Solubility and Conformational Studies of the Intrinsically Disordered HIV-1 Tat_{1-72} Protein

By

Taras Babiak

A thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of requirements of the degree of

Master of Science

Department of Chemistry
University of Manitoba
Winnipeg

Abstract

The HIV-1 Transactivator of transcription (Tat) protein is a transcriptional activator at the site of nascent viral mRNA. Tat interacts with the Transactivation Response element (TAR) of the 5' end long terminal repeat (LTR) of all newly formed HIV-1 mRNA transcripts and in cooperation with TAR, recruits the host positive transcription elongation factor b (P-TEFb) in a process that is essential for viral gene expression and viral replication. Tat has also been implicated in numerous pathogenic processes including neurotoxicity. Although neurons of the central nervous system are not initially infected by HIV-1, extracellular Tat can be transported across the blood brain barrier. Tat is a 101-residue protein encoded by two exons. The first exon encodes a 72-residue protein that is transcriptionally active. Tat_{1-72} is an intrinsically disordered protein (IDP) at pH 4.1 as previously indicated by NMR chemical shifts and coupling constants, and confirmed by analyses of the NMR ^{15}N-relaxation parameters. Owing to the seven cysteine residues, a high net positive charge at pH 7, and a hydrophobic core region, Tat_{1-72} is highly prone to oxidative cross-linking and aggregation. The time-dependent solution stability of Tat_{1-72} in various buffers approaching neutral pH in the absence and presence of reducing agent, NaCl and various detergents was probed to determine conditions in which Tat_{1-72} could be studied at neutral pH in the presence of binding partners. With and without reducing agent and detergent, Tat_{1-72} is soluble in weakly acidic conditions, whereas in the presence of anionic SDS detergent in the presence and absence of reducing agent, Tat_{1-72} is soluble at neutral pH. The presence of SDS elicits a conformational change to α-helicity in Tat_{1-72}. In the presence of the non-ionic DDM (dodecyl maltoside) detergent and zinc, Tat was found to be soluble at pH 4 when bound to TAR RNA; TAR binding also elicits a conformational shift to α-helicity in Tat_{1-72}. The β-sheet content of Tat_{1-72} is increased in the presence of NaCl. In similar conditions,

Tat_{1-72} aggregates stained with Congo Red displayed a yellow-green birefringence and a red-shift in the Congo Red absorbance at 540 nm that is typical of β-amyloid fibril. The web-based algorithm "WALTZ", used to predict amyloidogenic regions of proteins, identifies the majority of the Tat_{1-72} hydrophobic core region as amyloidogenic. Although Tat_{1-72} is an IDP, certain regions of Tat undergo disordered-to-ordered transitions when interacting with ligands. The helical propensity of Tat_{1-72} in 2,2,2-trifluoroethanol was determined to be most prominent in residues 40 – 51 of the cysteine region and residues 64 – 74 of the hydrophobic and arginine-rich regions. These residues may be molecular recognition features (MoRFs) that enable Tat to interact with biologically relevant binding partners, in which case they may be the target of future studies of small molecule inhibitors.

Acknowledgements

First and foremost I would like to thank Joe O'Neil for my experience as a graduate student in his group. Joe's passion for research and his great attitude has developed an inquisitive rapport amongst the researchers in his laboratory. Thank you.

I would like to warmly thank my mother Helena, brother Ihor, and sister Ivanka for taking care of me. Each of you have a very special place in my life. Thank you.

I thank Dr. Khajehpour's group, Dr. McKenna's group, Dr. Stetefeld's group, and Dr. Duckworth's group for the use of their laboratory, instrumentation, and for their advice. I have grown to know each of the members of these groups and I thank them for the comraderie.

I thank my committee members Dr. Brian Mark, Dr. Jörg Stetefeld, and Dr. Joe O'Neil for taking the time out of their busy lives to read my long thesis.

Contents

List of Figures

Figure 3.9 The 24 hour absorbance spectra of 72 μM Tat_{1-72} in 10 mM acetate, 10 mM HEPES, and 20 mM SDS at pH values 4.47 – 6.67.

Figure 3.10 CD spectrum of 14 uM Tat_{1-72}, 10 mM acetate, pH 4.05 (blue), 14 uM Tat_{1-72}, 10 mM acetate, pH 4.07, and 1.75 mM SDS (green).

Figure 3.11 48 μM The absorption spectrum of Tat_{1-72} solubilised by 100 mM SDS, with 2.5 mM TCEP at various pH values.

Figure 3.12 Absorption of twenty-four hour 48 μM Tat_{1-72} in 100 mM SDS, 2.5 mM TCEP, 50 mM acetate, and 50 mM MES, pH 7.

Figure 3.13 UV absorbance spectrum of 48 μM Tat_{1-72} in 10 mM LHGP, 20 mM sodium acetate, 20 mM MES, and 2.5 mM TCEP at various pH values.

Figure 3.14 The absorbance spectra of 190 μM Tat_{1-72} in 50 mM citrate, 100 mM sodium chloride at pH values of 5.4 and 6.2.

Figure 3.15 CD spectrum of 190 μM Tat_{1-72} in 50 mM citrate, 100 mM NaCl at pH 5.5.

Figure 3.16 CD spectrum of 220 μM Tat_{1-72} in 50 mM citrate, 100 mM NaCl at pH 5.35.

Figure 3.17 The absorption spectrum of 54 μM Tat in 10 mM acetate, pH 4.05, 30 % TFE.

Figure 3.18 An overlay of CD spectra of Tat_{1-72} in 10 mM acetate at an intitial of pH 4, diluted to final TFE concentrations of 0% - 90% TFE.

Figure 3.19 The apparent change in apparent pH with the addition of TFE to 10 mM acetate at an initial pH of 4.

Figure 3.20 The mean residue ellipticity at 190 nm of 100 μM Tat_{1-72} with increasing TFE.

Figure 3.21 The mean residue ellipticity at 222 nm of 100 μM Tat_{1-72} with increasing TFE.

Figure 3.22 Average secondary structure deconvolution of Tat_{1-72} as a function of solution TFE composition.

Figure 3.23 CD spectra of 100 μM Tat_{1-72} in 20% TFE, 8 mM TCEP, 8 mM acetate, apparent pH 4.2 from 50 °C – 80 °C.

Figure 3.24 CD Spectra of 140 μM Tat_{1-72} in 10 mM acetate, apparent pH 4.28, diluted to 30% TFE (v/v), in the presence of increasing concentrations of zinc(II).

Figure 3.25 The mean residue ellipticity at 205 nm of 140 μM Tat_{1-72} in 10 mM acetate, apparent pH 4.28, diluted to 30% TFE (v/v), in the presence of 0.5, 1, 2, 3, and 4 MEQ's of zinc (II).

Figure 3.26 CD Spectra of 100 μM Tat_{1-72} in the presence of 7 mM acetate, 7 mM MES, 30% TFE and 2 MEQ's of zinc(II) over a range of apparent pH from 4.69 to 7.20.

Figure 3.27 Absorption spectra of 110 μM Tat_{1-72} in 30% TFE, 7 mM acetate, 7 mM MES, 2 MEQ's Zn(II), apparent pH 7.2, and the same sample at pH 6.9 after 72 hours.

Figure 3.28 ^{1}H-^{15}N HSQC spectrum at 298 K of ^{15}N enriched, His-tagged Tat_{1-72} in 10 mM acetate at pH 4 in the presence of 10 mM TCEP reducing agent.

Figure 3.29 ^{1}H-^{15}N HSQC spectrum at 298 K of ^{15}N enriched, His-tagged Tat_{1-72} in 10 mM acetate at pH 4 in the presence of 10 mM TCEP reducing agent shown in Figure 3.28 with resonance assignments.

Figure 3.30 ^{1}H-^{15}N HSQC relative peak intensities of Tat_{1-72} in the presence of reducing agent at pH 4.

Figure 3.31 ^{15}N chemical shift differences of Tat_{1-72} HSQC cross-peaks (Figure 3.29) from random coil values corrected for local sequence effects.

Figure 3.32 $^{1}H^{N}$ chemical shift differences of Tat_{1-72} from random coil values corrected for local sequence effects.

Figure 3.33 ^{1}H-^{15}N HSQC spectra of backbone amide resonances for ^{15}N enriched Tat_{1-72} in 0% TFE and 7% TFE overlayed at 298 K.

Figure 3.34 ^{1}H-^{15}N-HSQC amide backbone resonance assignments at 298 K for ^{15}N enriched, His-tagged Tat_{1-72} in 9 mM TCEP with 7% TFE, apparent pH 4.01.

Figure 3.35 Relative ^{1}H-^{15}N HSQC peak intensities of His-tagged Tat_{1-72} in 7% (v/v) TFE, 9 mM acetate, 9 mM TCEP, apparent pH 4.01.

Figure 3.36 ^{15}N chemical shift differences from random coil values corrected for local sequence effects of Tat_{1-72} in 7% TFE.

Figure 3.37 $^{1}H^{N}$ chemical shift differences from random coil values corrected for local sequence effects of Tat_{1-72} in 7% TFE.

Figure 3.38 ^{1}H-^{15}N-HSQC spectra of ^{15}N enriched Tat_{1-72} in 7% and 15% TFE at 298 K.

Figure 3.39 ^{1}H-^{15}N HSQC backbone amide resonance assignments at 298 K for ^{15}N enriched, His-tagged Tat_{1-72} with 8.5 mM TCEP in 15% TFE at apparent pH 4.05.

Figure 3.40 Relative ^{1}H-^{15}N peak intensities of His-tagged Tat_{1-72} in 15% (v/v) TFE, 8.5 mM acetate, 8.5 mM TCEP, at apparent pH 4.05.

List of Tables

decreasing order, with the exception of unassignable residues.

