

ห้องสมุดงานวิจัย สำนักฯ: คณะกรรมการวิจัยแห่งชาติ



E46984



TARGET IDENTIFICATION OF HIV-1 INHIBITORS BY USING
PROTEIN-LIGAND DOCKING

MISS PINPISUT SENGCHA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE
(BIOINFORMATICS AND SYSTEMS BIOLOGY)

SCHOOL OF BIORESOURCES AND TECHNOLOGY AND
SCHOOL OF INFORMATION TECHNOLOGY
KING MONGRUT'S UNIVERSITY OF TECHNOLOGY THONBURI

2010



E46984

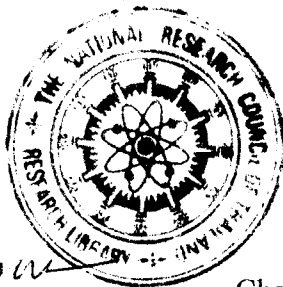
Target identification of HIV-1 inhibitors by using protein-ligand docking

Miss Pinpisut Sengcha B.Sc. (Computer Science)

A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Science (Bioinformatics and systems biology)
School of Bioresources and Technology and School of Information Technology

King Mongkut's University of Technology Thonburi

2010



Thesis Committee

.....
Supa Hannongbua
Chairman of Thesis Committee
(Assoc. Prof. Supa Hannongbua, Ph.D.)

.....
Marasri Ruengjitchatchawalya
Member and Thesis Advisor
(Asst. Prof. Marasri Ruengjitchatchawalya, Ph.D.)

.....
Ekachai Jenwitheesuk
Member and Thesis Co Advisor
(Researcher, Ekachai Jenwitheesuk, D.V.M., Ph.D.)

.....
Narongrit Waraporn
Member
(Lecturer, Narongrit Waraporn, Ph.D.)

.....
Rungtiva Palangsuntikul
Member
(Lecturer, Rungtiva Palangsuntikul, Dr. rer. nat.)

PREFACE

This thesis is written to carry out my master degree in field of Bioinformatics and Systems biology, King Mongkut's University of Technology Thonburi. This thesis is "Target identification of HIV-1 inhibitors by using protein-ligand docking" or thai title name is “การหาโปรตีนเป้าหมายของอนุพันธ์ต้านไวรัสเอชไอวีโดยใช้โมเลกุลาร์ docking” This thesis was done at King Mongkut's University of Technology Thonburi under collaboration with Biostatistics and Informatics laboratory, Genome Institute, BIOTEC, Thailand.

Thesis Title	Target identification of HIV-1 inhibitors by using protein-ligand docking
Thesis Credits	12
Candidate	Miss Pinpisut Sengcha
Thesis Advisor	Asst. Prof. Dr. Marasri Ruengjitchatchawalya
Thesis Co-advisor	Dr. Ekachai Jenwitheesuk
Program	Master of Science
Field of study	Bioinformatics and Systems biology
Faculty	School of Bioresources and Technology and School of Information Technology
B.E.	2553

E46984

Abstract

Acquired Immune Deficiency Syndrome (AIDS) is a global pandemic disease caused by human immunodeficiency virus type 1 (HIV-1). Currently, there is no therapeutic vaccine or drug that can cure this disease. Antiretroviral drug is the only option used to limit viral replication in patient’s body, thereby prolonging the development of full AIDS. To find a new antiretroviral drug, scientists have screened 42,689 chemical compounds against HIV-1 in vitro; 1,504 of which exhibited strong inhibitory activity. However, the inhibitory mechanisms of these inhibitors have not been elucidated. This study aimed to identify a viral protein target of these inhibitors by using protein-ligand docking approach. The X-ray crystal structures of seven HIV-1 proteins, i.e. protease, reverse transcriptase, integrase, gp41, capsid, nef, and tat were retrieved from the Protein Data Bank (PDB). The three dimensional structures of the inhibitors were downloaded from the ligand.info database. Each inhibitor structure (ligand) was then docked into the active site of each viral protein by using AutoDock 4.2 program. The binding affinity in terms of calculated inhibitory constant (K_i) was calculated for each protein-ligand pair, and was used as an indicator to identify the target of the inhibitor. The results showed that 116 ligands were predicted to fit with high affinity with HIV-1 proteins; 46 ligands (3.06%) fit with CAI binding site of the capsid, 41 ligands (2.73%) fit with CAP-1 binding site of the capsid, 37 ligands (2.46%) fit with the NRTI binding site of the reverse transcriptase, 36 ligands (2.39%) fit with gp41, 30 ligands (1.99%) fit with protease, 17 ligands (1.13%) fit with integrase, 4 ligands (0.27%) fit with the NNRTI binding site of the reverse transcriptase, and 6 ligands (0.40%) fit with nef

