國立臺灣大學理學院化學系

碩士論文

Department of Chemistry

College of Science

National Taiwan University

Master Thesis

甲基化賴胺酸對 α 螺旋與 β 折板結構穩定度

及對核糖核酸辨識與細胞穿透之影響

Effect of Lysine Methylation on

α-Helix Propensity, β-Sheet Propensity,

RNA Recognition, and Cell Penetration

劉牧群

Mu-Chun Liu

指導教授: 陳平 博士

Advisor: Richard P. Cheng, Ph.D.

中華民國 103 年 7 月

July, 2014

國立臺灣大學碩士學位論文
口試委員會審定書

（論文中文題目）甲基化賴胺酸對 α 螺旋與 β 折板結構穩定度

及對核糖核酸辨識與細胞穿透之影響

（論文英文題目）Effect of Lysine Methylation on α-Helix

Propensity, β-Sheet Propensity, RNA Recognition, and Cell

Penetration

本論文係＿＿劉牧群＿＿君（學號 R01223160 ）在國立臺灣大學
化學系完成之碩士學位論文，於民國 103 年 7 月 14 日承下列考試委員
審查通過及口試及格，特此證明。

口試委員：

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿（簽名）
（指導教授）

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

系主任、所長 ＿＿＿＿＿＿＿＿＿＿＿＿＿＿（簽章）

誌謝

　　首先要非常感謝　陳平老師自大三至碩二以來長年的指導，提供了很多實驗幫助、論文修改、對自己所修習的領域有更多的了解以及對未來的想法與幫助。碩士班兩年雖然在蛋白質領域還不能算上是非常精通，但經由老師的傾囊相授後也學了很多相關的知識以及實驗方法，更重要的是培養自己找需要資訊的能力，獲益良多！

　　另外也非常高興能夠認識實驗室一起同甘共苦的夥伴們：已經畢業的振勳大學長耐心地指導實驗方法以及解答許多課業上的問題，非常可靠！秀婷學姊在光譜上的大力幫助以及可口的台南名產都是撰寫這篇論文不可或缺的要素；身為同儕但能力與大學長並列哲豪在實驗和課業上的單方面幫忙讓我受益不少；與靖凱和清怡之間的瑣事暢談讓碩士生涯中的許多壓力舒緩了不少；碩一的學弟妹們云喬、東偉、伯宜、昌華讓這個實驗室總是不缺歡樂的氣氛。非常感謝實驗室的所有夥伴在這兩年之間的多方幫忙與照顧，非常榮幸能認識大家！最後要感謝我的父母提供他們的一切讓我能毫無牽掛的完成我的學業。

<div align="right">2014.07.01</div>

中文摘要

轉譯後修飾在蛋白質行為上的影響非常重要，而賴胺酸的甲基化是其中一種非常常見的範例，甲基化可能會造成蛋白質結構上的變化以及功能上的活化或抑制。賴胺酸甲基化在生物體中有三種形式：單甲基、雙甲基和三甲基，不同程度的甲基化可能對蛋白質也有不同層級和效果的影響。在結構方面，三種甲基化賴胺酸被置入於兩種基本二級結構的胜肽模板中來研究其對二級結構穩定度的影響：α-螺旋和β-折板。圓二色光譜儀用來測量胜肽的螺旋程度，而折板的結構資訊則使用二維核磁共振光譜來分析。

於功能方面，本論文則是探討於人類免疫缺乏病毒的感染過程中非常重要的調控蛋白：Tat。此蛋白中存在著一段富含正電荷胺基酸的區域（RKKRRQRR）並且其已被研究出可以高選擇性地與特定 RNA 結構結合藉以大幅度提高病毒蛋白的表現量。這段非常重要的序列同時也賦予 Tat 蛋白穿透細胞膜的能力來入侵其他細胞。而賴胺酸的甲基化被證實與此蛋白的活化和抑制息息相關。因此，我們合成這段胜肽並在 50 和 51 號分別置入三種甲基化賴胺酸來研究其對辨認特定 RNA 以及穿透細胞能力的影響。與 RNA 之間的解離常數由膠體電泳計算，而穿透細胞的能力則使用流式細胞儀來進行測試。

Abstract

Post-translational modification dominates many protein behaviors. Methylation of lysine impacts both protein function and structure. There are three variations of methylated lysines that were identified in proteins. It is logical to assume that different numbers of methyl groups attached onto the side chain amino group should have different degrees of effects on proteins. In this study, various types of methylated lysines are placed into two basic secondary structures: the α-helix and the simplest β-sheet model "β-hairpin", to investigate the effect of lysine methylation on structural stability. The fraction helix of the helical peptides was determined by circular dichorism spectroscopy. The structural information of the hairpin peptides was analyzed by 2D NMR.

Lysine methylation also plays an important part in many biological processes. The regulatory Tat protein contains a basic region (RKKRRQRRR, residue 49 to 57) which specifically binds to the trans-activating responsive (TAR) element to modulate HIV-1 RNA transcription. The binding between HIV-1 Tat protein and TAR RNA is essential for HIV-1 virus to efficiency produce full-length viral RNA. To study the effect of lysine methylation on RNA recognition and cellular uptake, two lysine residues Lys_{50} and Lys_{51} were replaced with monothylated, dimethylated, and trimethylated lysines.

The dissociation constant for the Tat derived peptide-TAR RNA complexes was determined by gel shift assay. The cellular uptake efficiency of Tat derived peptide into Jurkat cell was assessed by flow cytometry.
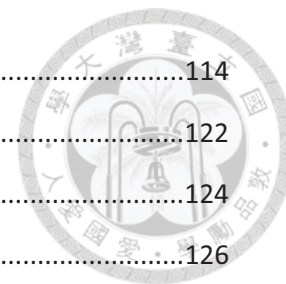
Table of Contents

List of Charts

List of Figures

List of Tables

List of Schemes

Abbreviation

α-SYN	α-Synuclein
Aβ	Amyloid β peptide
Ac	Acetyl
AD	Alzheimer's disease
AIDS	Acquired immune deficiency syndrome
Ala	Alanine
APS	Ammonium persulfate
Arg	Arginine
Bis-acrylamide	N,N'-methylene-bis-acrylamide
CCR5	C-C chemokin receptor type 5
CD	Circular dichroism
CD4	Cluster of differentiation 4
CDK9	Cyclin-dependent kinase 9
CPPs	Cell-penetrating peptides
Cys	Cysteine
CXCR4	C-X-C chemokine receptor type 4
DIEA	Diisopropylethylamine
DMF	Dimethylformamide
Dmk	Dimethyllysine
DNA	Deoxyribonucleic acid
DQF-COSY	Double-quantum filtered-correlated spectroscopy
E. coli tRNA	*Escherichia coli* transfer ribonucleic acid
EMSA	Electrophoretic mobility shift assay
FBS	Fetal bovine serum
Fl	6-Carboxy-fluorescein
Fmoc	N-9-Fluorenylmethoxycarbonyl
Gln	Glutamine
Gly	Glycine
gp120	Envelope glycoprotein GP120
HBTU	*O*-1H-benzotriazol-1-yl-1,1,3,3-tetramethyluronium hexafluorophosphate
HIV	Human immunodeficiency virus
HOBT	1-Hydroxybenzotriazole
Ile	Isoleucine
K_D	Dissociation constant
Leu	Leucine

LTR	Long terminal repeat
Lys	Lysine
MALDI-TOF	Matrix-assisted laser desorption ionization time-of-flight
MeOH	Methanol
Mmk	Monomethyllysine
NELF	Negative elongation factor
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
Orn	Ornithine
PD	Parkinson's disease
PMT	Photomultiplier tube
PRC2	Polycomb repressive complex 2
Pro	Proline
PrP	Prion protein
P-TEFb	Positive transcriptional elongation factor-b
PTMs	Post-transcriptional modifications
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
ROESY	Rotating-frame nuclear Overhauser effect correlation spectroscopy
RT	Reverse transcriptase
RP-HPLC	Reversed phase high-performance liquid chromatography
RPMI	Roswell Park Memorial Institute medium
SPPS	Solid phase peptide synthesis
TAR	Trans-activation response element
Tat	Trans-activator of transcription
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFA	Trifluoroacetic acid
Thr	Threonine
Tmk	Trimethyllysine
TOCSY	Total correlation spectroscopy
Tris	Tris (hydroxylmethyl)-aminomethane
Tyr	Tyrosine
UV-vis	Ultraviolet-visible
Val	Valine

Chapter 1

Introduction

Chapter 1. Introduction

1-1 Central Dogma of Molecular Biology

DNA, RNA, and proteins are the three crucial marcomolecules in living organisms. The central dogma was introduced by Crick to describe the process of producing proteins from DNA through RNA in 1958 (Figure 1-1).[1] DNA is a biopolymer that carries genetic information. DNA is duplicated before a cell undergoes self-replication. DNA is also used to produce pre-RNA through transcription. Both duplication and transcription of DNA occur in the cell nucleus. RNA is processed through RNA splicing to remove the non-coding regions before translation. The mature RNA is transported to the cytoplasm. Proteins are built based on the corresponding genetic code on the mature RNA through translation.

Cell nucleus   **Cytoplasm**

DNA self-replication
DNA $\xrightarrow{\text{Transcription}}$ pre-RNA $\xrightarrow{\text{RNA-splicing}}$ mature RNA $\xrightarrow{\text{Translation}}$ Proteins
Reverse Transcription
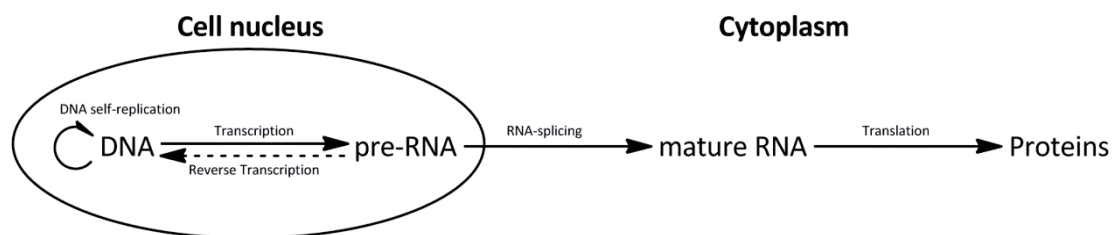
Figure 1-1. The central dogma of molecular biology is the genetic information flowing from DNA through RNA to proteins.[1] The solid arrows indicate the information flow that occurs in all eukaryotic cells. The dashed arrow indicates the information flow that occasionally occurs in viruses through reverse transcriptases.

1-2 Proteins

Proteins are the end products of the central dogma. Based on the unique genetic code carried by the RNA, each protein is composed of different types and number of amino acid. Most amino acids are L-α-amino acids. Proteins are linear biopolymers with peptide bonds linking an α-carboxyl group of one amino acid and an α-amino group of another. The peptide bond is planar with six atoms in the same plane. The length of a peptide bond is 1.32 Å, which is between a C-N single bond (1.49 Å) and a double bond (1.27 Å), suggesting partial double bond character.[2] Each amino acid contains a different side chain functional group, allowing proteins to perform various bioactivities. Proteins are essential elements that control nearly all cellular functions. There are several types of proteins differing in utility including structural components,[3] signal transduction,[4] catalysis,[5] and immune response.[6] Proteins are responsible for almost all bioactivities in the cell, and thus studies to enhance the fundamental knowledge on proteins should improve our understanding of nature, along with potential technological advancement.

1-3 Protein Folding and Function
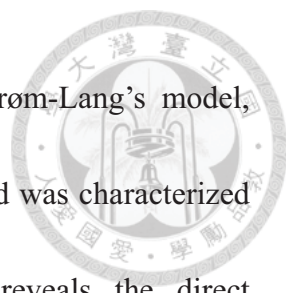
In order to perform various biological functions, proteins must fold into three-dimensional structures with high accuracy. Different protein structures give rise to

various protein functions.[3, 7] For example, at least 15 distinct enzyme families require a specific protein fold named α/β barrel to construct the appropriate active site geometry.[8] If proteins are denatured or mutated and cannot fold correctly into the corresponding three dimensional shape, proteins lose their functions or even lead to protein misfolding diseases such as Alzheimer's,[9] Parkinson's,[10] Huntington's,[11] and Crutzfeldt-Jacob (prion) diseases.[12] Alzheimer's disease (AD) is a clinical syndrome caused by neurodegeneration and was estimated that 24.3 million people suffered from it in 2001.[13] AD is related to the abnormal formation and accumulation of amyloid β peptide (Aβ) and tau protein.[14] Parkinson's disease (PD) is a common nerval syndrome caused by the abnormal aggregation of a stable tetrameric protein, α-synuclein (α-SYN), to form insoluble fibrils.[15] Prion disease is also caused by the aggregation of a helical-containing protein called prion protein (PrP).[16] These three diseases are all involved in peculiar protein stacking of once structurally diverse proteins into β-sheet structured amyloid fibrils. Importantly, the exact conformation of a protein plays an important role in its function. Thus, a thorough study of protein function at the molecular level requires detailed structural analysis.

1-4 Hierarchy of Protein Structure

In 1952, Linderstrøm-Lang proposed the hierarchy of protein structure with four

levels: primary, secondary, tertiary, and quaternary.[17] In Linderstrøm-Lang's model,

each level was constructed by the elements of the previous level and was characterized

by specific patterns of interactions.[17] The primary structure reveals the direct

composition of a protein in the unit of various types of amino acids, starting from the

amino-terminal end (N) to the carboxyl-terminal end (C'). The main-chain atoms are an

NH group of one residue bound to Cα, a central carbon atom (Cα) to which the side

chain (R) is attached, and a carbonyl group C'=O linked to the NH of another residue.

The backbone atoms are basically composed of a repeating unit $(NH\text{-}C\alpha H\text{--}C'\text{=}O)_n$,

which serve as the common framework of an amino acid (Figure 1-2). In order to

describe the structural properties of a protein, another method is introduced to

characterize the main chain. The original repeating unit can be viewed as one central

carbon $(C\alpha_{n+1})$ extending to its prior $(C\alpha_n)$ and subsequent central carbons $(C\alpha_{n+2})$. As

discussed earlier, the peptide C-N bond has partial double bond character.[18] This

character allows the peptide bond to arrange six main chain atoms

$(C\alpha_n\text{-}C'O\text{-}NH\text{-}C\alpha_{n+1}$ and $C\alpha_{n+1}\text{-}C'O\text{-}NH\text{-}C\alpha_{n+2})$ in a rigid planar structure.[2] Two

neighboring rigid planar structures are linked by the covalent bonds with the Cα atom,

rotating through N-Cα and Cα-C' bonds. The two conventional dihedral angles for these

two bonds are named phi (φ) and psi (ψ), respectively (Figure 1-2).
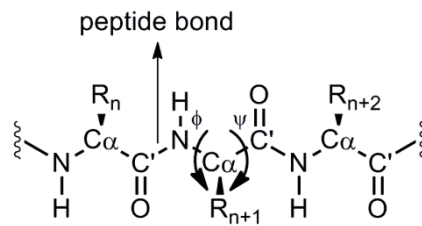
Figure 1-2. The peptide bond and the dihedral angles ϕ and ψ in the backbond.

The combinations of the dihedral angles are used to describe the structural properties of the main chain. Most of the combinations of φ and ψ angles are not allowed due to steric clashes between the peptide backbone and the side chains.[19] G. N. Ramachandran calculated and plotted the sterically allowed regions as Ramachandran plots with the dihedral angles ranging from -180° to 180° (Figure 1-3).[19] The allowed regions depend on the permitted van der Waals contact distance and the combination of dihedral angles.[19]
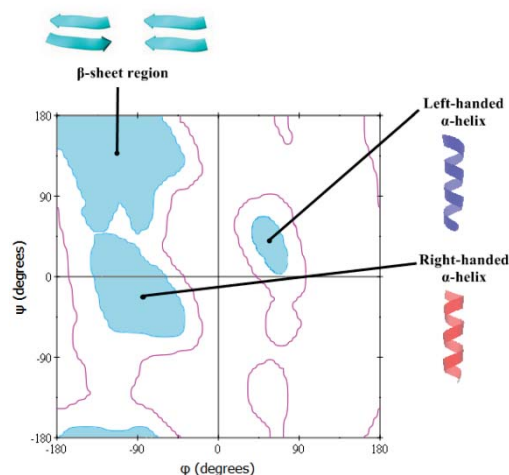
Figure 1-3. Ramachandran plot.[19] The X axis is φ and the Y axis is ψ angles, and the angle regions are from -180° to 180°.

Secondary structure is defined by patterns of hydrogen bonds between the backbone amide and carboxyl groups. The basic secondary structures are α-helix

and β-sheet.[20] The α-helix was first described by Pauling in 1951.[21] The α-helix is a right-handed coil with dihedral angles ϕ = -57° and ψ = -47°.[22, 23] The coil-like structure has 3.6 residues per turn and is characterized by consecutive, main-chain, $i \leftarrow i+4$ hydrogen bonds between each carbonyl oxygen (i) and an amide hydrogen ($i+4$) on the adjacent helical turn (Figure 1-4).[24] One third of all protein residues adopt an α-helix conformation, showing that helical proteins play important roles in living organism.[25]

Figure 1-4. The structure of an α-helix (an α-helix from a four-α-helix bundle, PDB 2I7U).

β-Sheet is another common secondary structure. It is a flat plate configuration containing multiple β-strands with inter-strand hydrogen bonds between backbone C'=O and N-H on neighboring strands. β-Sheets can be further categorized into two types: parallel and anti-parallel, distinguished by the arrangement of the hydrogen bond orientation.[26] A parallel β-sheet is characterized by a series of twelve-membered hydrogen-bonded rings, while an anti-parallel β-sheet is characterized by an alternating series of ten-and fourteen-membered hydrogen-bonded rings. The dihedral angles of parallel and anti-parallel β-sheets are (ϕ = -119°, ψ = +113°) and (ϕ = -139°, ψ = +135°), respectively. β-Hairpins are one of the simplest super-secondary structures, consisting of

two anti-parallel β-strands connected through a short loop region (Figure 1-5).[27-29]

Figure 1-5. The structure of a β-hairpin (the C-termini β-hairpin from GB1 protein, PDB 2PLP).

Tertiary structure refers to the stable three-dimensional structure formed by a polypeptide chain.[30] Various recurring secondary structures assemble to form the tertiary structure, which is required to perform different and precise protein functions. X-ray analysis has revealed significant relationship between function and structure. Domains are the fundamental units of tertiary structure, which are also closely related to protein function. The concept of a domain was first introduced by Wetlaufer after X-ray studies of hen lysozyme and papain,[31, 32] and proteolysis studies of immunoglobulins.[33, 34] Protein tertiary structures can be divided into four major classes based on their secondary structure content of the domain: all-α domains, all-β domains, α+β domains, and α/β domains.[35] According to an algorithm named "Structural Classification of Proteins (SCOP) Database", which investigates sequences and structures, these common folds account for 16.2%, 22.6%, 25.4%, and 23.4% of the total 87681 structural hits, respectively.[36] Pyruvate kinase is a phosphate group-transferring enzyme that plays an crucial role in glycolysis. It contains three major domains: an all-β regulatory domain,
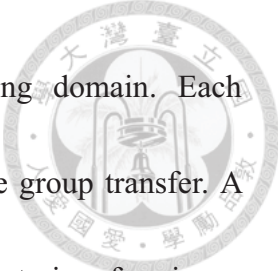
an α/β substrate binding domain, and an α/β nucleotide binding domain. Each structurally different domain serves a different purpose in phosphate group transfer. A typical tertiary structure has its nonpolar residues buried in the interior, forming a hydrophobic core.[37] Polar and charged residues are more frequently found on the surface, where proteins can interact with the aqueous environment through the hydrophilic side chains.[37]

Quaternary structure is the spatial assemble of multiple polypeptide chains.[38] Examples of proteins with quaternary structure include hemoglobin, DNA polymerase, and ion channels. Conformational change or re-orientation of individual polypeptides can induce changes in quaternary structure or connection between polypeptides. Through such structural changes, protein function can be regulated and exert their physiological function.

Each level of protein structure is held together by characteristic interactions and forces. Higher levels of proteins structure are assembled through the structural units of the lower level (Figure 1-6). Among the protein structure hierarchy, the secondary structural level plays a key role in protein folding. Therefore, research on the factors that affect the formation of secondary structure is important for understanding protein structure formation and prediction.

Primary   Secondary   Tertiary   Quaternary
Structure   Structure   Structure   Structure

Lys₁₈
Lys₁₉
Ser₂₀
Leu₂₁
Gly₂₂
Glu₂₃
Leu₂₄
Ile₂₅
His₂₆
Thr₂₇
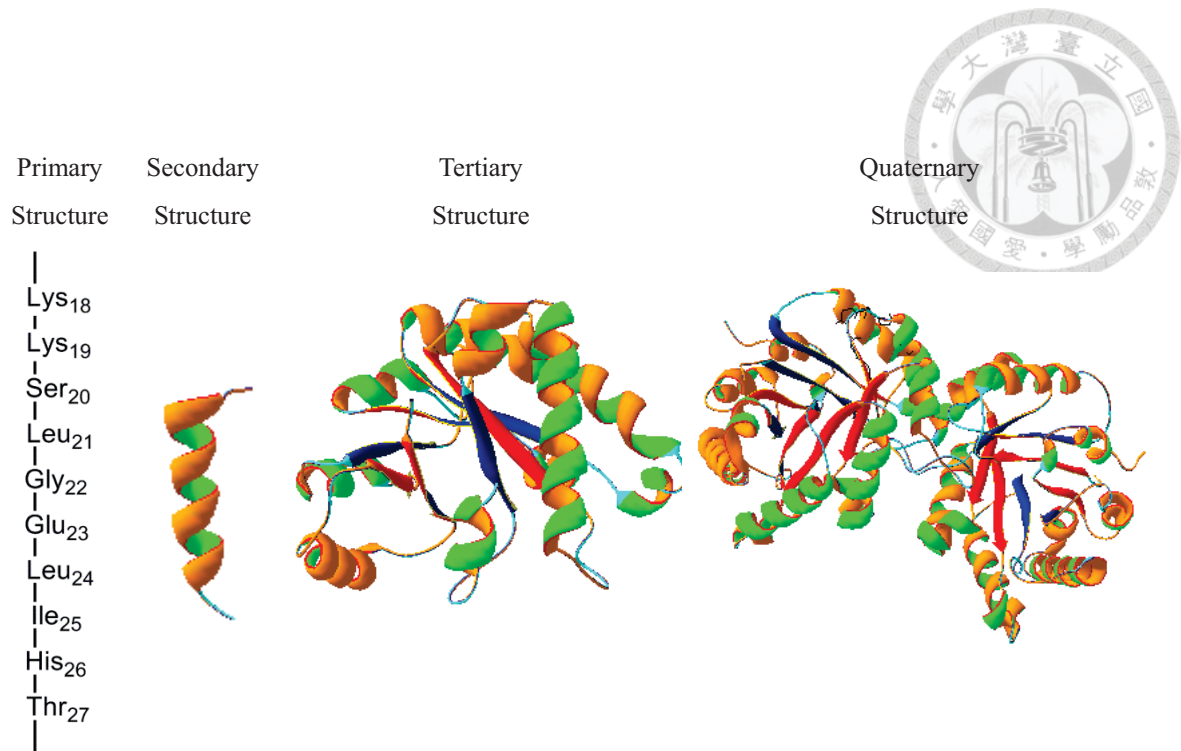
Figure 1-6. Four hierarchical levels of protein structure (triosephosphate isomerase, PDB 8TIM).

1-5 Driving Force of Protein Folding

Proteins must fold into the native structure to carry out its function. There are four dominant forces for protein folding and all these four forces are non-covalent in nature.[39] These four forces are hydrophobics, electrostatics interaction, hydrogen bonding, and van der Waals.[37, 39-46]

Protein residues can be divided into two groups, polar and non-polar, depending on their side chains. When a protein folds, most of the non-polar residues are buried inside and form a hydrophobic core, while polar residues are mostly exposed to solvent. This phenomenon is entropically favored and therefore leads to the increased stability of

proteins.[37, 47, 48] The hydrophobic effect was first described by Kauzmman in 1959.
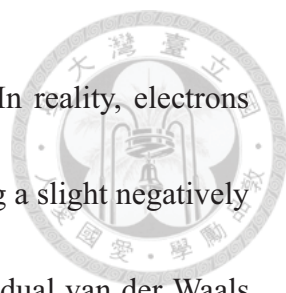
Polar residues are mostly charged and free to interact with their environment, including solvent molecules and other polar functional groups. Electrostatic interactions can be divided into three types: ion-ion, ion-dipole, and dipole-dipole.[41, 49] A charged side chain can interact with an oppositely charged functional group located on another residue or the protein terminus. Dipoles are formed by the asymmetric distribution of electrons due to the differences in electronegativity of the two atoms in a covalent bond. Electrostatic interactions through ionic charges or dipoles contribute to protein stability and the formation of protein structures.[50, 51]

A hydrogen bond is an interaction between a hydrogen atom in an X-H group and a highly electronegative atom Y such as nitrogen, oxygen, or fluorine.[40, 52] The partial positive charge on the H atom interacts with the partial negative charge on the Y atom.[40, 52] Such an interaction is important for stabilizing secondary and tertiary structures.[44, 53, 54] The backbone hydrogen bond $C=O\cdots H-N$ is the most prevalent (68.1%), with $C=O\cdots$ side chain (10.9%), $N-H\cdots$ side chain (10.4%), and side chain$\cdots$side chain hydrogen bond (10.6%) account for the remainder of the hydrogen bonds in protein structures.[44]
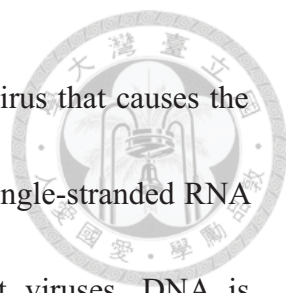
Another intermolecular interaction is van der Waals. Van der Waals force is a dispersion force caused by the fluctuating polarization of the nearby entities.[55] In a

symmetrical molecule, there is no charge distribution on average. In reality, electrons

are mobile and might more towards one end of the molecule, forming a slight negatively

charged end (δ-) and a slightly positively charged end (δ+).[55] Individual van der Waals

interactions are very weak, yet a massive number of such weak forces can still

significantly influence protein structure and stability.[56]

1-6 RNA Recognition
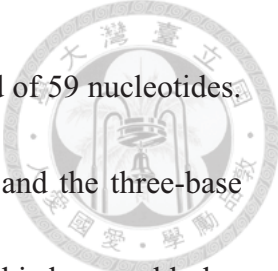
RNA-protein interactions are important in various fundamental biological processes,

including transcription, translation,[57] RNA processing and modification.[58] Both double

helical RNA and DNA are constructed by multiple complementary base pair such as

A-U, C-G and A-T, C-G.[59] There are three factors that control the binding affinity

between RNA and protein: electrostatic interaction between the protein positively

charged region and the negatively charged phosphate groups on the RNA backbone,

hydrogen bonding, and the interactions between the RNA groove and the protein side

chains. Specific proteins bind to specific sites on specific RNAs. The appropriate

binding of such proteins acts as a switch for RNA activation or repression. Therefore,

studies on RNA-protein recognition are important for understanding many diseases

related to RNA.

Human immunodeficiency virus (HIV) is a type of RNA retrovirus that causes the acquired immune deficiency syndrome (AIDS).[60] A retrovirus is a single-stranded RNA virus that targets a host cell as an obligate parasite.[61] In most viruses, DNA is transcribed into RNA, and RNA is translated into viral protein. In retroviruses, however, RNA is reverse-transcribed into DNA by a virally encoded reverse transcriptase, and then integrated into the genome of the host cell by a virally encoded integrase.[62] Most retroviruses contain three common genes in RNA genomes: *gag, pol,* and *env*. These genes contain the information necessary for building the structural proteins and important enzymes for new virus particles. The *gag* and *env* genes code for the core nucleocapsid polypeptides and surface-coat proteins of the virus, respectively.[63] The *pol* gene code for the viral reverse transcriptase and other enzymes.[64] In the HIV-1 viral RNA genome, there are six additional regulatory genes (*tat, rev, nef, vif, vpr*, and *vpu*) that code for proteins that control the infection by HIV and the production of new viral particles.[64] The *tat* gene encodes for the Tat protein, which serves as a transcriptional trans-activator by binding TAR RNA. The Tat protein is important for HIV-1 replication.

Trans-activator of transcription (Tat) protein contains a basic region that can recognize RNA: RKKRRQRRR (residue 49 to 57). The Tat protein targets the trans-activating responsive element (TAR) RNA located at the 5'end of nascent HIV-1

transcripts.[65] The TAR RNA contains a stem-loop structure composed of 59 nucleotides.

Two essential regions are the pentanucleotide loop ($^{+29}$CUGGG^{+33}) and the three-base

bulge ($^{+22}$UCU^{+24}) at the sites from +17 to +45. By interacting with this loop and bulge

region, Tat proteins alters the properties of the transcriptional complex and recruits

crucial enzymes, including the positive transcription elongation complex and RNA

polymerase II, for efficient production of full-length viral RNA.[66] The Tat-TAR binding

provides a positive feedback cycle and allows HIV to have an explosive response once

the threshold amount of Tat protein is reached.[67] Blocking this protein-RNA interaction

may repress the transcription of HIV-1 and serve as a potential treatment towards

AIDS.[68]

1-7 Post-Translational Modifications (PTMs)

Proteins are synthesized through the following biological steps: translation,

polymerization, termination, and processing.[69] There are only 20 amino acids encoded

by the triple nucleotide codons in mRNA. However, there are about 140 amino acids

derivatives that have been identified in different proteins.[70] These 20 encoded amino

acids must undergo various modifications to increase or even alter their functionalities.

Any modification that occurs after the completion of translation is considered a

post-translational modification (PTM).[70]

PTMs are a series of covalent processing events including peptide bond cleavage and functional group attachment onto individual amino acids. Some common PTMs are phosphorylation,[71] acetylation,[72] glycosylation,[73] acylation,[74] and methylation.[75] PTMs are responsible for protein function regulation and structural change.[76]

Protein methylation is a common post-translational modification that affects thermal stability,[77] cellular stress response,[78] protein aging,[79] ,gene regulation,[80-82] and transcriptional regulation.[83] Protein methylation typically takes place on arginine (Arg) or lysine (Lys) residues in the protein sequence.[75] Lysine can be methylated once, twice, or three times by lysine methyltransferases into monomethyllysine (Mmk), dimethyllysine (Dmk), and trimethyllysine (Tmk), respectively.[84] Lysine methylation leads to the increase of the positive charge effective radius and hydrophobicity. Such methylated lysines play an important role in protein-protein and protein-nucleic acid regulation.[85, 86] For the Tat protein, several post-translational modifications have been identified that modulate the interactions of Tat with TAR and other essential enzyme complexes.[87] These modifcations include lysine methylation at the residue adjacent to the basic region.[87] Accordingly, in this thesis, we investigate the effect of various types of lysine methylation on TAR RNA recognition by Tat_{47-57} derivatives.

1-8 Thesis Overview

Post-translational modifications are responsible for many protein behaviors. Lysine methylation alters the physiological properties of the residue and may impact both protein function and structure. There are three variations of methylated lysines that are identified in proteins. It is logical to assume that the different numbers of methyl groups attached on the side chain amino group should have different effects on proteins. In this study, various types of methylated lysines are placed into two basic secondary structures: α-helix and the simplest β-sheet model, "β-hairpin", to investigate the effect of lysine methylation on structural stability (Chapter 2).
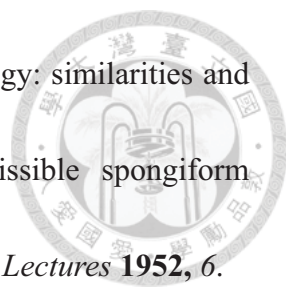
Lysine methylation also plays an important part in biological processes.[75] The regulatory Tat protein contains a basic region (RKKRRQRRR, residue 49 to 57), which can specifically bind to the trans-activating responsive (TAR) element to modulate transcription.[88] The binding between HIV-1 Tat protein and TAR RNA is essential for the HIV-1 virus to efficiently produce viral RNA.[88] To study the effect of lysine methylation on RNA recognition and cellular uptake, two lysine residues Lys_{50} and Lys_{51} were replaced with monothylated, dimethylated, and trimethylated lysines individualing in Chapter 3. The dissociation constant for the Tat derived peptide-TAR RNA complexes was determined by gel shift assay. The cellular uptake efficiency of Tat
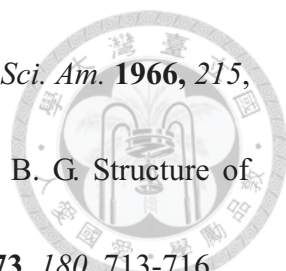
derived peptides into Jurkat cell was assessed by flow cytometry.

1-9 References

1. Crick, F. Central dogma of molecular biology. *Nature* **1970,** *227*, 561-563.
2. Pauling, L.; Corey, R. B. The Planarity of the Amide Group in Polypeptides. *J. Am. Chem. Soc.* **1952,** *74*, 3964-3964.
3. Hall, A. Rho GTPases and the actin cytoskeleton. *Science* **1998,** *279*, 509-514.
4. Nishizuka, Y. The Role of Protein Kinase-C in Cell-Surface Signal Transduction and Tumor Promotion. *Nature* **1984,** *308*, 693-698.
5. Radzicka, A.; Wolfenden, R. A Proficient Enzyme. *Science* **1995,** *267*, 90-93.
6. Aderem, A.; Ulevitch, R. J. Toll-like receptors in the induction of the innate immune response. *Nature* **2000,** *406*, 782-787.
7. Gavin, A. C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Jensen, L. J.; Bastuck, S.; Dumpelfeld, B.; Edelmann, A.; Heurtier, M. A.; Hoffman, V.; Hoefert, C.; Klein, K.; Hudak, M.; Michon, A. M.; Schelder, M.; Schirle, M.; Remor, M.; Rudi, T.; Hooper, S.; Bauer, A.; Bouwmeester, T.; Casari, G.; Drewes, G.; Neubauer, G.; Rick, J. M.; Kuster, B.; Bork, P.; Russell, R. B.; Superti-Furga, G. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **2006,** *440*, 631-636.
8. Wierenga, R. K. The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett.* **2001,** *492*, 193-198.
9. Georges, J. Alzheimer's disease in real life. *Eur. J. Neurol.* **2005,** *12*, 328-328.
10. Lee, J. C.; Gray, H. B.; Winkler, J. R. Copper(II) binding to α-synuclein, the Parkinson's protein. *J. Am. Chem. Soc.* **2008,** *130*, 6898-6899.
11. Bates, G. P. Huntington's disease - Exploiting expression. *Nature* **2001,** *413*, 691-694.
12. Prusiner, S. B.; Groth, D.; Serban, A.; Stahl, N.; Gabizon, R. Attempts to Restore Scrapie Prion Infectivity after Exposure to Protein Denaturants. *Proc. Natl. Acad. Sci. U. S. A.* **1993,** *90*, 2793-2797.
13. Ferri, C. P.; Prince, M.; Brayne, C.; Brodaty, H.; Fratiglioni, L.; Ganguli, M.; Hall, K.; Hasegawa, K.; Hendrie, H.; Huang, Y. Q.; Jorm, A.; Mathers, C.; Menezes, P. R.; Rimmer, E.; Scazufca, M.; Intl, A. D. Global prevalence of dementia: a Delphi consensus study. *Lancet* **2005,** *366*, 2112-2117.
14. Ballard, C.; Gauthier, S.; Corbett, A.; Brayne, C.; Aarsland, D.; Jones, E. Alzheimer's disease. *Lancet* **2011,** *377*, 1019-1031.
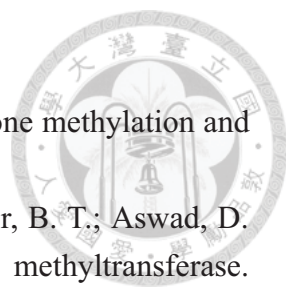
15. Kahle, P. J. α-synucleinopathy models and human neuropathology: similarities and differences. *Acta Neuropathol.* **2008,** *115*, 87-95.

16. Caughey, B.; Chesebro, B. Prion protein and the transmissible spongiform encephalopathies. *Trends. Cell Biol.* **1997,** *7*, 56-62.

17. Linderstrøm-Lang, K. U. Proteins and Enzymes. *Lane. Medical. Lectures* **1952,** *6*.

18. Invernizzi, G.; Papaleo, E.; Sabate, R.; Ventura, S. Protein aggregation: Mechanisms and functional consequences. *Int. J. Biochem. Cell B* **2012,** *44*, 1541-1554.

19. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963,** *7*, 95-99.

20. Richardson, J. S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **1981,** *34*, 167-339.

21. Pauling, L.; Corey, R. B.; Branson, H. R. The Structure of Proteins - 2 Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc. Natl. Acad. Sci. U. S. A.* **1951,** *37*, 205-211.

22. Arnott, S.; Wonacott, A. J. Atomic co-ordinates for an α-helix: refinement of the crystal structure of α-poly-l-alanine. *J. Mol. Biol.* **1966,** *21*, 371-383.

23. Barlow, D. J.; Thornton, J. M. Helix geometry in proteins. *J. Mol. Biol.* **1988,** *201*, 601-619.

24. Pauling, L.; Corey, R. B. The structure of synthetic polypeptides. *Proc. Natl. Acad. Sci. U. S. A.* **1951,** *37*, 241-250.

25. Cheng, R. P.; Girinath, P.; Suzuki, Y.; Kuo, H. T.; Hsu, H. C.; Wang, W. R.; Yang, P. A.; Gullickson, D.; Wu, C. H.; Koyack, M. J.; Chiu, H. P.; Weng, Y. J.; Hart, P.; Kokona, B.; Fairman, R.; Lin, T. E.; Barrett, O. Positional Effects on Helical Ala-Based Peptides. *Biochemistry* **2010,** *49*, 9372-9384.

26. Pauling, L.; Corey, R. B. The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl. Acad. Sci. U. S. A.* **1951,** *37*, 251-256.

27. Sibanda, B. L.; Thornton, J. M. β-hairpin families in globular proteins. *Nature* **1985,** *316*, 170-174.

28. Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. Conformation of β-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J. Mol. Biol.* **1989,** *206*, 759-777.

29. Sibanda, B. L.; Thornton, J. M. Conformation of β hairpins in protein structures: classification and diversity in homologous structures. *Methods Enzymol.* **1991,** *202*, 59-82.

30. Janin, J.; Chothia, C. Domains in proteins: definitions, location, and structural principles. *Methods Enzymol.* **1985,** *115*, 420-430.

31. Phillips, D. C. 3-Dimensional Structure of an Enzyme Molecule. *Sci. Am.* **1966,** *215*, 78-90.

32. Drenth, J.; Jansoniu.Jn; Koekoek, R.; Swen, H. M.; Wolthers, B. G. Structure of Papain. *Nature* **1968,** *218*, 929-932.

33. Porter, R. R. Structural Studies of Immunoglobulins. *Science* **1973,** *180*, 713-716.

34. Edelman, G. M. Antibody Structure and Molecular Immunology. *Science* **1973,** *180*, 830-840.

35. Levitt, M.; Chothia, C. Structural Patterns in Globular Proteins. *Nature* **1976,** *261*, 552-558.

36. Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **1995,** *247*, 536-540.

37. Pace, C. N.; Shirley, B. A.; McNutt, M.; Gajiwala, K. Forces contributing to the conformational stability of proteins. *FASEB J.* **1996,** *10*, 75-83.