List of Abbreviations

A	Absorbance
AIDS	Acquired Autoimmune Disorder
ARM	Arginine rich motif
BBB	blood-brain barrier
bHLHZip	basic helix-loop-helix-leucine zipper
BMEC	Brain microvascular endothelial cells
CA	Capsid protein
cAMP	cyclic adenosine monophosphate
CPMG	Carr-Purcell-Meiboom-Gill
CBP	CREB-response element binding protein
CD	Circular dichroism
Cdk9	Cyclin dependant kinase 9
CMC	critical micellar concentration
CNS	central nervous system
CPP	Cell penetrating peptide
CR	Congo Red
CRE	cAMP response element
CREB	cAMP response element-binding
CSI	Chemical shift indexing
CTD	carboxy terminal domain
CV	column volume

CycT1	Cyclin T1
ddH_2O	double deionized water
DDM	n-dodecyl beta-D-maltoside
DNase	deoxyribonuclease
DRIB	DRB sensitivity-inducing factor
drkN SH3	N-terminal SH3 domain from the adapter protein drk
DSS	2,2-dimethyl-2-silapentane-5-sulfonate
EIAV	Equine infectious anemia virus
Env	Envelope proteins
Gag	group-specific antigen
Gdm^+	Guanidinium cation
Gdn-HCl	guanidine-hydrochloride
GST	Glutathione S-transferase
HIV	Human Immunodeficiency Virus
HMQC	Heteronuclear Multiple Quantum Coherence
HSQC	Heteronuclear Single Quantum Coherence
HTLV	Human T-lymphotrophic Virus
IAPP	Islet amyloid polypeptide
IDP	Intrinsically disordered protein
IN	Integrase
INEPT	Insensitive nuclei enhanced by polarization transfer
IPTG	Isopropyl-b-D-thiogalactopyranoside
IR	infra red spectroscopy

LRP	Low-density lipoprotein receptor related
LTNPs	Long term non-progressors
LTR	Long Terminal Repeat
MA	Matrix protein
Max	myc-associated factor X
Meq	molar equivalents
MES	2-(N-morpholino)ethanesulfonic acid
MoRFs	Molecular recognition features
MRE	mean residue ellipticity
Myc	Myc oncoprotein
NC	Nucleocapsid protein
Nef	Negative Regulatory Factor
NELF	Negative elongation factor
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect spectroscopy
PIC	Pre-integration assembly
Ppm	parts per million
PR	Protease
P-TEFb	Positive Transcriptional Elongation Factor b
Rev	Regulator of virion
RCPL	Right circularly polarized light

RF	Radio frequency
RNAPII	RNA Polymerase II
RNase	Ribonuclease
RT	Reverse Transcriptase
S.D.	Standard deviation
snRNP	small nuclear ribonucleoprotein
SU	Surface unit
TAR	Transactivating response element
Tat	Transactivator of transcription
TB	Terrific Broth
TCEP	Tris-2-carboxyethyl phosphine
TFE	2,2,2-trifluoroethanol
TM	Transmembrane complex
TOCSY	Total correlation spectroscopy
TRX	Thioredoxin
UV	Ultraviolet
Vif	Viral infectivity factor
Vpr	Viral protein R
VSFS	Vibrational sum frequency spectroscopy
YPrA	Yeast aspartic proteinase A

Chapter 1

1. Introduction

1.1 The Human Immunodeficiency Virus

In 2009, there was an estimated 2.6 million people who became newly infected with Human Immunodeficiency Virus (HIV) and in 2010, the UNAIDS report estimated that 33.3 million people are infected with the HIV[4]. However, the annual numbers of new infections is declining and HIV carriers are living longer because of the reduction in HIV caused acquired immunodeficiency syndrome (AIDS) related deaths due to antiretroviral therapy[4]. Efforts to control the progression to AIDS have been focused on the biology, biochemistry, and structure biology of HIV and the interactions between viral components and drugs. Among the first viral inhibitors were targeted towards the viral reverse transcriptase and protease inhibitors, and treatment often involved a combination of inhibitors[5]. This type of treatment does not prevent HIV infection, but delays the onset of AIDS, and gives rise to drug resistant mutants[5]. Strains of HIV that are resistant to current HIV treatment develop due to the error-prone reverse transcription step of the viral lifecycle, and a very high frequency of recombination during viral integration. One study estimates that point mutations in HIV occur 10^4 to 10^5 time per day in the total viral load of a patient in a latent state of infection[6]. With the rise of mutant viruses that are unresponsive to current HIV treatment, and a lack of effective vaccines, it is crucial to develop HIV therapies that target interactions between viral proteins and host proteins.

HIV has been known to be the causative agent of AIDS since the early 1980's. An unknown virus was independently isolated from AIDS patients[7, 8] that showed similar

pathogenicity to T-cell viruses[9], and utitlized a magnesium dependant reverse transcriptase[10, 11]. By 1986, it became clear through phylogenetic studies that the cause of AIDS was a lentivirus[12-14]. HIV-1 (Figure 1.1) contains two, positive sense RNA strands encoding a reverse transcriptase, and can persist dormant as a latent infection through a host-integrated DNA proviral state[15]. It is CD4+ cell tropic and initial interaction occurs between the viral outer membrane protein complex and cells displaying the CD4+ molecule[16]. The host cell-derived membrane surrounding the virus is penetrated by a glycoprotein complex responsible for CD4+ cell specificity. The complex consists of a trimeric association between gp120 trimers and gp41 trimers[17]. A depression between the outer and inner domains of the gp120 trimer facilitates initial non-covalent interactions with a CD4+ molecule[18]. The outer membrane proteins are anchored in a spherical protein complex of matrix proteins (MA). Within the matrix exists the conical complex of capsid proteins (CA) enclosing key enzymes, nucleocapsid proteins (NC), and the viral genome [19].

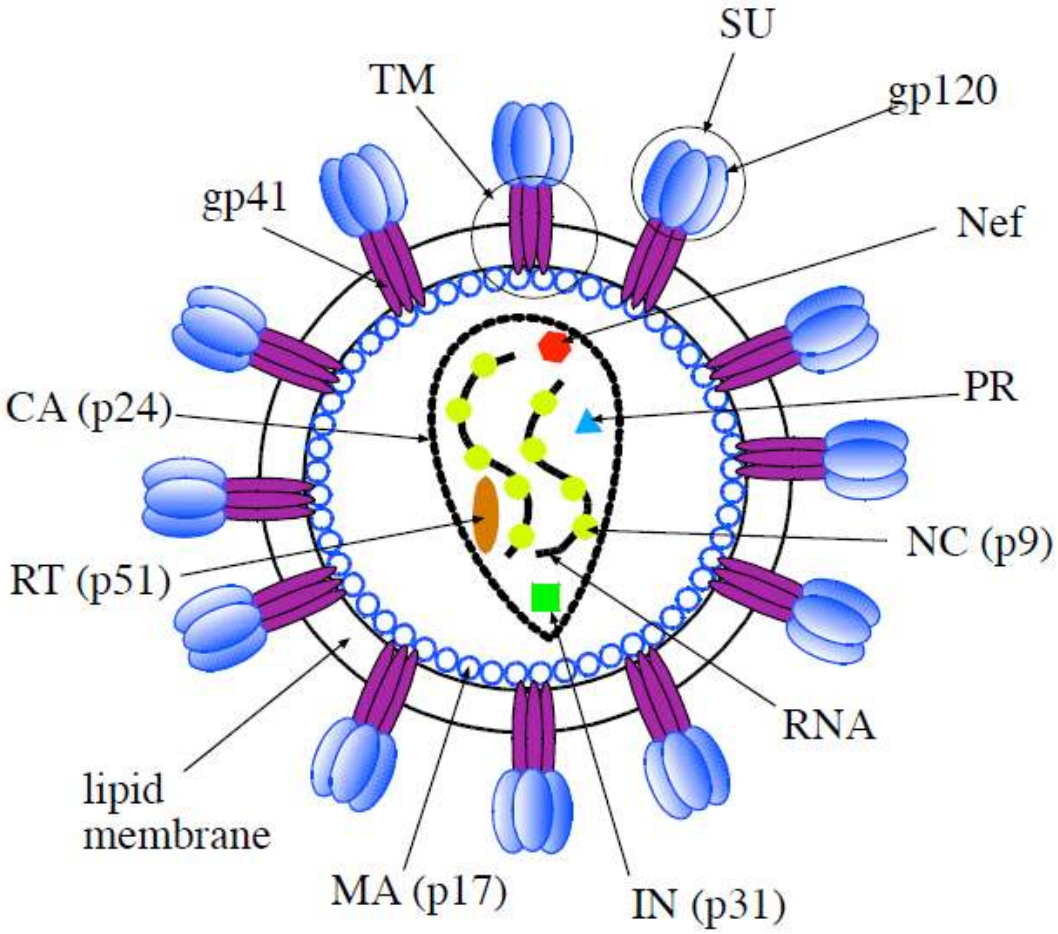

Figure 1.1 A schematic representation of the HIV-1 virion. The virion is encapsulated by a host cell-derived lipid membrane. The trimeric gp41 proteins make up the transmembrane (TM) complex which are attached to the surface unit (SU) made up of trimeric gp120. The matrix (MA) is beneath the lipid membrane and contains the conical capsid (CA) and a complement of viral enzymes, diploid, positive sense RNA, and nucleocapsid proteins (NC). The enzymes include reverse transcriptase (RT), integrase (IN), protease (PR), and the protein negative regulatory factor (Nef). Image reproduced with permission from S. Shojania[20].

The HIV-1 lifecycle (Figure 1.2) begins with the initial recognition and binding of the CD4+ molecule by gp120 on T-lymphocytes. A conformational change allows gp120 to interact with co-receptors (mainly CCR5 or CXCR4). Co-receptor specificity has been attributed to a variable loop (V3) recognition domain of gp120[21]. Residue mutations within the V3 loop region change the co-receptor specificity and vary cell tropism between cells of the immune system displaying the CD4+ receptor[21, 22]. The V3 loop domain of the virus may be a key determinant in viral transmission to other immune cells displaying the main CD4+ receptor, thus making it a target for therapeutic trials and vaccine development. In its inactive state, gp41 is concealed, but upon gp120 co-receptor interaction, conserved structural elements of gp41 are exposed and inserted into the host membrane for viral fusion[23]. It's possible that the structural concealment of the conserved gp41, and the variability of the gp120 recognition domain allows the virus to evade immune pressure[22].

