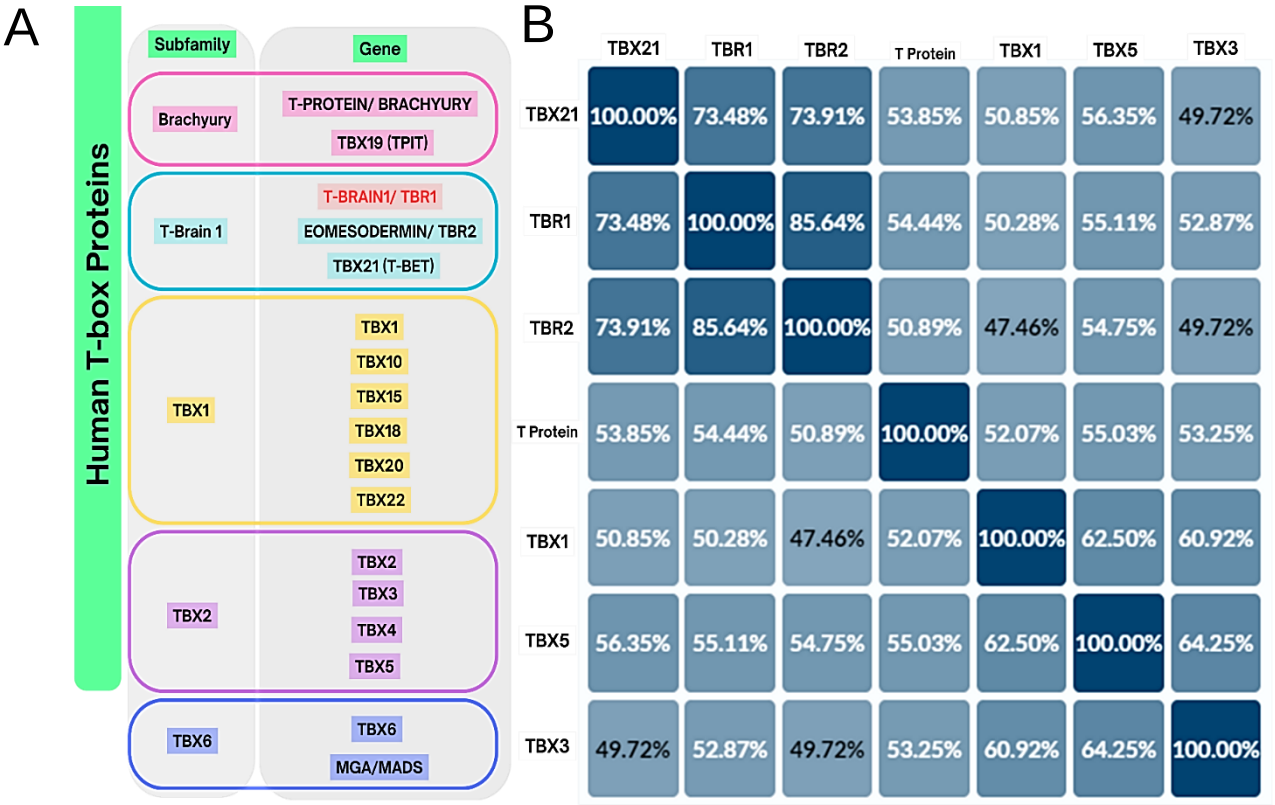


Figure 5: T-box Family Tree and Identity Matrix. A: The T-box protein superfamily in humans. The specific focus of this work, TBR1 (or T-Brain1), is highlighted in red and forms part of the T-Brain 1 subfamily. Other members of the Human T-box protein superfamily include Brachyury (in pink), TBX1 (in yellow), TBX2 (in light purple) and TBX6 (in dark purple). B: As TBR1 is a part of the T-Brain 1 subfamily, a percentage matrix of a multiple sequence alignment (UniProt Align) of the T-box domain is presented, showcasing its comparison with other members within this subfamily, such as TBR2 and TBX21. TBR1 is also contrasted with all other T-box proteins possessing crystallographically resolved structures, including TBXT (Brachyury), TBX1, TBX5, and TBX3, with all sequences sourced from Homo sapiens. TBR1 exhibits the highest similarity to TBR2 (EOMES) at 85.64%, followed by TBX21 at 73.48%, TBXT at 54.44%, TBX1 at 50.28%, TBX5 at 55.11%, and TBX3 at 52.87% (UniProt, *2023).*

The T-Brain subfamily, comprising TBR1, TBR2 (EOMES), and TBX21 (T-BET), is characterised by their expression in various brain regions and shared sequence similarities (Bulfone *et al.,* 1995). T-brain family members play roles in the formation of the brain, eye, immune system, mesoderm, and placenta. TBR2, for instance, contributes to both central nervous system and mesoderm development, along with influencing CD8+ T-cell differentiation (Lv *et al.,* 2019). On the other hand, TBX21 acts as a transcription factor regulating the immune system, specifically in the development of T helper 1 cells and exhibiting exclusive expression in olfactory bulb cells (Stolarczyk *et al.,* 2013). TBR1's crucial role in normal brain development involves the

differentiation and migration of neurons within specific regions of the cerebral cortex (Bulfone *et al.,* 1995). TBR1 expression peaks during embryogenesis and gradually declines postnatally (Bulfone *et al.,* 1995). Around time of birth, TBR1 expression is predominantly found in layer 6 of the cerebral cortex, with detectable levels in layers 2/3 and a subset of layer 5 neurons (Bulfone *et al.,* 1995; McKenna *et al.,* 2011). Here TBR1 ensures that corticospinal (CS) axons originate solely from layer 5 neurons by binding to the locus on FEZF2, thereby repressing *FEZF2* expression. This precise control contributes to organised neural circuitry and coordinated motor functions (Han *et al.,* 2011). Moreover, TBR1 regulates the expression of other genes, like RELN, GRIN2B, and AUTS2, which are closely associated with ASD and intellectual disability (Bedogni *et al.,* 2010).

1.5.1 TBR1 T-box Proteins

Twenty years ago, Dr. Rubenstein's laboratory first identified the significance of TBR1 in normal brain development, revealing a 10-fold higher expression of TBR1 in embryonic telencephalons compared to adult telencephalons (Bulfone *et al.,* 1995). Dr. Rubenstein and colleagues investigated the physiological function of TBR1 by studying Tbr1−/− mice, revealing that TBR1 is essential for survival, impacting neonatal lethality, neuronal migration patterns, and the development of projection neurons in the cerebral cortex, olfactory bulb, and amygdala (Bulfone *et al.,* 1998; Hevner *et al.,* 2001; Remedios *et al.,* 2007). TBR1 carries out its function by recognising and regulating the expression of genes containing a consensual DNA sequence, 5'-TCACACCT-3'. This conserved segment within the DNA binding sequence, denoting high functional relevance, remains consistent across the T-box protein family (Wilson and Conlon., 2002). Using the binding domain sequence, TBR1 target genes were identified by sequence analysis (Huang *et al.,* 2014). Due to shared DNA binding sequences within the T-box protein family, the method lacks specificity for TBR1, thus neuronal expression was selected as a primary criterion. This revealed that TBR1 regulates GRIN2B and RELN promoters, confirmed by functional assays, including electrophoretic mobility shift assay, chromatin immunoprecipitation, and luciferase reporter assay (Hevner *et al.*, 2001). Moreover, a literature search of the Simons Foundation Autism Research Initiative database reveals TBR1 regulates 23 genes associated with ASD. TBR1 could function as a central orchestrator, governing the expression of a cluster of ASD-associated genes, thereby exerting an influential role in neural development and function (Chuang *et al.,* 2014).

De novo mutations within the TBR1 gene are associated with moderate-to-severe developmental delays/intellectual disabilities (DD/ID), speech delays, autism, autistic traits, and aggression in affected individuals (McDermott *et al.,* 2018; Nambot *et al.,* 2020).This is due to alterations in transcription regulation, cellular distribution, and interactions with coregulatory factors such as CASK and FOXP2 (Deriziotis *et al.,* 2014). Likewise, mice studies have revealed that TBR1 haploinsufficiency results in abnormalities in axonal projections and deficits in various behaviours, including social interactions, ultrasonic vocalisation, associative memory, and cognitive flexibility (Huang *et al.,* 2014). Mouse models in which the 2q24 region, encompassing the TBR1 gene, has been conditionally deleted from cortical deep-layer glutamatergic neurons are also linked to intellectual disability and developmental delay in human patients (Traylor *et al.,* 2012). Thus, TBR1 emerges as a pivotal regulator in development (Herrmann., 1994). This regulatory process is intimately linked to the unique molecular architecture of TBR1, notably its DNA binding domain.

1.5.1.1 T-box DNA Binding Domain Structure

In general, transcription factors are classified according to the highly conserved structure of their DNA binding domain (Yang., 1998). The type of DNA binding domain includes helix-turn-helix (HTH), leucine zipper (LZ), zinc finger (ZF) and winged helix-turn-helix (W-HTH) (Yang., 1998) domains. The structure of the DNA binding domains contributes to the specificity of the TF interacting with a specific sequence of DNA (Wingender., 2013). One of the more unusual DNA binding domains is an immunoglobulin fold – which is found in the T-box domain. In the absence of a crystallographically solved structure, the AlphaFold predicted T-box domain of TBR1 is structurally aligned with TBX5 as a template (PDB: 2X6V), since it represents the closest family member with a crystallographically solved T-box domain (Figure 5B and 6). The crystal structure of human TBX5 T-box domain in DNA-bound and DNA-free form was solved in 2010 with a resolution of 2.20 Å (Stirnimann *et al.,* 2010).

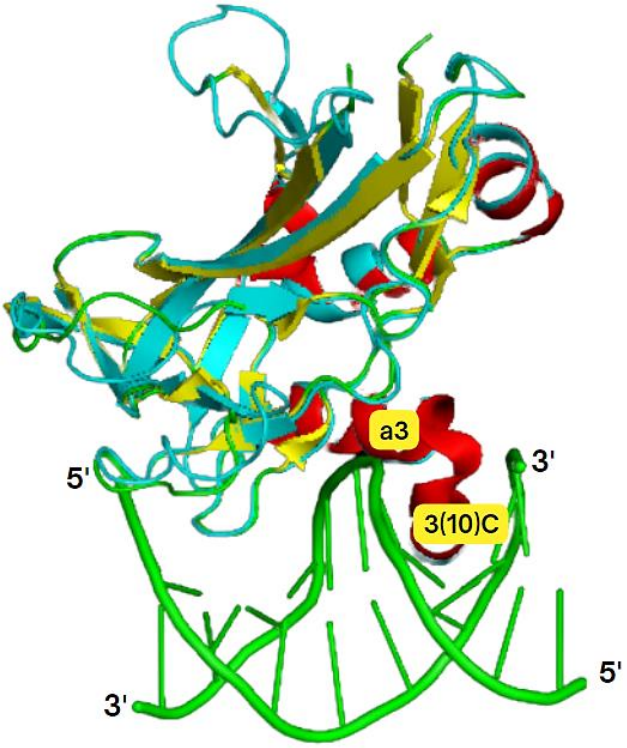


Figure 6: TBR1 T-box Domain Structure Comparison with TBX5. Structure alignment between the DNA binding domain of TBR1 in blue (UniProt ID: Q16650) and TBX5 in green (UniProt ID: Q99593) interacting with DNA (from the crystallised structure of TBX5 interacting with DNA). β-pleated sheets and β-barrels are shown as arrows. $\boldsymbol{\alpha}$-helix structures are shown in red. $\boldsymbol{\alpha}$3 and 3_{10}C helices (yellow) are both directly interacting with the DNA (green) and present in both TBR1 and TBX5. Image generated using PyMOL MatchAlign (PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC).

This immunoglobulin fold consists of a B-structured core, typically a B-barrel or B-sheet, with exposed coils and helices that are responsible for interacting with the minor and major grooves of DNA (Wingender., 2013). In figure 6, the T-box proteins exhibit a distinct 7-stranded B-barrel configuration, composed of two B-sheets arranged in a Greek key topology, providing a stable platform for DNA interaction (Bork *et al.,* 1994). In addition to the 7-stranded B-barrel, T-box proteins feature three additional beta sheets that form a lid-like structure on top of the B-barrel. This structural arrangement enhances their DNA binding capabilities. Within the T-box protein structure, four helices are present, each with specific functions. Helix 2, for instance, plays a crucial role in stabilising the protein's structure. Helix 1 is pressed against and nearly parallel to helix 3, further contributing to the overall stability of the protein. Helices 3 and 4, which lie perpendicular to each other, are responsible for interacting with the minor groove of DNA, facilitating the binding and regulation of target genes (Muller and Herrmann., 1997).

Unravelling the potential disruptions to the DNA binding ability of TBR1 poses a critical juncture in understanding the intricate regulatory mechanisms governing genes associated with ASD. This divergence underscores the urgent need for additional research to fill the gaps in current literature, specifically focusing on the impact of cannabis on the molecular mechanisms behind ASD development. Given these gaps, further exploration is essential. The intriguing possibility of CBD, with its potential to reach the foetal brain, interacting with high-confidence ASD-related proteins like TBR1, serves as a crucial starting point for assessing the potential connection between cannabis and ASD. This sets the groundwork for subsequent investigations and future explorations on whether CBD can disrupt the DNA binding ability of TBR1, which holds promise for shedding light on the nuanced influence of prenatal cannabis use on the etiology of ASD. It is crucial to note that ASD development is multifaceted, and this research does not assert a causal relationship between CBD and autism. Rather, it serves as a valuable contribution to informing pregnant women about the potential influences of these substances, emphasising the importance of informed decision-making and dispelling ignorance through education.

Study Rationale and Aims and Objectives

Study Rationale

In recent years, there has been a significant increase in the use of CBD among pregnant women (Wieckiewicz *et al.,* 2022). CBD is sought after for its potential to alleviate various pregnancy-related discomforts, such as pain, nausea and anxiety (Vanstone *et al.,* 2022). However, the rising use of CBD during pregnancy has raised serious concerns about its safety, particularly its effects on the developing foetal brain (Young-Wolff *et al*., 2019). Despite this growing trend, there is still a lack of information regarding the potential consequences of maternal CBD exposure on the foetus (Corsi *et al.,* 2020). The brain of a developing foetus is a complex system that requires precise genetic coordination for normal neurodevelopment. Disruptions in this process have been linked to various neurodevelopmental disorders, including ASD. One gene that is strongly associated with ASD is TBR1, due to its critical role in brain development (Bulfone *et al.,* 1995). TBR1 plays a crucial role in regulating the differentiation and functioning of neurons in the cerebral cortex, making it an interesting target for ASD research (Bulfone *et al.,* 1995). Dysfunctions in TBR1 have been linked to ASD pathogenesis (McDermott *et al*., 2018; Nambot *et al.,* 2020).

The purpose of this study is not to explore the complexities of CBD's effects on the foetal brain, but rather to take an essential first step in investigating whether CBD has the potential to bind to the TBR1 T-Box Domain, which is a fundamental protein in neurodevelopment. Ignorance about these potential risks cannot persist and even incremental research in this field is essential to foster growth, awareness and informed decision-making for expecting mothers. In this study, a combination of *in vitro* and *in-silico* methodologies is explored to determine the potential interaction between CBD and the TBR1 T-box domain. Predictions regarding the interaction are made by optimising the structures of the TBR1 T-box domain and CBD and simulating their interaction using Induced Fit Docking and Molecular Dynamics via Maestro Schrödinger. To confirm the interaction using *in vitro* methods, the TBR1 T-box domain is transformed, expressed and purified, and the potential binding, thermal stability and physiochemical properties are assessed using a pull-down assay, thermal shift assay and time-resolved Fourier-transformed infrared spectroscopy.

Study Significance:

Maternal CBD Exposure: As the use of CBD becomes more prevalent during pregnancy, understanding its consequences is imperative. The study addresses a crucial aspect of foetal brain development by focusing on the interaction with TBR1, a key neurodevelopmental regulator. Therefore, the study aims to provide essential insights for public health initiatives, preventive measures and targeted interventions to promote healthier pregnancies and reduce the incidence of ASD in the population.

Link to Neurodevelopmental Disorders: Contrasting research has raised questions about the potential link between prenatal cannabis use and neurodevelopmental disorders in offspring. Further research is needed to address this gap in the literature.

Foundational Exploration: The main goal of this study is to investigate the potential interaction between CBD and the TBR1 T-Box Domain. Specifically, the T-box domain will be assessed in isolation to determine whether CBD can directly interact with the DNA binding domain and will provide a basis for further, more comprehensive studies exploring the functional implications of this interaction.

Future Research: By confirming this interaction, future research can study the mechanisms, consequences and clinical relevance of CBD's relationship with ASD. This insight can guide future interventions and therapeutic strategies.

Aim:

In vitro and *in silico* identification and characterisation of the interaction between CBD (cannabidiol) and TBR1 T-box Domain

Objectives:

Objective 1: Assessing the Potential Interaction between TBR1 T-box Domain and CBD using Bioinformatics

- **Structural Analysis:** Compare the AlphaFold predicted structure of TBR1 T-box domain with the crystal structure of TBX5 T-box domain (PDB: 2X6V) through sequence and structural alignments using PyMOL. Through this, the DNA interacting residues and structures can be predicted.
- **Predicting Binding Domains**: Determine the location and volume of potential druggable binding sites within TBR1 T-box domain, using CASTp3.0, ProteinPlus and SiteMap Maestro Schrödinger.
- **CBD ADMET Properties**: Investigate the chemical structure of CBD and analyse its absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties to investigate the drug-likeness of CBD and its potential to interact with TBR1 T-box domain, using QikProp Maestro Schrödinger.
- **Structure Preparation**: Use Schrödinger's Maestro to prepare energy-minimised and optimised 3D structure of CBD and TBR1 T-box domain.
- **Induced Fit Docking**: Perform Induced Fit Docking (IFD) between CBD and TBR1 T-box domain, to assess potential interaction sites and increase the accuracy of the binding poses, using Maestro Schrödinger.
- **Molecular Dynamics:** Perform molecular dynamics simulations between TBR1 T-Box Domain and CBD to determine the dynamics and stability of the potential interaction and provide insights into structural conformation (RMSD), type and distribution of bonds, local variations (RMSF), overall compactness (rGyr) and change in secondary structure upon ligand binding, using Maestro Schrödinger.

Objective 2: Assessing the Potential Interaction between TBR1 T-box Domain and CBD using Experimental Approaches

- **Transformation**: Transform the T7 *E.coli* cells with pET-11a plasmid containing the codon-optimised sequence of TBR1 T-box domain
- **Expression**: Express the codon-optimised TBR1 T-box domain in T7 *E.coli* cells
- **Purification and Protein Purity**: Purify the overexpressed TBR1 T-box domain using Immobilised Metal Affinity Chromatography (IMAC), followed by Size Exclusion Chromatography (SEC) and assess the purity of TBR1 T-box domain using SDS-PAGE.
- **Circular Dichroism**: Investigate the secondary structure and folding status of TBR1 T-box domain using Far UV Circular Dichroism (CD) spectroscopy.
- **Pull-Down Assay**: Assess the potential binding between CBD and the TBR1 T-box domain through an Epoxy-Activated Agarose Pull-Down Assay.
- **Thermal Shift Assay**: Assess the impact of increasing concentrations of CBD on the thermal stability of the TBR1 T-box domain via a Thermal Shift Assay (TSA) using quantitative polymerase chain reaction (qPCR).
- **Fourier Transformed Infrared Spectroscopy:** Evaluate the physiochemical properties of the TBR1 T-box domain, CBD and TBR1 T-box domain-CBD complex using Time-resolved Fourier Transform Infrared Spectra (TR-FTIR).

2 CHAPTER 2: LITERATURE REVIEW

2.1 History of Cannabis

The earliest archaeological and historical evidence of Cannabis use dates back 4000 B.C., where the cannabis plant was cultivated in China. The stems of cannabis were used for fibres, which the Chinese used to manufacture various pieces of equipment. Specifically, textiles and papers made from these fibres were found in the tomb of Emperor Wu – dating back 104-87 B.C. (Li., 1974). However, the earliest mention of the medicinal potential of cannabis dates back as early as 2737 B.C. where it was reported in the Emperor Shén Nong pharmacopoeia– where he documented prescribing cannabis for the treatment of malaria, constipation, poor memory, and others (Touwn., 1981). The emperor was so impressed with the medicinal benefits the plant had to offer, that he considered it to be one of the Superior Elixirs of Immortality. Interestingly, Hua T'o – the founder of Chinese surgery – invented a mixture of cannabis and wine as an anaesthetic for patients (Li., 1974). He became well-known for his ability to perform complex and dangerous surgeries – including incisions into the body cavities – by the help of a plant-based aesthetic prepared using cannabis resin.

2.1.1 Cannabis in South Africa

The earliest written reference to cannabis in South Africa, was documented by Jan van Riebeeck – where he made notes about a certain herb, named 'dagga' (Crampton., 2015). However, by that time, the cultivation and use of cannabis had already been widespread in South Africa for centuries (Theron., 1974). 'Dagga' is the colloquial term used to describe cannabis in South Africa – as written by the Afrikaans people – and possibly originated from the word 'dachah' by the Khoikhoi (Theron., 1974). The Khoikhoi, and their use of dagga, is also explicitly documented in the journal of Jan van Riebeeck and it is believed that the Khoikhoi were the main cultivators in South Africa and it served as a previous trading item amongst the indigenous groups (Crampton., 2015).

In 1922, the first nationwide legislation was enacted, prohibiting the use, possession, import, export, transport and trade of cannabis, along with other 'habit-forming drugs' like morphine, cocaine, heroin and opium (Customs and Excise Duties Amendment Act 35 of 1922). Subsequently, cannabis cultivation in South Africa was prohibited in 1928 under the Medical, Dental, and Pharmacy Act, No. 13 (Medical, Dental, and Pharmacy Act 13 of 1928). It wasn't until

2000, that challenges were made to sections 4b of the Drugs and Drug Trafficking Act 140 of 1992 (Drugs Act) and 22A of the Medicines and The Medicines Act. This legal action led to the decriminalisation of the private use and cultivation of cannabis by adults (Minister of Justice V Prince, 2018). The Medicines Act divides pharmaceuticals and therapeutic substances into eight schedules, with Schedule 8 encompassing the most hazardous compounds with the highest restricted access. THC and cannabis (the complete plant or its components or derivatives) are Schedule 7 drugs. Authorities classify CBD as a Schedule 4 substance, except in specific instances where CBD may receive categorisation as a Schedule 0 compound (Unscheduled CBD Products). Schedule 4 drugs, like Schedule 6 substances, are only available with a prescription from an authorised doctor. Schedule 0 substances are available for purchase at clinics, shops, health stores, and convenience stores (The Medicines Act). Since the rescheduling of CBD, its market has experienced exponential growth, reaching $12.8 billion in 2021 and projected to achieve a 21.7% CAGR, reaching $56.3 billion by 2028 (Moltke and Hindocha., 2021). This increase is driven by heightened awareness of CBD's medical applications, resulting in an uptake in retail sales. As we delve into the intricate world of CBD, it becomes evident that comprehending its effects on the human body is now crucial. This nuanced understanding is imperative for exploring the full spectrum of potential benefits and side effects associated with CBD.

2.2 Cannabis in the Human Body

Cannabis has been used for medicinal and recreational purposes for centuries. The human body contains an ECS that regulates various physiological functions, including pain, appetite, and mood. Depending on the route of administration, cannabis consumption can lead to different pharmacokinetic and pharmacodynamic effects (Chayasirisobhon., 2021). Oral consumption is one of the most commonly used methods of administering cannabinoids, with options like edibles, tinctures, capsules, and oils (Grotenhermen and Russo., 2002) available. However, it's important to note that the absorption of cannabinoids in the body is slower when ingested orally, resulting in a more gradual onset and lower peak concentrations compared to other routes of administration (Law *et al.,* 1984). The effects of consuming cannabis orally tend to last longer than other methods, making it a suitable option for sustained symptom relief (Grotenhermen and Müller-Vahl., 2012). However, the slow onset time and variable absorption in the gastrointestinal tract may result in unpredictable effects and dosing errors, which can be a safety concern, especially for inexperienced users (MacCallum and Russo., 2018). About 80% to 90% of the cannabis

metabolites are excreted within five days, primarily in the form of hydroxylated and carboxylated metabolites (Lemberger *et al*., 1970). The pharmacokinetics of CBD are intricate, and when administered orally, its bioavailability is generally low across different species (Smith-Kielland *et al.,* 1999). The oral bioavailability of CBD is approximately 6% in humans (Smith-Kielland *et al.,* 1999) and it possesses a notable affinity for lipid substances, facilitating its rapid distribution among the brain, adipose tissue, and various organs (Devinsky *et al.,* 2014; Harvey., 1991).

CBD has been shown to have minimal affinity for CB1 and CB2 receptors and instead appears to exert its effects through other mechanisms (McPartland *et al.,* 2015). It also does not stimulate the firing of dopamine neurons in the VTA (French *et al.,* 1997), and lacks inherent rewarding properties (Katsidoni *et al.,* 2013; Parker *et al.,* 2004). Studies have shown that CBD can indirectly interact with the CB1 receptor as an agonist or antagonist (Wolf *et al.,* 2010). In fact, the study suggests that CBD may act as an indirect agonist of the CB1 receptor, as some in vivo effects of CBD are not observed in CB1 receptor knockout mice (Wolf *et al.,* 2010). CBD has pharmacological activity on various receptors besides the ECS system. These receptors have an affinity for various cannabinoid ligands and are considered an extended ECS system (McPartland et al. 2015). Two examples of these receptors are the orphan G protein-coupled receptor (GPCR) GPR55 and transient receptor potential (TRP) channels. GPR55 is a potential novel metabotropic cannabinoid receptor (Ross., 2009). TRP channels are ionotropic cannabinoid receptors (Di Marzo *et al*., 2002). CBD has been shown to interact with other receptors and signalling pathways in the body, such as the serotonin 5-HT1A receptor, to produce therapeutic effects without the psychoactive effects of THC (Russo., 2011). Additionally, CBD has been shown to inhibit the breakdown of anandamide, an endocannabinoid that can activate CB1 and CB2 receptors. This results in increased levels of anandamide in the body, which can lead to a variety of therapeutic effects such as pain relief and reduction in inflammation. Understanding the intricacies of CBD distribution provides valuable insights into its ability to modulate various physiological pathways, paving the way for targeted therapeutic applications across a spectrum of medical conditions.

2.3 Therapeutic Potential of Cannabis

Formerly, the constraints imposed by legal regulations made it challenging to conduct regulated clinical research on the impact of cannabis consumption on human behaviour. Medicinal cannabis, describing the physician-prescribed use of cannabis, involves the ingestion of cannabinoids, believed to confer various health benefits (Rahn and Hofmann., 2009). This herbal

medicine, deemed safe, self-administered and cost-effective compared to traditional western pharmaceuticals, has garnered increased interest for its potential medicinal advantages (Rahn and Hofmann., 2009). Despite this interest, the clinical utility of cannabis lacks robust research-based evidence and much of the existing knowledge relies on subjective accounts from recreational users. This shortage of evidence complicates physicians' ability to prescribe the herbal medicine, given the limited understanding of its physiological effects, potential side effects and effects at different life stages (Rahn and Hofmann., 2009).

2.3.1 Cannabis in the Treatment of Diseases

In 2017, the National Academy of Sciences, Engineering, and Medicine (NASEM, Washing DC) published a scientific review paper examining the use of medicinal cannabis for various medical conditions (NASEM., 2017). The comprehensive review assessed both the safety and efficacy of cannabis and related products in treating medical conditions. The evaluation considered the frequency of usage and the postulated success of the plant in treating specific diseases or disorders. Among the medical conditions studied, epilepsy, chronic non-cancer pain and multiple sclerosis were ranked highest based on the quality of evidence related to the efficacy, effectiveness and safety outcomes of medicinal cannabis (NASEM., 2017).

A study by Jugl *et al.,* (2021) expanded on the study conducted by NASEM in 2017, to include more recent research on the use of medicinal cannabis on approved conditions in the US from 2016 to 2019. Despite this expansion, several literature gaps were identified, with one of the least researched medical conditions being autism. During the NASEM literature review, only 17 eligible publications focused on the relationship between cannabis use and autism (Jugl *et al.,* 2021). In a randomised controlled trial aimed at addressing behavioural symptoms in children diagnosed with ASD, the application of CBD emerged as a promising intervention. The study's outcomes revealed that CBD was well-tolerated and yielded marked enhancements in anxiety management and communication skills among the participating children (Barchel *et al.,* 2019). In a separate investigation, researchers observed reduced plasma levels of anandamide, a vital endocannabinoid, in ASD-afflicted children compared to their typically developing counterparts. These findings suggest an underlying dysregulation within the ECS in the context of ASD (Karhson *et al.,* 2018). Scant scientific literature is available concerning the potential correlation between cannabis use and ASD. Previous studies have primarily focused on finding strategies to reduce ASD-related symptoms, however, there is evidence that suggest a connection between prenatal cannabis use and the emergence of ASD in the offspring (Corsi *et al.,* 2020; Wood *et al.,*

2015). Unfortunately, the available literature on this topic is inconsistent, with other studies showing no significant link between prenatal cannabis use and ASD-related behaviour in children (DiGuiseppi *et al.,* 2021; Nutor *et al.,* 2023). Consequently, the significance of the current investigation lies in its endeavour to address this.

As previously mentioned, there is a mounting body of evidence suggesting that cannabinoids, including CBD, could traverse the placental barrier and interact with the developing foetal brain during pregnancy. The reported prevalence of cannabis use among pregnant women varies from 2% to 36% (Brown *et al.,* 2017; Corsi *et al.,* 2019; Ko *et al.,* 2018). In the US alone, prenatal cannabis use has increased from 3.4% in 2002 to 7.0 % in 2017 (Young-Wolff *et al.,* 2021). Investigating the potential link between prenatal cannabis use and ASD is crucial not only to enhance our comprehension of associated risk factors but also to address the perceived safety of CBD during pregnancy. With emerging evidence suggesting connections between prenatal cannabis use and neurodevelopmental disorders in offspring (Committee Opinion., 2017), it is imperative to move beyond existing notions.

Children exposed to cannabis in utero may exhibit cognitive and behavioural deficits. Longitudinal studies have reported that these children may have lower scores on measures of cognitive function, increased impulsivity, and higher rates of behavioural problems (Goldschmidt *et al.,* 2012). These findings suggest that prenatal cannabis exposure can have lasting effects on a child's development and behaviour. Prenatal cannabis use has also been linked to adverse birth outcomes, including low birth weight, preterm birth and small-for-gestational-age infants (Gunn *et al.,* 2016). These outcomes are concerning as they can lead to further health complications and developmental challenges. Understanding the potential risks associated with prenatal cannabis use is crucial for informing public health policies and guidelines. As cannabis legalisation spreads, it is imperative to educate pregnant women about the possible impacts of cannabis on foetal development and encourage them to avoid its use during pregnancy.

2.3.1.1 In Silico Methods and Their Benefits in Biomedical Research

In addition to empirical research, *in silico* methods have become invaluable in the study of molecular interactions and drug discovery. These computational techniques allow for the prediction and analysis of biological systems through simulations and modelling. Molecular docking, a common *in silico* method, predicts how small molecules interact with target proteins at

the atomic level. This technique helps to identify binding sites and optimise binding affinities, which are critical steps in the drug development pipeline (Meng *et al.,* 2011).

Molecular dynamics (MD) simulations provide dynamic insights into the stability and conformational changes of protein-ligand complexes over time. This is particularly useful for understanding the behaviour of molecular interactions in a realistic biological context (Hollingsworth and Dror., 2018). Moreover, ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) predictions can be made, which are crucial for evaluating the pharmacokinetic and toxicological properties of compounds early in the research process. This helps in assessing the drug-likeness of compounds and predicting potential side effects, thereby reducing the likelihood of late-stage failures in drug development (Daina *et al.,* 2017).

3 CHAPTER 3: METHODOLOGY

Objective 1: Assess the Potential Interaction between TBR1 T-box Domain and CBD using Bioinformatics

With the increasing use of CBD among pregnant women, understanding its potential impact on foetal development has become imperative. Previous studies have associated prenatal cannabis use with the development of ASD in offspring, wherein TBR1 dysfunction is also implicated (Corsi *et al.,* 2020). Given this context, the study aims to assess whether CBD binds to TBR1 T-box domain, laying the groundwork for future investigations into the effect of this interaction on the DNA binding function of TBR1 T-box domain and as a result, possible implications in ASD. Using an *in silico* approach as a stepping stone, the AlphaFold predicted structure of the TBR1 T-box domain (UniProt ID: Q16650) was compared with the crystal structure of the TBX5 T-box domain (UniProt ID: Q99593) via PyMOL (PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC), aiming to predict DNA binding residues and structures. The identification of potential druggable binding sites within the TBR1 T-box domain, where CBD might bind, and the assessment of whether this coincides with the DNA binding residues, were achieved using a combination of CASTp3.0, ProteinPlus, and SiteMap. Furthermore, the assessment of the drug-like properties and the potential interaction of CBD with the TBR1 T-box domain was assessed based of an analysis on the ADMET properties of CBD (CID:644019) via QikProp in Maestro Schrödinger. This was succeeded by induced fit docking (IFD) to unveil potential binding poses, residues and binding affinities. Finally, the dynamics and stability of the interaction was determined using molecular dynamics, by analysing the Root-Mean Square Deviation (RMSD), Root-Mean Square Fluctuations (RMSF), Radius of Gyration (rGyr) and the secondary structure of both the APO and CBD-bound states of the TBR1 T-box domain. The bioinformatics investigation preceded the experimental work with the intent of determining the potential binding capacity of CBD and evaluating the feasibility of this interaction. The aim was to prompt a thorough exploration in the experimental domain, validating the insights obtained from the preceding bioinformatic analysis.

3.1 Structural Analysis

AlphaFold, a deep learning system developed by DeepMind Technologies, uses a neural network trained on a large dataset of known protein structures, to accurately predict the 3D structure of proteins, based on their amino acid sequence (Anfinsen *et al.,* 1961; Jumper *et al.,* 2021). AlphaFold has exhibited exceptional accuracy in predicting protein structures, as evidenced by its outstanding performance in the recent CASP14 (Critical Assessment of Protein Structure Prediction) blind prediction experiment (Jumper *et al.,* 2021). Surpassing all other predictors, AlphaFold achieved remarkable accuracy with an impressive RMSD of 0.8 Å. In comparison, the next best method showed a 2.8 Å deviation from experimental backbone structures, indicating AlphaFold's proximity to experimental accuracy (Jumper *et al.,* 2021). Nevertheless, AlphaFold-predicted structures should not substitute for experimental validation of protein structures, particularly when it comes to interactions not accounted for in the predictions. Additionally, it is crucial to consider the confidence predictions when employing an AlphaFold-predicted model (Terwilliger *et al.,* 2024). These confidence levels are predicted by utilising a complex neural network that integrates details from both protein sequences and structures and employs Amber refinement as a final step (Tian *et al.,* 2018).

Therefore, in the absence of a crystallographically solved TBR1 T-box domain, the AlphaFold predicted structure of the full TBR1 structure (UniProt ID: Q16650) in *Homo sapiens* was obtained from UniProt and the DNA binding domain (T-box domain) was isolated using PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC. Throughout the study, the T-box domain was the chosen focus – rather than the entire protein. The T-box domain is crucial for the functionality of TBR1 and stands out as the sole region in the AlphaFold predicted model with an exceptionally high confidence level. Simultaneously, the T-box domain of the crystallographically solved TBX5 (PDB: 2X6V) within *Homo sapiens* was isolated and juxtaposed with the T-box domain of TBR1. TBR1 was compared to TBX5 for several reasons, including the shared characteristic of both being a T-box transcription factor and containing a T-box DNA binding domain. TBX5 is the closest T-box family member to TBR1 that has a crystallographically solved structure in humans and the structure is solved in both the DNA-bound and DNA-free forms.

Specifically, a sequence alignment was performed using Clustal Omega (EMBL-EBI) to determine the degree of conservation between the structures. Thereafter, a structure alignment was conducted using PyMOL MatchAlign and the alignment was evaluated based on the RMSD values obtained. The combination of sequence and structure alignment was used to determine

corresponding DNA interacting residues and structures within TBR1 T-box domain, to assess whether CBD may interact with these regions. Therefore, potential binding pockets within the TBR1 T-box domain were identified to assess potential binding sites for CBD before initiating docking simulations.

3.2 Predicting Binding Domains

Binding sites for small ligands are commonly situated within cavities or openings on protein surfaces, serving as key points for interface recognition and promoting strong interactions (Sotriffer and Klebe., 2002). As a result, it is common to detect potential ligand binding sites based on specific geometric and energetic characteristics in protein surfaces. Computed Atlas of Surface Topography of proteins (CASTp3.0) is a cutting-edge tool that predicts binding pockets on both the surface and interior of protein structure. Its predictions are based on spatial geometric measurements, including surface area and volume (Tian *et al.,* 2018), which are calculated using three different techniques: Delaunay Triangulation, Alpha Shape, and Discrete Flow. The Delaunay Triangulation method forms triangles connecting points in space, creating a network that captures the overall structure of the object. Alpha shapes are mathematical constructs defining object boundaries and edges, considering the distance between points for connectivity. This helps represent the outer surface and internal features of the object. The Discrete Flow technique analyses the movement of a "flow" through the structure, identifying pockets, cavities, and precise boundaries within the object (Edelsbrunner and Mucke., 1994; Edelsbrunner., 1995; Edelsbrunner and Shah., 1996; Edelsbrunner *et al*., 1998). However, predicting binding domains within a protein involves considering other critical factors as well, such as hydrophobicity and shape complementarity. Due to this consideration, the binding domain within the TBR1 T-box domain was also predicted using an alternative prediction tool - Protein Plus, and the results were cross-referenced with those obtained from CASTp3.0. Protein Plus precalculates binding pockets by utilising Protoss and DoGSiteScorer. Protoss is a hydrogen prediction tool which determines an optimal hydrogen bonding network for molecular interactions (Lippert and Rarey., 2009; Bietz *et al.,* 2014). DoGSiteScorer detects potential binding pockets by using a Difference of Gaussian filter on the entire protein structure, which divides the structure into sub-pockets, based on size, shape and chemical properties (Volkamer *et al.,* 2010; Volkamer *et al.,* 2012). Therefore, the binding pocket location, size, volume and hydrophobicity was determined and compared based on the results obtained from both CASTp3.0 and Protein Plus. To address any differences

observed between Protein Plus and CASTp3.0, another prediction tool – SiteMap – was used to not only confirm the location of predicting binding domains, but also provide a more in-depth exploration into the characteristics of each binding domain. SiteMap (version 3.5, Schrödinger, LLC) identifies binding ‘sites’ suitable for ligand binding by strategically placing ‘site points’ on and around the proteins surface. Secondly, site maps are generated showing the presence of hydrophobic and hydrophilic regions. As a result, SiteMap ranks all the potential binding pockets according to a scoring function – SiteScore - which eliminates any pharmaceutically irrelevant sites. The SiteScore, which considers site size, exposure to solvent and hydrophilicity, aims for an average score of 1.0, with a score above 1 indicating promise (SiteMap, Schrödinger, LLC). Notably, a SiteScore of 0.80 accurately distinguishes between drug-binding and non-drug-binding sites. In addition to calculating the volume and size of the pocket, the Druggability Score (Dscore), using similar properties with different coefficients and no hydrophilic score cap, aids in discerning ‘difficult’ and ‘undruggable’ targets from "druggable" ones.

Therefore, to determine the potential binding pockets on TBR1 T-box domain, the AlphaFold predicted T-box domain was imported into the CASTp3.0 web interface as a PDB file. Following a detailed analysis, CASTp3.0 generated a visual representation of the predicted binding pockets, along with information on their area and volume. Similarly, the PDB file of the T-box domain was analysed on the Protein Plus web interface. The results were presented as a purple mesh that indicated the predicted binding pockets, along with a comprehensive results table containing key properties such as volume, area and hydrophobicity. Lastly, the isolated T-box domain was opened in Maestro Schrödinger and the ‘Binding Site Detection’ function was selected. The top-ranking potential receptor binding sites were pinpointed for all atoms within the workspace, using Maestro's default settings. The process involved the identification of a maximum of five sites while ensuring a minimum of 15 site points per reported site for accuracy. SiteMap generated a table presenting SiteScore, Dscore, volume and size of all the predicted binding pockets. The potential binding domains were predicted with the aim of determining the optimal location for generating the grid box in IFD. Moreover, this served as a preliminary step to determine if CBD may bind in the regions directly implicated in DNA binding, potentially influencing its functionality.

Schrödinger Maestro User Interface

Schrödinger's Maestro (Maestro, Schrödinger, LLC) software suite represents a pivotal tool in the realm of molecular modelling and computational biology, offering an extensive array of capabilities for researchers across diverse scientific disciplines. At its core, Maestro serves as an intuitive and versatile graphical interface, providing access to a plethora of potent molecular modelling techniques and simulations. It plays a pivotal role in the exploration of molecular interactions, prediction of binding affinities, and the investigation of structural dynamics within biomolecules.

Within the Maestro interface, several major elements contribute to its utility and functionality. The **Workspace** itself constitutes the primary arena for molecular modelling and simulation activities. Within this workspace, users can load molecular structures, engage in molecular docking studies, conduct molecular dynamics simulations, and visualise results. Maestro's visual and interactive environment empowers researchers to explore 3D molecular structures, investigate binding interactions, and simulate the dynamic behaviours of biomolecules. The **Workspace Navigation Pane** assumes a central role in project management and organization, offering researchers a hierarchical view of their projects and components, enhancing accessibility and the systematic arrangement of research endeavours. Complementing this is the **Workspace Configuration Toolbar**, prominently situated at the interface's summit. This dynamic toolbar provides swift access to essential functions and options, allowing users to tailor their workspace configurations with ease, thereby enhancing efficiency and streamlining research workflows.

Schrodinger's Maestro software suite finds applications spanning a broad spectrum of scientific domains, including drug discovery, materials science, and structural biology. Researchers harness Maestro's capabilities to predict protein-ligand interactions, simulate the dynamic behaviour of biomolecules over time, perform structural analyses, and scrutinize sequence data. Its versatility and user-friendly interface make it an indispensable tool for those seeking a deeper understanding of molecular interactions and their implications in various scientific investigations (Maestro, Schrödinger, LLC).

3.3 CBD ADMET Properties

Around 40% of potential drug candidates face setbacks during clinical trials due to inadequate Absorption, Distribution, Metabolism, and Excretion (ADMET) properties. These late-stage disappointments contribute significantly to the increasing costs of novel drug development. Identifying problematic candidates early on can help mitigate the wastage of time and valuable resources, streamlining the entire drug development process. QikProp is equipped to forecast an extensive range of pharmaceutical properties vital for decision-making in drug development (QikProp, version 4.4, Schrödinger). QikProp's algorithms are based on established physicochemical relationships and quantitative structure-activity relationships (QSAR) models. These models analyse a compound's molecular structure to predict its behaviour in biological systems. The software assesses properties such as aqueous solubility, blood-brain barrier permeation, overall CNS activity, permeability in Caco-2 and MDCK cell lines and binding to human serum albumin (QikProp., 2023). QikProp not only predicts pharmaceutically relevant properties of an organic molecule but provides an optimal range for these descriptors based off 95% of known drugs. This comprehensive analysis equips researchers to make well-informed judgments concerning a molecule's suitability for further research and development (QikProp., 2023).

Therefore, the 3D conformation of CBD was procured from PubChem (CID:644019) and retrieved as an SDF file. Subsequently, the structure was imported into the Maestro Schrödinger software suite and subjected to a computational assessment utilising QikProp. The ADMET properties were exported as a .csv file and analysed on MS Excel, which included the #stars, CNS activity, volume of CBD, hydrophobicity (QPlogPC16), lipophilicity (QPlogPo/w), intestinal absorption (QPPCaco), blood brain barrier (BBB) permeability (QPlogBB and QPPMDCK), blood transportability (QPlogKhsa) and potential for oral availability (RuleOfFive and RuleOfThree). By analysing the ADMET properties of CBD, a detailed assessment was conducted to understand its potential to enter the foetal bloodstream. Irrespective of the developmental stage of the foetal brain, the ADMET properties also provide insights into CBD's capacity to traverse the BBB and reach the developing brain. This analysis may offer valuable insights into the safety of CBD during pregnancy, irrespective of the foetal developmental stage.

3.4.1 Protein Preparation: TBR1 T-box domain

Protein structures sourced from databases like the Protein Data Bank (PDB), or predicted by tools like AlphaFold, often necessitate preprocessing to ensure accuracy for molecular modelling. The Protein Preparation Wizard (Protein Preparation Wizard, Epik, Schrödinger, LCC, 2023) in Maestro addresses this need by systematically refining these structures. It corrects common discrepancies such as co-crystallised ligands, metal ions, water molecules, multimeric states, missing atom connectivity and absent side chains (Sastry *et al.,* 2013). Even for AlphaFold-predicted structures, the wizard plays a crucial role by optimising the hydrogen bonding network, rectifying missing atoms and adjusting protonation states to reflect specific pH conditions (Skolnick *et al.,* 2021). This preprocessing includes assigning bond orders, adding hydrogens, repairing structural gaps, and performing energy minimisation to relieve steric clashes, ensuring the protein model is in a physically plausible and energetically favourable conformation (Sastry *et al.,* 2013). By enhancing structural integrity and accuracy, the Protein Preparation Wizard significantly bolsters the reliability of the protein models for subsequent computational analysis, such as molecular docking or dynamics simulations.

Subsequently, the structure was imported into Maestro Schrödinger to be prepared for docking simulations using the Protein Preparation Wizard. The Pre-processing window included three tabs, namely Import and Process, Review and Modify and Refine. In the Import and Process window, the default settings were selected and the main structural issues, including to fill in missing side chains using Prime, were selected (Protein Preparation Wizard, Epik, Schrödinger, LCC; Prime, Schrödinger, LCC). Prime is a computational tool that utilises a combination of homology modelling, structure refinement and binding energy estimation (Prime, Schrödinger, LCC). By incorporating information from known protein structures, Prime is capable of accurately predicting the most probable conformation of the missing side chains, thereby improving the overall accuracy and reliability of the protein structure. Following pre-processing in the Review and Modify tab, a thorough review of the structure revealed that no modifications were necessary, indicating that the AlphaFold predicted structure exhibited no missing or overlapping atoms. This coherence aligns with expectations, given the anticipated high quality of the predicted structure. The accuracy of AlphaFold predictions is credited to its advanced neural network architecture and the comprehensive training it undergoes, reinforcing the reliability of the generated protein structures. This led to the final step of refinement, in the Refine tab, using the OPLS_2005 force

field to optimise the energetics and stabilise the overall structure. Subsequently, structure optimisation and minimisation procedures were performed to further optimise the three-dimensional conformation of the TBR1 T-box domain. Minimisation includes adjusting any strain or tension in the protein's backbone and side chain atoms, including its length and angles. Moreover, it assessed any steric clashes, missing atoms and peptide/side chain planarity. These steps were carried out prior to docking simulations to ensure a more accurate representation of the protein's active site and improve the reliability of subsequent molecular interactions analysis.

3.4.2 Ligand Preparation: CBD (Cannabidiol)

LigPrep, a Schrödinger software application, is integral to the process of preparing low energy ligand structures for molecular modelling and simulations. Central to its functionality is the conversion of 2D structures into 3D models with precise stereochemistry, thereby ensuring computational accuracy in ligand behaviour analysis (Johnston *et al.,* 2023). This software corrects bonding pattern errors and rectifies ligand inconsistencies, such as inappropriate bond orders and missing hydrogens, based on Lewis structures (LigPrep, Schrödinger, LLC).

