IN SILICO ANALYSIS OF INTERACTION OF HUMAN POSITIVE TRANSCRIPTION ELONGATION FACTOR B (P-TEFB) WITH VIRAL PROTEINS

BY

NUR NADIAH BINTI CHE AZIZ

A thesis submitted in fulfilment of the requirement for the degree of Master of Science (Biotechnology)

Kulliyyah of Science
International Islamic University Malaysia

NOVEMBER 2021

ABSTRACT

Due to the small genome size, viruses do not have the ability to reproduce on their own, but rely on host cellular proteins to facilitate its replication and complete the life cycle. As a result, many host proteins complexes are manipulated during viral infections, and these proteins that interact with viral proteins are deemed as potential drug target. One of the host protein complexes is the positive transcription elongation factor b (P-TEFb), which comprises cyclin-dependent kinase 9 (CDK9) and cyclin T1 that has been of great interest due to its interaction with human immunodeficiency virus-1 (HIV-1) Tat protein. Due to the critical role of host P-TEFb in viral replication, it is suspected that P-TEFb may act as central role for interaction with a number of viral proteins other than HIV-1 Tat. This dissertation aims to uncover the interaction mode of P-TEFb and viral protein complexes, specifically the HIV and human Herpes Simplex Virus (HSV), in hopes of identifying responsible binding site using computational approaches. To achieve the first aim, an integrated protein-protein interaction networks of selected HIV and HSV viral proteins and P-TEFb was constructed to understand how the human viruses control the host's biological function. The second aim was to analyze the potential binding interfaces on viral proteins with P-TEFb, particularly between wild and mutant types of viral proteins by using molecular docking. The protein-protein docking revealed that HIV-1 Vpr protein has higher binding affinity for P-TEFb compared to HIV-1 Tat protein thus formed the best interaction. It was predicted that both HIV-1 Tat and Vpr bind to similar region of cyclin T1, suggesting that both viral proteins are unable to bind the host protein at the same time. Finally, the investigation of the key amino acids of viral proteins and their mutants by coarse-grained molecular dynamic (CGMD) revealed that mutation of both acidic/cysteine region of HIV-1 Tat as well as helix 3 of HIV-1 Vpr leads to the different orientation of protein-protein complexes when compared with the wild type, thus these regions may be responsible for the interaction with cyclin T1 domain. It is worth to note that both HIV-1 Tat and Vpr viral proteins are cooperatively in control of viral gene transcription and replication, as shown in the interaction network and in the *in silico* analyses. An additional analysis of HSV-1 viral proteins VP16 and ICP22 showed that both binding sites of the viral proteins have been identified as the transactivation domain and were predicted to bind at the same region of CDK9. It can be concluded that cyclin T1 (residue 162-259) and CDK9 subunit of P-TEFb (residue 46-260) could be potential target site for the drug design specifically to treat the human viral infections.