protein. Among the ligands that fit with the viral proteins, 54 ligands showed multitarget feature, i.e. the ability of the ligands that could bind with several proteins with high affinity. Most of the multitarget ligands had high affinity with protease and capsid since these proteins have large hydrophobic pockets that allow a wide range of ligands to easily access. The rest of the ligands (1,100 of 1,504) did not bind with significant affinity with viral proteins. This study suggests a list of potential target of HIV-1 inhibitors that may be experimentally verified in the laboratory. Further experiments such as mutagenesis study and ligand optimization may also be done to improve the potency of these inhibitors. This study also shows the utility of computational methods that can be applied to quickly identify potential targets of the inhibitors that the inhibitory mechanisms are unknown.

Keywords: AIDS/ Computational Docking Method/ Target Identification

หัวข้อวิทยานิพนธ์	การหาโปรตีนเป้าหมายของอนุพันธ์ต้านไวรัสเอชไอวีโดยใช้โมเลกุลาร์ด็อกกิง
หน่วยกิต	12
ผู้เขียน	นางสาวพินพิสทธิ์ เสง์ชา
อาจารย์ที่ปรึกษา	ผศ.ดร. มารศรี เรืองจิตชัชวาลย์
อาจารย์ที่ปรึกษาร่วม	ดร. เอกชัย เจนวิถีสุข
หลักสูตร	วิทยาศาสตร์มหาบัณฑิต
สาขาวิชา	ชีวสารสนเทศและชีววิทยาระบบ
คณะ	ทรัพยากรชีวภาพและเทคโนโลยี และเทคโนโลยีสารสนเทศ
พ.ศ.	2553

E46984

บทคัดย่อ

กลุ่มอาการภูมิคุ้มกันบกพร่องหรือเอดส์เป็นโรคระบาดที่แพร่ระบาดไปทั่วโลก ซึ่งเกิดจากเชื้อไวรัสเอชไอวี-1 ในปัจจุบันนี้เอดส์ยังไม่มีวัคซีนหรือยาที่สามารถรักษาให้หายขาดได้ ซึ่งยาที่ใช้ต้านเชื้อไวรัสในปัจจุบันถูกใช้ในการควบคุมการแพร่กระจายของเชื้อไวรัสในร่างกายของคนไข้ เพื่อไม่ให้มีการพัฒนาไปเป็นเอดส์ในระยะที่สมบูรณ์เท่านั้น กระบวนการหายาด้านเชื้อไวรัสชนิดใหม่นั้น นักวิทยาศาสตร์ได้ทำการคัดเลือกสารต้านไวรัสทั้งหมด 42,689 ชนิดจากการทดลองในห้องปฏิบัติการ และพบว่ามีสารต้านไวรัสทั้งหมด 1,504 ชนิดที่มีคุณสมบัติต้านเชื้อไวรัสเอชไอวี-1 ได้เป็นอย่างดี อย่างไรก็ตามยังไม่มีการศึกษาใดที่ทำการศึกษาระบวนการทำงานของโปรตีนกับสารต้านไวรัสเหล่านี้ ดังนั้นจุดมุ่งหมายของการวิจัยครั้งนี้เพื่อทำการหาโปรตีนเป้าหมายของสารต้านไวรัสดังกล่าวโดยใช้วิธี protein-ligand docking โดยได้นำโครงสร้างผลึกอิเล็กทรอนิกส์ของโปรตีนทั้งเจ็ดชนิดของไวรัส ได้แก่ protease, reverse transcriptase, integrase, gp41, capsid, nef, และ tat จาก Protein Data Bank และนำโครงสร้างสามมิติของสารต้านไวรัสจากฐานข้อมูลของ Ligand.info ซึ่งในที่นี้โครงสร้างของสารต้านไวรัสแต่ละชนิดจะถูกนำมา dock กับบริเวณเร่งปฏิกิริยาของโปรตีนแต่ละชนิดโดยใช้โปรแกรม autoDock เวอร์ชัน 4.2 เพื่อที่จะทำการพิจารณาค่า inhibitory constant (K_i) ในแต่ละคู่โปรตีนกับ ligand เพื่อบอกถึงประสิทธิภาพในการจับกันของโปรตีนและ ligand และใช้เป็นตัวบ่งชี้ในการหาโปรตีนเป้าหมายของสารต้านไวรัสดังกล่าว จากผลการวิจัยนี้พบว่า มีสารต้านไวรัสจำนวน 116 ชนิด มีประสิทธิภาพในการยับยั้งเชื้อไวรัสเอชไอวี-1 สูง โดยมี ligand จำนวน 46 ชนิด (3.06%) จับกับ capsid ที่บริเวณเร่งปฏิกิริยาของ CAI มี ligand จำนวน 41 ชนิด (2.73%) จับกับ capsid ที่บริเวณเร่งปฏิกิริยาของ CAP-1 มี ligand จำนวน 37 ชนิด (2.46%) จับกับ reverse transcriptase ที่บริเวณเร่งปฏิกิริยาของ NRTI มี ligand จำนวน 36 ชนิด (2.39%) จับกับบริเวณเร่งปฏิกิริยาของ gp41 มี ligand จำนวน 30 ชนิด (1.99%) จับกับ