38. Klotz, I. M.; Langerman, N. R.; Darnall, D. W. Quaternary structure of proteins. *Annu. Rev. Biochem.* **1970,** *39*, 25-62.

39. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990,** *29*, 7133-7155.

40. Hagler, A. T.; Huler, E.; Lifson, S. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *J. Am. Chem. Soc.* **1974,** *96*, 5319-5327.

41. Perutz, M. F. Electrostatic effects in proteins. *Science* **1978,** *201*, 1187-1191.

42. Barlow, D. J.; Thornton, J. M. Ion-pairs in proteins. *J. Mol. Biol.* **1983,** *168*, 867-885.

43. Nicholls, A.; Sharp, K. A.; Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **1991,** *11*, 281-296.

44. Stickle, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D. Hydrogen bonding in globular proteins. *J. Mol. Biol.* **1992,** *226*, 1143-1159.

45. Pace, C. N.; Grimsley, G. R.; Scholtz, J. M. Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J. Biol. Chem.* **2009,** *284*, 13285-13289.

46. Stigter, D.; Dill, K. A. Charge effects on folded and unfolded proteins. *Biochemistry* **1990,** *29*, 1262-1271.

47. Pace, C. N. Contribution of the hydrophobic effect to globular protein stability. *J. Mol. Biol.* **1992,** *226*, 29-35.

48. Pace, C. N.; Fu, H.; Fryar, K. L.; Landua, J.; Trevino, S. R.; Shirley, B. A.; Hendricks, M. M.; Iimura, S.; Gajiwala, K.; Scholtz, J. M.; Grimsley, G. R. Contribution of hydrophobic interactions to protein stability. *J. Mol. Biol.* **2011,** *408*,

514-528.

49. Yoder, C. H. Teaching ion-ion, ion-dipole, and dipole-dipole interactions. *J. Chem. Educ.* **1977,** *54*, 402-408.

50. Wada, A.; Nakamura, H. Nature of the charge distribution in proteins. *Nature* **1981,** *293*, 757-758.

51. Hol, W. G.; Halie, L. M.; Sander, C. Dipoles of the α-helix and β-sheet: their role in protein folding. *Nature* **1981,** *294*, 532-536.

52. Hagler, A. T.; Lifson, S. Energy functions for peptides and proteins. II. The amide hydrogen bond and calculation of amide crystal properties. *J. Am. Chem. Soc.* **1974,** *96*, 5327-5335.

53. Baker, E. N.; Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **1984,** *44*, 97-179.

54. McDonald, I. K.; Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994,** *238*, 777-793.

55. Feinberg, G.; Sucher, J. General Theory of the van der Waals Interaction: A Model-Independent Approach. *Phys. Rev. A* **1970,** *2*, 2395-2415.

56. Levitt, M.; Gerstein, M.; Huang, E.; Subbiah, S.; Tsai, J. Protein folding: the endgame. *Annu. Rev. Biochem.* **1997,** *66*, 549-579.

57. Matsuo, H.; Li, H.; McGuire, A. M.; Fletcher, C. M.; Gingras, A. C.; Sonenberg, N.; Wagner, G. Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. *Nat. Struct. Biol.* **1997,** *4*, 717-724.

58. Varani, G.; Nagai, K. RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.* **1998,** *27*, 407-445.

59. Seeman, N. C.; Rosenberg, J. M.; Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1976,** *73*, 804-808.

60. Weiss, R. A. How does HIV cause AIDS? *Science* **1993,** *260*, 1273-1279.

61. Yoshida, M. Discovery of HTLV-1, the first human retrovirus, its unique regulatory mechanisms, and insights into pathogenesis. *Oncogene* **2005,** *24*, 5931-5937.

62. Smith, J. A.; Daniel, R. Following the path of the virus: the exploitation of host DNA repair mechanisms by retroviruses. *ACS Chem. Biol.* **2006,** *1*, 217-226.

63. King, S. R. HIV - Virology and Mechanisms of Disease. *Ann. Emerg. Med.* **1994,** *24*, 443-449.

64. Greene, W. C. The molecular biology of human immunodeficiency virus type 1 infection. *N. Engl. J. Med.* **1991,** *324*, 308-317.

65. Weeks, K. M.; Ampe, C.; Schultz, S. C.; Steitz, T. A.; Crothers, D. M. Fragments of the HIV-1 Tat protein specifically bind TAR RNA. *Science* **1990,** *249*, 1281-1285.

66. Mujeeb, A.; Bishop, K.; Peterlin, B. M.; Turck, C.; Parslow, T. G.; James, T. L. NMR Structure of a Biologically-Active Peptide-Containing the RNA-Binding

Domain of Human-Immunodeficiency-Virus Type-1 Tat. *Proc. Natl. Acad. Sci. U. S. A.* **1994,** *91*, 8248-8252.

67. Cullen, B. R. Regulation of HIV-1 Gene-Expression. *FASEB J.* **1991,** *5*, 2361-2368.

68. Stevens, M.; De Clercq, E.; Balzarini, J. The regulation of HIV-1 transcription: Molecular targets for chemotherapeutic intervention. *Med. Res. Rev.* **2006,** *26*, 595-625.

69. Merrick, W. C. Mechanism and regulation of eukaryotic protein synthesis. *Microbiol. Rev.* **1992,** *56*, 291-315.

70. Uy, R.; Wold, F. Post-translational covalent modification of proteins. *Science* **1977,** *198*, 890-896.

71. Lipman, F. A.; Levene, P. A. Serinephosphoric acid obtained on hydrolysis of vitellinic acid. *J. Biol. Chem.* **1932,** *98*, 109-114.

72. Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **2009,** *325*, 834-840.

73. Moremen, K. W.; Tiemeyer, M.; Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell. Bio.* **2012,** *13*, 448-462.

74. Towler, D. A.; Gordon, J. I.; Adams, S. P.; Glaser, L. The Biology and Enzymology of Eukaryotic Protein Acylation. *Annu. Rev. Biochem.* **1988,** *57*, 69-99.

75. Paik, W. K.; Kim, S. Protein Methylation. *Science* **1971,** *174*, 114-119.

76. Seo, J.; Lee, K. J. Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *J. Biochem. Mol. Biol.* **2004,** *37*, 35-44.

77. Febbraio, F.; Andolfo, A.; Tanfani, F.; Briante, R.; Gentile, F.; Formisano, S.; Vaccaro, C.; Scire, A.; Bertoli, E.; Pucci, P.; Nucci, R. Thermal stability and aggregation of sulfolobus solfataricus β-glycosidase are dependent upon the N-epsilon-methylation of specific lysyl residues: critical role of in vivo post-translational modifications. *J. Biol. Chem.* **2004,** *279*, 10185-10194.

78. Desrosiers, R.; Tanguay, R. M. Methylation of Drosophila histones at proline, lysine, and arginine residues during heat shock. *J. Biol. Chem.* **1988,** *263*, 4686-4692.

79. Najbauer, J.; Orpiszewski, J.; Aswad, D. W. Molecular aging of tubulin: accumulation of isoaspartyl sites in vitro and in vivo. *Biochemistry* **1996,** *35*, 5183-5190.

80. Kramer Jamie, M. Epigenetic regulation of memory: implications in human cognitive disorders. *BioMol. Concepts* **2013,** *4*, 1-12.

81. Nakayama, J.; Rice, J. C.; Strahl, B. D.; Allis, C. D.; Grewal, S. I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science*

2001, *292*, 110-113.

82. Grewal, S. I.; Rice, J. C. Regulation of heterochromatin by histone methylation and small RNAs. *Curr. Opin. Cell. Biol.* **2004,** *16*, 230-238.

83. Chen, D.; Ma, H.; Hong, H.; Koh, S. S.; Huang, S. M.; Schurter, B. T.; Aswad, D. W.; Stallcup, M. R. Regulation of transcription by a protein methyltransferase. *Science* **1999,** *284*, 2174-2177.

84. Paik, W. K.; Paik, D. C.; Kim, S. Historical review: the field of protein methylation. *Trends. Biochem. Sci.* **2007,** *32*, 146-152.

85. Martin, C.; Zhang, Y. The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Bio.* **2005,** *6*, 838-849.

86. Zhang, X.; Wen, H.; Shi, X. B. Lysine methylation: beyond histones. *Acta Bioch. Bioph. Sin.* **2012,** *44*, 14-27.

87. Hetzer, C.; Dormeyer, W.; Schnolzer, M.; Ott, M. Decoding Tat: the biology of HIV Tat posttranslational modifications. *Microbes Infect.* **2005,** *7*, 1364-1369.

88. Debaisieux, S.; Rayne, F.; Yezid, H.; Beaumelle, B. The Ins and Outs of HIV-1 Tat. *Traffic* **2012,** *13*, 355-363.

Chapter2

Effect of Lysine Methylation on α-Helix and β-Sheet Propensity

Chapter 2.

2-1 Introduction

α-Helix

The most abundant secondary structure in proteins is the α-helix, which is adopted by nearly one third of all protein residues.[1] The α-helix is characterized by consecutive, main-chain, $i{\leftarrow}i{+}4$ hydrogen bonds between each carbonyl oxygen (i) and an amide hydrogen ($i{+}4$) on the adjacent helical turn.[2] α-Helix stability is determined by N- and C-capping effects, side chain-helix macrodipole interactions, side chain-side chain interactions, and the intrinsic structure forming tendencies of the constituting amino acids.[3] Relative occurring frequencies of each amino acid adopting different secondary structures were analyzed by Chou and Fasman.[4] These statistic results revealed that each amino acid has its own propensity for different secondary structures.[4] The thermodynamic helix propensities were determined by Baldwin and co-coworkers in alanine-based peptides with minimum side chain interaction based on circular dichroism spectra using modified Lifson-Roig theory.[5, 6] The thermodynamic propensies of the amino acids were converted to the free energy of helix formation/propagation. Also, the nucleation of a helix formation is thought to be more difficult than propagation, therefore capping effects were considered for helix formation.[6] As such, the Lifson-Roig

theory was modified by Baldwin and coworkers to incorporate N- and C-capping parameters.[5-7] The basic assumption is that helix propensity of each amino acid is position independent.[8]
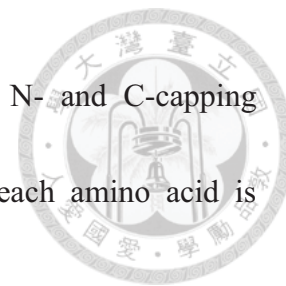
Lifson-Roig Theory

Statistical mechanical models had been used to describe the helix-coil equilibrium including Zimm-Bragg theory and Lifson-Roig theory.[9, 10] Both models assume that the helix-coil equilibrium of each residue is a two-state equilibrium. Residues can only adopt either helix (*h*) or coil (*c*) conformation. Zimm-Bragg theory introduced two parameters to describe the helix-coil equilibrium of each residue: nucleation (σ) and propagation (*s*).[9] The statistical weight for initiating a helical unit (*hc* or *ch*) is defined as σs. A statistical weight of two successive coil residues (*cc*) is unity (set to 1). The statistical weight of two successive helical residues (*hh*) is σs^2. Through the Zimm-Bragg model, the helicity of a peptide can be deduced by a partition function with a few parameters. However, the original Zimm-Bragg model neglected many other factors that affect helicity such as N- and C-capping, electrostatics, and macrodipole.[11] Modified Zimm-Bragg theory with additional parameters to include other important interactions had been proposed.[12, 13]

The Lifson-Roig theory, similar to Zimm-Bragg theory, employs two parameters
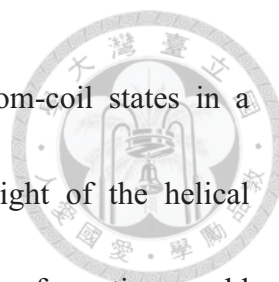
(w and v) to describe the equilibrium between α-helix and random-coil states in a statistical mechanical manner.[10] However, only the statistical weight of the helical conformation with at least three continuous residues adopting helix conformation would be considered a helix. This is due to the fact that a helix cannot exist without the stabilization of (i, i+4) hydrogen bonding. The statistical weight of each state of each residue is based on the residue's own state and the state of the two neighboring residues.[10] The Lifson-Roig model utilizes a 4x4 transfer matrix to describe the statistic weight of a residue, while Zimm-Bragg model uses a simpler 2x2 transfer matrix.

	hh	hc	ch	cc
hh	w	v	0	0
hc	0	0	1	1
ch	v	v	0	0
cc	0	0	1	1

The parameters were introduced to describe the different structural state of a residue: u, coil state; v, helix state adjacent to coil state; w, helix state located between helix states. The parameter v describes helix initiation,[10] which is an independent and energetically uphill process in helix formation. Baldwin and co-workers later discovered the different capping effects for N- and C- capping, thereby introducing two additional parameters to describe the N- and C-capping (n and c).[5] The modified 4x4 transfer matrix was:

	h<u>h</u>	h<u>c</u>	c<u>h</u>	c<u>c</u>
h<u>h</u>	w	v	0	0
h<u>c</u>	0	0	$\sqrt{nc}$	c
c<u>h</u>	v	v	0	0
c<u>c</u>	0	0	n	1
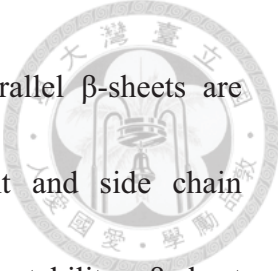
The fraction helix of a peptide can be described by the modified Lifson-Roig theory using the following equation:[5, 10, 14, 15]

$$
f_{helix} = \frac{1}{(N-2)} \frac{\displaystyle\sum_{i=1}^{N} \left[\begin{pmatrix} 0 & 0 & n & 1 \end{pmatrix} \prod_{j=1}^{j=i-1} \begin{bmatrix} w_j & v_j & 0 & 0 \\ 0 & 0 & \sqrt{n_j c_j} & c_j \\ v_j & v_j & 0 & 0 \\ 0 & 0 & n_j & 1 \end{bmatrix} \begin{bmatrix} w_i & v_i & 0 & 0 \\ 0 & 0 & 0 & 0 \\ v_i & v_i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \prod_{k=i+1}^{N} \begin{bmatrix} w_k & v_k & 0 & 0 \\ 0 & 0 & \sqrt{n_k c_k} & c_k \\ v_m & v_m & 0 & 0 \\ 0 & 0 & n_k & 1 \end{bmatrix} \begin{bmatrix} 0 \\ c \\ 0 \\ 1 \end{bmatrix} \right]}{\begin{pmatrix} 0 & 0 & n & 1 \end{pmatrix} \prod_{m=1}^{N} \begin{bmatrix} w_m & v_m & 0 & 0 \\ 0 & 0 & \sqrt{n_m c_m} & c_j \\ v_m & v_m & 0 & 0 \\ 0 & 0 & n_m & 1 \end{bmatrix} \begin{bmatrix} 0 \\ c \\ 0 \\ 1 \end{bmatrix}}
$$

β-Sheet

Another major secondary structure is the β-sheet. The relative flat structure consists of several β-strands. Neighboring strands are connected through hydrogen bonds from the N-H group of one strand to the C=O group of another strand.[16] There are two types of β-sheets according to the intrastrand hydrogen bonding pattern and the relative polypeptide orientation (N-terminus and C-terminus): parallel and anti-parallel β-sheet. Parallel β-sheets are characterized by a series of twelve-membered hydrogen bonded rings, while anti-parallel β-sheets are characterized by an alternating series of

ten- and fourteen-membered hydrogen bonded rings. The anti-parallel β-sheets are slightly more stable due to the hydrogen bonding arrangement and side chain orientation.[17] There are two main factors involved in β-sheet stability: β-sheet propensity of the constituting amino acids,[18-21] and cross-strand interactions.[22-25] The β-sheet propensity reflects the tendency of forming β-sheet structure for an amino acid. The difference between β-sheet and α-helix propensity is that sheet propensity is highly context dependent.[20] There are two distinct tertiary contexts in β-sheets: central strands and edge strands. There is no correlation for the thermodynamic of sheet formation at edge positions with statistically determined β-sheet propensity.[20] The β-sheet propensity has been determined by statistically examining both protein structure database and experimentally-based host-guest studies.[4, 19-21] Chou and Fasman first introduced this concept by applying conformational parameters for the 20 naturally occurring amino acids to compute the frequency of occurrence of each amino acid residue in different conformations from 15 proteins.[4] The results from several studies revealed that β-branched and aromatic amino acids are intrinsically favored for β-sheet formation.[4]

Instead of investigating a β-sheet, a β-hairpin is a more convenient model to examine the β-sheet stabilizing interactions because it is a reasonable minimalist model for β-sheets.[26] The β-hairpin is one of the simplest super-secondary structures composed of only two anti-parallel β-strands and a short loop region.[27] β-Hairpins also play

important roles in many proteins, including zinc-finger, and most likely HIV Tat. Due to the simpler structure compared to β-sheet, using β-hairpin as a model can exclude tertiary context effects and analyze intrinsic secondary structure propensity exclusively.
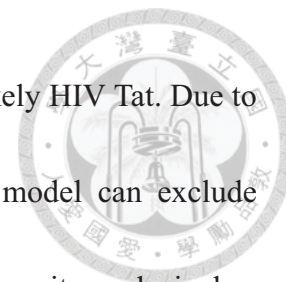
Lysine Methylation

Lysine is one of the most essential amino acids and plays important roles in living organisms.[28] Lysine is also a common substrate of methylation, which serves as a crucial regulatory process in various bioactivities.[29, 30] Lysine methylation has been mostly identified in the histone family and affects the epigenetic process through altering the binding of the surrounding DNA to histones.[31, 32] Many other non-histone proteins were later reported to have lysine methylation as well, indicating the importance of this post-translational modification in various protein functions.[33] Lysine methylation is achieved by lysine methyltransferases (PKMT), which employs S-adenosyl methionine as the methyl transferring unit.[34] Each lysine can be methylated once, twice, or three times, depending on which type of PKMTs are involved and can be distinguished by mass spectroscopy (Chart 2-1). Methylation on the side chain amino group mainly increases the hydrophobicity of the residue while retaining its positive charge in the physiological environment. The larger methyl group also increases the effective radius of the charge and reduces the strong potential electrostatic attraction.

Therefore it is reasonable to assume that different numbers of methyl groups attached to the amino group should have different effects on both function and structure. Previous studies mostly focused on the identification of methylation sites and the roles in bioactivities. As a result, this chapter will focus on the effect of lysine methylation on secondary structure stabilities.

Chart 2-1. Chemical Structure, Full Name and 3-Letter Code of Methylated Lysines

L-monomethyllysine   L-dimethyllysine   L-trimethyllysine
Mmk   Dmk   Tmk
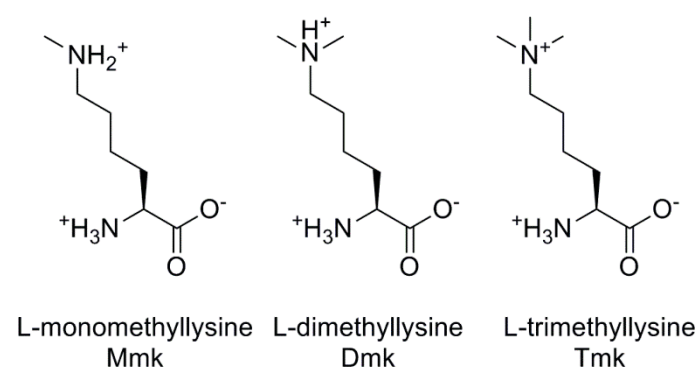
2-2 Results and Discussion

Peptide Design and Synthesis

For investigating helix propensity, a series of helical Ala-based peptides was designed based on the studies of Baldwin and co-workers (Table 2-1).[3, 5] Multiple Lys residues were evenly inserted to increase the aqueous solubility and minimize the aggregation of the peptides. Tyrosine (Tyr) was incorporated to facilitate peptide concentration determination by UV-vis.[35, 36] Glycine (Gly), a known "helix breaker", was used to separate Tyr from the helical part of the peptides to minimize the
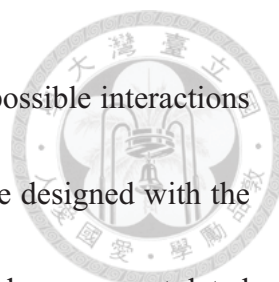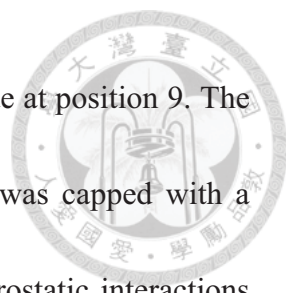
interference with CD measurements.[37] In order to exclude all other possible interactions

such as the interaction between Lys and modified Lys, peptides were designed with the

two residues separated by (i, $i+5$) spacing. The N-termini of all peptides were acetylated,

and the C-termini were designed to be a carboxyamide so there would be no bias

(created by charged termini) on the statistical mechanical capping parameters derived

for these residues. Each peptide contained one modified lysine monomethyllysine

(Mmk), dimethyllysine (Dmk), or trimethyllysine (Tmk) incorporated at position 9,

position 14, N-terminus, or at C-terminus to derive proper statistical mechanical

parameter w_9, w_{14}, n, or c, respectively, based on modified Lifson-Roig theory. [5, 10, 14, 15]

Table 2-1. Sequence of Ala-based Peptides for Determining the N-Cap Parameter, C-Cap Parameter, and Helix Propensity of Modified Lys Analogs

Peptide	Sequence
KXaa9	Ac-Tyr Gly Gly Lys Ala Ala Ala Ala **Xaa** Ala Ala Ala Ala Lys Ala Ala Ala Ala Lys-NH₂
KXaa14	Ac-Tyr Gly Gly Lys Ala Ala Ala Ala Lys Ala Ala Ala Ala **Xaa** Ala Ala Ala Ala Lys-NH₂
NCapXaa	Ac-**Xaa** Ala Ala Ala Ala Lys Ala Ala Ala Ala Lys Ala Ala Ala Ala Lys Gly Gly Tyr-NH₂
CCapXaa	Ac-Tyr Gly Gly Lys Ala Ala Ala Ala Lys Ala Ala Ala Ala Lys Ala Ala Ala Ala **Xaa**-NH₂

Xaa = monomethyllysine (Mmk), dimethyllysine (Dmk), trimethyllysine (Tmk)

For investigating sheet propensity, water-soluble monomeric β-hairpin peptides

were designed based on peptide YKL as described by Gellman (Table 2-2).[38] The

positively charged amino acids were incorporated to ensure water solubility and to

prevent aggregation.[22] In this study, the parent YKL peptide residue Tyr2 was replaced

with Thr to remove any possible diagonal interaction with the residue at position 9. The N-terminus was capped with an acetyl group and the C-terminus was capped with a carboxyamide to avoid unexpected and potentially interfering electrostatic interactions involving the backbone.[39] The guest site position 4 was a non-hydrogen bonding site near the center of the β-strands to avoid end fraying near the termini and excessive folding near the turn. Modified Lys was incorporated at position 4, and Ala was incorporated at position 9 to give the HPTXaaAla peptides (Figure 2-1). The peptides were named with the " HPT " prefix, representing **H**air**P**ins with **T**hr at position 2, followed by the three letter codes for the residues at positions 4 and 9. In order to determine percent folded for these peptides, the fully folded and fully unfolded reference peptides were necessary.[22, 23, 40, 41] Cysteines were introduced at both termini to give the fully folded reference peptides HPTFXaaAla through cyclized disulfide bond formation (Figure 2-1).[22, 23, 40, 41] Replacing ^{D}Pro6 with ^{L}Pro6 disfavored turn configuration for hairpin formation, gave the unfolded reference peptides HPTUXaaAla as negative controls (Figure 2-1).[22, 23, 42] The **F**olded and **U**nfolded reference peptides are named by adding the letters " F " and " U " after " HPT ", respectively.
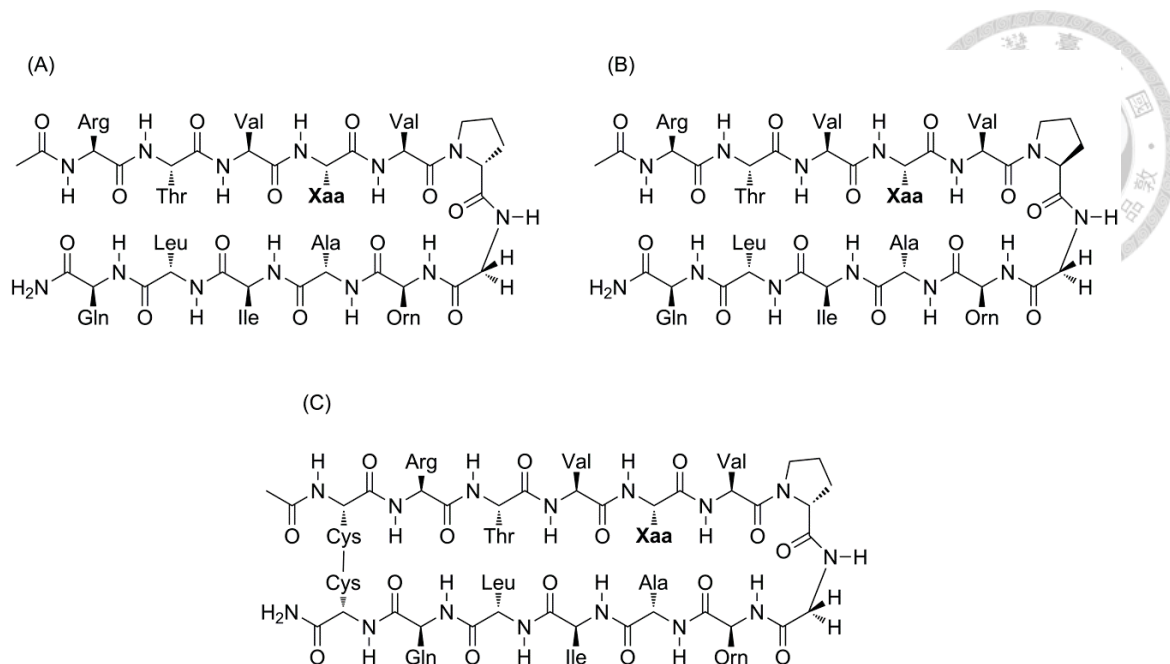
Figure 2-1. Chemical structure of the experimental (A), the fully unfolded (B), and the fully folded (C) hairpin peptides. Xaa was replaced with Mmk, Dmk, or Tmk

Table 2-2. Sequences for the Hairpin Peptides HPTXaaAla, the Unfolded Reference Peptides HPTUXaaAla, and the Folded Reference Peptides HPTFXaaAla Containing Modified Lys Analogs

Peptide	Sequence
HPTXaaAla	Ac-Arg Thr Val **Xaa** Val ^{D}Pro Gly Orn Ala Ile Leu Gln-NH$_2$
HPTUXaaAla	Ac-Arg Thr Val **Xaa** Val Pro Gly Orn Ala Ile Leu Gln-NH$_2$
HPTFXaaAla	Ac-Cys Arg Thr Val **Xaa** Val ^{D}Pro Gly Orn Ala Ile Leu Gln Cys-NH$_2$

Xaa= monomethyllysine (Mmk), dimethyllysine (Dmk), trimethyllysine (Tmk)

All the peptides were synthesized by solid phase peptide synthesis using Fmoc-based chemistry (Fmoc = 9-fluorenylmethoxycarbonyl).[43] The peptide was cleaved from the resin along with concomitant side chain deprotection using trifluoroacetic acid (TFA with 5% triisopropylsilane). Ethanedithiol was added to the cleavage of cysteine-containing folded reference peptides (TFA with 5%
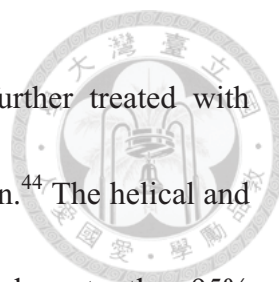
triisopropylsilane and ethanedithiol). The folded peptides were further treated with

charcoal-mediated air oxidation to perform intramolecular cyclization.[44] The helical and

hairpin peptides were purified by RP-HPLC to greater than 98.5% and greater than 95%

purity, respectively (Tables 2-3 and 2-4). All peptides were confirmed by MALDI-TOF.

Table 2-3. The Purity and Weight of the Helical Peptides
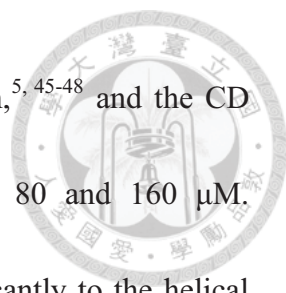
Peptide	Purity (%)	Weight (mg)
KMmk9	98.9	9.6
KDmk9	98.5	8.1
KTmk9	98.9	5.8
KMmk14	98.6	5.8
KDmk14	98.7	3.0
KTmk14	98.7	7.8
NCapMmk	98.6	8.9
NCapDmk	98.5	8.1
NCapTmk	99.1	13.2
CCapMmk	98.7	12.5
CCapDmk	98.6	1.5
CCapTmk	98.7	7.2

Table 2-4. The Purity and Weight of the Hairpin Peptides

Peptide	Purity (%)	Weight (mg)
HPTMmkAla	96.1	16.0
HPTDmkAla	95.6	5.8
HPTTmkAla	96.7	9.8
HPTUMmkAla	96.1	8.8
HPTUDmkAla	96.7	7.8
HPTUTmkAla	95.3	4.8
HPTFMmkAla	96.1	8.5
HPTFDmkAla	95.6	5.9
HPTFTmkAla	96.4	9.5

Circular Dichorism Spectroscopy

The concentration of the helical peptides was determined by the Edelhoch

method.[37] The circular dichroism (CD) spectrum for the helical peptides was acquired at

pH 7 and 273 K in the presence of 1 M NaCl to derive the fraction helix (f_{helix}) (Table

2-5). Each spectrum was acquired at least three times to confirm reproducibility.

Analogous peptides have been shown to be monomeric in solution,[5, 45-48] and the CD

spectrum of each peptide did not change significantly between 80 and 160 μM.

Therefore, intermolecular interactions should not contribute significantly to the helical

content of these peptides in solution. The mean residue ellipticity at 222 nm reflects the

helical content of a given peptide.[49] The helical content of KXaa9 peptides followed the

trend KMmk9 > KDmk9 > KTmk9 (Figure 2-2A). The helical content of KXaa14

peptides followed the trend KTmk14 > KDmk14 ~ KMmk14 (Figure 2-2B). The helical

content of the NCapXaa peptides followed the trend NCapDmk ~ NCapMmk ~

NCapTmk (Figure 2-2C), suggesting similar N-capping abilities among the Lys analogs.

The helical content of CCapXaa peptides followed the trend CCapTmk > CCapMmk >
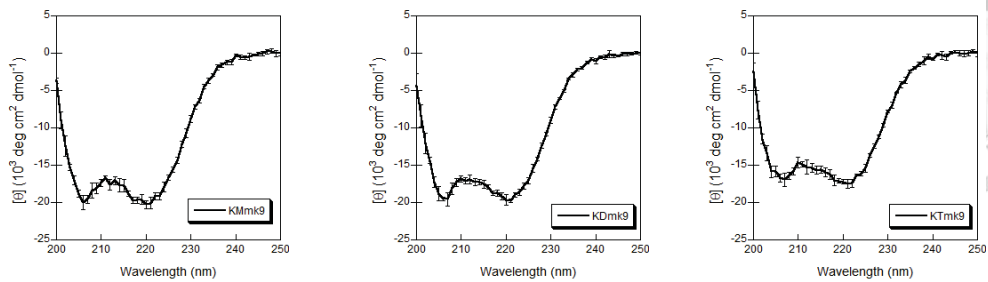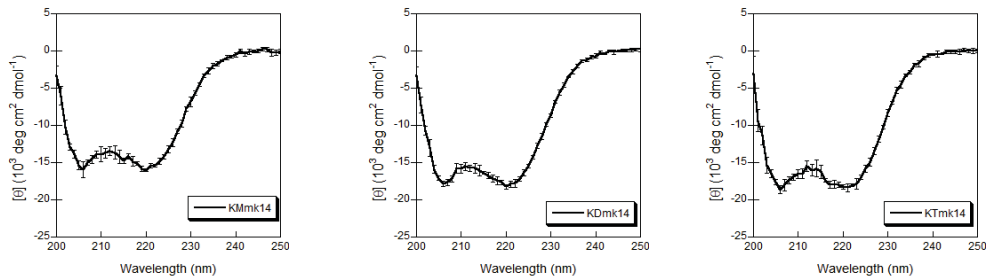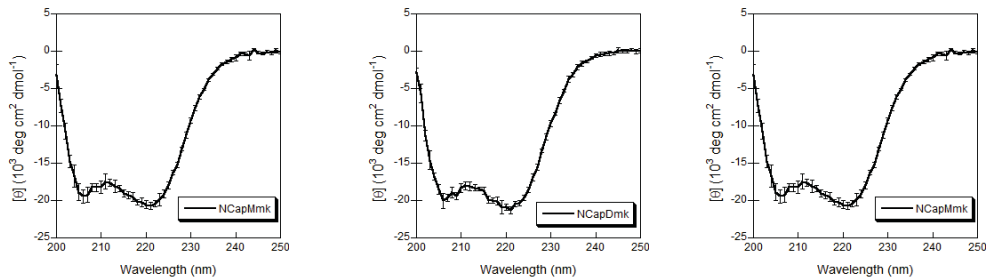
CCapDmk (Figure 2-2D).
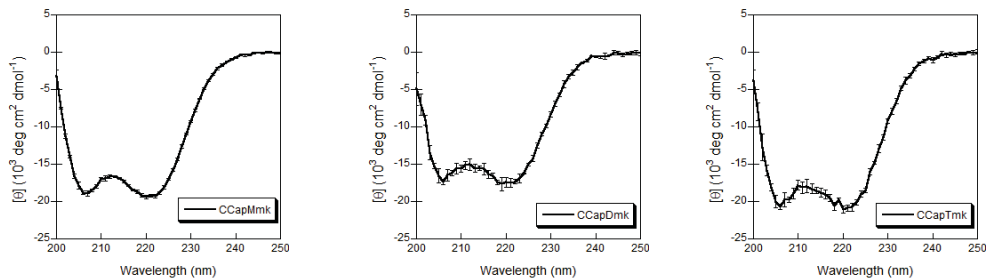
Figure 2-2. Circular dichorism spectra of the helical peptides at pH 7 (273 K) in 1 mM phosphate, borate, and citrate buffer with 1 M NaCl. (A) KXaa9 peptides, (B) KXaa14 peptides, (C) NCapXaa peptides, (D) CCapXaa peptides.

Table 2-5. Mean Residue Ellipticity at 222 nm and Fraction Helix (f_{helix}) of Helical Peptides Containing Methylated Lys

Peptide	$[\theta]_{222}$	f_{helix}
KMmk9	-19300 ± 300	0.553 ± 0.008
KDmk9	-19000 ± 400	0.543 ± 0.010
KTmk9	-17300 ± 300	0.494 ± 0.008
KMmk14	-17600 ± 500	0.504 ± 0.015
KDmk14	-17600 ± 400	0.503 ± 0.011
KTmk14	-18400 ± 500	0.525 ± 0.013
NCapMmk	-20500 ± 500	0.586 ± 0.014
NCapDmk	-20900 ± 400	0.596 ± 0.011
NCapTmk	-20000 ± 500	0.572 ± 0.013
CCapMmk	-19200 ± 200	0.548 ± 0.005
CCapDmk	-17200 ± 400	0.492 ± 0.012
CCapTmk	-20400 ± 400	0.582 ± 0.013

Helix Formation Parameters

The helix formation parameters including N-cap (n), C-cap (c), and helix propensity (w_9 and w_{14}) were derived from the CD data using modified Lifson–Roig theory (Table 2-6).[5, 10, 14, 15] The helix propensity w_9 followed the trend KMmk9 > KDmk9 > KTmk9. The helix propensity w_{14} followed the trend KTmk14 > KDmk14 ~ KMmk14. The difference between w_9 and w_{14} was possibly due to the N-fraying effect at position 14. The N-cap n parameter for all NCapXaa peptides was less than zero, suggesting extremely unfavorable energetics for placing positively charged residues at the N-terminus. However, values less than zero do not bear physical meaning, therefore the n values were set to zero. The C-cap c parameters of CCapXaa peptides followed the trend CCapTmk > CCapMmk > CCapDmk.

Table 2-6. Statistical Mechanical Helix Formation Parameters for Modified Lys Analogs Derived from Experimentally Measured Fraction Helix Based on Modified Lifson–Roig Theory

Residue	n	c	w_9	w_{14}
Mmk	0^a	2.385 ± 0.253	0.982 ± 0.059	0.687 ± 0.080
Dmk	0^a	0.254 ± 0.357	0.827 ± 0.061	0.642 ± 0.050
Tmk	0^a	4.585 ± 1.189	0.713 ± 0.040	0.859 ± 0.091

[a] The parameter initially converged to a negative probability (which carries no physical meaning), therefore the value was set to zero

Hairpin Structure Characterization

The peptides were analyzed by two-dimensional NMR spectroscopy including DQF-COSY,[50] TOCSY,[51] and ROESY.[52] Sequence specific assignments for all peptides were completed by using the aforementioned 2D-NMR spectra. Chemical shift dispersion correlates to the folded population; the higher the chemical shift dispersion, the higher the folded population.[53] The unfolded reference peptides HPTUXaaAla and disulfide-cyclized folded reference peptides HPTFXaaAla were assumed to represent the fully unfolded form and fully folded form of the experimental peptide HPTXaaAla, respectively.

Chemical shift deviation of the Hα signals ($\Delta\delta$ Hα), $^3J_{NH\alpha}$ spin−spin coupling constants, and cross-strand NOEs were used to confirm hairpin formation for the folded reference peptides and the experimental HPTXaaAla peptides. Chemical shift deviation of Hα ($\Delta\delta$ Hα) is defined as the difference between the observed chemical shift and the random coil value for the α-proton.[54] A positive deviation value is indicative of β-sheet
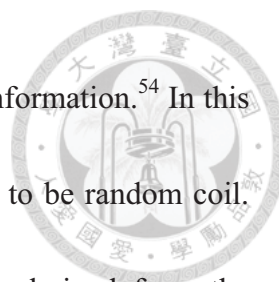
conformation, whereas a negative value is indicative of α-helical conformation.[54] In this study, the unfolded reference peptides HPTUXaaAla were assumed to be random coil. The Δδ Hα for the HPTXaaAla and HPTFXaaAla peptides were derived from the corresponding chemical shift values (Figures 2-3 and 2-4). The ratio of the chemical shift deviation for a given residue of the experimental peptide and the chemical shift deviation of the corresponding residue of the folded reference peptide gave the fraction folded for the residue of interest in the experimental peptide.