خلاصة البحث

صغر حجم المجين في الفايروسات ادى الى عدم قدرة الفايروسات للاعتماد على نفسها بالتكاثر, حيث تعتمد الفايروسات على وجود البروتينات الخلوية في المضيف لتسهيل عملية التكاثر و اكمال دورة الحياة للفيروس. يتم التلاعب بالعديد من مجمعات البروتين المضيف أثناء العدوى الفيروسية , هذه المعقدات البروتينية تستخدم كعلاج في بعض الاحيان. احد اهم معقدات البروتين هو عامل استطالة النسخ الموجب ب (P-TEFb) الذي يشمل كيناز 9 المعتمد على السيكلين(CDK9) و يشمل سايكلين ت1 الذي يعتبر ذات اهمية كبيرة لقدرته على التفاعل مع بروتين (HIV-1-Tat protein) لفايروس العوز المناعي البشري-1 (HIV-1). بسبب اهمية P-TEFb الذي يؤثر على تكاثر الفايروسات في المستضيف, لذلك يعتقد ان لدى P-TEFb دور مركزي في التفاعل مع العديد من بروتينات الفايروسات غير الHIV-1-Tat. هذه الاطروحة تهدف الى معرفة وضع التفاعل بين P-TEFb مع معقدات البروتين الفيروسي, و نخص بهذا فايروس العوز المناعي البشري (HIV) و فيروس الحلأ البسيط البشري (HSV), على امل ايجاد المنطقة المسؤولة على الارتباط باستخدام الاساليب الحسابية. من اجل تحقيق الهدف الاول, يتم دراسة التفاعلات بين البروتين للفايروسات المختارة HIVو HSV المدمجة مع بروتين P-TEFb للزيادة في توضيح كيفية الفايروسات البشرة تسيطر على الوظيفة الاحيائية في المضيف. اما الهدف الثاني هو من اجل التحقيق في اوجه الارتباط المحتملة بين بروتينات الفايروس مع P-TEFb, و بشكل خاص البروتينات المطورة و الغير مطورة للفايروسات باستخدام الالتحام الجزيئي. الالتحام بين البروتينات يكشف ان بروتين HIV-1 VPR لديه القدرة الارتباط عالية مع P-TEFb مقارنة ببروتين HIV-1 Tat. كان من المتوقع ان كلا بروتين HIV-1 Tatو بروتين VPR يرتبطان مع منطقة متشِابه لسايكلين ت1 و لذالك من غير الممكن ارتباطهما مع سايكلين ت1 في نفس الوقت. و اخيرا, عن طريق الديناميكية الجزيئية للحبيبات الخشنة (CGMD) اظهرت التحقيقات في الالحماض الامينية للبروتينات الفيروسية و البروتينات المهجنة, ان الطفرة الحامضة او/و سيستينية في منطقة HIV-1 Tat و كذلك Helix 3 لVpr HIV-1 تؤدي الترتيب مختلف في تكوين معقدات البروتين , لذلك هذه المنطقة ممكن تكون مسؤولة عن التفاعل مع نطاق سايكلين T1. من المهم ملاحظة ان بروتين Tat و بروتين Vpr كلاهما متعاونان في السيطرة على نقل الشفرات الجينية و التكاثر في الفايروس, كما اظهرته تحاليل نتائج السيليكو و الفاعلات الشبكية. تحاليل اضافية لبروتينات فايروس (HSV-1) التي هي VP16 و ICP22 , و قد اظهر نتائج التحاليل ان كليهما منطقة مواقع ارتباط لبروتين الفايروس, كما وقد عرفا هذين البروينين على انهما مناطق معاملات و يمكن ارتباطهما في نفس منطقة (CDK9). كخلاصة, سايكلين T1 (residue 162-259) و CDK9 الوحدة الفرعية من P-TEFb من المحتمل ان يستخدما كهدف للعلاج ضد الاصابة الفايروسية البشرية.

APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science (Biotechnology).

……
Nur Firdaus binti Isa
Supervisor

……
'Aisyah binti Mohamed Rehan
Co-Supervisor

……
Khairul Bariyyah binti Abd Halim
Co-Supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science (Biotechnology).

……
Tengku Haziyamin Tengku Abdul Hamid
Internal Examiner

……
Ramlah binti Kadir
External Examiner

This thesis was submitted to the Department of Biotechnology and is accepted as a fulfilment of the requirement for the degree of Master of Science (Biotechnology).

……
Tengku Haziyamin Tengku Abdul Hamid
Head, Department of Biotechnology

This thesis was submitted to the Kulliyyah of Science and is accepted as a fulfilment of the requirement for the degree of Master of Science (Biotechnology).

……
Jesni Shamsul Shaari
Dean, Kulliyyah of Science

DECLARATION

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Nur Nadiah binti Che Aziz

Signature …… Date 12.10.2021

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

IN SILICO ANALYSIS OF INTERACTION OF HUMAN POSITIVE TRANSCRIPTION ELONGATION FACTOR B (P-TEFB) WITH VIRAL PROTEINS

I declare that the copyright holders of this thesis are jointly owned by the student and IIUM.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may be used by others in their writing with due acknowledgement.