The viral capsid is released into the cell following fusion of the host cell-derived viral membrane and the host lipid membrane. Along with two copies of its 10 kb RNA genome[24], the uncoating of the viral capsid releases key proteins to its early lifecycle: a protease (PR), integrase (IN), reverse transcriptase (RT), nucleocapsid (NC) and three accessory proteins (Nef, Vif, Vpr). PR is responsible for cleaving viral poly-proteins into their active forms. It is structurally and functionally similar to other aspartyl proteases of the pepsin family[25], containing the catalytically active triad Asp25-Thr26-Gly27. PR is the most crystallized HIV

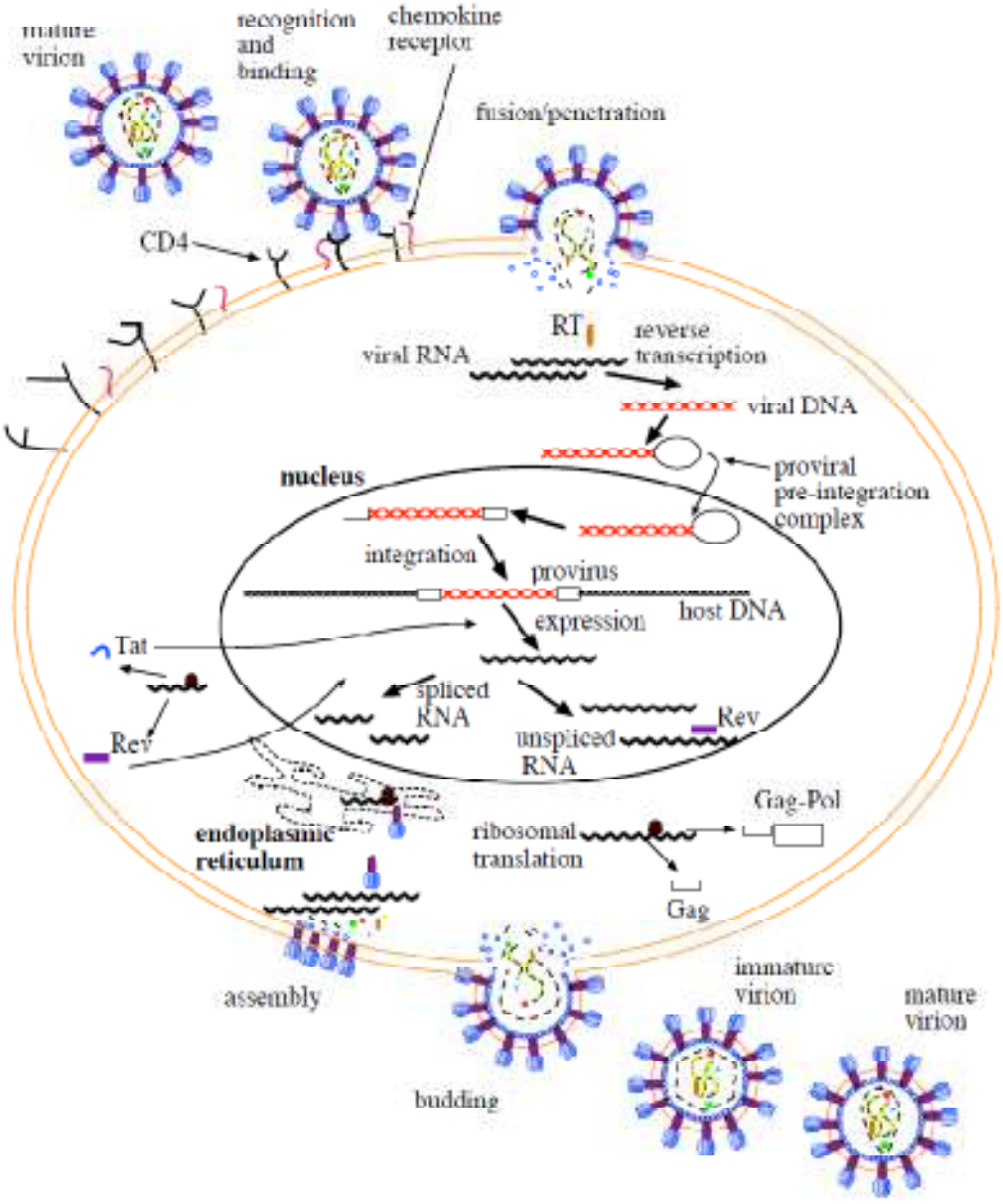

Figure 1.2 The lifecycle of HIV-1. Image reproduced with permission from S. Shojania[20].

protein, and many PR-inhibitor complexes have been solved, and have led to the development of structure-based drug design of PR inhibitors that are in wide clinical use[26]. RT is responsible for the reverse transcription of the viral RNA shortly after viral fusion. RT processes viral RNA to form viral DNA, but can use both RNA and DNA as template for DNA synthesis[27]. IN catalyzes the integration of viral DNA into the host genome[28]. Negative regulatory factor (Nef) is responsible for downregulating the presentation of the CD4+ receptor at the host surface by promoting the internalization of the receptor and trafficking from the membrane and Golgi apparatus to lysosomes for degradation[29]. This process has been implicated in the prevention of super-infection and in effective viral packaging by envelope (Env) proteins gp120 and gp41. Viral protein R (Vpr) acts as a nuclear localization signal; as the viral capsid is uncoated in the cytoplasm, a nucleoprotein complex believed to consist of MA, IN, NC, and RT bound to viral RNA and viral DNA (referred to as the pre-integration complex) is transported to the nucleus as mediated by Vpr, although there is disagreement about the involvement of MA and NC in this complex[30, 31]. Viral infectivity factor (Vif) binds to host defence factor human APOBEC3G and mediates its elimination through ubiquitination and degradation by the proteasomal pathway[32].

The HIV-1 genome, like all primate immunodeficiency viruses, has three main genes: *gag*, *pol*, and *env* (Figure 1.3). The existence of nine open reading frames and multiple protease cleavage sites account for fifteen distinct proteins[33]. Gag, Pol, and Env are synthesized as polyproteins and cleaved into their functional units by PR. Gag proteins are structural, and include NC, CA, MA and p6. p6 is believed to aid in viral assembly prior to budding[34]. The pol polyprotein is cleaved into RT, PR, and IN proteins. The env polyprotein gp160 is

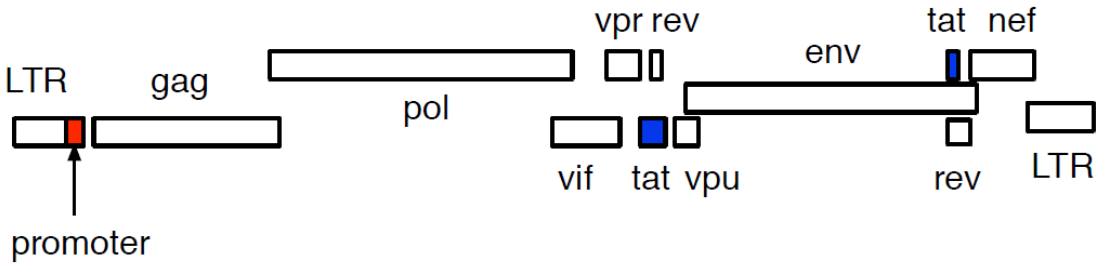

Figure 1.3 The open reading frames of the HIV-1 genome. Group-specific antigen (gag) encodes a polyprotein for viral structural components (matrix, capsid, and nucleocapsid proteins). Polyprotein (pol) encodes viral enzymes integrase, protease, and reverse transcriptase. Vif encodes viral infectivity factor; vpr encodes viral protein R; 2 tat exons (blue) encode the Transactivator of transcription; rev encodes regular expression of virus; vpu encodes viral protein U; nef encodes negative factor. Env encodes a polyprotein for the envelope proteins gp120 and gp41. Image reproduced with permission from S. Shojania[20].

proteolyzed to become the outermembrane protein gp120 and transmembrane protein gp41[35]. The additional six reading frames code for the regulatory proteins Tat and Rev, along with the accessory proteins Vif, Vpr, Vpu, and Nef. The Transactivator of transcription (Tat) is essential to viral replication, and is responsible for activating the transcription of viral DNA at the site of the Transactivation response element (TAR) stem-loop structure found in the 5' end long terminal repeat (LTR) regions of viral mRNA (Figure 1.3)[36]. It has been implicated in many viral processes and its multifunctional nature is the topic of this thesis. Regulator of virion (Rev) functions to export from the nucleus unspliced viral mRNA for the translation of full length viral proteins, and the packaging of full length viral RNA into progeny virions[37]. Rev interacts with the Rev response element (RRE) located in the env coding region[38] and via its leucine-rich

nuclear export signal[39], facilitates the export of full length, or singly spliced viral mRNA[37]. In the absence of Rev, mRNA becomes multiply spliced in the nucleus, encoding the Tat, Rev and Nef proteins. Viral protein U (Vpu) has been shown to increase the degradation of newly synthesized CD4 proteins in the endoplasmic reticulum, thus preventing internal interaction with gp120 and gp41, allowing their transport to the host cell surface for assembly[40].

HIV transcription will produce three classes of mRNA transcripts: full length transcripts of about 10 kb, singly spliced transcripts of about 4-5 kb, and multiply spliced transcripts of about 2 kb[41]. Different mRNAs are transcribed during different phases of the viral lifecycle. During the early phase, the regulator proteins Tat and Rev along with Nef are produced from multiply spliced mRNA's[42]. Tat upregulation at the site of the 5' end LTR mRNA greatly increases DNA processivity[43], while an increase in Rev promotes the export of full or singly spliced transcripts to the cytoplasm for synthesis of replicative enzymes and structural proteins for virion assembly[39]. The transition from low to high levels of viral replication coincides with a threshold level of Rev[44]. An abundance of full length and 4 kb transcripts predominates during the late phase of replication.

1.1.1 HIV-2 Subtype

Soon after the discovery of HIV-1, a separate sub-type was discovered in Western Africa[45]. Early sequence analysis showed that the genetic code for the Env proteins of the new sub-type differed by more than 55 %, whereas the Gag and Pol genes differed by about 40 %[46]. Given the fact that a high proportion of HIV-2 carriers are deemed long term non-progressors

(LTNPs) and that the virus is restricted endemically, it is not surprising that the majority of research has been focused on HIV-1, but the attenuated pathogenicity of HIV-2 may be relevant for HIV-1 therapy[47]. The relative transmission inefficiency of the virus is reflected by a long term study of Senegalese sex workers. From 1985 to 2003, HIV-2 prevalence has declined from 8 % to 5.5 %, while HIV-1 increased from 1 % to 13.8 %[48]. The relative transmission inefficiency of the virus can be simply observed by the rates of vertical viral transmission. In the absence of antiretroviral therapy, the transmission of HIV-2 from infected mothers to newborns is less than 4% compared to 24.4% for HIV-1[49]. Transmission efficiency is directly correlated to plasma viral load, and it has been observed that transmission of HIV-2 via sexual contact is lower in comparison to HIV-1 and that the levels of HIV-2 in semen are related to the plasma viral load[50].

Although the clinical syndromes of HIV-2 AIDS patients is indistinguishable from that of HIV-1 AIDS patients, there is a higher proportion of non-AIDS HIV-2 carriers, $CD4^+$ T cell decline is much slower, and there are fewer instances of the opportunistic Kaposi sarcoma[51]. Baseline plasma viral load has been a good predictor of survival, and in comparison of HIV-1 and HIV-2 LTNPs, most asymptomatic, low-risk HIV-2 patients have undetectable plasma viral RNA[52]. The notion that HIV-2 is an attenuated version of HIV-1 has been refuted by *in vitro* studies showing little difference in cytopathology[53]. *In vivo*, the viral load and plasma viral RNA levels during asymptomatic stages of infection are about 30-fold lower in HIV-2 than HIV-1[54]. Pro-viral loads at corresponding stages of $CD4^+$ T cell suppression are similar between the subtypes, but plasma viral load is higher in HIV-1[55]. This suggests that the sub-types have a similar efficiency in attaining a pro-viral state, but lower levels of viral mRNA and a lower

plasma viral load of HIV-2 suggests either tighter control of viral expression, or a lower replicative fitness.