(A)   (B)   (C)
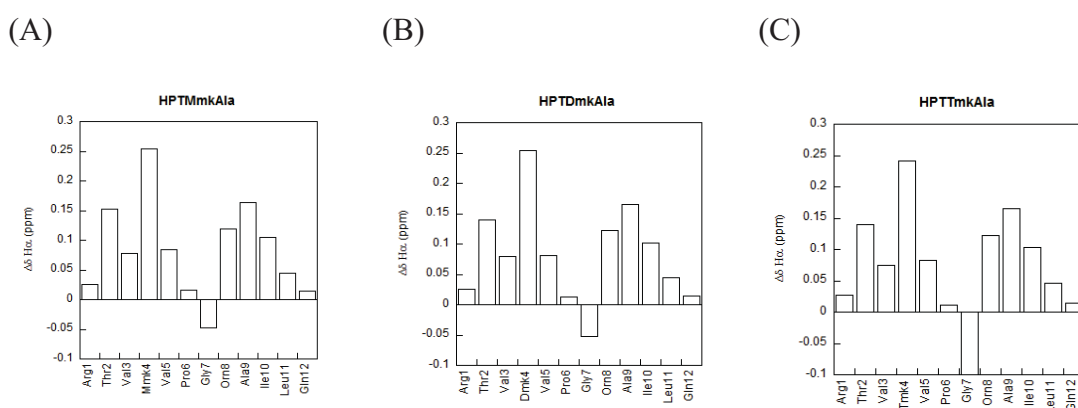
Figure 2-3. The Hα chemical shift deviation for peptides HPTMmkAla (A), HPTDmkAla (B), and HPTTmkAla (C). Reference fully unfolded peptides were HPTUMmkAla, HPTUDmkAa and HPTUTmkAla, respectively.

(A)   (B)   (C)
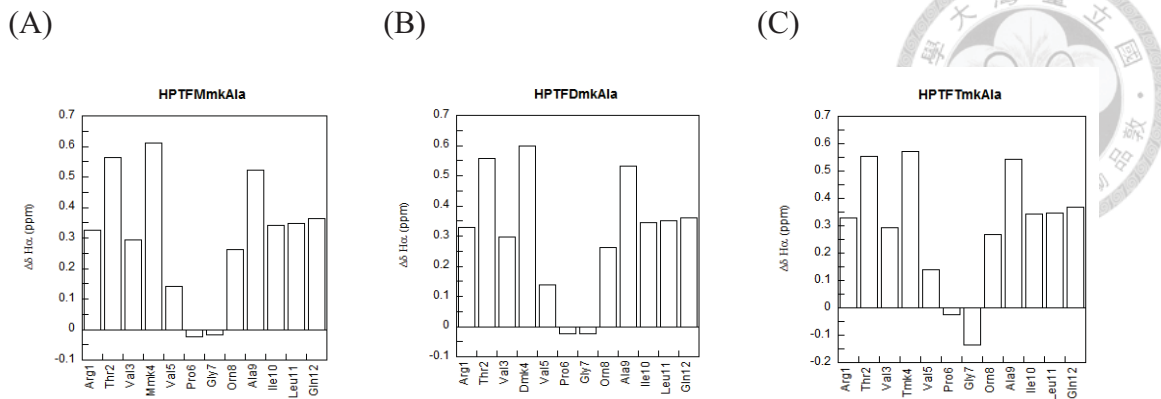
Figure 2-4. The Hα chemical shift deviation for peptides HPTFMmkAla (A) HPTFDmkAla, (B), and HPTFTmkAla (C). Reference fully unfolded peptides were HPTUMmkAla, HPTUDmkAa and HPTUTmkAla, respectively.

The three-bond coupling constant between the intra-residue alpha and amide protons ($^3J_{NH\alpha}$) is useful for secondary structure characterization because it is related directly to the dihedral angle φ of the backbone. The $^3J_{NH\alpha}$ coupling constants were determined from the absorptive (ν_a) and dispersive (ν_d) values in the DQF-COSY spectra. The dispersive spectrum was generated by a **90°** phase shift of the absorptive antiphase spectrum. The ν_a and ν_d were used to derive J from the following equation:[55]

$$J^6 - \nu_a^2 J^4 + \left(-\frac{9}{4}\nu_a^4 + \frac{3}{2}\nu_a^2\nu_d^2 + \frac{3}{4}\nu_d^4\right)J^2 + \frac{81}{64}\nu_a^6 - \frac{9}{16}\nu_a^4\nu_d^2 - \frac{21}{32}\nu_a^2\nu_d^4 - \frac{1}{16}\nu_d^6 + \frac{\nu_d^8}{64\nu_a^2} = 0$$

The $^3J_{NH\alpha}$ coupling constant values were greater than 7 Hz for the residues in all peptides, consistent with the β-hairpin structure (Tables 2-7, 2-8, and 2-9).

39

Table 2-7. The $^3J_{NH\alpha}$ Coupling Constant Values (Hz) of Peptide HPTFXaaAla

Residue	Xaa (Hz)		
	Mmk	Dmk	Tmk
Cys1	8.4	11	12
Arg2	11	12	12
Thr3	11	11	12
Val4	12	9.4	12
Xaa5	11	12	8.1
Val6	10	10	11
Gly8	n.d.	7.3	10
Orn9	11	12	11
Ala10	11	11	11
Ile11	12	11	12
Leu12	12	12	13
Gln13	12	11	11
Cys14	10	12	13

Table 2-8. The $^3J_{NH\alpha}$ Coupling Constant Values (Hz) of Peptide HPTXaaAla

Residue	Xaa (Hz)		
	Mmk	Dmk	Tmk
Arg1	8.6	11	11
Thr2	11	11	11
Val3	9.6	8.8	9
Xaa4	11	9.6	9.6
Val5	12	12	9.4
Gly7	8.3	8.3	9
Orn8	13	11	12
Ala9	8.8	11	8.1
Ile10	8.8	9.2	6.9
Leu11	12	9.8	9.8
Gln12	8.6	10	8.6

Table 2-9. The $^3J_{NH\alpha}$ Coupling Constant Values (Hz) of Peptide HPTUXaaAla

Residue	Xaa (Hz)		
	Mmk	Dmk	Tmk
Arg1	11	9.8	9.8
Thr2	11	8.6	9.4
Val3	9.7	11	9.0
Xaa4	9.8	9.8	8.8
Val5	11	10	9.6
Gly7	9.8	9.0	8.6
Orn8	11	10	11
Ala9	8.1	10	9.0
Ile10	9.7	9.5	9.0
Leu11	11	11	11
Gln12	8.6	8.1	9.4

NOEs are another piece of information to confirm the secondary structure of the peptide. NOEs are usually observed for protons separated by less than 5 Å. A strong NOE would be assigned for distances shorter than 2.5 Å (■■■), a medium NOE would be for the range of 2.5-3.5 Å (■■■) and a weak NOE would be for distances beyond 3.5 Å (———) (Figures 2-5~2-13). Illustrations were used to show the correlations (Figures 2-5~2-13). All sequential NH-Hα correlations in each peptide were observed. The fully folded reference peptides HPTFXaaAla exhibited more cross-strand correlations than the experimental peptides HPTXaaAla and the unfolded reference peptides HPTUXaaAla. The interproton distance (d_{NOE}) was derived via distance-volume calibrations from the volume integrals of ROESY cross peaks.
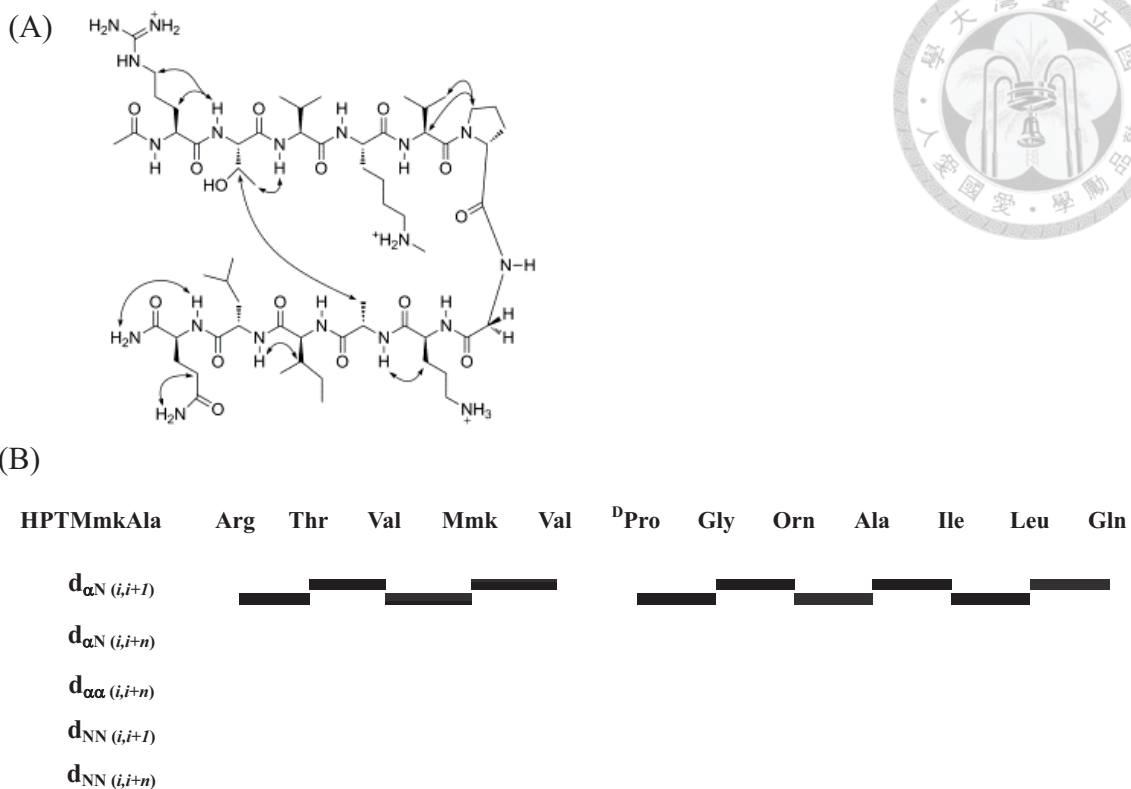
Figure 2-5. The NOEs (A) and Wüthrich diagram (B) of peptide HPTMmkAla. The thickness of the bands reflects the NOE intensity.
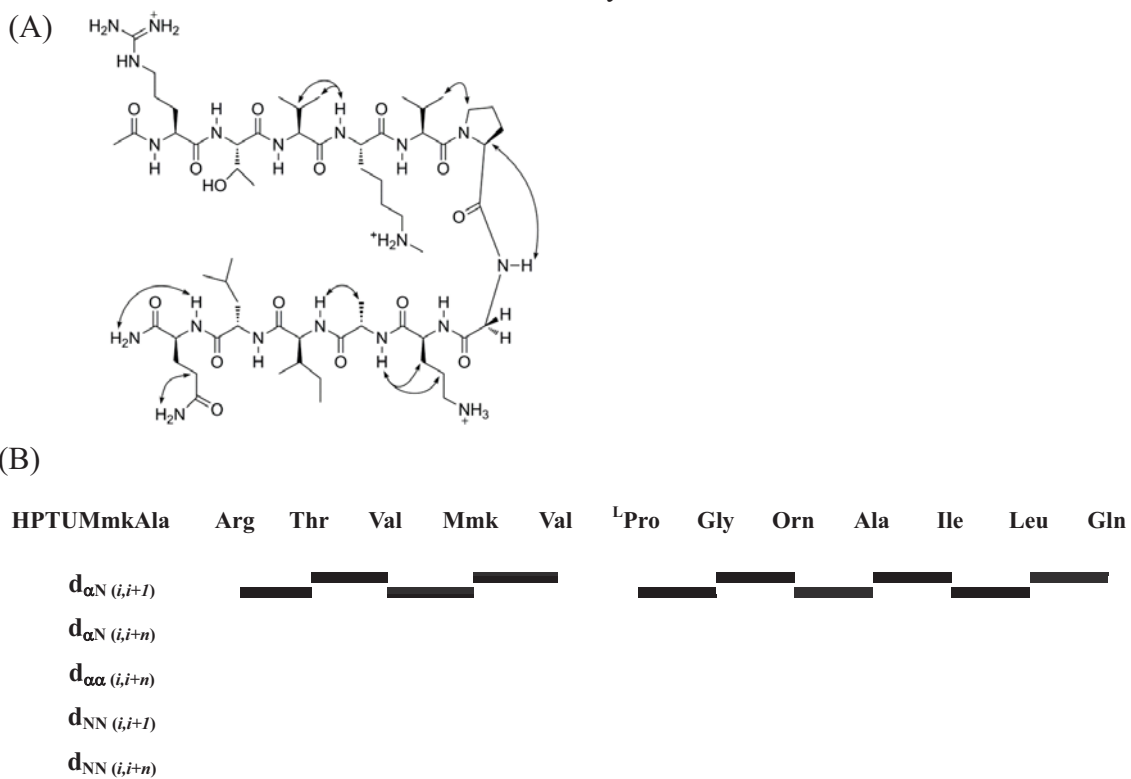
Figure 2-6. The NOEs (A) and Wüthrich diagram (B) of peptide HPTUMmkAla. The thickness of the bands reflects the NOE intensity.
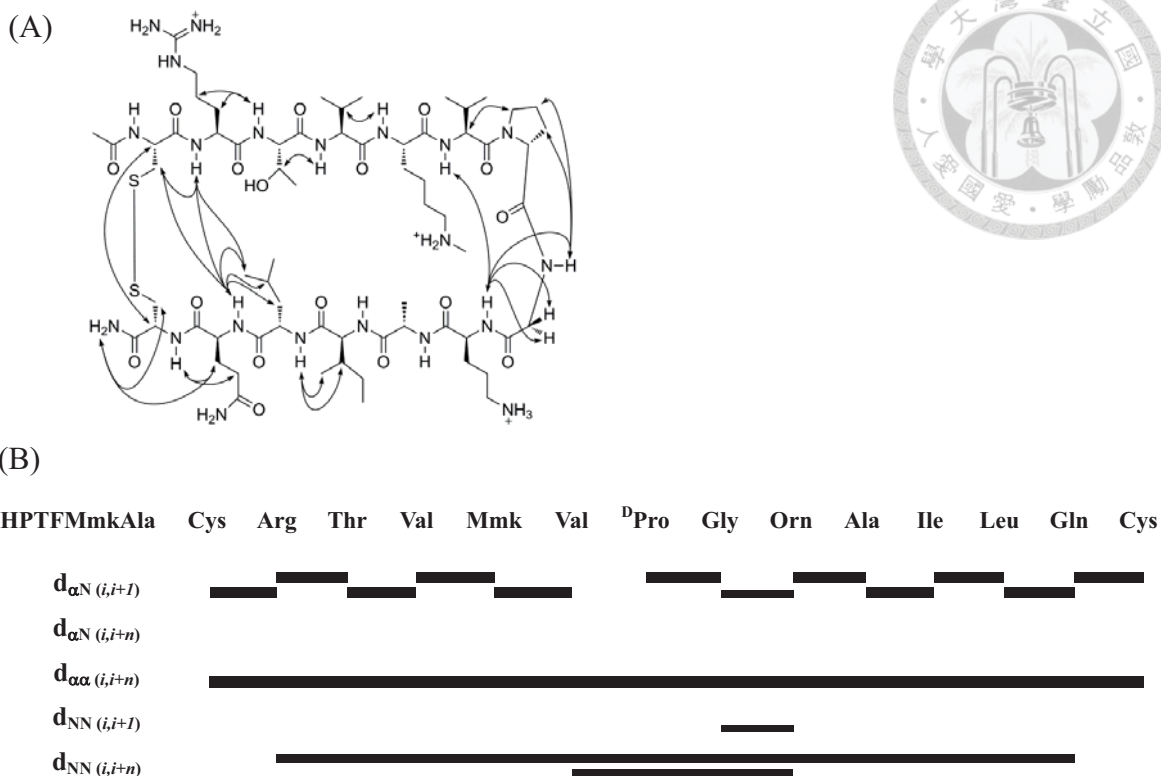
Figure 2-7. The NOEs (A) and Wüthrich diagram (B) of peptide HPTFMmkAla. The thickness of the bands reflects the NOE intensity.
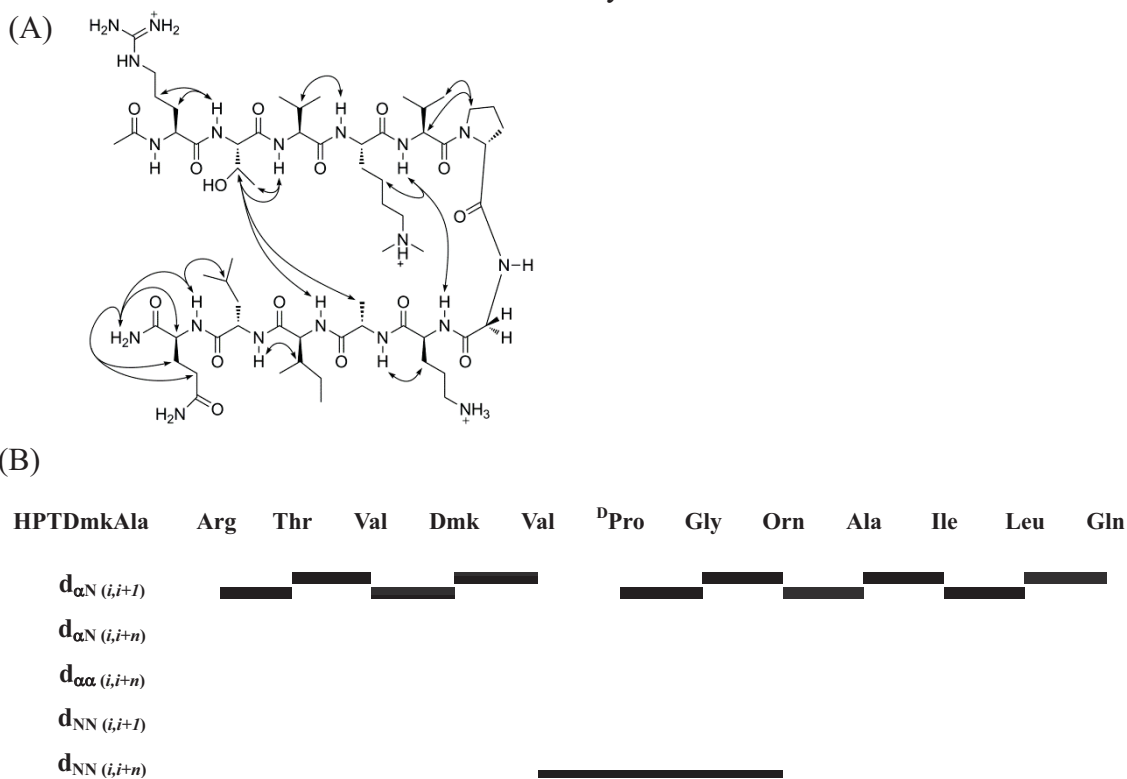
Figure 2-8. The NOEs (A) and Wüthrich diagram (B) of peptide HPTDmkAla. The thickness of the bands reflects the NOE intensity.

(A)

(B)

HPTUDmkAla	Arg	Thr	Val	Dmk	Val	LPro	Gly	Orn	Ala	Ile	Leu	Gln
$d_{\alpha N\ (i,i+1)}$												
$d_{\alpha N\ (i,i+n)}$												
$d_{\alpha\alpha\ (i,i+n)}$												
$d_{NN\ (i,i+1)}$												
$d_{NN\ (i,i+n)}$												
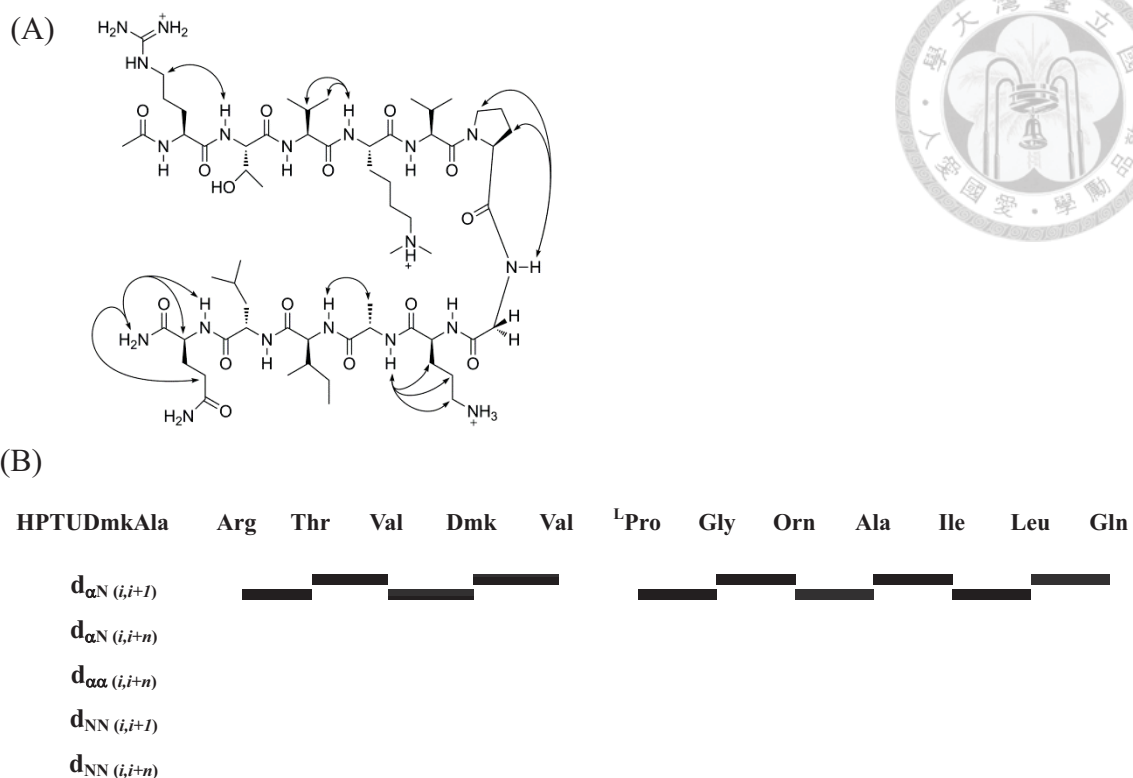
Figure 2-9. The NOEs (A) and Wüthrich diagram (B) of peptide HPTUDmkAla. The thickness of the bands reflects the NOE intensity.

(A)

(B)

HPTFDmkAla	Cys	Arg	Thr	Val	Dmk	Val	DPro	Gly	Orn	Ala	Ile	Leu	Gln	Cys
$d_{\alpha N\ (i,i+1)}$														
$d_{\alpha N\ (i,i+n)}$														
$d_{\alpha\alpha\ (i,i+n)}$														
$d_{NN\ (i,i+1)}$														
$d_{NN\ (i,i+n)}$														
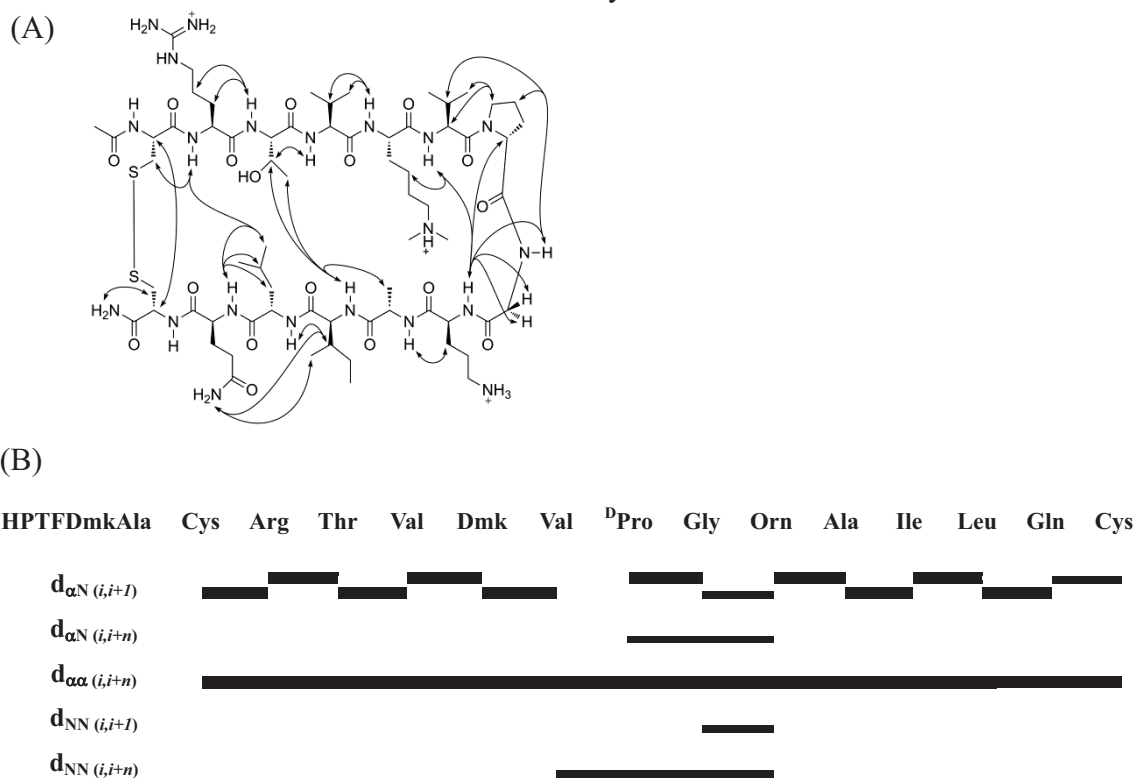
Figure 2-10. The NOEs (A) and Wüthrich diagram (B) of peptide HPTFDmkAla. The thickness of the bands reflects the NOE intensity.

(A)

(B)

HPTTmkAla	Arg	Thr	Val	Tmk	Val	^{D}Pro	Gly	Orn	Ala	Ile	Leu	Gln
$d_{\alpha N\,(i,i+1)}$												
$d_{\alpha N\,(i,i+n)}$												
$d_{\alpha\alpha\,(i,i+n)}$												
$d_{NN\,(i,i+1)}$												
$d_{NN\,(i,i+n)}$												
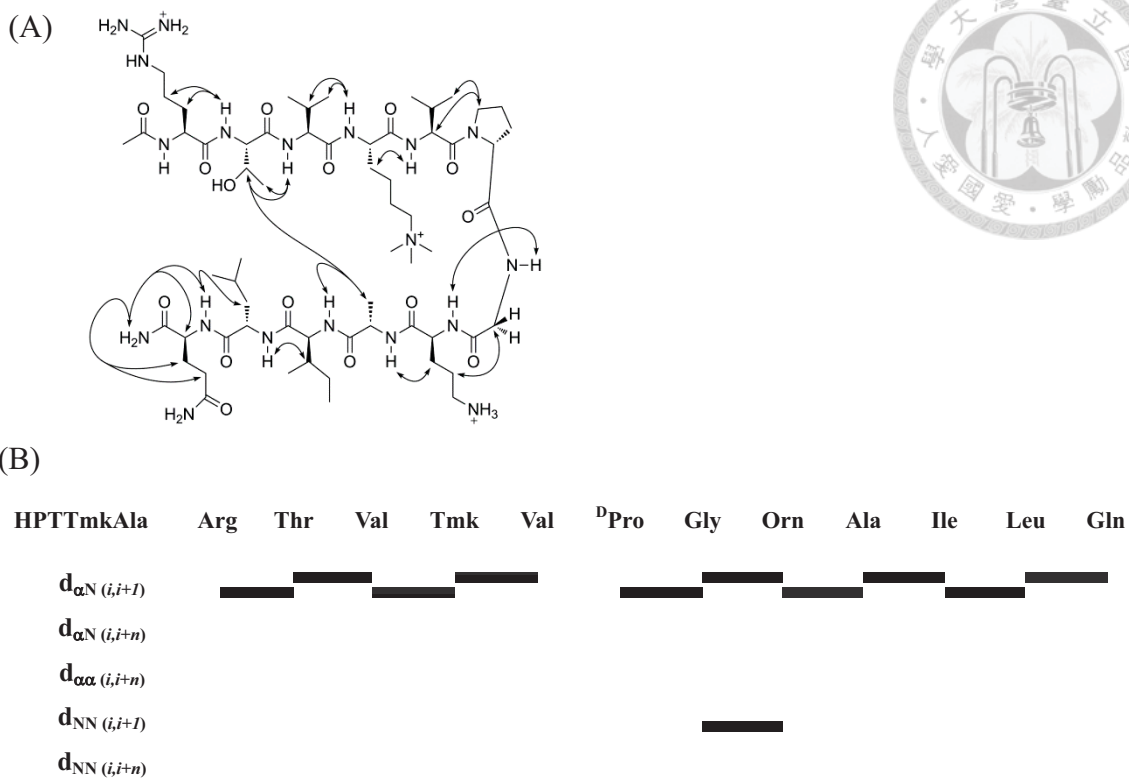
Figure 2-11. The NOEs (A) and Wüthrich diagram (B) of peptide HPTTmkAla. The thickness of the bands reflects the NOE intensity.

(A)

(B)

HPTUTmkAla	Arg	Thr	Val	Tmk	Val	^{L}Pro	Gly	Orn	Ala	Ile	Leu	Gln
$d_{\alpha N\,(i,i+1)}$												
$d_{\alpha N\,(i,i+n)}$												
$d_{\alpha\alpha\,(i,i+n)}$												
$d_{NN\,(i,i+1)}$												
$d_{NN\,(i,i+n)}$												
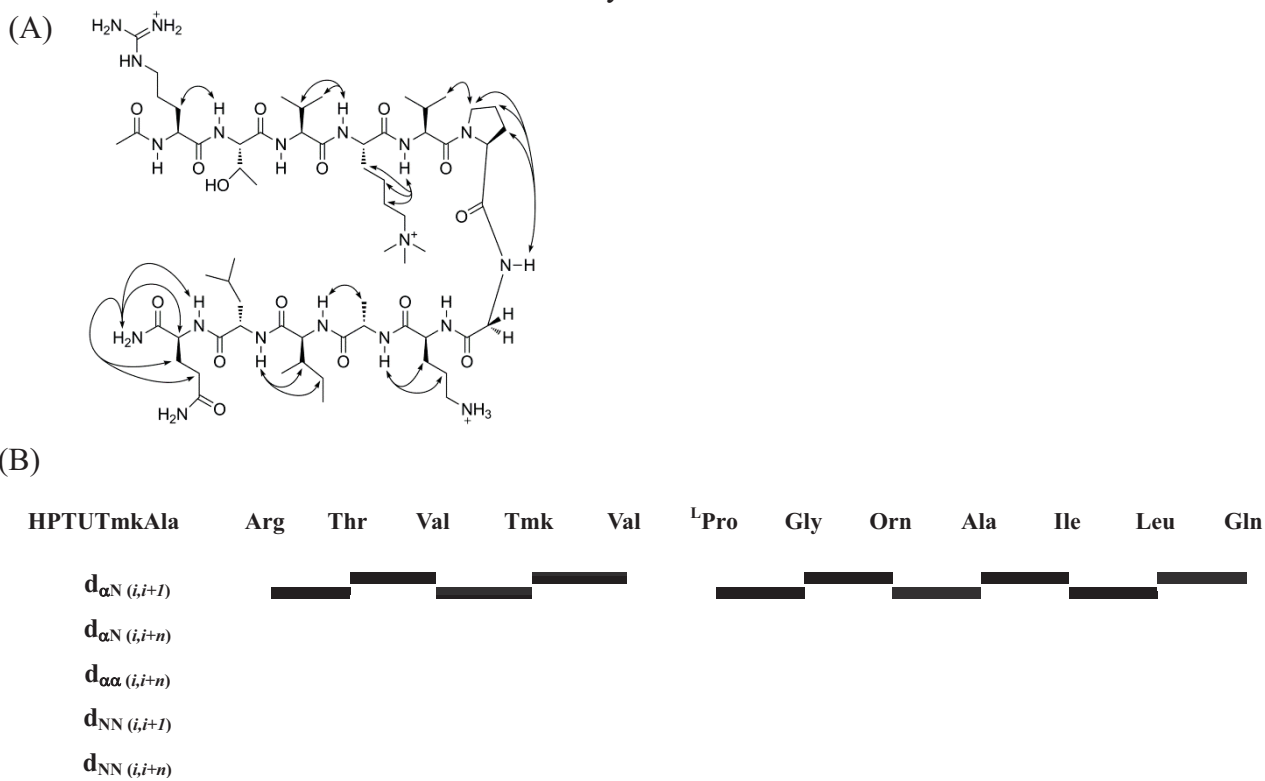
Figure 2-12. The NOEs (A) and Wüthrich diagram (B) of peptide HPTUTmkAla. The thickness of the bands reflects the NOE intensity.

(A)

(B)

HPTFTmkAla	Cys	Arg	Thr	Val	Tmk	Val	^{D}Pro	Gly	Orn	Ala	Ile	Leu	Gln	Cys
$d_{\alpha N\,(i,i+1)}$														
$d_{\alpha N\,(i,i+n)}$														
$d_{\alpha\alpha\,(i,i+n)}$														
$d_{NN\,(i,i+1)}$														
$d_{NN\,(i,i+n)}$														
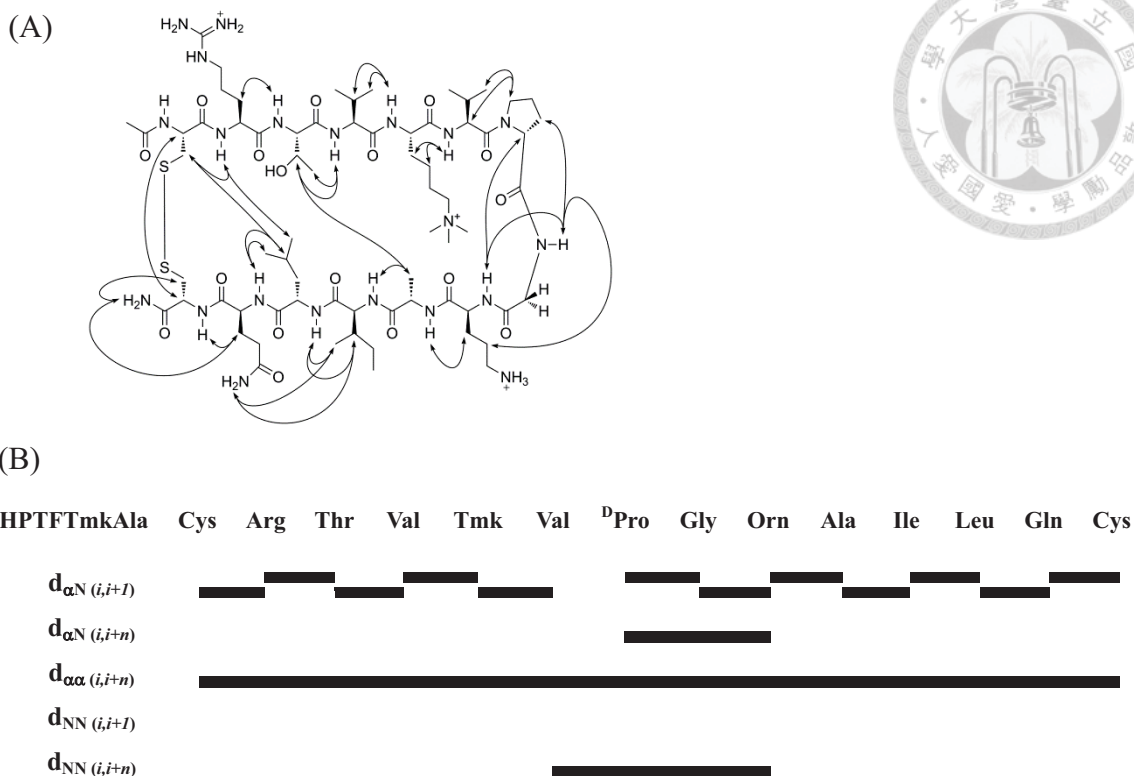
Figure 2-13. The NOEs (A) and Wüthrich diagram (B) of peptide HPTFTmkAla. The thickness of the bands reflects the NOE intensity.

Chemical shift deviation, $^3J_{NH\alpha}$ coupling constant, and NOEs provided information to confirm that all the peptides adopted a hairpin structure. The folding percentage for each residue in the peptide was derived from $\Delta\delta$ Hα data using the following equation:

$$\text{Folding Percentage} = \frac{\delta_{obs} - \delta_U}{\delta_F - \delta_U} \times 100\%$$

The equation was based on a two-stated folding process between the fully folded and fully unfolded states (Figure 2-14).[22, 23] The terminal residues and turn residues exhibited lower folding percentage than the central strand residues. The fraction folded for a given peptide was represented by the average of the fraction folded for residues at positions 2, 3, 9, and 10.[21, 46] These residues were selected to provide equal representation for the hydrogen bonded (3 and 10) and non-hydrogen bonded positions
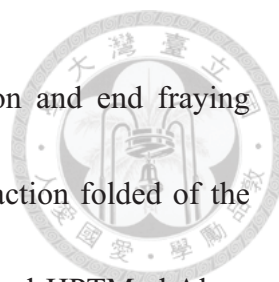
(2 and 9), avoid artifacts due to turn-promoted sheet conformation and end fraying effects, and account for unequal folding of the two strands. The fraction folded of the HPTXaaAla peptides with modified Lys analogues followed the trend HPTMmkAla ~ HPTDmkAla ~ HPTTmkAla (Figure 2-15). The results indicated that the effects of various methylated lysines on sheet stability were similar. The folding free energy (ΔG_{fold}) for each residue of the peptide was also derived from the chemical shift using the following equation:

$$\Delta G_{fold} = -RT \ln \frac{\delta_{obs} - \delta_U}{\delta_F - \delta_{obs}}$$

The ΔG_{fold} of each peptide was represented by the average of the ΔG_{fold} for residues at positions 2, 3, 9, and 10 (Table 2-10).[22, 23] The more positive the folding free energy indicates the less stable β-hairpin structure. The ΔG_{fold} followed the same trend as fraction folded: HPTMmkAla ~ HPTDmkAla ~ HPTTmkAla (Table 2-10).

(A)   (B)   (C)
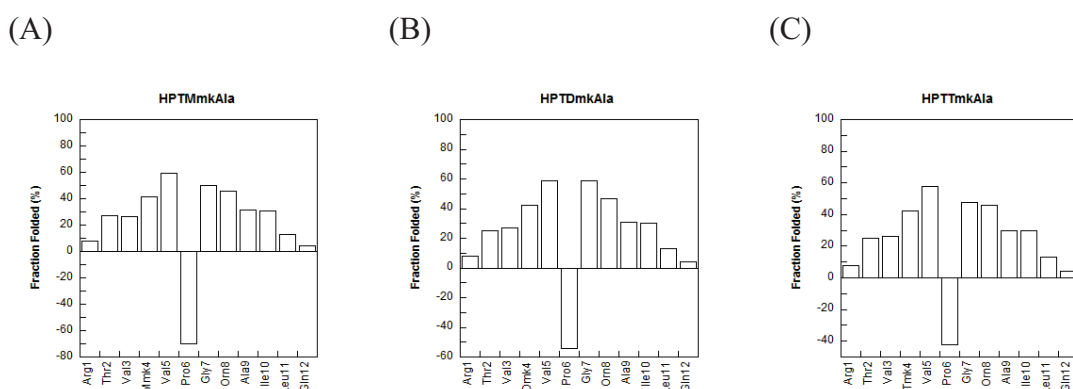
Figure 2-14. The folding percentage of each residue for peptide HPTMmkAla (A), HPTDmkAla (B), and HPTTmkAla (C).
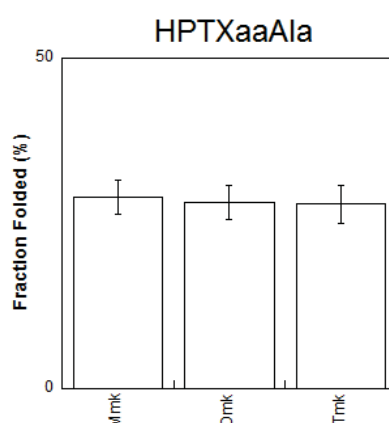
Figure 2-15. Fraction folded for the HPTXaaAla peptides (Xaa = Mmk, Dmk, Tmk).