2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purposes.

3. The IIUM library will have the right to make, store in a retrieved system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Nur Nadiah binti Che Aziz

…….. 12.10.2021

Signature Date

All of my hard works is dedicated to my beloved parents, family and friends for all their kindness, love and continuous support.

ACKNOWLEDGEMENTS

In the name of Allah the Most Gracious, the Most Merciful. All praise to Allah for the completion of this study and His blessing. I thank to the Almighty for all the possibilities, trials and strength that have been shown to me to finish writing the thesis. During the process, I experienced so much, not only form the academic angle, but also from my personality side. My sincere gratitude to the Holy Prophet Muhammad whose way of life had guided me continually.

First and foremost, I am most indebted to my supervisor, Dr Nur Firdaus Isa, for her guidance, empathy, patience and most importantly, for providing constructive motivation and a warm spirit to complete this thesis. It has been a great pleasure and honor to have her as my supervisor. I must give my high, respectful gratitude to my co-supervisor, Dr Khairul Bariyyah Abd Halim and Dr ‘Aisyah Mohamed Rehan for giving me opportunity to explore new knowledge of mine and giving some ideas as well as for giving me precious advices in order to improve myself to become a better person. My appreciation also extends to Prof Alexandre Bonvin, one of the person that guide me in Computational Structural Biology especially in Molecular Docking expertise. To Head of RUBIC, Dr Azzmer, I am extremely grateful for your assistance and suggestions throughout my research.

I would like to express my external appreciation towards my parents, Che Aziz Che Awang and Nor Hasimah Ibrahim and family who are always been there for me no matter where I am, for all unconditional supports and patience. Thank you for being ever so understanding and supportive. Lastly, my appreciation also extends to my laboratory colleagues; Bro Ariff, his early insights launched the greater part of Coarse-Grained Simulation. Thanks also to Bro Hakim for his guidance and helped sustain a positive atmosphere in RUBIC (Research Unit for Bioinformatics & Computational Biology), IIUM Kuantan.

- If you really want what you’re praying for, you’ll never tired of waiting. The Almighty knows. The beauty lies in struggle. Trust Him -

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Å	angstrom
μs	microsecond
t	time
%	percentage
kcal	kilocalorie
kJ	kilojoule
mol	mole
α	alpha
β	beta
γ	gamma

LIST OF ABBREVIATIONS

STIs	Sexually Transmitted Infections
ABCG2	ATP-binding cassette super-family G member 2
ACV	Acyclovir
AIDS	Acquired Immunodeficiency Syndrome
Ala	Alanine
APOBEC3G	Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3G
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
BCL2	B-cell lymphoma 2
BECN1	Beclin-1
BRD4	Bromodomain-containing protein 4
CADD	Computer-Aided Drug Design
CCNT1	Cyclin T1
CCR5	C-C chemokine receptor type 5
CDK12	Cyclin Dependent Kinase 12
CDK13	Cyclin Dependent Kinase 13
CDK9	Cyclin dependent Kinase 9
CG	Coarse-Grained
CHEK1	Checkpoint Kinase 1
CREB1	CAMP Responsive Element Binding Protein 1
CSNK2A1	Casein Kinase 1 Alpha 1
CTD	C-terminal domain
CTR9	CTR9 Homolog, Paf1/RNA polymerase II complex
CUL4A	Cullin 4A
CXCR4	C-X-C chemokine receptor type 4
Cys	Cysteine
DDB1	Damage Specific DNA Binding Protein 1
EP300	E1A Binding Protein P300
FAMCV	Famciclovir
Gln	Glutamine