บริเวณเร่งปฏิกิริยาของ protease มี ligand จำนวน 17 ชนิด (1.13%) จับกับบริเวณเร่งปฏิกิริยาของ integrase มี ligand จำนวน 6 ชนิด (0.40%) จับกับบริเวณเร่งปฏิกิริยาของ nef และมี ligand จำนวน 4 ชนิด (0.27%) จับกับ reverse transcriptase ที่บริเวณเร่งปฏิกิริยาของ NNRTI นอกจากนั้นยังพบว่าในจำนวน ligand ที่จับกับโปรตีนดังกล่าว มี ligand ทั้งหมด 54 ชนิดที่สามารถจับกับโปรตีนเป้าหมายได้หลายชนิด โดยส่วนใหญ่ ligand ในกลุ่มนี้สามารถจับกับ protease และ capsid ซึ่งโปรตีนทั้งสองมี hydrophobic pocket ที่ใหญ่พอที่จะให้ ligand เข้าไปได้อย่างสะดวก ส่วน ligand ที่เหลือจำนวน 1,100 ชนิด นั้นพบว่าไม่สามารถจับกับโปรตีนของไวรัสที่เรา拿來ศึกษาได้ ผลการศึกษาทั้งหมดในการวิจัยครั้งนี้นำมาซึ่งรายชื่อของโปรตีนเป้าหมายที่สามารถจับกับสารต้านไวรัสเอชไอวี-1 เพื่อนำมาศึกษาและตรวจสอบในห้องปฏิบัติการ เช่น ในการศึกษาการกลายพันธุ์ และการหา ligand ที่เหมาะสมในการพัฒนาเป็นสารต้านไวรัสชนิดใหม่ เป็นต้น นอกจากนี้ ผลการวิจัยครั้งนี้ยังแสดงถึงวิธีในการหาโปรตีนเป้าหมายของสารต้านไวรัสซึ่งยังไม่มี การทำการศึกษามาก่อนได้อย่างมีประสิทธิภาพและรวดเร็วโดยใช้คอมพิวเตอร์มาประยุกต์

คำสำคัญ: AIDS/ Computational Docking Method/ Target Identification

ACKNOWLEDGEMENTS

I would like to thank Asst. Prof. Dr. Marasri Ruengjitchatchawalaya, my advisor, for many helpful advices and guidance. Also, I would like to thank my co-advisor, Dr. Ekachai Jenwitheesuk who spent time and effort for introducing many important suggestion and commentation to precede thesis. Additionally, I really spend all fully thanks to my committees, Dr. Narongrit Waraporn and Dr. Rungtiva Palangsuntikul who contributed in my thesis especially in giving information for their time in various helpful discussion and suggestion. Other ones that I would like to thank are Ms. Alisa Wilantho, Ms. Wanwimon Mokmak, Mr. Chumpol Ngamphiw, Ms. Pavita Tipsombatboon and members of Biostatistics and Informatics laboratory, Genome Institute, BIOTEC for their kindness and many helpful in the fields of software, guidelines for docking calculation and output interpretation.