Table 2-10. Fraction Folded (%)[a] and ΔG_{fold} (kcal/mol)[b] of the Peptides HPTXaaAla

Peptide	Fraction Folded (%)[a]	ΔG_{fold} (kcal/mol)[b]
HPTMmkAla	29.0 ± 2.6	0.532 ± 0.074
HPTDmkAla	28.2 ± 2.6	0.556 ± 0.077
HPTTmkAla	27.9 ± 2.9	0.566 ± 0.086

[a]Average of fraction folded for residue 2, 3, 9, and 10.
[b]Average of ΔG_{fold} for residue 2, 3, 9, and 10.

2-3 Conclusions

Lysine methylation has various effects on different positions in helical peptides. The effect of methylation on position 9 followed the trend: Mmk > Dmk > Tmk; whereas the effects on position 14 have no significant difference between each modified residue, with the w_{14} value of Tmk slightly higher than the others. It is possible that the residue at position 14 would be affected more by the end fraying effect than the residue at position 9, which might account for the different and insignificant trend observed for peptides KXaa14. The w_9 values decreased with the increasing number of attached

methyl groups, suggesting that decreasing the hydrogen bonding capacity compromises helix formation. The available hydrogen bond acceptors on the Lys side chain potentially interact with the backbone and strengthen the helical structure.[56] Moreover, the w value of Lys (1.06 ± 0.03) was larger than any of the modified Lys.[1] This indicated that replacing hydrogen with methyl groups destabilized the helix conformation through removing possible hydrogen bonding, despite the increased hydrophobicty. The effects of all three N-capped peptides were energitcally unfavorable for helix formation. This is due to the repulsive interaction between two positively charged groups: side chain and N-terminal amino groups, and cannot be canceled by increasing hydrophobicity. The effect of methylation on the C-terminus followed the trend Tmk > Mmk > Dmk. The c value of Lys (2.85 ± 0.40) was in between the values observed form Tmk and Mmk.[1] The increased energetic preference for Tmk was possibly due to the increased hydrophobic interaction with the C-terminus.

Different number of methyl groups attached to the lysine amino group showed similar effects towards β-hairpin stabilization. The fraction folded and ΔG_{fold} of HPTLysAla has been derived by our group and were 28.8 ± 2.4 (%) and 0.554 ± 0.074 (kcal/mol), respectively.[18] Therefore, the effect of lysine methylation has no effect on β-sheet propensity in this hairpin peptide model, although methylation would potentially increase the hydrophobicity of the residue.

2-4 Future Aspects

Utilizing model peptides incorporated with the desired amino acid of interest could reveal the propensities of the amino acid for forming different secondary structures. However, the sequence of the model peptides are not typically found in by natural proteins, thus the resulting propensities may be bio-irrelevant. Therefore, investigating a natural protein sequence which adopts a specific secondary structure conformation in its natural form and also contains the amino acid of interest would be a more direct method to investigate the effects on secondary structure. For investigating the effect of lysine methylation on secondary structure, proteins with lysine methylation sites located in specific secondary structures were surveyed. In histone proteins, lysine methylation and demethylation serves as a switch to allow or block the access of transcriptional factors to DNA.[57] In histone 3.3, residues 49 to 59 (-Leu-Arg-Glu-Ile-Arg-Arg-Tyr-Gln-**Lys**-Ser-Thr-) adopt an α-helix conformation and lysine at residue 57 was found to be methylated.[57] For lysine methylation in β-sheet, methylated site Lys197 in polycomb protein EED was located in a strand structure ranging from residues 193 to 198 (-Ile-Asn-Glu-Leu-**Lys**-Phe-).[58] The polycomb protein EED is a core component of the polycomb repressive complex 2 (PRC2),[58] which exhibits methyltrasferase activity and binding towards methylated lysines in histone proteins.[58] The activity of PRC2 is regulated by EED lysine methylation.[58] These two

sequences could be used as models for investigating the effect of lysine methylation on helix and sheet stability in a more biological relevant context.

The helix and hairpin model peptides are designed based on the sequences in histone 3.3 and polycomb protein EED, respectively (Table 2-11).[57, 58] The neighboring strand in the hairpin peptide is based on the residues 204 to 211 in the polycomb protein EED (-Asn-Leu-Leu-Leu-Ser-Val-Ser-Lys-).[58] The sequence forms an anti-parallel β-sheet conformation with the strand of interest in the natural protein. The peptide will be synthesized via Fmoc-based chemistry and purified by reverse phase (RP)-HPLC.[43] For the helical peptide, the concentration will be determined by UV-vis spectroscopy.[37] The helicity will be determined by circular dichorism spectroscopy at pH 7.[49] The helicity should decrease upon lysine methylation because the histone structure was loosened after methylation in order to interact with transcriptional proteins.[32] For the hairpin peptide, the structural information will be obtained by by two-dimensional NMR spectroscopy including DQF-COSY,[50] TOCSY,[51] and ROESY.[52] The hairpin stability should increase upon methylation because a stable sheet structure was required for recognizing the methylated lysines in histone protein.[58]
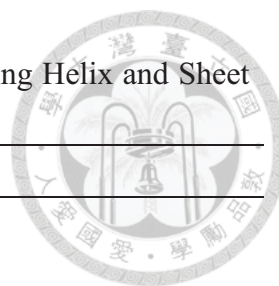
Table 2-11. Sequences for the Future Model Peptides for Investigating Helix and Sheet Stability

Peptide	Sequence
H3.3Xaa57	Ac-Leu Arg Glu Ile Arg Arg Tyr Gln **Xaa** Ser Thr-NH$_2$
EEDXaa197	Ac-Ile Asn Glu Leu **Lys** Phe ^{D}Pro Gly Asn Leu Leu Leu Ser Val Ser Lys-NH$_2$
EEDUXaa197	Ac-Ile Asn Glu Leu **Lys** Phe Pro Gly Asn Leu Leu Leu Ser Val Ser Lys-NH$_2$
EEDFXaa197	Ac-Cys Ile Asn Glu Leu **Lys** Phe ^{D}Pro Gly Asn Leu Leu Leu Ser Val Ser Lys Cys-NH$_2$

Xaa = lysine (Lys), monomethyllysine (Mmk), dimethyllysine (Dmk), trimethyllysine (Tmk)

However, these protein segments may not fold into their native structures because the neighboring parts of the protein or other ligands in the physiological environment may contribute to the formation of the structure. The length of the peptide may also require further adjustments. Moreover, there may exist multiple factors aside from secondary structure propensity that stabilize the structure, it would be difficult to measure only the contribution of the propensity. The position in the sequence may affect the propensity measurement. Side reactions may also occur during peptide synthesis due to the variety of side chain functional groups. Replacing original residues with amino acids exhibited minimum side chain interaction such as Ala in model sequence might be necessary for the helical peptides. Moreover, it is possible that the model peptides do not fold into their native form and thus some other proteins might be required.

2-5 Acknowledgement
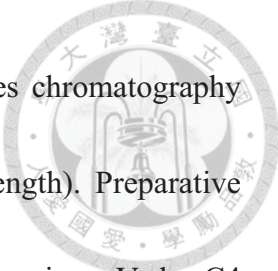
2-6 Experimental Section

General Materials and Methods

Reagents and solvents were used without purification. All of the chemical reagents were purchased from Sigma Aldrich unless indicated otherwise. Diisopropylethylamine (DIEA), piperidine, trifluoroacetic acid (TFA), and acetic anhydride (Ac_2O) were from Arcos. Dimethylformamide (DMF) and hexanes were from Mallinckrodt. Methanol and acetonitrile were from Merck. *N*-9-Fluorenylmethoxycarbonyl (Fmoc)-amino acids, 1-hydroxybenzotriazole (HOBt), *O*-1*H*-benzotriazol-1-yl-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), NovaSyn®TGR resin, and Fmoc-protected methylated lysine derivatives were from NovaBiochem. Fmoc-PAL-PEG-PS resin was from Applied Biosystems. Analytical

reverse phase (RP)-HPLC was performed on an Agilent 1200 series chromatography system using a Vydac C18 column (4.6 mm diameter, 250 mm length). Preparative RP-HPLC was performed on a Waters Breeze chromatography system using a Vydac C4 or C18 column (22 nm diameter, 250 mm length). Mass spectrometry of the peptides was performed on a matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) spectrometer (Bruker BIFLEX) using α-cyano-4-hydroxycinnamic acid as the matrix. Determination of peptide concentration was performed on a UV-vis spectrometer (Jasco V-650).

Peptide Synthesis

Fmoc-PAL-PEG-PS or NovaSyn®TGR resin (0.05 mmol) was swollen in *N,N*-dimethylformamide (DMF, 5 mL) for 30 minutes before the first coupling. The resin was then washed with DMF (5 mL, 5x1 min). This was followed by Fmoc deprotection with 20% piperidine/DMF (5 mL, 3x8 min). The resin was subsequently washed with DMF (5 mL, 5x1 min). A mixture of 3 equivalents of the appropriately protected Fmoc amino acid, HOBt, and HBTU was dissolved in DMF (1 mL). Diisopropylethylamine (DIEA, 8 equivalents) was then added to the solution. The solution was then mixed thoroughly and should give a fruity ester smell and should turn yellow. The solution was then applied to the resin. The vial that contained the solution was rinsed with DMF (2x1 mL), and the DMF was added to the reaction. The coupling

54

reaction was typically carried out for 45 minutes. The coupling times varied for different amino acids depending upon their positions in the sequence, the former residue, and the residue itself. The first amino acid was coupled for 8 hours. The 8th to 14th residues that were attached to the resin were coupled for 1.5 hours. For capping with acetic anhydride, a solution of Ac_2O (20 equivalents), DIEA (20 equivalents), and DMF (3 mL) was added to the resin. The reaction was shaken for 2 hours. The resin was subsequently washed with DMF (5 mL, 5x1 min) and then washed with MeOH (1mL, 1 min), and lyophilized overnight. The peptide was deprotected and cleaved off the resin by treating the resin with 95:5 trifluoroacetic acid (TFA) /triisopropylsilane for 2 hours. Additional 0.5 mL ethanedithiol was added during cleavage step when treating with cysteine-containing peptides (fully folded hairpin peptides). The reaction was then filtered through glass wool and the resin was washed with TFA. The combined filtrate was then evaporated by a gentle stream of N_2. The resulting oil was washed with 3 mL hexanes, dissolved in water, and lyophilized. The peptides (1 mg mL^{-1} aqueous solution) were analyzed using an analytical C18 column using 1 mL/min flow rate, linear 1%/min gradient from 100% A to 0% A (solvent A: 99.9% water, 0.1%TFA; solvent B: 90% acetonitrile, 10% water, 0.1% TFA). Appropriate linear solvent A/ solvent B gradients were used for purification on RP-HPLC preparative C4 and C18 columns. The identity of the peptides was confirmed by MALDI-TOF.

KMmk9

(Ac-Tyr-Gly-Gly-Lys-Ala-Ala-Ala-Ala-**Mmk-**Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-Lys-NH$_2$)

   The peptide was synthesized using 314.5 mg (0.057 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 388.0 mg of resin (67.2% yield). Large scale cleavage gave 81.3 mg of crude peptide (98.7% yield). Retention time on analytical RP-HPLC was 24.1 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 07_17) and C18 (PLG 13_22) columns. The final weight and purity were 9.6 mg and 98.9%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for C$_{76}$H$_{130}$N$_{24}$O$_{21}$ [MH]$^+$: 1715.99; observed : 1715.75.

KMmk14

(Ac-Tyr-Gly-Gly-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-**Mmk**-Ala-Ala-Ala-Ala-Lys-NH$_2$)

   The peptide was synthesized using 281.8 mg (0.051 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 337.3 mg of resin (56.6% yield). Large scale cleavage gave 70.9 mg of crude peptide (>99% yield). Retention time on analytical RP-HPLC was 24.0 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 7_16) and C18 (PLG 13_22) columns. The final weight and purity were 5.8 mg and 98.6%,

respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{76}H_{130}N_{24}O_{21}$ [MH]$^+$: 1715.98; observed : 1716.37.

NCapMmk

(Ac-**Mmk**-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-Lys-Gly-Gly-Tyr-NH$_2$)

The peptide was synthesized using 313.0 mg (0.056 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 403.8 mg of resin (83.4% yield). Large scale cleavage gave 89.8 mg of crude peptide (93.0% yield). Retention time on analytical RP-HPLC was 23.7 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 07_16) and C18 (PLG 13_22) columns. The final weight and purity were 8.9 mg and 98.6%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{76}H_{130}N_{24}O_{21}$ [MH]$^+$: 1715.98; observed : 1715.90.

CCapMmk

(Ac-Tyr-Gly-Gly-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-**Mmk**-NH$_2$)

The peptide was synthesized using 304.6 mg (0.055 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 380.7 mg or resin (71.8% yield). Large scale cleavage gave

96.6 mg of crude peptide (>99% yield). Retention time on analytical RP-HPLC was 24.2 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 7_16) and C18 (PLG 13_22) columns. The final weight and purity were 12.5 mg and 98.7%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{76}H_{130}N_{24}O_{21}$ $[MH]^+$: 1715.98; observed : 1715.60.

KDmk9

(Ac-Tyr-Gly-Gly-Lys-Ala-Ala-Ala-Ala-**Dmk-**Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-L

ys-NH$_2$)

The peptide was synthesized using 273.5 mg (0.052 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 331.3 mg of resin (57.1% yield). Large scale cleavage gave 91.8 mg of crude peptide (>99% yield). Retention time on analytical RP-HPLC was 25.0 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 09_19) and C18 (PLG 12_25) columns. The final weight and purity were 3.3 mg and 98.6%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{77}H_{132}N_{24}O_{21}$ $[MH]^+$: 1729.98; observed : 1731.70.

KDmk14

(Ac-Tyr-Gly-Gly-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-**Dmk-**Ala-Ala-Ala-Ala-L

ys-NH$_2$)

The peptide was synthesized using 266.0 mg (0.051 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 327.2 mg of resin (62.2% yield). Large scale cleavage gave 91.7 mg of crude peptide (>99% yield). Retention time on analytical RP-HPLC was 25.0 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 07_17) and C18 (PLG 12_25) columns. The final weight and purity were 3.0 mg and 98.7%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for C$_{77}$H$_{132}$N$_{24}$O$_{21}$ [MH]$^+$: 1729.98; observed : 1729.70.

NCapDmk

(Ac-**Dmk**-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-Lys-Gly-Gly-Tyr-NH$_2$)

The peptide was synthesized using 276.9 mg (0.053 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 317.5 mg of resin (39.6% yield). Large scale cleavage gave 76.5 mg of crude peptide (>99% yield). Retention time on analytical RP-HPLC was 25.0 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 07_17) and C18 (PLG 14_25) columns. The final weight and purity were 8.1 mg and 98.5%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for C$_{77}$H$_{132}$N$_{24}$O$_{21}$ [MH]$^+$: 1729.98; observed : 1729.95.

CCapDmk

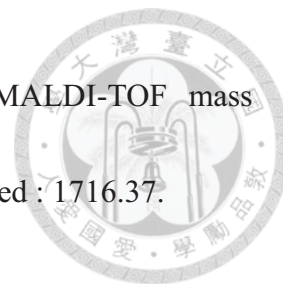(Ac-Tyr-Gly-Gly-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-**Dm k**-NH$_2$)

The peptide was synthesized using 274.2 mg (0.052 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 335.4 mg of resin (60.3% yield). Large scale cleavage gave 99.1 mg of crude peptide (>99% yield). The peptide was purified by preparative RP-HPLC using C4 (PLG 09_22) and C18 (PLG 14_25) columns. The final weight and purity were 1.5 mg and 98.6%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for C$_{77}$H$_{132}$N$_{24}$O$_{21}$ [MH]$^+$: 1730.01; observed : 1729.85.

KTmk9

(Ac-Tyr-Gly-Gly-Lys-Ala-Ala-Ala-Ala-**Tmk-**Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-L ys-NH$_2$)

The peptide was synthesized using 277.8 mg (0.050 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 354.4 mg of resin (78.1% yield). Large scale cleavage gave 73.5 mg of crude peptide (85.7% yield). Retention time on analytical RP-HPLC was 23.8 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 07_17) and C18 (PLG 14_23) columns. The final weight and purity were 5.8 mg and 98.9%,

respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{78}H_{135}N_{24}O_{21}^{+}$ [M]$^{+}$: 1745.03; observed : 1743.80.
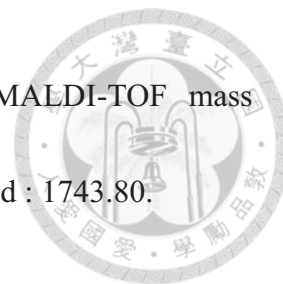
KTmk14
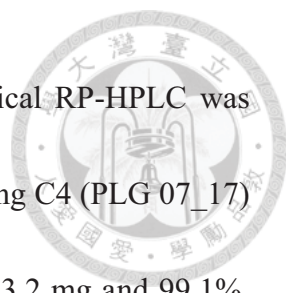
(Ac-Tyr-Gly-Gly-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-**Tmk**-Ala-Ala-Ala-Ala-Lys-NH$_2$)

The peptide was synthesized using 292.2 mg (0.053 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 353.6 mg of resin (59.5% yield). Large scale cleavage gave 71.9 mg of crude peptide (>99% yield). Retention time on analytical RP-HPLC was 23.8 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 07_17) and C18 (PLG 14_23) columns. The final weight and purity were 7.8 mg and 98.7%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{78}H_{135}N_{24}O_{21}^{+}$ [M]$^{+}$: 1745.03; observed : 1743.89.
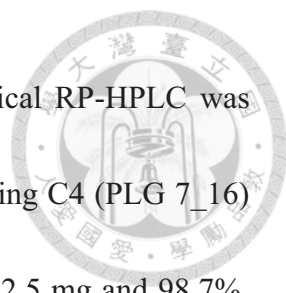
NCapTmk

(Ac-**Tmk**-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-Lys-Gly-Gly-Tyr-NH$_2$)

The peptide was synthesized using 303.6 mg (0.055 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 371.5 mg of resin (63.4% yield). Large scale cleavage gave

85.7 mg of crude peptide (>99% yield). Retention time on analytical RP-HPLC was 23.6 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 07_17) and C18 (PLG 12_22) columns. The final weight and purity were 13.2 mg and 99.1%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{78}H_{135}N_{24}O_{21}^{+}$ $[M]^{+}$: 1745.03; observed : 1745.14.

CCapTmk

(Ac-Tyr-Gly-Gly-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-Lys-Ala-Ala-Ala-Ala-**Tmk**-NH$_2$)

The peptide was synthesized using 312.0 mg (0.056 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 364.2 mg of resin (47.4% yield). Large scale cleavage gave 84.4 mg of crude peptide (>99% yield).Retention time on analytical RP-HPLC was 24.1 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 09_19) and C18 (PLG 14_23) columns. The final weight and purity were 7.2 mg and 98.7%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{78}H_{135}N_{24}O_{21}^{+}$ $[M]^{+}$: 1745.03; observed : 1745.23.

HPTMmkAla

(Ac-Arg-Thr-Val-**Mmk**-Val-^{D}Pro-Gly-Orn-Ala-Ile-Leu-Gln-NH$_2$)

The peptide was synthesized using 201.8 mg (0.050 mmol) of NovaSyn®TGR resin. The synthesis gave 291.0 mg of resin (94.9% yield). Large scale cleavage gave 59.0 mg of crude peptide (73.1% yield). Retention time on analytical RP-HPLC was 28.3 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 06_16) and C18 (PLG 16_26) column. The final weight and purity were 16.0 mg and 96.1%, respectively. The NMR sample concentration was 11.8 mM. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{61}H_{111}N_{19}O_{15}$ $[MH]^+$: 1350.86; observed : 1350.79.

HPTDmkAla

(Ac-Arg-Thr-Val-**Dmk**-Val-^{D}Pro-Gly-Orn-Ala-Ile-Leu-Gln-NH$_2$)

The peptide was synthesized using 215.3 mg (0.054 mmol) of NovaSyn®TGR resin. The synthesis gave 314.5 mg of resin (98.2% yield). Large scale cleavage gave 77.1 mg of crude peptide (85.9% yield). Retention time on analytical RP-HPLC was 28.4 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 06_18) and C18 (PLG 16_26) column. The final weight and purity were 5.8 mg and 95.6%, respectively. The NMR sample concentration was 8.5 mM. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{62}H_{113}N_{19}O_{15}$ $[MH]^+$: 1364.87; observed : 1364.80.

HPTTmkAla

(Ac-Arg-Thr-Val-**Tmk**-Val-^{D}Pro-Gly-Orn-Ala-Ile-Leu-Gln-NH$_2$)

The peptide was synthesized using 216.4 mg (0.054 mmol) of NovaSyn$^®$TGR resin. The synthesis gave 318.0 mg of resin (99.3% yield). Large scale cleavage gave 64.7 mg of crude peptide (70.4% yield). Retention time on analytical RP-HPLC was 28.3 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 11_21) column. The final weight and purity were 9.8 mg and 96.7%, respectively. The NMR sample concentration was 14.2 mM. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{63}H_{116}N_{19}O_{15}^+$ [M]$^+$: 1379.90; observed : 1378.72.

HPTUMmkAla

(Ac-Arg-Thr-Val-**Mmk**-Val-^{L}Pro-Gly-Orn-Ala-Ile-Leu-Gln-NH$_2$)

The peptide was synthesized using 185.5 mg (0.046 mmol) of NovaSyn$^®$TGR resin. The synthesis gave 260.6 mg of resin (87.0% yield). Large scale cleavage gave 55.9 mg of peptide crude (82.0% yield). Retention time on analytical RP-HPLC was 26.0 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 06_16) and C18 (PLG 14_24) column. The final weight and purity were 8.8 mg and 96.1%,

respectively. The NMR sample concentration was 13.0 mM. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{61}H_{111}N_{19}O_{15}$ $[MH]^+$: 1350.86; observed: 1350.84.

HPTUDmkAla

(Ac-Arg-Thr-Val-**Dmk**-Val-^{L}Pro-Gly-Orn-Ala-Ile-Leu-Gln-NH$_2$)

The peptide was synthesized using 193.7 mg (0.048 mmol) of NovaSyn$^®$TGR resin. The synthesis gave 274.6 mg of resin (89.0% yield). Large scale cleavage gave 49.0 mg of crude peptide (66.8% yield). Retention time on analytical RP-HPLC was 26.2 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 06_16) and C18 (PLG 14_24) columns. The final weight and purity were 7.8 mg and 96.7%, respectively. The NMR sample concentration was 11.4 mM. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{62}H_{113}N_{19}O_{15}$ $[MH]^+$: 1364.87; observed: 1364.82.

HPTUTmkAla

(Ac-Arg-Thr-Val-**Tmk**-Val-^{L}Pro-Gly-Orn-Ala-Ile-Leu-Gln-NH$_2$)

The peptide was synthesized using 203.8 mg (0.051 mmol) of NovaSyn$^®$TGR resin. The synthesis gave 280.1 mg of resin (79.2% yield). Large scale cleavage gave 31.9 mg
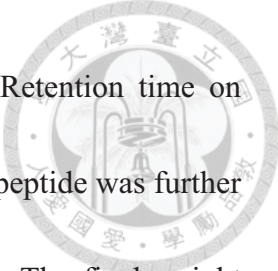
of crude peptide (46.0% yield). Retention time on analytical RP-HPLC was 26.2 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 06_16) and C18 (PLG 14_24) column. The final weight and purity were 4.8 mg and 95.3%, respectively. The NMR sample concentration was 7.0 mM. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{63}H_{116}N_{19}O_{15}^{+}$ $[M]^{+}$: 1379.90; observed : 1378.96.
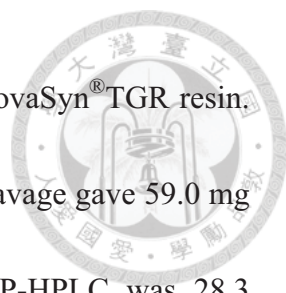

HPTFMmkAla

(Ac-Cys-Arg-Thr-Val-**Mmk**-Val-^{D}Pro-Gly-Orn-Ala-Ile-Leu-Gln-Cys-NH$_2$)

The peptide was synthesized using 245.6 mg (0.061 mmol) of NovaSyn$^{®}$TGR resin. The synthesis gave 398.4 mg of resin (97.5% yield). Large scale cleavage gave 110.5 mg of crude peptide (97.4% yield). Retention time on analytical RP-HPLC was 31.2 minutes. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{67}H_{121}N_{21}O_{17}S_2$ $[MH]^{+}$: 1556.925; observed : 1555.731. The peptide was purified by preparative RP-HPLC using C4 (PLG 11_23) column. The purified peptide was dissolved in pH 8 buffer (1 mM phosphate, 1 mM citrate, and 1 mM borate), at a concentration of 0.20 mg/mL. Granulated charcoal was added to the peptide solution, using ten times the weight of the peptide. The heterogeneous reaction mixture was stirred open to atmosphere for 2 hrs. Upon completion of the reaction, the mixture

was filtered through glass wool, and the filtrate was lyophilized. Retention time on analytical RP-HPLC after oxidation was 27.5 minutes. The oxidized peptide was further purified by preparative RP-HPLC using C18 (PLG 16_26) column. The final weight and purity were 8.5 mg and 96.1%, respectively. The NMR sample concentration was 10.9 mM. The identity of the cyclized peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{67}H_{119}N_{21}O_{17}S_2$ [MH]$^+$: 1554.925; observed : 1554.811.

HPTFDmkAla

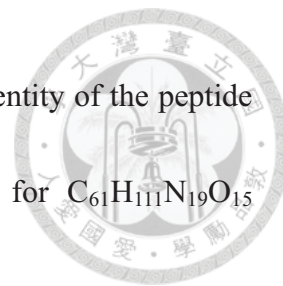(Ac-Cys-Arg-Thr-Val-**Dmk**-Val-^{D}Pro-Gly-Orn-Ala-Ile-Leu-Gln-Cys-NH$_2$)

The peptide was synthesized using 241.7 mg (0.060 mmol) of NovaSyn$^®$TGR resin. The synthesis gave 381.2 mg of resin (89.9% yield). Large scale cleavage gave 104.7 mg of crude peptide (>99% yield). Retention time on analytical RP-HPLC was 31.4 minutes. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{68}H_{123}N_{21}O_{17}S_2$ [MH]$^+$: 1570.925; observed : 1569.708. The peptide was purified by preparative RP-HPLC using C4 (PLG 11_23) column. The purified peptide was dissolved in pH 8 buffer (1 mM phosphate, 1 mM citrate, and 1 mM borate), at a concentration of 0.20 mg/mL. Granulated charcoal was added to the peptide solution, using ten times the weight of the peptide. The heterogeneous reaction mixture was stirred open to atmosphere for 2 hrs. Upon completion of the reaction, the mixture

was filtered through glass wool, and the filtrate was lyophilized. Retention time on analytical RP-HPLC after oxidation was 27.4 minutes. The oxidized peptide was further purified by preparative RP-HPLC using C18 (PLG 16_26) column. The final weight and purity were 5.9 mg and 95.6%, respectively. The NMR sample concentration was 7.5 mM. The identity of the cyclized peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{68}H_{121}N_{21}O_{17}S_2$ [MH]$^+$: 1568.925; observed : 1568.842.
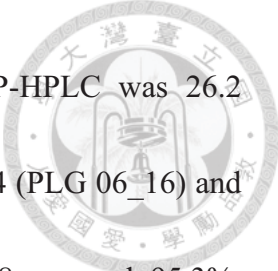
HPTFTmkAla

(Ac-Cys-Arg-Thr-Val-**Tmk**-Val-^{D}Pro-Gly-Orn-Ala-Ile-Leu-Gln-Cys-NH$_2$)

The peptide was synthesized using 243.0 mg (0.061mmol) of NovaSyn®TGR resin. The synthesis gave 383.5 mg of resin (89.6% yield). Large scale cleavage gave 101.7 mg of crude peptide (97.1% yield). Retention time on analytical RP-HPLC was 31.4 minutes. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{69}H_{126}N_{21}O_{17}S_2{}^+$ [M]$^+$: 1584.925; observed : 1583.656. The peptide was purified by preparative RP-HPLC using C4 (PLG 11_21) column. The purified peptide was dissolved in pH 8 buffer (1 mM phosphate, 1 mM citrate, and 1 mM borate), at a concentration of 0.20 mg/mL. Granulated charcoal was added to the peptide solution, using ten times the weight of the peptide. The heterogeneous reaction mixture was stirred open to atmosphere for 2 hrs. Upon completion of the reaction, the mixture was

filtered through glass wool, and the filtrate was lyophilized. Retention time on analytical RP-HPLC after oxidation was 27.4 minutes. The oxidized peptide was further purified by preparative RP-HPLC using C18 (PLG 16_26) column. The final weight and purity were 9.5 mg and 96.4%, respectively. The NMR sample concentration was 12.0 mM. The identity of the cyclized peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{69}H_{124}N_{21}O_{17}S_2^+$ $[M]^+$: 1582.925; observed : 1582.846.

Ultraviolet-Visible (UV-vis) Spectroscopy

The helical peptides were prepared as 10 mM stock solutions. The concentration of Ac-capped peptide stock solution was determined by the tyrosine absorbance in 6 M guanidinium chloride under 25 ^{o}C. 2 μL of peptide stock solution was gradually added to 10 μL into 400 μL 6 M guanidine hydrochloride. Highly concentrated guanidine hydrochloride (GdnHCl) served as a denaturant and should sabotage the peptide structure. The peptide would be in a random coil conformation and highly unfolded. The full spectra (200-500 nm) and single wavelength absorbance (276, 278, 280, 282, 335, and 400 nm) were acquired. Literature values were used for the molar absorption coefficients (ε value) of tyrosine (in 6 M GdnHCl) (ε276=1455, ε278=1395, ε280=1285, ε282=1220).[35, 37] Each addition of 2 μL peptide stock solution was stirred by pipette for 100 times. The mixed solution was equilibrated for 15 minutes before measurements. The absorbance of the four aforementioned wavelengths (276, 278, 280, and 282 nm)

were baseline corrected by the absorbance at 335 or 400 nm. The concentration of the

stock solution was derived through linear regression analysis. The R value higher than

0.999 was deemed acceptable.

Circular Dichorism Spectroscopy

CD data was collected using a 1 mm pathlength cell. CD measurements were

reported at peptide concentrations 70-80 μM in 1 M NaCl, 1 mM sodium phosphate, 1

mM sodium citrate, and 1 mM sodium borate (pH 7) at 0 °C. The data was analyzed

using Kaleidagraph 3.52 (Synergy Software CA). Each reported CD value was the mean

of at least 3 determinations. Data was expressed in terms of mean residue molar

ellipticity (deg·cm^2·dmol^{-1}). The mean residue molar ellipticity of the peptides was

independent of peptide concentration (80-160 μM). The fraction helix of each peptide

(f_{helix}) was calculated from mean residue molar ellipticity at 222 nm and the number of

backbone amides (N) using the following equation:[5]

$$f_{helix} = \frac{\theta_{222}}{40000(1 - \frac{2.5}{N})}$$

Helix Propensity and Capping Parameter Derivation

The statistical mechanical parameters for the residues were derived numerically

based on modified Lifson-Roig theory.[1, 6, 10, 45, 46] For all residues, v was set to 0.048.[5]

The N-cap parameter (n) and C-cap parameter (c) for a given amino acid Xaa were

derived from the experimentally determined f_{helix} of the corresponding NCapXaa and

CCapXaa peptides, respectively. The helix propensities (*w_9 and w_{14}*) for a given amino

acid Xaa were then derived from the f_{helix} of the corresponding KXaa9 and KXaa14

peptides based on modified Lifson-Roig theory, respectively. [5, 10, 14, 15] The free energy

for C-capping ($\Delta G_{C\text{-}cap}$) and helix formation (ΔG_{helix}) was derived by $\Delta G_{C\text{-}cap} = -RT\cdot\ln(c)$

and $\Delta G_{helix} = -RT\cdot\ln(w)$, respectively.

Hairpin Peptide Structure Analysis by 2D-NMR

The purified HPTXaaAla, HPTUXaaAla, and HPTFXaaAla peptides were

analyzed by NMR spectrometry at 25 °C. NMR samples were prepared by dissolving

the peptide in 50 mM sodium deuterioacetate in H_2O/D_2O (9:1 v/v) at pH 5.5

(uncorrected) with a concentration of approximately 5-10 mM. The

2,2-Dimethyl-2-silapentane-5-sulfonate (DSS) was added to the samples as an internal

reference. NMR experiments were performed on a Bruker AVIII 800 mHz Instrument

spectrometer. Phase-sensitive DQF-COSY,[50] TOCSY,[51] and ROESY[52] experiments were

performed. Solvent suppression was achieved by the WATERGATE solvent

suppression.[59, 60] TOCSY and ROESY experiments employed a spin locking field of 10

kHz. Mixing times of 150 ms were used in the NOESY experiments.

Table 2-12. The ^{1}H Chemical Shift Assignments for Peptide HPTMmkAla

Residue	HN	Hα	Hβ	Others
Ac		2.033		
Arg1	8.317	4.365	1.833,1.749	Hγ: 1.658,1.612 ; Hδ: 3.202 ; HNt:7.200
Thr2[a]	8.260	4.518	4.100	Hγ: 1.147
Val3[b]	8.389	4.193	2.030	Hγ: 0.898
Mmk4[c]	8.417	4.599	1.671	Hγ: 1.376,1.235; Hδ: 1.618; Hε:2.954
Val5	8.461	4.509	2.013	Hγ: 0.930
^{D}Pro6		4.414	2.331,1.975	Hγ:2.085,2.042 ; Hδ:3.832,3.802
Gly7[d]	8.453,8.451	3.940,3.878		
Orn8	8.049	4.482	1.870,1.794	Hγ:1.710 ; Hδ:3.017 ; HNt:7.616
Ala9	8.472	4.473	1.316	
Ile10[e]	8.421	4.223	1.848	Hγ:1.436,1.182,0.883 (Me); Hδ:0.825
Leu11	8.368	4.431	1.600	Hγ:1.355 ; Hδ:0.908,0.859
Gln12[f]	8.450	4.314	2.093,1.965	Hγ: 2.341 ; HNt:7.470,6.871
NH$_2$	7.625,7.115			

[a]The assignments for the minor Thr2 spin system are 8.239 (HN), 4.370 (Hα), 4.173 (Hβ), 1.180 (Hγ). [b]The assignments for the minor Val3 spin system are 8.314 (HN), 4.087 (Hα), 2.032 (Hβ), 0.898 (Hγ), 0.800 (Hγ). [c]Signal for the terminal HN and the terminal HMeN were not observed. [d]The assignments for the minor Gly7 spin system are 8.594 (HN), 3.998 (Hα), 3.949 (Hα). [e]The assignments for the minor Ile10 spin system are 8.183 (HN), 4.112 (Hα), 1.833 (Hβ), 0.897 (Hγ). [f]The assignments for the minor Gln12 spin system are 8.318 (HN), 4.297 (Hα), 2.362 (Hγ).

Table 2-13. The [1]H Chemical Shift Assignments for Peptide HPTDmkAla

Residue	HN	Hα	Hβ	Others
Ac		2.033		
Arg1	8.317	4.364	1.830,1.747	Hγ:1.630 ; Hδ:3.201; HNt:7.199
Thr2[a]	8.262	4.517	4.095	Hγ:1.146
Val3[b]	8.388	4.198	2.030	Hγ: 0.898
Dmk4[c]	8.421	4.607	1.673	Hγ:1.360,1.220 ;Hε:3.048
Val5	8.467	4.505	2.011	Hγ: 0.926
DPro6		4.413	2.329,1.973	Hγ:2.090,2.038 ; Hδ:3.830,3.801
Gly7[d]	8.452,8.421	3.938,3.875		
Orn8	8.045	4.483	1.872,1.791	Hγ:1.710 ; Hδ: 3.015 ; HNt:7.615
Ala9	8.481	4.475	1.316	
Ile10[e]	8.421	4.222	1.846	Hγ:1.435,1.180,0.881(Me); Hδ: 0.826
Leu11	8.369	4.430	1.597	Hγ:1.355 ; Hδ:0.907,0.857
Gln12[f]	8.450	4.313	2.093,1.958	Hγ:2.340 ; HNt:7.469,6.872
NH2	7.625,7.115			

[a]The assignments for the minor Thr2 spin system are 8.239 (HN), 4.369 (Hα), 4.163 (Hβ), 1.180 (Hγ). [b]The assignments for the minor Val3 spin system are 8.322 (HN), 4.090 (Hα), 0.898 (Hγ), 0.800 (Hγ). [c]Signal for the Hδ, the terminal HN, and the terminal HMeN were not observed. [d]The assignments for the minor Gly7 spin system are 8.596 (HN), 3.997 (Hα), 3.948 (Hα). [e]The assignments for the minor Ile10 spin system are 8.186 (HN), 4.113 (Hα), 1.832 (Hβ), 0.896 (Hγ). [f]The assignments for the minor Gln12 spin system are 8.339 (HN), 4.297 (Hα), 2.340 (Hγ)

Table 2-14. The ^{1}H Chemical Shift Assignments for Peptide HPTTmkAla

Residue	HN	Hα	Hβ	Others
Ac		2.036		
Arg1	8.319	4.364	1.831,1.748	Hγ:1.656 ; Hδ:3.201 ; HNt:7.203
Thr2[a]	8.265	4.513	4.100	Hγ: 1.148
Val3[b]	8.382	4.194	2.034	Hγ:0.901
Tmk4[c]	8.434	4.609	1.698	Hγ:1.358,1.222 ; Hε:3.241
Val5	8.471	4.506	2.018	Hγ: 0.931
^{D}Pro6		4.414	2.330,1.978	Hγ:2.065 ; Hδ:3.835,3.802
Gly7[d]	8.451,8.452	3.939,3.876		
Orn8	8.050	4.482	1.870,1.795	Hγ:1.716 ; Hδ:3.017 ; HNt:7.616
Ala9	8.488	4.472	1.318	
Ile10[e]	8.414	4.222	1.847	Hγ: 1.437,1.182,0.884(Me); Hδ:0.827
Leu11	8.370	4.431	1.603	Hγ:1.357 ; Hδ:0.907,0.858
Gln12[f]	8.447	4.313	2.100,1.958	Hγ:2.341 ; HNt:7.474,6.873
NH2	7.625,7.117			

[a]The assignments for the minor Thr2 spin system are 8.242 (HN), 4.369 (Hα), 4.164 (Hβ), 1.178 (Hγ). [b]The assignments for the minor Val3 spin system are 8.338 (HN), 4.092 (Hα), 0.898 (Hγ), 0.803 (Hγ). [c]Signal for the Hδ and the terminal HMeN were not observed. [d]The assignments for the minor Gly7 spin system are 8.599 (HN), 3.998 (Hα), 3.949 (Hα). [e]The assignments for the minor Ile10 spin system are 8.188 (HN), 4.114 (Hα), 1.833 (Hβ), 0.898 (Hγ). [f]The assignments for the minor Gln12 spin system are 8.339 (HN), 4.297 (Hα), 2.340 (Hγ).