Glu	Glutamic acid
Gly	Glycine
GO	Gene Ontology
GTF2B	General Transcription Factor IIB
GUD	Genital Ulcer Disease
HAART	Highly Active Antiretroviral Therapy
HCFC1	Host Cell Factor C1
HEX1M1	Hexamthylene Bis-Acetamide Inducible 1
His	Histidine
HIV	Human Immunodeficiency Virus
HPV	Human Papillomavirus
HSV	Herpes Simplex Virus
ICP	Infected Cell Protein
IL-6	Interleukin 6
Ile	Isoleucine
IRF3	IFN regulatory factor 3
JUN	Jun proto-oncogene
KAT2B	Lysine Acetyltransferase 2B
Leu	Leucine
LTR	Long Terminal Repeat
Lys	Lysine
Met	Methionine
MLL5	Mixed lineage leukemia 5
mRNA	Messenger Ribonucleic Acid
NNRTIs	Non-nucleoside Reverse Transcriptase Inhibitors
NRTIs	Nucleoside Reverse Transcriptase Inhibitors
Phe	Phenylalanine
PPI	Protein-Protein Interaction
PPP1CC	Protein Phosphatase 1 Catalytic Subunit Gamma
PPP2R5D	Protein Phosphatase 2 Regulatory Subunit B'Delta
Pro	Proline
PSMB7	Proteosome subunit beta type-7
PSMD7	26S proteosome non-ATPase regulatory subunit 7
P-TEFb	Positive Transcriptional Elongation Factor b
RBBP4	Histone-binding protein RBBP4

RBBP7	Histone-binding protein RBBP7
RBM8A	RNA-binding protein 8A
Ser	Serine
SMARCB1	SWI/SNF, Matrix Associated, Actin Dependent Regulator of Chromatin, Subfamily B, Member 1
SOCS3	Suppressor of cytokinase signaling 3
STAT2	Signal transducer and activator of transcription 2
SUB1	SUB1 Regulator of Transcription
SUPT6H	Human Suppressor of Ty Homologue-6
TAT	Trans-Activator of Transcription
TBK1	TANK binding kinase 1
TCEB1	Transcription Elongation Factor B Polypeptide 1
Thr	Threonine
TP53	Tumor Protein 53
TRL	Inverted repeats (IRL)
Trp	Tryptophan
Tyr	Tyrosine
VACV	Valaciclovir
Val	Valine
VP16	Virion Protein 16
VPR	Viral protein R
VPRBP	Vpr (HIV-1) binding protein
WDR73	WD repeat-containing repeat domain 73
WHO	World Health Organization

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND OF THE STUDY

Infectious disease is one of the leading causes of death particularly in the developing world. The sexual transmitted infections (STIs) are among the most prevalent and highly contagious diseases in the world, affecting the health and lives of women, men, and babies. Stigma, stereotyping, and shame are often faced by people with STIs and they are susceptible to gender-based abuse (Morris et al., 2014). It is known that more than 30 different bacteria, viruses and parasites are transmitted through sexual contact. The highest prevalence of sexually transmitted diseases is associated with eight known pathogens. Currently, four of these eight pathogens are curable; the syphilis, gonorrhea, chlamydia and trichomoniasis. The other four are incurable viral infections: hepatitis B, herpes simplex virus (HSV), human papillomavirus (HPV) and human immunodeficiency virus (HIV). While bacteria STIs can be treated and cured with widely available medications, symptoms or diseases due to the incurable viral infections can be only reduced and cured through antibiotics (WHO, 2020). In Malaysia, the full operationalization and effective implementation of national and local development plans were started to achieve the goal of reducing the number of new HIV cases. HIV infections have decreased from peak in 2002 (28 per 100,000) to a plateau since 2009 (11 per 100,000) till 2017 (10.3 per 100,000). However, to achieve the vision of ending AIDS by 2030, the successful implementation is dependent on discovering undiagnosed individuals, linking and keeping them in care (Anita & Chai, 2019).