The preparation of the CBD structure in Maestro involved using the LigPrep tool, accessible from the Task bar. After importing the CBD structure, LigPrep was employed with the OPLS_2005 (Optimised Potentials for Liquid Simulations) force field, renowned for its efficacy in simulating organic molecules in biologically relevant environments (Kaminski and Friesner., 2001). The focus was on generating various states of CBD at a target physiological pH of 7 ± 2. This was achieved using the Epik module (version 2.3, Schrödinger, LLC), which accurately predicts pKa values, thereby elucidating the molecule's behaviour in different environmental conditions (Shelley *et al.,* 2007). This prediction is critical for determining the most likely form of CBD under physiological conditions. LigPrep's capabilities extend to exploring multiple protonation and tautomeric forms, different ring conformations and generating diverse stereoisomers. It also performs a desalting process to eliminate non-ligand entities such as water molecules and counter ions (Johnston *et al.,* 2023). To optimise computational resources, the system was configured to produce five conformers per ligand, focusing on the top-scoring conformer for subsequent analyses. These conformers underwent energy minimisation under the OPLS_2005 force field, producing energetically favourable conformations, crucial for accurate and reliable downstream computational studies (LigPrep, Schrödinger, LLC).

3.5 Molecular Docking

Molecular docking is a key computational technique used to predict the preferred orientation of one molecule when bound to another, typically a small molecule ligand docking into a larger protein receptor. This process aids in understanding the molecular interactions and binding affinity between these two entities, which is critical in drug design and discovery. The principle behind molecular docking involves two main approaches: rigid docking and IFD (Figure 7) (Koshland., 1995; Morrison *et al.,* 2006).

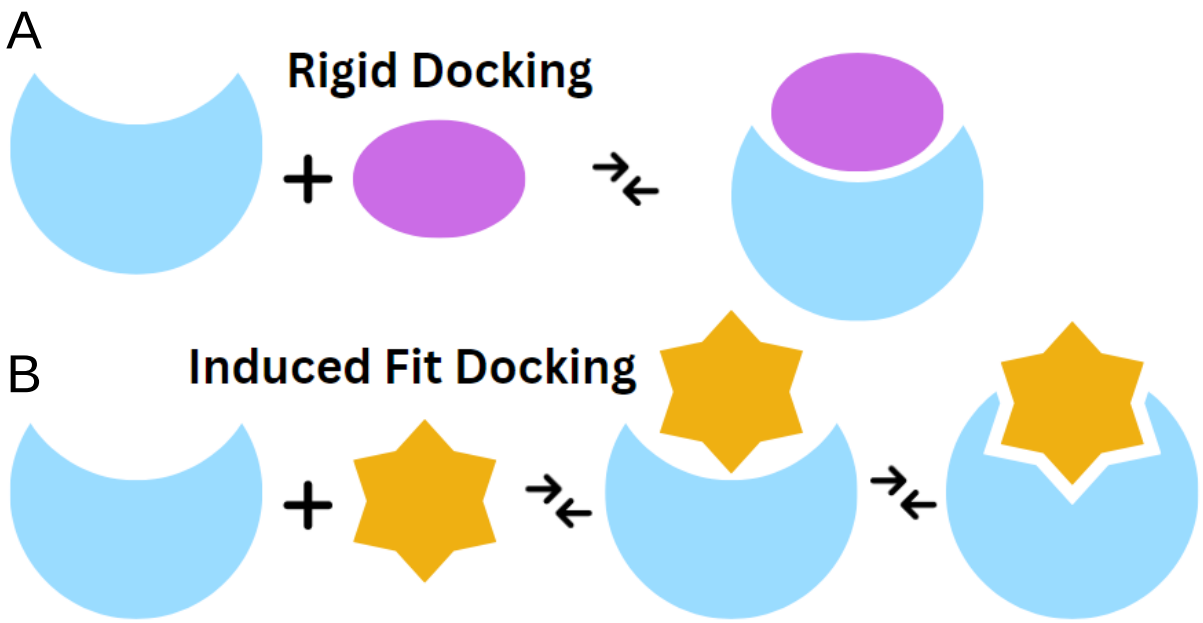


Figure 7: Comparing Rigid and Induced Fit Docking Models. A: In traditional virtual docking investigations, ligands are introduced into the binding site of a receptor under the assumption that the receptor remains rigid, referred to as the 'Lock-and-Key' model However, relying on a fixed receptor may lead to inaccurate outcomes, as numerous proteins in reality exhibit dynamic side-chain or backbone movements, or both, when interacting with ligands. B: These dynamic alterations enable the receptor to modify its binding site to better match the shape and binding manner of the ligand, commonly known as 'Induced Fit Model'. Image was generated using Canva. (Koshland., 1995; Morrison *et al., 2006).*

Molecular docking in Maestro Schrödinger utilises an advanced Induced Fit docking model, which combines the capabilities of both Glide and Prime algorithms to identify the most favourable binding conformations of ligands (Induced Fit Docking, Schrödinger, LLC; Sherman *et al.,* 2006). Prime is a molecular modelling software designed for protein structure prediction and refinement, optimising protein structures based on their amino acid sequence. It emphasises accurate modelling of interactions within the protein, enabling precise predictions of protein-ligand and

protein-protein interactions. To achieve a diverse set of ligand poses, the algorithm reduces the van der Waals radii in the ligand, allowing it to explore various compact conformations. Moreover, to avoid steric hindrance and unfavourable interactions, the algorithm temporarily removes highly flexible side chains, allowing the ligand to freely explore the active site. The goal is to generate a broad range of binding orientations and confirmations for a more comprehensive exploration of potential ligand-protein interactions.

On the other hand, Glide is a molecular docking software that predicts the interaction between small molecules (ligands) and protein active sites (Glide, Schrödinger, LCC, 2023). It employs a robust algorithm known as Grid-based Ligand Docking with Energetics (Glide), which efficiently identifies and ranks ligand poses based on their binding affinities (Yang *et al.,* 2021; Friesner *et al.,* 2006; Halgren *et al.,* 2004; Friesner *et al.,* 2004). Glide's scoring system, GlideScore, evaluates various energetic contributions including van der Waals interactions, Coulombic interactions, lipophilic interactions, hydrogen bonding and metal binding. It also considers specific molecular features, such as the freezing of rotatable bonds, assigning rewards or penalties accordingly (Glide, Schrödinger, LLC, New York, NY, 2023). Glide uses this series of filters to search for possible locations and orientations of ligand binding to an active site within the protein, within the selected grid box. This produces a variety of docked poses, ranked according to an Induced Fit Docking Score (IFDScore) (Yang *et al.,* 2021; Friesner *et al.,* 2006; Halgren *et al.,* 2004; Friesner *et al.,* 2004). The IFDScore is calculated as the sum of the glide docking score and 5% of the prime energy derived from the refinement step. More negative IFDScores are indicative of more favourable ligand-protein binding interactions. The glide docking score is composed of two components: the GlideScore and the Epik state penalty. The GlideScore reflects how effectively the ligand fits into the binding pocket of the protein, highlighting the ligand's suitability for the active site (Induced Fit Docking, Schrödinger, LLC). Meanwhile, the Epik state penalty accounts for any unfavourable alterations in the ligand's conformation or tautomeric state, predicting the ionisation/tautomeric states and measuring the associated energy cost with generating each state in a solution with a neutral pH. Consequently, the docking score serves as an estimation of the binding strength between CBD and TBR1 T-box domain, with more negative scores indicating more favourable binding interactions, signifying a stronger affinity (Yang *et al.,* 2021; Friesner *et al.,* 2006; Halgren *et al.,* 2004; Friesner *et al.,* 2004).

In the Maestro Schrödinger platform, the IFD method was used to simulate the interaction between energy-minimised and optimised structures of CBD and the TBR1 T-box domain. Opting

for the standard protocol, the docking box was configured to span the entire molecular chain, tailored to accommodate ligands up to 46 Å in length. A limit of 20 poses per ligand was set, with selection criteria based on a Coulomb-vdW score under 100 and an H-bond score over -0.05. This was followed by side-chain predictions for residues within 5 Å of any ligand pose, allowing for the flexibility in residue inclusion and the potential integration of an implicit membrane model. The Prime minimisation step was then applied to each pose, adjusting the receptor structure to the ligand's conformation. Glide redocking of the ligand within the adjusted receptor structure was performed, with a focus on the most energetically favourable configurations under the default energy threshold of 30 kcal/mol. The ligand underwent rigorous docking using the default Glide settings, which produced various CBD-docked TBR1 T-box domain poses, each evaluated by its IFDScore (Induced Fit Docking, Schrödinger, LLC; Farid *et al.,* 2006; Sherman *et al.,* 2006; Sherman *et al.,* 2006). Additionally, 2D ligand-interaction diagrams for all poses were analysed to determine the specific interactions between CBD and the TBR1 T-box domain, aiming to predict the most biologically relevant binding mode.

To assess the accuracy of these simulations, RMSD values were computed for each docked pose by isolating and superimposing the ligand and protein structures onto their respective native conformations. A RMSD value of 2 Å or less, is generally considered indicative of a well-matched docked pose (Castro-Alvarez *et al.,* 2017). RMSD represent the average atomic displacement, suggesting the degree of structural deviation. Lower RMSD values indicate better structural alignment between the native and docked structures and therefore induced minimal structural changes upon binding. The quantitative assessment of RMSD aids in identifying the degree of resemblance between the docked compound and the native conformation, thus offering valuable insights into the potential biological implications of the ligand-protein interaction.

3.6 Molecular Dynamics

Molecular dynamics (MD) simulations are a powerful tool in computational biology, allowing researchers to explore the dynamic behaviour of biomolecules at an atomic level (Hollingsworth and Dror., 2018). The fundamental principle of MD is to simulate the time-dependent evolution of a molecular system by solving Newton's equations of motion for atoms or molecules, thereby predicting their trajectories over time (Karplus and Mc Cammon., 2002). This process involves the application of a force field—a mathematical model representing the forces acting on the atoms, including bond stretching, angle bending and non-bonded interactions like van der Waals and electrostatic forces (Lindorff-Larsen *et al.,* 2012). In an MD simulation, the system, typically composed of proteins, ligands, solvents and ions, is first set up in a specific state and then the positions and velocities of the atoms are updated over discrete time steps (Hollingsworth and Dror., 2018). This approach allows for the observation of molecular movements and interactions over time, enabling researchers to study properties such as protein folding, conformational changes, ligand binding and reaction dynamics under varying conditions. Temperature and pressure are also simulated, making it possible to replicate different environmental conditions (Hollingsworth and Dror., 2018).

After importing the prepared IFD structures into Maestro Schrödinger, the ‘System Builder’ function from the taskbar was used to construct the model system (Maestro-Desmond Interoperability Tools, Schrödinger, 2023). For solvation modelling, the predefined TIP3P model was selected in the Solvation tab. The system’s periodic boundary conditions were established by specifying an orthorhombic box shape, ensuring a buffer distance around the solute structures. Solvent molecules were systematically duplicated and arranged, discarding those extending beyond the periodic box boundary. Default values were retained for box distances and angles, with subsequent minimisation of the box volume to reorient the solutes. The force field was configured to OPLS_2005 for the simulations. To achieve an electrically neutral system, ions were added. The system's charge was balanced by introducing a minimum number of sodium (Na+) or chloride (Cl-) ions, randomly positioned in the solvent. This ionisation was complemented by setting a salt concentration of 0.15 M, approximating physiological conditions of monovalent ions. The model system was then minimised to reorient the solutes within the optimised box volume, preparing it for the subsequent molecular dynamics simulations (Maestro-Desmond Interoperability Tools, Schrödinger, 2023).

Therefore, molecular dynamics of all IFD poses between CBD and TBR1 T-box domain, as well as the APO TBR1 T-box domain was simulated over 250 ns, with a total of 1000 trajectory frames to be captured. The NPT ensemble class was selected, with the temperature set at 300 K and pressure at 1.01325 bar, both adhering to default values. The simulations were run for the specified duration, after which a post simulation analysis was conducted (Desmond Molecular Dynamics System version 4.2; Maestro-Desmond Interoperability Tools, Schrödinger, LLC).

3.6.1 Post Simulation Analysis

Following the completion of MD simulations, a comprehensive post-simulation analysis was conducted to validate the accuracy and reliability of the results. This analysis encompassed several key aspects, including Simulation Quality Analysis, Protein Cα and Ligand-Fit-to-Protein (LP) RMSD, Protein-Ligand Contacts, Cα-RMSF, rGyr, and Secondary Structure Analysis. Each of these components played a vital role in providing a deeper understanding of the biomolecular interactions, structural stability and conformational changes of the TBR1 T-box domain in both its APO form and when bound to CBD. The APO simulation, acting as a control, provided insights into the protein's dynamics and behaviour when unbound. This baseline state was crucial for evaluating the protein's inherent flexibility, fluctuations and dynamics. These analyses collectively served to enhance our understanding of the molecular mechanisms at play and substantiate the docking and simulation outcomes.

3.6.1.1 Simulation Quality Analysis

To assess the reliability of the simulation, the Simulation Quality Analysis panel in Maestro was selected. This panel provided information on the energy (kcal/mol), potential energy (kcal/mol), temperature (K), pressure (bar) and volume ($Å^3$) of each system over the course of the simulation. Statistical analyses on each parameter were provided, including the average, standard deviations and slope (ps^{-1}). This post-simulation evaluation serves to verify the simulation's fidelity and accuracy, ensuring that the results are both reliable and reflective of realistic molecular behaviour.

3.6.1.2 Protein Cα and Ligand-Fit-to Protein Root Mean Square Deviation (RMSD)

Tracking the RMSD of a protein or ligand gives insight into its structural conformation over the course of the simulation (Kabsch., 1976). Specifically, the alpha carbon (Cα) RMSD quantifies structural similarities or differences between protein conformations by focusing on the Cα atoms, key components of the protein backbone (Kufareva and Abagyan., 2012). The RMSD calculation,

shown as Equation 1 below, involves aligning the corresponding Cα atoms in the superimposed structures and determining the root mean square of their positional differences (Maestro, Schrödinger, LCC, 2023). This calculation yields a numerical value reflecting the average atomic displacement, thereby elucidating structural changes or consistencies. Lower Cα-RMSD values indicate greater structural alignment and less deviation (Maestro, Schrödinger, LCC, 2023). Similarly, the LP- RMSD involves superimposing the docked ligand onto the corresponding region of the native protein, aligning the ligands heavy atoms (Maestro, Schrödinger, LCC, 2023). This analysis provides insights into conformational shifts upon ligand binding, impacting the stability and equilibrium of the simulation.

Equation 1:

$$RMSD_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i'(t_x)) - r_i(t_{ref}))^2}$$

Where N is the no. atoms, t_{ref} is typically $t = 0$, r' is the position of the atoms selected relative to the reference frame and x is the time recorded at time t (Maestro, Schrödinger, LCC, 2023).

The RMSD values were obtained by loading the trajectory of each individual simulation and opening the 'Simulation Interaction Analysis' in the task bar, followed by exporting the Simulation Interactions Diagram Report. The Cα-RMSD and LP-RMSD was plotted over the 250 ns simulation duration, using MS Excel. The average RMSD and standard deviation (reported as average ± std. dv) was calculated for each simulation.

3.6.1.3 Protein-Ligand Contacts

For every CBD-docked pose, a 2D ligand interaction diagram was generated using Maestro, highlighting the hydrogen bonds formed between CBD and specific amino acid residues persisting for over 30% of the simulation (Maestro, Schrödinger, LCC, 2023). These diagrams represent the spatial arrangement of bonds, offering insights into the nature and persistence of key interactions. Simultaneously, an interaction fraction plot (bond distribution plot) was generated for each amino acid involved in CBD interactions. These plots offer a quantitative representation of the distribution of bonds formed between the ligand and protein over the simulation time. These plots are particularly valuable for assessing the dynamic changes in the ligand-protein complex, validating binding modes across various poses. The 2D ligand interaction diagram and bond distribution plots were generated in the Simulation Interaction Analysis Report previously exported.

3.6.1.4 Cα-Root Mean Square Fluctuations (Cα-RMSF)

RMSF is a metric employed in MD simulations to assess the flexibility and fluctuations of individual atoms within a biomolecular system over time (Vazquez *et al.,* 2018). RMSF is calculated by measuring the fluctuations of Cα-atoms across the simulation frames, according to Equation 2, providing insights into the dynamic behaviour and inherent flexibility of the protein's backbone (Maestro, Schrödinger, LCC, 2023).

Equation 2:

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} < (r_i'(t)) - r_i(t_{ref}))^2 >}$$

Where T is the simulation time over which the RMSF is calculated, t_{ref} is typically $t = 0$, r' is the position of the atoms in residue i relative to the reference frame and the < / > indicate the mean of the squared distances is calculated across the chosen atoms within the residue (Maestro, Schrödinger, LCC, 2023).

The Cα-RMSF was generated for each simulation (APO and CBD-bound poses) in the Simulation Interaction Analysis Report previously exported and was plotted per residue number, using MS Excel. The average RMSF and standard deviation (reported as average ± std. dv) was determined for each simulation.

3.6.1.5 Radius of Gyration (rGyr)

The rGyr is a metric used to assess the overall compactness or spread of a molecular system (Shrivastava *et al.,* 2022). It is calculated as the root mean square distance of each atom in the system from its centre of mass (Maestro, Schrödinger, LCC, 2023). A smaller rGyr suggests a more compact and densely folded structure, while a larger radius indicates a more extended or open conformation (Shrivastava *et al.,* 2022). In the context of TBR1 T-box domain, analysing the rGyr is valuable for understanding the global structural characteristics and dynamic changes in the macromolecular structure, offering insights into its stability, folding and interactions with CBD.

The rGyr was generated for each simulation (APO and CBD-bound poses) in the Simulation Interaction Analysis Report previously exported and was plotted as the difference (ΔRg) between the APO TBR1 T-box domain and the CBD-docked poses, using MS Excel. The average rGyr and standard deviation (reported as average ± std. dv) was determined for each simulation.

3.6.1.6 Secondary Structure Analysis

Lastly, Maestro predicted the secondary structure of the APO and CBD-docked simulations. The α-helix, β-strand and random coil percentage was generated in the Simulation Interaction

Analysis Report (Maestro, Schrödinger, LCC, 2023). This analysis provided insights into the changes in secondary structure upon ligand binding and therefore, stability of the complex.

Objective 2: Assessing the Potential Interaction between TBR1 T-box domain and CBD using Experimental Approaches

In molecular biology, experimental validation is necessary for corroborating speculations drawn from bioinformatics simulations. This necessity stems from the limitations of computational models, which, despite their advanced algorithms, cannot fully encapsulate the intricate complexities of biological systems. Thus, the experimental aspect of this study is designed to empirically validate the interactions between CBD and the TBR1 T-box protein domain initially predicted through *in silico* techniques.

The pET-11a plasmid, containing the codon optimised TBR1 T-box gene, was used to transform T7 *E. coli* cells for protein expression. Subsequently, the expressed protein underwent purification through Immobilised Metal Affinity Chromatography (IMAC) and Size Exclusion Chromatography (SEC), methodologies chosen for their efficiency in isolating proteins with high specificity and purity. Following purification, Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) was used to determine the molecular size and purity of the recombinant TBR1 T-box domain. With adequate quantities and purity achieved, the study proceeded to the Pull-Down Assay (PDA) to explore the potential binding between CBD and the TBR1 T-box domain. In addition, a Thermal Shift Assay (TSA) was conducted to assess any changes in the thermal stability of the protein in response to CBD binding, providing insights into the binding-induced conformational changes. Furthermore, Time-resolved Fourier Transformed Infrared Spectroscopy (FTIR) was employed to analyse the functional groups in control samples compared to those incubated with CBD, aiming to confirm their interaction.

This experimental approach is designed not only to validate the results obtained from bioinformatics analysis but also to substantiate the hypothesised interaction between CBD and the TBR1 T-box domain, thereby contributing to a deeper understanding of their molecular relationship.

3.7 Transformation of pET-11a Plasmid into *E.coli* cells

Transformation involves introducing foreign DNA into a host cell, often a bacterium, to modify its genetic makeup. This process, crucial for studies in genomics and protein expression, requires the preparation of competent cells. These cells are treated chemically or through electroporation to make their cell walls permeable to DNA (Chung *et al.,* 1989). During transformation, these competent cells are exposed to exogenous DNA, like a plasmid carrying a specific gene of interest (Mandel and Higo., 1970). Post uptake, the foreign DNA either integrates into the host genome or remains as an independent entity, replicating and expressing the introduced genes (Lederberg and Cohen., 1974). The efficiency and success of transformation are often validated using selection markers, commonly antibiotic resistance genes embedded in the plasmid. Cells that successfully incorporate and express these genes survive in selective media (Chung *et al.,* 1989).

To explore the interaction between CBD and the TBR1 T-box protein, sufficient quantities of isolated TBR1 T-box domain was expressed and subsequently purified. The pET-11a plasmid obtained from GenScript (USA) was selected for this purpose, encoding the codon optimised sequence of TBR1 T-box domain. This expression system was chosen for its remarkable efficiency in harnessing cellular resources for the robust expression of the target gene, as evidenced by previous research indicating that proteins expressed via this system can constitute over 50% of the cell's total protein production in a brief period (Novagen., 1992). The gene encoding the protein of interest, as illustrated in Figure 8, was positioned between the Lac operator (LacO gene) and the T7 terminator, utilising the Nde1 and BamH1 restriction sites. Furthermore, the gene was strategically fused with a His-tag and a thrombin cleavage site, offering important advantages during the subsequent purification steps.

Protein	Name	Residues	Construct
TBR1	Wild Type T-box	203-396	HMGSA**WSHPQFEK**GSS**HHHHHH**SSG**LVPRGS** **–** T-box Sequence

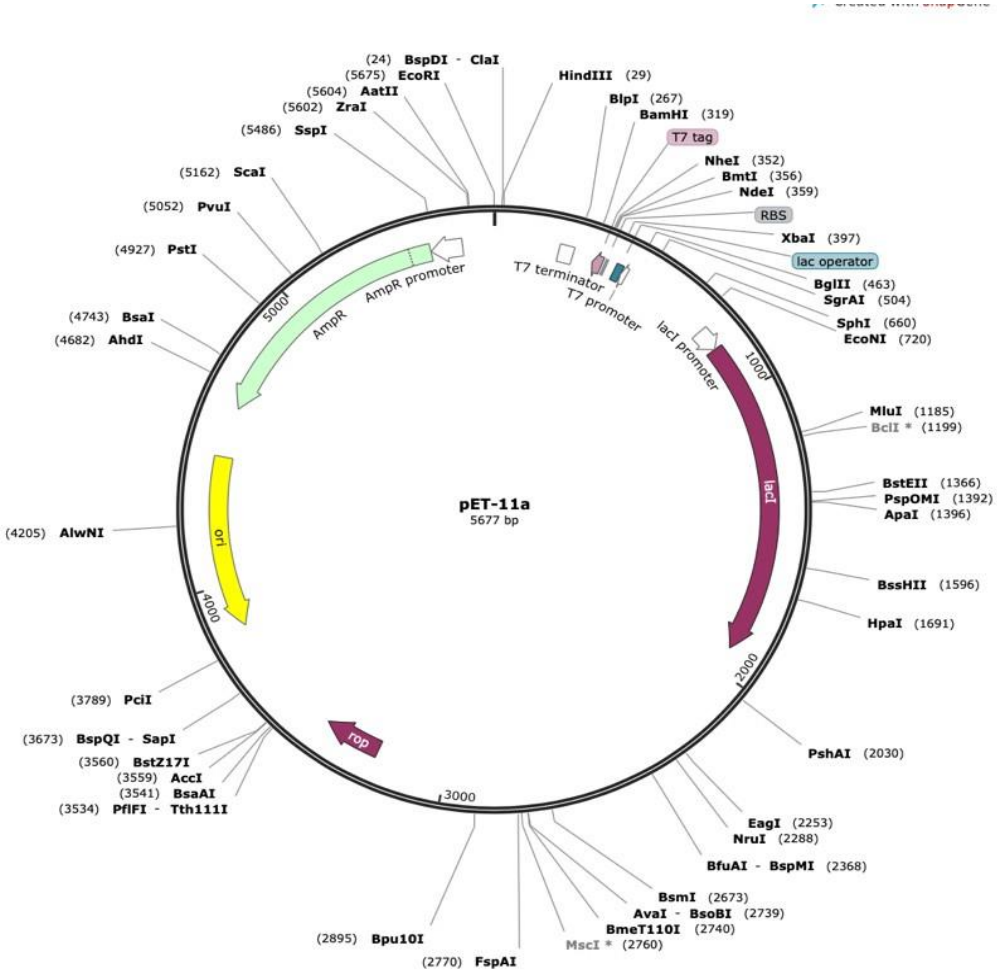


Figure 8: pET-11a Plasmid Utilised in Transformation: During the transformation process, the pET-11a plasmid was used. This plasmid incorporates a T7 RNA polymerase, which binds to the T7 lac promoter and initiates transcription in the presence of IPTG. The AmpR gene, on the other hand, encodes a β-lactamase enzyme responsible for the degradation of the β-lactam ring present in ampicillin. This ensures the survival of bacteria when grown in ampicillin-containing media. The TBR1 T-box construct is displayed above the pET-11a plasmid. It encompasses several distinct elements. Notably, a Strep tag is depicted in blue, a Histidine tag in green, and a Thrombin cleavage site in purple. Specifically, the Thrombin cleavage site (LVPRGS) is inserted after the Histidine tag, facilitating the removal of the tag during purification by cleaving arginine and glycine residues. Moreover, the inclusion of a Strep tag permits potential additional purification using streptavidin affinity chromatography if required. The black regions within the sequence represent linker residues, as indicated by Snapgene (2021).

Competent T7 Escherichia coli (*E. coli)* cells, obtained from New England Biosciences (USA), containing the pET-11a plasmid with the TBR1 T-box domain gene of interest, were transformed. A volume of 2 µl of the plasmid was added to 100 µl of competent cells and the mixture was incubated on ice for 30 minutes. Subsequently, a heat shock was implemented at a temperature of 42 °C for a duration of 45 seconds. This was done out to induce the permeabilisation of the cell membrane, facilitating the successful insertion of the plasmid into the cell. The brief exposure to elevated temperature causes a temporary depolarisation of the cell membrane, creating conditions conducive to drawing in the negatively charged DNA into the cells (Panja *et al*., 2006). The cells were then subjected to an additional incubation on ice for 5 minutes, followed by the addition of 950 µl of recovery media, specifically SOC media (Super Optimal Broth with Catabolite repression). The SOC media was composed of 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, and 40 mM glucose. This approach aimed to optimise the transformation efficiency and allow sufficient time for the bacterial cells to recover and generate ampicillin resistance.

The mixture was incubated at 37 ℃ for one hour, accompanied by vigorous shaking. Thereafter, a serial dilution was carried out on agar plates containing 0.1 mg/mL of ampicillin (Melford, UK). These plates were left to grow overnight at 37 ℃ to allow for the expression of the AmpR gene. The following day, isolated colonies were transferred to a nutrient-rich growth medium known as 2xYT media, comprising 1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl and 0.1 mg/mL of ampicillin and incubated overnight. Given that the AmpR gene confers resistance, only colonies possessing this gene would survive overnight, significantly reducing the risk of contamination. On the following day, colonies were isolated to generate a 1:1 ratio of cell culture to 80% glycerol stock, ensuring long-term preservation in the -80 ℃ freezer. A GeneJet miniprep kit (Thermo Fisher scientific, EU) was used to isolate the plasmid, according to the manufacturer's instructions. The isolated plasmid DNA was then sent to Inqaba Biotec (RSA) for Sanger sequencing to validate the accuracy of the sequence for subsequent experiments. Finally, the plasmid DNA was stored at -20 ℃ to maintain its integrity.

3.8 Expression of TBR1 T-box domain

To overexpress the codon-optimised TBR1 T-box domain gene, T7 *E.coli* cells were used to ensure maximum expression levels for a single gene, in a cost-effective and efficient manner (Rosano and Ceccarelli., 2014). In the T7 *E.coli* genome, the LacI gene plays a vital role in encoding the Lac Repressor, a protein that binds to the lac promoter (New England BioLab C3010I Datasheet). This binding action effectively prevents the *E. coli* RNA polymerase from initiating the transcription of the T7 RNA polymerase gene, thereby inhibiting its production. However, the addition of IPTG (Isopropyl β-D-1-thiogalactopyranoside) (Melford, UK) acts as a regulatory switch; it displaces the Lac Repressor from the lac promoter. This displacement allows the *E. coli* RNA polymerase to resume its function, leading to the transcription of the T7 RNA polymerase gene. The system is intricately designed so that T7 RNA polymerase interacts with its promoter only post-IPTG intervention, thus inducing protein expression, as shown in Figure 9. Such a regulatory mechanism significantly enhances the specificity and control of gene expression within the T7 *E. coli* system (New England BioLab C3010I Datasheet).

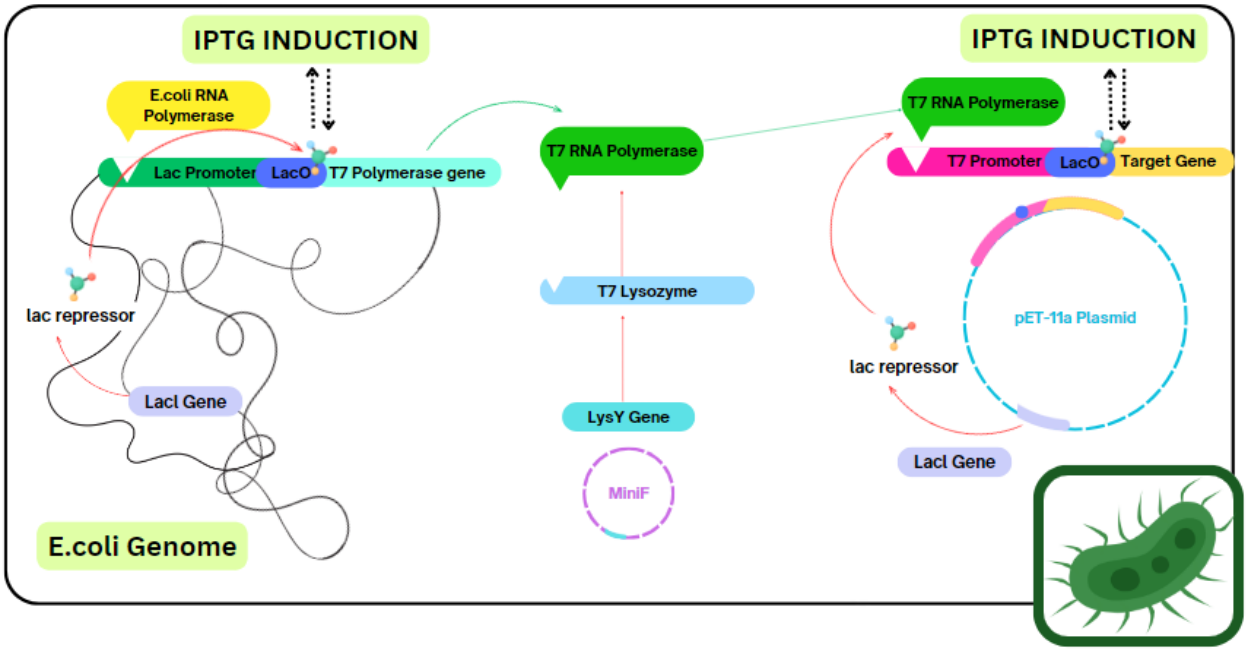


Figure 9: pET11-a Plasmid Expression System in T7 E.coli Cells. This figure illustrates the key regulatory elements within the T7 E.coli genome. Central to this regulation is the LacI gene, encoding the Lac Repressor, which binds to the lac promoter, preventing transcription of the T7 RNA polymerase gene by *E. coli* RNA polymerase. Introduction of IPTG disrupts this interaction, releasing the inhibition and enabling transcription of T7 RNA polymerase. Graphic created using Canva. (New England BioLab C3010I Datasheet).

A 100 mL overnight culture, containing 1% tryptone, 1.5% yeast and 1% NaCl, was created by introducing 2 mL of the glycerol stock, along with 100 μg/mL ampicillin. Subsequently, the culture was incubated at a temperature of 37 ℃, using an MRC Orbital Shaker Incubator, for a duration of 12-16 hours to allow the bacterial cells to reach the exponential phase of bacterial growth – in which the cells are healthiest and most uniform. Thereafter, 3.6 L of 2YT media was inoculated with a total of 3.6 mL of ampicillin (1 in 1000 dilution) and 12 mL of the overnight culture. The cells were incubated at 37 ℃ until an optical density (OD600) of 0.6 was attained. The OD600, measured using a spectrophotometer, is a widely employed method for determining the concentration of bacterial cells (Shao *et al.,* 2016). This measurement serves as a direct indicator of the stage of bacterial growth, enabling the optimal harvesting of cells. After reaching the desired OD600, the cells underwent a cold shock treatment for 30 minutes, followed by incubation with 0.2 mM IPTG. It is crucial to cool the solution prior to the addition of IPTG to mitigate the rapid synthesis of proteins, as higher temperatures could negatively impact protein folding and solubility (Gomes *et al.,* 2020). The media was immediately transferred to a 20 ℃ incubation for 24 hours at 230 rmp.

The following day, the overexpressed media was centrifuged at 5000 xg for 10 minutes at 4 ℃. Following the removal of the supernatant, the pellet was homogenised using 50 mL of equilibration buffer. The sample was supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) (Roche, Germany) and 0.1 mg/mL lysozyme (Sigma-Aldrich, USA). PMSF, an inhibitor that selectively binds to the active site of serine proteases, was employed to inhibit protease activity and safeguard the protein from degradation by endogenous proteases within the sample (Nemoto *et al.,* 2013). Lysis of the cells was facilitated by the addition of lysozyme, enabling the release of the protein from within the cells (Nemoto *et al.,* 2013). The following day, sonication was performed with 18x power, repeated 10-15 times, with intervals of ice incubation between each round. Subsequently, 50 μl of 0.01 mg/mL DNase (Merck, Germany) was added to the lysate and incubated for a duration of 30 minutes at room temperature on a slow spin. The lysate was filtered, resulting in the presence of soluble protein. The pellet was resuspended in equilibration buffer containing 30 mM imidazole, 50 mM NaCl, and 20 mM Tris-HCl at a pH of 7.5, followed by homogenisation. A small portion of both the pellet and the supernatant was retained and subjected to SDS-PAGE gel analysis.

The equilibration buffer was crucial for establishing the appropriate conditions required for subsequent chromatography steps, ensuring the system was equilibrated accordingly (Ullah.,

2012). Specifically, the inclusion of imidazole in the buffer served to disrupt weakly bound proteins in the IMAC column and allowed for the His-tagged protein to be eluted last with a high concentration of imidazole (Jiang *et al.,* 2006). NaCl was employed to mitigate non-specific interactions, while Tris-HCl regulated the pH of the buffer (Arakawa *et al.,* 2004).

3.9 Purification and Protein Purity

3.9.1 Immobilised Metal Affinity Chromatography

Immobilised Metal Affinity Chromatography (IMAC) was the first purification step used to purify TBR1 T-box domain. This is a widely used technique for protein purification based on the interaction between a metal ion (Co^{2+} or Ni^{2+}) and the recombinant protein – typically containing a poly histidine tag (His-tag) (Figure 10). The imidazole side chain of histidine exhibits high specificity for the chelated metals on the column, albeit with low affinity, effectively minimising non-specific interactions with the column. Consequently, the protein of interest would bind to the column, allowing unbound proteins to be eluted first. Through a series of wash steps, including a salt wash, any non-specifically bound molecules such as other proteins or DNA are eliminated. Subsequently, an elution buffer containing a high concentration of imidazole is introduced to facilitate the elution of the protein of interest. The high concentration of imidazole within the elution buffer effectively outcompetes the imidazole ring present in the His-tag, enabling the targeted protein to be released from the column (Arnau *et al.,* 2006).

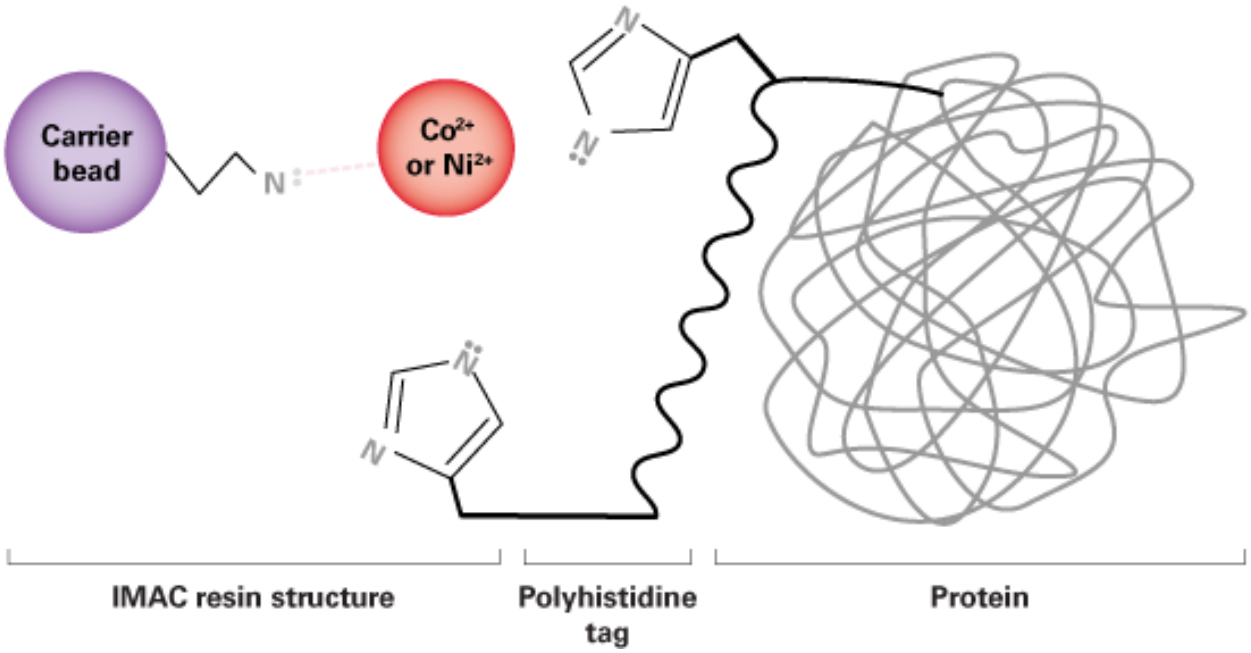


Figure 10: IMAC Resin Structure and Protein Purification Process. The diagram depicts the structure of an Immobilised Metal Affinity Chromatography (IMAC) resin containing cobalt (Co^{2+}) or nickel (Ni^{2+}) ions (red circle) and its interaction with a polyhistidine (His-tag) tag on a protein during the purification process. Each chelated ion has the ability to establish two hydrogen bonds with the nitrogen atoms in the imidazole side chain. Elution of the target protein is achieved by adjusting conditions such as imidazole concentration to disrupt the His-tag and metal ion interaction, releasing the purified protein from the column (Bornhorst and Falke., 2000).

The purification procedure incorporated a 5 mL HisTrap IMAC column (GE Healthcare, Sweden), specially loaded with cobalt ions. After cleaning the AKTA start protein purification system, the column was equilibrated with 5 volumes of equilibration buffer. This buffer consisted of 20 mM Tris, 500 mM NaCl, and 30 mM imidazole, at pH of 7.5. Subsequently, the supernatant was loaded into the super loop and the flow-through was collected. Following this, an additional equilibration buffer equivalent to 5 column volumes was introduced to re-establish equilibrium within the system. A detergent salt wash (20 mM Tris, 1.5 M NaCl, 30 mM imidazole and 0.1 % Tween20, at pH of 7.5) was used to wash the column with 5 column volumes. The inclusion of NaCl served to prevent non-specific binding to the column, while the incorporation of Tween20, a non-ionic detergent, significantly decreases non-specific interactions (Biocare., 2014). These precisely formulated buffers played a dual role in preserving the protein's pH balance and disrupting the electrostatic forces that may have existed between any DNA bound to the column and the target protein. Following another re-equilibration of the column, an elution buffer was introduced,

comprising 20 mM Tris, 500 mM NaCl, and 300 mM imidazole, at pH of 7.5, effectively releasing the proteins that had adhered to the column, with the target protein being the last to elute. Subsequently, the system underwent final equilibration and was stored in a 20% ethanol solution, followed by a thorough system wash. All collected fractions, in addition to the supernatant and pellet, were analysed by SDS-PAGE.

3.9.2 Size Exclusion Chromatography

Size exclusion chromatography (SEC) represents a purification technique to further refine the purification of the TBR1 T-box domain. The fundamental principle underlying SEC, shown in Figure 11, hinges upon the separation of molecules, primarily on their respective sizes and masses, collectively termed the Stokes radius (Carmack and Kelley., 1968; Sherman., 1999)

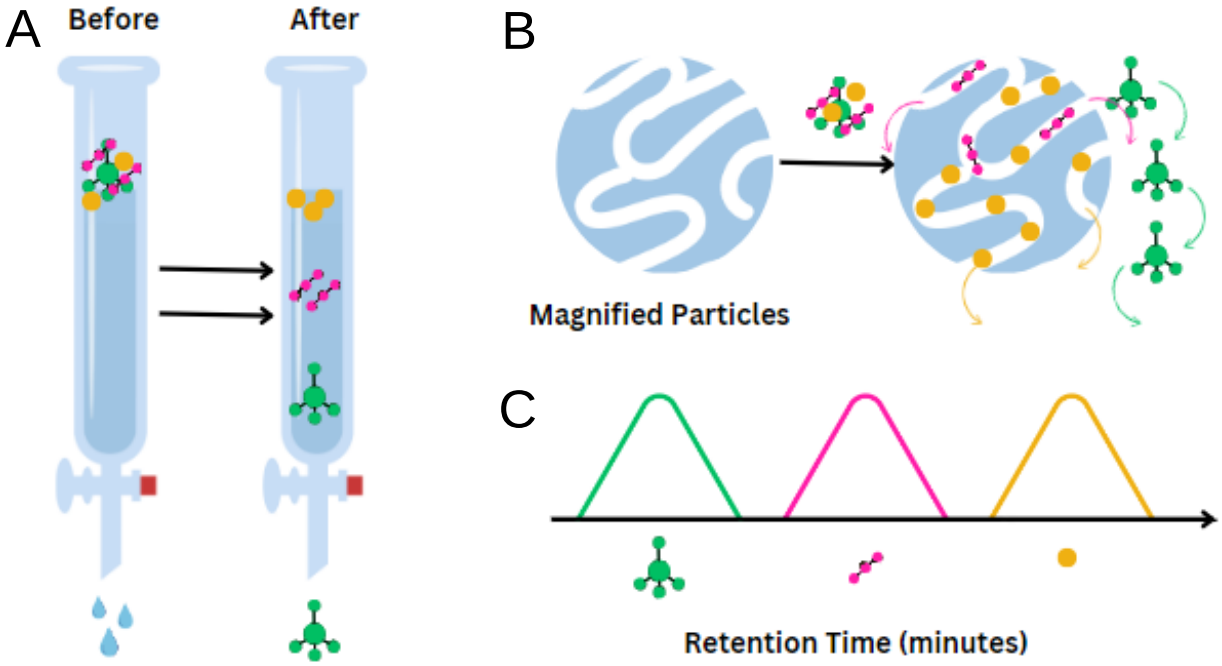


Figure 11: Principle for Size Exclusion Chromatography (SEC). A: As the solution traverses the column's stationary phase, predominantly comprised of beads, molecules undergo separation based on their Stokes radius. B: Larger molecules (green) are eluted first as they struggle to penetrate the porous matrix. In contrast, smaller molecules (orange) enter the pores, experiencing delayed elution due to their slower movement. C: Molecules of varying sizes exhibit distinct retention times, eluting according to their size (from large to small: Green, pink and orange). Those not interacting with the matrix are classified within the void volume, sharing an elution rate with the buffer. The image was created using Canva. (Carmack and Kelley., 1968; Sherman., 1999)

To purify the TBR1 T-box domain, a Superdex™ resin featuring covalently linked agarose porous beads (Sepharose) connected through dextran chains was chosen. This selection stemmed from its stability and lack of reactive or adsorptive properties (Merck., 2021). In addition to facilitating the purification of the TBR1 T-box domain, SEC is also used to determine the molecule's size relative to a set of established standards (Carmack and Kelley., 1968).

The protein concentration was increased using a VivaSpin 500 centrifugal concentrator before SEC. To accomplish this, 12 mL of the highest purity IMAC elutions were transferred into spin columns and subjected to centrifugation at 12000 xg for 12 minutes at 4 ℃. The flow-through was collected and designated for use in the SEC procedure. Furthermore, to refine the sample, it was filtered through the Amicon Ultra 0.5 mL filtration device. To prevent spontaneous dimer formation at positions 7 and 72 of the TBR1 T-box domain protein, which includes cysteine residues, 2 mM of DTT, commonly referred to as Cleland's reagent, was incubated with the sample at room temperature for 20 minutes (Carmack and Kelley., 1968).

To safeguard the equipment from potential damage and the substance's molar volume and the stationary phase's integrity (Merck., 2021) are maintained, a pressure limit of 0.3 kPa was set. Variations in pressure could exert an undue influence on these critical parameters. To facilitate the SEC process, a controlled flow rate of 5 mL/min was established on the ÄKTA Prime UHPLC System. The HiLoad 26/600 Superdex™ 75 prep grade column (GE Healthcare, USA) was equilibrated with an equilibration buffer throughout the night. This extended equilibration was necessary to prevent any unwarranted interactions between the protein and the column resin, thereby preserving the resolution's integrity. Subsequently, 2 mL of the prepared sample was loaded through the column. During this process, discernible peaks emerged as a direct consequence of the flow-through. These peaks were subjected to UV range analysis, yielding valuable insights into the presence of aromatic amino acids within the protein. The resultant flow-through was collected in numerous fractions and securely stored at a temperature of 4 ℃. Each of these fractions was analysed using SDS-PAGE.

3.9.3 SDS-PAGE

SDS-PAGE is an essential technique in the characterisation of biomolecules, particularly for evaluating protein purity and determining molecular sizes. SDS-PAGE relies on the migration of charged protein molecules through a porous polyacrylamide matrix in response to an applied electrical field (Fairbanks *et al.,* 1971). Proteins are denatured by sodium dodecyl sulphate (SDS), a detergent that uniformly coats the proteins with a negative charge proportional to their length. Consequently, when exposed to an electric field within the gel, these negatively charged proteins migrate towards the positively charged anode (Fairbanks *et al.,* 1971). The separation is primarily based on molecular size; smaller proteins traverse the gel matrix faster than larger ones. The gel's pore size, dictated by the acrylamide concentration, acts as a molecular sieve, allowing separation within specific size ranges (Fairbanks *et al.,* 1971). This technique employs a discontinuous system consisting of two distinct gel regions: a stacking gel and a resolving gel. The stacking gel allows all protein samples to converge at the same starting point before entering the resolving gel. This is achieved by adjusting the pH and composition of the stacking gel to ensure that all molecules align at the interface between the two gel segments (Fairbanks *et al.,* 1971).

One notable advantage of this discontinuous system over standard continuous SDS-PAGE lies in its ability to provide higher resolution in separating proteins based on size (Fairbanks *et al.,* 1971). The resolving gel itself is formulated to achieve this objective, comprising acrylamide/bis-acrylamide stock along with a gel buffer solution containing Tris and HCl, accompanied by SDS to disrupt protein-protein interactions. Additionally, glycine chloride is introduced into the gel system. During electrophoresis, chloride ions migrate most rapidly in the stacking gel, followed by the protein molecules and then glycine (Fairbanks *et al.,* 1971). Ultimately, these entities converge at the interface between the stacking and resolving gels. As the process continues, glycine outpaces the proteins in the resolving gel, thereby creating a uniform electrical field for the proteins to navigate, allowing for precise separation based on size (Stringer., 2004).