Table 2-15. The ^{1}H Chemical Shift Assignments for Peptide HPTUMmkAla

Residue	HN	Hα	Hβ	Others
Ac		2.041		
Arg1	8.330	4.340	1.836,1.749	Hγ:1.641 ; Hδ:3.205 ; HNt:7.203
Thr2	8.246	4.365	4.159	Hγ: 1.181
Val3	8.177	4.115	2.040	Hγ: 0.914
Mmk4[a]	8.425	4.345	1.767	Hγ:1.415,1.339 ; Hδ:1.698 Hε:3.006
Val5[b]	8.264	4.425	2.071	Hγ: 0.969,0.931
Pro6		4.398	2.307,1.958	Hγ:2.069 ; Hδ:3.861,3.710
Gly7[c]	8.413,8.410	3.988,3.913		
Orn8	8.208	4.363	1.871	Hγ:1.729 ; Hδ:3.014 ; HNt:7.611
Ala9	8.369	4.309	1.360	
Ile10	8.189	4.118	1.839	Hγ: 1.488,1.190(Me); Hδ:0.859
Leu11	8.323	4.386	1.648,1.591	Hγ:1.590 ; Hδ:0.927,0.867
Gln12	8.345	4.299	2.111,1.979	Hγ:2.364 ; HNt: 7.528,6.861
NH2	7.581,7.106			

[a]Signal for the terminal HN and the terminal HMeN were not observed.[b]The assignments for the minor Val5 spin system are 7.981 (HN), 4.244 (Hα), 0.903 (Hγ). [c]The assignments for the minor Gly7 spin system are 8.574 (HN), 4.002 (Hα), 3.881 (Hα).

Table 2-16. The [1]H Chemical Shift Assignments for Peptide HPTUDmkAla
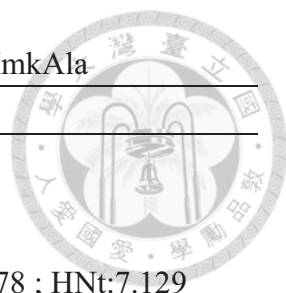
Residue	HN	Hα	Hβ	Others
Ac		2.041		
Arg1	8.329	4.339	1.833,1.750	Hγ:1.617 ; Hδ:3.204 ; HNt:7.204
Thr2	8.247	4.377	4.165	Hγ: 1.179
Val3	8.175	4.118	2.042	Hγ:0.914
Dmk4[a]	8.427	4.353	1.775	Hγ:1.334 ; Hδ:1.721; Hε:3.104
Val5[b]	8.269	4.424	2.073	Hγ:0.970,0.932
Pro6		4.400	2.309,1.946	Hγ:2.068 ; Hδ:3.862,3.710
Gly7[c]	8.415,8.412	3.990,3.915		
Orn8	8.209	4.360	1.871	Hγ:1.734 ; Hδ:3.015 ; HNt:7.613
Ala9	8.370	4.310	1.361	
Ile10	8.189	4.120	1.839	Hγ:1.489,1.190, 0.893 (Me); Hδ:0.859
Leu11	8.324	4.386	1.644,1.592	Hγ:1.589 ; Hδ:0.927,0.867
Gln12	8.345	4.299	1.978	Hγ:2.363 ; HNt:7.529,6.860
NH2	7.582,7.105			

[a]Signal for the terminal HN and the terminal HMeN were not observed. [b]The assignments for the minor Val5 spin system are 7.987 (HN), 4.257 (Hα), 0.903 (Hγ). [c]The assignments for the minor Gly7 spin system are 8.577 (HN), 4.003 (Hα), 3.884 (Hα).

Table 2-17. The ^{1}H Chemical Shift Assignments for Peptide HPTUTmkAla
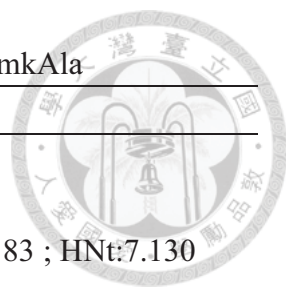
Residue	HN	Hα	Hβ	Others
Ac		2.040		
Arg1	8.330	4.337	1.835,1.749	Hγ:1.642 ; Hδ:3.203 ; HNt:7.205
Thr2	8.247	4.374	4.162	Hγ: 1.178
Val3	8.178	4.119	2.040	Hγ: 0.910
Tmk4[a]	8.438	4.368	1.802	Hγ:1.404 ; Hδ:1.738; Hε:3.291
Val5[b]	8.286	4.424	2.075	Hγ: 0.971,0.932
Pro6		4.403	2.310,1.965	Hγ:2.066 ; Hδ:3.866,3.715
Gly7[c]	8.412	3.989,3.911		
Orn8	8.211	4.359	1.870	Hγ:1.730 ; Hδ:3.013 ; HNt:7.613
Ala9	8.370	4.307	1.358	
Ile10	8.190	4.118	1.838	Hγ:1.488,1.189,0.893 (Me); Hδ:0.858
Leu11	8.330	4.385	1.641,1.592	Hγ:1.592 ; Hδ: 0.925,0.866
Gln12	8.346	4.298	2.107,1.974	Hγ:2.362 ; HNt:7.529,6.860
NH2	7.582,7.104			

[a]Signal for the terminal HMeN were not observed. [b]The assignments for the minor Val5 spin system are 8.003 (HN), 4.258 (Hα), 0.905 (Hγ). [c]The assignments for the minor Gly7 spin system are 8.576 (HN), 4.006 (Hα), 3.883 (Hα).

Table 2-18. The [1]H Chemical Shift Assignments for Peptide HPTFMmkAla

Residue	HN	Hα	Hβ	Others
Ac		2.083		
Cys1	8.449	5.202	3.168, 2.659	
Arg2	8.767	4.665	1.838, 1.681	Hγ:1.545 ; Hδ:3.178 ; HNt:7.129
Thr3	8.616	4.928	3.931	Hγ: 1.056
Val4	9.098	4.410	1.975	Hγ:0.816
Mmk5[a]	8.503	4.957	1.651	Hγ:1.349,1.183 ; δ:1.593 ;Hε:2.904
Val6	8.716	4.567	1.949	Hγ:0.909, 0.888
[D]Pro7		4.375	2.356, 1.955	Hγ:2.142, 2.043 ; Hδ:3.862, 3.774
Gly8	8.659	3.970, 3.843		
Orn9	7.940	4.624	1.831	Hγ:1.686 ; Hδ:3.006 ; HNt:7.608
Ala10	8.527	4.830	1.222	
Ile11	9.108	4.458	1.864	Hγ: 1.359, 1.144,0.861(Me); Hδ:0.796
Leu12	8.409	4.733	1.650	Hγ:1.493 ; Hδ: 0.817, 0.786
Gln13	9.211	4.664	2.073	Hγ:2.278, 2.219 ; HNt:7.314, 6.834
Cys14	8.983	5.074	3.138, 3.002	
NH2	7.607,7.246			

[a]Signal for the terminal HMeN were not observed.

Table 2-19. The ^{1}H Chemical Shift Assignments for Peptide HPTFDmkAla

Residue	HN	Hα	Hβ	Others
Ac		2.077		
Cys1	8.444	5.207	3.165,2.657	
Arg2	8.766	4.668	1.839,1.682	Hγ:1.543 ; Hδ:3.183 ; HNt:7.130
Thr3	8.620	4.933	3.926	Hγ: 1.061
Val4	9.089	4.415	1.988	Hγ:0.855,0.816
Dmk5[a]	8.498	4.953	1.667	Hγ:1.178 ; Hδ:1.324 ;Hε:3.016
Val6	8.708	4.562	1.943	Hγ:0.914,0.885
ᴰPro7		4.376	2.351,1.960	Hγ:2.135,2.039 ; Hδ:3.858,3.768
Gly8	8.663	3.965,3.847		
Orn9	7.939	4.623	1.843,1.823	Hγ:1.686 ; Hδ:3.007 ; HNt:7.613
Ala10	8.532	4.842	1.218	
Ile11	9.108	4.464	1.862	Hγ:1.358,1.145,0.855(Me); Hδ:0.789
Leu12	8.405	4.737	1.661,1.647	Hγ:1.497 ; Hδ:0.816,0.787
Gln13	9.210	4.659	2.085,1.886	Hγ:2.280,2.218 ; HNt:7.319,6.830
Cys14	8.982	5.075	3.134,3.007	
NH2	7.613,7.241			

[a]Signal for the terminal HMeN were not observed.
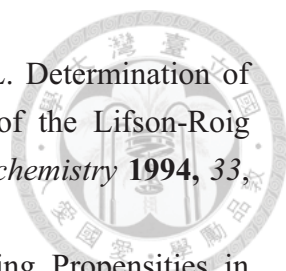
Table 2-20. The ^{1}H Chemical Shift Assignments for Peptide HPTFTmkAla

Residue	HN	Hα	Hβ	Others
Ac		2.078		
Cys1	8.449	5.212,5.193	3.159,2.660	
Arg2	8.767	4.665	1.839,1.682	Hγ:1.545 ; Hδ:3.179 ; HNt:7.129
Thr3	8.616	4.929	3.931	Hγ: 1.056
Val4	9.098	4.411	1.992	Hγ:0.856
Tmk5[a]	8.522	4.939	1.687	Hγ:1.164 ; Hδ:1.317 ;Hε:3.212
Val6	8.694	4.565	1.952	Hγ:0.910,0.885
DPro7		4.377	2.348,1.961	Hγ:2.142,2.044 ; Hδ:3.862,3.765
Gly8	8.664	3.970,3.853		
Orn9	7.941	4.626	1.831	Hγ:1.688 ; Hδ:3.003 ; HNt:7.627
Ala10	8.537	4.851	1.213	
Ile11	9.089	4.460	1.861	Hγ: 1.360,1.144,0.861(Me); Hδ:0.798
Leu12	8.410	4.733	1.648	Hγ:1.491; Hδ:0.812.0.783
Gln13	9.207	4.665	2.074,1.887	Hγ:2.279,2.220 ; HNt:7.315,6.836
Cys14	8.982	5.075	3.139,3.003	
NH2	7.246,7.608			

[a]Signal for the terminal HMeN were not observed.

2-7 References
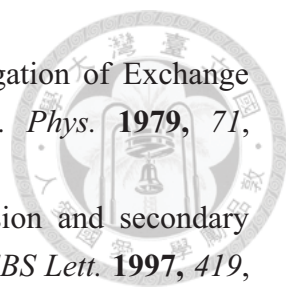
1. Cheng, R. P.; Girinath, P.; Suzuki, Y.; Kuo, H. T.; Hsu, H. C.; Wang, W. R.; Yang, P. A.; Gullickson, D.; Wu, C. H.; Koyack, M. J.; Chiu, H. P.; Weng, Y. J.; Hart, P.; Kokona, B.; Fairman, R.; Lin, T. E.; Barrett, O. Positional Effects on Helical Ala-Based Peptides. *Biochemistry* **2010,** *49*, 9372-9384.

2. Pauling, L.; Corey, R. B. The structure of synthetic polypeptides. *Proc. Natl. Acad. Sci. U. S. A.* **1951,** *37*, 241-250.

3. Doig, A. J. Recent advances in helix-coil theory. *Biophys. Chem.* **2002,** *101-102*, 281-93.

4. Chou, P. Y.; Fasman, G. D. Conformational Parameters for Amino-Acids in Helical, β-Sheet, and Random Coil Regions Calculated from Proteins. *Biochemistry* **1974,** *13*, 211-222.

5. Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. Helix Propensities of the Amino-Acids Measured in Alanine-Based Peptides without Helix-Stabilizing Side-Chain Interactions. *Protein Sci.* **1994,** *3*, 843-852.

6. Doig, A. J.; Chakrabartty, A.; Klingler, T. M.; Baldwin, R. L. Determination of Free-Energies of N-Capping in α-Helices by Modification of the Lifson-Roig Helix-Coil Theory to Include N-Capping and C-Capping. *Biochemistry* **1994,** *33*, 3396-3403.

7. Chakrabartty, A.; Doig, A. J.; Baldwin, R. L. Helix Capping Propensities in Peptides Parallel Those in Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1993,** *90*, 11332-11336.

8. Ermolenko, D. N.; Richardson, J. M.; Makhatadze, G. I. Noncharged amino acid residues at the solvent-exposed positions in the middle and at the C terminus of the α-helix have the same helical propensity. *Protein Sci.* **2003,** *12*, 1169-1176.

9. Zimm, B. H.; Bragg, J. K. Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains. *J. Chem. Phys.* **1959,** *31*, 526-535.

10. Lifson, S. Theory of Helix-Coil Transition in Polypeptides. *J. Chem. Phys.* **1961,** *34*, 1963-1974.

11. Sali, D.; Bycroft, M.; Fersht, A. R. Stabilization of Protein-Structure by Interaction of α-Helix Dipole with a Charged Side-Chain. *Nature* **1988,** *335*, 740-743.

12. Vila, J.; Alessandrini, J. L. Adsorption of an Ionic Homopolypeptide to a Flat Surface Induced by pH Change. *J. Theor. Biol.* **1988,** *134*, 445-449.

13. Robert, C. H. A Hierarchical Nesting Approach to Describe the Stability of α-Helices with Side-Chain Interactions. *Biopolymers* **1990,** *30*, 335-347.

14. Shalongo, W.; Stellwagen, E. Incorporation of Pairwise Interactions into the Lifson-Roig Model for Helix Prediction. *Protein Sci.* **1995,** *4*, 1161-1166.

15. Qian, H.; Schellman, J. A. Helix-Coil Theories - a Comparative-Study for Finite Length Polypeptides. *J. Phys. Chem.* **1992,** *96*, 3987-3994.

16. Pauling, L.; Corey, R. B. The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl. Acad. Sci. U. S. A.* **1951,** *37*, 251-256.

17. Richardson, J. S. β-Sheet topology and the relatedness of proteins. *Nature* **1977,** *268*, 495-500.

18. Kuo, L. H.; Li, J. H.; Kuo, H. T.; Hung, C. Y.; Tsai, H. Y.; Chiu, W. C.; Wu, C. H.; Wang, W. R.; Yang, P. A.; Yao, Y. C.; Wong, T. W.; Huang, S. J.; Huang, S. L.; Cheng, R. P. Effect of charged amino acid side chain length at non-hydrogen bonded strand positions on β-hairpin stability. *Biochemistry* **2013,** *52*, 7785-7797.

19. Kim, C. W. A.; Berg, J. M. Thermodynamic β-Sheet Propensities Measured Using a Zinc-Finger Host Peptide. *Nature* **1993,** *362*, 267-270.

20. Minor, D. L.; Kim, P. S. Measurement of the β-sheet-forming propensities of amino acids. *Nature* **1994,** *367*, 660-663.

21. Smith, C. K.; Withka, J. M.; Regan, L. A Thermodynamic Scale for the β-Sheet

Forming Tendencies of the Amino-Acids. *Biochemistry* **1994,** *33*, 5510-5517.

22. Syud, F. A.; Stanger, H. E.; Gellman, S. H. Interstrand side chain-side chain interactions in a designed β-hairpin: Significance of both lateral and diagonal pairings. *J. Am. Chem. Soc.* **2001,** *123*, 8667-8677.

23. Syud, F. A.; Espinosa, J. F.; Gellman, S. H. NMR-based quantification of β-sheet populations in aqueous solution through use of reference peptides for the folded and unfolded states. *J. Am. Chem. Soc.* **1999,** *121*, 11577-11578.

24. Espinosa, J. F.; Munoz, V.; Gellman, S. H. Interplay between hydrophobic cluster and loop propensity in β-hairpin formation. *J. Mol. Biol.* **2001,** *306*, 397-402.

25. Ciani, B.; Jourdan, M.; Searle, M. S. Stabilization of β-hairpin peptides by salt bridges: Role of preorganization in the energetic contribution of weak interactions. *J. Am. Chem. Soc.* **2003,** *125*, 9038-9047.

26. Ramirez-Alvarado, M.; Kortemme, T.; Blanco, F. J.; Serrano, L. β-hairpin and β-sheet formation in designed linear peptides. *Bioorg. Med. Chem.* **1999,** *7*, 93-103.

27. Sibanda, B. L.; Thornton, J. M. β-hairpin families in globular proteins. *Nature* **1985,** *316*, 170-174.

28. Young, V. R. Adult Amino-Acid-Requirements - the Case for a Major Revision in Current Recommendations. *J. Nutr.* **1994,** *124*, 1517-1523.

29. Paik, W. K.; Kim, S. Protein Methylation. *Science* **1971,** *174*, 114-119.

30. Cantoni, G. L. Biological Methylation - Selected Aspects. *Annu. Rev. Biochem.* **1975,** *44*, 435-451.

31. Nakayam, J.; Rice, J. C.; Strahl, B. D.; Allis, C. D.; Grewal, S. I. S. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **2001,** *292*, 110-113.

32. Grewal, S. I. S.; Rice, J. C. Regulation of heterochromatin by histone methylation and small RNAs. *Curr. Opin. Cell. Biol.* **2004,** *16*, 230-238.

33. Zhang, X.; Wen, H.; Shi, X. B. Lysine methylation: beyond histones. *Acta Bioch. Bioph. Sin.* **2012,** *44*, 14-27.

34. Chiang, P. K.; Gordon, R. K.; Tal, J.; Zeng, G. C.; Doctor, B. P.; Pardhasaradhi, K.; McCann, P. P. S-adenosylmethionine and methylation. *FASEB J.* **1996,** *10*, 471-480.

35. Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. How to Measure and Predict the Molar Absorption-Coefficient of a Protein. *Protein Sci.* **1995,** *4*, 2411-2423.

36. Chakrabartty, A.; Kortemme, T.; Padmanabhan, S.; Baldwin, R. L. Aromatic Side-Chain Contribution to Far-Ultraviolet Circular-Dichroism of Helical Peptides and Its Effect on Measurement of Helix Propensities. *Biochemistry* **1993,** *32*,

5560-5565.

37. Edelhoch, H. Spectroscopic Determination of Tryptophan and Tyrosine in Proteins. *Biochemistry* **1967,** *6*, 1948-1954.

38. Almeida, A. M.; Li, R.; Gellman, S. H. Parallel β-Sheet Secondary Structure Is Stabilized and Terminated by Interstrand Disulfide Cross-Linking. *J. Am. Chem. Soc.* **2012,** *134*, 75-78.

39. Searle, M. S.; Griffiths-Jones, S. R.; Skinner-Smith, H. Energetics of weak interactions in a β-hairpin peptide: Electrostatic and hydrophobic contributions to stability from lysine salt bridges. *J. Am. Chem. Soc.* **1999,** *121*, 11615-11620.

40. Tatko, C. D.; Waters, M. L. Selective aromatic interactions in β-hairpin peptides. *J. Am. Chem. Soc.* **2002,** *124*, 9372-9373.

41. Tatko, C. D.; Waters, M. L. The geometry and efficacy of cation-pi interactions in a diagonal position of a designed β-hairpin. *Protein Sci.* **2003,** *12*, 2443-2452.

42. Fisk, J. D.; Powell, D. R.; Gellman, S. H. Control of hairpin formation via proline configuration in parallel β-sheet model systems. *J. Am. Chem. Soc.* **2000,** *122*, 5443-5447.

43. Fields, G. B.; Noble, R. L. Solid-Phase Peptide-Synthesis Utilizing 9-Fluorenylmethoxycarbonyl Amino-Acids. *Int. J. Pept. Prot. Res.* **1990,** *35*, 161-214.

44. Volkmer-Engert, R.; Landgraf, C.; Schneider-Mergener, J. Charcoal surface-assisted catalysis of intramolecular disulfide bond formation in peptides. *J. Pept. Res.* **1998,** *51*, 365-369.

45. Chiu, H. P.; Suzuki, Y.; Gullickson, D.; Ahmad, R.; Kokona, B.; Fairman, R.; Cheng, R. P. Helix propensity of highly fluorinated amino acids. *J. Am. Chem. Soc.* **2006,** *128*, 15556-15557.

46. Doig, A. J.; Baldwin, R. L. N- and C-Capping Preferences for All 20 Amino-Acids in α-Helical Peptides. *Protein Sci.* **1995,** *4*, 1325-1336.

47. Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. Relative Helix-Forming Tendencies of Nonpolar Amino-Acids. *Nature* **1990,** *344*, 268-270.

48. Chiu, H. P.; Cheng, R. P. Chemoenzymatic Synthesis of (S)-Hexafluoroleucine and (S)-Tetrafluoroleucine. *Org Lett* **2007,** *9*, 5517-5520.

49. Chang, C. T.; Wu, C. S. C.; Yang, J. T. Circular Dichroic Analysis of Protein Conformation - Inclusion of β-Turns. *Anal. Biochem.* **1978,** *91*, 13-31.

50. Aue, W. P.; Bartholdi, E.; Ernst, R. R. 2-Dimensional Spectroscopy - Application to Nuclear Magnetic-Resonance. *J. Chem. Phys.* **1976,** *64*, 2229-2246.

51. Bax, A.; Grzesiek, S. Methodological Advances in Protein NMR. *Acc. Chem. Res.* **1993,** *26*, 131-138.

52. Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. Investigation of Exchange Processes by 2-Dimensional NMR-Spectroscopy. *J. Chem. Phys.* **1979,** *71*, 4546-4553.

53. Yao, J.; Dyson, H. J.; Wright, P. E. Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett.* **1997,** *419*, 285-289.

54. Dalgarno, D. C.; Levine, B. A.; Williams, R. J. P. Structural Information from NMR Secondary Chemical-Shifts of Peptide α-C-H Protons in Proteins. *Biosci. Rep.* **1983,** *3*, 443-452.

55. Kim, Y. M.; Prestegard, J. H. Measurement of Vicinal Couplings from Cross Peaks in Cosy Spectra. *J. Magn. Reson.* **1989,** *84*, 9-13.

56. Groebke, K.; Renold, P.; Tsang, K. Y.; Allen, T. J.; McClure, K. F.; Kemp, D. S. Template-nucleated alanine-lysine helices are stabilized by position-dependent interactions between the lysine side chain and the helix barrel. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93*, 4025-4029.

57. Yu, Y. X.; Song, C. Y.; Zhang, Q. Y.; DiMaggio, P. A.; Garcia, B. A.; York, A.; Carey, M. F.; Grunstein, M. Histone H3 Lysine 56 Methylation Regulates DNA Replication through Its Interaction with PCNA. *Mol. Cell* **2012,** *46*, 7-17.

58. Xu, C.; Bian, C. B.; Yang, W.; Galka, M.; Hui, O. Y.; Chen, C.; Qiu, W.; Liu, H. D.; Jones, A. E.; MacKenzie, F.; Pan, P.; Li, S. S. C.; Wang, H. B.; Min, J. R. Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107*, 19266-19271.

59. Piotto, M.; Saudek, V.; Sklenar, V. Gradient-Tailored Excitation for Single-Quantum NMR-Spectroscopy of Aqueous-Solutions. *J. Biomol. NMR* **1992,** *2*, 661-665.

60. Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V. Gradient-Tailored Water Suppression for H^1-N^{15} HSQC Experiments Optimized to Retain Full Sensitivity. *J. Magn. Reson., Ser A* **1993,** *102*, 241-245.

Chapter 3

Effect of Lysine Methylation on RNA Recognition and Cellular Uptake by Tat Derivatives

Chapter 3.

3-1 Introduction

Ribonucleic acid (RNA)

Ribonucleic acid (RNA) is a crucial biopolymer in living organisms. In the central dogma of molecular biology,[1] deoxyribonucleic acid (DNA) is transcribed to RNA, which is translated to protein. Proteins are the end result of the cellular genetic informational flow, serving as the basic fundamental building block of most cell activities. RNA is composed of nucleotides and is usually single-stranded. Four different nucleotides are used for building RNA, varying in the nucleobases attached to a ribose sugar: adenine (A), uracil (U), cytosine (C), and guanine (G) (Figure 3-1). The nucleobase and phosphate group are attached to the C1' and C5' carbon of the ribose, respectively. The C3' phosphate group of the former nucleotide and the C5' hydroxyl group of the latter one form a phosphodiester to link adjacent nucleotides (Figure 3-2).
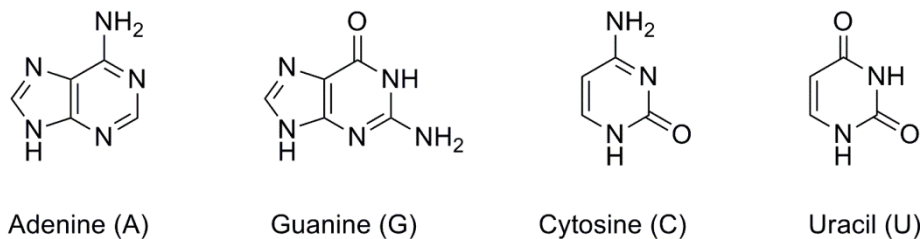
Figure 3-1. The The chemical structure of four nucleobases for RNA.
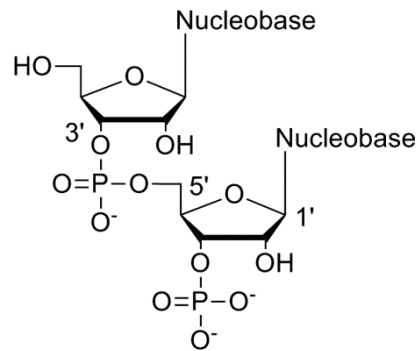
Figure 3-2. The basic constitution of a single-stranded RNA.

A single-stranded RNA can bend and form different three dimensional secondary structures through complementary base pairing.[2] These structures are generally divided into helices, various types of loops, and their combinations (Figure 3-3).[2] Three types of RNA are essential in protein synthesis: messenger RNA (mRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA). mRNA carries a series of codons to determine the protein sequence. rRNA and ribosomal proteins are responsible for protein synthesis in the ribosome. tRNA modulates the synthetic process via matching the codons on mRNA and the anticodons on tRNA in order to synthesize proteins with the correct sequence.
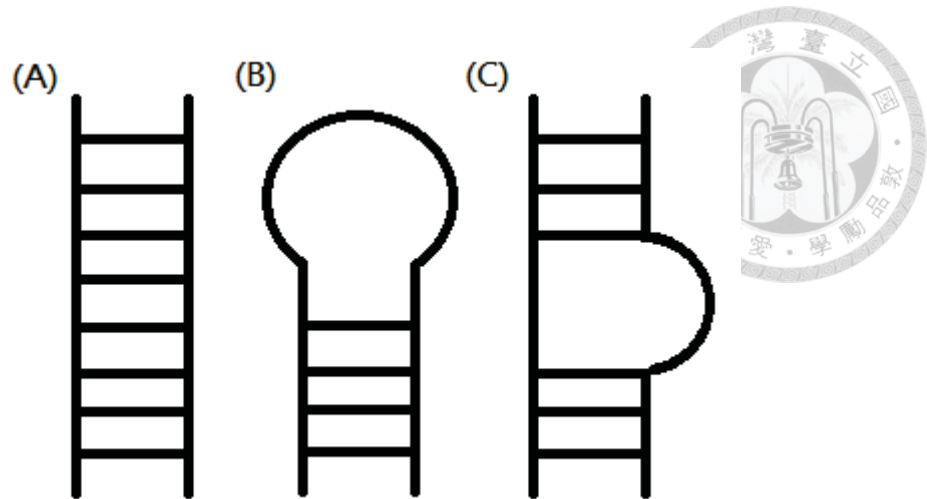
Figure 3-3. Various RNA secondary structures: helix (A), stem-loop (B), and bulge loop (C).

Human Immunodeficiency Virus (HIV)

HIV is a type of retrovirus that causes acquired immunodeficiency syndrome (AIDS).[3, 4] HIV infects vital immune cells such as helper T cells, marcophages, and dendritic cells.[4] The mechanism of HIV infection involves many steps.[5, 6] The surface trimeric envelope protein on the virus, gp120, engages an immune cell surface receptor CD4, inducing a conformational change in gp120.[7] This interaction results in the exposure of the gp120 chemokine binding domain, enabling the interaction with the host cell chemokine receptor (e.g., CCR5 and CXCR4).[7] The series of changes on gp120 give rise to the stable attachment with the host cell and allows the protein gp41 (fusion domain) to interact with the cell surface fusion receptor and penetrate the cell membrane.[6, 8-10] Once the virion enters the host cell, infection initiates.

87

In a retrovirus infection, viral RNA is reverse transcribed to the corresponding complementary DNA (cDNA) by a reverse transcriptase (RT).[11] HIV virus uses this pathway to convert the viral RNA genome into DNA.[12] This new DNA is further incorporated into the host cell genome by an intergrase and becomes a provirus.[5, 13] The viral genome is expressed along with the host cell genomes in a normal cell cycle, producing the viral proteins required to assemble new copies of the virus and infect other normal cells.[5, 14, 15]

Three common genes in the viral RNA genome are involved in retrovirus replication: *gag*, *pol*, and *env* (Figure 3-4).[16, 17] These genes are responsible for expressing the structural proteins and the reverse transcriptase for new virus particles.[18] There are additional six genes (*vif*, *vpu*, *vpr*, *tat*, *rev*, and *nef*) that code certain proteins to help the proliferation of HIV infection.[16] One of the regulatory genes *tat* encodes for the Tat protein. The RNA-protein interactions between the Tat protein and the trans-activation responsive element (TAR) on viral RNA regulate the RNA transcription activity by enhancing the transcription of the full-length viral RNA.[19]
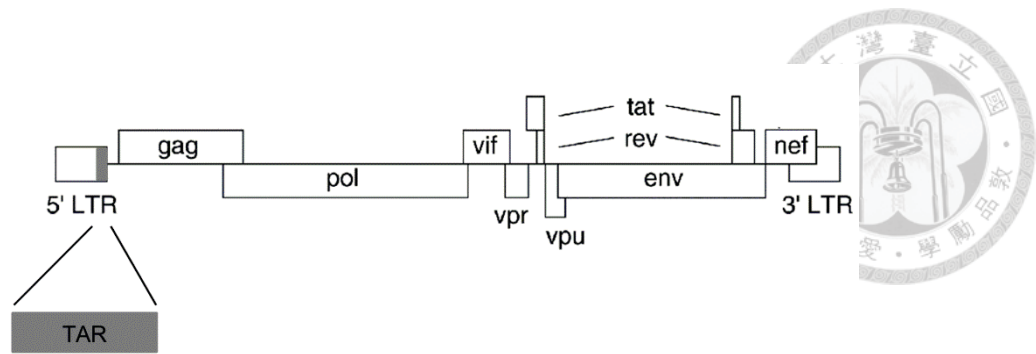
Figure 3-4. The landmark of HIV-1 genome consists of nine essential genes. The trans-activation response element (TAR, the fixed box on bottom left) located at the viral 5' LTR promoter and the trans-activator of transcription (Tat) protein displayed in the right hand side.[17] This figure was reprinted from Nucleic Acid Research, volume 33, by Westerhout, E. M.; Ooms, M.; Vink, M.; Das, A. T.; Berkhout, B. HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome, page 796-804, Copyright (2008), with permission from Oxford University Press.

Trans-Activation Response Element (TAR) RNA

A certain RNA region serves to stimulate gene expression during HIV-1 infection. This region is called the trans-activation response element (TAR) because it needs the Tat protein binding to mediate the trans-regulation. TAR RNA is a 59-nucleotide sequence located between positions -17 to +80 at the 5'-end of the viral long terminal repeat (5' LTR) promoter.[20-23] The TAR RNA possesses a stem-loop and a bulge, from positions +17 to +45.[24] The unique secondary structures are important for the transcription of the HIV virus.[24] The stem-loop contains a conserved 5'-CUGGG-3' sequence at positions +30 to +35, and the bulge contains a 5'-UCU-3' sequence at positions +23 to +25 (Figure 3-5).[24] The function of TAR RNA is closely related to its different structural regions.[25, 26] The trinucleotide bulge region and its adjacent base

pairs are responsible for both trans-activation and the specific binding with the Tat protein.[27-33] The loop region affects the level of transcription.[34-38]
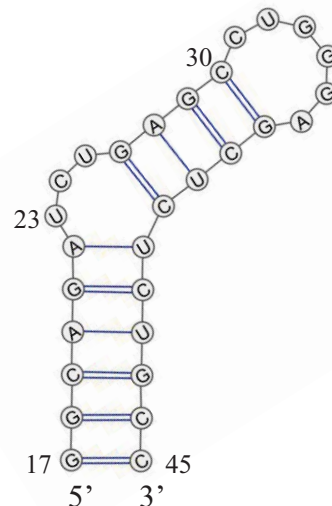
Figure 3-5. The sequence and secondary structure of HIV-1 from +17 to +45. This region, TAR RNA, contains a bulge and a loop structures, from +23 to +25 and from +30 to +35, respectively.[24]

Trans-Activator of Transcription (Tat) Protein

The Tat protein is a crucial element for viral transcription conserved among all lentiviruses.[39] The Tat protein is a small nuclear protein consisting between 86 and 101 amino acids depending on the subtype.[40] The Tat protein can be mainly divided into five domains:[41, 42] the N-terminal segment (residues 1-21), rich in acidic residues; the cysteine-rich region (residues 22-37); the core region (residues 38-48); the basic region (residues 49-72); and the C-terminal segment (Figure 3-6). The basic region contains a positive charge rich sequence RKKRRQRRR, which is responsible for both specific
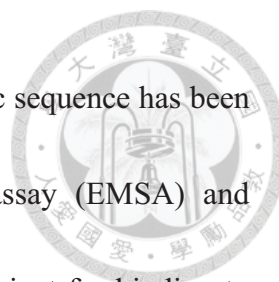
binding to TAR RNA and nuclear localization.[19, 32, 43, 44] This specific sequence has been

studied for Tat-TAR binding by electrophoretic mobility shift assay (EMSA) and

showed that the short peptides containing residues 49 to 57 is sufficient for binding to

TAR RNA, indicating its importance for the Tat protein.[32] Other highly positively

charged peptides has been tested and failed to bind TAR RNA, indicating that the

specificity of Tat-TAR interaction was not merely from the basic nature of the

sequence.[32] Another characteristic besides TAR regulation is the translocation across the

cell membrane.[45] Tat protein contains a protein transduction domain, allowing it to leave

the host cell and enter neighboring cells.[45] The extracellular Tat protein can induce

apoptosis of neighboring uninfected immune cells and the pathogenesis of HIV-1.[46]
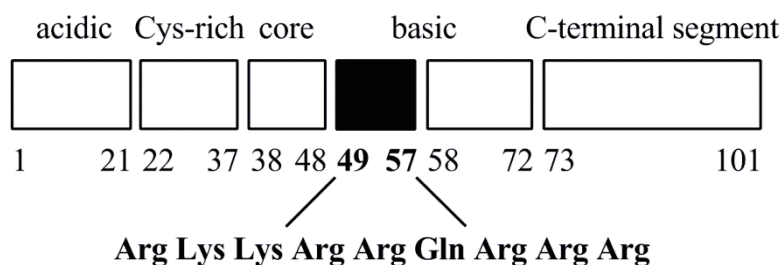
Figure 3-6. A schematic illustration of the Tat protein.[40, 41]

Tat-Mediated Transcription

Viral DNA is transcribed into mRNA by host cell RNA polymerase II (RNAPII)

during HIV infection. However, the transcription cannot produce full-length viral RNA

efficiently due to a stable RNA stem-loop structure called the trans-activation response

(TAR) element.[47] This structure is located at the 5' end of the HIV mRNA in nascent viral transcripts and would block the production of full-length viral RNA.[25, 47] Only a small number of RNA transcripts can be transcribed before Tat protein reaches a certain threshold amount. The complete elongation acquires the presence of Tat protein to interact with TAR RNA.[47, 48]

The bulge and loop regions of TAR RNA are involved in Tat protein recognition.[47, 49] Tat protein recruits a positive transcription elongation complex (P-TEFb), composed of cyclin-dependent kinase (CDK9) and cyclin T1.[50] Upon binding to TAR RNA, the C-terminal domain of RNAPII and two negative elongation factors DSIF and NELF are phosphorylated by CDK9.[16] The phosphorylation of the Tat-P-TEFb complex enhances the transcriptional elongation and allows the production of the full-length HIV RNA.[51, 52]

The Tat-TAR binding is important for HIV-1 proliferation (Figure 3-7).[47, 53] Therefore, targeting this interaction would be a potential treatment by developing inhibitors towards either Tat protein or TAR RNA.[54] The Tat_{49-57} derived peptides can be competitive inhibitors towards Tat-TAR binding.[55-63] In this chapter, a series of Tat_{49-57} peptide derivatives were designed and synthesized to study the binding affinity between TAR RNA and these derived peptides by EMSA.
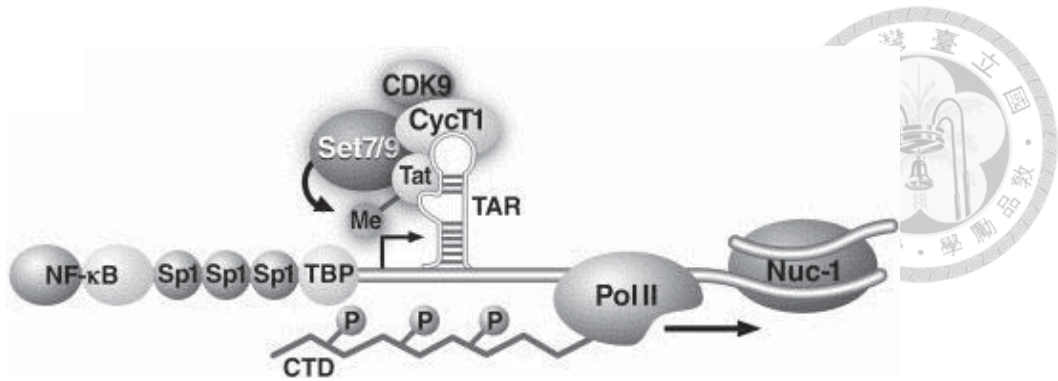
Figure 3-7. Trans-activated transcription of HIV-1 via Tat-TAR binding.[53] This figure is reprinted from Cell Hosts & Microbe, volume 7, by Pagans, S.; Kauder, S. E.; Kaehlcke, K.; Sakane, N.; Schroeder, S.; Dormeyer, W.; Trievel, R. C.; Verdin, E.; Schnolzer, M.; Ott, M. The Cellular Lysine Methyltransferase Set7/9-KMT7 Binds HIV-1 TAR RNA, Monomethylates the Viral Transactivator Tat, and Enhances HIV Transcription, page 234-244, Copyright (2010), with permission from Elsevier.