1.2 The HIV-1 Transactivator of Transcription

1.2.1 Tat's Role as a Regulator

Tat is a two exon (Figure 1.3), 101 amino acid protein (Figure 1.4). The first exon codes for the first 72 amino acids, and the second exon codes for the last 29 amino acids. Tat is conserved among lentiviruses, and is essential to viral replication[56]. An 86 amino acid version of Tat arose from passaging in strains HXB2 and LAI, although no version of this protein has been isolated *in vivo*[57]. This truncated version of Tat may be an artifact of a point mutation introducing a premature termination codon[58]. Tat arises from three different multiply spliced viral mRNA's in the host nucleus early on in the viral lifecyle[59]. In the absence of Tat, host RNA Polymerase II (RNAP II) is inefficient in processing viral DNA, resulting in the premature termination of viral transcripts[60]. In the presence of Tat, processivity is enhanced over a thousand-fold.

Tat is an intrinsically disordered protein[20] composed of functional regions that do not constitute independent structural domains and are largely based on amino acid composition (Figure 1.4). The first exon consists of an acidic, proline-rich N-terminal region (1-21), a cysteine-rich region (22-37), a hydrophobic core region (38-47), a basic, arginine-rich region (48-57), and a glutamine-rich region (58-72). The first exon product has been shown to be

Residues 1 -21, Proline-rich Region
MEPVDPRLEPWKHPGSQPKTA

Residues 22 – 37, Cysteine-rich Region
CTNCYCKKCCFHCQVC

Residues 38 – 48, Hydrophobic core Region
FITKALGISYG

Residues 49 – 59, Arginine-rich Region
RKKRRQRRRPP

Residues 60 – 72, Glutamine-rich Region
QGSQTHQVSLSKQ

Residues 73 – 101, Exon 2 Segment

Figure 1.4 The amino acid sequence of the HIV-1, BH10 isolate Tat[61].

transcriptionally active, where conservation in the cysteine rich region and the basic region are critical to transcriptional efficiency[62]. Although the first exon is transcriptionally active, it has been suggested that the second exon enhances transactivation[63], is critical for nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) activation[63], is involved in blood-brain barrier (BBB) penetration by cellular uptake[64], and enhances T-cellular apoptosis[65].

Tat specifically interacts with the TAR element of the 5' end LTR region of all viral transcripts by its arginine rich motif (ARM) of the basic region (residues 48-57 of the first exon). TAR is a 59 base stem-loop structure (Figure 1.5) that forms in nascent viral transcripts. It contains a 6 nucleotide loop with a characteristic uridine23-cytidine24-uridine25 bulge that is important for Tat recognition[66]. Tat binds specifically by its basic region to the bulge, and there

is evidence that this region folds into a helix, allowing specific presentation of basic residues into the groove of the TAR stem-loop[67].

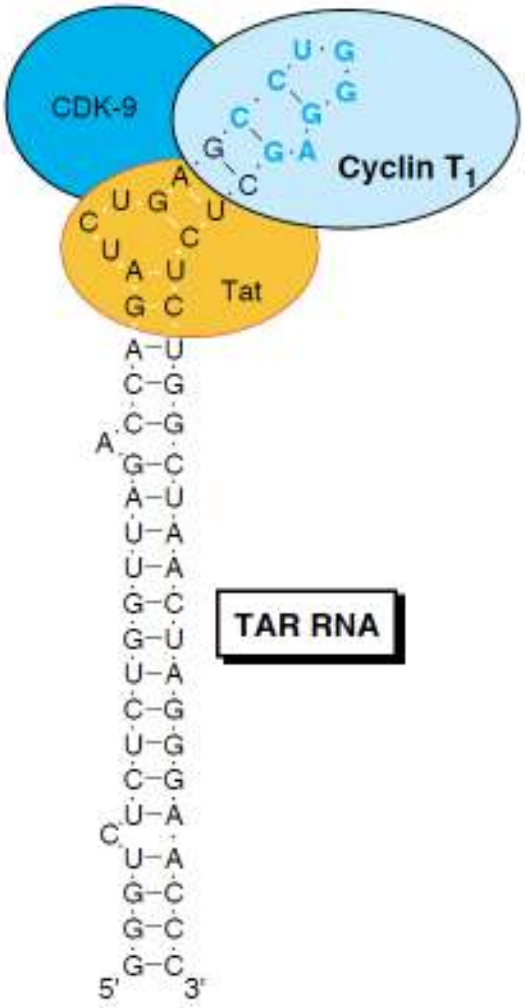

Figure 1.5 The TAR/Tat/P-TEFb interaction complex. Reproduced with permission from Elsevier and J. Karn[68].

Tat is a transcriptional activator at the site of viral mRNA because it acts to increase processivity of RNAPII. Tat recruits the positive transcription elongation factor b (P-TEFb) by a highly co-operative interaction with the TAR element (Figure 1.5). The role of P-TEFb in uninfected cells is to promote transcriptional elongation by activating RNAPII that is paused at

promoter sites and a feature common to about 70 % of all human genes[69]. P-TEFb is composed of Cyclin T1 (CycT1) and Cyclin dependant kinase 9 (Cdk9). X-ray diffraction studies have shown that P-TEFb binds to Tat by interaction with the proline-rich, cysteine-rich zinc-finger, core, and basic regions of Tat, and mediated with numerous Van der Waal and water contacts between the pair[70]. Another X-ray diffraction study shows a tripartite interaction between CycT1-Tat-TAR, mediated by the basic region of Tat[67]. The Cdk9 component of P-TEFb activates RNAPII by hyperphosphorylation at its C-terminal domain. Early studies have shown that Tat-TAR specificity is governed by specific interactions of Tat's arginine-rich region and the flanking glutamine and hydrophobic regions with TAR[71]. Phosphorylation of the carboxy terminal domain (CTD) of RNAPII by Cdk9 of P-TEFB is achieved when Tat bound to P-TEFb in a pre-formed complex interacts with the TAR element of halted mRNA transcripts[72]. Tat acetylation by p300/ CREB-response element binding protein (CBP) and p300/CBP-associated factor (PCAF) at Lys-50 inhibits the Tat/TAR interaction[73]. In another study of HIV-1, Tat acetylation at Lys-28 by PCAF enhances the stablility of the CycT1-Tat-TAR complex[74].

Tat's role as a regulator of transcription at the level of mRNA is accomplished by its ability to activate Cdk9 kinase to phosphorylate the CTD of RNAPII (Figure 1.6)[75]. Recently, new research shows the manner in which P-TEFb activates transcriptional elongation by Tat recruitment[72, 76, 77]. In its inactive form, P-TEFb is associated with the 7SK small nuclear ribonucleoprotein (snRNP) in a complex including the inhibitory protein Hexim1 and/or Hexim2, LARP7, SART3, and MePCE (a 5'-end 7SK capping enzyme)[78]. Hexim1 and/or Hexim2 association with a 5'-end hairpin of the 7SK snRNP represses the kinase activity of P-TEFb[78]. In the absence of Tat, it has been shown that the 7SK snRNP assembly can associate

with the pre-integration complex (PIC) consisting of transcriptional factors on the HIV-1 promoter[76]. During initial, inefficient HIV-1 transcription, Tat is synthesized and associates with the 7SK snRNP complex. Tat may associate with the 7SK snRNP complex before

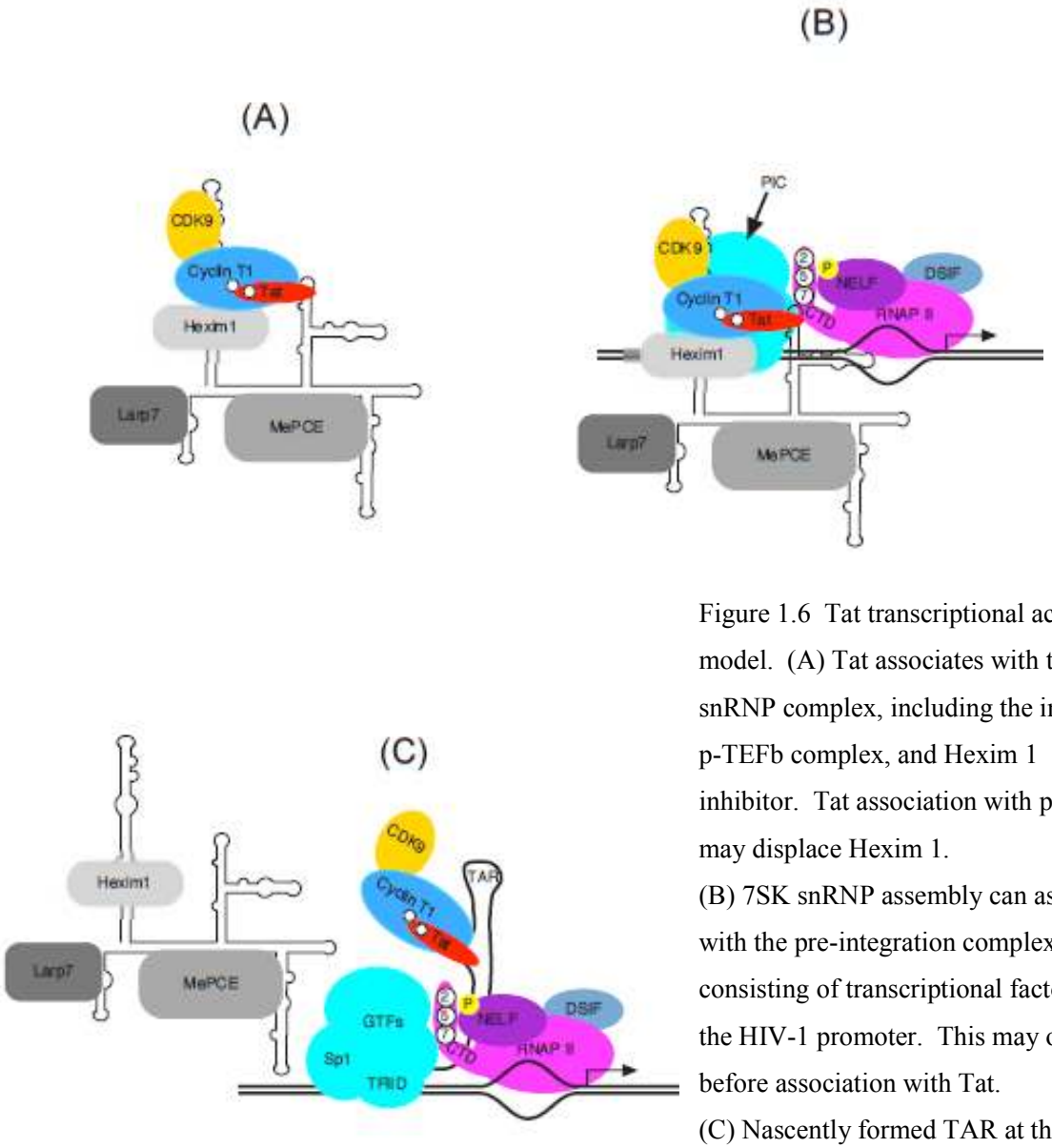

Figure 1.6 Tat transcriptional activation model. (A) Tat associates with the 7SK snRNP complex, including the inactive p-TEFb complex, and Hexim 1 inhibitor. Tat association with p-TEFb may displace Hexim 1.
(B) 7SK snRNP assembly can associate with the pre-integration complex consisting of transcriptional factors on the HIV-1 promoter. This may occur before association with Tat.
(C) Nascently formed TAR at the 5'-end of viral mRNA transcripts forms a tripartite interaction with Tat/p-TEFb mediated by Tat. The 7SK snRNP is displaced from the pre-integration complex.