The 4% stacking gel was prepared by combining 30% monomer solution, stacking buffer, 10% SDS and distilled water within a falcon tube. For the 12.5% separating gel, 30% monomer solution, separating buffer, 10% SDS and distilled water were combined. The Bio-Rad Mini PROTEAN^{TM} Tetra Cell electrophoresis system was used to conduct the SDS-PAGE. 10% APS (Ammonium Persulfate) and TEMED (Tetramethyl ethylenediamine) was introduced into the separating gel mixture, which is subsequently poured between two glass plates. Following gel

polymerisation, 10% APS and TEMED are added to the stacking gel mixture and a comb was inserted to create the wells for sample loading. These steps ensure the structural integrity of the gel and precise formation of wells for sample loading. To initiate electrophoresis, the glass slides containing the polymerised gels were transferred into tanks filled with Tank buffer, composed of 250 mM Tris-HCl, 192 mM glycine, and 0.1 % (w/v) SDS at a pH of 8.3. This buffer system is crucial for providing the appropriate ionic environment for electrophoresis and for maintaining pH conditions conducive to effective protein migration (Oswald., 2016).

Before loading samples onto the gel, a mixture of 10 µl of the sample and 10 µl of 2x reducing sample buffer (comprising 20% (v/v) glycerol and 10% (v/v) β-mercaptoethanol at pH 6.8) was prepared. The addition of SDS serves to impart a negative charge to proteins, ensuring uniform charge-to-mass ratios among the samples. β-Mercaptoethanol plays a crucial role in denaturing proteins by breaking disulfide bonds, further facilitating uniform electrophoretic behaviour (Fairbanks *et al.,* 1971). Upon completion of electrophoresis, Coomassie blue is used to stain the gel (Shapiro., 2017). The inclusion of Tris-HCl in the system maintains the pH environment and enhances the rate of ion movement, contributing to the overall success of the SDS-PAGE analysis (Oswald., 2016).To ensure effective denaturation of proteins, the samples were vortexed and boiled for a duration of 5 minutes. This combined approach serves to efficiently denature the proteins, ensuring their linearisation. Subsequently, 5 µl aliquots of the treated samples were loaded into the wells of the SDS-PAGE gel. Initiating the electrophoresis process, a voltage of 80 V was applied until the dye front reached the gel interface, at which point the voltage was increased to 160 V. This continued until the dye front reached the far end of the gel, effectively separating the proteins by size. A 5 µl Thermo scientific PageRuler^{TM} Prestained Protein Ladder, 10 to 180 kDa (Thermo Fisher Scientific, EU) was loaded into the first lane to serve as a reference for protein molecular weights. Following electrophoresis, the gels were carefully removed for staining with Coomassie stain solution, consisting of 0.1 % (w/v) Coomassie dye, 1 % (v/v) acetic acid, 5 % (v/v) methanol and 94 % (v/v) water. Staining was conducted using a microwave for approximately 1 minute, ensuring uniform dye distribution. The gel then underwent an overnight de-staining process in a solution identical to the staining solution but without the Coomassie dye. Multiple de-staining steps were performed to achieve clear visualisation and quantification of the protein bands.

3.9.4 Protein Concentration and Purity

UV-visible spectroscopy is an analytical technique for determining protein concentration and assessing purity (Sharma., 2004). The principle of spectroscopy for protein analysis relies on the absorbance of ultraviolet light by proteins, primarily due to the aromatic amino acids like tryptophan and tyrosine (Edelhoch., 1967). At 280 nm, these amino acids exhibit strong absorbance, which can be quantitatively measured to determine protein concentration. This method's efficacy is grounded in the Beer-Lambert Law, which is shown as Equation 3. The Beer-Lambert law stipulates that the quantity of electromagnetic radiation absorbed or transmitted by a specimen is directly proportional to the concentration of said specimen and the distance traversed by the radiation. Following its passage through the specimen, the intensity of the incident radiation surpasses the emitted radiation (Christian and O'Reilly., 1986). The purity of protein samples can also be evaluated using spectroscopy by measuring the ratio of absorbance at 280 nm (A280) and 260 nm (A260). Proteins absorb at 280 nm, while nucleic acids absorb strongly at 260 nm (Stoscheck., 1990). A lower A260/A280 ratio indicates higher protein purity, with a ratio of approximately 1.8 being generally indicative of pure protein (Grimsley and Pace., 2004).

Protein sample concentrations were quantified using a Thermo Scientific™ NanoDrop™ One Spectrophotometer. The absorbance spectrum was measured at 280 nm with the protein buffer used as a blank.

Beer-Lambert Law Equation 3: $A=\varepsilon l c$

This law establishes a linear relationship between the absorbance (A) of a substance in solution, its molar absorptivity (ε), the path length (l) of the cuvette, and the concentration (c) of the substance. For TBR1 T-box domain, ε is reported as 36565 M^{-1} cm^{-1}.

3.10 Structural Analysis

3.10.1 Far-UV Circular Dichroism

Circular dichroism (CD) spectrophotometry serves as a pivotal technique, offering valuable insights into the conformational aspects of proteins based on the differential absorption of left and right-handed circularly polarized light, a phenomenon exhibited by optically active, asymmetric

molecules (Woody., 1995). Proteins, with their intricate three-dimensional structures, inherently possess chiral elements due to the presence of chiral carbon centres within their backbone (Kelly *et al.,* 2005). The principle behind CD is explained in Figure12.

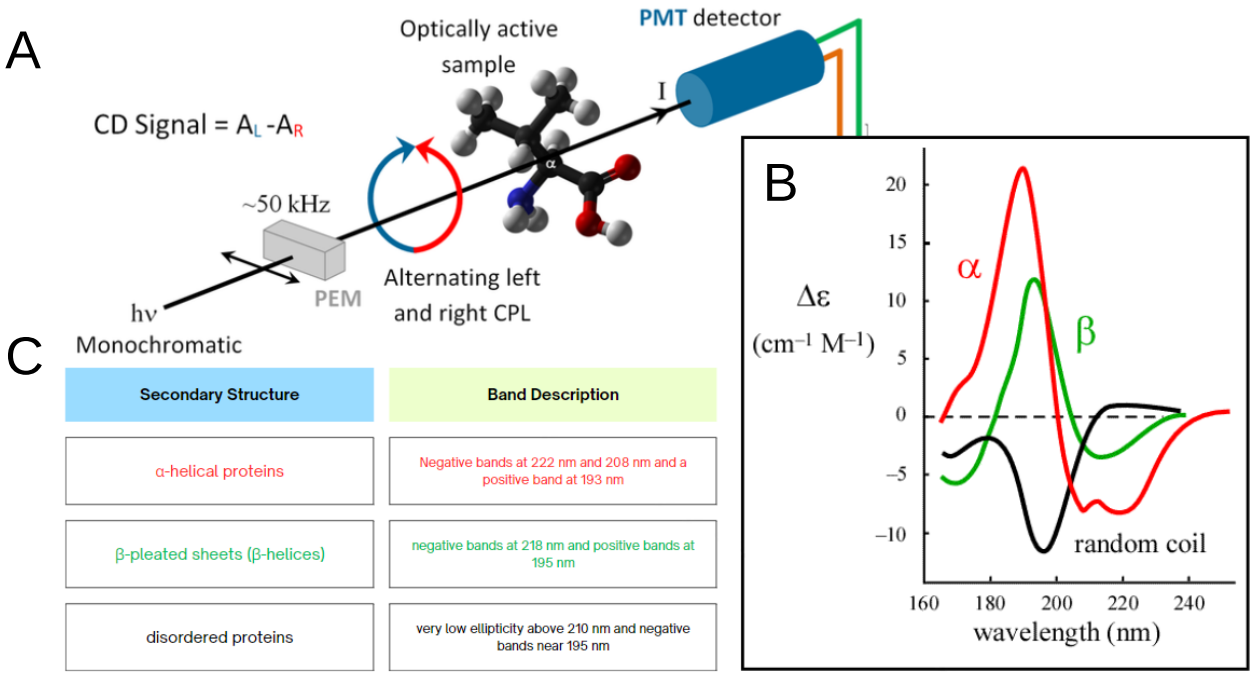


Figure 12: Principle of Far-UV CD. A: Linearly polarised light traverses a Photo Elastic Modulator (PEM), where it undergoes a transformation into alternating left and right circularly polarised light (CPL). This CPL comprises oscillating electric and magnetic fields, both rotating around the direction of propagation, thus forming sinusoidal wave patterns. The light is differently absorbed and detected by a Photo Multiplier Tube (PMT). Chiral molecules, including proteins, exhibit the capability to absorb this rotating light, a property that forms the basis for CD spectroscopy. B: CD spectra provide distinct signatures for various secondary structural elements of proteins, C: such as alpha helices (bands at 222 nm, 208 nm and 193 nm), beta sheets (bands at 218 nm and 195 nm), and random coils/ disordered proteins (bands at 210 nm and 195 nm). These spectral patterns offer a means to identify the presence and relative abundance of secondary structural motifs within the protein. Image created using Canva (Kelly *et al*., 2005*).*

The analysis of recombinantly expressed protein samples, such as the TBR1 T-box domain, often requires evaluating their secondary structure to ensure proper folding and structural integrity. Before commencing the experimental procedures, the recombinant TBR1 T-box domain was dialysed into a lower concentration of salt using a HEPES buffer solution (20 mM HEPES, 100 mM NaCl and 2 mM DTT). High chloride concentrations can potentially introduce interference with the emission spectra, leading to noisy data (Berndt., 1996). To unveil the secondary structure of TBR1 T-box domain, UV CD spectra were obtained in triplicate, using a Jasco J-810 Circular Dichroism Spectropolarimeter. This process involved using a quartz cuvette with protein at a fixed

concentration of 10 μM and a spectral range spanning from 200 to 260 nm. To further reduce the concentration of chloride ions, the protein solution was diluted with MiliQ water prior to measurements.

3.11 Pull-Down Assay

The PDA is a widely used technique in molecular biology to study protein-ligand interactions, offering a direct method to investigate and confirm the physical binding between two molecules. This method, rooted in the principle of affinity chromatography, involves immobilising a target molecule onto a solid support and subsequently exposing it to a sample containing the molecule of interest, referred to as the prey protein. This technique was slightly modified for the purpose of this study, where a ligand was chosen as the immobilised target. Therefore, this technique offers a high degree of specificity in isolating target proteins based on their affinities for the immobilised ligand (Lapetina and Gil-Henn., 2017). In conventional PDA, the bait protein would contain a tag fused to the protein (e.g. His-tag) and form an interaction with the metal-coated resin (similar to IMAC). However, in this study – the ligand interacts with the epoxide groups on the resin and form stable pH dependant linkages, as seen in Figure 13. After the incubation period, the unbound components are washed away, leaving only the complexes formed between the immobilised ligand and any interacting proteins (Lapetina and Gil-Henn., 2017). The bound proteins can then be eluted and analysed, often using techniques like SDS-PAGE and Western blotting, to identify and characterise the proteins that specifically interact with the ligand.

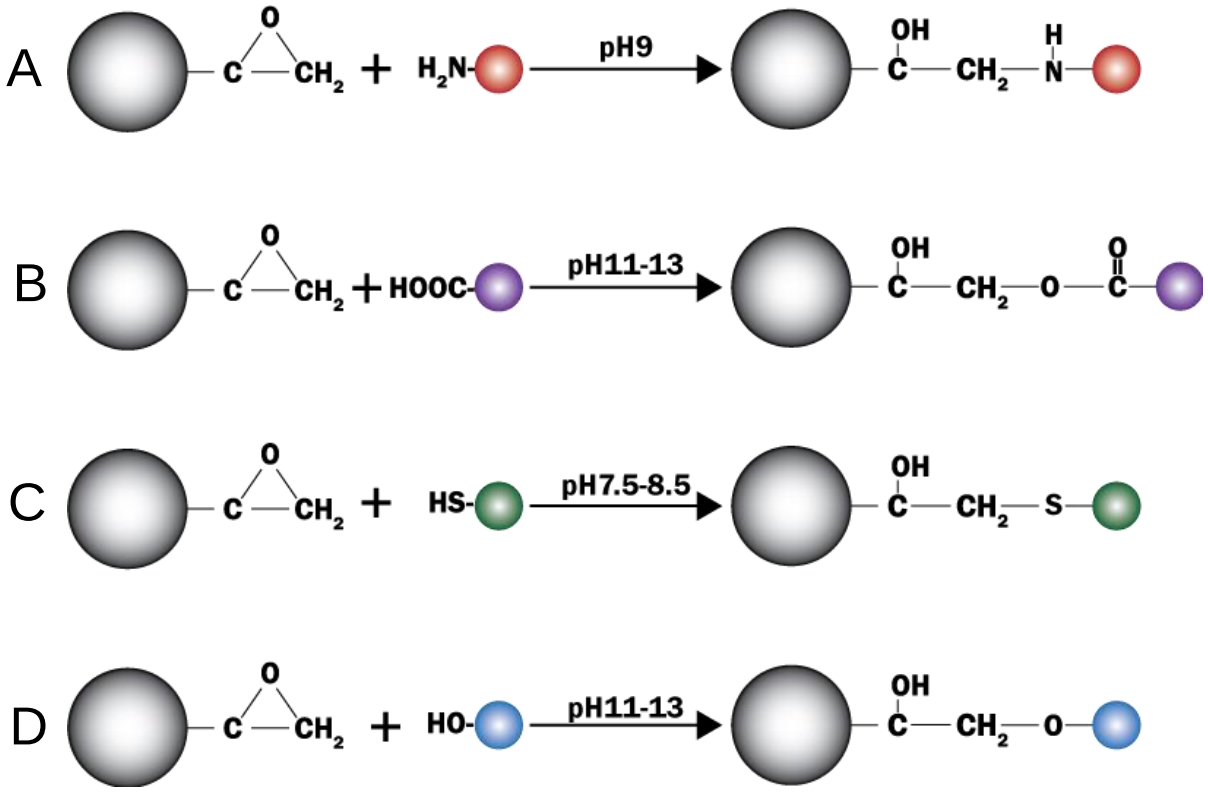


Figure 13: Epoxy Activated Agarose Pull-Down Assay: Epoxide-activated agarose is characterised by a dense distribution of epoxide groups, which engage in a ring-opening mechanism when encountering nucleophiles. These epoxide moieties readily undergo reactions with (A) primary amines, (B) carboxylic acids, (C) thiols and (D) hydroxyl groups, resulting in the formation of stable (A and B) secondary amide, (C) thioester and (D) ether linkages, respectively. Notably, thiol groups particularly favour these reactions, requiring a buffered system near physiological pH (pH 7.5-8.5). In contrast, epoxide groups necessitate elevated pH conditions to react with hydroxyl groups (pH 11-13) and moderately alkaline conditions for engaging amine groups (pH 9).

The epoxy activated agarose resin was prepared by accurately weighing 35 mg aliquots into three separate 2 mL Eppendorf tubes. Each tube was filled with MiliQ water and centrifuged at 1000 xg for 1 minute. The supernatant was discarded after each centrifugation and this washing step was repeated thrice to ensure thorough cleaning of the resin. A solution of 200 µM/mL CBD in a coupling buffer (100 mM sodium carbonate and 100 mM NaCl at pH 11) was prepared and incubated with the resin for 12 hours at 40°C. This pH was specifically chosen due to the hydroxyl group (OH) present in the phenol ring of CBD. Post incubation, the coupling reaction was quenched with 10 volumes of the coupling buffer, followed by centrifuging for 1 minute at 1000 xg at 4°C. To further quench the reaction, 10 volumes of 1 M ethanolamine at pH 8 was added and

incubated for 4 hours at 40 °C. Subsequently, the resin was washed with alternating low and high pH buffers to remove unbound substances. This involved three rounds of washing with 10 volumes each of 1 M sodium acetate (pH 4 with 0.5 M NaCl) and 0.1 M Tris buffer (pH 8 with 0.5 M NaCl). Each round consisted of filling the tubes with the respective pH buffer, shaking, and centrifuging for 1 minute at 1000 xg at 4°C.

Next, 50 µl of the resin slurry was collected and loaded into a spin-down column. The column was centrifuged at 2000 xg for 1-2 minutes at 4 °C to remove any residual buffer. Then, 600 µl of protein buffer (Equilibration buffer) was added to the column and centrifuged for 1 minute at 1000 xg at 4°C and repeated three times. Following the washes, 300 µl of TBR1 T-box protein at 120 uM was added to the resin. After these washes, 300 µl of TBR1 T-box protein at a concentration of 120 µM was added to the resin for interaction studies. This step was replicated in three distinct experiments: two involving CBD-coupled resins, with incubation periods of 1 hour and 24 hours, respectively, at 4 ℃, and a control experiment with a 1-hour incubation using resin without CBD. These varied incubation times allowed for the assessment of interaction strength under different conditions. Additionally, an absorbance spectrum of the CBD stock solution was measured to confirm the presence of CBD.

Post-incubation, each column underwent centrifugation at 2000 xg for 1-2 minutes. The initial flow-through collected at the column's base, termed Sample A, was expected to primarily contain unbound residual protein. Subsequently, the column was washed with 600 µl of protein buffer and centrifuged again, collecting the subsequent flow-through as Sample B. This wash step, repeated three times and cumulating all flow-through in Sample B, aimed to remove non-specifically bound proteins and contaminants. Following the washes, buffer was added to the resin and a fraction of this was collected as Sample C, which was anticipated to contain the target protein if it had successfully bound to the CBD-coupled resin. For analysis, the Thermo Scientific™ NanoDrop™ was used to examine the results. Unlike traditional PDA analysis, typically conducted using gel electrophoresis for protein-protein interactions, the absorbance spectrum method was used instead. This choice was informed by the need to detect both CBD and protein; while gel-based methods are limited to protein detection, the absorbance spectrum can effectively identify both components. This approach in PDA was strategically chosen to explore the potential binding between CBD and the recombinant TBR1 T-box domain, corroborating the bioinformatics predictions with empirical evidence.

3.12 Thermal Shift Assay

The TSA, a reliable method for determining the melting temperature (Tm) of proteins, was conducted to assess the impact of CBD binding on the thermal stability of the recombinant TBR1 T-box domain. This assay used SYPRO Orange dye in combination with the Bio-Rad CFX96 Touch Real-Time PCR Detection System to monitor the protein's thermal denaturation process.

SYPRO Orange, a hydrophobic dye, initially experiences maximum quenching with negligible fluorescence due to its surrounding water molecules in the solution. However, as the protein unfolds upon heating, hydrophobic regions are exposed. The dye then binds non-specifically to these exposed hydrophobic patches, resulting in a decrease in quenching and a consequent increase in fluorescence intensity. This interaction is pivotal, as the dye's fluorescence intensifies in proportion to its binding with the unfolding protein, leading to an inflection peak on the fluorescence curve that corresponds to the protein's Tm (Niesen *et al.,* 2007). The Tm, indicative of the temperature at which 50% of the protein is thermally denatured, is a crucial parameter for understanding protein stability (see Figure 14). An increase in Tm generally denotes enhanced protein stability, often observed in the presence of stabilising agents like interacting ligands (Bruylants *et al.,* 2005; Huynh and Partch., 2015). Therefore, this assay not only determines the Tm but also provides insights into the stabilising effects of CBD on the recombinant TBR1 T-box domain.

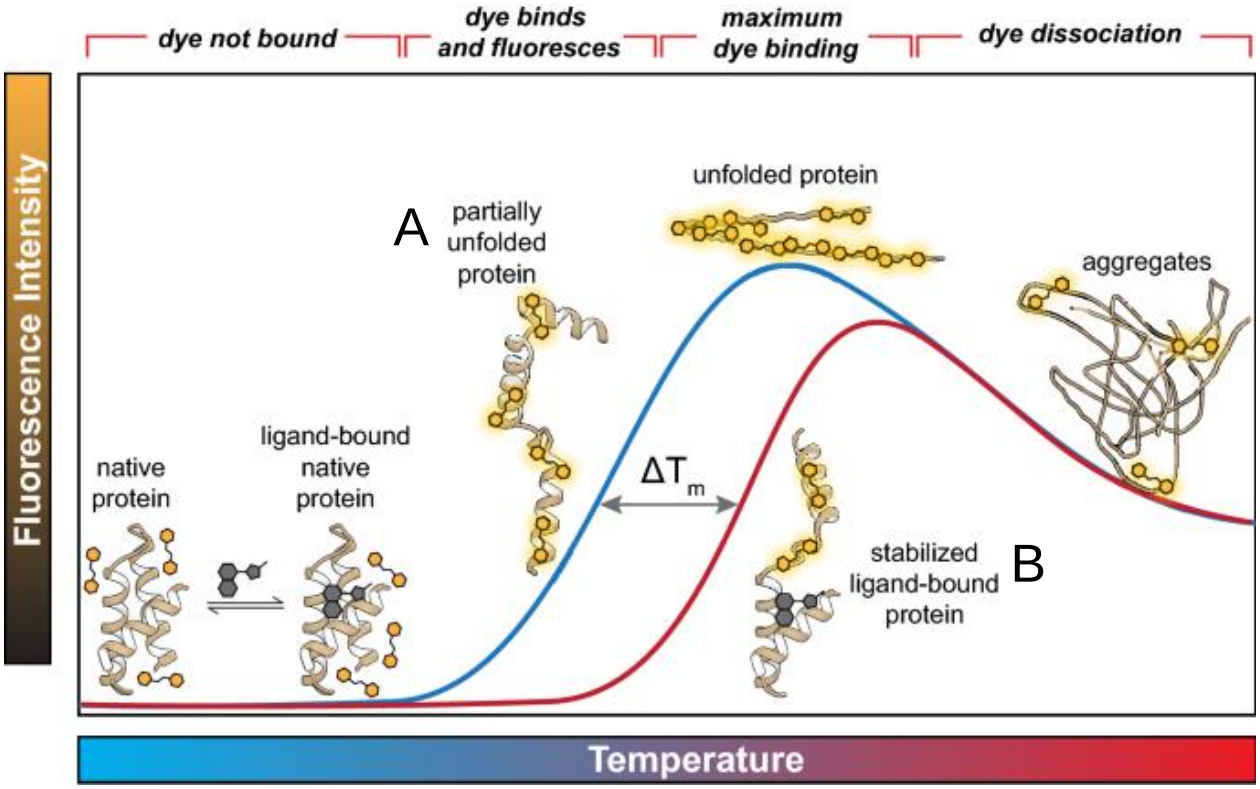


Figure 14: TSA Principle. A TSA measures the melting temperature (Tm) of the protein, which reflects the temperature at which half of the protein population transitions from a folded to an unfolded state upon increase in temperature. A: The exposure of hydrophobic residues, due to protein unfolding, results in increased fluorescence of SYPRO Orange dye – which binds to those hydrophobic patches. B: Compound binding to the target protein can enhance its stability, consequently elevating the Tm. Image was sourced from Samuel *et al*., (2021).

To conduct the TSA, a stock solution of the TBR1 T-box domain was prepared at a concentration of 66 µM. The assay was performed in a 96-well PCR microplate provided by Bio-Rad. For each well, the setup was designed to achieve a total volume of 25 µL. The first lane of the microplate was allocated for control solutions, which included wells containing only HEPES buffer, the TBR1 T-box domain protein, and isolated CBD, each in separate wells. Subsequent wells were allocated for the assay samples, each containing 10 µM of the TBR1 T-box domain protein in HEPES buffer, accompanied by a range of CBD concentrations varying from 3 µM to 200 µM. This gradient allowed for the assessment of CBD's effect on protein stability across a spectrum of concentrations. To track protein denaturation, 5 µL of SYPRO Orange dye was added to each well. The remaining volume in each well was completed with buffer, ensuring uniformity across all samples.

The microplate was securely sealed using a Microseal 'B' PCR Plate Sealing Film (Bio-Rad). The plate was vortexed and centrifuged for 30 seconds. The samples were analysed using a Real-Time PCR Detection System. To capture the thermal denaturation profile, a melting curve was generated by subjecting the samples to a temperature range extending from 20 °C to 80 °C. The temperature increments used were set at 0.5 °C, with each increment lasting 10 seconds. Upon completion of the TSA, the relative fluorescence unit (RFU) and the derivative of the fluorescence signal concerning temperature (-d(RFU)/dT) were accessed using CFX Maestro. To ensure the accuracy of the findings, the average RFU and -d(RFU)/dT values were corrected by subtracting the averaged blank values. The degree of protein unfolding was determined using the corrected average of the RFU. Both the fraction unfolded and the corrected average of (-d(RFU)/dT) were plotted individually against temperature.

3.13 Time-resolved Fourier Transformed Infrared Spectroscopy

FTIR is an analytical technique used to obtain an infrared spectrum of absorption or emission of various states of matter - solids, liquids or gases. This technique detects the vibrational properties of chemical functional groups in a specimen when exposed to infrared radiation (Barth., 2007). When infrared light interacts with the specimen, it is absorbed through vibrational transitions, where incident light frequencies align with molecular vibrational frequencies. Therefore, the chemical bonds vibrate at a frequency characteristic to their chemical environment, where they undergo stretching, contracting and bending motions (Barth., 2007). In the mid- and far-infrared spectral regions, absorption occurs when light and vibration frequencies match, accompanied by changes in the molecular dipole moment. The strength and polarity of vibrating bonds, influenced by intra- and intermolecular factors, affect this absorption (Barth., 2007). The position of an infrared absorption band depends on vibrating masses, bond type and electron effects in the molecular environment. Vibrational frequency and absorption probability rise with stronger and more polar vibrating bonds. Consequently, a specific wavenumber range is absorbed by a chemical functional group, irrespective of the molecule's overall structure. This correlation between the wavenumber position of a band and the chemical structure is used for functional group identification within a specimen, as shown in Figure 15. By monitoring the evolving IR spectra, researchers can deduce the step-by-step progression of molecular events during a chemical reaction or a biological process.

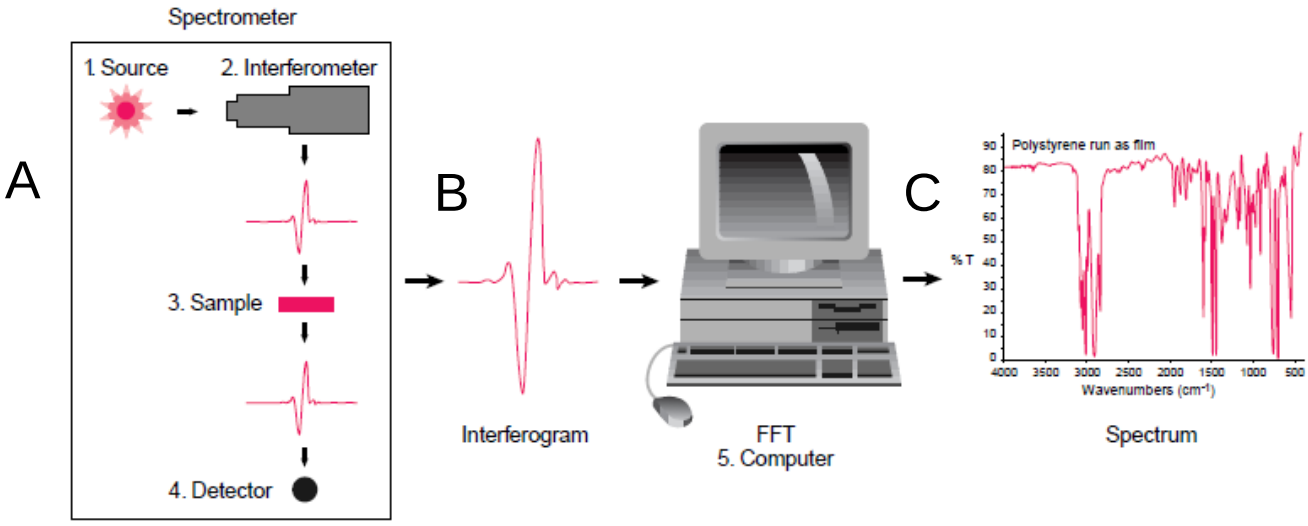


Figure 15: The fundamental principle underpinning Fourier Transform Infrared (FT-IR) spectroscopy. A: FT-IR leverages an interferometer—an optical device that bifurcates the incoming infrared beam into two optical beams. One of these beams reflects off a fixed mirror, while the other bounces off a movable mirror, capable of minute displacements from the beam splitter. B: As these two beams reconverge, they undergo interference, generating an interferogram. This unique signal encapsulates all the infrared frequencies originating from the source. Direct interpretation of this interferogram for sample analysis is not feasible. C: A computer executes a Fourier transformation, unmasking a comprehensive spectrum of intensity values at each frequency. Image sourced from Thermo Nicolet (Hasegawa., 2017).

FTIR spectroscopy plays a pivotal role in protein analysis, offering unique insights into the secondary structure composition of proteins (Ye *et al.,* 2014). Unlike X-ray crystallography and NMR spectroscopy, which primarily emphasise the tertiary aspects of protein structural data, FTIR spectroscopy specialises in uncovering the secondary structural characteristics of these biomolecules. This includes the determination of the Amide I (1600-1700 cm^{-1}), Amide II (1510-1580 cm^{-1}) and Amide III (1250-1350 cm^{-1}) bands, which originate from the amide linkages connecting amino acids (Ye *et al.,* 2014), shown in Figure 16. Compared to Far UV-CD, FTIR spectra offer a more intricate insight into the impact of ligand binding by furnishing detailed information about the specific vibrational modes influenced. In contrast, Far UV-CD primarily signifies alterations in the overall secondary structure.

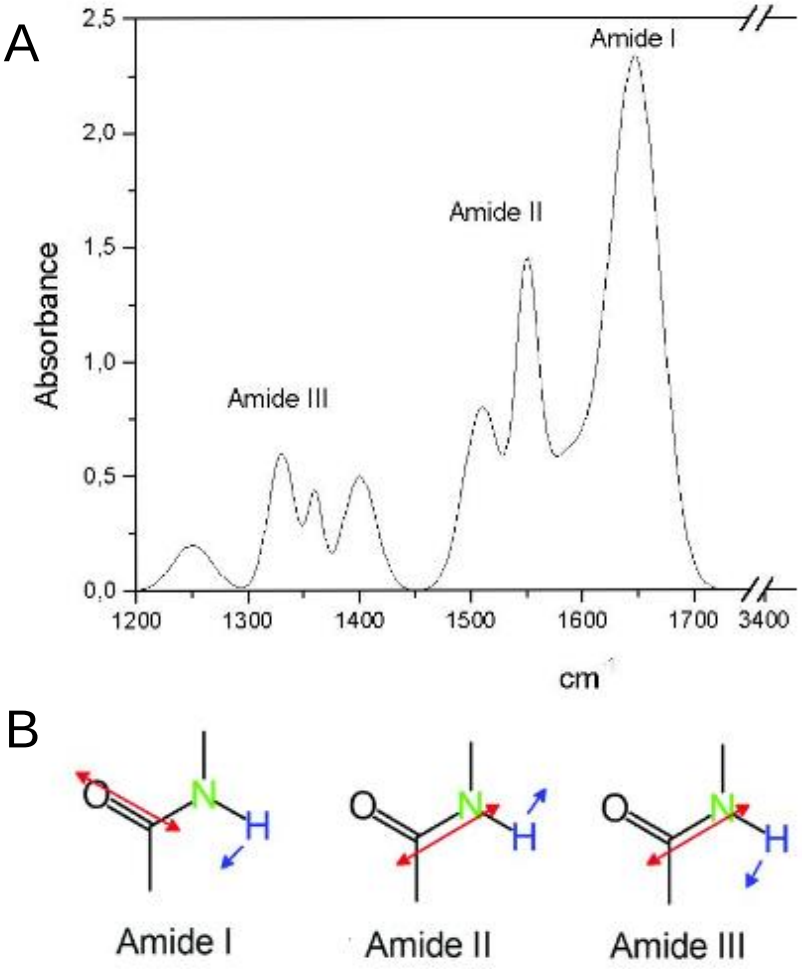


Figure 16: Amide I, Amide II and Amide III Vibrations in FTIR. A: Location of amide I (1600-1700 cm^{-1}), amide II (1510-1580 cm^{-1}), and amide III (1250-1350 cm^{-1}) bands in an FTIR spectrum. B: Red arrows represent C=O and N-H stretching. Blue arrows represent N-H bending. The Amide I band predominantly arises from the stretching vibrations of the C=O bond, constituting 70% to 85% of the potential energy within the amide bond. Its determination is primarily governed by the backbone conformation, showing independence from the amino acid sequence, as well as charge or hydrophobic/phillic properties. The Amide II band primarily results from bending vibrations of the N—H bond (40 – 60% of the bonds potential energy), C-N stretching vibration (18 – 40% of potential energy) and C-C vibrations (roughly 10% of the potential energy). Amide II bands are responsive to variations in the conformation of a structure. Amide III bands are very complex bands and therefore, offer limited structural information. Image adapted from Ichikawa *et al.,* (2023). (Chirgadze and Nevskaya., 1976; Venyaminov and Kalnin., 1990; Ji *et al.,* 2020).

Time-Resolved Fourier Transform Infrared Spectroscopy (TR-FTIR) is a specialised form of FTIR that provides dynamic insights into molecular interactions and changes over time (Possenti *et al.,* 2021). In the context of studying protein-ligand interactions, TR-FTIR is particularly valuable because it can capture rapid conformational changes and interactions at the molecular level

(Lorenz-Fonfria., 2020). When a ligand binds to a protein, it often induces conformational changes, which can lead to alterations in the vibrational frequencies of specific functional groups. TR-FTIR tracks these changes over time, providing real-time data on how the binding event affects the protein's structure and dynamics (Lorenz-Fonfria., 2020).

The use of Attenuated Total Reflectance (ATR) in TR-FTIR further enhances its capabilities. ATR is a sampling technique in which infrared radiation is passed through a crystal in contact with the sample (Possenti *et al.,* 2021). The radiation reflects at the crystal-sample interface, and the evanescent wave created can be absorbed by the sample, allowing for the measurement of its infrared spectrum. ATR-TR-FTIR is especially useful for studying protein-ligand interactions as it requires minimal sample preparation, can handle samples in various states (solid, liquid, or gel), and is highly sensitive to surface phenomena.

Using an ATR-FTIR spectrophotometer (PerkinElmer Spectrum One with Spectrum V5.00 software, PerkinElmer Instruments, CT, USA), the change in vibrational properties of chemical functional groups in response to interactions with infrared light was analysed between TBR1 T-box domain, CBD and TBR1 T-box domain-CBD complex. A 120 μM stock concentration of TBR1 T-box domain was prepared in equilibration buffer, along with a 1 in 4 dilution of CBD in equilibration buffer (containing 30mM imidazole, 500mM NaCl, and 20mM tris). Three control spectra were separately analysed, namely: equilibration buffer, TBR1 T-box domain, and CBD. Afterward, the TBR1 T-box domain and CBD stock solutions were incubated for 70 minutes, and measurements were taken every 10 minutes. The unique infrared signal resulting from the measurements carries a distinct signature for each sample. Through Fourier Transformation, interferograms were converted into infrared spectra, depicting transmittance as a function of wavenumber. A background spectrum was captured and subtracted from the sample spectra. Spectra were collected through 64 scans in transmission mode from 4000 to 400 cm^{-1} range. The resolution was set to 4 cm^{-1} and analysis was carried out at room temperature. The data was exported and plotted as Transmittance (%) as a function of Wavenumber (cm^{-1}) using MS-Excel.

4 CHAPTER 4: RESULTS

In light of increasing CBD usage among pregnant women, concerns have escalated regarding its potential impact on foetal neurodevelopment, particularly given the suggested link between prenatal cannabis exposure and Autism Spectrum Disorder (ASD) (Corsi *et al.,* 2020). TBR1, a gene encoding the transcription factor TBR1, is often associated with sporadic forms of ASD and intellectual disability due to protein dysfunction. This research project aims to investigate whether CBD can interact with the TBR1 T-box domain, a DNA binding domain of TBR1, establishing a basis for future studies on the effects of CBD binding on TBR1 function, potentially influencing normal neuronal development and ASD. The study's objectives were achieved by *in silico* and *in vitro* experimentation, to conclude whether an interaction between CBD and TBR1 T-box domain is possible. The *in silico* study used the AlphaFold model of TBR1 T-box domain (UniProt ID: Q16650) compared it TBX5 (UniProt ID: Q99593), to identify corresponding DNA binding residues. Thereafter, CASTP3.0, Protein Plus and SiteMap software analysed the structure of TBR1 T-box domain to determine potential binding pockets. These results were used to identify the regions with the highest potential for CBD binding, particularly those associated with DNA binding, as a preliminary step before conducting docking and molecular dynamics studies. Maestro Schrödinger software was used to analyse the ADMET properties of CBD (CID:644019), perform docking simulations and molecular dynamics between CBD and TBR1 T-box domain. The ADMET properties of CBD offer insights into the oral absorption potential for CBD to enter foetal circulation and potentially reaching the foetal brain. Docking simulations were employed to identify potential interaction sites and residues between CBD and the TBR1 T-box domain. The aim was to elucidate whether they interact, understand the nature, location and assess the stability and strength of the proposed interaction. Molecular dynamics simulations were conducted to further unpack the interaction between CBD and the TBR1 T-box domain. By analysing the RMSD, RMSF and rGyr, this study aimed to understand the dynamic behaviour and stability of the CBD-TBR1 T-box domain complex. These simulations provided insights into the conformational changes and binding interactions over time, offering a deeper understanding of how CBD potentially influences the structural and functional integrity of the TBR1 T-box domain. These *in silico* studies laid a solid foundation for subsequent *in vitro* experiments, which were designed to validate the interaction between CBD and the TBR1 T-box domain, thereby contributing to the broader understanding of CBD's role in neuronal development and its potential implications in ASD.

Objective 1: Assessing the Potential Interaction between TBR1 T-box Domain and CBD using Bioinformatics

4.1 Structure Analysis

In the absence of a crystallographically solved TBR1 T-box domain (UniProt ID: Q16650), the predicted AlphaFold structure was compared with the crystallographically solved structure of TBX5 in both DNA-bound and DNA-free forms (PDB: 2X6V) to identify potential DNA binding residues within the TBR1 T-box domain.

The T-box domain was separated from the overall predicted structure for two primary reasons: firstly, it represents the region of interest, as the aim is to investigate the potential binding of CBD to this specific region and its impact on the DNA binding ability of TBR1; secondly, this region boasts a remarkably high-confidence model, with a per-residue model confidence score (pLDDT) exceeding 90%, as depicted by the dark green region in Figure 17.

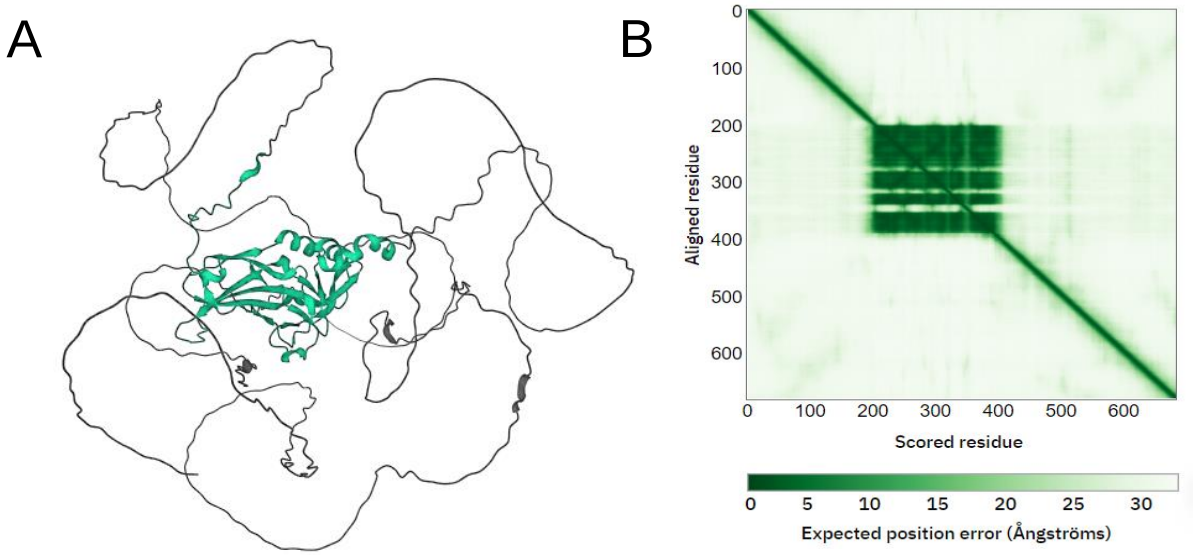


Figure 17: AlphaFold Predicted Structure of TBR1 T-box Domain. A: AlphaFold predicted structure of TBR1, with the T-box domain isolated in green. The Tbox domain is represented as having a very high confidence level (pLDDT > 90%) as shown by (B). B: Predicted aligned error (PAE) which assesses the reliability of the relative positioning and orientations of various domains within the predicted structure, offering insights into the confidence level in these aspects (Jumper *et al.,* 2021).

The T-box domain of the AlphaFold predicted structure of TBR1 was isolated, using PyMOL Molecular Graphics System (version 2.5 Schrödinger, LCC) and juxtaposed with the crystal structure of the T-box domain of TBX5 bound to DNA so as to pinpoint potential DNA binding residues. This comparative analysis aims to assess whether CBD might interact with these regions in forthcoming docking simulations, potentially affecting the protein's direct binding to DNA. It's crucial to note that this remains a predictive analysis, necessitating subsequent experimental validation. Therefore, serving as a predictive model, utilising a combination of sequence and structural alignment between the T-box domain of TBR1 and TBX5, the potential DNA binding residues of TBR1 is shown in Figure 18.

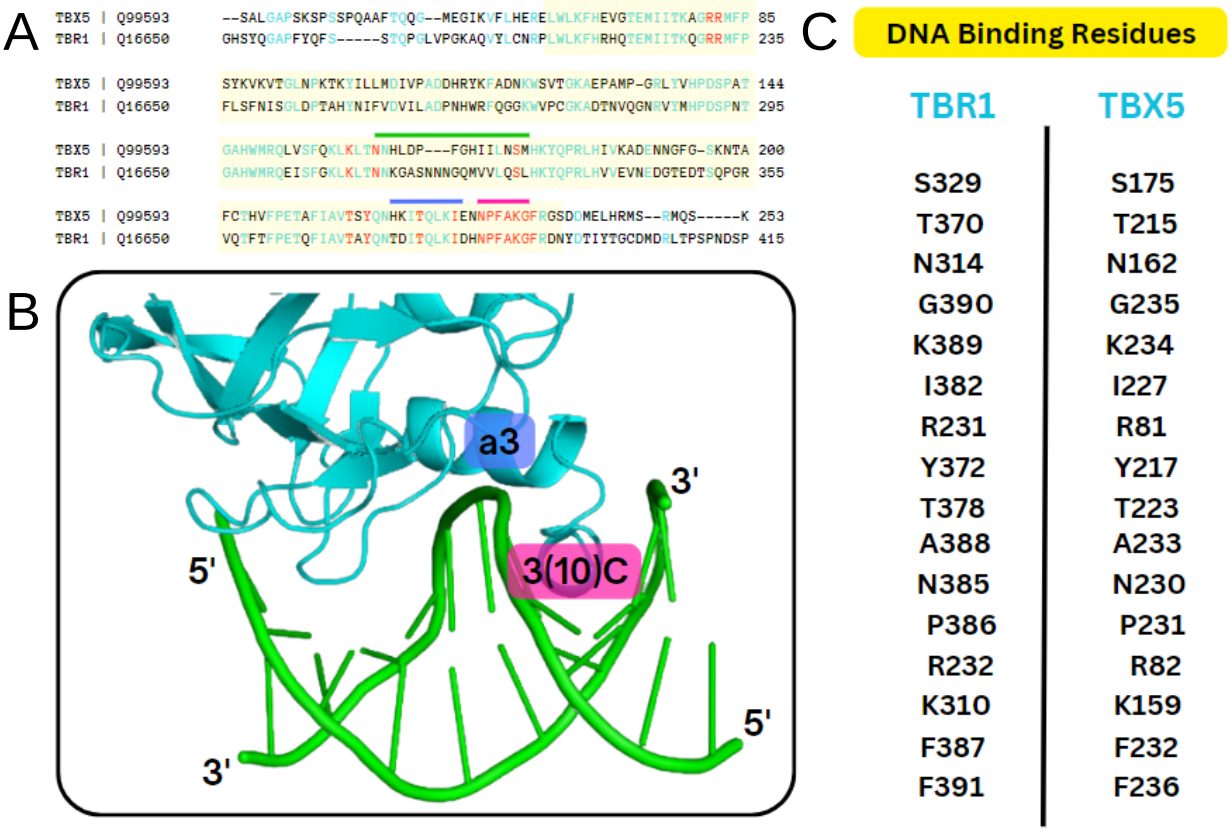


Figure 18:: Sequence and Structure Alignment between the T-box domains of TBR1 and TBX5. A: Sequence Alignment between the complete TBR1 (UniProt ID: Q16650) and TBX5 (UniProt ID: Q99593), where (A) shows only the region of interest – the T-box DNA binding domain. This domain is highlighted in yellow. Similar residues are shown in light blue, whereas the DNA interacting residues of TBX5 and the corresponding residues of TBR1 is shown in red. The structural alignment was performed using Clustal Omega. B: Structure alignment between the DNA binding domain of TBR1 (blue) interacting with DNA (from the crystallised structure of TBX5 interacting with DNA PDB: 2X6V) where only the α3 (purple) and 3_{10}C helices are present and predicted to interact with DNA. The structure alignment was performed using MatchAlign in PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LCC. C: DNA binding residues of TBX5 and the corresponding TBR1 residues, as predicted from the sequence alignment (UniProt., 2023)

A sequence alignment conducted with Clustal Omega between the T-box domains of TBR1 and TBX5 resulted in a 54.86% identity matrix, indicative of a moderate degree of conservation. Subsequently, a structural alignment was performed using PyMOL's MatchAlign tool on the T-box domains of TBR1 (comprising 181 amino acids) and TBX5 (comprising 176 amino acids). Despite three refinement cycles, only 141 atoms were successfully aligned, possibly due to structural variations or flexibility in specific regions. This alignment yielded a Cα RMSD of 0.438 Å, denoting a small average displacement of atoms and suggesting a close structural alignment between the T-box domains of TBR1 and TBX5. Regardless, each DNA-interacting residue in TBX5 corresponded precisely to the equivalent residue in TBR1. The sequence alignment between the T-box domain of TBR1 and TBX5 revealed the DNA interacting residues for TBX5 (S175, T215, N162, G235, K234, I227, R81, Y217, T223, A233, N230, P231, R82, K159, F232 and F236) corresponds to the DNA interacting residues of TBR1 (S329, T370, N314, G390, K389, I382, R231, Y372, T378, A388, N385, P386, R232, K310, F387 and F391 respectively). These findings were compared to the location and key residues of the predicted binding pockets located in the T-box of TBR1.

4.2 Predicting Binding Domains

The potential CBD-binding domains in TBR1 T-box were predicted using CASTp3.0 and Protein Plus. CASTp3.0 predicts the binding pockets within a protein based on geometric features, while Protein Plus considers hydrophobicity and shape complementarity, shown in Figure 19. This analysis aimed to provide insights into the possible interaction of CBD with these pockets and specific residues.