Lysine Methylation in Tat Protein

Protein lysine methylation is an important post-translational modification commonly found in various transcriptional regulations.[64] The methyltransferases Set7/9-KMT7 and SETDB1-KMT1E are associated with Tat protein and methylate both Lys50 and Lys51 in the basic region.[65] Monomethylation of Lys51 enhances transcriptional activity.[53] The methylation and acetylation at these two positions are responsible for the regulation of recruiting important protein complexes for HIV-1 transcription.[66]

Cell Penetration

Cell-penetrating peptides (CPPs) are short peptides that facilitate cellular uptake

and can carry various desired molecules through chemical linkage.[67] The amino acid composition of CPPs mainly contains abundant positively charged amino acids or hydrophobic amino acids. Others types are proline-rich CPPs, β-peptides, *etc* (Figure 3-8).[67] Tat peptide from the basic region of Tat protein: residues 49-57 contains many lysines and arginines and is thus a cationic CPP. As such, the Tat protein is capable of crossing the cell membrane to enter neighboring cells for HIV-1 proliferation.[68]

The mechanism for cellular uptake can be classified into endocytosis and non-endocytosis. Endocytosis is an energy-dependent process by which a cell absorbs larger polar molecules by engulfing them to form protein-coated vesicles. Another pathway, non-endocytosis is an energy-independent process in the form of fusion, penetration, microinjection, *etc*.[69]
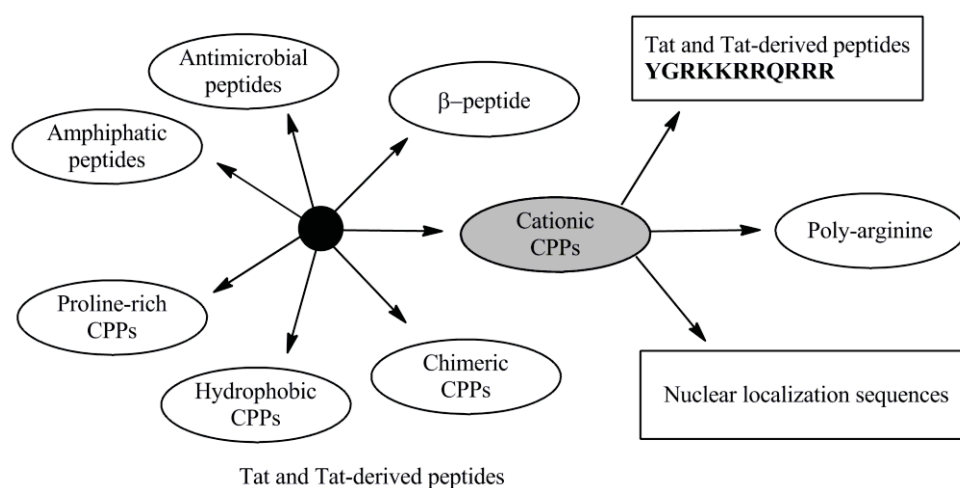
Figure 3-8. The classification of cell-penetrating peptides. [67]

94

The mechanisms by which CPPs enter the cells have not been completely understood.[70] However, the positively charged residues in Tat peptide must interact with the anionic cell membrane constituents such as carboxylates, phosphates, and sulfates.[71, 72] The cellular uptake efficiency of poly-arginine and Tat protein were all attenuated after the replacement of arginine with methylated arginines.[73] It is possible that the methylation at the other positively charged residue in the basic region: lysine, would also alter the cell penetrating ability.

3-2 Results and Discussion

Peptide Design and Synthesis

Peptides were designed based on the basic region of the wild-type Tat protein. The nine amino acids long sequence in the basic region, RKKRRQRRR (residue 49 to 57), is responsible for TAR recognition and crossing the cell membrane.[74] We replaced each individual lysine residue in the basic region of Tat protein with monomethyllysine (Mmk), dimethyllysine (Dmk), or trimethyllysine (Tmk) (Chart 3-1 and Table 3-1). Ac-capped peptides were for EMSA to determine the dissociation constant between TAR RNA and each Tat-derived peptide. In order to determine the cellular uptake efficiency of Tat-derived peptides by flow cytometry, the N-terminus of the peptides were coupled with β-alanine and 6-carboxy-fluorescein (Figure 3-9 and Table 3-2).[75]

All the peptides were synthesized by solid phase peptide synthesis using Fmoc-based chemistry (Fmoc = 9-fluorenylmethoxycarbonyl).[76, 77] The peptide was cleaved from the resin with concomitant side chain deprotection using trifluoroacetic acid (TFA) with 5% triisopropylsilane. All the peptides were purified by RP-HPLC to greater than 95% purity and confirmed by MALDI-TOF (Table 3-3). Peptide concentration was determined by Uv-vis spectroscopy.

Chart 3-1. Chemical Structure, Full Name and 3-Letter Code of Methylated Lysines

L-monomethyllysine   L-dimethyllysine   L-trimethyllysine
        Mmk   Dmk   Tmk
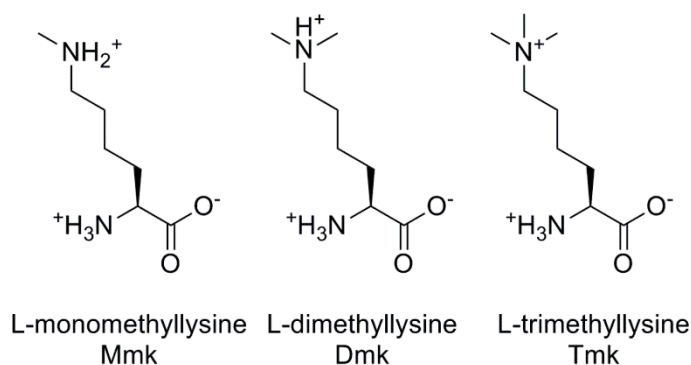
Table 3-1. The Sequence of Tat-Derived Peptides Capped with an Acetyl Group

Peptide	Sequence
ArgTat	Ac-Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg-NH$_2$
Ac-Mmk50-Tat	Ac-Tyr Gly Arg **Mmk** Lys Arg Arg Gln Arg Arg Arg-NH$_2$
Ac-Dmk50-Tat	Ac-Tyr Gly Arg **Dmk** Lys Arg Arg Gln Arg Arg Arg-NH$_2$
Ac-Tmk50-Tat	Ac-Tyr Gly Arg **Tmk** Lys Arg Arg Gln Arg Arg Arg-NH$_2$
Ac-Mmk51-Tat	Ac-Tyr Gly Arg Lys **Mmk** Arg Arg Gln Arg Arg Arg-NH$_2$
Ac-Dmk51-Tat	Ac-Tyr Gly Arg Lys **Dmk** Arg Arg Gln Arg Arg Arg-NH$_2$
Ac-Tmk51-Tat	Ac-Tyr Gly Arg Lys **Tmk** Arg Arg Gln Arg Arg Arg-NH$_2$
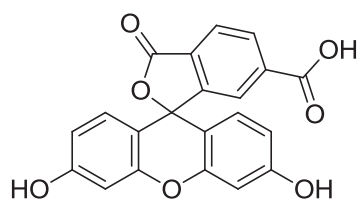
Figure 3-9. The chemical structure of 6-carboxy-fluorescein.

Table 3-2. The Sequence of Tat-Derived Peptides Capped with 6-Carboxy-Fuorescein

Peptide	Sequence[a]
Fl-ArgTat	Fl-βAla Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg-NH$_2$
Fl-Mmk50-Tat	Fl-βAla Tyr Gly Arg **Mmk** Lys Arg Arg Gln Arg Arg Arg-NH$_2$
Fl-Dmk50-Tat	Fl-βAla Tyr Gly Arg **Dmk** Lys Arg Arg Gln Arg Arg Arg-NH$_2$
Fl-Tmk50-Tat	Fl-βAla Tyr Gly Arg **Tmk** Lys Arg Arg Gln Arg Arg Arg-NH$_2$
Fl-Mmk51-Tat	Fl-βAla Tyr Gly Arg Lys **Mmk** Arg Arg Gln Arg Arg Arg-NH$_2$
Fl-Dmk51-Tat	Fl-βAla Tyr Gly Arg Lys **Dmk** Arg Arg Gln Arg Arg Arg-NH$_2$
Fl-Tmk51-Tat	Fl-βAla Tyr Gly Arg Lys **Tmk** Arg Arg Gln Arg Arg Arg-NH$_2$

[a]Fl = 6-carboxy-fluorescein.

Table 3-3. The Purity and Weight of the Tat-Derived Peptides

Peptide	Purity (%)	Weight (mg)
Ac-Mmk50-Tat	96.2	4.6
Ac-Dmk50-Tat	96.2	2.3
Ac-Tmk50-Tat	96.3	5.1
Ac-Mmk51-Tat	95.5	2.2
Ac-Dmk51-Tat	95.7	2.1
Ac-Tmk51-Tat	95.1	2.1
Fl-Mmk50-Tat	96.2	3.2
Fl-Dmk50-Tat	95.3	2.7
Fl-Tmk50-Tat	96.5	4.0
Fl-Mmk51-Tat	97.3	1.9
Fl-Dmk51-Tat	95.9	4.6
Fl-Tmk51-Tat	95.5	2.2

Reductive Methylation on Lysine

All the fluorescein-capped peptides were successfully synthesized by solid phase

peptide synthesis except for peptide Fl-Dmk51-Tat. A side reaction apparently occurred

between the modified amine side chain of Dmk and fluorescein. As such, it was difficult

97

to obtain sufficient amount of pure peptide (Figure 3-10A). This was because the side chain of commercial Dmk was unprotected (Chart 3-2). The side chains of other two Lys derivatives were in the form of a carbamate and a salt, respectively, and thus was not nucleophic (Chart 3-2).

Chart 3-2. The Chemical Structures of Commercially Available Methylated Lysines

Fmoc-Lys(Me,Boc)-OH   Fmoc-Lys(Me)₂-OH   Fmoc-Lys(Me₃Cl)-OH

(A)   (B)
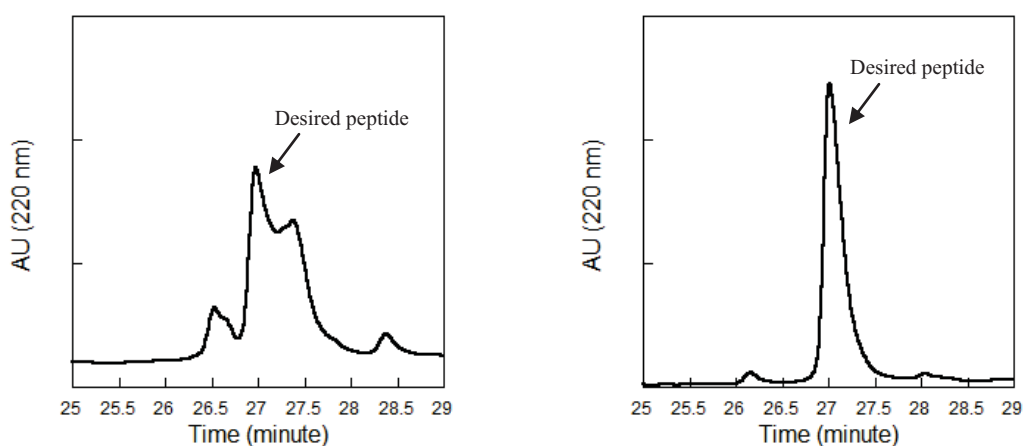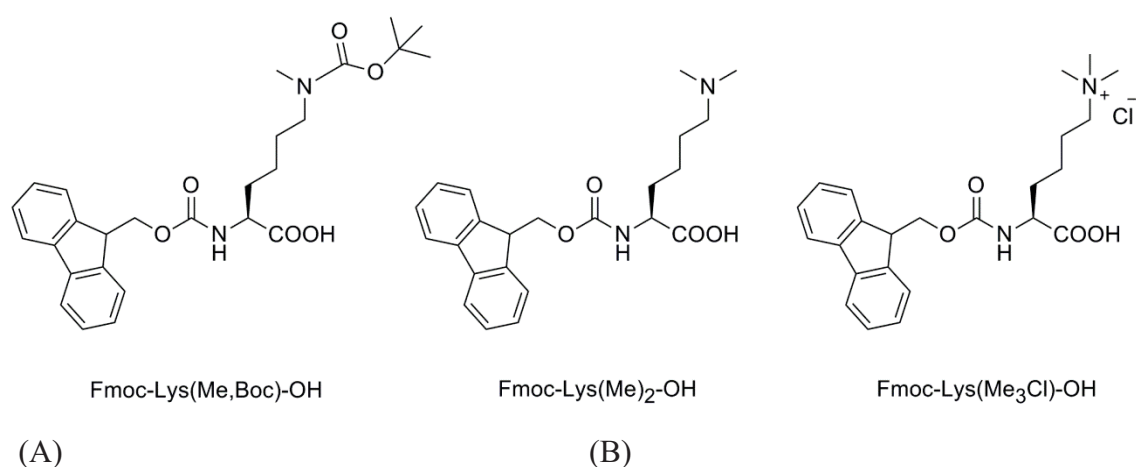
Figure 3-10. The analytical RP-HPLC chromatogram of peptide Fl-Dmk51-Tat synthesized using commercially available Dmk (A), and synthesized by reductive methylation (B).

A procedure was adopted to selectively introduce two methyl groups onto the desired Lys residue after solid phase peptide synthesis (SPPS) without the unwanted side reaction with the fluorescein moiety (Scheme 3-1). The precursor peptide Fl-Lys(ivDde)51-Tat was synthesized. The hydroxy group on the 6-carboxyfluorescein moiety was protected by reacting with trityl chloride.[78] The ivDde group was then selectively cleaved with hydrazine, and thus the further reactions can only occur on the desired position Lys51. The reductive methylation method was based on the procedures developed by Kaljuste and Undén in 1993.[79] Formaldehyde (CH_2O) was used as the carbon transfer unit in the presence of sodium cyanoborohydride ($NaBH_3CN$). All the reactions were conducted on the solid phase. The reaction would undergo twice and result in a tertiary amine if no additional protecting group existed on the substituted amine (Scheme 3-2). From this mechanism it is clear that there will be no further reaction after the attachment of two methyl group because it is impossible for a tertiary amine to form another imine or imine ion and become a quaternary ammonium salt. The analytical chromatogram of the product showed that the amount of impurities greatly decreased in the synthesis of peptide Fl-Dmk51-Tat by applying this method (Figure 3-10B).

Scheme 3-1. Synthesis of Fl-Dmk51-Tat via Reductive Methylation
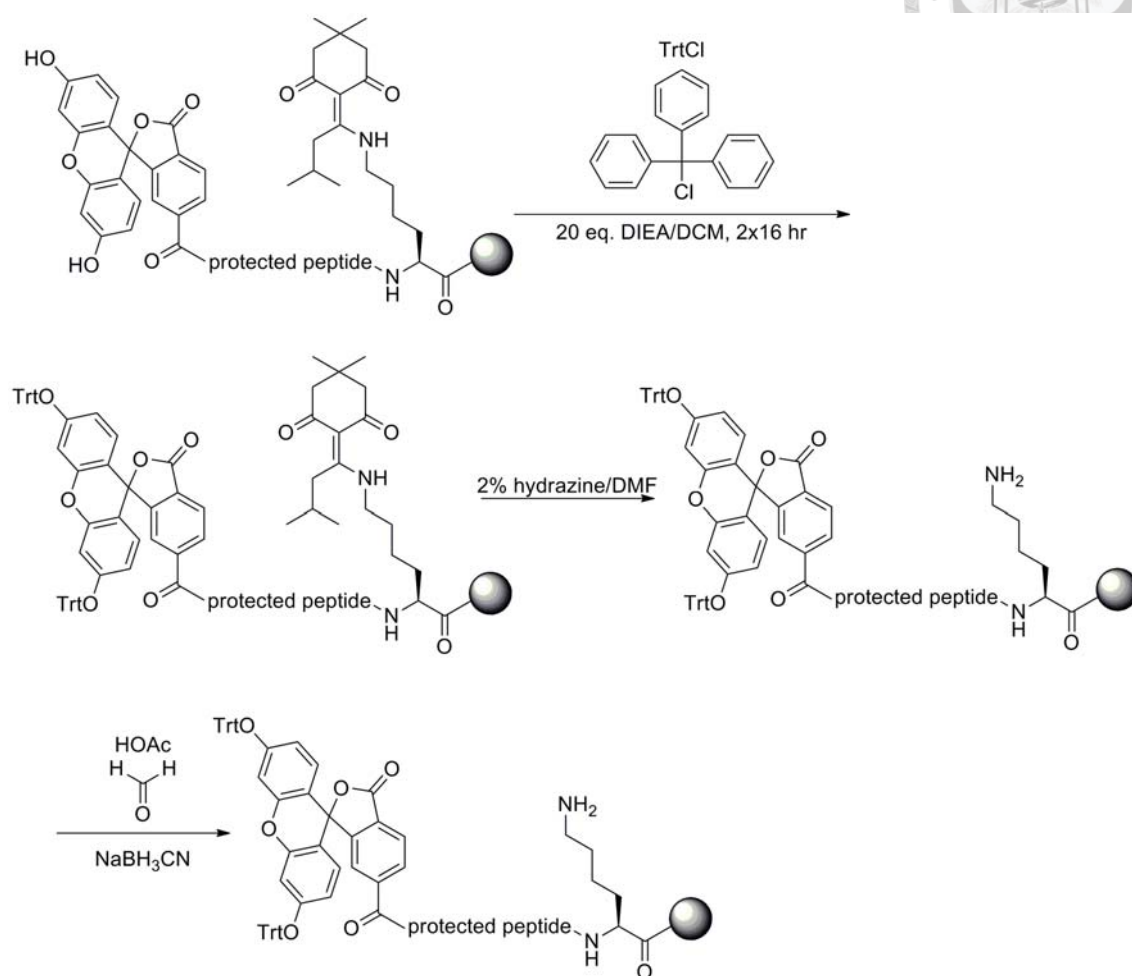
Scheme 3-2. Mechanism of Reductive Methylation
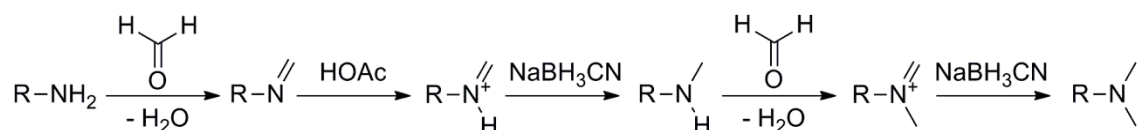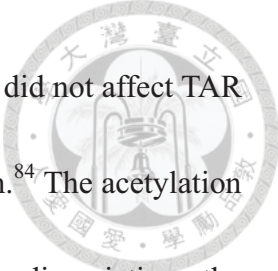
Electrophoretic Mobility Shift Assay in the Presence of Bulk *E. coli* tRNA

Electrophoretic mobility shift assay (EMSA), also known as band-shift assay or gel retardation assay, has been used to study the binding activity of purified DNA/RNA-binding proteins.[80] The basis of the assay is that DNA/RNA-protein complexes migrate more slowly than the free DNA/RNA fragment by gel

electrophoresis.[80, 81] EMSA was performed to determine the dissociation constant for Tat-derived peptides-TAR RNA complexes. There are many negatively charged molecules that exist in the cell and m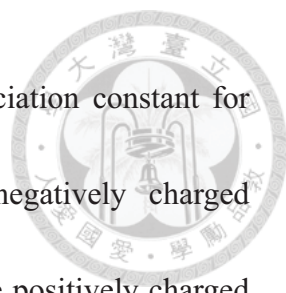ay bind non-specifically to the positively charged Tat-derived peptides. As such, the bulk *E. coli* tRNA were added to imitate the native cellular condition for investigating the binding between Tat-derived peptides and TAR RNA.[82] The assays were performed in 12% polyacrylamide under 140 V for 75 minutes at room temperature. Images of typical gels of EMSA for Tat-derived peptides are shown in Figure 3-11. The Tat-peptide concentration-fraction TAR RNA bound curve was fit globally to derive the apparent dissociation constants (Figure 3-12). The binding of Tat-derived peptides to TAR RNA was assumed to be 1:1 in stoichiometry.[83] The bulk *E. coli* tRNA was used as a non-specific competitor. Therefore, the apparent dissociation constant of Tat-derived peptides-TAR RNA complexes from EMSA in the presence of bulk *E. coli* tRNA should represent the specific binding of the Tat-derived peptide towards TAR RNA. The specific binding of the Tat-derived peptide to TAR RNA followed the trend: Ac-Tmk50-Tat ~ Ac-Dmk50-Tat > Ac-Mmk50-Tat ~ Ac-Dmk51-Tat > Ac-Mmk51-Tat > Ac-Tmk51-Tat. The Z and P values for comparing the apparent dissociation constant were calculated (Table 3-3). The K_D value of wild type Tat peptide (ArgTat) was previously determined by our lab to be 258 ± 27.[82]

Dimethylation and trimethylation of Lys50 resulted in higher binding affinity

towards TAR RNA than ArgTat, where as monomethylation of Lys50 did not affect TAR RNA binding. Interestingly, Lys50 is also acetylated in the Tat protein.[84] The acetylation of Lys50 is a later process in trans-activation transcription, dissociating the transcriptional complex by the suspension of the electrostatic interactions and the increased steric hindrance.[85] Lysine methylation strengthened the electrostatic interaction with RNA phosphate group via increase in charge effective radius.[85] Therefore, post-translational modification on residue Lys50 might possibly be a checkpoint in viral transcription, where methylation strengthened the Tat-TAR binding and acetylation dissociated the complex formed by Tat, TAR, and P-TEFb.

There was insignificant difference in TAR RNA binding between the wild type peptide ArgTat and the peptide with Lys51 dimethylated, whereas peptides with position 51 monomethylated and trimethylated exhibited decreased binding affinity towards TAR RNA. Accordingly, the Lys51 side chain may have hydrogen bonding interactions with oxygen atoms on several bases in the RNA loop section.[85, 86] Therefore, methylation should abolish this putative interaction when the hydrogen atom is replaced with methyl group. Multiple methylation of Lys51 has been shown to inhibit complex formation.[53] However, previous studies also suggested that methylation of Lys51, especially monomethylation, could increase the transcription level.[53] The residue Lys51 plays a central role in the trans-activation cycle which methylation resulted in either
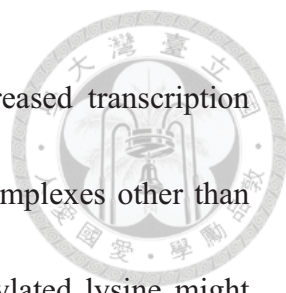
inhibitory or activating effects.[87] The weakened binding and increased transcription might be due strengthened binding to the transcriptional protein complexes other than with TAR RNA upon monomethylation of Lys51. The monomethylated lysine might create a new binding site for other transcriptional protein complexes in a TAR-independent manner. The increased dissociation constant may be compensated by the binding of monomethyltransferase Set7/9-KMT7 on TAR RNA, which monomethylates Lys51 and also binds to TAR RNA as a coactivator.[53] The dimethylation and trimethylation of Lys51 may interfere with the binding of the protein complex due to the steric hindrance and disabled hydrogen bonding. Multiple methylation has also been identified as a transcriptional silencer in histone regulation.[88] Therefore, Lys51 could serve as both an activation and an inhibition site regulated by different degrees of methylation.
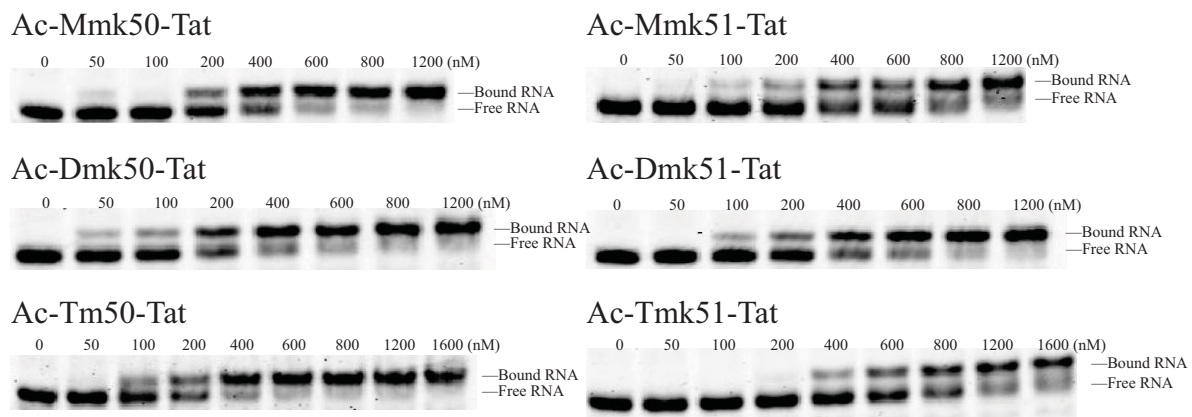
Figure 3-11. Images of typical gels of electrophoretic mobility shift assay (EMSA) for Tat-derived peptides. All lanes contain 100 nM fluorescein-labeled HIV-1 TAR RNA in the presence of 10 μg/mL bulk *E. coli* tRNA.
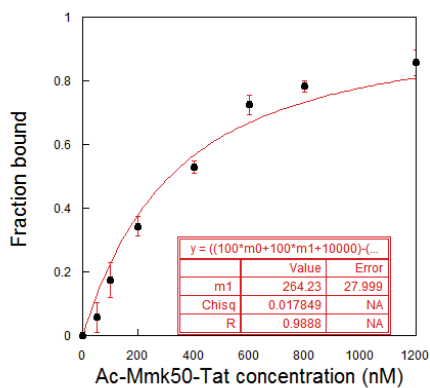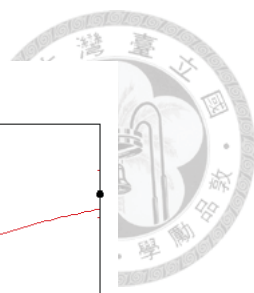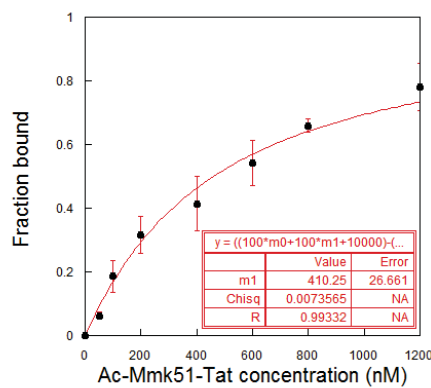
K_D= 264 ± 28 nM K_D= 410 ± 27 nM
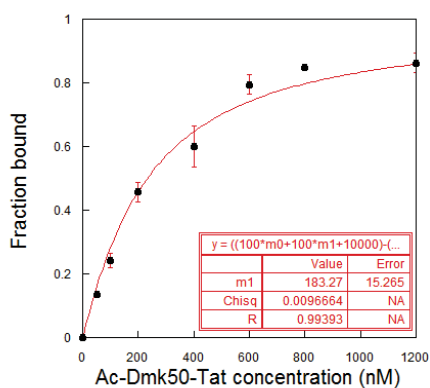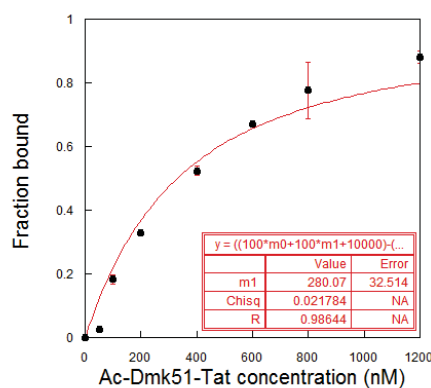
K_D= 183 ± 15 nM K_D= 280 ± 33 nM
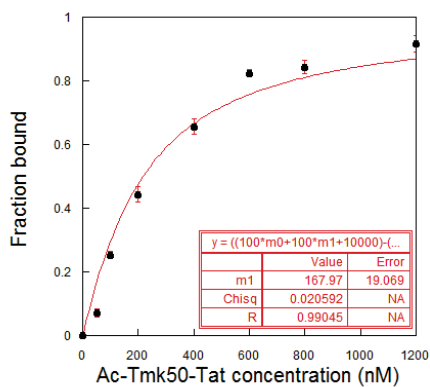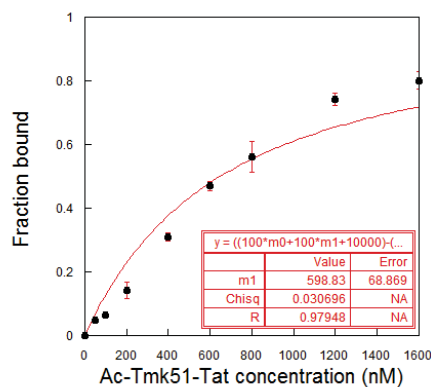
K_D= 168 ± 19 nM K_D= 599± 69 nM

Figure 3-12. The global fitting results of Tat-derived peptides binding to TAR RNA in the presence of 10 μg/mL bulk *E. coli* tRNA.

Table 3-4. The Apparent Dissociation Constants for Tat-Derived Peptides-TAR RNA Complexes in the Presence of 10 μg/mL bulk *E. coli* tRNA. TAR RNA concentration was 100 nM.

Peptide	Dissociation constant (nM)
Ac-Mmk50-Tat	264 ± 28
Ac-Dmk50-Tat	183 ± 15
Ac-Tmk50-Tat	168 ± 19
Ac-Mmk51-Tat	410 ± 27
Ac-Dmk51-Tat	280 ± 33
Ac-Tmk51-Tat	599 ± 69
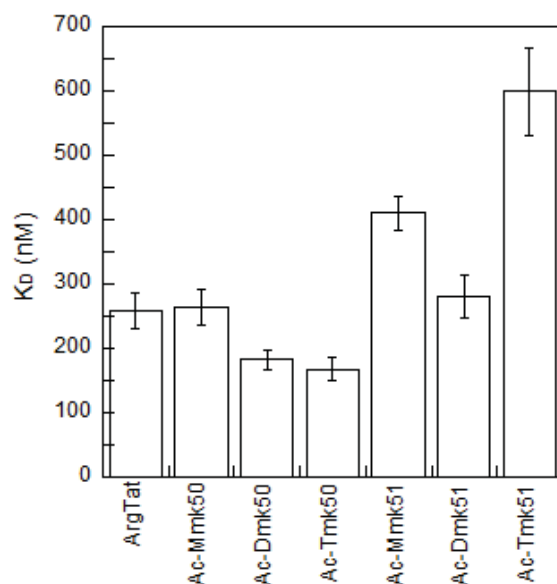
Figure 3-13. Apparent dissociation constants for Tat-derived peptide-TAR RNA complexes as determined by EMSA in the presence of 10 μg/mL bulk *E. coli* tRNA. TAR RNA concentration was 100 nM.

Table 3-5. The Z and P values for Comparing the Apparent Dissociation Constants of Wild Type Peptide and Tat-Derived Peptides in the Presence of 10 μg/mL bulk *E. coli* tRNA.

Peptides		Absolute Z value	P value
ArgTat	Ac-Mmk50-Tat	0.15	4.4×10^{-1}
ArgTat	Ac-Dmk50-Tat	2.4	7.6×10^{-3}
ArgTat	Ac-Tmk50-Tat	2.7	3.2×10^{-3}
ArgTat	Ac-Mmk51-Tat	4.0	3.5×10^{-5}
ArgTat	Ac-Dmk51-Tat	0.52	3.0×10^{-1}
ArgTat	Ac-Tmk51-Tat	4.6	$< 10^{-5}$
Ac-Mmk50-Tat	Ac-Dmk50-Tat	2.6	5.4×10^{-3}
Ac-Mmk50-Tat	Ac-Tmk50-Tat	2.8	2.3×10^{-3}
Ac-Mmk50-Tat	Ac-Mmk51-Tat	3.8	8.7×10^{-5}
Ac-Mmk50-Tat	Ac-Dmk51-Tat	0.37	3.6×10^{-1}
Ac-Mmk50-Tat	Ac-Tmk51-Tat	4.5	$< 10^{-5}$
Ac-Dmk50-Tat	Ac-Tmk50-Tat	0.62	2.7×10^{-1}
Ac-Dmk50-Tat	Ac-Mmk51-Tat	7.4	$< 10^{-5}$
Ac-Dmk50-Tat	Ac-Dmk51-Tat	2.7	3.7×10^{-3}
Ac-Dmk50-Tat	Ac-Tmk51-Tat	5.9	$< 10^{-5}$
Ac-Tmk50-Tat	Ac-Mmk51-Tat	7.3	$< 10^{-5}$
Ac-Tmk50-Tat	Ac-Dmk51-Tat	2.9	1.6×10^{-3}
Ac-Tmk50-Tat	Ac-Tmk51-Tat	6.0	$< 10^{-5}$
Ac-Mmk51-Tat	Ac-Dmk51-Tat	3.1	1.1×10^{-3}
Ac-Mmk51-Tat	Ac-Tmk51-Tat	2.6	5.4×10^{-3}
Ac-Dmk51-Tat	Ac-Tmk51-Tat	4.2	1.5×10^{-5}

Circular Dichorism Spectroscopy

Circular dichorism (CD) spectroscopy is a convincing tool for studying the secondary structure of peptides. The CD spectra for all the Tat-derived peptides were acquired at pH 7 and 25 °C (Figure 3-14). The results suggested that all the peptides were random coil, which was consistent with the previous reports on the natural Tat peptide structure.[89]
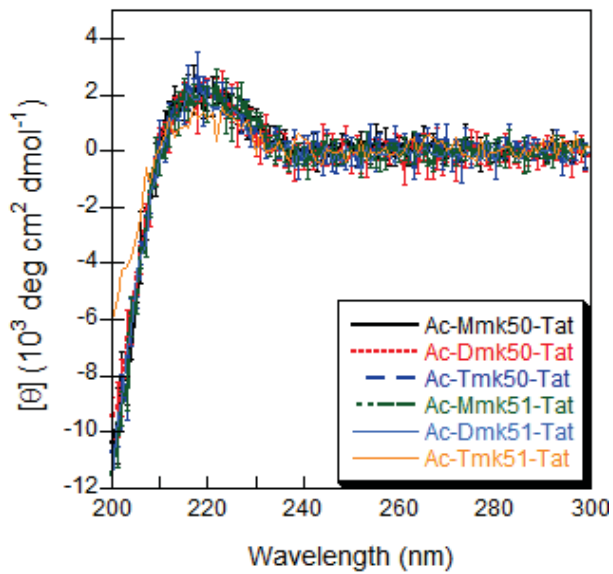
Figure 3-14. CD spectra between 200 to 300 nm of the Tat-derived peptides. The spectra were acquired in 10 mM Tris buffer at pH 7 and 25 $^\circ$C. Peptide concentration was 50 μM.

Cellular Uptake Assay

The Jurkat cells treated with 6-carboxy-fluorescein coupled Tat-derived peptides were analyzed by fluorescence-assisted cell sorting (FACS) to quantitatively measure the uptake of labeled Tat-derived peptides. The cellular uptake efficiency of Tat-derived peptides was based on the fluorescence intensity to measure the number of cells that have taken up the labeled peptides. The mean cellular fluorescence intensities measured upon incubating with 7 μM and 120 μM peptide are shown in Table 3-6. The results showed that the mean cellular fluorescence intensity was directly proportional to the incubated peptide concentration.

Table 3-6. Cellular Uptake of Tat-Derived Peptides Treated into Jurkat Cells in PBS. Mean fluorescence intensity for each peptide with 7 µM and 120 µM

Peptide	Mean fluorescence intensity (a.u.) 7 µM	Mean fluorescence intensity (a.u.) 120 µM
Fl-ArgTat	164 ± 5	8856 ± 607
Fl-Mmk50-Tat	154 ± 18	8403 ± 972
Fl-Dmk50-Tat	137 ± 20	9145 ± 515
Fl-Tmk50-Tat	163 ± 20	13112 ± 638
Fl-Mmk51-Tat	117 ± 16	6415 ± 564
Fl-Dmk51-Tat	89 ± 8	8037 ± 392
Fl-Tmk51-Tat	92 ± 15	8200 ± 511
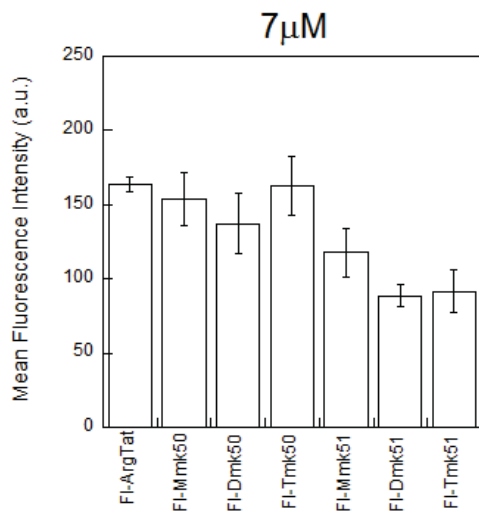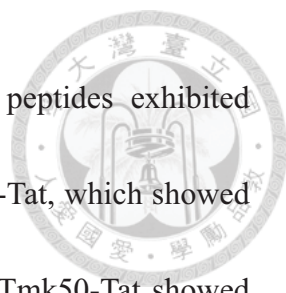
(A)   (B)
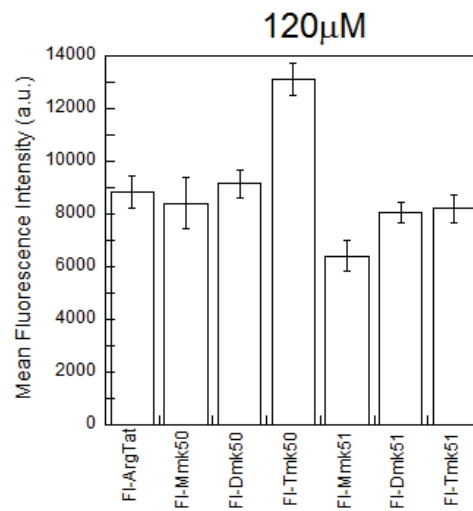
Figure 3-15. The mean fluorescence intensity of Jurkat cells treated with 7 µM (A) and 120 µM (B) Tat-derived peptides in cellular uptake assays.

The Z and P values of the mean cellular fluorescence intensity upon incubating with 7 µM and 120 µM peptide in cellular uptake assays are shown in Tables 3-7 and 3-8. Incubating with 7 µM peptide, the Tat-derived peptides exhibited less uptake compared to Fl-ArgTat except for peptide Fl-Tmk50-Tat, which appeared to be the same

as Fl-ArgTat. Incubating with 120 μM peptide, the Tat-derived peptides exhibited similar uptake compared to Fl-ArgTat except for peptide Fl-Mmk51-Tat, which showed less uptake compared to Fl-ArgTat. On the other hand, peptide Fl-Tmk50-Tat showed increased fluorescence intensity compared to Fl-ArgTat.

The uptake results for peptides with methylated Lys50 suggest that both hydrogen bond donors and hydrophobic interactions contributed to cellular uptake efficiency. The hydrogen bonding between the positively charged residues and the cell surface has been proved to be important for cellular uptake.[73] The interaction between the hydrophobic residues and the cell surface interaction also help non-polar CPPs to enter cells.[90] Introducing a methyl group reduces the hydrogen bonding ability and potentially disturbs the hydrogen bonding network formed around neighboring arginine residues. However, the increased hydrophobicity seemed to be able to compensate when all three hydrogens were replaced with methyl groups, and even resulted in significantly higher uptake efficiency at high peptide concentration (120 μM). In addition, the hydrophobic interactions might be site-specific because the Tat-derived peptides with methylated Lys51 exhibited less cellular uptake compared to Fl-ArgTat and Tat-derived peptides with methylated Lys50.