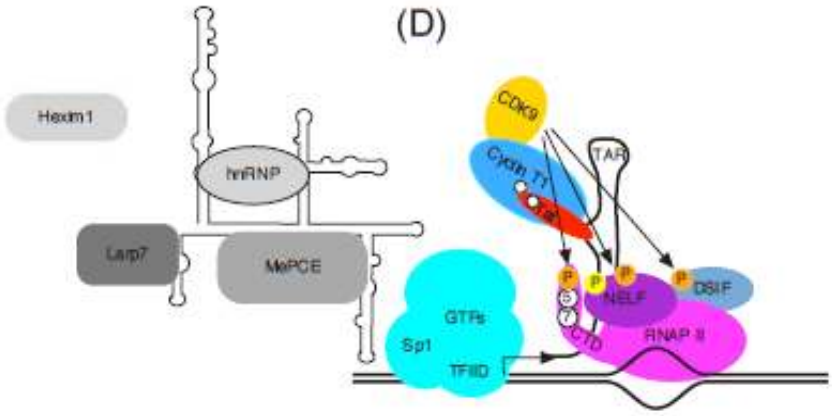

Figure 1.6 continued: (D) Carboxy-terminal domain phosphorylation at Ser-2 of RNAPII, DSIF, and NELF by Cdk9 initiates elongation. Figure reproduced with permission from S. Shojania[79].

association with the PIC in compartmentalized regions of the nuclear space, referred to as nuclear speckles, where P-TEFb and 7SK snRNP association takes place[80]. Tat relieves P-TEFb inhibition by displacing Hexim1 or altering the conformation of P-TEFb, which in turn displaces Hexim1. A multisubunit negative elongation factor (NELF) and DSIF/Spt4-Spt5 (DRB sensitivity inducing factor) arrests RNAPII transcriptional elongation after promoter clearance[81]. The nascently formed TAR at the 5'-end of viral mRNA transcripts forms a tripartite interaction with P-TEFb mediated by Tat. It is believed that once the P-TEFb/Tat/TAR interaction is formed, the 7SK snRNP complex is displaced from the PIC and phosphorylation at Ser-2 of RNAPII, by Cdk9 initiates elongation. Phosphorylation of DSIF and NELF also regulates transcriptional activation by relieving their inhibition.

1.2.2 Other Biological Roles of Tat

The scientific literature on HIV-1 Tat contains thousands of reports of the effects of Tat on cells, metabolic and signaling pathways and of direct interactions with many proteins. A proteomics study involving the 1-86 amino acid Tat variant was conducted by subjecting Jurkat T-cell nuclear extracts to affinity chromatography containing immobilized glutathione S-transferase (GST) tagged 1-86 amino acid Tat, and a control column containing only GST[(82)]. Following extensive washes to remove non-specific interactions, the proteins were eluted under denaturing conditions, separated by 1-D sodium dodecyl sulfate polyacrylamide gel electrophoresis, digested by trypsin, and submitted to tandem mass spectrometry analysis for identification. Following an *in silico* analysis, an interaction profile showed that Tat interacts with 183 proteins in total. A motif analysis revealed that Tat interaction involves proteins containing domains for mediating protein, DNA, RNA, ATPase, and helicase activities. This suggests that Tat's biological role is one involving gene replication, expression and regulation, chromatin remodeling and organization, and nuclear organization with the most represented process being transcription (39 %).

Neuronal loss and neurogenic inflammation occurs in patients infected with HIV, although neurons are not initially infected[(83)]. One hypothesis that has been put forward is that neurotoxicity is mediated by HIV-1 Tat. Tat is actively secreted into the extracellular environment by infected CD4+ presenting T-cells and macrophages, and can be taken up by neighbouring, uninfected cells like neurons, astrocytes and glial cells[(84)]. Tat can also be transported across the BBB[(85)], is neurotoxic, and capable of inducing oxidative stress in a dose-dependent manner[(86)]. Tat has been detected in the sera of HIV-infected patients at low

nanomolar concentrations, which may be an underestimation since Tat may become associated with heparin sulfate (a protein widely expressed on cell surfaces)[87], whereas *in vivo* measurements are difficult because Tat antibodies have weak affinity[88]. Tat mRNA levels are elevated in brain extracts from individuals with HIV-1 dementia and Tat has been hypothesized to be a contributor to HIV-dementia[89, 90].

The aggregation and adhesion of rat cerebellum neurons was promoted by full length Tat *in vitro*, an effect that appears to be caused by the basic, arginine-rich region[91]. Monoclonal antibodies raised against the arginine-rich region prevented and reversed the aggregation of rat cerebellum neurons *in vitro*, but antibodies raised against the proline-rich region had no effect[91]. The regions of the brain that are particularly susceptible to Tat toxicity are the striatum, hippocampal dentate gyrus, and the CA3 region of the hippocampus[92, 93]. Tat-induced neurodegeneration is believed to involve an excitotoxic mechanism involving calcium. When interacting extracellularly, Tat is capable of depolarizing human cortical neurons, increasing intracellular Ca^{2+}, and initiating neuronal apoptosis[94-96]. Of particular interest to us is that the central nervous system (CNS) regions susceptible to Tat toxicity and the pathological effects of the protein are similar to those observed for β-amyloid in Alzheimer's disease[97].

HIV-infected cells enter the CNS by a proposed "Trojan horse" strategy[98], and the disease related deterioration of the CNS is referred to as NeuroAIDS. HIV-infected monocytes in peripheral tissue migrate across the BBB. The virus begins to replicate efficiently once the host monocyte differentiates to a macrophage[99]. Following CNS invasion, infected macrophages actively produce and release HIV, as well as a variety of proinflammatory factors,

which adversely affect the function of surrounding cells, resulting in HIV-associated encephalopathy and HIV-associated dementia. Once the CNS becomes infected, inflammatory conditions associated with NeuroAIDS can increase the rate of leukocyte migration, stimulating more infected monocytes to cross the BBB[100]. CNS infection occurs concurrently with the dysregulation of tight junction proteins produced by brain microvascular endothelial cells (BMEC) that are responsible for maintaining the integrity of the BBB. Blood-borne Tat has been shown to cross the BBB and has been implicated in disrupting the BBB[101]. A study has shown that Tat can induce the phosphorylation of focal endothelial kinase, leading to increased BMEC permeability by altering focal adhesion assembly[102].

Tat neurotoxicity can occur by specific interaction with the neuronal cell membrane when exposed extracellurarly, but when introduced intracellularly, it does not alter the neuronal membrane potentials[103]. Tat binds to low-density lipoprotein receptor related-protein (LRP) on neurons, becomes internalized, and transported to the neuronal nuclei[104]. The arginine-rich, basic region of Tat is classified as a cell penetrating peptide and contains a nuclear localization signal[105]. Any deletion to this basic region (47-57) reduces the membrane translocation activity of the protein[106]. Peptides designed to contain the arginine-rich motif have been used to transport into cells a variety of materials including proteins, genes, antibodies, magnetic nanoparticles, drugs, and liposomes[107].

Intracellular and extracellular activities of Tat have been attributed to the unique sequence domains contained within the protein. Residues 1-24 of the Tat protein encompassing the whole of the acidic, proline-rich region and residues 22-24 of the cysteine-rich region interact

with the co-activator and acetyltransferase CBP (CREB-response element binding protein) KIX domain binding site[108]. CREB (cAMP response element-binding) is a transcriptional factor influenced by cAMP[109]. CREB influences transcription by recognizing the upstream cAMP response elements (CRE) of DNA and is co-activated by CBP. CBP interacts with numerous mammalian and viral transcription activators, coactivators, corepressors and the transcriptional machinery[110]. The presence of numerous, independently functioning domains of CBP has been shown to be a key regulator of basic cellular processes[111]. Aberrant influence of the co-activator by competition between cell regulatory proteins and viral regulators during infectivity may activate apoptotic pathways. The interaction between Tat and CBP may be an important mechanism for histone acetylation leading to chromatin re-modeling for the accessibility of transcription machinery to integrated proviral DNA[112].

Zinc co-ordination by cysteine and histidine groups often stabilizes structures that confer specificity to nucleic acids and proteins. Zinc co-ordination occurs in proteins involved in cell cycle regulation, gene expression and repair, cell growth, differentiation, signaling, and metabolism. They are important in stabilizing both structural components and higher energy transition state complexes of oxidoreductases[113], esterases[114], and reductases[115]. The Tat protein contains a cysteine-rich region (residues 22-37) that behaves like a zinc-finger. The protein has been shown to form homo-dimeric complexes with zinc[116] and forms a zinc-bridged interaction with CycT1[117]. A mutation in any six of the seven cysteines eliminates transactivation[118]. Structurally, Tat interaction with the human P-TEFb complex requires a specific complementary Cys residue at position 261 of P-TEFb that is not conserved among murine elongation factors[119]. C261Y P-TEFb mutants form low affinity interactions with Tat

that are independent of zinc, and cause poor transactivation *in vivo*. Tat can induce apoptosis in non-infected cells by a mechanism dependent on zinc, and by a specific interaction between host microtubule assembly and four conserved residues of the core domain[120-122].

The necessity for studying the Tat/zinc interaction stems from the *in vitro* evidence that Tat forms a zinc-bridged association with the human cyclin T1 of the P-TEFb required for successful synthesis of viral mRNA. Tat interaction with zinc by its cysteine-rich region may prevent Tat oligomer formation by covalent cross-linking between cysteine groups and subsequent precipitation. Zinc co-ordination with deprotonated thiol groups of cysteine has been shown to be effective in preventing unwanted oxidation[123].