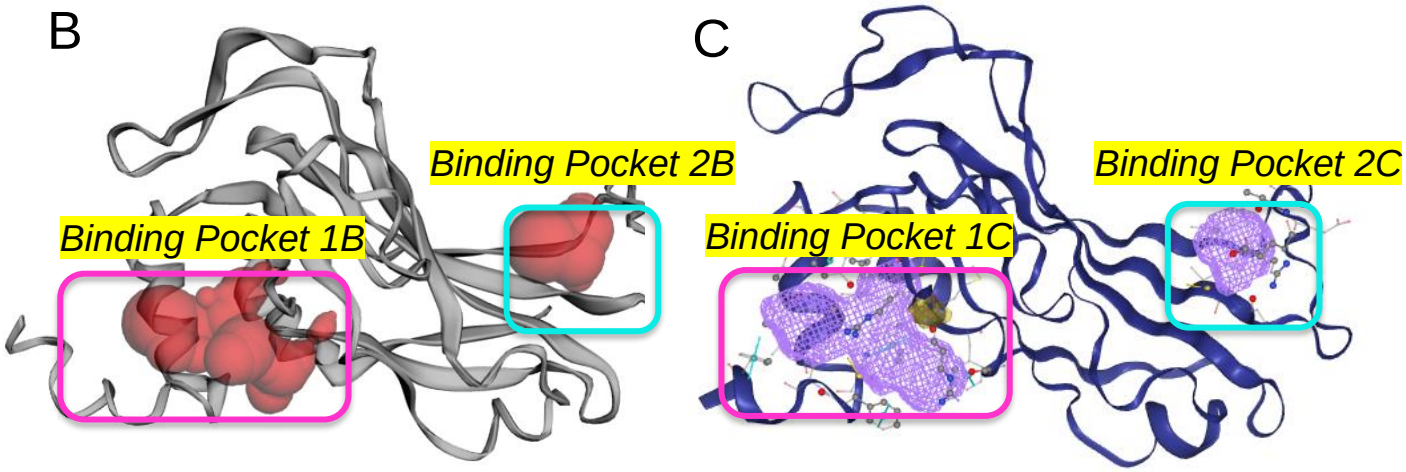


Figure 19: TBR1 T-box Domain Binding Pocket Prediction. A: Amino acid sequence of TBR1 T-box domain showing the secondary structure elements. Purple: 3_{10} helices. Pink: π-helix. Blue: turn. Green: bend. Grey: coil. Black: no structure and yellow: $\boldsymbol{\beta}$- pleated sheets. The residues highlighted in light blue are key residues within the binding cavity. B: Predicted binding pockets by CASTP3.0 indicated in red, based on spatial geometric measurements such as area and volume. Pink square: largest binding pocket (labelled as Binding Pocket 1B) with an area of 212.038 $Å^2$ and volume of 91.819 $Å^3$. Blue square: smaller binding pocket (labelled as Binding Pocket 2B) with an area of 71.645 $Å^2$ and a volume of 37.121 $Å^3$. Any other predicted binding pockets occupied a volume of less than 8.950 $Å^3$. C: Predicted binding pockets by ProteinsPlus indicated in purple, based on a combination of Protoss, DoGSiteScorer, and Geomine. This approach considers concavity, shape complementarity and hydrophobicity. Pink square: largest binding pocket (labelled as Binding Pocket 1C) with an area of 676.29 $Å^2$, volume of 260.10 $Å^3$ and hydrophobicity of 0.74. Blue square: smaller binding pocket (labelled as Binding Pocket 2C) with an area of 287.08 $Å^2$, a volume of 57.34 $Å^3$ and a hydrophobicity of 0.68. The larger binding pocket in both (pink) is located between α3 helix and the 3_{10}C-helix of the T-box domain, which is known to directly interact with DNA, whereas the smaller binding pocket (blue)is located near the e' beta pleated sheet, far from direct DNA binding regions (Tian *et al.,* 2018; Schöning-Stierand *et al.,* 2022).

CASTp3.0 found 20 key residues that reside within the binding pockets (M224, I225, R231, R232, M233, F234, F236, F306, G307, L309, L311, Y372, T378, K381, I382, N385, P386, F387, A388, F391), of which R231, R232, Y372, T378, I382, N385, P386, F387, A388 and F391 correspond to the predicted DNA binding residues in Figure 18. Additionally, the analysis conducted with CASTp3.0 and Protein Plus revealed the presence of two substantial binding pockets in the TBR1 T-box domain, designated as Binding Pocket 1 (with a volume ranging from 91.819 $Å^3$ to 260.10 $Å^3$) and Binding Pocket 2 (with a volume ranging from 37.121 $Å^3$ to 57.34 $Å^3$). Significantly, the binding pocket 1 is strategically located in proximity to the α3 helix and the 3_{10}C-helix of the T-box domain, situated towards the terminal region of the structure. Figure 18 reveals these regions to interact directly with the DNA.

While numerous researchers have found tools like CASTp3.0 and Protein Plus to yield accurate and reliable predictions (Binkowski *et al.,* 2003; Yahaya *et al.,* 2021; Tian *et al.,* 2018), these tools often lack detailed information on the properties and ranking of individual binding sites. To address this limitation, SiteMap (version 3.5, Schrödinger, LLC) was employed to assess the presence of hydrogen bond acceptors/donors and hydrophilic/hydrophobic regions within each binding site. SiteMap also provided valuable parameters such as SiteScore, Size, DScore, and Volume for each predicted binding site, as outlined in Table 1. Notably, SiteMap identified the same binding sites (labelled binding pocket 1 and binding pocket 2) as predicted by CASTp3.0 and Protein Plus.

Table 1: TBR1 T-box Domain Predicted Binding Pocket Analysis by SiteMap

Binding Pocket	**SiteScore**	**DScore**	**Volume**	**Size**
Binding Pocket 1	0.973	0.943	2000.655	95
Binding Pocket 2	0.573	0.515	95.354	24

SiteMap analysis provided a comprehensive analysis of the predicted binding pockets within the TBR1 T-box domain. In Binding Pocket 1, the SiteScore reached an impressive value of 0.973, designating it as a potential drug-binding site. This pocket exhibited significant hydrophilic occupancy (694.398 Å²) and a smaller hydrophobic region (57.442 Å²). The H-acceptor and H-donor regions measured 404.168 Å² and 299.854 Å², respectively, resulting in a total surface area of 1393.328 Å². Conversely, Binding Pocket 2 displayed a lower SiteScore of 0.515, classifying it as a non-drug-binding site. It also showed predominant hydrophilic occupancy (522.105 Å²), a smaller hydrophobic region (16.774 Å²), and H-acceptor and H-donor regions measuring 161.171 Å² and 351.421 Å², respectively, with a total surface area of 878.019 Å². Consistent with SiteScore, these measurements classified Binding Pocket 1 as druggable and Binding Pocket 2 as undruggable. Comparing results with CASTp3.0, both binding pockets' volumes (Binding Pocket 1: 2000.655 Å³, Binding Pocket 2: 95.354 Å³) exceeded those obtained from CASTp3.0 and Protein Plus. Lastly, the size represents a measure of the site, where roughly 3 site points correspond to each atom of the bound ligand and therefore is a good indicator of the preferred binding site. In light of these parameters, binding pocket 1 is the preferred choice for binding. However, due to the absence of literature regarding CBD binding to the T-box domain and the observation that the predicted binding domains exhibit high hydrophilicity (while CBD is

hydrophobic), the decision was made to use the entire DNA binding domain as the grid box for docking purposes.

These results were further unpacked using Schrödinger's Maestro to delve into the ADMET properties of CBD. Investigating absorption, distribution, metabolism and excretion helps determine how CBD reaches and affects the foetal brain, which is crucial for its neurological impact. Assessing CBD's toxicity, especially in pregnant women, ensures safety. Integrating ADMET data with computational predictions offers a comprehensive view of CBD's interaction with TBR1, crucial for accurate conclusions about its effects on neurodevelopment and conditions like ASD.

4.3 CBD Structure and ADMET Properties

ADMET properties, which encompass Absorption, Distribution, Metabolism, Excretion and Toxicity of a drug, are critical factors determining a drug's behaviour and safety within the human body. These attributes are integral in understanding the pharmacokinetics and safety profile of a compound. In the context of this study, the ADMET analysis of CBD was specifically tailored to examine its journey through the body under the assumption of oral administration. The analysis focused on several key stages: CBD's absorption through the gastrointestinal tract, its distribution in the maternal bloodstream, the transplacental passage into the foetal bloodstream and eventual reach to the foetal brain. This detailed evaluation of CBD's pathway through these physiological stages is of paramount importance for assessing its safety during pregnancy. It provides a comprehensive perspective on how CBD might affect not only maternal health but also foetal development. Such an understanding is essential for making informed decisions about the potential therapeutic use of CBD, especially considering the possible implications for neurodevelopment and disorders like ASD.

After importing the 3D structure of CBD (CID:644019) into Schrödinger software suite's Maestro, QikProp generated the ADMET properties (QikProp, version 4.4, Schrödinger) listed in Table 2. These properties include #stars, CNS activity, volume, QPlogPC16, QPlogPo/w, QPPCaco, QPlogBB, QPPMDCK, QPlogKhsa, RuleofThree and RuleofFive.

Table 2: Table Describing the ADMET Properties of CBD

ADMET Property	**Value**	**Description**	**Acceptable values**	**Analysis**
#stars	1	Assesses a molecule's drug-likeness	0-5	CBD is a suitable drug-like molecule
CNS	0	Predicted central nervous system activity	–2 (inactive) to +2 (active) scale.	CBD has limited CNS activity
Volume	1157.7		N/A	CBD occupies a volume of 1157.7 Å^3
QPlogPC16	10.5	A measure of hydrophobicity	4.0 – 18.0	CBD is highly hydrophobic
QPlogPo/w	5.3	Provides information about lipophilicity or hydrophobicity	–2.0 – 6.5	CBD is highly lipophilic
QPPCaco	2405.6	Predict the ability of a compound to permeate through the Caco-2 cell monolayer, a representation of intestinal absorption.	<25 poor, >500 great	CBS has excellent intestinal absorption
QPlogBB	-0.5	Estimates potential to traverse the blood-brain barrier. QikProp predictions are geared towards orally administered drugs.	–3.0 – 1.2	CBD has limited ability to transverse the blood brain barrier
QPPMDCK	1277.7	Apparent MDCK cell permeability in nm/sec. MDCK cells are a good mimic for the blood brain barrier.	<25 poor, >500 great	CBD can transverse the cell membranes efficiently

QPlogKhsa	1.1	Prediction of binding to human serum albumin.	- 1.5 to 1.5	CBD has moderate affinity for binding to Human serum albumin
RuleOfFive	1	Count of deviations from Lipinski's rule of five, which encompasses criteria: molecular weight below 500, QPlogPo/w less than 5, no more than 5 hydrogen bond donors, and no more than 10 hydrogen bond acceptors.	An orally active drug-like compound should not have more than one violation of the following criteria as stated by Lipinski's rule:	CBD violates 1 Lipinski rule
RuleOfThree	2	Count of infringements against Jorgensen's rule of three, QPlogS greater than -5.7, QP PCaco exceeding 22 nm/s, and a count of primary metabolites less than 7.	Compounds with fewer, preferably zero, breaches of these standards tend to have better potential for oral availability.	CBD violates 2 Jorgensen’s rules

#stars, indicating the drug-likeness of CBD, evaluates various factors like molecular weight, dipole moment, polarizability, lipophilicity, solubility and more. CBD score of 1 implies that CBD diverges from the standard range in only one descriptor, a range that 95% of established pharmaceuticals typically adhere to. (Jorgensen and Duffy., 2000). Furthermore, Lipinski's Rule of 5 (Ro5) has served as a widely acknowledged guideline for assessing drug-like properties (Lipinski., 2000). This rule, considered a general "rule of thumb," provides a valuable framework for predicting the Absorption, Distribution, Metabolism and Excretion performance of drugs, particularly those intended for oral administration. CBD exhibits a partial violation of Lipinski's Rule of Five, specifically by surpassing the criteria for fewer than 5 hydrogen bond donors (donorHB) and less than 10 hydrogen bond acceptors (accptHB), suggesting potential challenges in achieving oral drug success. The rule is formulated on the observation that most therapeutic drugs are small and highly lipophilic molecules.

CBD exhibited distinct hydrophobic characteristics, as evidenced by QPlogPC16 and QPlogPo/w values of 10.512 and 5.32, respectively. These metrics provided precise measurements of CBD's affinity for lipid environments, indicating its inclination toward hydrophobic interactions. The assessment of CBD's permeability across Caco-2 cell monolayers, a model for intestinal absorption, yielded a high QPPCaco value of 2405.6. This outcome signifies favourable permeability through the gastrointestinal tract, suggesting efficient absorption of CBD across the intestinal wall and into the bloodstream. Predicted binding affinity to human serum albumin (HSA) is high (QPlogKhsa = 1.0523), suggesting CBD forms stable interactions with HSA, which could lead to prolonged circulation time in the bloodstream and therefore increase the likelihood of interaction with other tissues. The QPlogBB value of -0.496, signifying the Quantitative Prediction of Blood-Brain Barrier Penetration, prompts consideration that CBD is predicted to exhibit restricted capability to transverse the blood-brain barrier. This aligns with a central nervous system activity value of 0, further suggesting limited activity in the central nervous system. In contrast, CBD demonstrates excellent cell membrane penetration capabilities as indicated by QPPMDCK (1277.658), which serves as a model for the BBB. These contrasting results highlight the challenges in predicting BBB permeability accurately and may arise from differences in the predictive models used for these parameters.

Finally, considering the recorded volume of CBD is 1157.7 Å³ and the predicted volumes of the larger binding pockets within the TBR1 T-box domain (forecasted by CASTp3.0, Protein Plus, and SiteMap) range from 91.819 Å³ to 2000.665 Å³, the substantial size difference and the

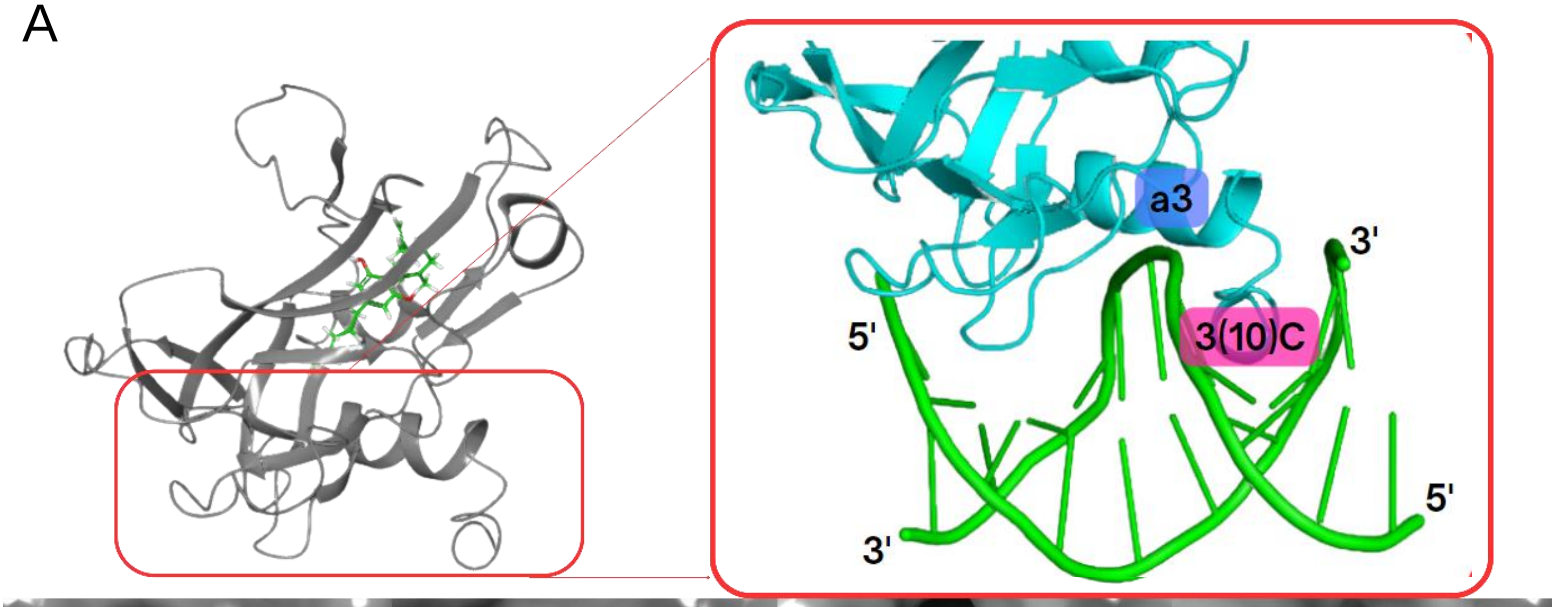
predominantly hydrophilic nature of these pockets raise doubts about the feasibility of CBD binding within them. This concern prompts a more thorough evaluation through docking simulations focused on the entire DNA binding domain.

4.4 Molecular Docking

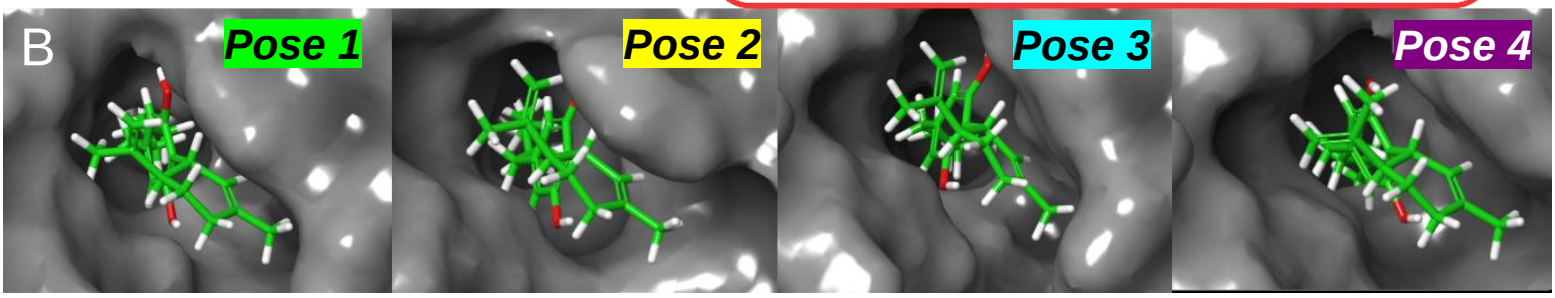
After evaluating the drug-like properties of CBD and conducting energy minimisation of both CBD and the AlphaFold-predicted TBR1 T-box domain, an IFD procedure was implemented using the OPLS_2005 force field. The initial step involved ensuring the structural integrity of the TBR1 T-box domain, which confirmed the absence of steric clashes and validated the quality of the input structure. The structural metric further confirmed the refined quality, with an average bond angle RMSD of 2. 06 ° and average bond length RMSD of 0.02 Å – signifying minimal adjustments needed for optimisation. The average peptide planarity measured 173.17 ° and the average side chain planarity was 0.009, denoting a flat and planar arrangement for the backbone and side chains, respectively. No missing atoms were identified, further contributing to the reliability of AlphaFold-predicted structures. However, these also highlight the potential benefits of further refinement and energy minimisation before proceeding to docking simulations.

Preliminary identification of potential binding sites within the TBR1 T-box domain aimed to inform the selection of the most appropriate region for the docking grid box. Despite these efforts, the examination of CBD's ADMET properties highlighted a potential challenge: the hydrophobic nature of CBD might hinder its interaction with the predominantly hydrophilic predicted binding domains. Therefore, to accommodate this possibility and the lack of specific literature on CBD's exact binding site, the entire T-box domain was included within the docking grid box rather than targeting a specific section. Molecular docking was conducted in a pH-dependent ionisation mode to mimic the physiological conditions of the human body. The most favourable docking poses, selected by the software, are graphically depicted in Figure 20.

A

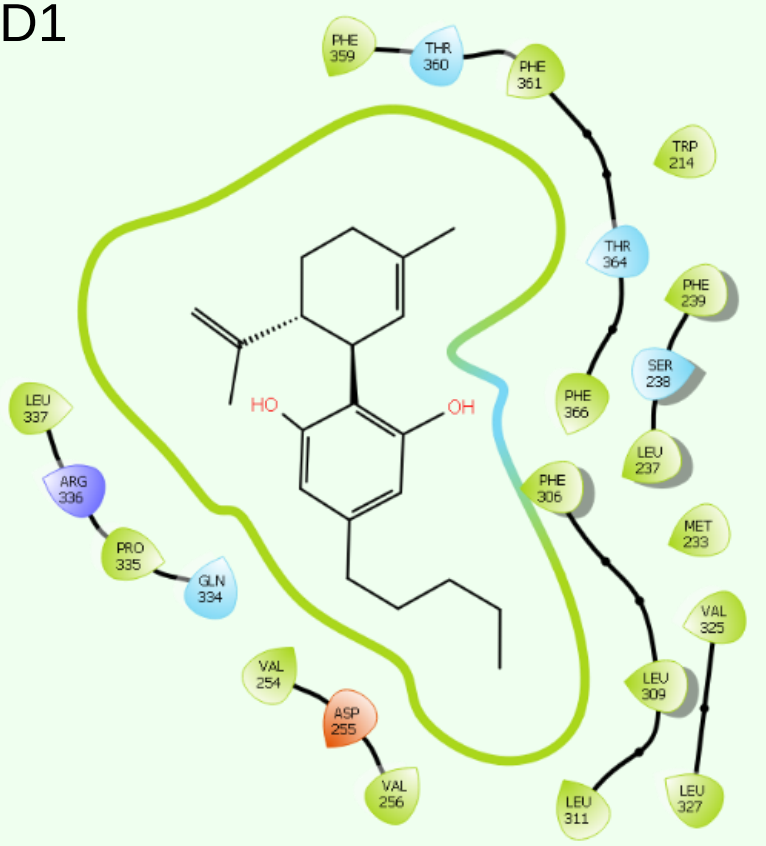
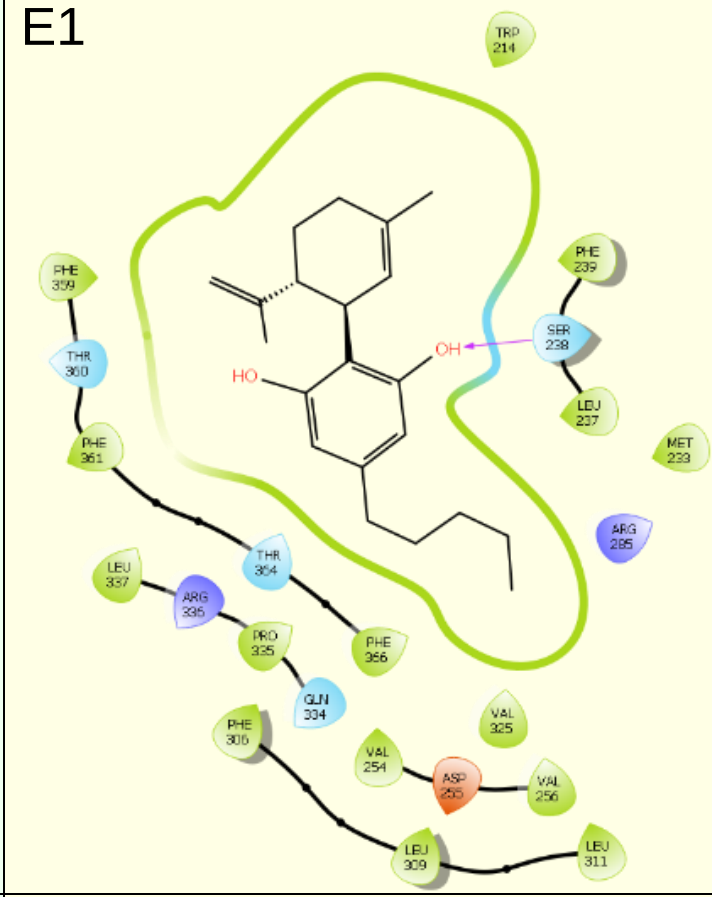
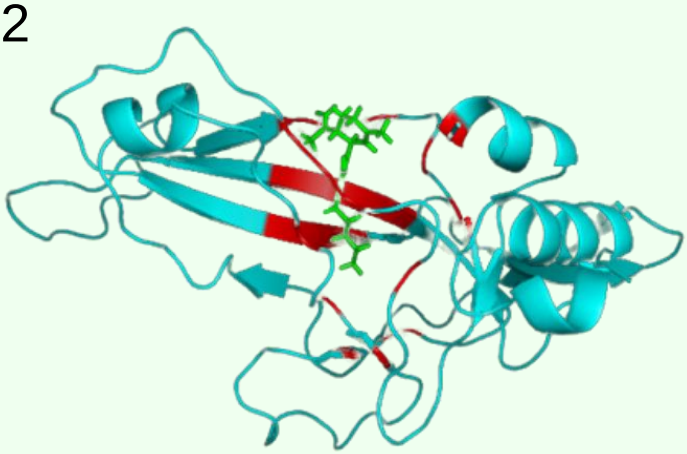
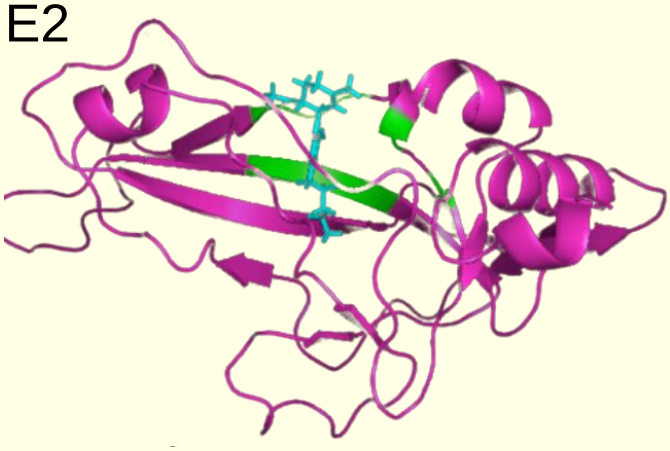


B



C

	Docking Score (kcal/mol)	Glide score (kcal/mol)	Glide emodel	Glide energy	IFDScore (kcal/mol)
Pose 1	-9.432	-9.432	-63.410	-8121.3	-415.50
Pose 2	-9.823	-9.823	-63.307	-8112.7	-415.46
Pose 3	-9.514	-9.515	-62.886	-8106.5	-414.84
Pose 4	-8.495	-8.496	-58.265	-8108.4	-413.92

Pose 1	Pose 2
D1 	E1 
D2 	E2 
Pose 1: No direct interaction between CBD and TBR1 T-box doman.	**Pose 2: Hydrogen bond formation between Serine 238 of TBR1 and CBD's hydroxyl group (OH).**

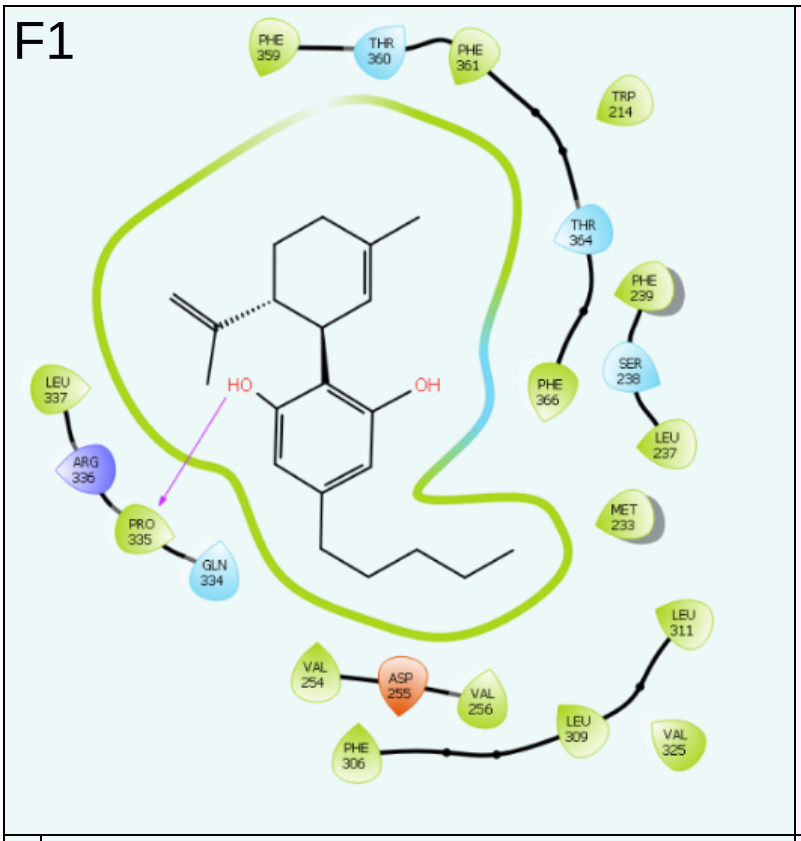
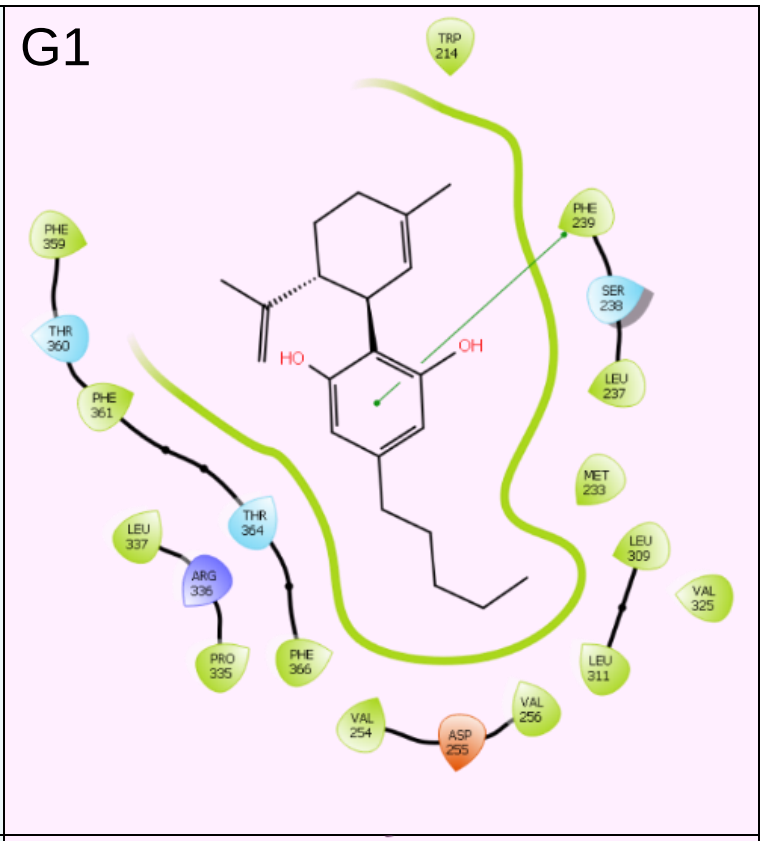
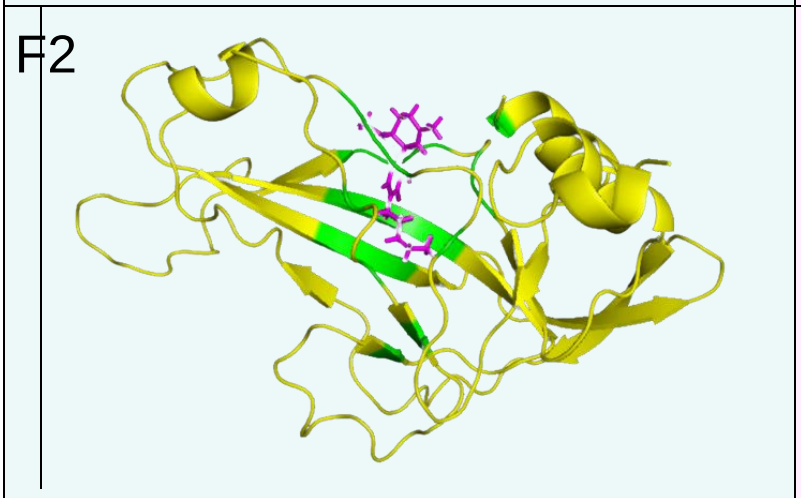
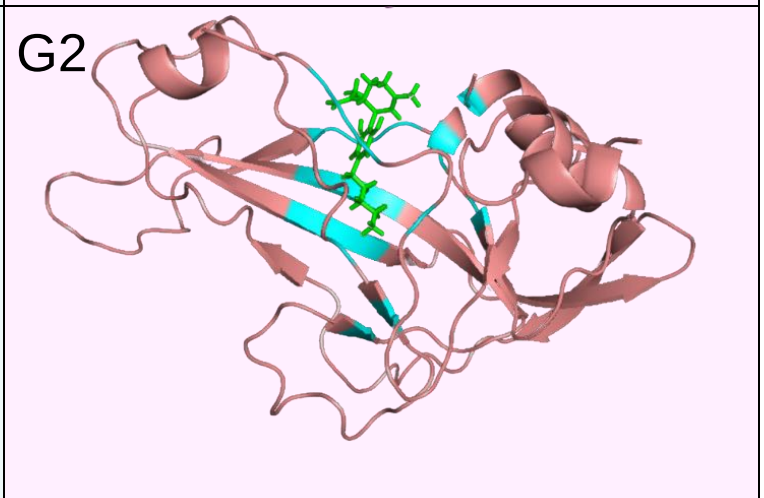
Pose 3	Pose 4
	
	
Pose 3: Hydrogen bond formation between Proline 335 of TBR1 and CBD's hydroxyl group (OH).	**Pose 4: π-π interactions between Phenylalanine 239 of TBR1 and the central phenol ring in CBD**

Figure 20: Various Poses Generated from the Induced Fit Docking between TBR1 T-box Domain and CBD. A: TBR1 T-box Domain is presented in grey, while the CBD molecule docked within it is shown in green. The red block highlights the DNA binding site of the TBR1 T-box domain, with the close-up view on the right showcasing the interacting secondary structures, specifically helices e', a3, and 3_{10}C. B: This series of images illustrates the binding poses of CBD within the predicted binding pocket of the TBR1 T-box

domain, ranked from highest to lowest, arranged from left to right. These images vividly display the conformational alterations of both CBD (in green) and the binding pocket (in grey). C: Docking Scores for TBR1 T-box Domain and CBD as Calculated using the Induced Fit Protocol in Schrödinger's Maestro. The figures, labelled D to G in both 2-dimensional (2D) and 3-dimensional (3D) formats, present the top docking poses selected by Maestro. These poses are listed in descending order of ranking, from D to G, with the highest-ranked pose labelled as D. In the 2D representations (D1-G1), CBD's structure is centred, encircled by protein residues. Residues outside the threshold range are depicted as black dots. Teardrop symbols illustrate the amino acid residues of the TBR1 T-box Domain, with the direction of the teardrop indicating the orientation of the residue side chain. Lines surrounding the CBD ligand are color-coded to match the nearest residue's colour. These lines represent the interaction formed between CBD and TBR1 t-box domain. Gaps in these lines signify solvent exposure within the pocket. Non-polar amino acids are displayed in green, neutral polar amino acids in blue, positively charged polar residues in red, and negatively charged polar residues in purple. Direct interactions between CBD and the TBR1 T-box domain are represented by lines connecting the ligand and the protein, signifying hydrogen bonds, salt bridges, pi-pi stacking, pi-cation interactions, and metal-coordination's. Figures D2-G2 depict the 3D structures of TBR1 T-box domain in the top 4 poses ranked from D2, to G2. The amino acids that stabilise the interaction with CBD are coloured differently), in its varying conformations within the binding pocket (PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC; Induced Fit Docking, Schrödinger, LLC; Farid *et al.,* 2006; Sherman *et al.,* 2006; Sherman *et al.,* 2006).

Various docking poses were generated to investigate the binding interactions between CBD and TBR1 T-box Domain using the IFD Protocol in Schrödinger's Maestro software. These docking poses were assessed and ranked based on several scoring parameters to gain insights into the binding affinity and structural compatibility of CBD with the protein. One such parameter considered was the glide score, which is measured in kcal/mol (Induced Fit Docking, Schrödinger, LLC).

Pose 2 emerged as the most favourable, evidenced by its best glide score of -9.823, suggesting a strong binding affinity. This is likely due to the formation of a hydrogen bond between the hydroxyl group (OH) of Ser238 in the TBR1 T-box domain and the OH group of CBD. Here, the hydrogen atom in serine acts as a donor and the oxygen in CBD as an acceptor. Pose 3, while also forming a hydrogen bond with Proline 335, yielded a slightly less favourable glide score of -9.514. This difference in affinity could be attributed to the rigid nature of Proline's cyclic structure, affecting the interaction dynamics. Pose 4 ranked lowest in terms of glide score, showcases a π-π interaction between the benzene ring in Phe239 of the TBR1 T-box domain and CBD's central phenol ring. In this interaction, the aromatic rings of phenylalanine and CBD align in a parallel or near-parallel fashion. The interaction is primarily governed by dispersion forces, which result from fluctuations in electron distribution around the aromatic rings. π electrons from the aromatic rings induce dipoles in one another, leading to an attractive force. Conversely, Pose 1, despite showing no direct interaction, still achieved a notable position in the rankings, possibly due to other

favourable non-bonded interactions or solvent effects. The analysis also revealed that in Pose 1, CBD is surrounded by hydrophobic regions of the protein, as opposed to Pose 4, where it is more exposed to the solvent. This finding is crucial as it suggests that hydrophobic interactions play a significant role in determining the stability of the CBD-protein complex. The glide scores are comparable to the docking scores, considering both are in kcal/mol. For most poses, both scores align, except pose 3 and 4 which display a -0.001 Epik state penalty, resulting in a lower docking score. The negative penalty suggests energetically favourable ionisation/tautomeric states, reducing the overall energy cost (Sastry *et al.,* 2013). Poses 1 and 2 exhibit a stable state for CBD, reflecting a penalty of 0 and no additional energy cost for these specific states.

The Glide Emodel assigns a substantial weight to forcefield parameters, facilitating the comparison of molecular shapes and arrangements within the binding site. Combining GlideScore with nonbonded interaction energy and excess internal energy for flexible docking, Emodel assesses the overall energetics and stability of ligand-protein complexes during molecular docking. A negative Emodel score signifies a more favourable interaction, with lower values indicating higher stability. In the ranking of Glide Emodel scores (-63.410 for Pose 1, -63.307 for Pose 2, -62.886 for Pose 3, and -58.265 for Pose 4), Pose 1 achieves the highest score, indicating the most energetically favourable conformation despite lacking direct interactions. Pose 1 may achieve optimal packing within the hydrophobic pocket, minimising steric clashes and maximising favourable nonbonded interactions, such as van der Waals forces. While Pose 1 lacks direct interactions, it may also avoid unfavourable interactions, such as steric clashes or repulsive forces, contributing to its overall stability. Pose 2 closely follows with a slightly lower score but benefits from hydrogen bonding. Pose 3 exhibits moderate stability with a score slightly less favourable than Poses 1 and 2. Conversely, Pose 4, with the lowest Emodel score, represents the least stable conformation. The numerical values provided (-8121.3, -8112.7, -8106.5, -8108.4) represent the Glide energy scores for Poses 1 to 4. Glide Energy is focused on the overall interaction energy, considering the total energy associated with the ligand-protein complex. More negative values correspond to stronger and more favourable binding interactions, with Pose 1 exhibiting the most negative score. Finally, the IFDScore was used to assess the overall fit of CBD within the binding site. Maestro ranks poses 1 to 4 based off the IFDScore (-415.50 for Pose 1, -415.46 for Pose 2, -414.84 for Pose 3 and -413.92 for Pose 4) from most to least favourable binding interaction (more negative to less negative).

Each of the four poses (Pose 1 to Pose 4) prominently exhibit CBD's engagement in the hydrophobic binding pocket, facilitated by the insertion of CBD's alkyl side chain. Table 3 provides a comprehensive overview of conformational variances observed in both the native CBD and CBD within each docking pose, as well as in the native DNA binding domain in comparison to each docking pose, by evaluating the Cα-RMSD value.

Table 3: Cα-RMSD Comparison Between Native and Docked Ligands, and Native and Docked T-Box Domain Poses

	RMSD (Ligand)	RMSD (Protein)
Pose 1	0.4974	1.7591
Pose 2	0.9912	1.6393
Pose 3	1.0843	1.7467
Pose 4	1.3226	1.6718

In the absence of a crystallographically solved structure, the RMSD in Table 3 measures the structural differences between the AlphaFold predicted structure of the T-Box domain (termed native in this context) and the docked T-Box domains for each pose obtained from Maestro. All 4 poses recorded minimal deviation (RMSD $\leq$ 2 Å). Upon visual inspection, the secondary structural elements of the T-box domain remained stable, with minimal deviations (Asn294 and Thr295) observed in the positions of random coils situated near the predicted binding pocket. When comparing the RMSD between the native CBD ligand and docked CBD ligands, all poses exhibited an RMSD $\leq$ 2 Å. Pose 1 and Pose 2 achieved an even more favourable alignment, with RMSD values ≤ 1 Å. Therefore, CBD experiences minor structural changes upon binding. While analysing RMSD values with the crystal structure of CBD within the T-box domain could offer valuable insights, it's important to acknowledge that the solved structure of CBD represents its most stable conformation and may be indicative of CBD being well accommodated within the T-box domain. For a more in-depth exploration of stability, molecular dynamics provides a comprehensive view by illustrating how the interaction between CBD and the T-box domain persists over time.

4.5 Molecular Dynamics: Post Simulation Analysis

The Molecular Dynamics (MD) simulations were conducted for the best-scoring IFD poses (Pose 1 to 4) between CBD and TBR1 T-box domain, in comparison to the unbound TBR1 T-box domain (APO). To assess the quality of these simulations, a comprehensive analysis was performed, including Protein Cα and LP-RMSD, Protein-Ligand Contacts, Protein Residue Cα-RMSF, ΔRg and Change in Secondary Structure. These analyses provide detailed insights into the structural dynamics, binding interactions, local variations, overall compactness and secondary structure alterations between CBD-bound and APO states.

4.5.1 Simulation Quality Analysis

Upon the completion of the 250-nanosecond MD simulation for both the unbound TBR1 T-box domain (APO Protein) and the four previously CBD-docked systems (Pose 1–4), an analysis of simulation quality was provided, seen in Table 4. The energy, potential energy, temperature, pressure and volume over the entire 250 ns MD simulation of all systems, can be seen in Table 4.

Table 4: Molecular Dynamics Simulation Quality Analysis After Equilibration

Model System	**Parameter**	**Average**	**Std. Dev.**	**Slope (ps^{-1})**
Total Energy (kcal/mol)	APO Protein	-72398.157	72.123	-0.000
	Pose 1	-72229.713	70.966	-0.000
	Pose 2	-72292.120	70.862	-0.000
	Pose 3	-72184.797	72.721	-0.000
	Pose 4	-72291.664	71.315	-0.000
Potential Energy (kcal/mol)	APO Protein	-88856.439	61.982	-0.000
	Pose 1	-88703.204	60.768	-0.000
	Pose 2	-88775.881	61.091	-0.000
	Pose 3	-88643.644	62.839	-0.000
	Pose 4	-88779.032	61.222	-0.000

Model System	Parameter	Average	Std. Dev.	Slope (ps^{-1})
Temperature (K)	APO Protein	298.734	0.643	-0.000
	Pose 1	298.735	0.641	0.000
	Pose 2	298.728	0.637	0.000
	Pose 3	298.728	0.642	-0.000
	Pose 4	298.728	0.642	0.000
Pressure (bar)	APO Protein	1.063	60.062	-0.000
	Pose 1	1.086	60.641	0.000
	Pose 2	1.394	60.470	0.000
	Pose 3	1.465	60.360	0.000
	Pose 4	0.470	60.004	0.000
Volume ($Å^3$)	APO Protein	269805.674	313.922	-0.000
	Pose 1	269937.148	325.991	0.001
	Pose 2	270066.700	315.875	-0.000
	Pose 3	269593.503	315.164	-0.000
	Pose 4	270141.753	314.278	0.000

Temperature (K), representing the system's kinetic energy, remained nearly constant (298.73 K) for both the APO protein and various docking poses. This confirms that the temperature was uniform during the simulations, as indicated by the small standard deviations observed in Table 4. However, the pressure varied across all systems, as shown by a standard deviation of approximately 60 atm. Such variability is not uncommon in NPT simulations, where volume adjustments are permissible to sustain an average constant pressure, akin to Maestro's default value of 1.101325 bars. All the systems share a similar average pressure, except for Pose 4 (0.470 bars), which may reflect dynamic changes within the system. Importantly, a consistent slope of 0 was observed in all simulations, which signifies that the system had reached equilibrium

and stability. Total energy (kcal/mol), which encompasses the overall energy, remained relatively stable across the APO protein and four docking poses (Pose 1 - Pose 4), with values ranging from approximately -72184.797 to -72398.157 kcal/mol. The standard deviation was relatively small (¬70 kcal/mol) compared to the overall scale of the total energy, which indicates that the system had reached equilibrium and stability, along with the slope of 0. The potential energy (kcal/mol), responsible for particle interactions, displayed a similar pattern, maintaining stability between -88643.644 and -88775.881 kcal/mol for Pose 1 through Pose 4, indicating a well-behaved system. Lastly, the volume ($Å^3$) parameter representing occupied space, revealed relatively consistent volumes for the APO protein compared to the docked poses.

Overall, the simulations are well-behaved showing consistent energy and temperature profiles across all the systems. All the systems reached equilibrium and are stable, indicating reliability and suitability for further analysis.

4.5.2 Molecular Dynamics Analysis

In earlier docking simulations, four distinct binding poses emerged when examining the interaction between CBD and the TBR1 T-box domain. Traditionally, researchers tend to opt for the top-scoring pose (Pose 1) for subsequent molecular dynamic simulations. Nevertheless, a more comprehensive understanding of the potential interactions and stability between CBD and the TBR1 T-box domain can be gained by analysing all four poses. This comprehensive analysis incorporates various parameters:

1. **Protein Cα and Ligand Fit-to-Protein RMSD:** This provides insights into the structural conformation of the protein and ligand upon binding.
2. **Protein-Ligand Contacts:** Describing the type and distribution of bonds formed during binding.
3. **Protein Residue Cα-RMSF:** Assessing local variations across the protein chain.
4. **Radius of Gyration Difference:** Offering information on the overall compactness or spread of the protein structure between the CBD-bound and APO poses.
5. **Change in Secondary Structure:** Examining alterations in secondary structure between different CBD-bound poses compared to the APO protein.

This collectively contributes to a comprehensive assessment of the MD simulations and their underlying binding interactions, facilitating a robust and insightful analysis of the model systems' behaviour and stability.

4.5.2.1 Protein Cα and Ligand-Fit-to Protein Root Mean Square Deviation (RMSD)

The RMSD is a cornerstone metric in the assessment of biomolecular structural dynamics, as it calculates the average change in atomic displacement between a protein and a reference frame, across the entire trajectory. The RMSD calculated over 250 ns for the protein backbone (Cα-RMSD) and for the ligand bound to the protein (ligand-fit-to protein) is shown in Figure 21 below.