I would like to give a special thank for all my bioinformatics friends who encourage me with many words and cheerful. They always made every time that I spent with them to be a special occasion and that made me thought positive to overcome any troubles. Additionally, my sincere appreciation is expressed to all teachers and both National Center for Genetic Engineering and Biotechnology (BIOTEC), Thailand and King Mongkut's University of Technology Thonburi (KMUTT) for the full scholarship to allow me carry out my master's degree study.

Finally, I would like to thank my family who encourage and understand everything that I did and support me no matter what I need. With the best thanks, their kindness and happiness that they give to me, made me being in the right way and help me to overcome many problem through this thesis.

CONTENTS

	PAGES
PREFACE	i
ENGLISH ABSTRACT	ii
THAI ABSTRACT	iv
ACKNOWLEDGEMENTS	vi
CONTENTS	vii
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF TECHNICAL VOCABULARY AND ABBREVIATIONS	xiv

CHAPTER

1. INTRODUCTION	1
1.1 Background and Rationale	1
1.2 Objective	4
1.3 Scope of work	4
1.4 Expected Result	4
 2. LITERATURE REVIEWS	 5
2.1 Human immunodeficiency virus	5
2.1.1 HIV life cycles	6
2.1.2 Essential protein of HIV	7
2.2 Target identification	18
2.2.1 Available software for target identification	20
2.3 Force field, free energy and inhibitory constant	20
2.3.1 Force field energy	20
2.3.2 The binding free energy	21
2.3.3 Inhibitory constant	22
2.4 AutoDock suite	22
2.4.1 Larmackian genetic algorithm	23
2.4.2 Scoring function in AutoDock	25
2.4.3 Grid maps	28

3.	MATERIALS AND METHODS	29
3.1	Materials and methods	29
3.1.1	Preparation of ligands and protein structures for docking	30
3.1.2	Protein-ligands docking	32
3.1.3	Analysis of the result	35
4.	RESULTS AND DISCUSSIONS	39
4.1	Preparation of ligands and protein structures for docking	39
4.1.1	Preparation of ligand structures	39
4.1.2	Preparation of protein structure	40
4.2	Result of protein-ligand docking	41
4.2.1	Validate docking system base on RMSD	42
4.2.2	Histogram based on inhibitory constant value for program validation	43
4.3	Identified target for positive set	46
4.3.1	Identified drug target for positive ligands set base on inhibitor constant	46
4.4	Discussion the results by using Venn diagram	52
4.4.1	Maximum and minimum number of ligands fit with set of protein	52
4.4.2	Large set of ligand fit with RT (NRTIs) apart from other protein based on venn diagram	54
4.4.3	Set of intersect ligand which fit with protease, integrase and nef protein	54
4.4.4	Set of intersect ligand which fit with capsid (CAI), capsid (CAP-1), gp41 and tat protein	55
4.5	Interaction of LigandID941-protein complex	56
4.5.1	Interaction between ligandID941 and protease binding site	56
4.5.2	Interaction between ligandID941 and RT at NNRTIs binding site	60
4.5.3	Interaction between ligandID941 and RT at NRTIs binding site	63

4.5.4	Interaction between ligandID941 and capsid (CAI) binding site	64
4.5.5	Interaction between ligandID941 and capsid (CAP-1) binding site	66
4.5.6	Interaction between ligandID941 and nef protein binding site	69
4.5.7	Interaction between ligandID941 and gp41 binding site	70
4.5.8	Interaction between ligandID941 and integrase binding site	71
4.5.9	Interaction between ligandID941 and tat protein binding site	73
4.5.10	Conclusion of ligandID941 with nine binding site	75
4.6	Database of predicted protein targets for HIV-1 inhibitors	76
4.6.1	Customer view	76
5.	CONCLUSION AND RECOMMENDATION	91
5.1	Conclusion	91
5.2	Recommendations	92
	REFERENCES	93
	APPENDIX	
A.	Script for running molecular docking and extract the results	96
	CURRICULUM VITAE	106