The glutamine-rich region (residues 60-72) has been implicated in T cell apoptosis[124]. A comparison of sequence homology of Tat between a rapid progressing patient and a long term survivor reveals a mutation in the glutamine region[125]. Equine infectious anemia virus (EIAV) Tat contains a glutamine-rich region that is truncated in comparison to HIV-1 Tat[67]. In a study of the 86 residue version of Tat, the glutamine region showed the most long range Nuclear Overhauser Effects (NOE)[125], suggesting a conformational homogeneity. A structure-function comparison between subtype D Tat proteins from patients with different rates of disease progression revealed a short α-helix in the glutamine-rich segment of Tat isolated from rapid disease progressors[126]. In this study, it was observed that the rate of uninfected T-cell apoptosis was greater in the presence of Tat isolated from a rapid disease progressor. Structural heterogeneities were shown by circular dichroism (CD) studies and molecular modeling[126].

1.3 Intrinsically Disordered Proteins

Since the inception of three-dimensional study of proteins to an atomic resolution, there has been an assumption that the biological function of a protein is rooted in its three-dimensional structure, this being central to the structure-function paradigm[127]. Presently we understand that proteins can occupy diverse states of conformation ranging from well folded structures with tight hydrophobic packing, to completely disordered proteins, with extended backbone conformations. The term intrinsically disordered proteins (IDPs) has been applied to proteins that range from partially disordered to entirely disordered (Figure 1.7)[127]. IDPs may contain terminal regions lacking any stabilizing contacts, long loops or hinge regions separating folded domains, whole domains of disorder, or they may be entirely disordered. The Protein Data bank contains almost 70, 000 experimentally derived protein structures using nuclear magnetic resonance spectroscopy or x-ray crystallography. About 10 % of these structures contain regions of intrinsic disorder greater than thirty residues in length and about 40 % contain regions between ten to thirty resides of intrinsic disorder[128]. Low sequence complexity, along with regions of amino acid bias and predicted flexibility, are criteria for the classification of intrinsically disordered proteins[129]. Low amino acid variability and repetition of residues within a short region defines low sequence complexity, and can indicate intrinsic disorder[130]. A preference for charged over hydrophobic residues is also observed in IDPs[127]. Smaller residues that do not restrict backbone movement appear in IDPs at higher frequencies. The dependence of disorder on amino acid sequence complexity, charge, hydropathy, and secondary structure propensity allows for the prediction of disorder based on protein sequence analysis alone[129, 131]. These observations have led to the development of a number of algorithms to predict disordered segments in proteins[132-134].

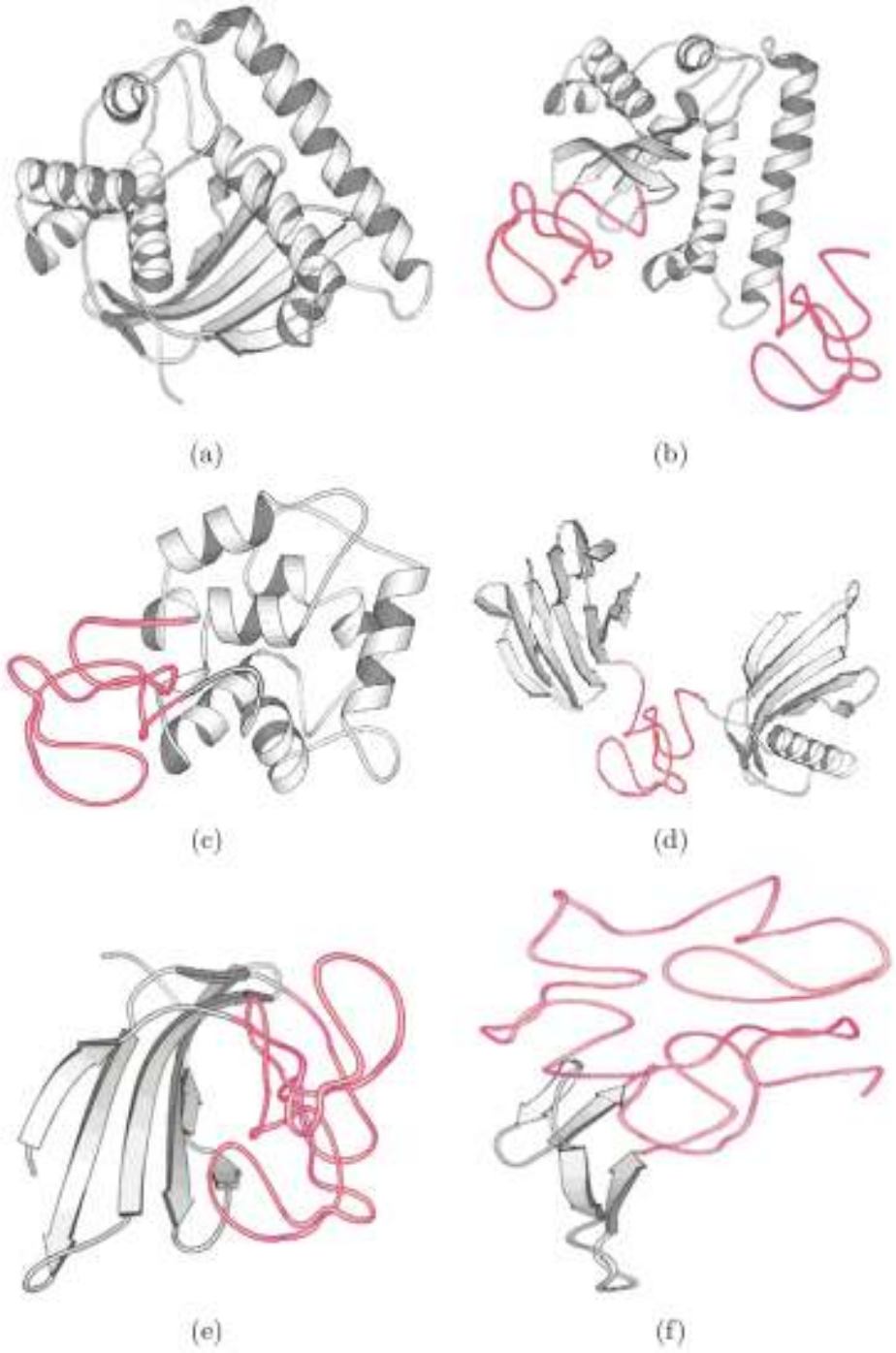

Figure 1.7 Various degrees of intrinsic protein disorder: (a) Folded protein with no disorder; (b) Terminal regions of disorder; (c) Disordered loops; (d) Disordered linker; (e) Disordered domain; (f) Mainly disordered protein with some defined structure.

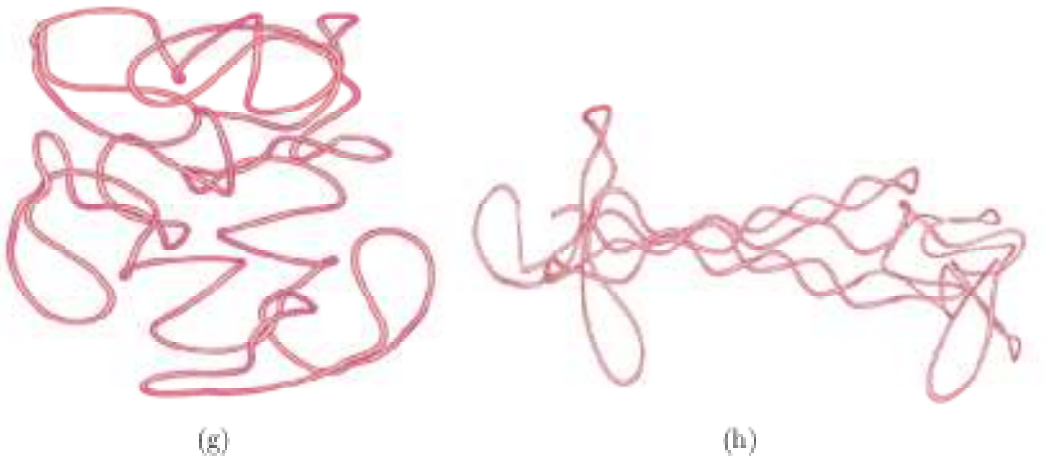

Figure 1.7 continued: (g) Completely disordered protein in collapsed state; (h) Completely disordered protein in an extended state. Reproduced with permission from V. N. Uversky.

A recent genomic analysis of the three kingdoms of life predicted that about 33 % of eukaryotic proteins contain stretches of thirty or more residues lacking well-defined tertiary structure, as compared to 2 % of archaeans and 4.2 % of eubacteria, and that disorder is often associated with functions of nucleic acid interaction and kinase activity[135]. The frequency of IDPs in higher life forms suggests a complexity in their function and it is evident that protein disorder and flexibility are important characteristics in the functional role of proteins. Sequence disorder as a functional role in proteins has been suggested to allow for multiple biological interactions with cellular environments containing an abundance of signaling proteins. A different proteomics study shows that 37 % of viral proteins contain intrinsic disorder, compared to 19% for archaea, 21.5 % for eukaryota and 30 % for bacteria[136]. The higher proportion of multi-ligand binding IDPs found in viruses could explain how a genetically simple virus is able to influence a complicated network of host cell growth factors for efficient viral replication.

Protein intrinsic disorder must be a persistent evolutionary advantage that allows for binding to numerous partners, or multiple substrates of similar structure.

1.3.1 The Roles of IDPs

Early classification of IDPs identified six categories: *effectors*, proteins that inhibit or activate cellular activities like transcription and translation; *chaperones*, auxiliary proteins that have protective roles when associated with other proteins or genetic material; *assemblers* of macromolecular structures; display sites for post-translational modification; *scavengers* of small molecules, ligands, or ions; and *entropic chains*, involved with dynamic linkers between functional domain and regulation of active sites or pores[137]. With the exception of entropic chains, further classification of the six categories of IDPs is based on whether ligand interactions cause transient conformational changes (display sites and chaperones) or permanent conformational changes (scavengers, effectors and assemblers)[138].

There have been several advantages proposed that explain why IDPs may be superior for certain roles based on their properties[130]. IDPs, or regions of intrinsic disorder, display larger surface areas compared to folded sequences of the same molecular weight. Extended IDPs form large interaction surfaces that may be able to bind multiple ligands, or capable of surrounding the binding partner[139]. Upon recognition of a ligand, IDPs often undergo an entropic loss by interaction-induced folding with a binding partner, referred to as a disorder-to-order transition[127, 140]. This entropic loss has been proposed to be part of an interaction mechanism that is characterized by a low affinity between partners, but a survey of folded and disordered proteins

shows that their binding affinities are within the same range[141]. The entropic loss in a disordered-to-ordered transition may be overcome by the release of a large number of water molecules from the binding site surface and the burial of hydrophobic surfaces[142].