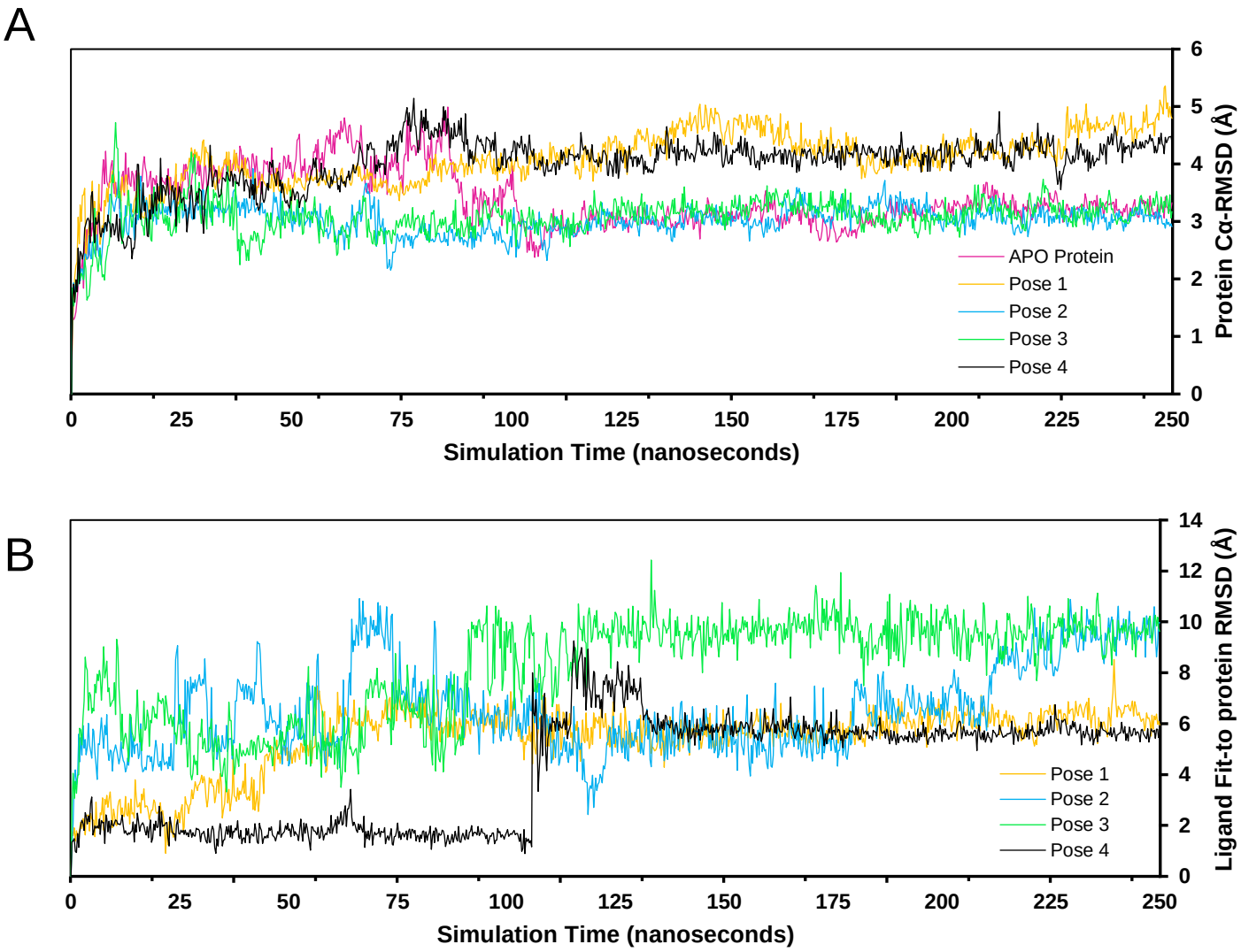


Figure 21: Protein Cα and LP-RMSD Plots for TBR1 T-box domain and CBD for the APO and CBD-docked TBR1 T-box domain systems. A: Protein backbone Cα-RMSD calculated for the APO protein (pink), CBD-docked Pose 1 (orange), Pose 2 (blue), Pose 3 (green) and Pose 4 (black) showing the average atomic deviation of Cα. The AlphaFold predicted structure of TBR1 T-box domain is used as reference. B:LP-RMSD calculated for Pose 1 (orange), Pose 2 (blue), Pose 3 (green) and Pose 4 (black). The CBD reference structure was obtained from PubChem (CID: 644019). Greater RMSD values correspond to greater deviations from the reference structure and conformational changes – reflecting stability of the complex. Graphs were generated using MS Excel and RMSD was calculated using Maestro Schrödinger.

As a reference, the APO Cα-RMSD represents the stability of TBR1 T-box domain in the absence of ligands. It shows a larger Cα-RMSD (between 4 Å and 5 Å) in the initial 90 ns of the simulation, followed by a 10 ns drop to ¬3 Å for the remainder of the simulation. Fluctuations in the range of 1-3 Å are entirely acceptable when analysing smaller proteins (Desmond, Schrödinger, LCC,

2023), especially when focusing solely on the DNA binding domain of TBR1 rather than the entire protein. This RMSD fluctuation corresponds to structural changes in the protein's backbone, whereafter the structure may have stabilised to a more consistent conformation. While the APO Cα-RMSD never attains the conventionally accepted value of ≤2 Å (Castro-Alvarez *et al.,* 2017), this does not pose an issue, considering the substantial dependence of RMSD on the choice of the reference structure—in this case, the AlphaFold predicted structure of TBR1 T-box domain. Deviations from the reference structure may not inherently signify instability or inaccuracies in the simulation, given that proteins inherently display natural flexibility and undergo conformational changes. The APO protein stabilises at ¬105 ns with an average Cα-RMSD value of 3.38 ± 0.52 Å.

Pose 2 and 3 exhibit the lowest average Cα-RMSD values of 3.02 ± 0.28 Å and 3.07 ± 0.33 Å, respectively. These poses closely resemble the AlphaFold predicted structure, after its structural fluctuations stabilise after 75 ns - indicating minimal structural changes upon ligand binding. The LP-RMSD in Figure 21B shows overall high variability in the ligand conformation upon binding – suggesting the ligand undergoes more significant conformational changes upon protein binding than the protein itself. It is evident that, after an initial period of approximately 130 ns, each pose reaches a relatively stable equilibrium. However, Pose 2 deviates from this trend, showing significant fluctuations around the 210 ns mark, resulting in a substantial RMSD increase from 5 Å to 10 Å. The considerably larger LP-RMSD values observed in comparison to the Cα-RMSD suggest that the ligand has diffused away from its initial binding site toward the end of the simulation. While Pose 2 exhibits an average LP-RMSD of 6.55 ± 1.73 Å, it does not hold the highest average; Pose 3 surpasses this with an average LP-RMSD of 8.15 ± 2.00 Å.

In comparison, CBD-docked Poses 1 and 4 demonstrate a somewhat elevated Cα -RMSD, ranging between 4 Å and 5 Å, while maintaining stability within this interval. Despite Pose 1 initially representing the most stable docked conformation, there is an observed increase in its Cα-RMSD from approximately 3 Å to 5 Å over the course of the simulation, with an average Cα-RMSD value of 4.09 ± 0.49 Å. Interestingly, the LP-RMSD of Pose 1 indicates a relatively constant ligand conformation after the first 60 ns of the simulation, with an average LP-RMSD value of 5.34 ± 1.35 Å. Overall, Pose 1 represents the highest Cα-RMSD, which suggests that the ligand might induce increased flexibility and dynamics within the protein-ligand complex. Pose 4 exhibited consistent RMSD stability until the 75 ns mark, where a notable increase from 4 Å to 5 Å occurred. Subsequently, at 90 ns, the RMSD decreased to 4 Å and maintained this value for the remainder

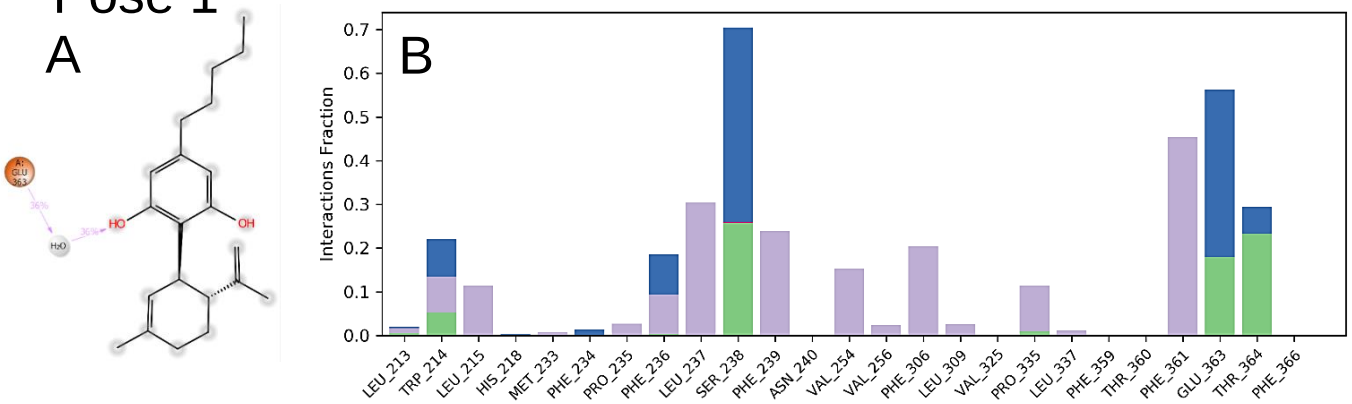
of the 250 ns simulation. The observed fluctuations in Pose 4 may signify a specific phase of interaction or adjustment between CBD and the TBR1 T-box domain. The average Cα-RMSD of Pose 4 is 4.00 ± 0.49 Å. The LP-RMSD remains most stable during the initial 100 ns, followed by a substantial increase from approximately 1 Å to 6 Å. Importantly, during this period, the protein Cα-RMSD does not exhibit significant changes, suggesting minimal structural fluctuations in response to the larger ligand structural changes. On average, the LP-RMSD of Pose 4 is 4.15 ± 2.15 Å, representing the lowest LP-RMSD of all the poses.

Ultimately, all four CBD-docked poses depict CBD situated in a singular binding pocket within the TBR1 T-box domain and reach equilibrium, indicating a stable interaction. Figure 22 highlights the specific protein residues interacting with the ligand, providing a more in-depth exploration of the formed bonds and the stability of these interactions.

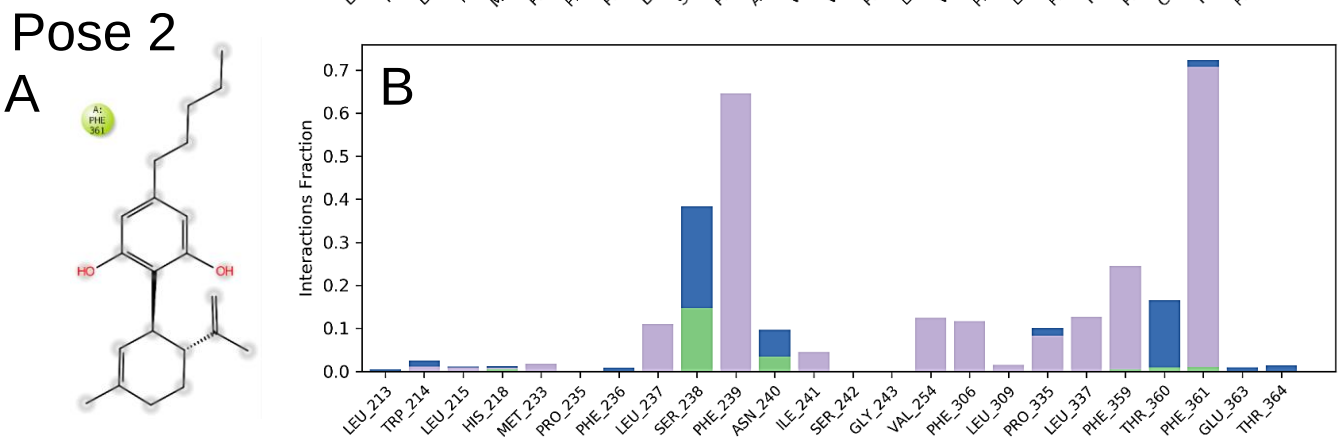
4.5.2.2 Protein-Ligand Contacts

Figure 22 depicts 2D ligand interaction diagrams, illustrating interactions that occurred in over 30% of the 250 ns simulation. Furthermore, it presents details on protein-interacting residues and the distribution of bond types across all four poses.

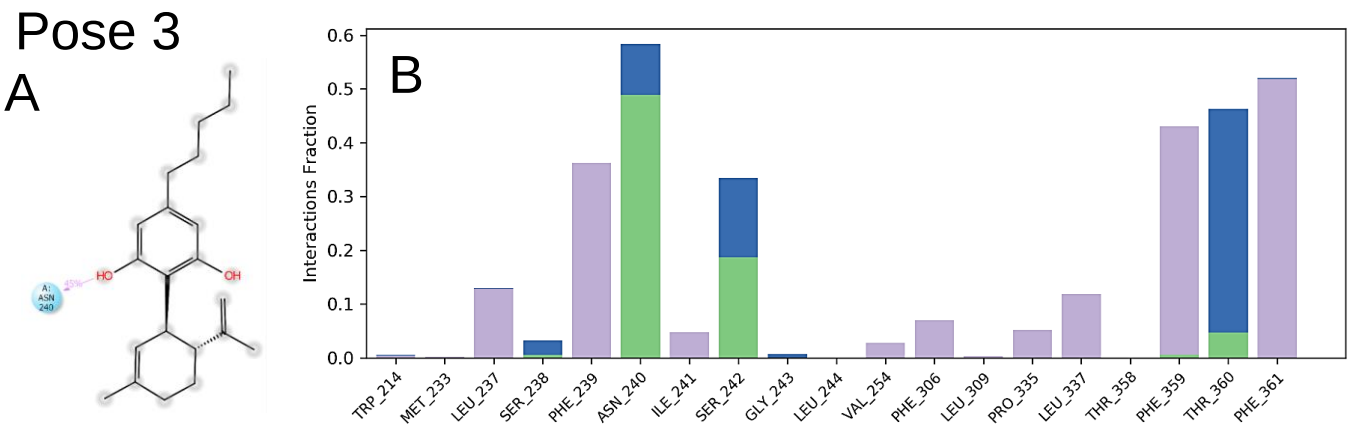
Pose 1



Pose 2



Pose 3



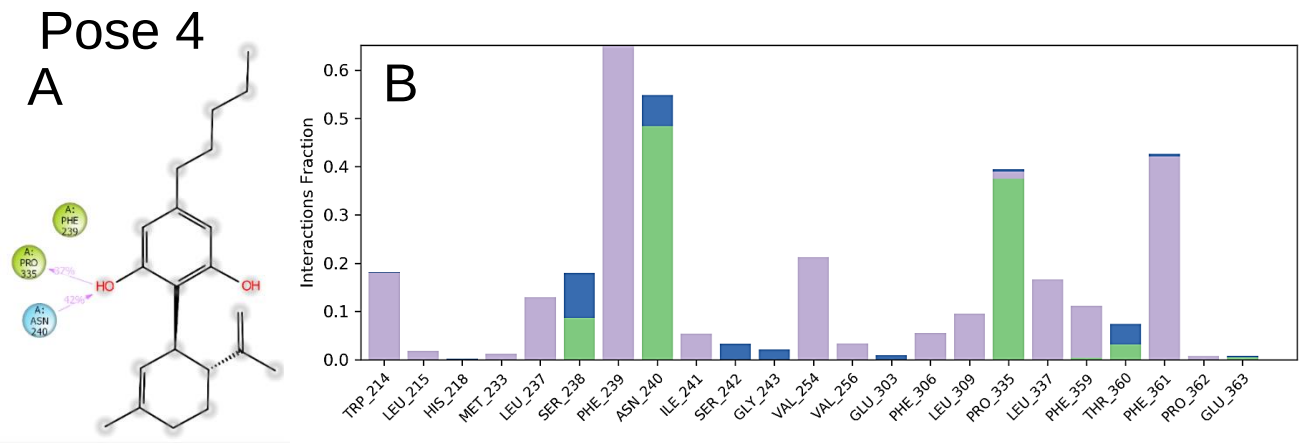


Figure 22: 2D Ligand Interaction Diagrams and Protein-Residue Bonds Distribution Across all Four Poses in 250 ns Simulation. A: The CBD ligand along with its interactions with the protein residues is observed, for the CBD-docked poses 1-4. The protein residues are represented by coloured circles, showing hydrophobic (green), negatively charged (red), and hydrophilic residues (blue). To highlight the significance of these interactions, the percentage of time they persist during the 250 ns simulation is displayed, with interactions exceeding 30% of the simulation time indicated in purple – representing hydrogen bonding. B: All the ligand-interacting protein residues with their respective bond distributions, including hydrogen bonds (green), hydrophobic interactions (purple), ionic bons (pink) and water bridges (blue). Images and graphs generated using Maestro Schrödinger.

Figure 22A highlights ligand interacting protein residues that persist for over 30% of the simulation time, although additional interacting residues exist throughout the simulation below this threshold (Figure 22B). Upon initial examination, it is evident that poses 1 through 4 predominantly manifest hydrophobic interactions with CBD, contributing to the stabilisation of CBD within a hydrophobic pocket in the protein structure. In Pose 1, a water-mediated hydrogen bond occurs between Glu363 and the hydroxyl group (OH) of CBD. This interaction involves the oxygen atom within the side-chain carboxylate group of Glu363 (sidechain hydrogen bond acceptor) and is sustained for over 30% of the simulation. Pose 1 also shows hydrogen bonding between Ser238, Thr364, Trp214, Pro335, Leu213, Phe236 and CBD, in order of decreasing interaction frequency. In Pose 2, Phe361 stabilises CBD through a hydrophobic interaction for majority of the interaction, along with Phe239 (although not shown in the 2D ligand interaction diagram). Additionally, the interaction is stabilised through hydrogen bonding - seen between Ser238, Asn240, His218, Thr360, Phe316 and Phe359 and CBD, in order of decreasing interaction frequency. In Pose 3, Asn240 interacts with the CBD OH group for 45% of the simulation time, as the side-chain NH group of Asn acts as a H-bond donor and the OH of CBD as the hydrogen bond acceptor. Additional hydrogen bonds are seen in Ser242, Thr360, Ser238 and Phe359 in order of decreasing interaction frequency. Pose 4 displays dominating hydrophobic interactions with Phe239 and additional hydrogen interactions with Pro335 and Asn246 for 37% and 42% of the

time, respectively. Specifically, the NH group in Asn serves as a side chain hydrogen bond donor to the OH of CBD. As a result, the OH group of CBD serves as a hydrogen bond donor and the side chain oxygen of Pro as the hydrogen bond acceptor. Additional hydrogen bonds are observed in Ser238, Thr360 and Glu363 in order of decreasing interaction frequency. A common trend in hydrogen bond formation is seen across poses 1 – 4 in the following protein residues: Ser238, Pro335, Thr360, Glu363 and Asn240.

To delve deeper into the dynamics and flexibility of the interactions observed, Figure 23 presents the RMSF of each residue in the protein, with specific focus on the those exhibiting hydrogen bonds common to all poses. This analysis aims to provide insights into how individual residues contribute to the overall structural stability and fluctuations observed in the interactions.

4.5.2.3 Cα-Root Mean Square Fluctuations (Cα-RMSF)

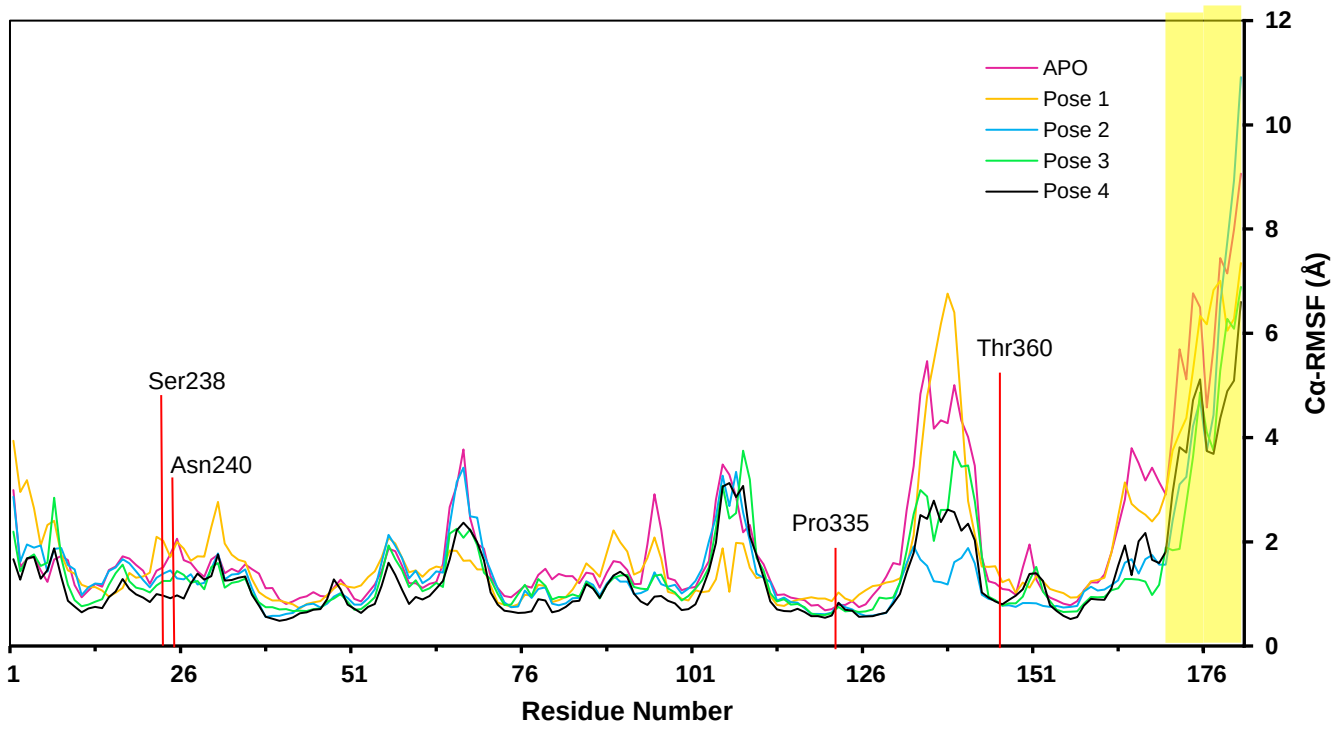


Figure 23: Cα-RMSF for each residue in the TBR1 T-box domain, emphasising residues involved in hydrogen bond formation with the ligand across the entire 250 ns simulation. The Cα RMSF value for each residue in the APO Protein (pink), Pose 1 (orange), Pose 2 (blue), Pose 3 (green) and Pose 4 (black) are shown. The CBD-interacting residues showing a common trend in the four poses are shown by red lines and are labelled. The yellow highlighted region represents the residues associated with direct DNA binding and thus functionality. The graph was generated using MS Excel based off calculations from Maestro Schrödinger.

The Cα-RMSF is a valuable metric for assessing local variations along the protein chain. This calculation considers the square distance averages for a specific residue over a given trajectory time. In Figure 23, larger peaks on the RMSF plot signify regions within the protein that exhibit the greatest fluctuations during the simulation. The biggest fluctuations are observed at the protein's tails (N- and C-terminal), where the DNA interacting residues are located (yellow highlighted). On average, the APO protein shows the greatest average Cα-RMSF value of 1.96 ± 1.49 Å, closely followed by Pose 1 with a value of 1.87 ± 1.41 Å. Although the variability in these values is large, the average fluctuations per residue decrease in Pose 2, Pose 3 and Pose 4 as well – with Cα-RMSF values of 1.53 ± 1.30 Å, 1.47 ± 1.04 Å and 1.36 ± 0.99 Å, respectively. The lower values observed in the CBD-docked poses show an overall decrease in local fluctuations of TBR1 T-box domain upon CBD binding – suggesting a more stable conformation. Typically, secondary structural elements such as α-helices and β-strands tend to be more rigid, resulting in reduced fluctuations compared to the less structured portions, like loop regions. This phenomenon is depicted in Figure 23, where notable fluctuations are evident in residues 66-67, 105-108, and 135-141, along with the protein terminals. These regions exhibit the highest RMSF values, indicating their active involvement in loop formations, which is confirmed in the 3D structure of TBR1 T-box domain.

Residues involved in hydrogen bond formation with the ligand, common in more than one simulation, are shown in red. This includes Ser238, Pro335, Thr360, Glu363 and Asn240. These residues are in low fluctuating areas, indicating these residues are in relatively stable regions with limited structural variations. By focusing on bonds that are common to all poses highlights stable and consistent interactions. None of these interacting residues are found in the regions predicted to interact with DNA (3_{10}C and α3 helix) and not found in highly variable regions (loops). This may indicate a rigid binding site that is well suited for the ligand. The influence of ligand binding on the overall compactness or spread of a protein can be analysed via the rGyr, offering a deeper insight into the conformational changes induced upon ligand binding (Figure 24).

4.5.2.4 Radius of Gyration (rGyr)

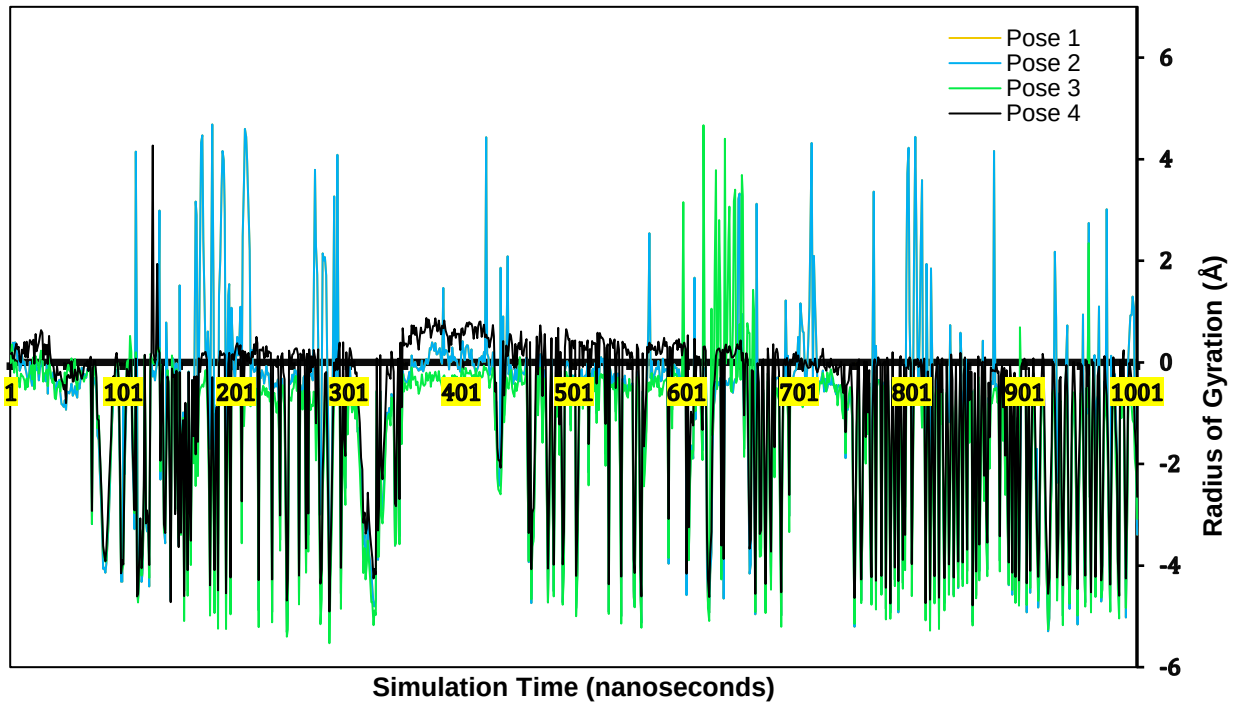


Figure 24: ΔRg between CBD-Bound and Unbound TBR1 T-box Domain over the 250 ns simulation. The rGyr for each pose is plotted as a difference of the APO protein: Pose 1 (orange), Pose 2 (blue), Pose 3 (green) and Pose 4 (black). Pose 1 overlaps exactly with Pose 2 and can therefore not be visualised on the graph. The Rg provides information about the overall structural response of protein to ligand binding, with negative values indicating a more compact structure and positive values indicating a more flexible structure. The graph was generated using MS Excel based off calculations from Maestro Schrödinger.

ΔRg between the CBD-docked poses (Pose 1 to 4) and APO TBR1 T-box domain is shown in Figure 24 The rGyr serves as a measure of the overall structural compactness or spread of the protein, and calculating the difference plot between ligand-docked poses and the APO form provides valuable insights by highlighting specific conformational changes induced by ligand binding. A positive ΔRg signifies an increase in the rGyr upon ligand binding, indicating heightened flexibility in the protein structure. Conversely, a negative ΔRg denotes a decrease in Rg, suggesting increased compactness following ligand binding. All the poses (Pose 1 to 4) show an overall negative ΔRg, indicating the protein experiences ligand-induced stabilisation in specific regions upon binding, highlighting the effectiveness of CBD binding in promoting structural compactness. On average, the Rg decreases by 0.92 ± 1.81 Å, 0.92 ± 1.81 Å, 1.32 ± 1.59 Å and 0.82 ± 1.54 Å for poses 1 to 4, respectively, with Pose 3 showing the largest average ΔRg. None of the poses show a stable Rg throughout and have a considerable degree of variability. The large

changes in ΔRg could be influences by alterations in the stability of secondary structural motifs. Therefore, the secondary structure analysis of the APO protein, compared to CBD-docked poses 1 to 4 are shown in Table 5.

4.5.2.5 Secondary Structure Analysis

Table 5: Secondary Structure Prediction after MD Simulations

Simulation	Helix (%)	Strand (%)	Coil (%)
APO Protein	4.83	28.83	66.34
Pose 1	4.04	28.95	67.01
Pose 2	2.92	28.88	68.20
Pose 3	2.82	27.33	69.85
Pose 4	3.86	29.42	66.72

Across all CBD-docked poses (Pose 1 to 4), there is a consistent decrease in helix content and a slight increase in coil content compared to the APO form. The increase in coil content suggests that CBD binding induces structural changes leading to a more extended or flexible conformation in the protein. Pose 3 exhibits the most substantial decrease in helix content and increase in coil content, suggesting a pronounced impact on the protein's secondary structure in this pose.

Objective 2: Determine the Potential Interaction between TBR1 T-box Domain and CBD using Experimental Approaches

The experimental phase of this study was initiated to augment an understanding of the potential interaction between CBD and the TBR1 T-box domain, complementing the insights gained from the bioinformatics segment. This investigation is of paramount importance, given the growing interest in CBD's usage during pregnancy and its potential implications on foetal development, particularly in the context of neurodevelopmental disorders like ASD. Existing research on CBD's safety during pregnancy is inconclusive, necessitating a comprehensive exploration of its molecular interactions. To comprehensively evaluate CBD's impact, a multifaceted experimental approach was employed, encompassing key experiments. These encompassed the expression and subsequent purification of the TBR1 T-box domain through IMAC and SEC. The purity of the isolated protein was assessed using SDS-PAGE. Additionally, CD spectroscopy was employed to analyse the protein's structural characteristics, ensuring its appropriate folding. Further experiments included a TSA, PDA and TR-FTIR to probe potential binding interactions between CBD and the TBR1 T-box domain. This multifaceted experimental approach underscores the necessity of empirical data to complement and validate the outcomes of the preceding bioinformatics analyses. Unfortunately, most of the experiments were not conducted in triplicate due to lack of CBD availability. This limitation resulted in the inability to conduct statistical analyses on the data.

4.6 Transformation and Expression

pET-11a plasmids containing inserts encoding the gene sequences of TBR1 T-Box domain were procured from GenScript (USA) and served as the basis for transforming T7 *E. coli* cells. A multiple sequence alignment between the codon optimised TBR1 T-box sequence and the sequence used for experimental work is shown in Figure 25. The TBR1 T-box domain was expressed under the induction of 0.2 mM IPTG for a duration of 20-22 hours at a controlled temperature of 20 ℃.This incubation period was chosen to ensure an adequate yield of protein for subsequent structural and functional investigations.

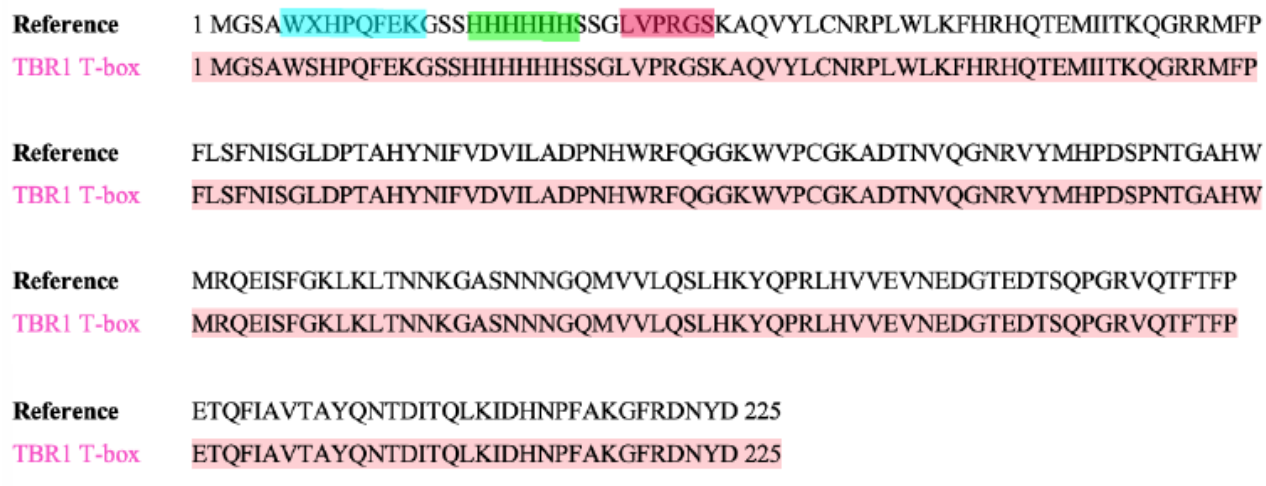


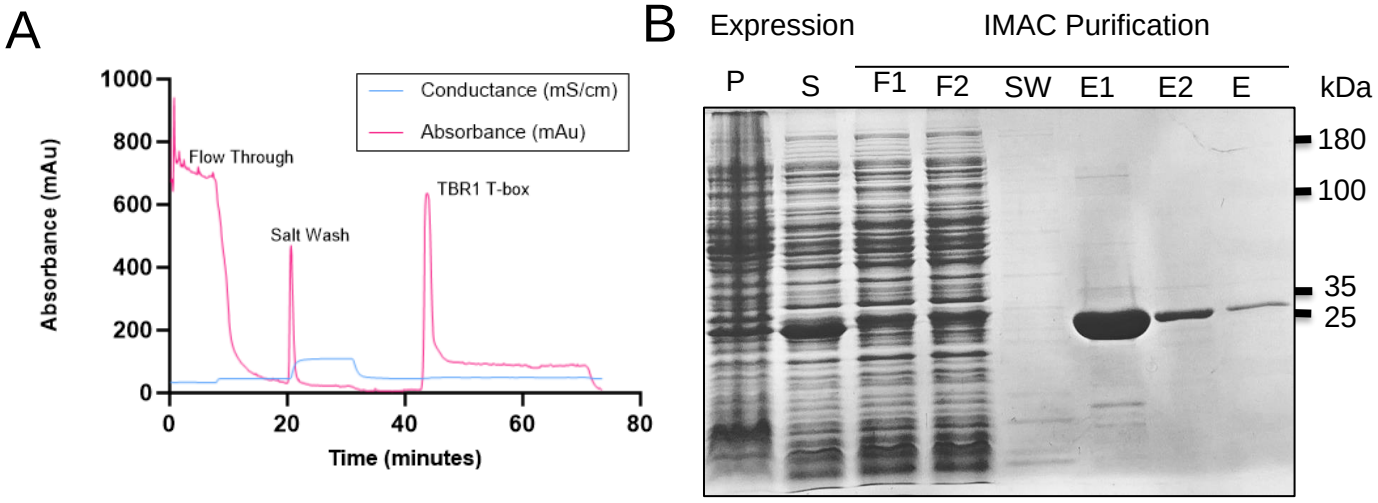
Figure 25: Codon Optimised Sequence and Experimental Sequence of TBR1 T-box Domain Sequence Alignment. Sanger sequencing, conducted by Inqaba Biotech (RSA) confirmed the accurate presence of the TBR1 T-box DNA sequence. The codon optimised sequence is represented as 'reference' and the experimental sequence used is represented as 'TBR1 T-box'. Visualisation of these sequences was accomplished using FinchTV software and subsequently translated using ExPasy. A sequence alignment, using Clustal Omega (EMBL-EBI), exhibited a 100% similarity. Noteworthy features, such as the Strep tag highlighted in blue, the histidine tag in green, and the thrombin cleavage site in red, are shown. This information adds to the assurance of the DNA sequences accuracy and underscores the reliability of the experimental data.

4.7 Purification and Protein Purity

After successfully expressing the TBR1 T-box domain protein, it underwent purification through two vital techniques: Immobilised Metal Affinity Chromatography (IMAC) and Size Exclusion Chromatography (SEC). Purification eliminates impurities and contaminants that may be present in the protein sample after expression. These impurities can interfere with subsequent experiments and structural analyses, potentially leading to inaccurate results. The TBR1 T-box domain protein, like other biomolecules, can exist in various conformations and states of aggregation. Purification via SEC allows us to isolate the specific conformation or oligomeric state of interest, ensuring that subsequent experiments are conducted with a homogenous protein population. SDS-PAGE was employed to evaluate the purity of the protein by separating it based on size and visualising any impurities or contaminants.

4.7.1 Immobilised Metal Affinity Chromatography (IMAC)

The soluble cell lysates, following the overexpression of TBR1 T-box, were introduced into the IMAC column. Multiple fractions were gathered, comprising the flow-through, salt wash, and eluted proteins. Figure 26 illustrates the elution pattern of TBR1 T-box and a gel showing the purification profile.



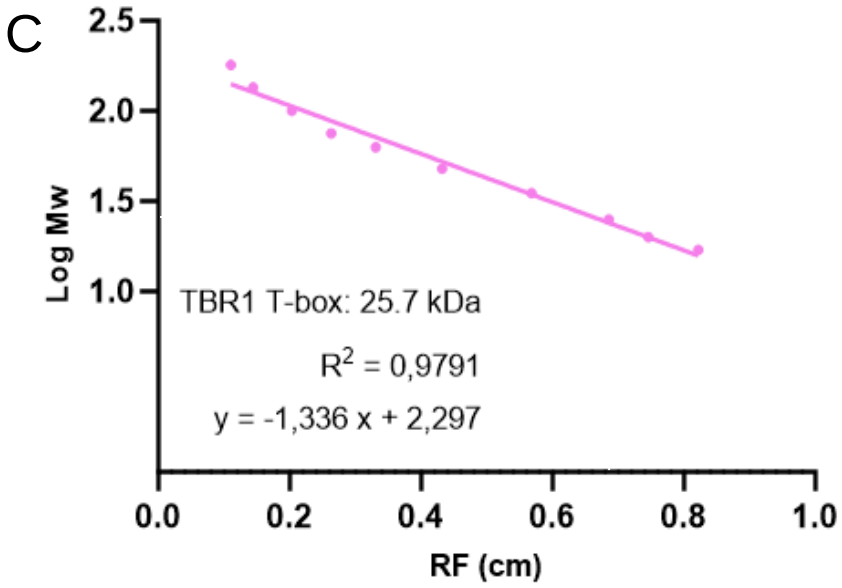


Figure 26: Purification and Purity Profile of TBR1 T-box Domain using IMAC. A: The elution profile of TBR1 T-box domain using IMAC. The graph displays the measure of absorbance (mAu) over time (represented in pink), along with the conductance (Ms/cm) (represented in blue). The various fractions collected during the IMAC process are labelled: the flow-through contains contaminants not bound to the column, the salt wash (1.5 M NaCl and 0.1 % Tween 20) includes any non-specifically bound compounds, and the elution of the TBR1 T-box domain (using imidazole) is marked as "TBR1 T-box". B: 12.5 % glycine SDS PAGE gel, under reducing conditions, representing both the overexpression and IMAC purification of the TBR1 T-box domain. Labels indicate the pellet (P), supernatant (S), Flow Through 1 (F1), Flow Through 2 (F2), Salt Wash (SW), and Elutions 1 to 3 (E1-E3) containing the TBR1 protein obtained from IMAC purification. The molecular weight marker is shown on the right (kDa) (Thermo Scientific PageRuler™ Prestained Protein Ladder, 10 to 180kDa). C: Standard curve of the TBR1 T-box domain, determined based on the molecular weight marker bands from SDS PAGE for its IMAC purification band on the gel. This standard curve estimates the size of the protein, based on the straight-line equation: y = -1.336x + 2.297 as approximately 25.7 kDa. Data was plotted using GraphPad Prism 10.0.0.

The IMAC elution profile of the TBR1 T-box domain demonstrates the elution of the protein due to an increased imidazole concentration in the elution buffer. The gel analysis suggests that the eluted fractions are reasonably pure (E1-E3). However, to further enhance purity, a secondary purification step was carried out using SEC, as depicted in Figure 27. The estimated molecular weight of the TBR1 T-box domain is approximately 25.7 kDa, determined through a standard curve generated from the equation $y = -1.336x + 2.297$. T-box proteins typically exhibit a size variation within the range of 50 kDa to 78 kDa (Hermann *et al.,* 1990), with the T-box domain constituting approximately one-third of this range (17-25 kDa) (Ruvinsky and Silver., 1997). Using the optimised sequence of the TBR1 T-box domain submitted to ExPasy, the theoretical molecular weight is calculated to be 25.657 kDa, aligning closely with the obtained experimental value of 25.7 kDa. To further purify the protein, the first IMAC elution was purified through SEC.

4.7.2 Size Exclusion Chromatography (SEC)

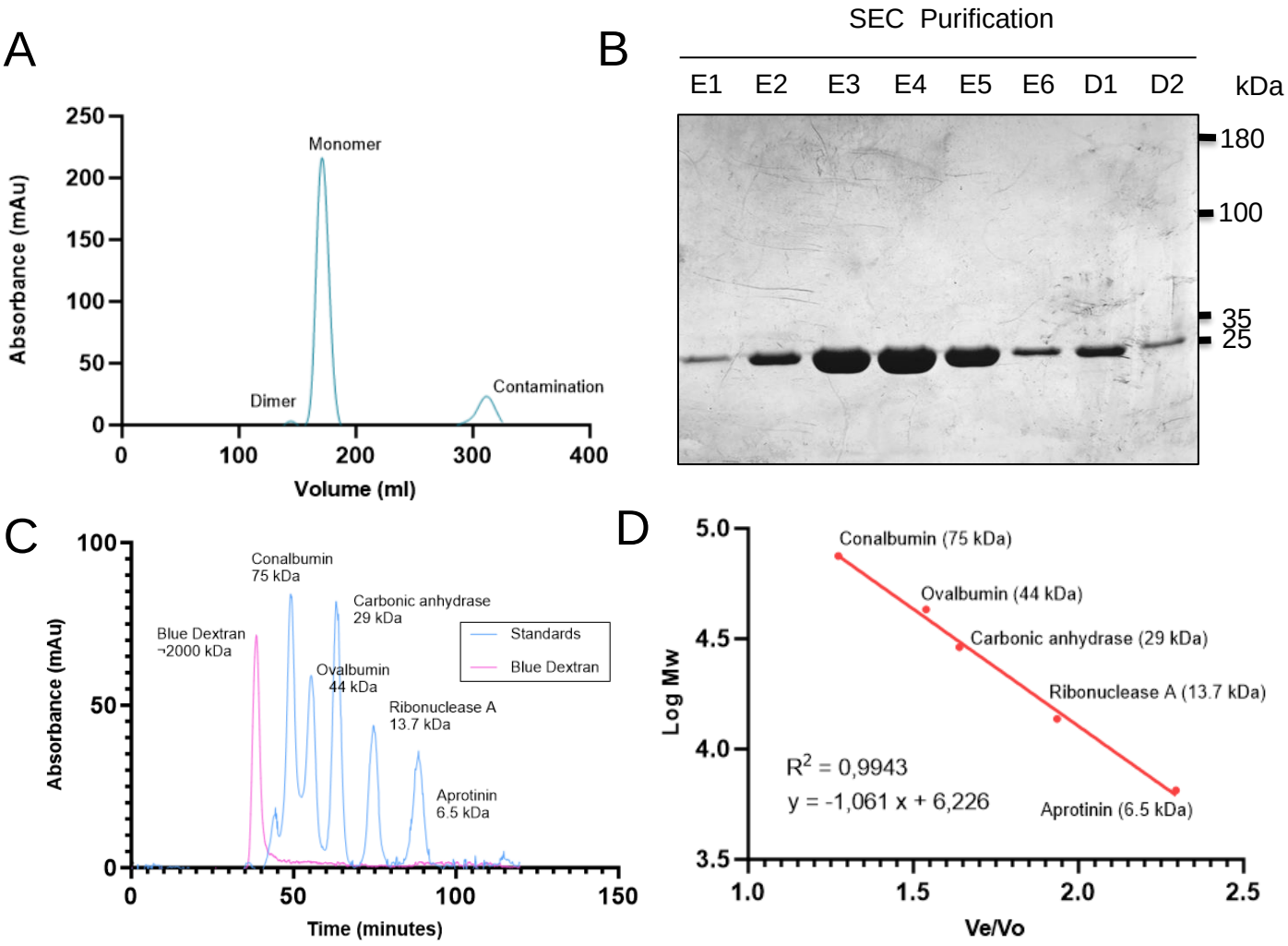


Figure 27: Purification and Purity of TBR1 T-box domain using SEC. A: The elution profile of the TBR1 T-box domain displays the sequential elution of the protein dimer, monomeric form of the protein, and lastly, some residual contamination. B: A 12.5% glycine SDS-PAGE gel, conducted under reducing conditions, exhibits the eluted fractions obtained from the purification process. E1-E6 corresponds to the elutions of TBR1 T-box domain collected in its monomeric form, while D1-D2 represents the eluted fractions collected in its dimeric form. Molecular weight markers are displayed on the right side of the gel (kDa). C: A standard curve for the TBR1 T-box domain is constructed based on the observed molecular weight marker bands on the SDS-PAGE Gel. The equation y = -1.275x + 2.225 is utilised, estimating the molecular weight of the TBR1 T-box domain monomer to be approximately 25.2 kDa. D: A set of GE Healthcare gel filtration standards is subjected to separation on the HiLoad™ 26/600 Superdex™ 75 prep-grade column. Blue dextran, with an exceptionally high molecular weight of approximately 2000 kDa, is run separately from the TBR1 standards. The TBR1 standards include Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic anhydrase (29 kDa), Ribonuclease A (13.7 kDa), and Aprotinin (6.5 kDa). E: The standard curve generated from the SEC standards data results in a linear relationship with the equation y = -1.061x + 6.2263. Utilising

this equation, the molecular weight of TBR1 T-box domain is calculated to be approximately 25.6 kDa. Data was plotted using GraphPad Prism 10.0.0.

The SEC elution profile displays three distinct peaks identified as dimer, monomer, and contamination. The dimer peak appears first in the elution sequence, which can be attributed to the limited penetration of larger molecules into the porous structure of the chromatography resin. Despite the pre-incubation of the TBR1 T-box solution with 2 mM DTT before SEC, the presence of a small peak suggests that some of the sample might not have undergone complete reduction of disulfide-linked dimers. Alternatively, there is a possibility of the formation of non-disulfide linked dimers. The second peak reveals TBR1 T-box domain is in its monomeric state. The results from the 12.5% glycine SDS-PAGE gel show that both the dimer and monomer solutions are highly pure. Using data from the gel, a standard curve was generated, estimating the molecular weight of TBR1 T-box domain to be 25.2 kDa in its monomeric and reduced dimeric configuration, as expected. These calculations are derived from a set of established standards with well-defined molecular weights. The final contamination peak, which manifested around 300 minutes, was intentionally not collected, and subsequently discarded. This decision was based on the prior knowledge that the elution time for TBR1 T-box was expected around 170 minutes and thus, the peak appearing at 300 minutes was deemed irrelevant for collection and analysis.

4.8 Structural Analysis of TBR1 T-box Domain

CD analysis was conducted to affirm the folded structure of the TBR1 T-box domain. CD spectroscopy provides insights into the secondary structure of the TBR1 T-box domain, aiding in the validation of the protein's structure as predicted by AlphaFold. Notably, the AlphaFold prediction suggests a predominant composition of β-sheets within the protein's structure.

4.8.1 Far-UV Circular Dichroism (CD)

To explore the secondary structure of the TBR1 T-box, a CD spectroscopy experiment was conducted using the Jasco J-1500 circular dichroism spectrophotometer, as shown in Figure 28. The voltage during the experiment did not exceed 600V, therefore the data is reliable for the full spectrum recorded.