The most popular drug prescribed for HIV infections is the reverse transcriptase inhibitor, in which the drug acts by preventing initial replication in conjunction with other drugs such as protease inhibitors that inhibit the invasion of new viruses in already infected cells (Rai et al., 2018). However, in certain cases, HIV becomes drug resistant, and the exact combination of drugs prescribed for a HIV patient needs to be changed frequently. The problem lies in the reverse transcriptase of HIV, an enzyme which is particularly prone to errors (Sebastián-Martín et al., 2018). Many different viral strains of HIV are continuously produced within the same infected person as a result of errors in the reverse transcribed viral DNA. This high genetic variability of HIV also presents a challenge for the development of effective HIV vaccines (Servín-Blanco et al., 2016). Finally, the strains that are prone to mutagenesis and immune or resistant to a specific drug replicates uncontrollably and manifests their effect in the infected host. This condition also applies to viruses other than HIV, for example the HSV. The spontaneous mutations with HSV genome are introduced by errors during DNA replication (Cohen et al., 2020). The fact that viruses have become part of the human body, without directly affecting the patient as a result of latent infection, make them hard to cure or remove. The reluctance of some people to seek diagnosis and treatment, asymptomatic infections in which the individual is unaware that they are infected, latency and persistence of infection, diagnostic tests not being performed and the overuse of antibiotics are the main reasons that explain why clinical trial design for vaccines against STIs is problematic (Mcintosh, 2020).

Host proteins are critical for viral replication and very important targets for the production of antiviral drugs and drug repurposing (Ullah et al., 2019). The human positive transcription elongation factor b (P-TEFb), which interacts with viral proteins, is one of the host protein known to play critical roles in viral life cycle (Fujinaga, 2020).

The viral proteins hijack cellular processes through physical interaction with the P-TEFb. A structural theory was proposed by Franzosa and the group pertaining human-virus protein-protein interaction whereby the viral proteins have capabilities to interact with human proteins by imitating and vying the interaction interfaces of their binding host counterpart (Franzosa & Xia, 2011). It is intriguing to know whether the selected viral proteins follow the proposed theory when it comes to the interaction with host P-TEFb.

The human protein P-TEFb is a heterodimer complex that performs an important role in the regulation of RNA polymerase II transcription in eukaryotes (Ott et al., 2011). It is a cyclin dependent kinase that comprises CDK9, the catalytic subunit and cyclin T1, the regulatory subunit. P-TEFb is associated with other host proteins, including BRD4 bromodomain-containing protein 4, and also associated with a large protein complex called super elongation complex. Many viral factors affect transcription by recruiting or modulating the activity of P-TEFb. The viruses have evolved the strategies to hijack the host protein through their own regulatory proteins and P-TEFb functions act as a central player for the virus-host interaction especially in viral replication cycle. Notably, P-TEFb interacts with and targeted by HIV Tat protein to promote the viral gene expression, which bypasses cellular P-TEFb regulation (Tahirov et al., 2010). Knowing that the P-TEFb and HIV Tat interaction is well-studied in the literature therefore serve as a solid reference for the study, the thesis here presents analyses that focus on the protein-protein interactions (PPIs) between the viral proteins of HIV-1 (Tat and Vpr proteins) and HSV-1 (ICP22 and VP16 proteins) and their targeted host proteins P-TEFb to predict and uncover how these viral proteins coordinate the interactions that may take place in the course of infection. Many years ago, a finding has indicated that HIV Vpr synergistically promotes the transcriptional

activity of the HIV-1 long terminal region (LTR) through structural and functional interaction with Tat, which is an active viral regulatory protein (Sawaya et al., 2000). However, the precise domains of interaction in Vpr protein have yet to be clarified. To verify the Tat-Vpr and their cooperative interaction in HIV infection, more studies are still required because to the best of our knowledge, there is only a single research study that was done to investigate this interaction (G. Li & De Clercq, 2016).