Table 3-7. The Z and P Value of the Mean Fluorescence Intensity at 7 µM for All Peptides of Cellular Uptake Assays

Peptides		Absolute Z value	P value
Fl-ArgTat	Fl-Mmk50-Tat	0.54	3.0×10^{-1}
Fl-ArgTat	Fl-Dmk50-Tat	1.3	9.5×10^{-2}
Fl-ArgTat	Fl-Tmk50-Tat	0.05	4.8×10^{-1}
Fl-ArgTat	Fl-Mmk51-Tat	2.8	2.5×10^{-3}
Fl-ArgTat	Fl-Dmk51-Tat	8.0	$< 10^{-5}$
Fl-ArgTat	Fl-Tmk51-Tat	4.6	$< 10^{-5}$
Fl-Mmk50-Tat	Fl-Dmk50-Tat	0.63	2.6×10^{-1}
Fl-Mmk50-Tat	Fl-Tmk50-Tat	0.33	3.7×10^{-1}
Fl-Mmk50-Tat	Fl-Mmk51-Tat	1.5	6.2×10^{-2}
Fl-Mmk50-Tat	Fl-Dmk51-Tat	3.3	5.0×10^{-4}
Fl-Mmk50-Tat	Fl-Tmk51-Tat	2.7	4.1×10^{-3}
Fl-Dmk50-Tat	Fl-Tmk50-Tat	0.92	1.8×10^{-1}
Fl-Dmk50-Tat	Fl-Mmk51-Tat	0.78	2.2×10^{-1}
Fl-Dmk50-Tat	Fl-Dmk51-Tat	2.2	1.3×10^{-2}
Fl-Dmk50-Tat	Fl-Tmk51-Tat	1.8	3.6×10^{-2}
Fl-Tmk50-Tat	Fl-Mmk51-Tat	1.8	3.6×10^{-2}
Fl-Tmk50-Tat	Fl-Dmk51-Tat	3.4	3.0×10^{-4}
Fl-Tmk50-Tat	Fl-Tmk51-Tat	2.8	2.3×10^{-3}
Fl-Mmk51-Tat	Fl-Dmk51-Tat	1.6	5.9×10^{-2}
Fl-Mmk51-Tat	Fl-Tmk51-Tat	1.1	1.3×10^{-1}
Fl-Dmk51-Tat	Fl-Tmk51-Tat	0.18	4.3×10^{-1}

Table 3-8. The Z and P Value of the Mean Fluorescence Intensity at 120 μM for All Peptides of Cellular Uptake Assays

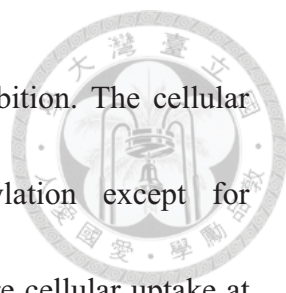| Peptides		Absolute Z value	P value
Fl-ArgTat	Fl-Mmk50-Tat	0.40	3.5×10^{-1}
Fl-ArgTat	Fl-Dmk50-Tat	0.36	3.6×10^{-1}
Fl-ArgTat	Fl-Tmk50-Tat	4.8	$< 10^{-5}$
Fl-ArgTat	Fl-Mmk51-Tat	3.0	1.6×10^{-3}
Fl-ArgTat	Fl-Dmk51-Tat	1.1	1.3×10^{-1}
Fl-ArgTat	Fl-Tmk51-Tat	0.83	2.0×10^{-1}
Fl-Mmk50-Tat	Fl-Dmk50-Tat	0.67	2.5×10^{-1}
Fl-Mmk50-Tat	Fl-Tmk50-Tat	4.1	2.6×10^{-5}
Fl-Mmk50-Tat	Fl-Mmk51-Tat	1.8	3.8×10^{-2}
Fl-Mmk50-Tat	Fl-Dmk51-Tat	0.35	3.6×10^{-1}
Fl-Mmk50-Tat	Fl-Tmk51-Tat	0.18	4.3×10^{-1}
Fl-Dmk50-Tat	Fl-Tmk50-Tat	4.8	$< 10^{-5}$
Fl-Dmk50-Tat	Fl-Mmk51-Tat	3.6	2.0×10^{-4}
Fl-Dmk50-Tat	Fl-Dmk51-Tat	1.7	4.4×10^{-2}
Fl-Dmk50-Tat	Fl-Tmk51-Tat	1.3	9.6×10^{-2}
Fl-Tmk50-Tat	Fl-Mmk51-Tat	7.9	$< 10^{-5}$
Fl-Tmk50-Tat	Fl-Dmk51-Tat	6.8	$< 10^{-5}$
Fl-Tmk50-Tat	Fl-Tmk51-Tat	6.0	$< 10^{-5}$
Fl-Mmk51-Tat	Fl-Dmk51-Tat	2.4	9.1×10^{-3}
Fl-Mmk51-Tat	Fl-Tmk51-Tat	2.4	9.5×10^{-3}
Fl-Dmk51-Tat	Fl-Tmk51-Tat	0.25	4.0×10^{-1}

3-3 Conclusion

The CD results showed that the Tat-derived peptides are random coil. The binding specificity of the Tat-derived peptide towards TAR RNA followed the trend: Ac-Tmk50-Tat ~ Ac-Dmk50-Tat > Ac-Mmk50-Tat ~ Ac-Dmk51-Tat > Ac-Mmk51-Tat > Ac-Tmk51-Tat. Methylation of Lys50 generally increased the binding specificity. Methylation of Lys51 exhibited varying effects on TAR RNA binding, suggesting that

methylation of this position may regulate both activation and inhibition. The cellular uptake at 7 μM Tat-derived peptide decreased upon methylation except for Fl-Mmk50-Tat and Fl-Tmk50-Tat. Also, Fl-Tmk50-Tat showed more cellular uptake at 120 μM compared to Fl-ArgTat. Importantly, trimethylation of Lys50 not only enhanced Tat-TAR binding but increased cellular uptake efficiency at 120 μM peptide. Therefore, this modification may be potentially useful for suppressing HIV-1 transcription.

3-4 Acknowledgement

3-5 Experimental Section

General Materials and Methods

Reagents and solvents were used without purification. All of the chemical reagents were purchased from Sigma Aldrich. Diisopropylethylamine (DIEA), piperidine,
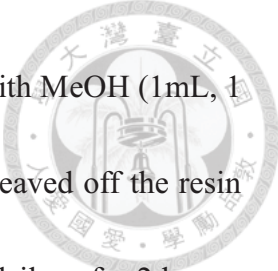
trifluoroacetic acid (TFA), and acetic anhydride (Ac$_2$O) were from Arcos. Dimethylformamide (DMF) and hexanes were from Mallinckrodt. Methanol and acetonitrile were from Merck. *N*-9-Fluorenylmethoxycarbonyl (Fmoc)-amino acids, 1-hydroxybenzotriazole (HOBt), *O*-1*H*-benzotriazol-1-yl-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), NovaSyn®TGR resin, and methylated lysines were from NovaBiochem. Fmoc-PAL-PEG-PS resin was from Applied Biosystems. Analytical reverse phase (RP)-HPLC was performed on an Agilent 1200 series chromatography system using a Vydac C18 column (4.6 mm diameter, 250 mm length). Preparative RP-HPLC was performed on a Waters Breeze chromatography system using a Vydac C4 or C18 column (22 nm diameter, 250 mm length). Mass spectrometry of the peptides was performed on a matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) spectrometer (Bruker BIFLEX) using α-cyano-4-hydroxycinnamic acid as the matrix. Determination of peptide concentration was performed on a UV-vis spectrometer (Jasco V-650). The EMSA results were monitored by a Typhoon TRIO$^+$ gel imager with 526 nm emission wavelength. Jurkat Cells were incubated using a CO$_2$ incubator (Thermo Scientific, Forma steri-cycle CO$_2$ incubator). Cells were counted using a hemacytometer (Reichert Bright-Line, hemacytometer 1490). The mean fluorescence intensity of Jurkat cells treated by 6-caroxy-fluorencein labled Tat-derived peptides were measured by a flow cytometer (Becton Dickinson, FACS Canto^{TM} II).

Peptide Synthesis

NovaSyn®TGR resin (0.05 mmol) was swollen in *N,N*-dimethylformamide (DMF, 5 mL) for 30 minutes before the first coupling. The resin was then washed with DMF (5 mL, 5x1 min). This was followed by Fmoc deprotection with 20% piperidine/DMF (5 mL, 3x8 min). The resin was subsequently washed with DMF (5 mL, 5x1 min). A mixture of 3 equivalents of the appropriately protected Fmoc amino acid, HOBt, and HBTU was dissolved in DMF (1 mL). Diisopropylethylamine (DIEA, 8 equivalents) was then added to the solution. The solution was then mixed thoroughly and should give a fruity ester smell and should turn yellow. The solution was then applied to the resin. The vial that contained the solution was rinsed with DMF (2x1 mL), and the DMF was added to the reaction. The coupling reaction was typically carried out for 45 minutes. The coupling times varied for different amino acids depending upon their positions in the sequence, the former residue, and the residue itself. The first amino acid was coupled for 8 hours. The 8th to 14th residues that were attached to the resin were coupled for 1.5 hours. For Ac-capped peptides, a solution of Ac_2O (20 equivalents), DIEA (20 equivalents), and DMF (3 mL) was added to the resin. The reaction was shaken for 2 hours. For fluorophore containing peptides, β-alanine and 6-carboxy-fluorescein (Fl) were coupled using standard protocol. When coupling the Fl, the reaction vessel should be wrapped with aluminium foil to avoid light. The resin was

subsequently washed with DMF (5 mL, 5x1 min) and then washed with MeOH (1mL, 1 min), and lyophilized overnight. The peptide was deprotected and cleaved off the resin by treating the resin with 95:5 trifluoroacetic acid (TFA) /triisopropylsilane for 2 hours. The reaction was then filtered through glass wool and the resin was washed with TFA. The combined filtrate was then evaporated by a gentle stream of N_2. The resulting oil was washed with 3 mL hexanes, dissolved in water, and lyophilized. The peptides (1 mg mL^{-1} aqueous solution) were analyzed using an analytical C18 column using 1 mL/min flow rate, linear 1%/min gradient from 100% A to 0% A (solvent A: 99.9% water, 0.1%TFA; solvent B: 90% acetonitrile, 10% water, 0.1% TFA). Appropriate linear solvent A/ solvent B gradients were used for purification on RP-HPLC preparative C4 and C18 columns. The identity of the peptides was confirmed by MALDI-TOF.

Ac-Mmk50-Tat

(Ac-Tyr-Gly-Arg-**Mmk**-Lys-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The peptide was synthesized using 98.0 mg (0.025 mmol) of NovaSyn®TGR resin. The synthesis gave 177.9 mg of resin (96.2% yield). Large scale cleavage gave 38.8 mg of crude peptide (79.5% yield). Retention time on analytical RP-HPLC was 17.3 minutes. The peptide was purified by Sep-Pak® Plus tC18® and then preparative RP-HPLC using a C18 (PLG 15_26) column. The final weight and purity were 4.6 mg

and 96.5%, respectively. The identity of the peptide was confirmed by MALDI-TOF

mass spectrometry. Calculated for $C_{67}H_{123}N_{33}O_{14}$ $[MH]^+$: 1614.89; observed : 1614.25.

Ac-Dmk50-Tat

(Ac-Tyr-Gly-Arg-**Dmk**-Lys-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The peptide was synthesized using 94.6 mg (0.024 mmol) of NovaSyn®TGR resin.

The synthesis gave 173.4 mg of resin (97.9% yield). Large scale cleavage gave 26.9 mg

of crude peptide (56.0% yield). Retention time on analytical RP-HPLC was 17.4

minutes. The peptide was purified by Sep-Pak® Plus tC18 and then preparative

RP-HPLC using a C18 (PLG 15_26) column. The final weight and purity were 2.3 mg

and 96.2%, respectively. The identity of the peptide was confirmed by MALDI-TOF

mass spectrometry. Calculated for $C_{68}H_{125}N_{33}O_{14}$ $[MH]^+$: 1628.89; observed : 1628.371.

Ac-Tmk50-Tat

(Ac-Tyr-Gly-Arg-**Tmk**-Lys-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The peptide was synthesized using 88.3 mg (0.022 mmol) of NovaSyn®TGR resin.

The synthesis gave 160.4 mg of resin (95.6% yield). Large scale cleavage gave 25.3 mg

of crude peptide (57.7% yield). Retention time on analytical RP-HPLC was 17.2

minutes. The peptide was purified by Sep-Pak® Plus tC18 and then preparative

RP-HPLC using a C18 (PLG 15_26) column. The final weight and purity were 5.1 mg

and 96.3%, respectively. The identity of the peptide was confirmed by MALDI-TOF

mass spectrometry. Calculated for $C_{69}H_{128}N_{33}O_{14}^{+}$ [M]$^{+}$: 1643.03; observed : 1641.48.

Ac-Mmk51-Tat

(Ac-Tyr-Gly-Arg-Lys-**Mmk**-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The peptide was synthesized using 115.0 mg (0.029 mmol) of NovaSyn®TGR resin.

The synthesis gave 194.6 mg of resin (81.7% yield). Large scale cleavage gave 35.7 mg

of crude peptide (71.9% yield). Retention time on analytical RP-HPLC was 17.0

minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 00_06) and

C18 (PLG 06_15) columns. The final weight and purity were 2.2 mg and 95.5%,

respectively. The identity of the peptide was confirmed by MALDI-TOF mass

spectrometry. Calculated for $C_{67}H_{123}N_{33}O_{14}$ [MH]$^{+}$: 1614.89; observed : 1615.10.

Ac-Dmk51-Tat

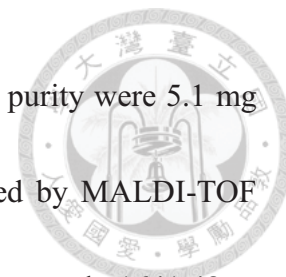(Ac-Tyr-Gly-Arg-Lys-**Dmk**-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The peptide was synthesized using 118.0 mg (0.030 mmol) of NovaSyn®TGR resin.

The synthesis gave 194.4 mg of resin (76.1% yield). Large scale cleavage gave 34.4 mg

of crude peptide (73.6% yield). Retention time on analytical RP-HPLC was 17.2

minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 00_06) and C18 (PLG 06_15) columns. The final weight and purity were 2.1 mg and 95.7%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{68}H_{125}N_{33}O_{14}$ [MH]$^+$: 1628.89; observed : 1630.20.

Ac-Tmk51-Tat

(Ac-Tyr-Gly-Arg-Lys-**Tmk**-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The peptide was synthesized using 120.6 mg (0.030 mmol) of NovaSyn$^®$TGR resin. The synthesis gave 217.2 mg of resin (93.8% yield). Large scale cleavage gave 37.0 mg of crude peptide (62.6% yield). Retention time on analytical RP-HPLC was 17.8 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 00_06) and C18 (PLG 06_15) columns. The final weight and purity were 2.1 mg and 95.1%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{69}H_{128}N_{33}O_{14}{}^+$ [M]$^+$: 1643.03 observed : 1643.15.
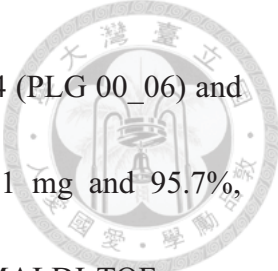
Fl-Mmk50-Tat

(Fl-βAla-Tyr-Gly-Arg-**Mmk**-Lys-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The peptide was synthesized using 170.0 mg (0.043 mmol) of NovaSyn$^®$TGR resin. The synthesis gave 300.7 mg of resin (81.5% yield). Large scale cleavage gave 51.2 mg

of crude peptide (51.5% yield). Retention time on analytical RP-HPLC was 27.5 minutes. The peptide was purified by Sep-Pak® Plus tC18 and then preparative RP-HPLC using a C18 (PLG 15_26) column. The final weight and purity were 3.2 mg and 96.2%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{89}H_{136}N_{34}O_{20}$ $[MH]^+$: 2001.23; observed : 2002.85.

Fl-Dmk50-Tat

(Fl-βAla-Tyr-Gly-Arg-**Dmk**-Lys-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The peptide was synthesized using 172.8 mg (0.043 mmol) of NovaSyn®TGR resin. The synthesis gave 320.6 mg of resin (90.3% yield). Large scale cleavage gave 55.5 mg of crude peptide (52.0% yield). Retention time on analytical RP-HPLC was 27.5 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 05_14) and then C18 (PLG 15_26) columns. The final weight and purity were 2.7 mg and 95.3%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{90}H_{138}N_{34}O_{20}$ $[MH]^+$: 2015.08; observed : 2017.55.
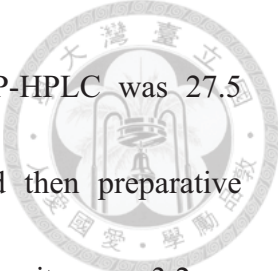
Fl-Tmk50-Tat

(Fl-βAla-Tyr-Gly- Arg-**Tmk**-Lys-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The peptide was synthesized using 171.1 mg (0.043 mmol) of NovaSyn®TGR resin.

The synthesis gave 304.6 mg of resin (82.1% yield). Large scale cleavage gave 63.5 mg of crude peptide (62.7% yield). Retention time on analytical RP-HPLC was 27.5 minutes. The peptide was purified by Sep-Pak® Plus tC18 and then preparative RP-HPLC using a C18 (PLG 15_26) column. The final weight and purity were 4.0 mg and 96.5%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{91}H_{141}N_{34}O_{20}{}^{+}$ [M]$^{+}$: 2030.11; observed : 2030.59.

Fl-Mmk51-Tat

(Fl-βAla-Tyr-Gly-Arg-Lys-**Mmk**-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The peptide was synthesized using 171.9 mg (0.043 mmol) of NovaSyn®TGR resin. The synthesis gave 280.5 mg of resin (66.9% yield). Large scale cleavage gave 48.9 mg of crude peptide (58.5% yield). Retention time on analytical RP-HPLC was 27.2 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 05_14) and C18 (PLG 15_26) columns. The final weight and purity were 1.9 mg and 97.3%, respectively. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{89}H_{136}N_{34}O_{20}$ [MH]$^{+}$: 2001.23; observed : 2002.22.
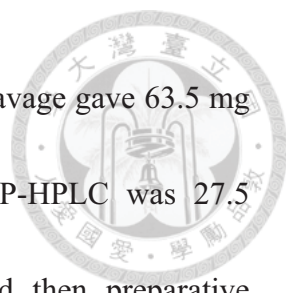
Fl-Dmk51-Tat

(Fl-βAla-Tyr-Gly-Arg-Lys-**Dmk**-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The corresponding precursor peptide Fl-Lys(ivDde)51-Tat was synthesized using 241.9 mg (0.060 mmol) of NovaSyn®TGR resin. The synthesis gave 449.2 mg of resin (98.5% yield). Retention time on analytical RP-HPLC was 34.8 minutes. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{101}H_{152}N_{34}O_{22}$ [MH]$^+$: 2194.19; observed : 2195.87. The precursor resin was treated with 20 equivalents trityl chloride and 20 equivalents DIEA in 1 mL DCM to protect the fluorescein moiety (2 x 16 hrs).[78] The reaction vessel was wrapped in aluminum foil. Then the protecting group ivDde was removed by suspending the resin in 2% hydrazine in DMF (4 mL, 8 x 8 min) and shaking at room temperature. The resin was washed with DMF (1 mL, 1 x 5 min). A formaldehyde solution was prepared with 10 mL 37% formalin and 90 mL DMF. The solution was then treated with 30 g $MgSO_4$ and stirred overnight. For 50 mg crude resin, the resin was treated with 3 mL of the formaldehyde solution and shaken for 5 minutes. Additional 3 mL formaldehyde solution and 30 μL of acetic acid were added, and the resin was shaken for 5 min. 30 mg of sodium cyanoborohydride was added and the resin was shaken for 1 hr. The reductive methylation was applied and gave 361.3 mg of resin (52.5% yield). Large scale cleavage gave 64.5 mg of crude peptide (72.1% yield). Retention time on analytical RP-HPLC was 27.4 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 02_16) and C18 (PLG 15_27) columns. The final weight and purity were 4.6

mg and 95.9%, respectively. The identity of the peptide was confirmed by MALDI-TOF

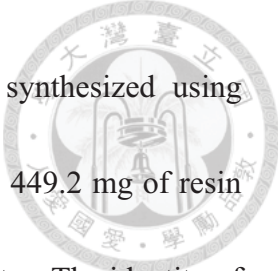mass spectrometry. Calculated for $C_{90}H_{138}N_{34}O_{20}$ [MH]$^+$: 2015.08; observed : 2016.58.

Fl-Tmk51-Tat

(Fl-βAla-Tyr-Gly-Arg-Lys-**Tmk**-Arg-Arg-Gln-Arg-Arg-Arg-NH$_2$)

The peptide was synthesized using 234.0 mg (0.059 mmol) of NovaSyn®TGR

resin. The synthesis gave 408.6 mg of resin (78.5% yield). Large scale cleavage gave

70.4 mg of crude peptide (52.2% yield). Retention time on analytical RP-HPLC was

27.2 minutes. The peptide was purified by preparative RP-HPLC using C4 (PLG 05_14)

and C18 (PLG 15_26) columns. The final weight and purity were 2.2 mg and 95.5%,

respectively. The identity of the peptide was confirmed by MALDI-TOF mass

spectrometry. Calculated for $C_{91}H_{141}N_{34}O_{20}{}^+$ [M]$^+$: 2030.11; observed : 2030.28.

Ultraviolet-Visible (UV-vis) Spectroscopy

Ac-capped peptides and Fl-capped peptides were prepared as 10 mM and 5 mM

stock solutions, respectively. The concentration of Ac-capped peptide stock solution was

determined by the tyrosine absorbance in 6 M guanidinium chloride under 25 °C. 2 μL

of peptide stock solution was gradually added to 10 μL into 400 μL 6 M guanidine

hydrochloride. Highly concentrated guanidine hydrochloride (GdnHCl) served as a

denaturant and would sabotage peptides structure. The peptide would be in the random
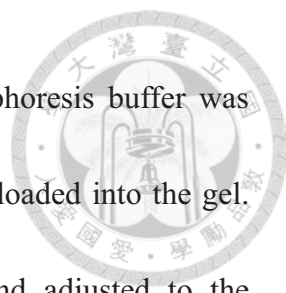
coil state and highly unfolded. The full spectra (200-500 nm) and single wavelength absorbance (276, 278, 280, 282, 335, and 400 nm) were measured. Literature values were used for the molar absorption coefficients (ε value) of tyrosine (in 6 M GdnHCl) ($\varepsilon_{276}=1455$, $\varepsilon_{278}=1395$, $\varepsilon_{280}=1285$, $\varepsilon_{282}=1220$ ($M^{-1}cm^{-1}$)).[91, 92] Each addition of 2 μL peptide stock solution was stirred by pipette for 100 times. The mixed solution was equilibrated for 15 minutes before measurements. The absorbance of the four aforementioned wavelengths (276, 278, 280, and 282 nm) were baseline corrected by the absorbance at 335 or 400nm. The concentration of the stock solution was derived through linear regression analysis. The R value higher than 0.999 was deemed acceptable.

For Fl-capped peptides, the concentration was determined in pH 9 buffer solution (1 mM borate, 1 mM citrate, and 1 mM phosphate) at 25 ^{o}C. 2 μL of peptide stock solution was gradually added to 4 μL into the pH 9 buffer solution. The full spectra (200-700 nm) and single wavelength absorbance (492 nm) were measured. The molar absorption coefficient of 6-carboxy-fluorescein in pH 9 buffer is 81000 ($M^{-1}cm^{-1}$).[93] Each addition of 2 μL peptide stock solution was stirred by pipette for 100 times. The mixed solution was equilibrated for 15 minutes before measurements.

Electrophoretic Mobility Shift Assay

The apparatuses (20 mL vial, stir bar, glass plates, combs and casting stand) were cleaned using ethanol and doubly distilled-H_2O. Acrylamide-bis-acrylamide 7.7 mL, TB buffer (0.4 M Tris-HCl and boric acid) 3 mL were added in the 20 mL vial and swirled gently (Table 3-9). 10 % Ammonium persulfate (APS) and *N, N, N',N'*-tetramethylethylenediamine (TEMED) were added in the vial to initialize polymerization. Total volume was 15 mL for the two gel format (10.5 cm x 9 cm x 0.1 cm each gel). Acrylamide is monomer and bis-acrylamide is cross-linker. The ratio of acrylamide to bis-acrylamide was 23.5:1.5. The APS is a strong oxidizing agent and radical initiator, initializing the polymerization of acrylamide and bis-acrylamide. TEMED is catalysis for APS polymerization reaction. The separating gel solution was poured onto the glass plates then the combs were placed and the gels were solidified for 30 minutes. Pipette tips and eppendorfs were sterilized for sample preparation. Bulk *E. coli* tRNA (10 µg/mL), 3'-modified fluorescein TAR RNA (1 µg/mL), 20% glycerol were added into eppendorfs with TKT buffer (contain 50 mM Tris-HCl, 50mM KCl, and 0.05% Triton X-100. The pH value was adjusted to 7.4 with KOH). These four materials were mixed into eppendorfs and different concentrations of Tat-derived peptide were added subsequently (Table 3-10). The eppendorfs were vortex, centrifuged, and then equilibrated for 30 minutes. After removing the combs, the gel and electrode

assembly were placed in the electrophoresis tank and the electrophoresis buffer was added (50 mL TB buffer and 450 mL dd-H$_2$O). The samples were loaded into the gel. The electrophoresis tank was connected to the power supply and adjusted to the appropriate voltage (140 V) to perform the gel shift assay (75 minutes) at room temperature. Each Gel shift assay was performed more than three times independently to confirm the reproducibility. Dried gels were scanned by a Typhoon TRIO$^+$ Variable Mode Imager with excitation wavelength 488 nm and emission wavelength 526 nm. The PMT was set to 620 V. Bands corresponding to the free (unbound) and bound RNA were quantified using the Image Quant software. The fraction bound RNA was derived by dividing the band intensity of bound RNA by the sum of the band intensities for the free (unbound) and bound RNA. The apparent dissociation constants were globally derived from the quantified data assuming a 1:1 binding stoichiometry using the following equation in Kaleidagraph 3.52. (Synergy Software).

$$\text{Fraction Bound RNA} = \frac{([Pep]_T + [RNA]_T + K_d) \pm \sqrt{([Pep]_T + [RNA]_T + K_d)^2 - 4([Pep]_T + [RNA]_T)}}{2[RNA]_T}$$

where $[Pep]_T$ is the total peptide concentration (including unbound peptide and bound peptide), $[RNA]_T$ is the total RNA concentration (including unbound RNA and bound RNA), and K_d is the apparent dissociation constant for the RNA-peptide complex.
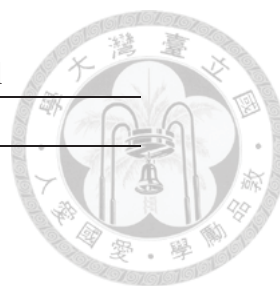
Table 3-9. Amount of Reagents for Preparation of the Separating Gel

Reagent	Volume
dd-H$_2$O	4.1 mL
X5 TB buffer	3.0 mL
23.5% acrylamide/1.5% bis-acrylamide	7.7 mL
10% APS	125 μL
TEMED	12.5 μL

Table 3-10. Amount of Reagents for Preparation of Samples with Different Concentrations

Peptide Concentration (nM)	dd-H$_2$O (μL)	2XTKT[a] buffer (μL)	t-RNA 1 μM (μL)	Tar-RNA 1 μM (μL)	20% Glycerol(μL)	Peptide (μL)
0	2.4	6	1.2	1.2	1.2	0
50	1.2	6	1.2	1.2	1.2	1.2 (500 nM)
100	1.2	6	1.2	1.2	1.2	1.2 (1 μM)
200	0	6	1.2	1.2	1.2	2.4 (1 μM)
400	1.2	6	1.2	1.2	1.2	1.2 (4 μM)
600	1.2	6	1.2	1.2	1.2	1.2 (6 μM)
800	0	6	1.2	1.2	1.2	2.4 (4 μM)
1200	0	6	1.2	1.2	1.2	2.4 (6 μM)
1600	1.2	6	1.2	1.2	1.2	1.2 (16 μM)

[a]TKT buffer: 50 mM Tris-HCl, 50 mM KCl, and 0.05% Triton X-100. Adjust the pH value to 7.4 with KOH).

Circular Dichroism Spectroscopy

2 μL of the peptide stock solution was gradually added into the 10 mM Tris buffer

(pH 7) to 4 μL (50–100 μM). Each sample was scanned between 200 and 300 nm at 25

^{o}C. Each reported CD value was the mean of at least 3 determinations. Data was

expressed in terms of mean residue molar ellipticity (deg·cm^2·dmol^{-1})

Cells and Cell Cultures

Jurkat cells were maintained in Roswell Park Memorial Institute (RPMI) medium,

supplemented with 2 mg/mL sodium bicarbonate (NaHCO$_3$), 10% (vol/vol) fetal bovine serum (FBS), 1% (vol/vol) penicillin streptomycin (Pen/Strep) in 5% CO$_2$/95% humidified air at 37°C. Antibody in FBS was deactivated by warming for 30 minutes at 56 °C.
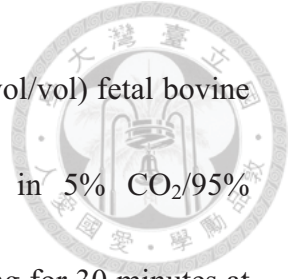
Cellular uptake Assay

All apparatuses were sterilized by an autoclave sterilizer and 70% ethanol. All procedures were performed in a laminar flow hood. The cell number was counted by a hemacytometer. There were 9 squares with a 1.0 mm^2 area and 0.1 mm depth in the hemacytometer. The cell number in the four corners was counted and averaged. The average was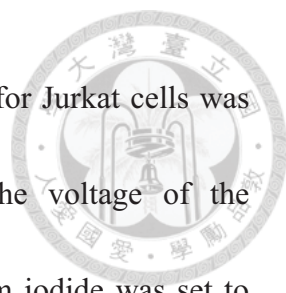 multiplied by 10^4 to obtain the cell number in 1 mL. The cell number was set to 8 x 10^5 in the cellular uptake experiment. Jurkat cells were incubated with the peptides at concentrations 7 and 120 μM at 37°C, 5% CO$_2$ for 15 minutes. Cells were washed with PBS (KCl (2 g/L), KH$_2$PO$_4$ (2 g/L), NaCl (80 g/L), Na$_2$PO$_4$ (11.5 g/L)) (2 x 350 μL) to remove the fetal bovine serum which might interfere the activity of trypsin. The cells were incubated with 0.05% trypsin/EDTA in PBS for 5 minutes to remove the peptides which adhered to the cell surface rather than entry to the cell.[96] The cells were washed with PBS (2 x 350 μL). The cells were resuspended in 300 μL PBS and transferred into the flow tube. The cells were terminated by adding 0.05% Triton-X 100 in the dead control group. Propidium iodide (PI) was added to all samples to stain the

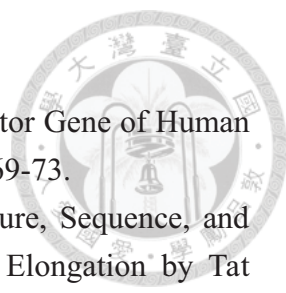dead cells except for the live control group. Fluorescence analysis for Jurkat cells was performed with a FACScan (Becton Dickinson Bioscience). The voltage of the photomultiplier tube for forward scatter, side scatter, and propidium iodide was set to 230, 420, and 350, respectively. The voltage of the photomultiplier tube for fluorescein isothiocyanate at 7 μM and 120 μM were set to 230 and 217, respectively. The cells which contained appropriate forward scatter and side scatter values were selected and gated as the P1 region as normal and live cells in the live control group. The minimum propidium iodide fluorescence intensity in the P1 region was set to the threshold value of dead cell in the death control group. The fluorescence of 6-carboxy-fluorescein was considered when the cell-morphology was in the P1 region and the propidium iodide fluorescence intensity was lower than the threshold value. The 6-carboxy-fluorescein fluorescence intensity was acquired for 10000 events at room temperature. The data were presented in mean cellular fluorescence intensity.
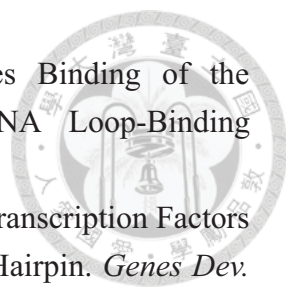
3-6 References

1. Crick, F. Central dogma of molecular biology. *Nature* **1970,** *227*, 561-563.
2. Fresco, J. R.; Alberts, B. M.; Doty, P. Some Molecular Details of the Secondary Structure of Ribonucleic Acid. *Nature* **1960,** *188*, 98-101.
3. Varmus, H. Retroviruses. *Science* **1988,** *240*, 1427-1435.
4. Weiss, R. A. How Does HIV Cause AIDS. *Science* **1993,** *260*, 1273-1279.
5. Greene, W. C.; Peterlin, B. M. Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. *Nat. Med.* **2002,** *8*, 673-680.
6. Chan, D. C.; Kim, P. S. HIV entry and its inhibition. *Cell* **1998,** *93*, 681-684.
7. Levy, J. A. Pathogenesis of Human-Immunodeficiency-Virus Infection. *Microbiol.*

Rev. **1993,** *57*, 183-289.

8. Doms, R. W.; Moore, J. P. HIV-1 membrane fusion: Targets of opportunity. *J. Cell. Biol.* **2000,** *151*, 9-13.

9. Caffrey, M. HIV envelope: challenges and opportunities for development of entry inhibitors. *Trends. Microbiol.* **2011,** *19*, 191-197.

10. Mehellou, Y.; De Clercq, E. Twenty-Six Years of Anti-HIV Drug Discovery: Where Do We Stand and Where Do We Go? *J. Med. Chem.* **2010,** *53*, 521-538.

11. Central Dogma Reversed. *Nature* **1970,** *226*, 1198-1199.

12. Hu, W. S.; Temin, H. M. Retroviral Recombination and Reverse Transcription. *Science* **1990,** *250*, 1227-1233.

13. Miller, M. D.; Farnet, C. M.; Bushman, F. D. Human immunodeficiency virus type 1 preintegration complexes: Studies of organization and composition. *J. Virol.* **1997,** *71*, 5382-5390.

14. Stoff, D. M.; Khalsa, J. H.; Monjan, A.; Portegies, P. HIV/AIDS and aging - Introduction. *AIDS* **2004,** *18*, 1-2.

15. Frankel, A. D.; Young, J. A. T. HIV-1: Fifteen proteins and an RNA. *Annu. Rev. Biochem.* **1998,** *67*, 1-25.

16. Stevens, M.; De Clercq, E.; Balzarini, J. The regulation of HIV-1 transcription: molecular targets for chemotherapeutic intervention. *Med. Res. Rev.* **2006,** *26*, 595-625.

17. Westerhout, E. M.; Ooms, M.; Vink, M.; Das, A. T.; Berkhout, B. HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome. *Nucleic Acids Res.* **2005,** *33*, 796-804.

18. Greene, W. C. Mechanisms of Disease - the Molecular-Biology of Human-Immunodeficiency-Virus Type-1 Infection. *N. Engl. J. Med.* **1991,** *324*, 308-317.

19. Ruben, S.; Perkins, A.; Purcell, R.; Joung, K.; Sia, R.; Burghoff, R.; Haseltine, W. A.; Rosen, C. A. Structural and Functional-Characterization of Human Immunodeficiency Virus Tat Protein. *J. Virol.* **1989,** *63*, 1-8.

20. Sodroski, J.; Rosen, C.; Wongstaal, F.; Salahuddin, S. Z.; Popovic, M.; Arya, S.; Gallo, R. C.; Haseltine, W. A. Trans-Acting Transcriptional Regulation of Human T-Cell Leukemia-Virus Type-III Long Terminal Repeat. *Science* **1985,** *227*, 171-173.

21. Sodroski, J.; Patarca, R.; Rosen, C.; Wongstaal, F.; Haseltine, W. Location of the Trans-Activating Region on the Genome of Human T-Cell Lymphotropic Virus Type-III. *Science* **1985,** *229*, 74-77.

22. Rosen, C. A.; Sodroski, J. G.; Haseltine, W. A. The Location of Cis-Acting Regulatory Sequences in the Human T-Cell Lymphotropic Virus Type-III

(HTLV-III/LAV) Long Terminal Repeat. *Cell* **1985,** *41*, 813-823.

23. Arya, S. K.; Guo, C.; Josephs, S. F.; Wongstaal, F. Trans-Activator Gene of Human T-Lymphotropic Virus Type-III (HTLV-III). *Science* **1985,** *229*, 69-73.

24. Selby, M. J.; Bain, E. S.; Luciw, P. A.; Peterlin, B. M. Structure, Sequence, and Position of the Stem Loop in Tar Determine Transcriptional Elongation by Tat through the HIV-1 Long Terminal Repeat. *Genes Dev.* **1989,** *3*, 547-558.

25. Muesing, M. A.; Smith, D. H.; Capon, D. J. Regulation of Messenger-Rna Accumulation by a Human-Immunodeficiency-Virus Transactivator Protein. *Cell* **1987,** *48*, 691-701.

26. Peterlin, B. M.; Luciw, P. A.; Barr, P. J.; Walker, M. D. Elevated Levels of Messenger-Rna Can Account for the Transactivation of Human-Immunodeficiency-Virus. *Proc. Natl. Acad. Sci. U. S. A.* **1986,** *83*, 9734-9738.

27. Weeks, K. M.; Crothers, D. M. RNA Recognition by Tat-Derived Peptides - Interaction in the Major Groove. *Cell* **1991,** *66*, 577-588.

28. Weeks, K. M.; Ampe, C.; Schultz, S. C.; Steitz, T. A.; Crothers, D. M. Fragments of the HIV-1 Tat Protein Specifically Bind TAR RRA. *Science* **1990,** *249*, 1281-1285.

29. Roy, S.; Delling, U.; Chen, C. H.; Rosen, C. A.; Sonenberg, N. A Bulge Structure in HIV-1 TAR RNA Is Required for Tat Binding and Tat-Mediated Transactivation. *Genes Dev.* **1990,** *4*, 1365-1373.

30. Muller, W. E. G.; Okamoto, T.; Reuter, P.; Ugarkovic, D.; Schroder, H. C. Functional-Characterization of Tat Protein from Human-Immunodeficiency-Virus - Evidence That Tat Links Viral RNA to Nuclear Matrix. *J. Biol. Chem.* **1990,** *265*, 3803-3808.

31. Dingwall, C.; Ernberg, I.; Gait, M. J.; Green, S. M.; Heaphy, S.; Karn, J.; Lowe, A. D.; Singh, M.; Skinner, M. A. HIV-1 Tat Protein Stimulates Transcription by Binding to a U-Rich Bulge in the Stem of the TAR RNA Structure. *EMBO J.* **1990,** *9*, 4145-4153.

32. Cordingley, M. G.; Lafemina, R. L.; Callahan, P. L.; Condra, J. H.; Sardana, V. V.; Graham, D. J.; Nguyen, T. M.; Legrow, K.; Gotlib, L.; Schlabach, A. J.; Colonno, R. J. Sequence-Specific Interaction of Tat Protein and Tat Peptides with the Transactivation-Responsive Sequence Element of Human-Immunodeficiency-Virus Type-1 Invitro. *Proc. Natl. Acad. Sci. U. S. A.* **1990,** *87*, 8985-8989.

33. Churcher, M. J.; Lamont, C.; Hamy, F.; Dingwall, C.; Green, S. M.; Lowe, A. D.; Butler, P. J. G.; Gait, M. J.; Karn, J. High-Affinity Binding of TAR RNA by the Human-Immunodeficiency-Virus Type-1 Tat Protein Requires Base-Pairs in the RNA Stem and Amino-Acid-Residues Flanking the Basic Region. *J. Mol. Biol.* **1993,** *230*, 90-110.

34. Wu, F.; Garcia, J.; Sigman, D.; Gaynor, R. Tat Regulates Binding of the Human-Immunodeficiency-Virus Transactivating Region RNA Loop-Binding Protein Trp-185. *Genes Dev.* **1991,** *5*, 2128-2140.

35. Sheline, C. T.; Milocco, L. H.; Jones, K. A. 2 Distinct Nuclear Transcription Factors Recognize Loop and Bulge Residues of the HIV-1 TAR RNA Hairpin. *Genes Dev.* **1991,** *5*, 2508-2520.