LIST OF TABLES

TABLES		PAGES
2.1	The list of docking programs are widely used	20
3.1	Represent our interesting protein	31
3.2	Represent attribute in each table in customer view	38
4.1	Represent list of proteins and their binding site	40
4.2	Show ranges for free energy and inhibitory constant value graphs	41
4.3	Estimated color scale range with log scale inhibitor value of each protein	47
4.4	Represent volume and flexibility bond of capsid (CAI) and RT (NNRTI)	53
4.5	Represent interaction of LigandID941-protease with H bond	57
4.6	Represent interaction of U0E-protease with H bond	57
4.7	Represent interaction of LigandID941-RT (NNRTIs) with H bond	60
4.8	Represent interaction of LigandID941-RT (NRTIs) with H bond	63
4.9	Represent interaction of LigandID941-capsid at CAP-1 binding site by H bond	66
4.10	Represent interaction of JPR-capsid at CAP-1 binding site by H bond	66
4.11	Represent interaction of LigandID941-nef binding site	69
4.12	Represent interaction of ligandID941-integrase with H bond	71
4.13	Represent interaction of reference-integrase with H bond	71
4.14	Represent interaction of ligandID941-tat protein with H bond	74
4.15	Represent the conclusion of ligandID941 with nine binding site	75

LIST OF FIGURES

FIGURES		PAGES
2.1	Represent life cycle of HIV-1 and schematic of a mature HIV virion	6
2.2	Schematic of a mature HIV virion	8
2.3	Structure of reverse transcriptase; p66 and p51 subunits	9
2.4	Represent HIV-1 protease with inhibitor	10
2.5	Represent HIV-1 IN activity	11
2.6	Represent HIV-1 IN	11
2.7	Represent overall view of N36 and C34 complex	13
2.8	Represent the structure of arrays of the viral CA (capsid protein) protein	13
2.9	Represent Hexagonal Lattice of the HIV-1 CA Protein	14
2.10	Represent the binding site of both inhibitor a.) Related with CAP-1 inhibitor.) Related with CAI inhibitor	15
2.11	Represent Nef protein binding site	16
2.12	Represent major groove view of TAR RNA without tat (left) and TAR RNA with tat protein bound (right). Residue Ala22, Sec23, and Gly26 (yellow highlighted) and arginine side-chain is not shown	17
2.13	Represent binding site of Tat protein	17
2.14	Represent drug discovery pipeline vs. computer-aided drug design (CADD) tools	18
2.15	Represent diagram for target identification by using reverse docking approach	18
2.16	Represent virtual screening and high-throughput screening	19
2.17	Illustration of the fundamental force field energy terms	21
2.18	Schematic representation of genetic algorithm	23
2.19	Represent dock parameter file of AutoDock	24
3.1	Overview all method	29
3.2	Represent values of the grid box parameters	32
3.3	Values of the genetic algorithm parameters	34
3.4	The validated histogram of each protein	35
3.5	Created the cut-off for receive the bound ligand with our protein and receive high binding affinity	36
3.6	Example of the binding pattern of positive set	37
4.1	Represent conversion of ligand in .sdf format to .pdb format via OpenBable	39
4.2	Represent co-crystal structure ligand (purple) and redocked ligand (green)	42
4.3	Created normal estimated inhibitory constant histogram	43
4.4	Estimated histogram of entry protein.	45