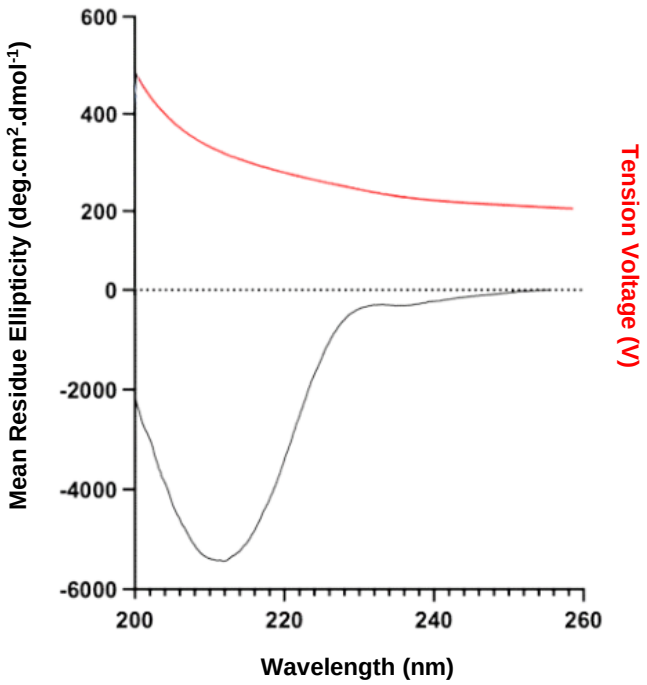


Figure 28: Far UV Circular Dichroism Spectrum for TBR1 T-box Domain. Black: The CD Spectrum of the TBR1 T-box Domain using a Jasco J-1500 circular dichroism spectrophotometer, serves to evaluate the protein's secondary structure elements. The graph illustrates Mean Residue Ellipticity (deg.cm^2.dmol^{-1}) plotted against wavelength (nm). Notably, a distinct trough at 218 nm is observed, which is indicative of the presence of majority of β-sheets within the protein's secondary structure. Red: The Graph of Tension Voltage (V) in relation to wavelength (nm) demonstrates that the data acquired is devoid of noise. It's important to note that the voltage, represented here, never exceeds 600 V throughout the measurement. Mean Residue Ellipticity, in this context, reflects the molar ellipticity of the protein per amino acid residue, and Voltage pertains to the electrical tension applied during the CD spectroscopy experiment. Data was plotted using GraphPad Prism 10.0.0.

CD spectroscopy was used to assess the secondary structure of the TBR1 T-box domain. The CD spectrum exhibits a distinct trough at 218 nm, which is a characteristic signature of β-sheets within the protein's secondary structure. This observation aligns well with the structural predictions generated by AlphaFold (Figure 18), which predominantly indicates the presence of β-sheets in the TBR1 T-box domain. The congruence between the experimental CD data and the AlphaFold predictions reinforces the validity of the predicted protein structure. Moreover, this confirms the structural integrity of the protein. When a protein possesses the correct secondary structure, it serves as an indication that the protein has likely undergone proper folding, ensuring it maintains its functional state in its native form. This is crucial for subsequent studies, including PDA, TSA and TR-FTIR. These functional assays will provide deeper insights into the potential

binding between CBD and the TBR1 T-box domain, offering information on the thermal stability of this interaction and identifying specific functional groups that may arise upon interaction.

4.9 Pull-Down Assay

The Epoxy-Activated Agarose PDA is a well-established technique extensively used in molecular biology to probe protein-ligand interactions. In this study, epoxy-activated agarose resin was coupled with 200 µM/ mL CBD and incubated with 120 µM TBR1 T-box protein for two separate incubation periods (1 hour and 24 hours). Concurrently, a control experiment was conducted, involving a 1-hour incubation with resin devoid of CBD. The amount of protein in each collected sample was determined using a Thermo Scientific™ NanoDrop™, as depicted in Figure 29, to assess whether the protein had bound to the CBD-coupled resin.

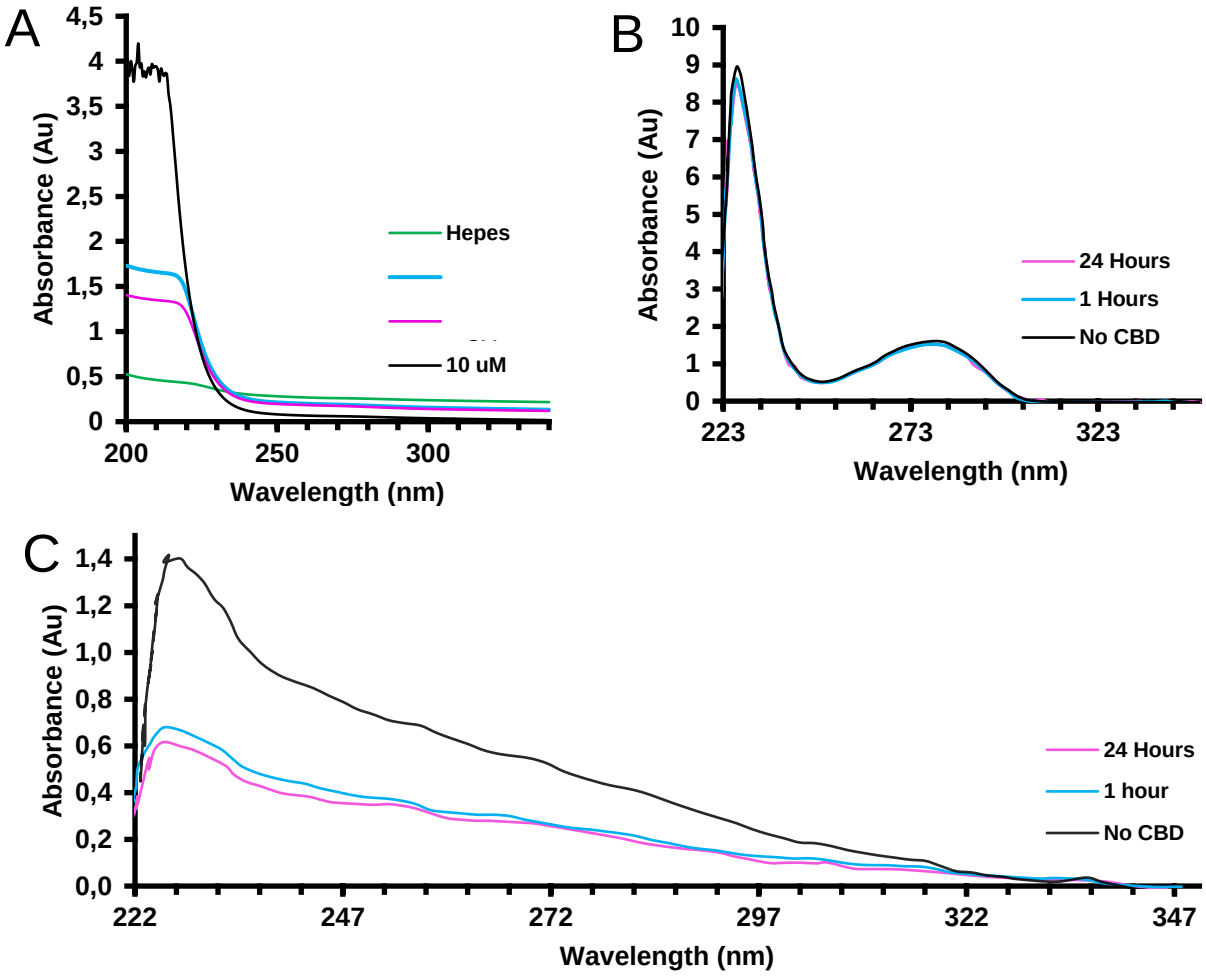


Figure 29: Absorbance Spectrum of CBD and Results of the PDA with TBR1 T-box Domain at 1 Hour and 24 Hours. A: Absorbance spectrum of 1 mg/mL CBD in methanol, diluted to 10 µM (black), 5 µM (blue) and 2.5 µM (pink) using HEPES buffer (0.2 M NaCl and 20 mM HEPES, pH 7.5). The spectrum was measured using a JASCO V-630 Spectrophotometer, showing absorbance (Au) plotted against wavelength

(nm) in the UV-Vis range of 200 nm to 350 nm. CBD exhibits an absorbance maximum from 200 nm to 220 nm. HEPES buffer is used as a control (green), as it was used to dilute the stock CBD solution and was used consistently as the proteins buffer. Maintaining a consistent buffer choice reduced variations in the experimental conditions, creating a stable environment for the protein in later studies. B: Absorbance readings of 'Sample A' obtained from the flow-through of the CBD-coupled resin from the PDA using a Thermo Scientific™ NanoDrop™. Absorbance is measured from 220 nm to 350 nm, for the flow-through obtained after 24 hours incubation (pink), 1 hours incubation (blue) and 1 hours incubation with a control (resin containing no CBD) (black). Absorbance maxima are observed at ¬230 nm and ¬280 nm for all samples. C Absorbance readings of 'Sample C' obtained from the CBD-coupled resin from the PDA using a Thermo Scientific™ NanoDrop™. Absorbance is measured from 220 nm to 350 nm, for the resin obtained after 24 hours incubation (pink), 1 hours incubation (blue) and 1 hours incubation with a control (resin containing no CBD) (black). Absorbance maxima are observed at ¬230 nm for all samples. Data was plotted using GraphPad Prism 10.0.0.

In Figure 29A, the absorbance spectrum of CBD is depicted, serving the purpose of confirming the presence of CBD in the solution and ensuring the absence of any contaminants **before** proceeding with subsequent experiments. Solubilised in methanol, CBD typically demonstrates two absorbance maxima, a larger peak around 210-220 nm and another small at 270-280 nm (Seccamani *et al.,* 2021). However, in Figure 29A's spectrum, CBD's peak absorbance falls within the range of 200-220 nm, and notably, no spectrum is recorded around 280 nm. Fortunately, subsequent FTIR experimentation validated the presence of CBD in the solution, as illustrated in Figure 31C.

During the PDA, three samples were collected – namely Sample A, Sample B and Sample C. Sample A comprised the flow-through collected post-incubation, while Sample B consisted of the flow-through obtained after subsequent washing steps. Sample C was directly obtained from the resin. This protocol was consistently followed for each experimental condition, involving both a 1-hour and a 24-hour incubation of the TBR1 T-box domain with the CBD-coupled resin, along with a control lacking CBD. In Figure 29B, showing the absorbance spectrum of Sample A for each experiment, an absorbance maximum is observed at 230 nm and 280 nm. It is common to detect some residual protein in the flow-through, evident from the absorbance at 280 nm, which serves as an indicator of the presence of aromatic amino acids such as tryptophan and tyrosine (Schmid., 2001). The TBR1 T-box domain encompasses 14 aromatic amino acids, including 4 tryptophan's, 5 tyrosine's, and 5 phenylalanine's. To assess the purity of the TBR1 T-box domain in this solution, the A260/A280 ratio is examined to identify potential nucleic acid contamination. TBR1 T-box domain displays a ratio of 0.6, which indicates highly pure protein. Importantly, no peak is observed at 340 nm, suggesting an absence of protein aggregation. This confirms that protein was present in the flow-through – representing protein that did not bind to the CBD-coupled resin.

Moving to Figure 29C, which represents 'Sample C' directly derived from the resin, the assessment aimed to determine if the TBR1 T-box domain interacted with CBD and if CBD successfully coupled to the resin. The absence of protein in the resin, as confirmed by the absence of a peak at 280 nm in the absorbance spectrum, suggests that the TBR1 T-box domain did not bind to the resin, and all the protein incubated with the resin was collected in the flow-through. Additionally, this spectrum shows that the resin without CBD exhibits higher absorbance compared to those allegedly containing CBD, implying that CBD did not successfully couple to the resin. However, according to Figure 29B, there is no indication that CBD was present in the flow-through (Sample A) either, as indicated by the similar peaks present for both supposed CBD-coupled and CBD absent experiments. Consequently, the absence of protein bound to the resin should not be interpreted as evidence that the protein does not bind to CBD. This lack of binding is attributed to the absence of CBD on the resin during the experimental process.

4.10 Thermal Shift Assay

Traditionally, small molecule ligands are known to stabilise protein structures upon binding, which is often linked to an increase in the melting temperature (Tm) of the protein-ligand complex (Huynh and Partch., 2015). A TSA was conducted between CBD and TBR1 T-box domain, to determine alterations in the proteins melting temperature upon ligand binding, to indicate potential ligand binding and consequently, the impact of binding on the thermal stability of the protein. Throughout the assay, the concentration of TBR1 T-box Domain remained constant at 10 µM, while increasing concentrations of CBD was introduced, ranging from 3 µM to 200 µM. Data analysis was conducted utilising the CFX Maestro software, and the results are shown in Figure 30.

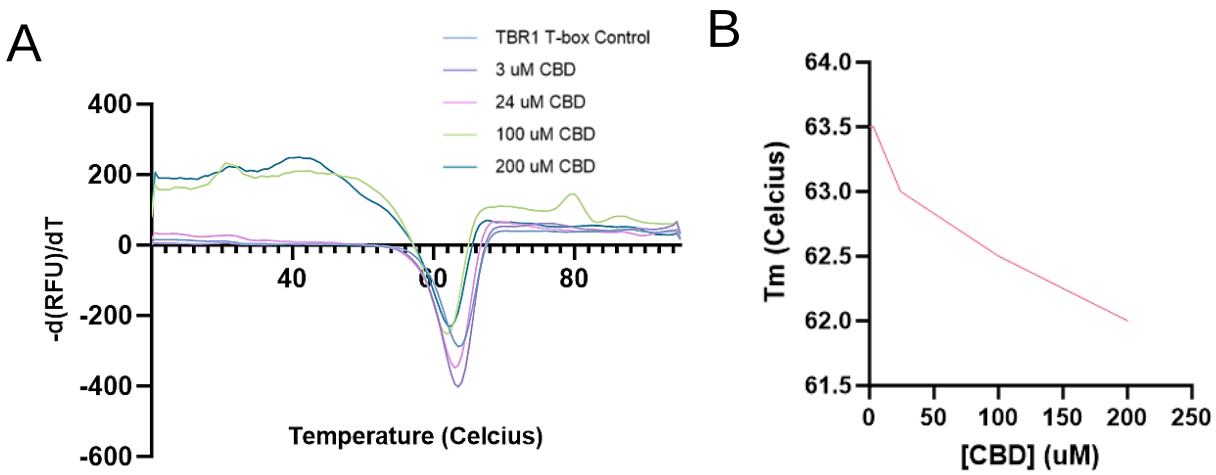


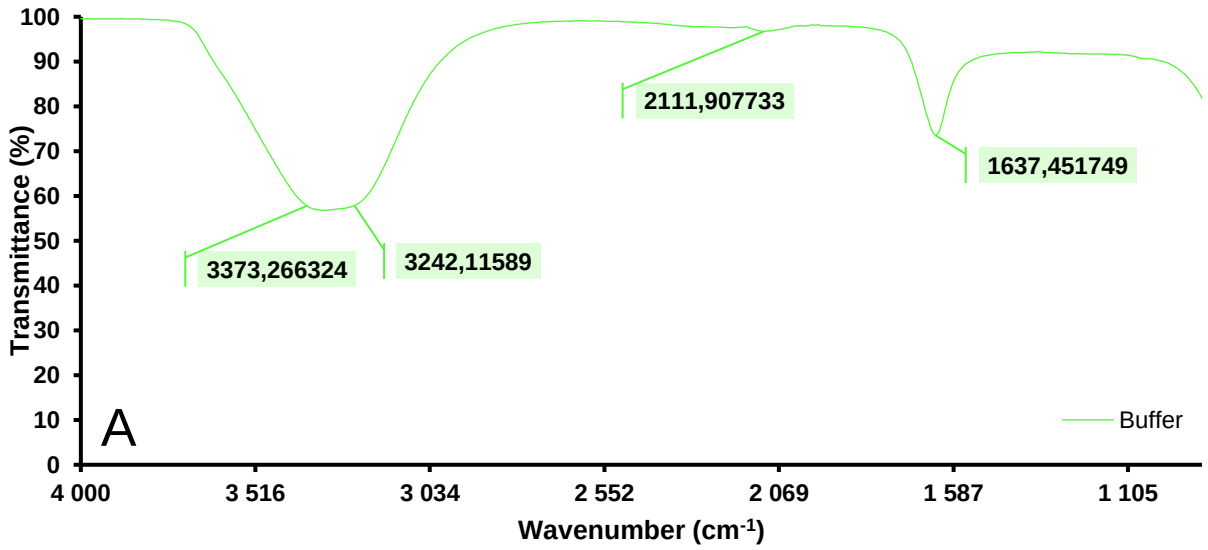
Figure 30: Thermal Melting Curves for TBR1 T-box Domain upon CBD Binding. The experiment was conducted with HEPES buffer, isolated TBR1 T-box Domain and CBD solutions utilised as control entities. Furthermore, the experimental series involved the TBR1 T-box Domain maintained at a consistent concentration of 10 µM, alongside varying concentrations of CBD, ranging from 3 µM to 200 µM. A: The melting curve profiles from the thermal denaturation behaviour of selected concentrations of TBR1 T-box domain, showing changes in melting temperature (Tm) at increasing CBD concentrations (3 µM, 24 µM, 100 µM, and 200 µM) each juxtaposed with a control solution of TBR1 T-box Domain devoid of CBD. Higher concentrations of CBD resulted in noisy data. B: Graph showing the relationship between melting temperature (Tm) and CBD concentration for the solutions delineated in graph B. Illustrating a reduction in the melting temperature as the concentration of CBD increases. Data was plotted using GraphPad Prism 10.0.0.

Figure 30 reveals that the TBR1 T-box domain, in the absence of CBD binding, exhibits a melting temperature (Tm) of 63.5 ℃. Upon introducing CBD, the Tm experienced a slight decline to 63.5, 63, 62.5, and 62 ℃ at varying CBD concentrations of 3 µM, 24 µM, 100 µM and 200 µM, respectively. Although the results indicate a decrease in the thermal stability of TBR1 T-box domain in the presence of CBD, this experiment does not confirm CBD binding. Given the

inconclusive evidence from both the PDA and TSA regarding a potential interaction between CBD and the TBR1 T-box domain, a more definitive approach was pursued. TR-FTIR was employed not only to establish conclusive evidence regarding the potential interaction between CBD and the TBR1 T-box domain but also to validate the integrity of each individual control solution (CBD and TBR1 T-box domain). Note that even though the identity of the TBR1 T-box domain was previously confirmed, the absorbance spectrum for CBD yielded inconclusive results. Through the FTIR experiment specifically conducted on CBD, the presence of CBD was successfully confirmed and verified, ensuring precision in the experimental setup.

4.11 Time-resolved Fourier Transform Infrared Spectra

To confirm the potential interaction between CBD and the TBR1 T-box Domain, a TR-FTIR utilising an FTIR-ATR spectrophotometer (PerkinElmer Spectrum One with Spectrum V5.00 software, PerkinElmer Instruments, CT, USA) was used to monitor alterations in the vibrational characteristics of functional groups. This was done by incubating a 1 in 4 dilution of CBD (1 mg/mL in methanol) with 120 μM of the TBR1 T-box Domain in an equilibration buffer over a period of 70 minutes. The resulting spectral data, spanning the range from 4000 to 400 cm^{-1}, is presented in the figures below. Figure 31 shows the control spectra for the equilibration buffer, protein and CBD; whilst Figure 32 represents the interaction between TBR1 T-box domain and CBD.



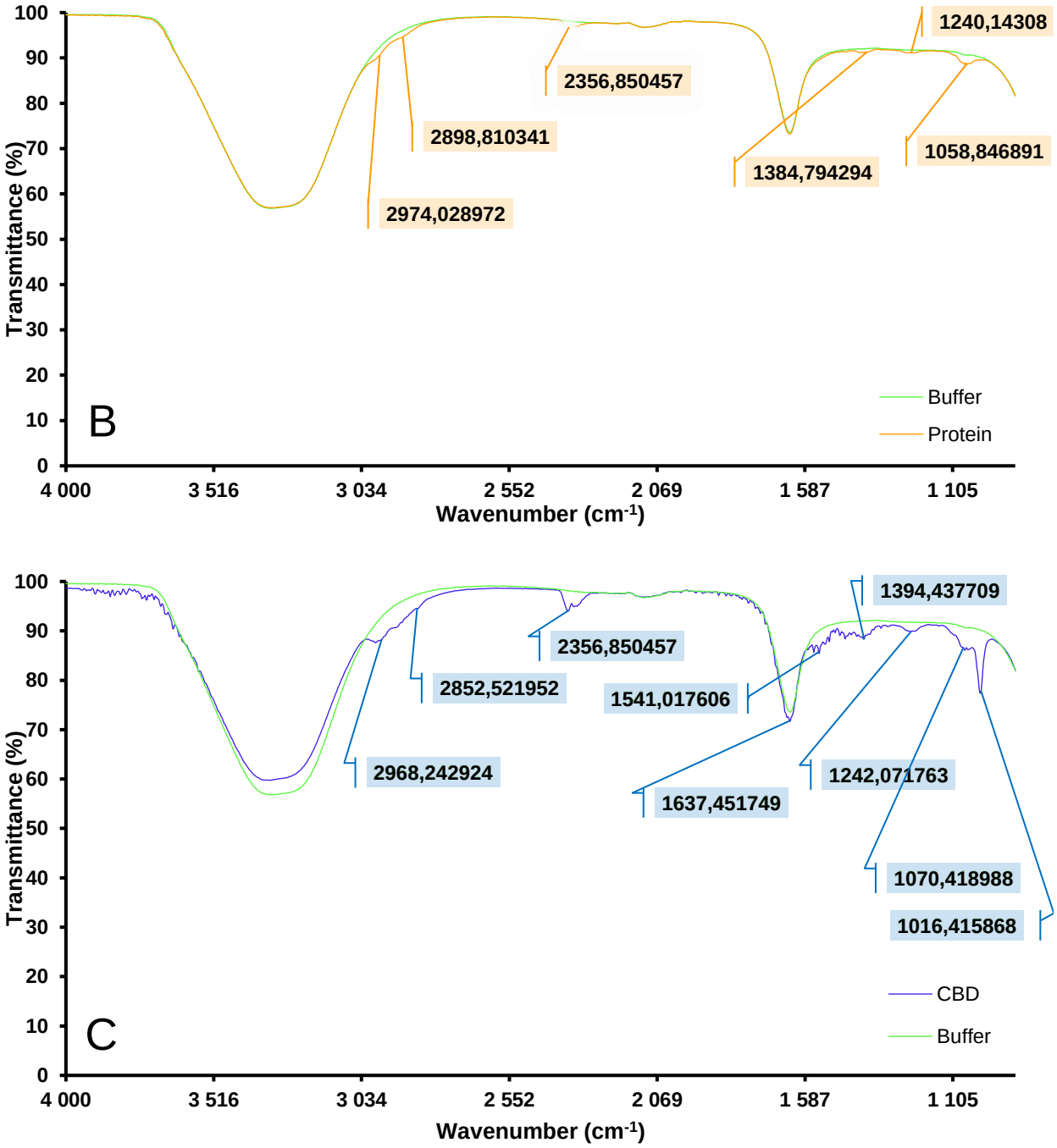


Figure 31: FTIR Spectra of Equilibration Buffer, CBD and TBR1 T-box Domain Controls. A: The green FTIR spectrum of equilibration buffer (30 mM imidazole, 500 mM NaCl and 20 mM Tris). Strong and broad bands observed between 3373.27 and 3242.12 cm^{-1}, a medium peak at 1637.45 cm^{-1} and a small band at 2111.91 cm^{-1}. B: Comparison FTIR spectra of equilibration buffer (green) and TBR1 T-box domain in equilibration buffer (orange). Peaks are observed at 2974.03 cm^{-1}, 2898.81 cm^{-1} , 2356.85 cm^{-1} , 1384.79 cm^{-1}, 1240.14 cm^{-1} and 1058.84 cm^{-1}. C: Comparison FTIR spectrum of a 1 in 4 dilution of CBD (1 mg/mL in methanol) in equilibration buffer (dark blue) and equilibration buffer (green). Bands are observed from

2968.24, 2852.52 cm^{-1} , 2356.85 cm^{-1} ,1541.02 cm^{-1}, 1242.07 cm^{-1}, 1395.44 cm^{-1},1637.45 cm^{-1}, 1070.41 cm^{-1} and 1016.42 cm^{-1}. The graphs were constructed using MS Excel (Sigma, Merck).

The FTIR spectrum in Figure 31 shows the control spectra for equilibration buffer, CBD and TBR1 T-box domain. Figure 31A represents the spectra obtained for the equilibration buffer (30 mM imidazole, 500 mM NaCl and 20 mM Tris). The presence of tris (Tris(hydroxymethyl)aminomethane) is verified by the pronounced presence of O-H stretching vibrations, evidenced by strong and broad peaks between 3373.27 and 3242.12 cm^{-1}, which aligns with the expected hydroxyl (OH) groups in its structure (Sigma, Merck; Wang *et al.,* 2023). Moreover, medium bands in the range of 3400-3300 cm^{-1} represent aliphatic primary amine N-H stretching vibrations, which are a characteristic feature of tris. Analysis of the spectrum further revealed characteristics attributed to imidazole including a medium peak at 1637.45 cm^{-1} which indicates N-H bending vibrations and C=C stretching vibrations, reflecting the presence of an amine group and conjugated alkenes within imidazole's structure, respectively (Sigma, Merck; Wang *et al.,* 2023). Additionally, C-H bending vibrations in the 2000-1650 cm^{-1} range further substantiate the presence of aromatic compounds in imidazole (Sigma, Merck; Wang *et al.,* 2023). Therefore, Figure 31A validates the presence of equilibration buffer.

The FTIR spectrum obtained for the TBR1 T-box domain in equilibration buffer (Figure 31B) reveals several distinctive peaks indicative of specific functional groups within the protein. Notably, two peaks at 2974.03 cm^{-1} and 2898.81 cm^{-1} signify C-H stretching vibrations commonly associated with aliphatic hydrocarbon chains present in amino acid side chains and the protein backbone (Sigma, Merck; Wang *et al.,* 2023). Additionally, a peak at 2356.85 cm^{-1}, while likely attributed to the presence of carbon dioxide (CO_2) in the sample, is a common observation in FTIR spectra exposed to atmospheric air (Sigma, Merck; Wang *et al.,* 2023). The peak at 1384.79 cm^{-1} indicates C-H bending vibrations, often associated with methyl groups (CH_3) and other aliphatic side chains of amino acids. Furthermore, the peak at 1240.14 cm^{-1} potentially reflects amide III vibrations, which involve a combination of C-N stretching and N-H bending, characterising the peptide bonds in the protein backbone (Sigma, Merck; Wang *et al.,* 2023). Lastly, a peak at 1058.84 cm^{-1} may signify C-N stretching vibrations associated with peptide bonds, particularly in the amide I region. Collectively, these spectral features provide valuable insights into the composition and potential structural elements of the TBR1 T-box domain (Sigma, Merck; Wang *et al.,* 2023). While there is a lack of existing literature on the FTIR spectrum of the TBR1 T-box domain, prior experiments involving SDS-PAGE gel, standard curves, and CD spectra have consistently affirmed the presence and purity of the TBR1 T-box domain. Thus,

Figure 31B stands as a reliable representation of the FTIR spectra of the protein, albeit somewhat overshadowed by the prevalent buffer signals.

Lastly, Figure 31C shows the FTIR spectrum of a 1 in 4 dilution of CBD (1 mg/mL in methanol) in equilibration buffer, compared with the equilibration buffer spectra. The FTIR spectrum of the CBD solution displayed several distinctive bands and peaks, each revealing specific molecular features confirming the presence of CBD in solution. Firstly, a broad and intense band spanning from 2968.24 to 2852.52 cm^{-1} was observed, which can be attributed to C-H stretching vibrations (Sigma, Merck; Wang *et al.,* 2023). This suggests the presence of aliphatic hydrocarbon chains in the CBD molecule. Note that the equilibration buffer exhibits C-H bending vibrations occurring at lower wavelengths, while CBD displays C-H stretching vibrations at higher wavelengths, making them distinguishable. Another band was observed at 2356.85 cm^{-1}, which indicates the presence of atmospheric CO_2 in the sample. The spectrum revealed distinctive peaks at 1541.02 cm^{-1}, 1242.07 cm^{-1}, and 1395.44 cm^{-1}. These peaks hint at potential features within the CBD molecule: C-H bending in the aliphatic chain or aromatic ring, O-H bending from the phenol group, and C-O bonds, suggesting the possible presence of phenolic and ether functional groups (Sigma, Merck; Wang *et al.,* 2023). A particularly interesting finding was the band at 1637.45 cm^{-1}, which appeared slightly more intense in the CBD solution compared to the equilibration buffer alone. This observation could signify the presence of C=C stretching vibrations, potentially originating from an alkene group within the CBD molecule. While the prior absorbance spectra of CBD in methanol (Figure 29A) did not conclusively establish the presence of CBD, Figure 31C using FTIR provides robust confirmation through the detection of distinctive C=C bonds. Methanol, a component of the CBD solution, does not contain any C=C bonds and although these bonds are present in equilibration buffer, the distinct bands observed are discernible from those in the equilibration buffer. These rules out the possibility of the observed bands originating from equilibration buffer functional groups, thus affirming the validity and existence of CBD in the solution also used in previous experimentations (PDA and TSA). Lastly, two distinct bands were observed at 1070.41 cm^{-1} and 1016.42 cm^{-1}. These bands may be indicative of C-O stretching vibrations, possibly arising from the presence of alcohol functional groups in the CBD structure (Sigma, Merck; Wang *et al.,* 2023).

To observe the potential interaction between CBD and TBR1 T-box domain, the compounds were incubated for 70 minutes, with FTIR readings taken every 10 minutes, as shown in Figure 32.

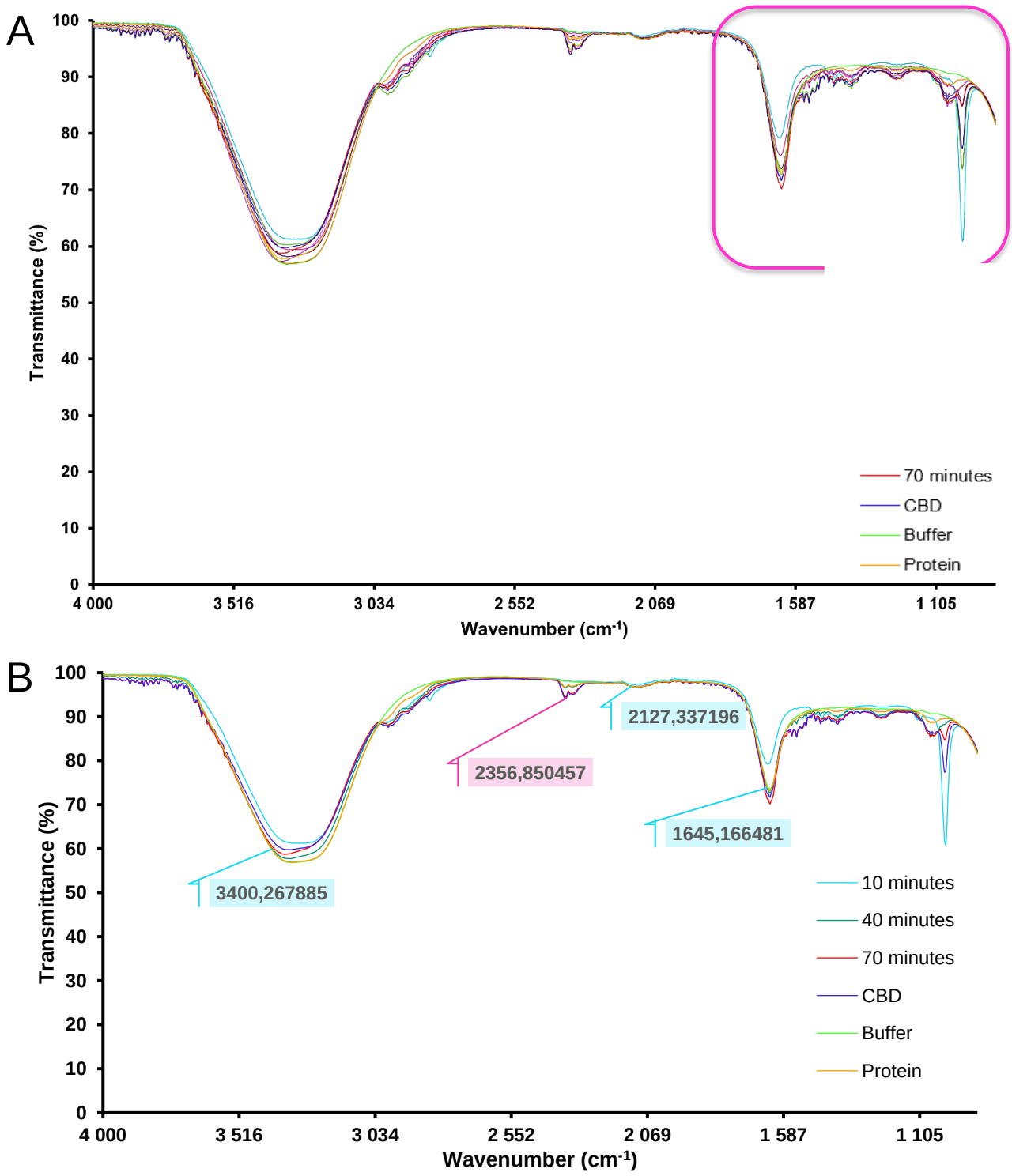


Figure 32: Time-resolved FTIR Spectrum of TBR1 T-box domain and CBD over 70 minutes. A: A solution containing 120 µM of the TBR1 T-box Domain was subjected to incubation with a 1 in 4 dilution of 1 mg/mL CBD in methanol, spanning a duration of 70 minutes. Measurements were acquired at 10-minute

intervals throughout the incubation period. Comparative spectra were generated for several controls, including CBD (dark purple), Equilibration buffer (neon green), TBR1 T-box Domain alone (orange), and at specific time points during the incubation (10 minutes in blue-grey, 20 minutes in dark pink, 30 minutes in purple, 40 minutes in yellow, 50 minutes in pink, 60 minutes in green, and 70 minutes in red). The neon pink square highlights the spectral regions corresponding to Amide I (1600-1800 cm^{-1}), Amide II (1470-1570 cm^{-1}), and Amide III (1250-1350 cm^{-1}), which are indicative of the amide groups within proteins and are the primary areas of interest. B: The FTIR spectra depicted in this include selected time stamps – including 10 minutes, 40 minutes, and 70 minutes – compared to the control spectra. The spectra illustrate the presence of O-H bonding (indicated by blue at 3400.27, 2127.34, and 1645.17 cm^{-1}) and CO_2 (indicated by pink at 2356.85 cm^{-1}). These graphs were constructed using MS Excel (Sigma, Merck).

Figure 32 shows the alterations in vibrational characteristics of functional groups within the solution comprising CBD and the TBR1 T-box domain. These spectral changes were tracked over a duration of 70 minutes. On the other hand, Figure 32B offers insights into the presence of functional groups that extend beyond the specific confines of the amide region. This broader analysis encompasses the detection of H_2O molecules, with particular emphasis on O-H stretching vibrations, manifested by peaks at 3400.27, 2127.34, and 1645.17 cm^{-1} (Sigma, Merck; Wang *et al.,* 2023). The intensity of these peaks correlates with the concentration of hydroxyl groups. Moreover, these hydroxyl groups may be representative of Tris, found in the equilibration buffer. The peak at 1645.17 cm^{-1} may also represent C=N stretching vibrations from the imidazole present in the equilibration buffer. Lastly, discernible peaks at 2356.85 cm^{-1} suggest the presence of carbon dioxide (CO_2) gas, a common occurrence when FTIR analysis involves exposure to atmospheric air (Sigma, Merck; Wang *et al.,* 2023). Figure 32A shows the region of primary interest, the amide region, highlighted in neon pink, which is discussed in greater detail in Figure 33.

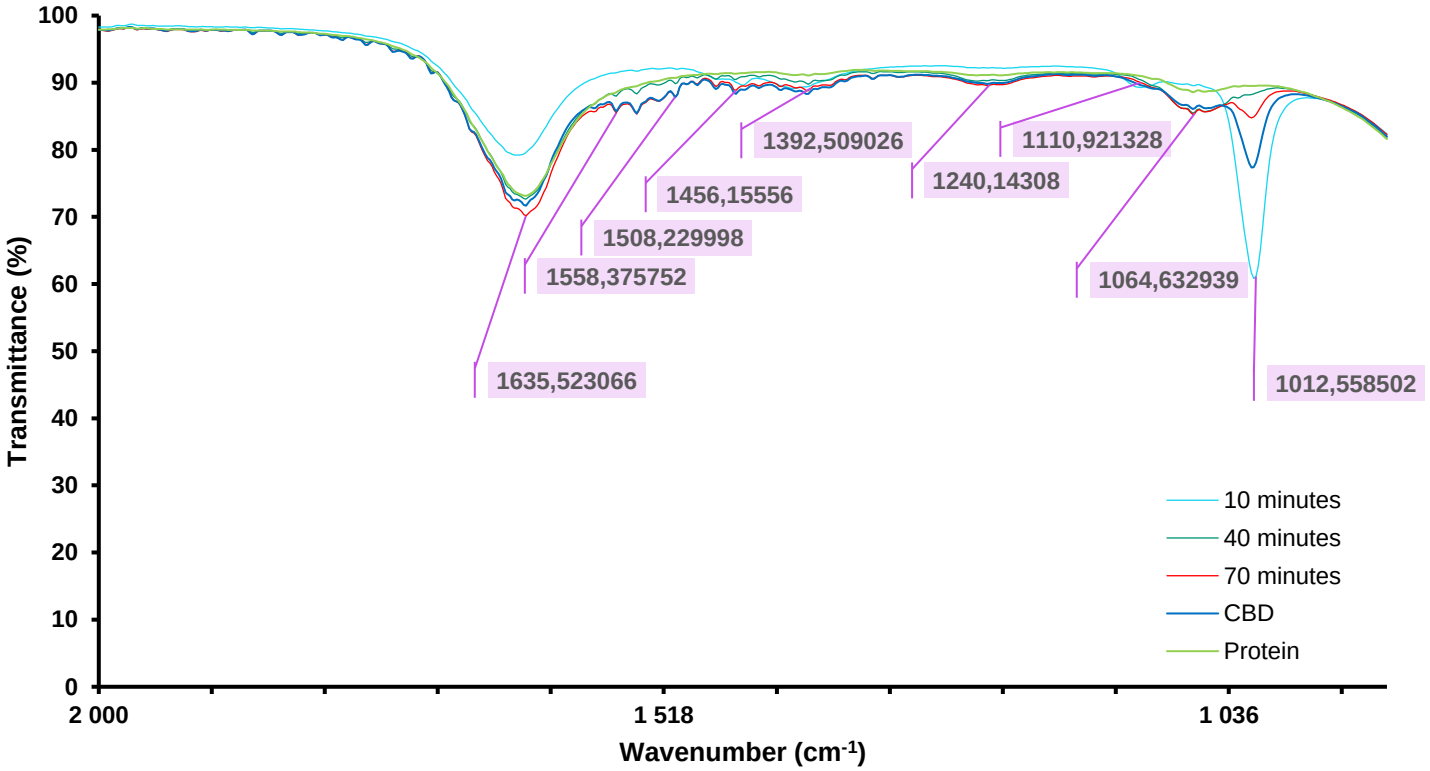


Figure 33: Time-resolved FTIR spectrum of the Interaction between CBD and TBR1 T-box Domain in the Amide region. This figure illustrates the dynamic changes observed in the Amide I, Amide II, and Amide III regions of the FTIR spectrum during the incubation of TBR1 T-box domain with CBD over varying time intervals. Notably, the Amide I band at 1635.52 cm^{-1} exhibits increasing absorbance with extended incubation. In the Amide II region, multiple bands at 1558.37, 1508.23, and 1456.16 cm^{-1} show escalating absorbance levels with incubation time, while the Amide III region displays a broad band at 1240.12 cm^{-1}. The interactions and changes in absorbance in these spectral regions offer valuable insights into the binding dynamics between CBD and the TBR1 T-box domain. These graphs were constructed using MS Excel.

Previous investigations by Venyaminov and Kalnin., (1990) discovered that within the set of 20 proteinogenic amino acids, only nine (Asp, Asn, Glu, Gln, Lys, Arg, Tyr, Phe, His) manifest notable absorbance specifically in the amide I and amide II regions. Considering the sequence of TBR1 T-box domain contains 181 amino acids, the 9 mentioned were primarily investigated in the FTIR spectrum. Within the Amide I region, a broad band extends from 1650 cm^{-1} to 1620 cm^{-1}, reaching its peak at 1635 cm^{-1}. This region is linked to secondary structure elements, particularly β-sheets (Venyaminov and Kalnin., 1990). Specifically, the 1645 cm^{-1} band is indicative of parallel β-sheets, while the bands at 1635 cm^{-1} and 1626 cm^{-1} are associated with anti-parallel β-sheets. A smaller peak at arises within the broader peak, at 1649 cm^{-1}, which represents the vibration of α-helical structures (Venyaminov and Kalnin., 1990). In comparison to the CBD control spectra, the band at 1635.52 cm^{-1} decreases upon incubation with the protein and displays an increase with

increasing incubation time. After 70 minutes, the intensity of this band surpasses that of the CBD control spectra. This observed trend hints at a potential alteration in the stretching vibrations linked to C=C bonds. Furthermore, this band is linked to symmetrical stretching vibrations in the guanidine group of arginine, along with terminal N-H asymmetrical bending vibrations (Venyaminov and Kalnin, 1990). Finally, the band observed at 1626 cm^{-1} can be associated with the N-H asymmetrical bending vibrations of lysine and the N-H bending vibrations of asparagine. In the Amide II region, distinctive bands appearing at 1558 cm^{-1}, 1541 cm^{-1}, 1522 cm^{-1} and 1518 cm^{-1} are identified. Specifically, the bands at 1558 cm^{-1}, 1522 cm^{-1} and 1518 cm^{-1} can be ascribed to glutamine's -COO asymmetrical stretching vibration and terminal N-H bending vibration, lysine's N-H symmetrical bending vibration, and tyrosine's ring hydroxyl vibrations, respectively. None of these bands are evident in the protein control spectra; instead, they emerge only after the 40-minute incubation period and exhibit a continual increase with prolonged incubation time. Additionally, the band at 1541 cm^{-1} only appears after 40 minutes incubation and continues to increase until T70, representing vibrations of α- helical structures because of hydrogen bonding and peptide backbone interactions (Venyaminov and Kalnin., 1990).

Within the Amide III region, a singular broad band is discerned at 1240.12 cm^{-1} after 70 minutes, potentially indicating the presence of C-O or C-N stretching vibrations associated with either aromatic ester or aromatic amine groups (Sigma, Merck; Wang *et al.,* 2023). Additionally, a subtle increase in absorbance at 1110.92 cm^{-1}, compared to the protein control spectra, suggests a modest enhancement in C-O stretching. This increase is followed by a subsequent decrease as the incubation time advances. The band at 1064.63 cm^{-1} increases as incubation time increases, indicative of C-O stretching vibrations. Finally, a strong band emerges at 1012.56 cm^{-1} upon incubation, becoming noticeable after 10 minutes. This band, absent in the protein control spectra, is evident in the CBD control spectra, albeit less prominently. Subsequently, this band decreases during incubation, signifying a decline in C-O stretching or C=C bending vibrations (Sigma, Merck; Wang et al., 2023).

5 CHAPTER 5: DISCUSSION

The growing popularity of cannabis and the increased use of CBD as a health remedy has made a big impact on modern healthcare. In this transformative cultural shift, the FDA has vigilantly monitored the ever-evolving statistics related to CBD usage, particularly its prevalence among expectant mothers. The FDA firmly recommends abstaining from CBD, THC, or marijuana in any form during pregnancy and lactation (U.S. Food and Drug Administration., 2019). This advice is rooted in the lack of comprehensive research exploring the effects of CBD on the developing foetus, expectant mothers and breastfed infants. Furthermore, the recent decriminalisation of cannabis for personal consumption in South Africa has sparked significant public health concerns. Notably, in-depth investigations among pregnant women in Cape Town have shown that approximately 3% to 11% (Petersen-Williams *et al.,* 2014; Western Cape Department of Health, 2019) self-reported cannabis consumption during pregnancy, a statistic that may even underestimate the actual prevalence (Petersen-Williams *et al.,* 2014). These statistical insights form a compelling narrative, reflecting the growing societal demand for alternative solutions to alleviate pain, anxiety and the various discomforts often accompanying pregnancy. Beyond addressing these immediate concerns, the data underscores the pressing need to understand the potential consequences of cannabinoid exposure during pregnancy on the developing foetal brain and the potential to cause neurodevelopmental disorders, such as ASD (Corsi *et al*., 2020). Additionally, studies have identified the TBR1 gene as a high-confidence ASD risk gene (Bedogni *et al.,* 2010). The loss of TBR1 function could lead to abnormalities in the formation and function of neural circuits, which have been implicated in the pathophysiology of ASD (Huang *et al.,* 2014).

Therefore, this study aimed to assess the potential interaction between CBD and TBR1 T-box domain to lay the groundwork for future investigations into its effect on the DNA binding function of TBR1 and possible implications in ASD development in offspring. An *in silico* study identified potential CBD binding pockets within TBR1 T-box and assessed the dynamics and stability of this interaction. The experimental results confirmed this interaction; however, further research may be necessary to thoroughly characterise this interaction under different conditions. As ASD is a complex condition with multiple contributing factors, research in this area cannot establish a causal relationship between CBD and autism. However, it can provide valuable insights for public health initiatives, preventive measures and targeted interventions to promote healthier pregnancies.

The T-box is the smallest part of the T-box protein that is required for binding to specific DNA sequences (Papaioannou., 2001). In the absence of a crystallographically solved structure of TBR1 T-box domain, the predicted AlphaFold structure (UniProt ID: Q16650) is marked as a high confidence level model with a pLDDT > 90 % (Figure 17). The structure was superimposed with the crystallographically solved structure of TBX5 (PDB: 2X6V) with a Cα-RMSD of 0.438 Å over 141 atoms, implying functional conservation in the DNA-binding domains. The structures provide to be predominantly β-sheets (Figure 17 and Table 5), which was confirmed by experimental CD studies, indicating a distinct trough at 218 nm (Figure 28). These comparisons reinforce the validity of the predicted TBR1 T-box structure used during the *in silico* studies.

DNA Binding Residues

Of the DNA-interacting residues identified in TBX5, certain residues such as K159, N162, S175, Y217, T223, I227 and N230 exhibit minimal conformational changes upon DNA binding (Stirnimann *et al.,* 2010), the corresponding residues in TBR1 (S329, T370, N314, G390, K389, I382, R231, Y372, T378, A388, N385, P386, R232, K310, F387, and F391) are predicted to play a similar role, indicating that TBR1 maintains structural integrity crucial for DNA interaction (Figure 18). Moreover, the significant conformational shifts in TBX5's 3_{10}C helix, particularly in residues F232–R237, which are vital for docking into the DNA minor groove (Stirnimann *et al.,* 2010), suggest that the corresponding helical region in TBR1 (F387 and F391) may exhibit comparable flexibility and undergo similar conformational changes to facilitate DNA binding. Similarly, residues R231 and R232 in TBR1 may also undergo minor adjustments, as observed in TBX5's R81 and R82, during DNA engagement (Stirnimann *et al.,* 2010). The dynamic nature of TBX5's 3_{10}C helix, which is critical for the positioning of its transactivation domain, hints that the corresponding TBR1 helix may influence the orientation and function of its subsequent transactivation domain, mirroring the regulatory role seen in TBX5 (Stirnimann *et al.,* 2010). While the structural alignment suggests a shared DNA-binding motif between TBR1 and TBX5 (Stirnimann *et al.,* 2010), variations in amino acid sequences might confer unique binding properties or regulatory functions specific to TBR1, underscoring the potential for both structural conservation and divergence.

The Potential for CBD to reach TBR1 T-box domain

A study by Weisbeck *et al.,* (2021) concluded that expectant mothers who consume cannabis tend to hold lower perceptions of risk regarding the effects on the foetus in comparison to pregnant women who abstain from cannabis use. Considering CBD is frequently consumed orally and is readily absorbed by the gastrointestinal tract (QPPCaco = 2405.6 in Table 2), there is significant potential for systemic exposure in pregnant women. Despite its low human bioavailability (6%)—attributable to factors like high lipophilicity (QPlogPo/w = 5.3), poor solubility (Grifoni *et al.,* 2022) and extensive first-pass metabolism (Perucca and Bialer, 2020)—CBD effectively diffuses through the maternal-placental circulation (Ochiai *et al.,* 2021). The presence of certain characteristics, as well as deviations from Lipinski's RuleOfFive and Jorgensen's RuleOfThree (as shown in Table 2), indicate difficulties in developing effective oral CBD formulations. This can affect the absorption of CBD and its transfer to the foetus, as well as prolonged exposure during critical developmental stages.