In the case of HSV, studies have shown that for successful lytic HSV-1 infection, transcriptional activation of the immediate early (IE) genes is necessary. Meanwhile the failure of IE gene activation is associated with latent infection. The viral IE genes namely ICP0, ICP4, ICP22, ICP24, and ICP27 are important for the transcriptional activation in viral early and late genes. Interestingly, it has recently been shown that ICP22 interacts physically with the CDK9 subunit of P-TEFb, and this interaction induces a loss of serine-2 phosphorylation on RNA polymerase II carboxyl-terminal consensus sequence YSPTSPS (Guo et al., 2012). HIV-1 Tat interacts with P-TEFb to promote HIV viral transcription, however the HSV ICP22 interacts with P-TEFb to repress the HSV gene transcription (Guo et al., 2012). Further studies have shown that VP16, which is a HSV tegument protein has relieved repression of immediate early gene transcription by ICP22 and this finding suggests a potential biological interaction between ICP22 and VP16 which able to perform as a pair of important regulators during HSV life cycle (Cun Wei et al., 2013). However, how these two HSV viral proteins coordinate to hijack a single host protein P-TEFb is not known.

Computational analysis based on coarse-grained molecular dynamic tool has been used to stimulate inhibitors entering the HIV-1 protease binding cavity (Li et al., (2009). The coarse-grained molecular dynamic can also provide critical molecular dynamics analysis between protein-protein interactions (Z. R. Xie et al., 2017). At

present, most of the viral proteins studies has focused on transcription activation for example, how HIV-1 hijacks the host protein with the help of the Tat protein, to support its own transcription which is necessary for robust and efficient viral expression (Asamitsu et al., 2018). Mutation on critical residues of viral proteins may differentially modified the function of viral proteins, thus, affecting the establishment of viral-host protein interaction. A comprehensive analysis of how viral proteins form a complex with host protein may aid in a better understanding of how viruses advance their life cycle and the contact interface of the virus-host complex may provide a clue for development of targeted drug to treat sexually transmitted infections.

1.2 PROBLEM STATEMENT

Viral proteins form a complex with host protein in which these host proteins are part of a larger interaction network and perform many critical processes in the cells. Therefore the virus requires the interaction with the host proteins to manipulate the biological processes readily available in the infected host cells. Genetic evolution suggests that viral and host proteins are continuously changing their interface residues, either to prevent or modify their binding capabilities (Jubb et al., 2017). Research on virus host protein-protein interactions is useful to offer some in-depth understanding of viral infection. Present antiviral drugs primarily include antiviral drugs targeting either or both the virus and the host proteins. Viral targeting drugs are intended to inhibit the biological activity of viral proteins such as viral polymerase and protease, whereas host targeting drugs are capable to interfere with the functional of host proteins that are involved in the life cycle of viral (Magden & Ahola, 2005). Due to continuous change and mutation of most viruses, the sensitivity to virus-targeting drugs frequently reduces their effectiveness in the clinic, particularly for infections caused by RNA viruses.

However, host-targeting drugs may minimize these effects due to the extremely slow evolution rate of host proteins (Sanjua, 2016). There is an ample knowledge available on sequences diversity and structures, which will help us understand better how viruses establish their interaction and manipulated the cellular mechanisms in their infected hosts. As a kickstart, computational approaches such as molecular dynamic simulation (MD) are required to help uncover specific characteristics of protein structures that would be difficult to obtain through experimental techniques as well as to understand on how they are affected by mutations through simulation action (Kmiecik et al., 2016).

1.3 RESEARCH OBJECTIVES

This presented work primarily applies *in silico* computational methods to understand how viral proteins synergistically cooperate in targeting the host protein P-TEFb. In order to optimize the effectiveness of this analysis, we will use all available information of viral-host protein interaction found in literature, along with the available computational resources. Three specific objectives were highlighted for this study which are:

1. To understand the coverage of large protein interaction by building protein-protein interaction network of HIV-1 Tat and Vpr, HSV-1 ICP22 and VP16 viral proteins with host protein P-TEFb.
2. To characterize the potential binding interfaces for the interaction between viral proteins HIV-1 Tat and Vpr, HSV-1 ICP22 and VP16 with P-TEFb.
3. To investigate the key amino acids between viral proteins HIV-1 Tat and Vpr, HSV-1 ICP22 and VP16 and its mutant with the P-TEFb using coarse-grained molecular dynamic (CGMD).