4.5	Represent three color scales and the cut-off per each	46
4.6	Represent binding pattern to identify the target using inhibitor constant	49
4.7	Represent number of ligands in each level from 1,504 ligands	51
4.8	Represent venn diagram from VennMaster program based on our information	52
4.9	Represent maximum number of ligands fit with capsid (CAI) and minimum number of ligands fit with RT at NNRTIs binding site	52
4.10	Represent 33 ligands fit with RT (NRTIs) apart from other protein	54
4.11	Represent similarity ligand show in set of protease, integrase and nef protein	54
4.12	Represent the similarity cylinder shape of protease (PDB ID: GNO1)	55
4.13	Represent similarity ligand show in set of capsid (CAI), capsid (CAP-1), gp41 and tat protein	55
4.14	Represent 2D structure of ligandID941	56
4.15	Illustration of ligandID941-protease interaction (A) and U0E-protease interaction (B)	59
4.16	Illustration of ligandID941-RT (NNRTIs) interaction (A) and NVP-RT (NNRTIs) interaction(B)	62
4.17	Illustration of ligandID941-RT (NRTIs) interaction	64
4.18	Illustration of ligandID941-capsid (CAI) interaction	65
4.19	Illustration of ligandID941-capsid (CAP-1) interaction (A) and NVP-capsid (CAP-1) interaction (B)	68
4.20	Illustration of ligandID941-nef protein interaction	69
4.21	Illustration of ligandID941-gp41 interaction	70
4.22	Illustration of ligandID941-integrase interaction	73
4.23	Illustration of ligandID941-tat protein interaction	74
4.24	Represent the index of target identification of HIV-1 inhibitor web-site	77
4.25	Represent the search information paper	78
4.26	Represent the result when we search information with interested ligand	79
4.27	Represent result when we search all ligand with interested protein	80
4.28	Represent result when search all ligand with interested protein and set range of value	81
4.29	Represent search score for compare page	82
4.30	Show the result of search scale of muti-protein and one ligand interaction	83
4.31	Show the result of search scale of muti-protein and all ligand interaction	84

4.32	Show the result of search scale of multi-protein and one ligand interaction and selected specific color range scale	85
4.33	Show the result of search scale of multi-protein and all ligand interaction and selected specific color range scale	86
4.34	Represent search the highest score page	87
4.35	Represent search the highest score page	88
4.36	Represent search the highest score page result	88
4.37	Represent search compare	89
4.38	Represent search compare result	90

LIST OF TECHNICAL VOCABULARY AND ABBREVIATIONS

3D	3 Dimension
3D QSAR	3 Dimension Quantitative Structure Activity Relationship
ADIS	Acquired Immune Deficiency Syndrome
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
CA	Capsid
CADD	Computer Aided Drug Designing
CAI	Capsid Assembly Inhibitor
CAP-1	N-(3-chloro-4-methylphenyl)-N'-{2-[(5-[(dimethylamino)-methyl]-2-furyl)-methyl]-sulfanyl}-ethyl}urea
CD4	Cluster of differentiation 4
CTD	C-Terminal Domain
Cys	Cysteine
DNA	Deoxyribonucleic acid
dsDNA	double-stranded Deoxyribonucleic acid
E	Enzyme
Env	Envelope
FDA	U.S. Food and Drug Administration
FF	Force Field
GA	Genetic Algorithm
Glu	Glutamic acid
Gly	Glycine
Gp 41	Transmembrane envelope glycoprotein
Gp120	Surface envelope glycoprotein
HAART	High Active Antiretroviral Therapy
His	Histidine
HIV-1	Human Immunodeficiency Virus type 1
HIV-1 RT	Human Immunodeficiency Virus type 1 Reverse Transcriptase
I	inhibitor
IC ₅₀	Median inhibitory concentration
Ile	isoleucine
IN	Integrase
kDa	kilodalton
K _i	Inhibition constants
LBDD	Ligand based drug design
Leu	Leucine
LGA	Larmackian Genetic Algorithm
LTR	Long terminal repeat
Lys	Lysine

MA	Matrix
MC	Monte Carlo
mRNA	messenger Ribonucleic acid
NC	Nucleocapsid
Nef	Negative regulatory factor
NMR	Nuclear magnetic resonance
NNRTIs	Non Nucleoside reverse transcriptase inhibitor
NRTIs	Nucleoside reverse transcriptase inhibitor
NTD	N-Terminal Domain
PBPK	Physiologically-based Pharmacokinetic
PDB	Protein Data Bank
PEOE	Partial Equalization of Orbital Electronegativities
Phe	Phenylalanine
PHP	Personal Home Pages
PIs	Protease Inhibitor
PR	Protease
Pro	Proline
QSAR	Quantitative Structure Activity Relationship
RNA	Ribonucleic acid
RNase H	Ribonuclease H
RMSD	root mean square deviation
RT	Reverse Transcriptase
SA	Simulated Annealing
SBDD	Structure-Based Drug Design
SD	Standard deviation
SDF	structure-data file
Ser	Serine
ssRNA	single-stranded Ribonucleic acid
SQL	Structured Query Language
SU	Surface glycoprotein
SW	Solic and Wet
TAR	Transactivation response region
Tat	Trans-activator of transcription
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
TM	Transmembrane glycoprotein
Val	valine