Despite these absorption challenges, CBD readily diffuses through the maternal-placental circulation (Ochiai *et al.,* 2021), with its affinity for Human Serum Albumin (HSA) (QPlogKhsa = 1.0523) suggesting prolonged circulation in the bloodstream. This is further supported by its extended half-life, estimated to be between 18 to 32 hours (Devinsky *et al.,* 2014). As a small lipophilic molecule, CBD can easily transverse the placental barrier and enter foetal circulation, with research indicating up to a fifth of consumed CBD ends up in the foetal compartment (Erez *et al.,* 2023). CBD is metabolised by CYP3A, although with less efficient in the foetal liver (Hyotylainen *et al.,* 2024), suggesting that CBD might linger longer in the foetal system compared to adults. Once in circulation, it can be distributed throughout the foetal organs with minimal liver metabolism, especially during the first trimester (early pregnancy) where the foetal organs (including the brain) are still being formed (Hyötyläinen *et al.,* 2024).

A study by Ochiai *et al.,* (2021) concluded that 66.9% of CBD present in maternal blood was transferred into foetal circulation within 15 minutes of administration. CBD showed a preferential distribution, predominantly accumulating in the foetal brain and liver. The half-life of CBD in the foetal brain was determined to be 7.2 ± 2.4 hours, which is notably longer than in the foetal liver, where the half-life was 4.2 ± 0.4 hours (Ochiai *et al.,* 2021) This suggests a prolonged presence of CBD in the brain compared to the liver. Table 2 predicts CBD to have limited CNS activity (CNS = 0) and limited BBB permeability (QPlogBB = -0.5). However, even after the foetal brain has fully developed, CBD can transverse the BBB and interact with extracellular endocannabinoid

receptors (Tang *et al.,* 2016). Recent research has shown that CBD can bind to the PPAR-γ nuclear receptor, suggesting it can cross the lipid cell membrane (supported by a QPPMDCK = 1277.7 for good cell permeability in Table 2) through diffusion or Hsp70 transportation and be delivered to the nucleus for transcriptional regulation (O'Sullivan., 2012). Additionally, a study by Downer *et al.,* (2012) reported that CBD can activate the peroxisome proliferator-activated receptors (PPARs), which are TFs involved in metabolic regulation and anti-inflammatory responses (Downer *et al.,* 2012). Nonetheless, it's essential to acknowledge that this field is continually evolving and further research is needed to fully elucidate the precise mechanisms and consequences of CBD's impact on TF activity, such as with TBR1.

CB1 receptors are present throughout the brain, with the highest concentrations in areas like the cerebellum, hippocampus, basal ganglia, and cerebral cortex (Herkenham *et al.,* 1991). The TBR1 TF is mainly expressed in the developing cerebral cortex (Hevner *et al.,* 2001), primarily during the early stages of brain development, and decreases significantly after birth (Bedogni *et al.,* 2010). The overlap in the distribution of CB1 receptors and TBR1 expression indicates that regions with high CB1 receptor density, which are potential targets for CBD, are also critical areas for TBR1 activity. These findings further support the importance of researching the possible interaction between CBD and TBR1 T-box domain and its influence on prenatal brain development.

CBD Meets TBR1 T-box domain: A Computational Choreography

CBD Predicted Binding Pocket

There are no existing studies on the interaction between CBD and TBR1 T-box domain and thus, no indication on where or if CBD can bind to the protein. Computational tools such as CASTp3.0, Protein Plus and SiteMap all predicted a druggable binding pocket (Figure 19) distinct from where CBD ultimately bound during molecular docking and molecular dynamics simulations. CASTp3.0 identified residues R231, R232, Y372, T378, I382, N385, P386, F387, A388 and F391 within the top scoring druggable Binding Pocket 1 (SiteScore = 0.973 in Table 1), that also correspond to the predicted DNA binding residues in Figure 18. Binding Pocket 1 had a suitable volume (2000 Å^3) considering the size of CBD (1157.7 Å^3 in Table 2) and was strategically placed near the regions predicted to interact directly with DNA (α3 helix and the 3_{10}C-helix). Initially, this overlap

suggested CBD binding to this binding pocket may influence the DNA binding function of TBR1. However, its large hydrophilic nature (694.398 Å²) and small hydrophobic region (57.442 Å²) raised doubts on being a suitable site for the hydrophobic CBD (QPlogPC16 = 10.5 in Table 2).

The stabilising effect of CBD

The 2D ligand interaction diagrams (Figure 20) from Maestro Schrödinger's IFD Protocol, encompassing CBD-docked poses (Pose 1 to Pose 4), reveal a pronounced preference of CBD for a distinct hydrophobic cavity. This cavity is different from the initially predicted Binding Pocket 1 in Figure 19. Such a preference indicates that the interaction of CBD with the TBR1 T-box domain is primarily driven by hydrophobic forces, supported by the observation of predominantly hydrophobic protein residues surrounding the ligand in Figure 20. Furthermore, the MD simulations confirmed the hydrophobic interactions to be dominant in all 4 poses across the 250 ns simulation (Figure 22B). The predictive models' failure to foresee this specific CBD binding site (Figure 19) might suggest that the ligand induces conformational changes in the protein, aligning with the Induced Fit model. This hypothesis is further supported by the higher average Cα-RMSD for the protein (1.70 Å) compared to the ligand (0.97 Å) during the docking simulations (Table 3) indicating structural adjustments in the protein to optimally accommodate the ligand. These results point to increased specificity for the ligand by reducing non-specific binding (Celej *et al.,* 2003). However, the ligand is predicted to exhibit greater flexibility in a dynamic environment (LP-RMSD = 6.3 $\pm$ 2 Å on average) compared to the protein (Cα-RMSD = 3.5 $\pm$ 0.4 Å on average) (Figure 21), which is more representative of a biological environment. This suggests that CBD does not fit in a pre-formed binding site in a ‘lock-and-key’ manner, rather it undergoes significant structural changes to achieve an optimal fit within the binding site. This is further indicative of an ‘induced fit’ model.

However, the ligand and protein of all 4 poses eventually reach equilibrium during the 250 ns simulation (Figure 21), indicating that a stable interaction is formed. Pose 2 is an exception where, towards the end of the simulation, the ligand aligns more closely with Pose 3, as evidenced by a higher LP-RMSD (Figure 21B). This convergence between Pose 2 and Pose 3 suggests potential similarities in their binding configurations. Pose 2 and 3 consistently exhibit similar Cα-RMSD values throughout the entire simulation, closely resembling the APO RMSD. This observation implies that these poses potentially correspond to more natural or favourable conformations and could be the preferred binding mode. The APO Cα-RMSD reaches equilibrium at 105 ns, while the CBD-docked poses exhibit stabilisation at 75 ns (Figure 21A). This suggests that CBD’s

interaction may enhance the protein's stability, leading to a conformation that is more energetically favourable. Therefore, in the absence of a ligand, the APO protein exhibits inherent flexibility, a characteristic typical of many proteins, which allows for dynamic interactions and functional versatility. This flexibility is also represented in the RMSF analysis, where the APO protein exhibited the highest average Cα-RMSF value of 1.96 ± 1.49 Å. In the CBD-docked poses there is a general decrease in RMSF, with average values of 1.56 ± 1.19 Å (Figure 23). This suggests that CBD binding leads to reduction in local fluctuations in TBR1 T-box domain, indicating a more stable protein conformation. This may represent an allosteric binding mechanism, where CBD binds to a site distinct from the active site, resulting in conformational changes that can alter the proteins activity. This is further supported by Figure 24, where the ΔRg upon CBD binding (– 0.99 ± 1.69 Å) suggests that the conformational impact of CBD extends beyond the immediate vicinity of the binding site. CBD binding promotes structural compactness and reduced fluctuations. This suggests the protein becomes more stable upon CBD binding which implies effective and specific binding. However, Table 5 shows the increase coil content and decreased helix content upon CBD interaction. Therefore, CBD binding appears to induce localised structural changes which may influence the flexibility of these specific areas – but the overall tertiary structure may become more stable in a compact manner. Remarkably, this stabilising effect of CBD on TBR1 mirrors similar observations in the CB1-CBD complex, where CBD also exerts a stabilising influence without causing major conformational disruptions (Aviz-Amador *et al.,* 2021; Ji *et al.,* 2020). Such consistency in CBD's interaction across different protein targets underscores its role as a stabilising agent, which may enhance TBR1's activity or modulate its gene regulatory functions.

These localised structural changes, indicated by the high fluctuating areas, are seen between residues 66-67, 105-108 and 135-141, along with the protein terminals. It appears that none of these regions, except for the protein terminals which demonstrate the largest fluctuations in both the APO and CBD-docked poses, directly participate in DNA binding. This implies that CBD may not have a large influence on these sites and may not significantly impact the DNA binding function of the TBR1 T-box domain. However, experimental validation is required to address the potential or lack of allosteric modulation mechanism by CBD. The only major alteration occurred between Asn294 and Thr295, transitioning from a random coil to a β-pleated sheet. Such a shift could be attributed to the inherent flexibility of proteins, a characteristic often observed under the induced fit protocol. These residues are not predicted to interact directly with DNA, as indicated in Figure 18.

Affinity Matters: CBD's Interaction with TBR1 T-box domain

In comparison to its interaction with known targets, CBD's binding to the endocannabinoid receptor CB1 involves the insertion of its alkyl side chain, a mode of interaction mirrored in the TBR1 docked poses. A study by Aviz-Amador *et al.,* (2021) showed CBD binding to CB1 and CB2 receptors with binding affinities of -8.5 kcal/mol and -9.3 kcal/mol, respectively. The average docking score for all the poses generated (Pose 1 to Pose 4) was -9.3 kcal/mol, with Pose 2 achieving the highest docking score of -9.8 kcal/mol. This suggests a comparably low affinity of CBD for TBR1 T-box domain, considering CBD is known to have low affinity for both CB1 and CB2 (Aviz-Amador *et al.,* 2021). The CBD-docked poses show minimal direct interacting residues. Hydrogen bonds are observed in Pose 2 and 3, with residues S238 and Pro335, respectively. Although Pose 1 did not show any direct binding, Pose 4 exhibited π-π interactions with Phe239. Similarly, CBD's interaction with CB1 and CB2 also predominantly exhibiting hydrophobic interaction, with some alkyl π and π-π interactions (Aviz-Amador *et al.,* 2021). MD simulations, although showing predominant hydrophobic interactions stabilising the CBD interaction, also exhibited hydrogen bonding across the 250 ns simulation. All four poses exhibited a common trend in hydrogen bond formation between Ser238, Pro335, Thr360, Glu363 and Asn240. This corresponds to docking simulations for Pose 2 and 3 (hydrogen bonding with residues S238 and Pro335 respectively). The recurring involvement of these residues suggests these are key sites for ligand interaction. Interestingly, these residues are situated in areas of lower fluctuations (Figure 23), which suggests these interactions sites are relatively stable. This stability reinforces the importance of these interaction in maintaining a strong protein-ligand complex.

The process of CBD interacting with TBR1 is hypothesised to be a two-step mechanism. Initially, CBD is likely attracted and anchored into the TBR1 T-box domain's binding site through hydrophobic interactions. These interactions are crucial for overcoming the solvation barrier, facilitating CBD's entry into the binding pocket. Following this, the proximity to the correct residues allows the formation of the hydrogen bonds observed in Figure 23, providing additional stability and specificity to the interaction. This sequential stabilisation mechanism ensures that CBD is not only correctly positioned but also securely fit within the binding site. Furthermore, none of these hydrogen-forming residues correspond to the predicted DNA interacting residues. However, Pro335 is conserved between TBR1 and TBX5 (Figure 18) suggesting a functionally important region, supporting the biological relevance of this interaction.

Biologically relevant CBD-docked pose

The comprehensive bioinformatics analysis of CBD's interaction with the TBR1 T-box domain highlights the significance of both Pose 2 and Pose 3, each demonstrating unique aspects of potential biological relevance. Pose 2 is distinguished by its strong and stable binding to the TBR1 T-box domain through a key hydrogen bond with Serine 238, which persists throughout the MD simulations. This interaction is further supported by hydrophobic contacts, contributing to the pose's overall stability. This combination of hydrogen bonding and hydrophobic interactions is reflected in Pose 2's high docking score (-9.8 kcal/mol in Figure 20C), suggesting a robust and specific binding mode. The MD simulations corroborate this stability, with the Cα-RMSD showing minimal conformational changes over time (Figure 21A). In contrast, Pose 3 exhibits a dynamic and potentially more functionally impactful interaction. It is marked by pronounced conformational changes, indicated by the largest decrease in the ΔRg (1.32 ± 1.59 Å) and a shift in secondary structure towards increased coil content from 66.35 % in the APO form to 69.85 % in Pose 3 (Table 5). This dynamic interaction suggests that CBD binding in Pose 3 could lead to increased structural adaptations in specific regions in the TBR1 T-box domain, which might have more profound implications for functionality. However, the decreased fluctuations (Cα-RMSF = 1.47 ± 1.04 Å in Figure 23) and stable binding (Cα-RMSD = 3.07 ± 0.33 Å in Figure 21) suggest an overall increase in stability of the whole tertiary structure upon CBD binding.

While Pose 1 was identified as the best scoring in the docking simulations, this may not always correspond to the experimentally observed binding orientation (Ramirez *et al.,* 2018). MD simulations allowed a deeper exploration into a biologically relevant interaction. Pose 2's stability and specificity make it a strong candidate for a biologically relevant conformation, while Pose 3's dynamic and potentially functionally impactful interaction could offer deeper insights into the protein's modulation by CBD. These insights are crucial for comprehending the molecular dynamics of CBD's interaction with TBR1 and have significant implications for developing therapeutic strategies. Future experimental validation and functional studies are essential to fully ascertain and understand the biological and therapeutic relevance of these poses.

In conclusion, the comprehensive bioinformatics analysis of CBD's interaction with the TBR1 T-box domain presents a multifaceted view of its binding mechanism and consequent structural influence. The induced fit model is particularly evident, as CBD's binding leads to significant but localized conformational changes in the TBR1 protein, suggesting a specific and dynamic interaction. This is further supported by the decrease in the ΔRg, indicating an overall compaction

of the protein structure upon CBD binding, and the RMSF analysis, which reveals decreased fluctuations in the protein's overall structure. These structural alterations point towards an allosteric modulation mechanism, wherein CBD binding at a site distinct from the active site induces conformational changes that potentially affect the protein's function. Notably, the areas with the most significant hydrogen bonding interactions—specifically with residues Ser238, Pro335, Thr360, Glu363, and Asn240—lie in regions of lower fluctuations, highlighting their stability and suggesting their crucial role in the binding process. Despite these changes, it is predicted that CBD binding might not significantly impact the DNA binding function of the TBR1 T-box domain, as the major conformational changes do not directly involve the DNA binding regions. This hypothesis, coupled with the recurring involvement of specific residues in the interaction, underscores the complexity of CBD's influence on TBR1's structure and function, suggesting a selective and potentially therapeutically relevant modulation of the protein's activity without disrupting its primary DNA-binding capabilities. Nevertheless, to ascertain the significance of this interaction, experimental validation is imperative.

CBD and TBR1 T-Box Domain: Time to Dance in the Lab

The *in silico* predictions, suggesting a potential interaction between CBD and the TBR1 T-box domain, necessitate empirical verification. To facilitate this, adequate amounts of the TBR1 T-box domain were successfully expressed and purified, as evidenced by Figures 26 and 27. The high purity of the protein was indicated by an A260/A280 ratio of 0.6 (Figure 29B). However, efforts to experimentally confirm the interaction between the TBR1 T-box domain and CBD encountered challenges, resulting in outcomes that were unreliable and inconclusive.

The Pull-Down Assay

During the PDA, a significant concentration of 200 µM/mL CBD was immobilised on epoxy-activated agarose resin. This resin was then exposed to a 120 µM solution of the TBR1 T-box domain for varied durations of 1 and 24 hours, as depicted in Figure 29. The assay, however, did not demonstrate any binding of the TBR1 T-box domain to the resin (Figure 29C). While some presence of protein in the flow-through post-incubation is typically expected due to unbound proteins being washed through, the concentration of protein in both the CBD-coupled and the resin without CBD was found to be identical. This uniformity in protein concentration suggests that all of the protein incubated with the resin ended up in the flow-through, as indicated in Figure 29B. These results do not conclusively rule out the possibility of an interaction between CBD and

TBR1. Instead, they point towards potential issues in the experimental setup, particularly in the effective coupling of CBD to the resin.

The efficacy of a PDA for ligand-protein interactions is heavily influenced by several factors, including the inherent properties of the ligand, its stability upon immobilisation and the nature of its interaction with the target protein. In the case of CBD, its hydrophobic nature and limited solubility in aqueous solutions likely posed challenges in achieving effective interaction with the hydrophilic epoxy-activated agarose resin. This solubility mismatch could have resulted in inadequate exposure of CBD's reactive groups to the resin, as the CBD molecules would aggregate to minimise their exposure to water. This may have masked the reactive groups in CBD that are essential for coupling to the resin. Additionally, these aggregates might have been inadvertently removed during the washing steps. Moreover, the use of a high pH (pH = 11) in the coupling buffer significantly deviates from CBD's optimal stability range, which lies between pH 4 and 6 (Jaidee *et al.,* 2022). At this elevated pH, CBD is prone to structural modifications, such as isomerisation at the Δ^6-CBD position, which involves a change in the position of the double bond within the cyclohexane ring in the CBD molecule (Wagner., 1966). Additionally, CBD is prone to degradation under basic conditions (Jaidee *et al.,* 2022), which may have resulted in the formation of compounds undetectable within the specified absorbance range. Such alterations can critically impact CBD's reactivity, potentially compromising its ability to effectively bind to the resin.

Exposure to air during preparation and handling is another factor that could have contributed to the oxidative degradation of CBD (Jaidee *et al.,* 2022). CBD contains several chemical groups, such as phenolic rings, that are prone to oxidation. Oxidation involves the loss of electrons from these groups, leading to the formation of new compounds. This oxidative process can substantially alter CBD's chemical structure and its reactivity, a crucial aspect to consider given that the CBD solution in this experiment was neither prepared nor stored in an oxygen-free environment. CBD's stability is also affected by its sensitivity to acidic conditions. In this experiment, CBD was subjected to alternating high and low pH environments during the wash steps. Exposure to acidic conditions could lead to further degradation or transformation of CBD, thereby impacting its structural integrity and its potential for binding. This aspect is particularly relevant given the use of buffers of varying pH and the subsequent quenching with 1M ethanolamine, which could have inadvertently contributed to the removal of CBD from the resin or negatively affected its ability to immobilise efficiently.

Following the inconclusive outcomes of the PDA, the focus shifts to the TSA to further investigate the potential interaction between CBD and the TBR1 T-box domain. The TSA offered a different perspective by assessing changes in the protein's thermal stability in the presence of the ligand, which can be indicative of an interaction. Although TSA is an indirect method and cannot confirm binding with certainty, it provides valuable insights into how the presence of CBD might influence the structural properties of TBR1 T-box domain.

Thermal shift assay

In the case of CBD and the TBR1 T-box domain, the TSA revealed a dose-dependent decrease in Tm, from 63.5°C in the control to 62°C at 200 µM CBD. This gradual decrease might suggest an interaction with CBD, potentially influencing the protein's stability. However, a decrease in Tm is linked to structural destabilisation, necessitating a detailed. Shifts of 1°C or more are generally viewed as substantial, though this threshold may vary with each protein's conformational sensitivity (Niesen *et al.,* 2007). Strikingly, this perceived destabilisation contrasts with bioinformatics analyses (RMSD in Figure 27, RMSF in Figure23 and rGyr in Figure 24) which indicated an increase in stability upon CBD binding. Bioinformatic analyses primarily focus on the energetics and geometry of the binding site at equilibrium, contrasting with TSA's holistic evaluation of the entire protein. This suggests that CBD binding might induce partial unfolding or destabilisation in certain regions of the protein leading to a decreased Tm.

However, several indirect factors could also be influencing these findings. For instance, alterations in the protein's solvation shell, changes in ionic strength, or other environmental factors arising from CBD's presence could affect stability. Moreover, CBD's solubility behaviour in HEPES buffer and its tendency to aggregation may indirectly affect the protein stability. At higher CBD concentrations, which may lead to increased aggregation or solvent effects, the true nature of CBD's interaction with the TBR1 T-box domain might not be accurately reflected. Additionally, the use of methanol as a solvent for CBD introduced additional complexities. With elevated CBD concentrations, there is a corresponding increase in methanol concentration, potentially exerting an indirect influence on protein stability. Methanol, even in moderate concentrations, is known to impact protein structure and dynamics. It can strengthen hydrogen bonds within the protein's polar backbone and weaken hydrophobic interactions in the protein's core (Hwang *et al.,* 2011). This dual effect could result in an expanded hydrophobic core and an increased rGyr, thereby destabilising the protein's structure. Thus, while CBD might interact with the TBR1 T-box domain, the role of methanol as a solvent cannot be overlooked. None of these factors were accounted

for in the bioinformatics results either. To better understand this, control experiments with varying concentrations of methanol, independent of CBD, should be conducted. These would help isolate the solvent's effects and provide a clearer picture of CBD's impact on the TBR1 T-box domain's thermal stability. It is important to note that these CBD concentrations are beyond what is physiologically achievable in the brain (Deiana *et al.,* 2012). This implies that the destabilisation observed in the TSA, if attributed to CBD interactions, might not accurately reflect the *in vivo* brain conditions. Considering the lack of information provided by the TSA on the potential interaction between CBD and TBR1 T-box domain, a TR-FTIR was used to ultimately assess the potential interaction at various time intervals.

Time-Resolved FTIR

TR-FTIR proved advantageous over the PDA and TSA, as it offered real-time observation of the protein-ligand interaction by monitoring changes in the vibrational frequencies of functional groups over time.

A 1 in 4 dilution of 1mg/mL CBD was incubated with 120 μM TBR1 T-box domain in equilibration buffer, over 70 minutes. Figure 31 represents the control spectra examined for the equilibration buffer (A), CBD (B) and TBR1 T-box domain (C) – prior to any interactions. From these graphs, it is clear that the equilibration buffer dominates the FTIR spectra with noticeable peaks emerging around 2272.27 cm^{-1} to 3242.12 cm^{-1} and another around 1635 cm^{-1}. These peaks primarily arise due to the presence of Tris and imidazole in the equilibration buffer. However, in comparison to the TBR1 T-box domain spectrum in the equilibration buffer, distinct peaks at 2974 cm^{-1}, 2898.81 cm^{-1}, 2356.9 cm^{-1}, 1384.79 cm^{-1}, 1240.14 cm^{-1} and 1058.85 cm^{-1} become apparent. These peaks correspond to functional groups involving C-H, C-N, and N-H found within the protein's backbone and side chains (as shown in Figure 31B). Furthermore, the spectrum of isolated CBD in the equilibration buffer exhibits additional peaks at 2968.24 cm^{-1}, 2852.52 cm^{-1}, 2356.85 cm^{-1}, 1541.02 cm^{-1}, 1637.45 cm^{-1}, 1394.44 cm^{-1}, 1242.07 cm^{-1}, 1070.41 cm^{-1} and 1016.42 cm^{-1}, corresponding to functional groups such as C-H, O-H, C-O, and C=C present in CBD (Figure 31C). This is consistent with previous studies, where distinctive peaks are observed at 1620 cm^{-1}; 1638 cm^{-1} and 1590 cm^{-1} for CBD (Vlad *et al.,* 2021), thus confirming the viability of the CBD solution used.

Comparatively, the FTIR spectrum was measured using Equilibration buffer, whereas the PDA and TSA was analysed using HEPES buffer. The equilibration buffer, with its higher ionic strength

(500 mM NaCl) compared to HEPES (100 mM NaCl), could better stabilise protein structures and maintain the integrity of ligand binding sites. This stabilisation is crucial for preserving the native conformation of the protein, ensuring that any observed interactions and structural changes are reflective of the protein's natural state in a cellular environment. By using the equilibration buffer, the FTIR analysis might provide a more accurate representation of the protein-ligand interaction dynamics, potentially revealing interactions or conformational states that could be altered or masked in the less stabilising HEPES buffer.

To assess CBD's binding to the TBR1 T-box domain, these results were juxtaposed with the spectra acquired during a 70-minute incubation (Figure 32). This analysis specifically focuses on the Amide I (1600-1800 cm^{-1}), Amide II (1470-1570 cm^{-1}) and Amide III (1250-1350 cm^{-1}) regions (Figure 33), which represent the backbone of the protein and are indicative of potential binding events. In the Amide I region, the analysis revealed the presence of parallel β-sheets, anti- parallel β-sheets and α-helical structures within the TBR1 T-box domain. These structural features align with the structure of other T-box (Bork *et al.,* 1994) and are further confirmed by the predicted AlphaFold structure (Figure 17A and Figure 18B). In the observed trends, peak intensities within the specified region initially decrease upon exposure to CBD, then subsequently increase as incubation time increases. This pattern implies that CBD binding induces immediate changes in the protein's secondary structure, leading to a destabilising conformation followed by a stabilisation phase Such changes, potentially affecting the hydrogen bonds that maintain secondary structures, may have temporary functional consequences. This is consistent with bioinformatic predictions of an Induced Fit mechanism for CBD interaction. In contrast, Table 5 shows a general reduction in β-sheets and α-helices upon CBD binding. However, FTIR analysis revealed an initial decrease in these structures, followed by an increase over time, suggesting the protein attains a new, more stable conformation in the presence of CBD. This two-phase structural alteration aligns with CBD's role as an allosteric modulator for the CB1 receptor (Laprairie *et al.,* 2015). Such allosteric modulation involves CBD binding to a site separate from the active site, inducing conformational changes that modify the protein's activity.

Similarly, the vibrational changes in arginine (1635.52 cm^{-1}), lysine (1626 cm^{-1}) and asparagine (1626 cm^{-1}) are observed in this region , which follow a similar two-phase trend. This may also indicate an initial disruption in their environment upon CBD binding, which may involve changes in specific bonds (e.g. hydrogen bonds or electrostatic interactions), followed by stabilisation (Tian *et al.,* 2004). The observed vibrational changes might reflect a potential interaction between CBD

and the side chain of these residues. Interestingly, asparagine (Asn240) was predicted to form stable hydrogen bonding between CBD and TBR1 T-box domain (Figure 23). Furthermore, the Amide II region also show changes in glutamine (1558 cm^{-1}), lysine (1522 cm^{-1}) and tyrosine (1518 cm^{-1}) upon CBD binding. This may suggest that CBD interacts with these residues through sided chain selectivity (carboxyl group in glutamine, the amine group in lysine, and the phenolic hydroxyl group in tyrosine). However, these bands only emerged 40-minutes post incubation, indicating a slow conformational change in which rearrangements in the protein tertiary structure leads to exposure of these amino acid residues. Considering these bands are not observed in the protein control spectra suggest the changes are a direct consequence of CBD interaction and not an inherent property of the protein. Interestingly, glutamine (Glu363) was predicted to form hydrogen bonds with CBD in pose1, 2 and 4 – although these interactions did not exceed 30% of the simulation time (Figure 22). This may suggest this interaction may take time to develop, supported by the band only emerging after 40 minutes.

The Amide III region, notable bands are seen at 1064.63 cm^{-1} and 1012.56 cm^{-1}, corresponding to C-O stretching and C=C bending vibrations (Figure 33) (Sigma, Merck; Wang *et al.,* 2023). These peaks are absent in the protein control spectra but undergo noticeable alterations during incubation, suggesting changes in the local environment surrounding C-O stretching and C=C bending groups within the protein structure. During the initial 10-minute incubation, the peak at 1064.63 cm^{-1} disappear entirely, which may signify a short-term response to the presence of CBD. However, the peak re-emerges after 40 minutes, resembling the CBD control spectra. The significance of the peak at 1012.56 cm^{-1} lies in its dynamic behaviour during the incubation process. It exhibits a substantial increase in absorbance after 10 minutes, nearly disappears after 40 minutes and then reappears at the 70-minute mark, albeit with significantly reduced intensity compared to its initial appearance. This suggest that the initial interaction with CBD leads to an immediate and substantial change in the electronic environment of aromatic amino acids. Alternatively, these observed change may also reflect changes in the C=C bonds within CBD as it interacts with the protein. The reappearance of this band, albeit with reduced intensity, suggests the protein-CBD complex established a new equilibrium state which may represent a more stabilised interaction different from the initial binding configuration.

Ultimately, these shifts in FTIR spectra serve as clear indicators of the structural adaptations occurring within proteins due to binding interactions (Volkov and Svergun., 2003). The dynamic character of these interactions is underscored by the time-dependent spectral changes,

highlighting a gradual binding process that possibly becomes more stable over time. It's apparent that longer incubation times are required for more pronounced structural alterations to take place. In this context, the reappearance and intensification of these peaks suggest an ongoing interaction between CBD and TBR1 T-box domain.

In summary, computational analyses proposed a specific hydrophobic interaction between CBD and the TBR1 T-box domain, characterised by an 'induced fit' binding mechanism. Despite the ADMET properties of CBD suggesting its capability to traverse the maternal bloodstream and potentially reach the foetal brain, it is theorised that this interaction does not significantly alter the DNA binding function of TBR1, which is vital for brain development. Nonetheless, the experimental evidence, including the PDA, TSA and TR-FTIR, presented a more complex scenario. The PDA results were inconclusive and while the TSA indicated a potential interaction through a decrease in TBR1's thermal stability in the presence of CBD, these findings also hinted at potential structural destabilisation. This observation contrasts with the computational prediction of a stabilising effect. It is important to note that these discrepancies might be attributable to the specific conditions of the experimental setups. Crucially, the TR-FTIR demonstrated alterations in the protein’s secondary structure upon exposure to CBD, confirming an interaction.

CONCLUSION

The increasing use of CBD by pregnant women as a remedy for symptoms like pain, nausea and anxiety raises significant concerns about its safety, especially its impact on foetal brain development. The capacity of CBD to traverse the placental barrier and enter the foetal bloodstream is a critical concern, compounded by the limited understanding of maternal CBD exposure's effects on foetal development. This study was designed to address these research gaps, aiming to provide expectant mothers with trustworthy, scientifically grounded insights into the safety of CBD use during pregnancy. It specifically focused on the interaction between CBD and the TBR1 T-box domain, a transcription factor implicated in ASD.

The study employed a blend of bioinformatic and experimental techniques to explore the potential interaction between CBD and the TBR1 T-box domain. Bioinformatics analysis suggested that CBD might bind within a hydrophobic pocket of the TBR1 T-box domain, stabilised by additional hydrogen bonding interactions with residues Ser238, Pro335, Thr360, Glu363 and Asn240. The computational findings proposed favourable binding under an induced-fit model, potentially

following an allosteric binding mechanism. Molecular dynamics analysis, including RMSD, RMSF, and ΔRg, pointed to a stabilising effect of CBD on the structure of the TBR1 T-box domain. However, transitioning from computational predictions to empirical validation faced hurdles. However, transitioning from computational predictions to empirical validation faced hurdles. The PDA and TSA did not provide definitive confirmation of the interaction. Fortunately, TR-FTIR emerged as a crucial technique, confirming the CBD-TBR1 T-box binding interaction. Additionally, the bioinformatic results did not conclusively demonstrate any direct implication of CBD binding on the DNA binding ability of the TBR1 T-box domain. Supported by ADMET properties, the potential of CBD reaching and affecting the foetal brain is evident, but the precise nature and consequences of its interaction with TBR1, critical in brain development, remain ambiguous and necessitate further research. These results collectively underscore the importance of a cautious approach towards the use of CBD during pregnancy. Although direct effects of CBD on foetal brain development, especially regarding neurological structures, are not definitively established, the potential interactions highlighted by this research underscore the urgency for additional studies. This work also stresses the importance of informed decision-making and regulatory oversight in the context of CBD use by pregnant individuals.

Limitations

The study encountered several limitations, chief among them being the inconclusive experimental validation of in silico predictions, which suggested a potential interaction between CBD and the TBR1 T-box domain. Empirical verification faced considerable challenges, leading to outcomes that were unreliable and left the proposed hypotheses unconfirmed. Specifically, the Pull-Down Assay (PDA) struggled with effectively coupling CBD to the epoxy-activated agarose resin due to CBD's hydrophobic nature and limited solubility, resulting in inconclusive evidence of interaction. The Thermal Shift Assay (TSA) further complicated matters by presenting ambiguous results: a decrease in the thermal stability of the TBR1 T-box domain in the presence of CBD, contradicting the increased stability suggested by bioinformatics analyses and indicating potential influences from external factors like solvent effects and CBD's aggregation behaviour. Moreover, while Time-Resolved Fourier Transformed Infrared Spectroscopy (TR-FTIR) provided valuable insights into the CBD-TBR1 interaction over time, further experimentation is needed to fully grasp the complexities of this interaction. Lastly, the study would have benefited from completing the experiments in triplicate, given a larger supply of CBD was available for use.

Future Recommendations

Future research should focus on refining experimental techniques to confirm the interaction between CBD and the TBR1 T-box domain, exploring alternative analytical methods such as surface plasmon resonance or nuclear magnetic resonance spectroscopy for enhanced characterisation. Integrating computational predictions with experimental data and expanding the scope of study to investigate additional molecular targets implicated in neurodevelopmental disorders is crucial. Longitudinal studies and clinical trials are necessary to assess the safety and efficacy of CBD use during pregnancy, while educational initiatives and public health interventions are needed to raise awareness about potential risks among healthcare providers and expectant mothers.

References:

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders. 5th ed. Arlington: American Psychiatric Publishing. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3777342/

Anderson LC (1980) Leaf variation among Cannabis species from a controlled garden. Harvard University Botanical Museum Leaflets. 28(1):61–69. http://www.jstor.org/stable/41762825.

Anfinsen, C. B., Haber, E., Sela, M., White, F. H. (1961). The Kinetics of Formation of Native Ribonuclease during Oxidation of the Reduced Polypeptide Chain. *Proc. National Academy of Science.* 47: 1309–1314. DOI:10.1073/pnas.47.9.1309

Arakawa, T., Philo, J.S., Tsumoto, K., Yumioka, R. Ejima, D. (2004). Elution of antibodies from a Protein-A column by aqueous arginine solutions. *Protein Expression and Purification.* 36(2):244–248. DOI: 10.1016/j.pep.2004.04.009.

Arnau, J., Lauritzen, C., Petersen, G. E., Pedersen, J. (2006). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expression and Purification. 48(1): 1–13. DOI: 10.1016/j.pep.2005.12.002.

Autism National Institute of Environmental Health Sciences. (2020). Available from: https://www.niehs.nih.gov/health/topics/conditions/autism/index.cfm. Accessed 29 November 2023.

Aviz-Amador, A., Contreras-Puentes, N., Mercado-Camargo, J. (2021). Virtual screening using docking and molecular dynamics of cannabinoid analogs against CB1 and CB2 receptors. *Computational Biology and Chemistry*. 95: 107590. DOI: 10.1016/j.compbiolchem.2021.107590

Babalonis, S., Haney, M., Malcolm, R. J., Lofwall, M. R., Votaw, V. R., Sparenborg, S., & Walsh, S. L. (2017). Oral cannabidiol does not produce a signal for abuse liability in frequent marijuana smokers. *Drug and Alcohol Dependence*. 172: 9-13. DOI: 10.1016/j.drugalcdep.2016.11.030

Baban, B., Khodadadi, H., Salles, É. L., Costigliola, V., Morgan, J. C., Hess, D. C., Vaibhav, K., Dhandapani, K. M., Yu, J. C. (2021). Inflammaging and Cannabinoids. *Ageing Research Reviews*. 72: 101487. DOI: 10.1016/j.arr.2021.101487.

Badowski, S., Smith, G. (2020). Cannabis use during pregnancy and postpartum. *Canadian Family Physician*. 66(2): 98-103. PMC7021337

Baker, D., Pryce, G., Croxford, J.L., Brown, P., Pertwee, R.G., Huffman, J.W., Layward, L. (2000). Cannabinoids control spasticity and tremor in a multiple sclerosis model. *Nature*. 404(6773): 84–87. DOI:10.1038/35003583.

Barchel, D., Stolar, O., De-Haan, T., Ziv-Baran, T., Saban, N., Fuchs, D. O. (2019). Oral cannabidiol use in children with autism spectrum disorder to treat related symptoms and improve the quality of life. *Frontiers in Pharmacology*. 9(9): 1521. DOI: 10.3389/fphar.2018.01521

Barth, A. (2007). Infrared spectroscopy of proteins. *Biochimica et biophysica acta*. 1767(9): 1073–1101. DOI: 10.1016/j.bbabio.2007.06.004

Batkai, S., Pacher, P., Osei-Hyiaman, D., Radaeva, S., Liu, J., Harvey-White, J., Kunos, G. (2007). Endocannabinoids acting at cannabinoid-1 receptors regulate cardiovascular function in hypertension. *Circulation*. 110(14): 1996-2002. DOI: 10.1161/01.CIR.0000143230.23252.D2

Beatty, J. R., Svikis, D. S., Ondersma, S. J. (2012). Prevalence and perceived financial costs of marijuana versus tobacco use among urban low-income pregnant women. *Journal of Addictive Behaviors.* 3(4): 1000135. DOI: 10.4172/2155-6105.1000135

Bedogni, F., Hodge, R. D., Elsen, G. E., Nelson, B. R., Daza, R. A. M., Beyer, R. P., Bammler, T. K., Rubenstein, J. L. R., Hevner, R. F. (2010). Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex. *Proceedings of the National Academy of Sciences of the United States of America.* 107(29): 13129-13134. DOI:10.1073/pnas.1002285107.

Berghuis, P., Rajnicek, A. M., Morozov, Y. M., Ross, R. A., Mulder, J., Urbán, G. M., Monory, K., Marsicano, G., Matteoli, M., Canty, A., Irving, A. J., Katona, I., Yanagawa, Y., Rakic, P., Lutz, B., Mackie, K., Harkany, T. (2007). Hardwiring the brain: endocannabinoids shape neuronal connectivity. *Science*. 316(5828): 1212–1216. DOI: 10.1126/science.1137406

Berman, E., Erenburg, N., Beloosesky, R., Eyal, S., Kovo, M. (2023). Placental disposition of cannabidiol: An ex vivo perfusion study. *Epilepsia*. 64(12): 3354–3364. DOI: 10.1111/epi.17778

Berndt, K. (1996). A guide to simple, direct, and quantitative in vitro binding assays. *Journal of Biological Methods*. 4(1): 62. DOI: 10.14440/jbm.2017.161

Bidwell, C.L., York Williams, S.L., Mueller, R.L., Bryan, A.D., Hutchison, K.E. (2018). Exploring cannabis concentrates on the legal market: user profiles, product strength, and health-related outcomes. *Addictive Behaviors Reports*. 8: 102–106. DOI: 10.1016/j.abrep.2018.08.004

Bietz, S., Urbaczek, S., Schulz, B. (2014). Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. Journal Cheminform. 6: 12. DOI: 10.1186/1758-2946-6-12

Binkowski, T. A., Naghibzadeh, S., Liang, J. (2003). CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic acids research*. 31(13): 3352–3355. DOI: 10.1093/nar/gkg512

Biocare (2014). 'Biocare'. Available at: https://biocare.net/wp-content/uploads/TWN20.pdf.

Bork, P., Holm, L., Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. *Journal of Molecular Biology*. 242(4): 309-320. DOI: 10.1006/jmbi.1994.1582.

Bornhorst, J. A., Falke, J. J. (2000). Purification of Proteins Using Polyhistidine Affinity Tags. *Methods Enzymology*. 326: 245–254. DOI: 10.1016/s0076-6879(00)26058-8.

Brown, Q. L., Sarvet, A. L., Shmulewitz, D., Martins, S. S., Wall, M. M., Hasin, D. S. (2017). Trends in marijuana use among pregnant and nonpregnant reproductive-aged women, 2002–2014. *JAMA*. 317(2): 207–209. DOI: 10.1001/jama.2016.17383

Brown, Q. L., Sarvet, A. L., Shmulewitz, D., Martins, S. S., Wall, M. M., Hasin, D. S. (2017). Trends in Marijuana Use Among Pregnant and Nonpregnant Reproductive-Aged Women, 2002-2014. *JAMA*. *317*(2): 207–209. DOI: 10.1001/jama.2016.17383

Bruylants, G., Wouters, J., Michaux, C. (2005). Differential scanning calorimetry in life science: thermodynamics, stability, molecular recognition and application in drug design.*Current medicinal chemistry.*12(17):2011–2020. DOI: 10.2174/0929867054546564

Bulfone, A., Smiga, S. M., Shimamura, K., Peterson, A., Puelles, L., Rubenstein, J. L. (1995). T-brain-1: a homolog of Brachyury whose expression defines molecularly distinct domains within the cerebral cortex. *Neuron*. 15(1): 63-78. DOI:10.1016/0896-6273(95)90065-9.

Bulfone, A., Wang, F., Hevner, R., Anderson, S., Cutforth, T., Chen, S., Meneses, J., Pedersen, R., Axel, R., Rubenstein, J. L. (1998). An olfactory sensory map develops in the absence of normal

projection neurons or GABAergic interneurons. *Neuron*. 21(6): 1273-1282. DOI: 10.1016/s0896-6273(00)80647-9.

Cabral, G. A., Griffin-Thomas, L. (2009). Emerging role of the cannabinoid receptor CB2 in immune regulation: therapeutic prospects for neuroinflammation. *Expert Reviews in Molecular Medicine.* 11: 3. DOI: 10.1017/S1462399409000957

Cabral, G. A., Jamerson, M. (2014). Marijuana use and brain immune mechanisms.. *International Review of Neurobiology*. 118: 199-230. DOI: 10.1016/B978-0-12-801284-0.00008-7

Calapai, F., Cardia, L., Sorbara, E. E., Navarra, M., Gangemi, S., Calapai, G., Mannucci, C. (2020). Cannabinoids, Blood-Brain Barrier, and Brain Disposition. *Pharmaceutics*. 12(3): 265. DOI: 10.3390/pharmaceutics12030265.

Campolongo, P., Trezza, V., Ratano, P., Palmery, M., Cuomo, V., Trabace, L. (2011). Developmental consequences of perinatal cannabis exposure: Behavioral and neuroendocrine effects in adult rodents. *Psychopharmacology*. 214(1): 5-15. DOI: 10.1007/s00213-010-1892-x

Carmack, M., Kelley, C.J. (1968). The Synthesis of the Optically Active Cleland Reagent[(—)-1,4-Dithio-Lg-threitol. *Journal of Organic Chemistry*. 33(5): 2171–2173. DOI: 10.1021/jo01269a123.

Castro-Alvarez, A., Costa, A.M., Vilarrasa, J. (2017). The Performance of Several Docking Programs at Reproducing Protein-Macrolide-Like Crystal Structures. *Molecules*. 22(1):136. DOI:10.3390/molecules22010136

Celej, M. S., Montich, G. G., Fidelio, G. D. (2003). Protein stability induced by ligand binding correlates with changes in protein flexibility. *Protein Science : A Publication of the Protein Society*. *12*(7): 1496-1506. DOI: 10.1110/ps.0240003

Chayasirisobhon, S. (2021). Mechanisms of Action and Pharmacokinetics of Cannabis. The Permanente Journal. 25(1): 1-3. DOI: 10.7812/TPP/19.200

Christian, G., O'Reilly, J. (1986). 'Instrumental analysis (Book, 1986) [WorldCat.org]', WorldCat.

Chuang, H.C., Huang, T.N., Hsueh, Y.P. (2014). Neuronal excitation upregulates Tbr1, a high-confidence risk gene of autism, mediating Grin2b expression in the adult brain. *Front Cell Neuroscience.* 8(10): 280. DOI: 10.3389/fncel.2014.00280.

Chung, C. T., Niemela, S. L., Miller, R. H. (1989). One-step preparation of competent Escherichia coli: Transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences of the United States of America*. *86*(7): 2172-2175. DOI: 10.1073/pnas.86.7.2172

Clancy, S. (2008). DNA Transcription. *Nature*. 1(1): 41. Available at: https://www.nature.com/scitable/topicpage/dna-transcription-426/

Committee on Obstetric Practice. (2017). Committee opinion No. 722: marijuana use during pregnancy and lactation. *Obstet Gynecol*. 130(4): 205-209. DOI: 10.1097/AOG.0000000000002354

Cooper, Z. D., Haney, M. (2008). Cannabis reinforcement and dependence: role of the cannabinoid CB1 receptor. *Addict Biology.* 13(2): 188-95. DOI: 10.1111/j.1369-1600.2007.00095.x

Corsi, D. J., Donelle, J., Sucha, E., Hawken, S., Hsu, H., El-Chaâr, D. (2020). Maternal cannabis use in pregnancy and child neurodevelopmental outcomes. National Medicine. 26(10): 1536–1540. DOI: 10.1038/s41591-020-1002-5

Corsi, D. J., Hsu, H., Weiss, D., Fell, D. B., Walker, M. (2019). Trends and correlates of cannabis use in pregnancy: a population-based study in Ontario, Canada from 2012 to 2017.*Canadian journal of public health. 110*(1): 76–84. DOI: 10.17269/s41997-018-0148-0

Corsi, D. J., Walsh, L., Weiss, D., Hsu, H., El-Chaar, D., Hawken, S., Hutton, E. K. (2020). Association between self-reported prenatal cannabis use and maternal, perinatal, and neonatal outcomes. JAMA. 324(15): 159-169. DOI: 10.1001/jama.2019.8734

Crampton, H. (2015). *Dagga: A short history*. Eastern Cape: Jacana Media

Crick, F. (1970). Central dogma of molecular biology. Nature. 227(5258): 561–562. DOI: 10.1038/227561a0.

Customs and Excise Duties Amendment Act 35 of 1922.

Daina, A., Michielin, O., Zoete, V. (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*. 7(42717). DOI: 10.1038/srep42717.

Dayanandan, P., and Kaufman, P.B. (1976) Trichomes of Cannabis Sativa L. (CANNABACEAE). *American Journal of Botany*. 63(5): 578–591. DOI: 10.2307/2441821

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., Buxbaum, J. D. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. 515(7526): 209-215. DOI: 10.1038/nature13772

Deiana, S., Watanabe, A., Yamasaki, Y., Amada, N., Arthur, M., Fleming, S., Woodcock, H., Dorward, P., Pigliacampo, B., Close, S., Platt, B., Riedel, G. (2012). Plasma and brain pharmacokinetic profile of cannabidiol (CBD), cannabidivarine (CBDV), delta(9)-tetrahydrocannabivarin (THCV) and cannabigerol (CBG) in rats and mice following oral and intraperitoneal administration and CBD action on obsessive-compulsive behaviour. *Psychopharmacology*. 219(3): 859–73. DOI: 10.1007/s00213-011-2415-0

Deriziotis, P., O'Roak, B.J., Graham, S.A., Estruch, S.B., Dimitropoulou, D., Bernier, R.A., Gerdts, J., Shendure, J. (2014). De novo TBR1 mutations in sporadic autism disrupt protein functions. Nature Communications. 5. DOI: 10.1038/ncomms5954.

Devinsky, O., Cilio, M. R., Cross, H., Fernandez-Ruiz, J., French, J., Hill, C., Katz, R., Di Marzo, V., Jutras-Aswad, D., Notcutt, W. G., Martinez-Orgado, J., Robson, P. J., Rohrback, B. G., Thiele, E., Whalley, B., Friedman, D. (2014). Cannabidiol: pharmacology and potential therapeutic role in epilepsy and other neuropsychiatric disorders. *Epilepsia*. 55(6): 791-802. DOI: 10.1111/epi.12631.

DiGuiseppi, C., Crume, T., Van Dyke, J., Sabourin, K. R., Soke, G. N., Croen, L. A., Daniels, J. L., Lee, L. C., Schieve, L. A., Windham, G. C., Friedman, S., Robinson Rosenberg, C. (2021). Peri-pregnancy cannabis use and autism spectrum disorder in the offspring: Findings from the study to explore early development. *Journal of Autism and Developmental Disorders*. 52(11): 5064 – 5071. DOI: 10.1007/s10803-021-05339-4.