It has been observed that disordered proteins are frequently involved in transcriptional and translational regulation[136]. Disordered proteins are able to quickly conform to numerous signaling factors. In some cases, a low affinity allows for rapid association and dissociation from binding partners supporting a model of the role of IDPs in a finely tuned cellular mechanism requiring very specific and rapid responses to stimuli, ie. rapid on and off rates that allow specific binding to numerous transcription and translation factors.

Disordered-to-ordered structural transitions are known to occur in protein-nucleic acid interactions. A study of the yeast transcriptional activator GCN4 showed a concentration dependent, α-helical transition from unordered monomers to coiled-coil dimers, and another transition to α-helical structure in the basic, unordered region upon addition of DNA containing the specific AP-1 binding site[143]. As a part of the same study, a similar folding transition occurred in the presence of nucleic acid with a similar binding site containing an additional central base pair, showing that in this case, the specificity of the DNA may not be important. This suggests that the free energy barrier to folding is partially overcome by complementary ionic interactions between the negatively charged phosphate groups of the DNA and the basic regions of the leucine-zipper. The specificity of interaction is achieved as folding occurs in response to specific DNA sequences, and it has been shown that a greater change in α-helical content occurs between basic regions and their specific DNA binding sites as opposed to non-

specific DNA binding sites[144]. Nuclear Magnetic Resonance (NMR) measurements of the ^{15}N longitudinal relaxation rates, ^{15}N transverse relaxation rates, and ^{1}H -^{15}N nuclear Overhauser effects of the GCN4 protein showed that the basic, DNA interacting regions were highly dynamic, occupying ensembles of transient α-helical character[145]. This dynamic behavior appears to be inherent to the disordered domain, and may be the mechanism allowing rapid and specific interaction between nucleic acid and protein. NMR studies have revealed the existence of specific "pre-ordered" regions of IDPs that are involved in ligand interaction[146, 147]. These regions interconvert between bound and unbound forms in the absence of ligand on μs – ms timescales[127, 148, 149]. The binding partner selects the ligand when it is in the "bound" conformation.

Structural ordering in response to zinc binding has been observed in the DNA binding domain of the human vitamin D receptor. α-helical content of the receptor increased in the presence of zinc, and NMR suggested a structural re-arrangement[150]. It is quite possible that the cysteine region of Tat becomes ordered upon zinc binding. An NMR study of the α-amylase inhibitor tendamistat fused to a Tat peptide showed α-helical tendency in the cysteine-rich region, and that the two proteins behaved as an ideal fusion protein; both of the proteins showed no mutual interaction[151]. This is supported by recent crystallographic evidence of a short $_{3\text{-}10}$ helix formation in the cysteine region of Tat[70]. Circular dichroism has shown minor structural changes in Tat in the presence of two molar equivalents of zinc, and some α-helical induction by interaction with the human co-activator and histone acetyltransferase CREB binding protein (CBP)[108].

IDPs and regions of intrinsic disorder may possess significant amounts of non-random structure[152]. This is referred to as "transient structure" or "structural propensity"[153]. As opposed to the robust changes in sequence that folded proteins can experience while still being reasonably well described by homology models[154], it is unclear how sensitive structural propensity is to mutations in the sequence of IDPs. A single point mutation in the IDP α-synuclein favours a folded structure by disrupting a disordered region found in the disease-causing state[155]. Changes in the chemical shifts between the fibrillar and non-fibrillar variants of α-synuclein suggest that the helical propensity of the fibrillar disease-causing variant is sequence specific[156]. β- and γ-synuclein do not possess the same fibrillar tendency as α-synuclein, most likely due to changes in sequence[157]. The helical propensity of α-synuclein differs from β-synuclein by containing an extra eleven residues in its amyloid forming region. Although the sequence identity and similarity of γ-synuclein are similar (66 % sequence identity and 74 % sequence similarity), the amyloidegenic properties of γ-synuclein are significantly less considerable; *in vitro* fibrillization occurs at much higher concentrations and at a much slower rate[157]. It was determined by NMR chemical shift measurements that γ-synuclein has a greater α-helical propensity in the amyloid forming region. The helical secondary structure of this region may be responsible for masking the β-sheet propensity of α-synuclein that is characteristic of amyloid fibrils[153].

1.3.2 IDPs and Drug Design

Many drugs target enzymes or receptors by mimicking the size, interactions, and contacts of natural substrates[158]. Rational drug design is based on the idea that 3-dimensional shape of

the enzyme active site or the receptor-binding pocket can be used to guide the design of substrate mimics or inhibitors. The nature of IDPs suggests that rational drug design is unfeasible because IDPs do not conform to stable, 3-dimensional structures, and that specific, targeted small molecule binding cannot occur. IDPs are abundant in major disease pathways that are associated with signaling, cancer, neurodegeneration, cardiovascular disease, and diabetes, and therefore are therapeutic targets[159]. In a study of yeast, it was found that about 80% of overexpressed genes were not harmful, however, overexpression of many proteins containing intrinsic disorder was harmful[160]. The authors propose that although IDPs exhibit high specificity, their toxicity is dose dependent. High concentrations of IDPs can lead to multiple ligand interactions that disrupt cellular pathways, where specificity of binding is overcome by the force of mass-action.

IDPs lack 3-dimensional structure, therefore drug development must rely on the primary amino acid sequence, but more specifically, regions within the sequence that undergo binding-induced transitions, referred to as molecular recognition features (MoRFs)[161]. MoRFs have been classified into three categories according to their structures in the bound state: α-MoRFs form α-helices, β-MoRFs form β-sheets, and ι-MoRFs form structures without any regular pattern of backbone hydrogen bonding[162]. An example of a MoRF that has been targeted by small molecule inhibitors that resulted in the abolition of activity is the basic-helix-loop-helix-leucine zipper (bHLHZip) in the ligand pair c-Myc oncoprotein (Myc) and myc-associated factor X (Max)[163]. Myc is a transcription factor that regulates genes associated with cellular growth, differentiation, metabolism, and apoptosis, and Myc is overexpressed in most human cancers[164]. Alone, Myc is a disordered monomer but it undergoes a coupled folding and binding transition in the presence of Max. Small molecule inhibitors targeting the Myc/Max dimer showed that the

ordered bHLHZip interaction region could be disrupted by small molecule inhibitor binding to the bHLHZip of Myc, causing a reversion of Myc to its disordered, monomeric form[163]. The dissociation constant was determined to be 2.8 ± 0.7 μM for the most effective inhibitor[163]. Three independent sites for small molecule inhibitors have been identified on Myc and it has been suggested that a higher drug affinity can be obtained by designing a multivalent inhibitor[161].

1.3.3 The Structure-Function Paradigm

It is clear that the classic structure-function paradigm emphasizing that protein function is rooted in a well defined, 3-dimensional structure does not accurately describe IDPs. It has been advocated to revise a model to include intrinsically disordered proteins as a functional class along with well-ordered, 3-dimensional protein structures[137]. The model for globular proteins describes an equilibrium between four states:

globular ↔ molten globule ↔ pre-molten globule ↔ completely unfolded

A globular protein exists in a state defined by a well folded structure, with a hydrophobic core consisting of tightly packed, aliphatic side chains. The molten-globule state is characterized by native-like secondary structure and a native-like folding pattern. In this state, the hydrodynamic radius has increased, and the hydrophobic core is accessible to hydrophobic fluoroprobes. The globular protein shows the existence of weak tertiary contacts that resemble a well defined

globular structure, but lacks the compactness, and can be said to be partially denatured. The pre-molten globule state is characterized by a secondary structure that resembles that of the native state, but shows an extended conformation in between that of the molten globule state, and a completely extended, unfolded state[165]. The pre-molten globular state interacts with hydrophobic, fluoroprobes with a lower affinity than a molten globule, signifying the existence of hydrophobic clusters in the pre-molten globule, but with considerably more solvent accessibility than hydrophobic clusters in a molten globule[166]. The pre-molten globule is a partially ordered form of a random coil. The final population of globular proteins is the completely unfolded form, consisting of a random coil, with the backbone completely exposed to solvent. To add functionally active disordered proteins (where disorder can exist on any order, from short spans of peptides, to wholly disordered proteins) would mean to change the structure/function paradigm to include intrinsically disordered proteins within these equilibria, not as denatured forms of proteins, but as biologically active agents. Functional proteins exist in each of these four states.

1.4 Structure and Dynamics Studies of Tat

A conformational and dynamic study of ^{15}N and $^{15}N/^{13}C$-labelled HIV-1 Tat_{1-72} using multinuclear NMR spectroscopy by Dr. Shaheen Shojania and Dr. Joe D. O'Neil determined that at pH 4.1 Tat_{1-72} exists in a random coil conformation based on the measurement of chemical shift, coupling constant measurements, and backbone NMR relaxation data[20]. Data for backbone assignments for the $^{13}C/^{15}N$-labelled histidine-tagged Tat_{1-72} protein were obtained from a 600-MHz INOVA spectrometer at 20 °C which allowed the assignment of 80 of the 84

non-proline resonances. 3-dimensional heteronuclear triple resonance experiments, that use one and two-bond scalar couplings to connect the backbone atoms, were used to determine the H^N, N, carbonyl carbon, C^α, and C^β chemical shifts, and the $^3J_HN_H^\alpha$ coupling constants[167]. Chemical shifts are sensitive indicators of conformation, and in comparison to random coil values, they show that the entire protein exists as a random coil with a slight tendency to α-helicity. The chemical shifts of the cysteine residues show that the protein is unambiguously reduced at pH 4.1. NMR relaxation parameters were analyzed by Lipari-Szabo and spectral density approaches, confirming that the protein lacks structure. The dynamics analysis did show that the most likely region to fold is the cysteine-rich region. ^{15}N-^1H NOE measurements for Tat were more negative compared to the average ^{15}N-^1H NOEs for a folded protein of similar size, indicating less restriction in dynamics on the ns – ps timescale[167]. Across the amino acid sequence, the ^{15}N-^1H NOEs showed a flattened bell-shaped pattern that is typical of a disordered polypeptide[168]. Circular dichroism studies of Tat showed that the CD signal has a strong negative ellipticity at 198 nm that is characteristic of disordered proteins[108, 169-172]. The addition of zinc to Tat in solution results in minor structural changes resulting in a CD spectrum that is still dominated by an intense negative band at 198 nm (Figure 1.8)[108, 172].