36. Marciniak, R. A.; Garciablanco, M. A.; Sharp, P. A. Identification and Characterization of a Hela Nuclear-Protein That Specifically Binds to the Trans-Activation-Response (TAR) Element of Human-Immunodeficiency-Virus. *Proc. Natl. Acad. Sci. U. S. A.* **1990,** *87*, 3624-3628.

37. Gaynor, R.; Soultanakis, E.; Kuwabara, M.; Garcia, J.; Sigman, D. S. Specific Binding of a Hela-Cell Nuclear-Protein to RNA Sequences in the Human Immunodeficiency Virus Transactivating Region. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86*, 4858-4862.

38. Gatignol, A.; Kumar, A.; Rabson, A.; Jeang, K. T. Identification of Cellular Proteins That Bind to the Human Immunodeficiency Virus Type-1 Trans-Activation-Responsive Tar Element Rna. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86*, 7828-7832.

39. Jeang, K. T. G., A. Comparison of regulatory features among primate lentiviruses. *Curr. Top. Microbiol. Immunol.* **1994,** *188*, 123-144.

40. Rana, T. M.; Jeang, K. T. Biochemical and functional interactions between HIV-1 Tat protein and TAR RNA. *Arch. Biochem. Biophys.* **1999,** *365*, 175-185.

41. Frankel, A. D.; Biancalana, S.; Hudson, D. Activity of Synthetic Peptides from the Tat Protein of Human Immunodeficiency Virus Type-1. *Proc. Natl. Acad. Sci. U. S. A.* **1989,** *86*, 7397-7401.

42. Kuppuswamy, M.; Subramanian, T.; Srinivasan, A.; Chinnadurai, G. Multiple Functional Domains of Tat, the Trans-Activator of HIV-1, Defined by Mutational Analysis. *Nucleic Acids. Res.* **1989,** *17*, 3551-3561.

43. Dang, C. V.; Lee, W. M. F. Nuclear and Nucleolar Targeting Sequences of C-Erb-a, C-Myb, N-Myc, P53, Hsp70, and HIV Tat Proteins. *J. Biol. Chem.* **1989,** *264*, 18019-18023.

44. Hauber, J.; Perkins, A.; Heimer, E. P.; Cullen, B. R. Transactivation of Human-Immunodeficiency-Virus Gene-Expression Is Mediated by Nuclear Events. *Proc. Natl. Acad. Sci. U. S. A.* **1987,** *84*, 6364-6368.

45. Schwarze, S. R.; Hruska, K. A.; Dowdy, S. F. Protein transduction: unrestricted delivery into all cells? *Trends. Cell. Biol.* **2000,** *10*, 290-295.

46. Campbell, G. R.; Pesquier, E.; Watkins, J.; Bourgarel-Rey, V.; Peyrot, V.; Esquieu, D.; Barbier, P.; de Mareuil, J.; Braguer, D.; Kaleebu, P.; Yirrell, D. L.; Loret, E. P.

The glutamine-rich region of the HIV-1 Tat protein is involved in T-cell apoptosis. *J. Biol. Chem.* **2004,** *279*, 48197-48204.

47. Kao, S. Y.; Calman, A. F.; Luciw, P. A.; Peterlin, B. M. Anti-Termination of Transcription within the Long Terminal Repeat of HIV-1 by Tat Gene-Product. *Nature* **1987,** *330*, 489-493.

48. Dayton, A. I.; Sodroski, J. G.; Rosen, C. A.; Goh, W. C.; Haseltine, W. A. The Transactivator Gene of the Human T-Cell Lymphotropic Virus Type-III Is Required for Replication. *Cell* **1986,** *44*, 941-947.

49. Wei, P.; Garber, M. E.; Fang, S. M.; Fischer, W. H.; Jones, K. A. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **1998,** *92*, 451-462.

50. Tahirov, T. H.; Babayeva, N. D.; Varzavand, K.; Cooper, J. J.; Sedore, S. C.; Price, D. H. Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature* **2010,** *465*, 747-751.

51. Ping, Y. H.; Rana, T. M. DSIF and NELF interact with RNA polymerase II elongation complex and HIV-1 Tat stimulates P-TEFb-mediated phosphorylation of RNA polymerase II and DSIF during transcription elongation. *J. Biol. Chem.* **2001,** *276*, 12951-12958.

52. Fujinaga, K.; Irwin, D.; Huang, Y. H.; Taube, R.; Kurosu, T.; Peterlin, B. M. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol. Cell. Biol.* **2004,** *24*, 787-795.

53. Pagans, S.; Kauder, S. E.; Kaehlcke, K.; Sakane, N.; Schroeder, S.; Dormeyer, W.; Trievel, R. C.; Verdin, E.; Schnolzer, M.; Ott, M. The Cellular Lysine Methyltransferase Set7/9-KMT7 Binds HIV-1 TAR RNA, Monomethylates the Viral Transactivator Tat, and Enhances HIV Transcription. *Cell Host Microbe* **2010,** *7*, 234-244.

54. Gallego, J.; Varani, G. Targeting RNA with small-molecule drugs: Therapeutic promise and chemical challenges. *Acc. Chem. Res.* **2001,** *34*, 836-843.

55. Athanassiou, Z.; Dias, R. L. A.; Moehle, K.; Dobson, N.; Varani, G.; Robinson, J. A. Structural mimicry of retroviral Tat proteins by constrained, β-hairpin peptidomimetics: Ligands with high affinity and selectivity for viral TAR RNA regulatory elements. *J. Am. Chem. Soc.* **2004,** *126*, 6906-6913.

56. Davidson, A.; Patora-Komisarska, K.; Robinson, J. A.; Varani, G. Essential structural requirements for specific recognition of HIV TAR RNA by peptide mimetics of Tat protein. *Nucleic Acids. Res.* **2011,** *39*, 248-256.

57. Hamy, F.; Felder, E. R.; Heizmann, G.; Lazdins, J.; AboulEla, F.; Varani, G.; Karn, J.; Klimkait, T. An inhibitor of the Tat/TAR RNA interaction that effectively

suppresses HIV-1 replication. *Proc. Natl. Acad. Sci. U. S. A.* **1997,** *94*, 3548-3553.

58. Huq, I.; Wang, X. L.; Rana, T. M. Specific recognition of HIV-1 TAR RNA by a D-Tat peptide. *Nat. Struct. Biol.* **1997,** *4*, 881-882.

59. Kesavan, V.; Tamilarasu, N.; Cao, H.; Rana, T. M. A new class of RNA-binding oligomers: Peptoid amide and ester analogues. *Bioconjugate Chem.* **2002,** *13*, 1171-1175.

60. Simon, R. J.; Kania, R. S.; Zuckermann, R. N.; Huebner, V. D.; Jewell, D. A.; Banville, S.; Ng, S.; Wang, L.; Rosenberg, S.; Marlowe, C. K.; Spellmeyer, D. C.; Tan, R. Y.; Frankel, A. D.; Santi, D. V.; Cohen, F. E.; Bartlett, P. A. Peptoids - a Modular Approach to Drug Discovery. *Proc. Natl. Acad. Sci. U. S. A.* **1992,** *89*, 9367-9371.

61. Tamilarasu, N.; Huq, I.; Rana, T. M. High affinity and specific binding of HIV-1 TAR RNA by a Tat-derived oligourea. *J. Am. Chem. Soc.* **1999,** *121*, 1597-1598.

62. Tamilarasu, N.; Huq, I.; Rana, T. M. Targeting RNA with peptidomimetic oligomers in human cells. *Bioorg. Med. Chem. Lett.* **2001,** *11*, 505-507.

63. Wang, X. L.; Huq, I.; Rana, T. M. HIV-1 TAR RNA recognition by an unnatural biopolymer. *J. Am. Chem. Soc.* **1997,** *119*, 6444-6445.

64. Zhang, X.; Wen, H.; Shi, X. B. Lysine methylation: beyond histones. *Acta Bioch. Bioph. Sin.* **2012,** *44*, 14-27.

65. Van Duyne, R.; Easley, R.; Wu, W. L.; Berro, R.; Pedati, C.; Klase, Z.; Kehn-Hall, K.; Flynn, E. K.; Symer, D. E.; Kashanchi, F. Lysine methylation of HIV-1 Tat regulates transcriptional activity of the viral LTR. *Retrovirology* **2008,** *5*, 40-53.

66. Sakane, N.; Kwon, H. S.; Pagans, S.; Kaehlcke, K.; Mizusawa, Y.; Kamada, M.; Lassen, K. G.; Chan, J.; Greene, W. C.; Schnoelzer, M.; Ott, M. Activation of HIV Transcription by the Viral Tat Protein Requires a Demethylation Step Mediated by Lysine-specific Demethylase 1 (LSD1/KDM1). *PLoS Pathog.* **2011,** *7*.

67. Fischer, R.; Fotin-Mleczek, M.; Hufnagel, H.; Brock, R. Break on through to the other side - Biophysics and cell biology shed light on cell-penetrating peptides. *Chembiochem* **2005,** *6*, 2126-2142.

68. Vives, E.; Brodin, P.; Lebleu, B. A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *J. Biol. Chem.* **1997,** *272*, 16010-16017.

69. Khalil, I. A.; Kogure, K.; Akita, H.; Harashima, H. Uptake pathways and subsequent intracellular trafficking in nonviral gene delivery. *Pharmacol. Rev.* **2006,** *58*, 32-45.

70. Madani, F.; Lindberg, S.; Langel, U.; Futaki, S.; Graslund, A. Mechanisms of cellular uptake of cell-penetrating peptides. *J. Biophys.* **2011,** *2011*, 414729.

71. Chao, T. Y.; Lavis, L. D.; Raines, R. T. Cellular Uptake of Ribonuclease A Relies on

Anionic Glycans. *Biochemistry* **2010,** *49*, 10666-10673.

72. Suzuki, T.; Futaki, S.; Niwa, M.; Tanaka, S.; Ueda, K.; Sugiura, Y. Possible existence of common internalization mechanisms among arginine-rich peptides. *J. Biol. Chem.* **2002,** *277*, 2437-2443.

73. Rothbard, J. B.; Jessop, T. C.; Lewis, R. S.; Murray, B. A.; Wender, P. A. Role of membrane potential and hydrogen bonding in the mechanism of translocation of guanidinium-rich peptides into cells. *J. Am. Chem. Soc.* **2004,** *126*, 9506-9507.

74. Calnan, B. J.; Biancalana, S.; Hudson, D.; Frankel, A. D. Analysis of Arginine-Rich Peptides from the HIV Tat Protein Reveals Unusual Features of Rna Protein Recognition. *Genes Dev.* **1991,** *5*, 201-210.

75. Gelman, M. A.; Richter, S.; Cao, H.; Umezawa, N.; Gellman, S. H.; Rana, T. M. Selective binding of TAR RNA by a Tat-derived β-peptide. *Org. Lett.* **2003,** *5*, 3563-3565.

76. Atherton, E.; Fox, H.; Harkiss, D.; Logan, C. J.; Sheppard, R. C.; Williams, B. J. Mild Procedure for Solid-Phase Peptide-Synthesis - Use of "Fluorenylmethoxycarbonylamino-Acids. *J. Chem. Soc. Chem. Comm.* **1978**, 537-539.

77. Fields, G. B.; Noble, R. L. Solid-Phase Peptide-Synthesis Utilizing 9-Fluorenylmethoxycarbonyl Amino-Acids. *Int. J. Pept. Prot. Res.* **1990,** *35*, 161-214.

78. Fischer, R.; Mader, O.; Jung, G.; Brock, R. Extending the applicability of carboxyfluorescein in solid-phase synthesis. *Bioconjugate Chem.* **2003,** *14*, 653-660.

79. Kaljuste, K.; Unden, A. New Method for the Synthesis of N-Methyl Amino-Acids Containing Peptides by Reductive Methylation of Amino-Groups on the Solid-Phase. *Int. J. Pept. Prot. Res.* **1993,** *42*, 118-124.

80. Garner, M. M.; Revzin, A. A Gel-Electrophoresis Method for Quantifying the Binding of Proteins to Specific DNA Regions - Application to Components of the Escherichia-Coli Lactose Operon Regulatory System. *Nucleic Acids. Res.* **1981,** *9*, 3047-3060.

81. Wang, J. H.; Huang, S. Y.; Choudhury, I.; Leibowitz, M. J.; Stein, S. Use of a polyethylene glycol-peptide conjugate in a competition gel shift assay for screening potential antagonists of HIV-1 Tat protein binding to TAR RNA. *Anal. Biochem.* **1995,** *232*, 238-242.

82. Wu, C. H.; Weng, M. H.; Chang, H. C.; Li, J. H.; Cheng, R. P. Effect of each guanidinium group on the RNA recognition and cellular uptake of Tat-derived peptides. *Bioorg. Med. Chem.* **2014,** *22*, 3016-3020.

83. Puglisi, J. D.; Chen, L.; Blanchard, S.; Frankel, A. D. Solution Structure of a

Bovine Immunodeficiency Virus Tat-TAR Peptide-RNA Complex. *Science* **1995,** *270*, 1200-1203.

84. Hauber, J.; Malim, M. H.; Cullen, B. R. Mutational Analysis of the Conserved Basic Domain of Human Immunodeficiency Virus Tat Protein. *J. Virol.* **1989,** *63*, 1181-1187.

85. Anand, K.; Schulte, A.; Vogel-Bachmayr, K.; Scheffzek, K.; Geyer, M. Structural insights into the Cyclin T1-Tat-TAR RNA transcription activation complex from EIAV. *Nat. Struct. Mol. Biol.* **2008,** *15*, 1287-1292.

86. Lee, J. C.; Gutell, R. R. Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs. *J. Mol. Biol.* **2004,** *344*, 1225-1249.

87. Ott, M.; Geyer, M.; Zhou, Q. The Control of HIV Transcription: Keeping RNA Polymerase II on Track. *Cell Host Microbe* **2011,** *10*, 426-435.

88. Schultz, D. C.; Ayyanathan, K.; Negorev, D.; Maul, G. G.; Rauscher, F. J. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **2002,** *16*, 919-932.

89. Mucha, P.; Rekowski, P.; Szyk, A.; Kupryszewski, G.; Giel-Pietraszuk, M.; Barciszewski, J. Circular dichroism studies of the interaction of Tat analogues substituted in the Arg(52) position with TAR RNA HIV-1. *Lett. Pept. Sci.* **1998,** *5*, 345-348.

90. Milletti, F. Cell-penetrating peptides: classes, origin, and current landscape. *Drug Discov. Today* **2012,** *17*, 850-860.

91. Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. How to Measure and Predict the Molar Absorption-Coefficient of a Protein. *Protein Sci.* **1995,** *4*, 2411-2423.

92. Edelhoch, H. Spectroscopic Determination of Tryptophan and Tyrosine in Proteins. *Biochemistry* **1967,** *6*, 1948-1954.

93. Haugland, R. P. *Handbook of Fluorescent Probes and Research Chemicals*. 6th ed.; Eugene, OR, USA, 1996.

Appendix

The statistical data of monomethyllysine (Mmk), dimethyllysine (Dmk), and trimethyllysine (Tmk), were obtained from UniProt Knowledgebase (UniProtKB). The experimentally confirmed methylation sites were included. Each site was classified into 4 groups according to the secondary structure assigned in Protein Data Bank (PDB): unknown for a site without any structural information, helix for a site identified in α-helix, sheet for a site identified in β-sheet, and random coil for other conformations.

Table A-1. The Secondary Structural Occurrence of Methylated Lysine in Natural Proteins

Modified Lysine	α-Helix	β-Sheet	Random Coil	Unknown	Total
Mmk	28	20	90	75	159
Dmk	10	16	54	13	93
Tmk	14	12	59	74	213

The full raw and summarized statistics for the various post-translational modifications were downloaded from Proteome-Wide PTM Statistics Curator (http://selene.princeton.edu/PTMCuration).[1] The "Entry name" for the proteins containing monomethyllysine (Mmk), dimethyllysine (Dmk), or trimethyllysine (Tmk), were compiled from the downloaded file "byidexperimental.txt". Each entry name was searched in the UniProt Knowledgebase (UniProtKB; http://www.uniprot.org). The methylation type and site were obtained from "Amino acid modifications" under "Sequence annotation (Features)". Only the experimentally confirmed methylation sites were analyzed. The sites without RCSB-PDB data in "3D structure databases" under

"Cross-references" were classified as "Unknown". For the sites with an RCSB-PDB entry, the "Entry" for the corresponding "Positions" containing the methylation site was used to access the entry information in the Protein Data Bank (http://www.rcsb.org). In the PDB entry information, the secondary structure for each methylation site was obtained from the "Annotations" under "Sequence". The secondary structure was defined using DSSP (Define Secondary Structure of Proteins). Sites annotated by DSSP as "alpha helix" were categorized as "Helix". Sites annotated by DSSP as "beta strand" were categorized as "Sheet". Sites annotated by DSSP as "turn", "bend", "beta bridge", "3/10 helix", "pi helix" or "No secondary structure assigned" were categorized as "Random coil". Sites annotated by DSSP as "No data for this region" were classified as "Unknown".

1. Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **2011**, *1*, 1-5.

Table A-2. The Z and P Values for w_9 Value of KXaa9

Peptides		Absolute Z value	P value
KLys9	KMmk9	1.2	1.2×10^{-1}
KLys9	KDmk9	3.4	3.0×10^{-4}
KLys9	KTmk9	7.0	$< 10^{-5}$
KMmk9	KDmk9	2.0	3.4×10^{-2}
KMmk9	KTmk9	4.0	8.1×10^{-5}
KDmk9	KTmk9	1.6	5.9×10^{-2}
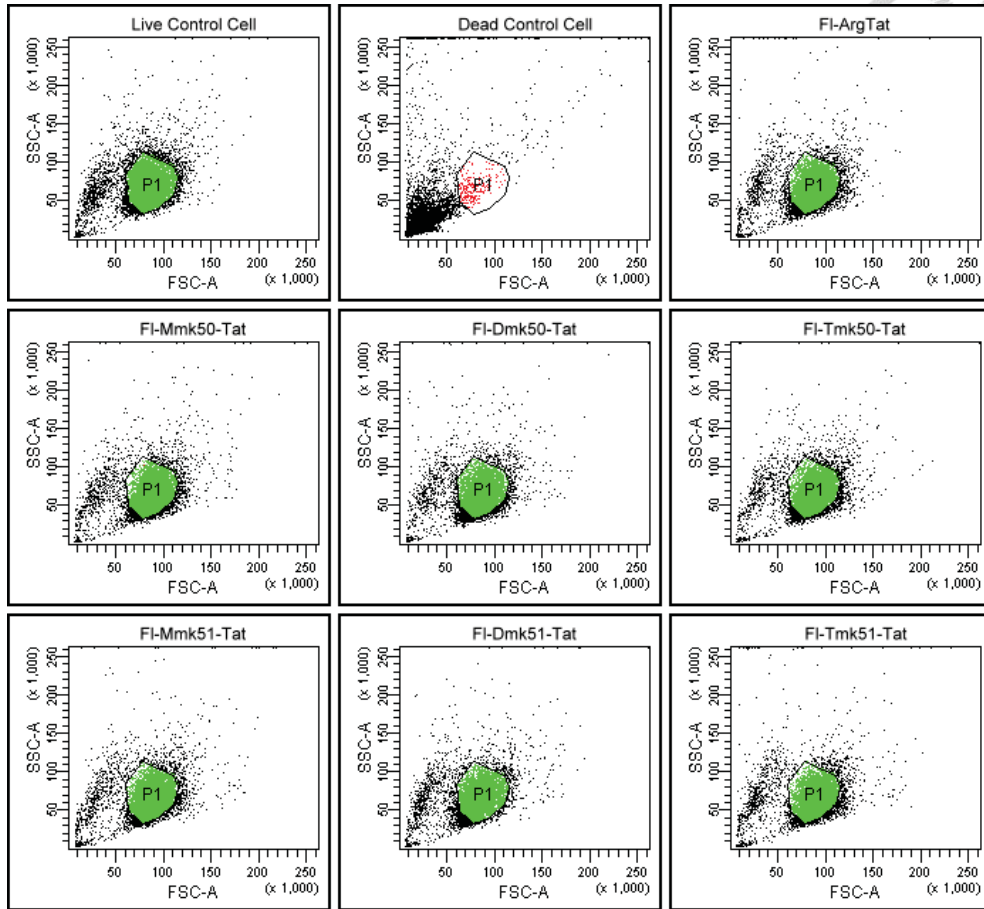
Figure A-1. Flow cytometry results showing the side scattered light plotted against the forward scattered light for live control cells, dead control cells, and cells incubated with 7 μM Tat-derived peptides. The gate used to restrict the population of cells analyzed is shown and labeled as P1.
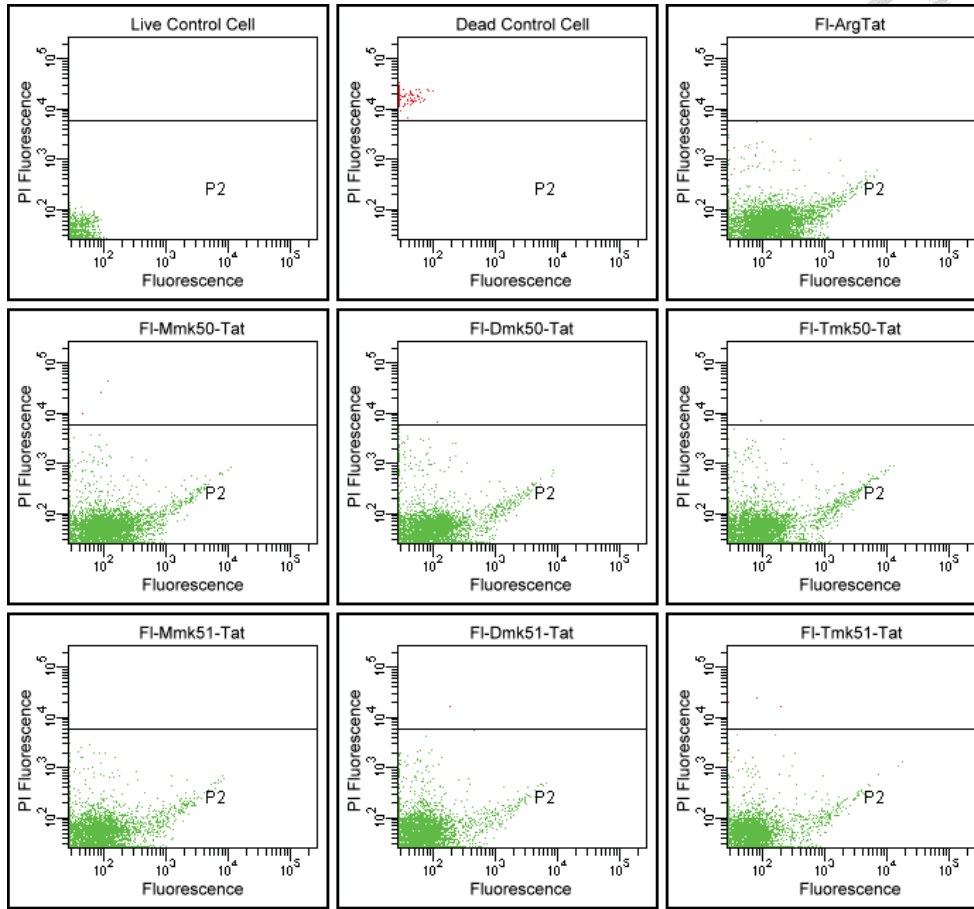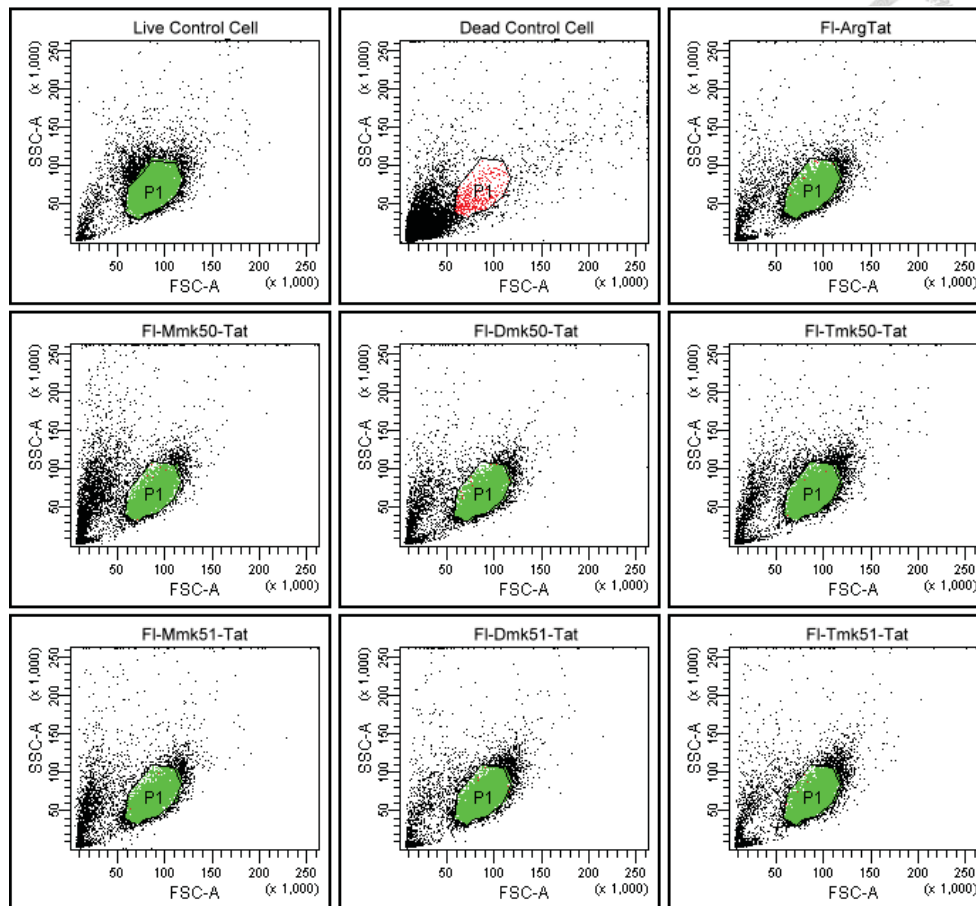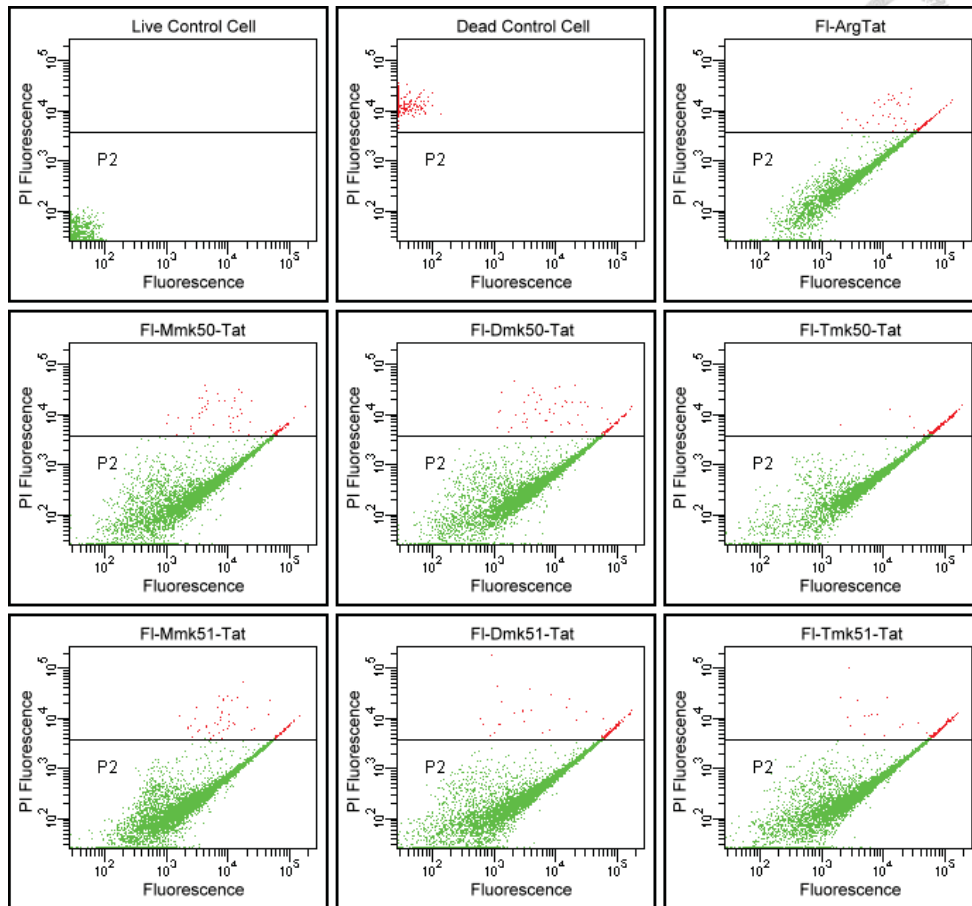
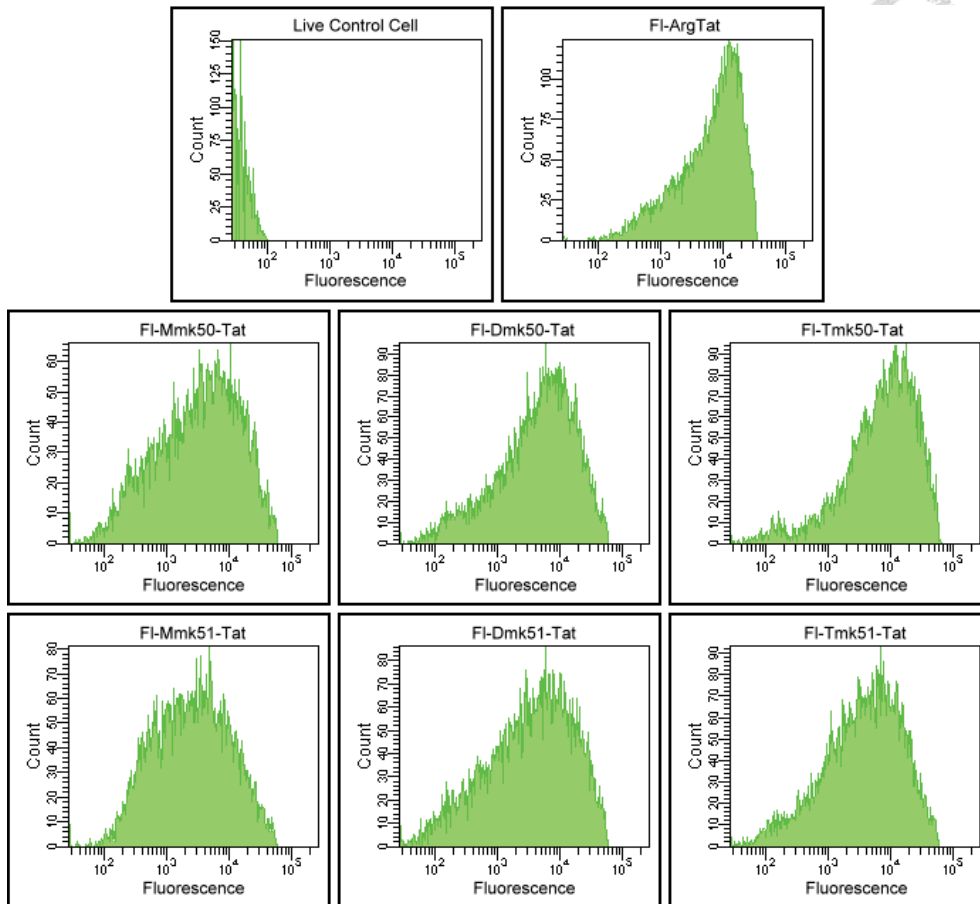**Figure A-2.** Flow cytometry results showing the propidium iodide fluorescence against the fluorescein fluorescence for live control cells, dead control cells, and cells incubated with 7 μM Tat-derived peptides. The gate used to restrict the fluorescence of cells analyzed is shown and labeled as P2.
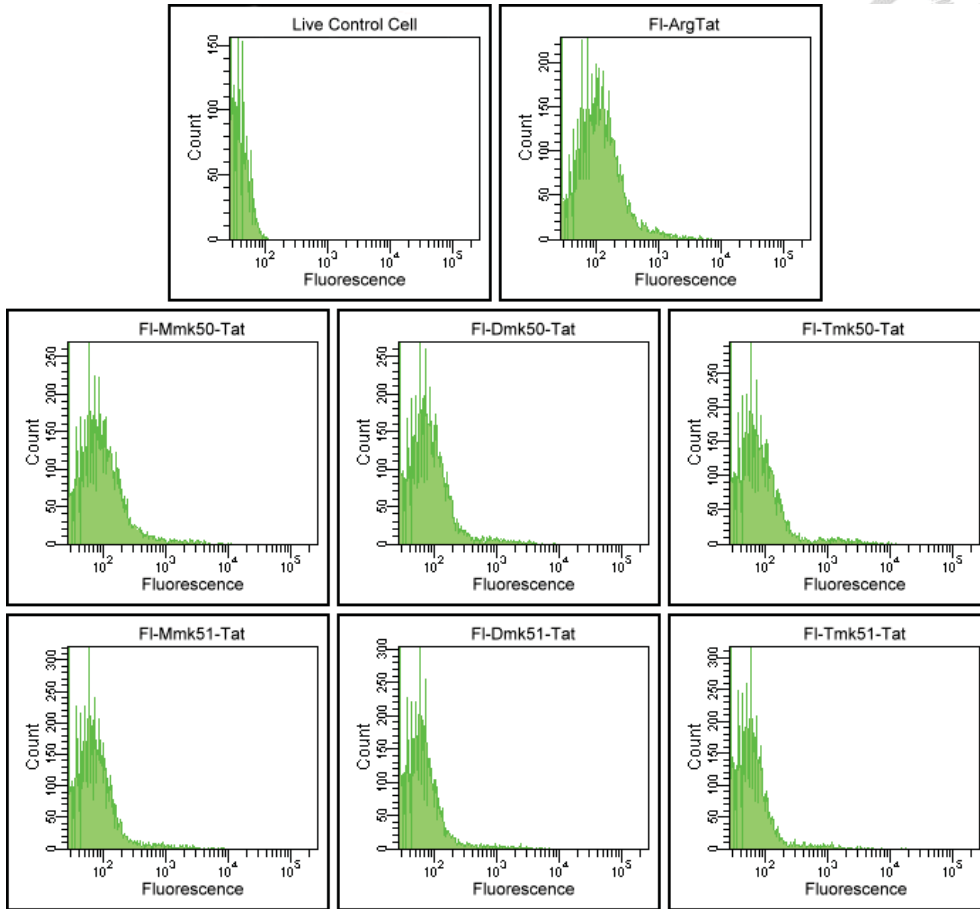
Figure A-3. Flow cytometry results showing the fluorescein fluorescence for live control cells, and cells incubated with 7 μM Tat-derived peptides for 15 minutes at 37 °C.

Figure A-4. Flow cytometry results showing the side scattered light plotted against the forward scattered light for live control cells, dead control cells, and cells incubated with 120 μM Tat-derived peptides. The gate used to restrict the population of cells analyzed is shown and labeled as P1.

Figure A-5. Flow cytometry results showing the propidium iodide fluorescence against the fluorescein fluorescence for live control cells, dead control cells, and cells incubated with 120 μM Tat-derived peptides. The gate used to restrict the fluorescence of cells analyzed is shown and labeled as P2.

Figure A-6. Flow cytometry results showing the fluorescein fluorescence for live control cells, and cells incubated with 120 μM Tat-derived peptides for 15 minutes at 37 °C.
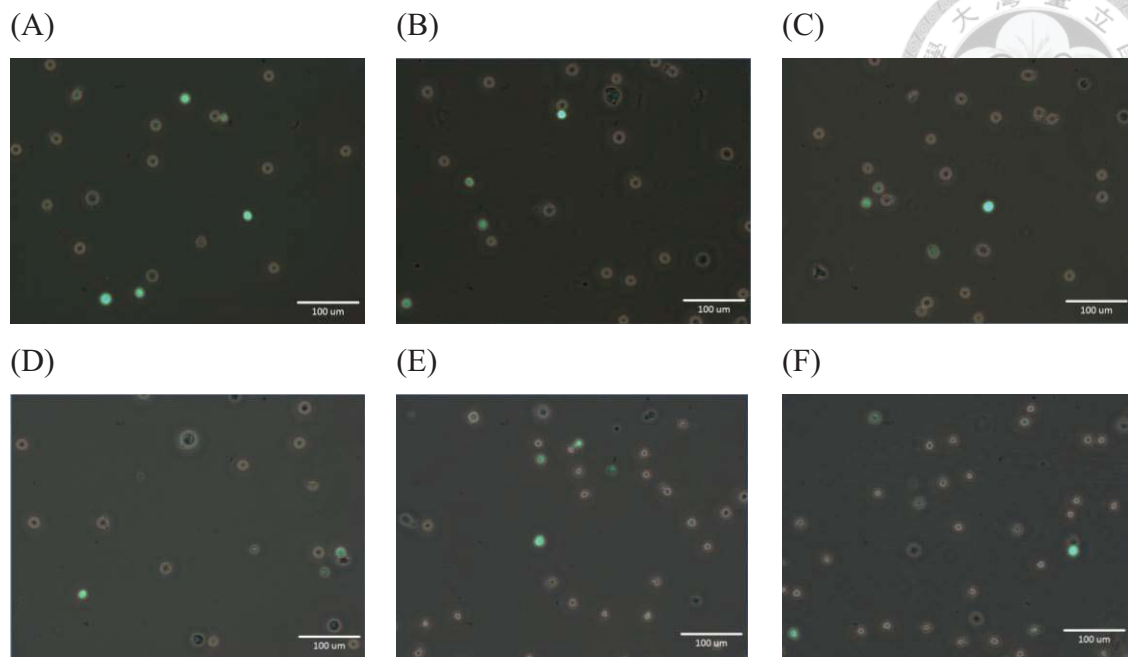
Figure A-7. The overlaid bright-foeld and fluorescence microscopy images of Jurkat cells incubated with 7 μM Fl-Mmk50-Tat (A), Fl-Dmk50-Tat (B), Fl-Tmk50-Tat (C), Fl-Mmk51-Tat (D), Fl-Dmk51-Tat (E), and Fl-Tmk51-Tat (F) for 15 minutes at 37 °C in the presence of fetal bovine serum, washed and treated with trypsin at 37 °C for 5 minutes.

Figure A-8. The license agreement for Figure 3-4

145

Review Order

This is a License Agreement between Mu-Chun Liou ("You") and Oxford University Press ("Oxford University Press") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Oxford University Press, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3421150021683
License date	Jul 03, 2014
Order Content Publisher	Oxford University Press
Order Content Publication	Nucleic Acids Research
Order Content Title	HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome:
Order Content Author	Ellen M. Westerhout, Marcel Ooms, Monique Vink, Atze T. Das, Ben Berkhout
Order Content Date	01/01/2005
Type of Use	Thesis/Dissertation
Institution name	None
Title of your work	Effect of Lysine Methylation on RNA Recognition and Cellular Uptake by Tat Derivatives
Publisher of your work	n/a
Expected publication date	Jul 2014
Permissions cost	0.00 USD
Value added tax	0.00 USD
TotalTotal	**0.00 USD**
TotalTotal	**0.00 USD**
Terms and Conditions	

Figure A-9. The license agreement for Figure 3-7