Downer, E. J., Clifford, E., Amu, S., Fallon, P. G., Moynagh, P. N. (2012). The Synthetic Cannabinoid R(+)WIN55,212-2 Augments Interferon-β Expression via Peroxisome Proliferator-activated Receptor-α. *The Journal of Biological Chemistry*. 287(30): 25440-25453. DOI: 10.1074/jbc.M112.371757

Drugs and Drug Trafficking Act 140 of 1992.

Edelhoch, H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*. 6(7): 1948–1954. DOI: 10.1021/bi00859a010

Edelsbrunner, H. (1995). The union of balls and its dual shape. *Springer Link*. 13(1): 415-440. DOI: 10.1007/BF02574053

Edelsbrunner, H., Facello, M., Fu, P., Liang, J. (1995). Measuring proteins and voids in proteins. *Semantic Scholar.* 88(5): 83-102. DOI: 10.1109/HICSS.1995.375331

Edelsbrunner, H., Facello, M., Liang, J. (1998). On the definition and the construction of pockets in macromolecules. *ScienceDirect*. 88(3): 83-102. DOI: 10.1016/S0166-218X(98)00067-5

Edelsbrunner, H., Mucke, E. P. (1994). Three-dimensional alpha shapes. *ACM Transactions on Graphics.* 13(1): 3-72. DOI: 10.1145/174462.156635

Edelsbrunner, H., Shah, N. R. (1996). Incremental topological flipping works for regular triangulations. Springer Link. 15(1): 223-241. DOI: 10.1007/BF01975867

Elmes, M.W., Kaczocha, M., Berger, W.T., Leung, K.; Ralph, B.P.; Wang, L.; Sweeney, J.M.; Miyauchi, J.T.; Tsirka, S.E.; Ojima, I. (2015). Fatty acid-binding proteins (FABPs) are intracellular carriers for Δ9-tetrahydrocannabinol (THC) and cannabidiol (CBD). *J. Biol. Chem*. 290: 8711–8721. DOI:10.1074/jbc.M114.618447.

ElSohly M. A., Slade D. (2005). Chemical constituents of marijuana: the complex mixture of natural cannabinoids. Life Sciences. 78(5): 539–548. DOI: 10.1016/j.lfs.2005.09.011

Fairbanks, G., Steck, T.L, Wallach, D.F.H. (1971). Electrophoretic Analysis of the Major Polypeptides of the Human Erythrocyte Membrane. *Biochemistry*. 10(13):2606–2617. DOI: 10.1021/bi00789a030.

Fride, E., Gobshtis, N., Dahan, H., Weller, A., Giuffrida, A., Ben-Shabat, S. (2009). The endocannabinoid system during development: Emphasis on perinatal events and delayed effects. *Vitamins and Hormones*. 81: 139–158. DOI: 10.1016/s0083-6729(09)81006-6.

Fried, P. A., Smith, A. M. (2001). A literature review of the consequences of prenatal marihuana exposure. An emerging theme of a deficiency in aspects of executive function. *Neurotoxicology and teratology*. 23(1): 1–11. DOI: 10.1016/s0892-0362(00)00119-7

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shaw, D. E., Shelley, M., Perry, J. K., Francis, P., Shenkin, P. S. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*. 47: 1739–1749. DOI: 10.1021/jm0306430

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., Mainz, D. T. (2006). Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *Journal of Medicinal Chemistry*. 49: 6177–6196. DOI: 10.1021/jm051256o

Fride, E., Gobshtis, N., Dahan, H., Weller, A., Giuffrida, A., Ben-Shabat, S. (2009). The endocannabinoid system during development: emphasis on perinatal events and delayed effects. *Vitamins and hormones*. 81: 139–158. DOI: 10.1016/S0083-6729(09)81006-6

Galve-Roperh, I., Chiurchiù, V., Díaz-Alonso, J., Bari, M., Guzmán, M., & Maccarrone, M. (2013). Cannabinoid receptor signaling in progenitor/stem cell proliferation and differentiation. *Progress in lipid research*. 52(4): 633–650. DOI: 10.1016/j.plipres.2013.05.004

Goldschmidt, L., Richardson, G.A., Willford, J., Severtson, S.G., Day, N.L. (2008) Prenatal marijuana exposure and intelligence test performance at age 6. *Journal of the American Academy of Child and Adolescent Psychiatry*. 51(3): 247-263. DOI: 10.1097/CHI.0b013e318160b3f0.

Gomes, L., Monteiro, G., Mergulhão, F. (2020). The Impact of IPTG Induction on Plasmid Stability and Heterologous Protein Expression by Escherichia coli Biofilms. *International Journal of Molecular Science*. 21(2):576. DOI:10.3390/ijms21020576

Grandl, J., Kim, S. E., Uzzell, V., Bursulaya, B., Petrus, M., Bandell, M., Patapoutian, A. (2010). Temperature-induced opening of TRPV1 ion channel is stabilized by the pore domain. *Nature Neuroscience*. 13(6): 708. DOI: 10.1038/nn.2552

Grifoni, L., Vanti, G., Donato, R., Sacco, C., Bilia, A. R. (2022). Promising Nanocarriers to Enhance Solubility and Bioavailability of Cannabidiol for a Plethora of Therapeutic Opportunities. *Molecules*. *27*(18). DOI: 10.3390/molecules27186070

Grimsley, G. R., Pace, C. N. (2004). Spectrophotometric determination of protein concentration. *Current protocols in protein science*. 3: 3.1.1–3.1.9. DOI: 10.1002/0471140864.ps0301s33

Grotenhermen, F., Müller-Vahl, K. (2012). The therapeutic potential of Cannabis and cannabinoids. *Dtsch Arztebl Int*. 109(29-30): 495-501. DOI: 10.3238/arztebl.2012.0495

Grotenhermen, F., Russo, E. (2002). Cannabis and cannabinoids: pharmacology, toxicology, and therapeutic potential. New York: Psychology Press.

Gunn, J.K.L., Rosales, C.B., Center, K.E., Nuñez, A., Gibson, S.J., Christ, C., Ehiri, J.E. (2016) Prenatal exposure to cannabis and maternal and child health outcomes: a systematic review and meta-analysis. *BMJ Open*. 6(4): e009986. DOI: 10.1136/bmjopen-2015-009986.

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry.* 47: 1750–1759. DOI: 10.1021/jm030644s

Hampson, A.J., Grimaldi, M., Axelrod, J., Wink, D. (1998). Cannabidiol and (-) 9-tetrahydrocannabinol are neuroprotective antioxidants. *Proceedings of the National Academy of Sciences*. 95(14): 8268–8273. DOI:10.1073/pnas.95.14.8268.

Han, W., Kwan, K. Y., Shim, S., Lam, M. M. S., Shin, Y., Xu, X., Zhu, Y., Li, M., Sestan, N. (2011). TBR1 directly represses Fezf2 to control the laminar origin and development of the corticospinal tract. *Proceedings of the National Academy of Sciences of the United States of America*. 108(7): 3041-3046. DOI:10.1073/pnas.1016723108.

Hanuš L.O., Hod Y. (2020). Terpenes/Terpenoids in Cannabis: Are They Important? Med. Cannabis Cannabinoids. 3:25–60. DOI: 10.1159/000509733.

Hanuš, L. O., Tchilibon, S., Ponde, D. E., Breuer, A., Fride, E., Mechoulam, R. (2005). Enantiomeric cannabidiol derivatives: synthesis and binding to cannabinoid receptors. Organic & Biomolecular Chemistry. 3(6):1116-23. DOI: 10.1039/b416943c

Harkany, T., Keimpema, E., Barabás, K., Mulder, J. (2008). Endocannabinoid functions controlling neuronal specification during brain development. *Molecular and Cellular Endocrinology.* 286: S84-90. DOI: 10.1016/j.mce.2008.02.011.

Harvey, D. J., Brown, N. K. (1991). Comparative in vitro metabolism of the cannabinoids. *Pharmacology, Biochemistry, and Behavior.* 40(3): 533-540. DOI: 10.1016/0091-3057(91)90359-a.

Hasegawa, T. (2017). Fundamentals of FT-IR. In: Quantitative Infrared Spectroscopy for Understanding of a Condensed Matter. Springer. DOI: 10.1007/978-4-431-56493-5_2

Haworth, K., Putt, W., Cattanach, B., Breen, M., Binns, M., Lingaas, F., Edwards, Y. H. (2001). Canine homolog of the T-box transcription factor T; failure of the protein to bind to its DNA target leads to a short-tail phenotype. *Mamm Genome*. 12(3): 212–218. DOI: 10.1007/s003350010245

Herkenham, M., Lynn, A. B., Johnson, M. R., Melvin, L. S., de Costa, B. R., Rice, K. C. (1991). Characterization and localization of cannabinoid receptors in rat brain: a quantitative in vitro autoradiographic study. *The Journal of neuroscience*. 11(2): 563–583. DOI: 10.1523/JNEUROSCI.11-02-00563.1991

Herrmann, B. G., Kispert, A. (1994). The T genes in embryogenesis. *Trends Genetics*. 10: 280–286. DOI: 10.1016/0168-9525(94)90042-6

Herrmann, B. G., Labeit, S., Poustka, A., King, T. R., Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse. *Nature*. 343: 617–622. DOI: 10.1038/343617a0

Hevner, R. F., Shi, L., Justice, N., Hsueh, Y., Sheng, M., Smiga, S., Bulfone, A., Goffinet, A. M., Campagnoni, A. T., Rubenstein, J. L. (2001). Tbr1 regulates differentiation of the preplate and layer 6. *Neuron*. 29(2): 353-366. DOI:10.1016/s0896-6273(01)00211-2.

Hevner, R. F., Shi, L., Justice, N., Hsueh, Y., Sheng, M., Smiga, S., Bulfone, A., Goffinet, A. M., Campagnoni, A. T., Rubenstein, J. L. (2001). Tbr1 regulates differentiation of the preplate and layer 6. *Neuron*. 29(2): 353–366. DOI: 10.1016/s0896-6273(01)00211-2

Hollingsworth, S. A., Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron*. *99*(6): 1129. DOI: 10.1016/j.neuron.2018.08.011

Hollister, L. E. (1974). Structure-Activity Relationships in Man of Cannabis Constituents and Homologs and Metabolites of Δ9-tetrahydrocannabinol. Pharmacological Reviews. 11(1): 3-11. DOI: 10.1159/000136462.

Huang, T.N., Chuang, H.C., Chou, W.H., Chen, C.Y., Wang, H.F., Chou, S.J., Hsueh, Y.P. (2014). Tbr1 haploinsufficiency impairs amygdalar axonal projections and results in cognitive abnormality. *Nature Neuroscience*. 17(2): 240-247. DOI:10.1038/nn.3626.

Huynh, K., Partch, C. L. (2015). Analysis of protein stability and ligand interactions by CD shift assay. *Current Protocols in Protein Science*. 79 (28): .1–28.9.14. DOI: 10.1002/0471140864.ps2809s79.

Hwang, S., Shao, Q., Williams, H., Hilty, C., Gao, Y. Q. (2011). Methanol Strengthens Hydrogen Bonds and Weakens Hydrophobic Interactions in Proteins – A Combined Molecular Dynamics and NMR study. *Journal of Physical Chemistry B*. 115(20): 6653–6660. DOI: 10.1021/jp111448a.

Hyman, S. L., Levy, S. E., Myers, S. M., Council on Children with Disabilities Section on Developmental and Behavioral Pediatrics. (2020). Identification, evaluation, and management of children with autism spectrum disorder. *National Library of Medicine*. 145(1): 20193447. DOI: 10.1542/peds.2019-3447.

Hyötyläinen, T., McGlinchey, A., Salihovic, S., Schubert, A., Douglas, A., Hay, D. C. (2024). In utero exposures to perfluoroalkyl substances and the human fetal liver metabolome in Scotland: a cross-sectional study. *8(1):* E5-E17. DOI: 10.1016/S2542-5196(23)00257-7.

Ichikawa, M., Masashi, O., Goto, H. (2023). Polyaniline Hybrids with Biological Tissue, and Biological Polymers as Physiological—Electroactive Materials. *Microbiology*. 3: 172-191. DOI: 10.3390/micro3010013.

Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2023; Prime, Schrödinger, LLC, New York, NY, 2023.

Iossifov, I., Levy, D., Allen, J., Ye, K., Ronemus, M., Lee, Y.-H., Yamrom, B., Wigler, M. (2015). Low load for disruptive mutations in autism genes and their biased transmission. *Proceedings of the National Academy of Sciences of the United States of America*. 112(41): 5600-5607. DOI:10.1073/pnas.1516376112

Izzo, A. A., Borrelli, F., Capasso, R., Di Marzo, V., Mechoulam, R. (2009). Non-psychotropic plant cannabinoids: new therapeutic opportunities from an ancient herb. *Trends in Pharmacological Sciences.* 30(10): 515-527. DOI: 10.1016/j.tips.2009.07.006

Jaidee, W., Siridechakorn, I., Nessopa, S., Wisuitiprot, V., Chaiwangrach, N., Ingkaninan, K., Waranuch, N. (2022). Kinetics of CBD, Δ9-THC Degradation and Cannabinol Formation in Cannabis Resin at Various Temperature and pH Conditions. *Cannabis and Cannabinoid Research*. *7*(4): 537-547. DOI: 10.1089/can.2021.0004

Ji, Y., Yang, X., Ji, Z., Zhu, L., Ma, N., Chen, D., Jia, X., Tang, J., Cao, Y. (2020). DFT-Calculated IR Spectrum Amide I, II, and III Band Contributions of N-Methylacetamide Fine Components. *ACS Omega*. 5(15): 8572-8578. DOI: 10.1021/acsomega.9b04421.

Jiang, C., Glorioso, J.C., Ataai, M. (2006). Presence of imidazole in loading buffer prevents formation of free radical in immobilized metal affinity chromatography and dramatically improves the recovery of herpes simplex virus type 1 gene therapy vectors. *Journal of Chromatography* A. 1121(1):40–45. DOI: 10.1016/j.chroma.2006.04.071.

Johnston, R. C.; Yao, K.; Kaplan, Z.; Chelliah, M.; Leswing, K.; Seekins, S.; Watts, S.; Calkins, D.; Chief Elk, J.; Jerome, S. V.; Repasky, M. P;. Shelley, J. C. (2023). Epik: pKa and protonation state prediction through machine learning. *J. Chem. Theory Computational. 19 (8)*: 2380–2388. DOI: 10.1021/acs.jctc.3c00044.

Jones, G., Pertwee, R. G., Gill, E.W., Paton, W. D., Nilsson, I. M., Widman, M. Agurell, S. (1974) Relative pharmacological potency in mice of optical isomers of i:11tetrahydrocannabinol, Biochemical Pharmacology. 23: 439-446

Jorgensen, W. L., Duffy, E. M. (2000). Prediction of drug solubility from Monte Carlo simulations. *Bioorganic & medicinal chemistry letters*. 10(11): 1155–1158. DOI: 10.1016/s0960-894x(00)00172-4

Jugl, S., Okpeku, A., Costales, B., Morris, E. J., Alipour-Haris, G., Hincapie-Castillo, J. M., Goodin, A. (2021). A Mapping Literature Review of Medical Cannabis Clinical Outcomes and Quality of Evidence in Approved Conditions in the USA from 2016 to 2019. *Medical Cannabis and Cannabinoids*. 4: 21-42. DOI: 10.1159/000515069

Jumper, J., Evans, R., Pritzel, A. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*. 596: 583–589. DOI:10.1038/s41586-021-03819-2.

Jumper, J., Evans, R., Pritzel, A. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*. 596, 583–589. DOI: 10.1038/s41586-021-03819-2

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. Acta Crystallophraphica. 32(5): 922-923. DOI: 10.1107/S0567739476001873

Kaminski, G., Friesner, R., Tirado-Rives, J., Jorgensen, W. (2001). Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate

Quantum Chemical Calculations on Peptides †. *The Journal of Physical Chemistry* B. 105: 6474-6487. DOI: 10.1021/jp003

Karhson, D. S., Krasinska, K. M., Dallaire, J. A., Libove, R. A., Philips, J. M., Chien, A. S. (2018). Plasma anandamide concentrations are lower in children with autism. *Molecular Autism*. 9(1): 18. DOI: 10.1186/s13229-018-0203-y

Karplus, M., McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature structural biology*. 9(9): 646–652. DOI: 10.1038/nsb0902-646

Kelly, S. M., Jess, T. J., Price, N. C. (2005). How to study proteins by circular dichroism. *Proteins and Proteomics*. 1751(2): 119–139. DOI: 10.1016/j.bbapap.2005.06.005.

Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw, "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters," Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, November 11-17, 2006.

Ko, J. Y., Tong, V. T., Bombard, J. M., Hayes, D. K., Davy, J., Perham-Hester, K. A. (2018). Marijuana use during and after pregnancy and association of prenatal use on birth outcomes: A population-based study. *Drug and alcohol dependence.* 187: 72–78. DOI: 10.1016/j.drugalcdep.2018.02.017

Koshland Jr, D. E. (1995) The key–lock theory and the induced fit theory. *Angew. Chem. Int. Ed..* 33: 2375–2378. DOI: 10.1002/anie.199423751

Lapetina, S., Gil-Henn, H. (2017). A guide to simple, direct, and quantitative in vitro binding assays. Journal of Biological Methods. 4(1): 62. DOI:10.14440/jbm.2017.161.

Laprairie, R. B., Bagher, A. M., M Kelly, M. E. (2015). Cannabidiol is a negative allosteric modulator of the cannabinoid CB1 receptor. *British Journal of Pharmacology*. *172*(20): 4790-4805. DOI: 10.1111/bph.13250

Law, B., Mason, P. A., Moffat, A. C., Gleadle, R. I., King, L. J. (1984). Forensic aspects of the metabolism and excretion of cannabinoids following oral ingestion of cannabis resin. *Journal of Pharm Pharmacology*. 36(5): 289-294. DOI: 10.1111/j.2042-7158.1984.tb04376.x.

Lawrence, D.K., Gill, E.W. The effects of Δ1-tetrahydrocannabinol and other cannabinoids on spin-labeled liposomes and their relationship to mechanisms of general anesthesia. (1975). Molecular Pharmacology. 11(5) :595–602. PMID: 170503.

Lederberg, E. M., Cohen, S. (1974). Transformation of Salmonella typhimurium by plasmid deoxyribonucleic acid. *Journal of bacteriology*. 119(3): 1072-4. DOI:10.1128/jb.119.3.1072-1074.1974

Lemberger, L., Silberstein, S.D., Axelrod, J., Kopin, I.J. (1970). Marihuana: Studies on the Disposition and Metabolism of Delta-9-Tetrahydrocannabinol in Man. *Science*. 170(3964): 1320–1322. DOI:10.1126/science.170.3964.1320.

Leone, D. P., Srinivasan, K., Chen, B., Alcamo, E., McConnell, S. K. (2008). The determination of projection neuron identity in the developing cerebral cortex. *Current Opinion in Neurobiology*. 18(1): 28-35. DOI:10.1016/j.conb.2008.05.006.

Li, H.L., Lin, H. (1974) An archaeological and historical account of cannabis in China. *Economic Botany.* 28(4): 437-447. http://www.jstor.org/stable/4253540.

LigPrep, Schrödinger, LLC, New York, NY, 2023.

Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O., Shaw, D. E. (2012). Systematic Validation of Protein Force Fields against Experimental Data. *PLOS ONE*. *7*(2): 32131. DOI: 10.1371/journal.pone.0032131

Lipinski, C. A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. *Journal of pharmacological and toxicological methods*. 44(1): 235–249. DOI: 10.1016/s1056-8719(00)00107-6

Lippert, T., Rarey, M. (2009). Fast automated placement of polar hydrogen atoms in protein-ligand complexes. *Journal Cheminform*. 1: 13. DOI:10.1186/1758-2946-1-13

Lorenz-Fonfria, V. A. (2020). Infrared Difference Spectroscopy of Proteins: From Bands to Bonds. *Chemical Reviews*. 120(7): 3466-3576. DOI: 10.1021/acs.chemrev.9b00449.

Lv, X., Ren, S., Zhang, X., Shen, Z., Ghosh, T., Xianyu, A., Gao, P., Li, Z., Lin, S., Yu, Y., Zhang, Q., Groszer, M., Shi, S. (2019). TBR2 coordinates neurogenesis expansion and precise microcircuit organization via Protocadherin 19 in the mammalian cortex. *Nature Communications.* 10(1): 1-15. DOI: 10.1038/s41467-019-11854-x

MacCallum, C. A., Russo, E. B. (2018). Practical considerations in medical Cannabis administration and dosing. *European Journal of Internal Medicine*. 49: 12-19. DOI: 10.1016/j.ejim.2018.01.004

Maestro, Schrödinger, LLC, New York, NY, 2023.

Maestro-Desmond Interoperability Tools, version 4.2, Schrödinger, New York, NY, 2015.

Malfait, A.M., Gallily, R., Sumariwalla, P.F., Malik, A.S., Andreakos, E., Mechoulam, R., Feldmann, M. (2000). The nonpsychoactive cannabis constituent cannabidiol is an oral anti-arthritic therapeutic in murine collagen-induced arthritis. *Proceedings of the National Academy of Sciences*. 97(17): 9561–9566. DOI:10.1073/pnas.160105897.

Mandel, M., Higo, A. (1970). Calcium-dependent bacteriophage DNA infection. *Journal of molecular biology*. 53(1): 159-62. DOI:10.1016/0022-2836(70)90051-3

McDermott, J. H., Study, D. D. D., Clayton-Smith, J., Briggs, T. A. (2018). The TBR1-related autistic-spectrum-disorder phenotype and its clinical spectrum. *European journal of medical genetics*. *61*(5): 253–256. DOI: 10.1016/j.ejmg.2017.12.009

McKenna, W. L., Betancourt, J., Larkin, K. A., Abrams, B., Guo, C., Rubenstein, J. L. R., Chen, B. (2011). Tbr1 and Fezf2 regulate alternate corticofugal neuronal identities during neocortical development. *Journal of Neuroscience*. 31(2): 549-564. DOI:10.1523/JNEUROSCI.4131-10.2011.

Medical, Dental and Pharmacy Act 13 of 1928.

Medicines and Related Substances Control Act 101 of 1965.

Meng, X.Y., Zhang, H.X., Mezei, M., Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*. 7(2): 146-157. DOI: 10.2174/157340911795677602.

Merck (2021). Setting Column Pressure Limits for Size Exclusion Chromatography. Merck. Available at: https://www.sigmaaldrich.com/ZA/en/technical-documents/technical-article/analytical-chemistry/purification/setting-columnpressure-limits.

Merck. IR Spectrum Table & Chart. Available at: https://www.sigmaaldrich.com/ZA/en/technical-documents/technical-article/analytical-chemistry/photometry-and-reflectometry/ir-spectrum-table

Metz, T.D., Silver, R.M., McMillin, G.A., Allshouse, A.A., Jensen, T.L., Mansfield, C. (2019). Prenatal marijuana use by self-report and umbilical cord sampling in a state with marijuana legalization. *Obstetrics & Gynecology*. 133: 98–104. DOI: 10.1097/AOG.0000000000003028.

Minister of Justice and Constitutional Development and Others v Prince (Clarke and Others Intervening); National Director of Public Prosecutions and Others v Rubin; National Director of Public Prosecutions and Others v Acton (CCT108/17) [2018] ZACC 30; 2018 (10) BCLR 1220 (CC); 2018 (6) SA 393 (CC); 2019 (1) SACR 14 (CC).

Moltke, J., Hindocha, C. (2021). Reasons for cannabidiol use: a cross-sectional study of CBD users, focusing on self-perceived stress, anxiety, and sleep problems. *Journal of cannabis research.* 3(1): 5. DOI: 10.1186/s42238-021-00061-5

Morrison, J. L., Breitling, R., Higham, D. J. and Gilbert, D. R. (2006) A lock-and-key model for protein-protein interactions. *Bioinformatics*. 22: 2012–2019. DOI:.1093/bioinformatics/btl338

Muller, C., Herrmann, B. (1997). Crystallographic structure of the T domain – DNA complex of the Brachyury transcription factor. *Nature*. 389(6653): 884–888. DOI: 10.1038/39929.

Nambot, S., Faivre, L., Mirzaa, G., Thevenon, J., Bruel, A. L., Mosca-Boidron, A. L., Masurel-Paulet, A., Goldenberg, A., Le Meur, N., Charollais, A., Mignot, C., Petit, F., Rossi, M., Metreau, J., Layet, V., Amram, D., Boute-Bénéjean, O., Bhoj, E., Cousin, M. A., Kruisselbrink, T. M., Thauvin-Robinet, C. (2020). De novo TBR1 variants cause a neurocognitive phenotype with ID and autistic traits: report of 25 new individuals and review of the literature. *European journal of human genetics* : 28(6): 770–782. DOI: 10.1038/s41431-020-0571-6

National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Committee on the Health Effects of Marijuana: An Evidence Review and Research Agenda. (2017). Prenatal, perinatal, and neonatal exposure to cannabis. National Academies Press. Available at: https://www.ncbi.nlm.nih.gov/books/NBK425751/

National Academies of Sciences, Engineering, and Medicine. (2017). The Health Effects of Cannabis and Cannabinoids: The Current State of Evidence and Recommendations for Research. National Academies Press.

Nemoto, W., Sato, T., Nakagawasai, O., Yaoita, F., Silberring, J., Tadano, T., Tan-No, K. (2013). Phenylmethanesulfonyl fluoride, a serine protease inhibitor, suppresses naloxone-precipitated

withdrawal jumping in morphine-dependent mice. *Neuropeptides*. 47(3):187-191. DOI: 10.1016/j.npep.2012.11.002.

Nevskaya, N. A., Chirgadze, Y. N. (1976). Infrared spectra and resonance interactions of amide-I and II vibration of alpha-helix. *Biopolymers*. 15(4): 637–648. DOI: 10.1002/bip.1976.360150404

NIDA. (2023). Can marijuana use during and after pregnancy harm the baby?. Retrieved from https://nida.nih.gov/publications/research-reports/marijuana/can-marijuana-use-during-pregnancy-harm-baby on 2023, September 9.

Niesen, F. H., Berglund, H., Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*. 2(9): 2212-2221. DOI: 10.1038/nprot.2007.321.

Niesen, F., Berglund, H. Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc.* 2: 2212–2221 (2007). DOI: 10.1038/nprot.2007.321

Nutor, C., Dunlop, A., Sadler, O., Brennan, P. A. (2023). Prenatal Cannabis Use and Offspring Autism-Related Behaviors: Examining Maternal Stress as a Moderator in a Black American Cohort. *Journal of autism and developmental disorders*. 1–13. DOI: 10.1007/s10803-023-05982-z

O'Sullivan, S. (2012). Cannabinoid activation of peroxisome proliferator-activated receptors: An update and review of the physiological relevance. *Wiley Interdisciplinary Reviews: Membrane Transport and Signaling*. 2(1): 17-25. DOI: 10.1002/wmts.73

Ochiai, W., Kitaoka, S., Kawamura, T., Hatogai, J., Harada, S., Iizuka, M. (2021). Maternal and fetal pharmacokinetic analysis of cannabidiol during pregnancy in mice. *Drug Metab Dispos,* 49: 337–343. DOI: 10.1124/dmd.120.000270.

Oswald, N. (2016). 'How SDS-PAGE Works', Bitesizebio. Available at: https://bitesizebio.com/580/how-sds-page-works/.

Pacher, P., Batkai, S., Kunos, G. (2006). The Endocannabinoid System as an Emerging Target of Pharmacotherapy. *Trends in Pharmacological Sciences*. 58(3): 389-462. DOI: 10.1124/pr.58.3.2

Pagano, C., Pilon, C., Calcagno, A., Urbanet, R., Rossato, M., Milan, G., Federspil, G. (2007). The endogenous cannabinoid system stimulates glucose uptake in human fat cells via phosphatidylinositol 3-kinase and calcium-dependent mechanisms. *Journal of Clinical Endocrinology & Metabolism*. 92(12): 4810-4819. DOI: 10.1210/jc.2007-0768

Panja, S., Saha, S., Jana, B., Basu, T. (2006). Role of membrane potential on artificial transformation of E. coli with plasmid DNA. *Journal of Biotechnology*. 127(1): 14-20. DOI: 10.1016/j.jbiotec.2006.06.008.

Papaioannou, V. E. (2001). T-box genes in development: from hydra to humans*. International Review of Cytology*. 207: 1–70. DOI: 10.1016/S0074-7696(01)07003-4

Papaioannou, V. E., Silver, L. M. (1998). The T-box gene family. *BioEssays*. 20(1):9–19. DOI: 10.1002/(SICI)1521-1878(199801)20:1<9::AID-BIES4>3.0.CO;2-Q

Perucca, E., Bialer, M. (2020). Critical Aspects Affecting Cannabidiol Oral Bioavailability and Metabolic Elimination, and Related Clinical Implications. *CNS drugs*. 34(8): 795–800. DOI: 10.1007/s40263-020-00741-5.

pET System Manual Novagen 1. (1992).

Petersen-Williams, P., Jordaan, E., Mathews, C., Lombard, C., Parry, C. D. H. (2014). Alcohol and Other Drug Use during Pregnancy among Women Attending Midwife Obstetric Units in the Cape Metropole, South Africa. *Advances in Preventive Medicine.* 871427. DOI: 10.1155/2014/871427.

Pflugfelder, G. O., Roth, H., Poeck, B. (1992). A homology domain shared between Drosophila optomotor-blind and mouse Brachyury is involved in DNA binding. *Biochemical Biophysical Research Communications.* 186 (2): 918–925. DOI: 10.1016/0006-291x(92)90833-7.

Possenti, E., Colombo, C., Realini, M., Song, C. L., Kazarian, S. G. (2021). Time-Resolved ATR–FTIR Spectroscopy and Macro ATR–FTIR Spectroscopic Imaging of Inorganic Treatments for Stone Conservation. *Analytical Chemistry*. 93(44): 14635–14642. DOI: 10.1021/acs.analchem.1c02392.

QikProp, Schrödinger, LLC, New York, NY, 2023.

Rahn, E. J., Hohmann, A. G. (2009). Cannabinoids as pharmacotherapies for neuropathic pain: from the bench to the bedside. *Neurotherapeutics*. 6(4): 713-737. DOI: 10.1016/j.nurt.2009.08.002

Ramírez, D., Caballero, J. (2018). Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry*. *23*(5):1038. DOI: 10.3390/molecules23051038

Remedios, R., Huilgol, D., Saha, B., Hari, P., Bhatnagar, L., Kowalczyk, T., Hevner, R. F., Suda, Y., Aizawa, S., Ohshima, T., Stoykova, A., Tole, S. (2007). A stream of cells migrating from the caudal telencephalon reveals a link between the amygdala and neocortex. *Nature Neuroscience*. 10(9): 1141-1150. DOI:10.1038/nn1955.

Rosano, G., Ceccarelli, E. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*. 5(1). DOI: 10.3389/fmicb.2014.00172

Russo, E. and Guy, G.W. (2006). A tale of two cannabinoids: the therapeutic rationale for combining tetrahydrocannabinol and cannabidiol. *Medical hypotheses*. 66(2): 234–46. DOI:10.1016/j.mehy.2005.08.026.

Russo, E.B. (2017). Cannabidiol Claims and Misconceptions. *Trends in Pharmacological Sciences*. 38(3): 198–201. DOI:10.1016/j.tips.2016.12.004.

Ruvinsky, I., Silver, L.M: (1997). Newly identified paralogous groups on mouse chromosomes 5 and 11 reveal the age of a T-box cluster duplication. *Genomics*. 40: 262-266. DOI: 10.1006/geno.1996.4591.

Sajdey, R., Brown, J. D., Goodin, A. J. (2021). Perinatal Cannabis Exposures and Autism Spectrum Disorders. *Med Cannabis Cannabinoids*. 4(1): 67–71. DOI: 10.1159/000515871

Samuel, E.L.G., Holmes, S.L. Young, D.W. (2021). Processing binding data using an open-source workflow.*Journal of Cheminform*. **13**: 99. DOI: 10.1186/s13321-021-00577-1

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*. 27(3): 221–234. DOI: 10.1007/s10822-013-9644-8.

Schmid, X. (2001). Biological Macromolecules: UV-visible Spectrophotometry. DOI: 10.1038/npg.els.0003142

Schöning-Stierand, K., Diedrich, K., Ehrt, C., Flachsenberg, F., Graef, J., Sieg, J., Penner, P., Poppinga, M., Ungethüm, A., Rarey, M. (2022). ProteinsPlus: A comprehensive collection of web-based molecular modeling tools. *Nucleic Acids Research*. *50*: 611. DOI: 10.1093/nar/gkac305

Schrödinger Suite 2012 Protein Preparation Wizard; Epik version 2.3, Schrödinger, LLC, New York, NY, 2012; Impact version 5.8, Schrödinger, LLC, New York, NY, 2012; Prime version 3.1, Schrödinger, LLC, New York, NY, 2012.

Schurman, L. D., Lu, D., Kendall, D. A., Howlett, A. C., Lichtman, A. H. (2020). Molecular Mechanism and Cannabinoid Pharmacology. *Handbook of Experimental Pharmacology*. 258: 323–353. DOI: 10.1007/164_2019_298

Seccamani, P., Franco, C., Protti, S. (2021). Photochemistry of Cannabidiol (CBD) Revised. A Combined Preparative and Spectrometric Investigation. *Journal of Natural Products*. 84(11):2858-2865. DOI:10.1021/acs.jnatprod.1c00567

Shao, J., Xiang, J., Axner, O., Ying, C. (2016). Wavelength-modulated tunable diode-laser absorption spectrometry for real-time monitoring of microbial growth. *Applied Optics*. 55(9):2339. DOI: 10.1364/ao.55.002339.

Shapiro, A. (2017). What is the purpose of using stain and destain in SDS-SDS-PAGE gel?. ResearchGate. Available at: https://www.researchgate.net/post/What-is-the-purpose-of-using-stain-and-destain-in-SDS-SDS-PAGE-gel.

Sharma, Y.R. (2004). Ultraviolet and visible spectroscopy. *Elementary Organic spectroscopy*. 16: 9-59. DOI: 10.4236/opj.2018.85012

Shelley, J., Cholleti, A., Frye, L., Greenwood, J., Timlin, M.R., Uchimaya, M. (2007). Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *Journal of Computational Aided Mol Des*. 21(12):681-91. DOI: 10.1007/s10822-007-9133-z.

Sherman, J.D. (1999). Synthetic zeolites and other microporous oxide molecular sieves. National Academy of Sciences. 96(7):3471–3478. DOI: 10.1073/pnas.96.7.3471.

Sherman, W., Beard, H. S., Farid, R. (2006). Use of an Induced Fit Receptor Structure in Virtual Screening. *PubMed*. 67(1): 83-84. DOI: 10.1111/j.1747-0285.2005.00327.x.

Shorey-Kendrick, L.E., Roberts, V.H.J., D'Mello, R.J., Sullivan, E.L., Murphy, S.K., Mccarty, O.J.T., Schust, D.J., Hedges, J.C., Mitchell, A.J., Terrobias, J.J.D., Easley IV, C.A., Spindel, E.R., Lo, J.O. (2023). Prenatal delta-9-tetrahydrocannabinol exposure is associated with changes in rhesus macaque DNA methylation enriched for autism genes. *Clinical Epigenetics.* 15(1): 104. DOI: 10.1186/s13148-023-01519-4

Shrivastava, A., Mathur, K., Verma, R. K., Magani, S. K. J., Vyas, D. K., Singh, A. (2022). Molecular dynamics study of tropical calcific pancreatitis (TCP) associated calcium-sensing receptor single nucleotide variation. Frontiers in Molecular Biosciences. 9: 982831. DOI: 10.3389/fmolb.2022.982831

SITEMAP SiteMap, version 3.5, Schrödinger, LLC, New York, NY, 2015.

SiteMap, version 3.5, Schrödinger, LLC, New York, NY, 2015.

Skolnick J., Gao M., Zhou H., Singh S. (2021). AlphaFold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. *Journal Chem. Inf. Model.* 61:4827–4831.

Small, E., Cronquist, A. (1976). A Practical and Natural Taxonomy for Cannabis. *Taxon*. 25(4): 405–435. DOI:10.2307/1220524.

Smart, R., Pacula, R. L. (2019). An economic analysis of U.S. drug policy. RAND Corporation.

Smith, J. (1999). T-box genes: what they do and how they do it. *Trends Genetics*. 15(4): 154–158. DOI: 10.1016/s0168-9525(99)01693-5.

Smith-Kielland, A., Skuterud, B., Mørland, J. (1999). Urinary excretion of 11-nor-9-carboxy-delta9-tetrahydrocannabinol and cannabinoids in frequent and infrequent drug users. *J Anal Toxicology*. 23(5): 323-332. DOI: 10.1093/jat/23.5.323.

SnapGene software (from Insightful Science; available at snapgene.com)

Sonawane, A.R., Platig, J., Fagny, M., Chen, C.Y., Paulson, J.N., Lopes-Ramos, C.M., DeMeo, D.L., Quackenbush, J. (2017). Understanding Tissue-Specific Gene Regulation. *Cell Reports*. 21(4):1077–1088. DOI: 10.1016/j.celrep.2017.10.001.

Sotriffer, C., Klebe, G. (2002). Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design. Farmacology. 57(3): 243–251. DOI: 10.1016/s0014-827x(02)01211-9

Stirnimann, C. U., Ptchelkine, D., Grimm, C., Müller, C. W. (2010). Structural basis of TBX5-DNA recognition: the T-box domain in its DNA-bound and -unbound form. *Journal of Molecular Biology.* 400(1): 71-81. DOI: 10.1016/j.jmb.2010.04.052

Stolarczyk, E., Lord, G. M., Howard, J. K. (2013). The immune cell transcription factor T-bet: A novel metabolic regulator. *Adipocyte*. 3(1): 58-62. DOI: 10.4161/adip.26220

Stoscheck, C. M. (1990). Quantitation of protein. *Methods in enzymology*. 182: 50–68. DOI: 10.1016/0076-6879(90)82008-p

Stringer, R. (2004). Electrophoresis – Overview. Encyclopaedia of Analytical Science: Second Edition. *Elsevier* : 356–363. DOI: 10.1016/B0-12-369397-7/00120-5.

Tahir, M.N., Shahbzi, F., Rondeau-Gagne, S., Trant, J. (2021). The Biosynthesis of the Cannabinoids. *Journal of Cannabis Research*. 3:7. DOI: 10.1186/s42238-021-00062-4

Tang, J., Chen, Q., Guo, J., Yang, L., Tao, Y., Li, L., Miao, H., Feng, H., Chen, Z., Zhu, G. (2016). Minocycline Attenuates Neonatal Germinal-Matrix-Hemorrhage-Induced Neuroinflammation and Brain Edema by Activating Cannabinoid Receptor 2*. Molecular neurobiology*. 53(3): 1935–1948. DOI: 10.1007/s12035-015-9154-x

Terwilliger, T.C., Liebschner, D., Croll, T.I. (2024). AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Natural Methods*. 21; 110–116. DOI: 10.1038/s41592-023-02087-4

Tian *et al., (*2018*).* CASTP3.0. Nucleic Acids Research.

Tian, J., Liu, J., He, W., Hu, Z., Yao, X., Chen, X. (2004). Probing the Binding of Scutellarin to Human Serum Albumin by Circular Dichroism, Fluorescence Spectroscopy, FTIR, and Molecular Modeling Method. *Biomacromolecules*. 5(5): 1956-1961. DOI: 10.1021/bm049668m.

Touwn M. (1981) The religious and medicinal uses of Cannabis in China, India, and Tibet. *Journal of Psychoactive Drugs*. 13(1): 23-34. DOI: 10.1080/02791072.1981.10471447

Traylor, R. N., Dobyns, W. B., Rosenfeld, J. A., Wheeler, P., Spence, J. E., Bandholz, A. M., Bawle, E. V., Carmany, E. P., Powell, C. M., Hudson, B. (2012). Investigation of TBR1 hemizygosity: four individuals with 2q24 microdeletions. *Mol Syndromology*. 3(3): 102–112. DOI: 10.1159/000342008

U.S. Food and Drug Administration. (2019). What You Should Know About Using Cannabis, Including CBD, When Pregnant or Breastfeeding. Accessed from https://www.fda.gov/consumers/consumer-updates/what-you-should-know-about-using-cannabis-including-cbd-when-pregnant-or-breastfeeding#:~:text=FDA%20strongly%20advises%20that%20during,or%20marijuana%20in%20any%20form.

Ullah, H. (2012) The Role of Ion Exchange Chromatography in Purification and Characterization of Molecules. Ion Exchange Technologies. DOI: 10.5772/52537.

UniProt Consortium. (2023). UniProt: The universal protein knowledgebase in 2023. *Nucleic Acids Research*. 51:D523-D531

Vanstone, M., Panday, J., Popoola, A., Taneja, S., Greyson, D., McDonald, S. D., Pack, R., Black, M., Murray-Davis, B., Darling, E. (2022). Pregnant People's Perspectives On Cannabis Use During Pregnancy: A Systematic Review and Integrative Mixed-Methods Research Synthesis. *Journal of midwifery & women's health*. 67(3): 354–372. DOI: 10.1111/jmwh.13363

Vázquez, P., Hermosilla, P., Guallar, V., Estrada, J., Vinacua, A. (2018). Visual analysis of protein-ligand interactions. *Computer graphics forum*. 37(3): 391-402. DOI: 10.1111/cgf.13428.

Venyaminov SYu, Kalnin, N. N. (1990). Quantitative IR spectrophotometry of peptide compounds in water (H2O) solutions. II. Amide absorption bands of polypeptides and fibrous proteins in alpha-, beta-, and random coil conformations. *Biopolymers*. 30(13-14): 1259–1271. DOI: 10.1002/bip.360301310

Venyaminov, A., Kalnin, N. N. (1990). Quantitative IR spectrophotometry of peptide compounds in water (H2O) solutions. I. Spectral parameters of amino acid residue absorption bands. *Biopolymers*. 30(13-14): 1243–1257. DOI: 10.1002/bip.360301309

Volkamer, A., Griewel, A., Grombacher, T., Rarey, M. (2010). Analyzing the topology of active sites: on the prediction of pockets and subpockets*. Journal of chemical information and modelling*. 50(11): 2041–2052. DOI: 10.1021/ci100241y

Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F., Rarey, M. (2012). Combining global and local measures for structure-based druggability predictions. *Journal of chemical information and modelling*. 52(2): 360–372. DOI: 10.1021/ci200454v

Volkov, V. V., Svergun, D. I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography*. 36(1): 860-864. DOI: 10.1107/S0021889803000268

Vučković, S., Srebro, D., Vujović, K.S., Vučetić, Č. Prostran, M. (2018). Cannabinoids and Pain: New Insights From Old Molecules. *Frontiers in Pharmacology*. 9(1259). DOI:10.3389/fphar.2018.01259.

Wang, T., Tan, Y., Chen, Y. Z., Tan, C. (2023). Infrared Spectral Analysis for Prediction of Functional Groups Based on Feature-Aggregated Deep Learning. *Journal of Chemical Information and Modeling.* 63(15): 4615–4622. DOI: 10.1021/acs.jcim.3c00749

Weisbeck, S. J., Bright, K. S., Ginn, C. S., Smith, J. M., Hayden, K. A., Ringham, C. (2021). Perceptions about cannabis use during pregnancy: a rapid best-framework qualitative synthesis. Canadian Journal of Public Health. 112(1): 49-59. DOI:10.17269/s41997-020-00346-x.

Western Cape Department of Health. (2019). How substance abuse affects your unborn baby.

Wieckiewicz, G., Stokłosa, I., Stokłosa, M., Gorczyca, P., Pudlo, R. (2022). Cannabidiol (CBD) in the Self-Treatment of Depression-Exploratory Study and a New Phenomenon of Concern for Psychiatrists. *Frontiers in Psychiatry*. *13*: 837946. DOI: 10.3389/fpsyt.2022.837946

Willsey, A. J., Sanders, S. J., Li, M., Dong, S., Tebbenkamp, A. T., Muhle, R. A., State, M. W. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. 155(5): 997-1007. DOI:10.1016/j.cell.2013.10.020.

Wilson., V., Conlon., F. L. (2002). The T-box family. *Genome Biology*. 3(6): 300-308. DOI: 10.1186/gb-2002-3-6-reviews3008.

Wingender, E. (2013). Criteria for an updated classification of human transcription factor dna-binding domains. *Journal Bioinformatics Computational Biology*. 11(1): 1340007. DOI: 10.1142/S0219720013400076.

Wood, A. G., Nadebaum, C., Anderson, V., Reutens, D., Barton, S., O'Brien, T. J., Vajda, F. (2015). Prospective assessment of autism traits in children exposed to antiepileptic drugs during pregnancy. *Epilepsia*. 56(7): 1047–1055. DOI: 10.1111/epi.13007

Woody, R.W. (1995). Circular Dichroism. *Methods in Enzymology*. 246: 34–71. DOI: 10.1016/0076-6879(95)46006-3.

World Health Organization. (2023). WHO Expert Committee on Drug Dependence. Cannabidiol (CBD). Pre-Review Report. WHO.

Yahaya, A.A., Sanusi, S., Malo, F.U. (2021). Computer-Assisted Multi-Epitopes T-Cell Subunit Covid-19 Vaccine Design. *Biomedical and Biotechnology Research Journal*. 5(1): 27-34. DOI:10.4103/bbrj.bbrj_5_21.

Yang, V.W. (1998). Eukaryotic transcription factors: Identification, characterization and functions. *Journal of Nutrition*. 128(11):2045–2051. DOI: 10.1093/jn/128.11.2045.

Yang, Y., Yao, K., Repasky, M. P., Leswing, K., Abel, R., Shoichet, B. K., Jerome, S. V. (2021). Efficient Exploration of Chemical Space with Docking and Deep Learning. *Journal of Chemical Theory and Computation.* 17(11): 7106–7119. DOI: 10.1021/acs.jctc.1c00810.

Ye, S., Li, H., Yang, W., Luo, Y. (2014). Accurate determination of interfacial protein secondary structure by combining interfacial-sensitive amide I and amide III spectral signals. *Journal of the American Chemical Society*. *136*(4): 1206–1209. DOI: 10.1021/ja411081t

Young-Wolff, K. C., Foti, T. R., Green, A., Altschuler, A., Does, M. B., Jackson-Morris, M. (2022). Perceptions about cannabis following legalization among pregnant individuals with prenatal cannabis use in California. *JAMA Network Open*. 5(12): 2246912. DOI: 10.1001/jamanetworkopen.2022.46912

Young-Wolff, K. C., Sarovar, V., Tucker, L.-Y., Conway, A., Alexeeff, S., Weisner, C., Armstrong, M. A., Goler, N. (2019). Self-reported Daily, Weekly, and Monthly Cannabis Use Among Women Before and During Pregnancy. JAMA network open. 2(7): 196471. DOI: 10.1001/jamanetworkopen.2019.6471

Young-Wolff, K.C., Ray, G.T., Alexeeff, S.E. (2021). Rates of Prenatal Cannabis Use Among Pregnant Women Before and During the COVID-19 Pandemic. *Journal of the American Medical Association.* 326(17): 1745–1747. DOI: 10.1001/jama.2021.16328.

Appendix A: Turn-it-in Report

Chanel_Final Dissertation_R2

by yosra abdalla

Submission date: 27-Jan-2024 06:18PM (UTC+0200)
Submission ID: 2277374856
File name: Turn_It_In_27-01-24.docx (17.19M)
Word count: 48643
Character count: 284895



ASSESSING THE POTENTIAL INTERACTION BETWEEN CBD AND TBR1 T-BOX DOMAIN

Chanel Blignaut (1831350)



A Dissertation submitted to the Faculty of Health Science, University of Witwatersrand, in fulfilment of the requirements of the degree of Masters of Medicine

Supervisor:

Dr. Sylvia Fanucchi

Dr. Samson Adeyemi

Chanel_Final Dissertation_R2

ORIGINALITY REPORT

16%	11%	12%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

3%

★ hdl.handle.net

Internet Source

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		