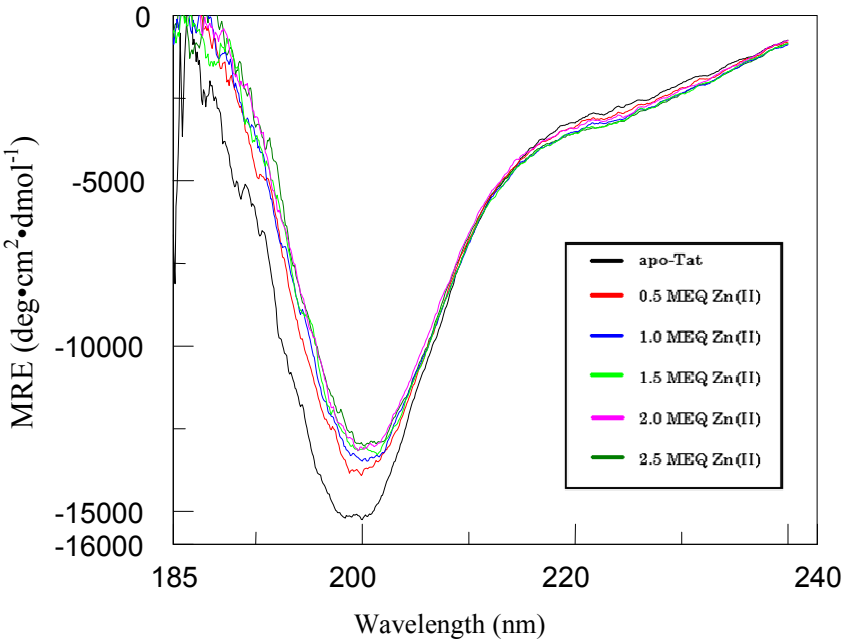

Figure 1.8 Circular Dichroism Spectra of a Zn(II) titration of 190 µM Tat at pH 5.0. Image reproduced with permission from T. Vo[172].

The aforementioned crystallography studies of Tat are described in this section with greater detail. An x-ray diffraction study of residues 13 – 69 of the equine infectious anemia virus (EIAV) Tat peptide fused C-terminally via a flexible glycine linker region to residues 8 – 263 of the cyclin box domain of equine CycT1 co-crystallized with and without a 22 nucleotide TAR stem-loop has been reported (Figure 1.9 (A))[67]. EIAV Tat only contains two cysteine residues, simplifying expression and purification. In the absence of TAR, no electron density could be observed for EIAV Tat in the complex, indicating that CycT1 could not induce a structure for EIAV Tat[173]. In the EIAV CycT1 – EIAV Tat – TAR complex, Tat covers an area of 3,164 Å^2, denoting the large interaction surface area of Tat upon assembly of the protein-RNA complex[67]. The region of Tat whose electron density could be assigned corresponds to residues

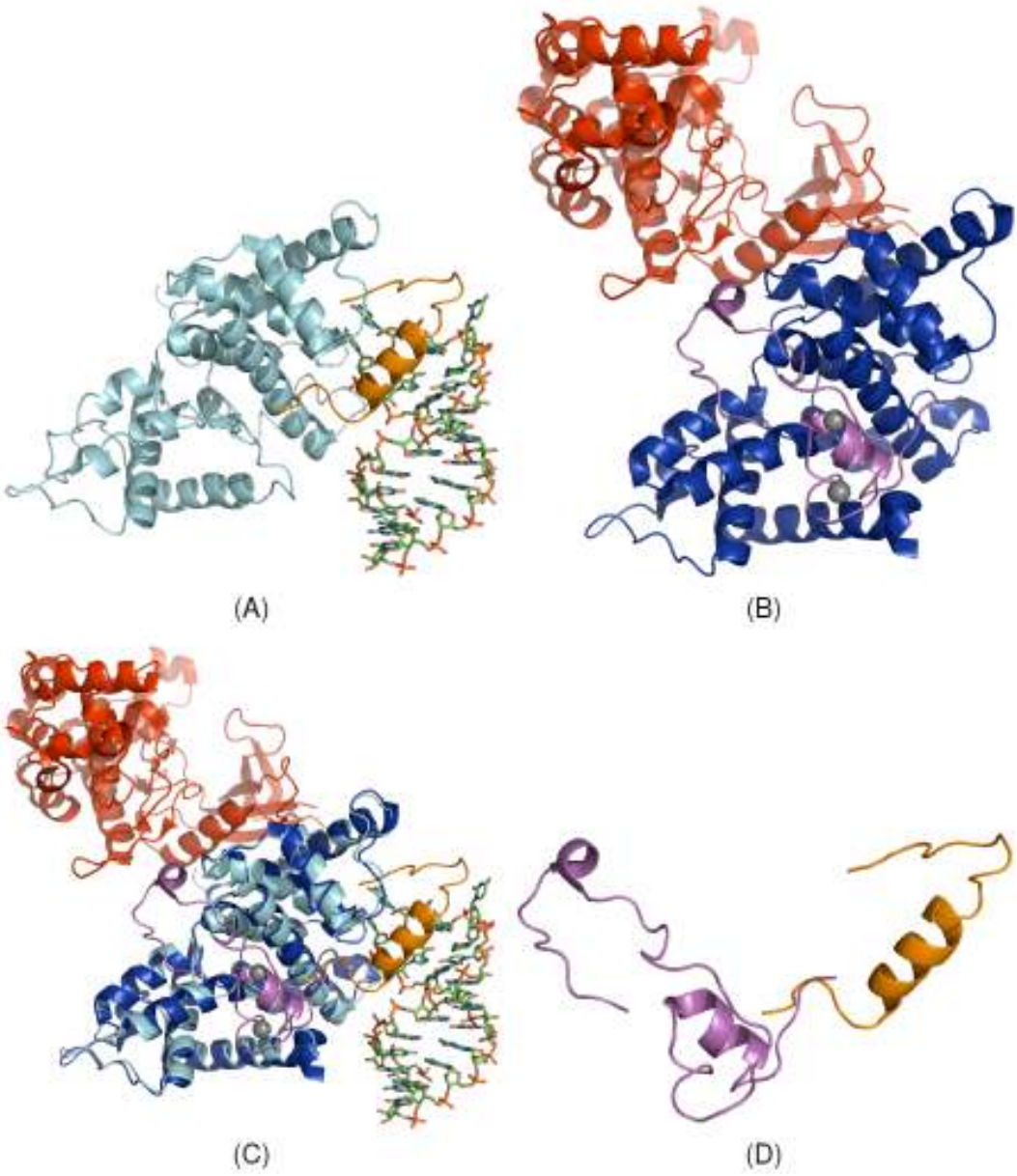

Figure 1.9 X-ray diffraction structures. (A) The EIAV CycT1-Tat-TAR complex showing CycT1 (teal), residue 41-69 Tat (yellow), and a stick representation of the major groove of TAR. (B) The human P-TEFb-Tat complex showing CycT1 (blue), Cdk9 (red), Tat (purple), and zinc atoms represented by grey spheres. (C) An alignment of the two structures shown in (A) and (B). (D) An alignment of residue 1 - 49 HIV-1 Tat with residue 41 – 69 EIAV Tat. Reproduced and adapted with permission from Anand[67] and Tahirov[70].

41 – 69. Residues 41 – 47 of the core region are in an extended conformation that is parallel to helix1 of the first cyclin box repeat of CycT1. Residues 48 – 59 (ARM) adopts a helix that fits into the major groove of TAR. The affinity of the ARM of Tat for TAR can be abolished by acetylation of Lys50[174]. The following residues Glu61 and Gly62 form a tight turn so that the C-terminus folds back around the cyclin box repeat. The C-terminus of Tat is stabilized by the insertion of hydrophobic residues Leu68 and Leu69 into a hydrophobic groove formed by helix4, helix5, and the interconnecting loop[67]. This X-ray diffraction study shows how the specific regions flanking the ARM interact with CycT1 in order to present the helix adopting ARM for interaction with TAR.

The other X-ray diffraction study involved the 86 residue human HIV-1 Tat protein crystallized with P-TEFb to elucidate structural characteristics of the Tat/P-TEFb complex that could not be elucidated in the absence of Cdk9 (Figure 1.9 (B))[70]. The electron density of residues 1 – 49 of Tat are defined, and shows that Tat adopts mainly an extended conformation that has an interaction area of 3,499 $Å^2$. 88 % of this interaction is with CycT1 and 12 % is with Cdk9. The importance of the presence of Cdk9 for the interaction of Tat with P-TEFb is evident in the lack of electron density in the N-terminal residues of the EIAV CycT1/Tat/TAR study[67], and is made clear by noting that Tat is inserted in a groove at the CycT1/Cdk9 interface[70]. The first segment of the N-terminal proline-rich region is in an extended, U-shaped conformation that interacts with a wide depression between the cyclin box repeats of CycT1. The second segment of the proline rich region adopts a β-turn and interacts with the T-loop of Cdk9. It is this region that is at the interface between CycT1 and Cdk9. The cysteine-rich and core regions form a more compact structure with the first zinc finger coordinated by Cys22, His33, Cys34, and

Cys37, separated by residues 29-33 forming a short 3_{10}-helix (zinc ions are denoted by grey spheres in Figure 1.9, B and C). The second zinc finger is coordinated by Cys25, Cys27, Cys30 of Tat and Cys261 of CycT1 (Figure 1.9, B and C). A short α-helix follows the zinc fingers, spanning residues 35 – 43[70]. Interestingly, Tat binding results in the disordering of a CycT1 α-helix spanning residues 253 – 256 and a 4.6 Å displacement of the C-terminal portion of the preceeding α-helix H5'. This disordering is proposed as the mechanism for displacing Hexim1 and relieving its inhibition.

No aqueous solution state studies to date have shown the ARM region of Tat to adopt a helix in the presence of TAR. A study of a 14 amino acid Tat peptide corresponding to the basic region showed that it forms a β-turn that sits in the major groove of the helical TAR hairpin of bovine immunodeficiency virus (BIV)[175] in contrast to the aforementioned crystallographic study of the fusion of CycT1-Tat protein and TAR of EIAV showing the basic region assuming a helical conformation[67].

A previous study of Tat has employed TFE as a co-solvent[176]. Residue assignments for the unlabelled, 75 amino acid EIAV Tat protein were determined by two-dimensional double quantum-filtered correlation spectroscopy, total correlation spectroscopy (TOCSY) with suppression of nuclear overhauser effect type cross-peaks, nuclear overhauser effect spectroscopy (NOESY), ^{15}N resolved three-dimensional NOESY heteronuclear multiple quantum coherence (HMQC) spectroscopy and TOCSY-HMQC spectroscopy. A tight, type II β-turn was observed by the hydrogen bonding between Ser44-Ile47, separating the helical, hydrophobic core residues Gln38-Arg43 and helical, arginine-rich region residues Asp48-Ala64 (Figure 1.10)[176].

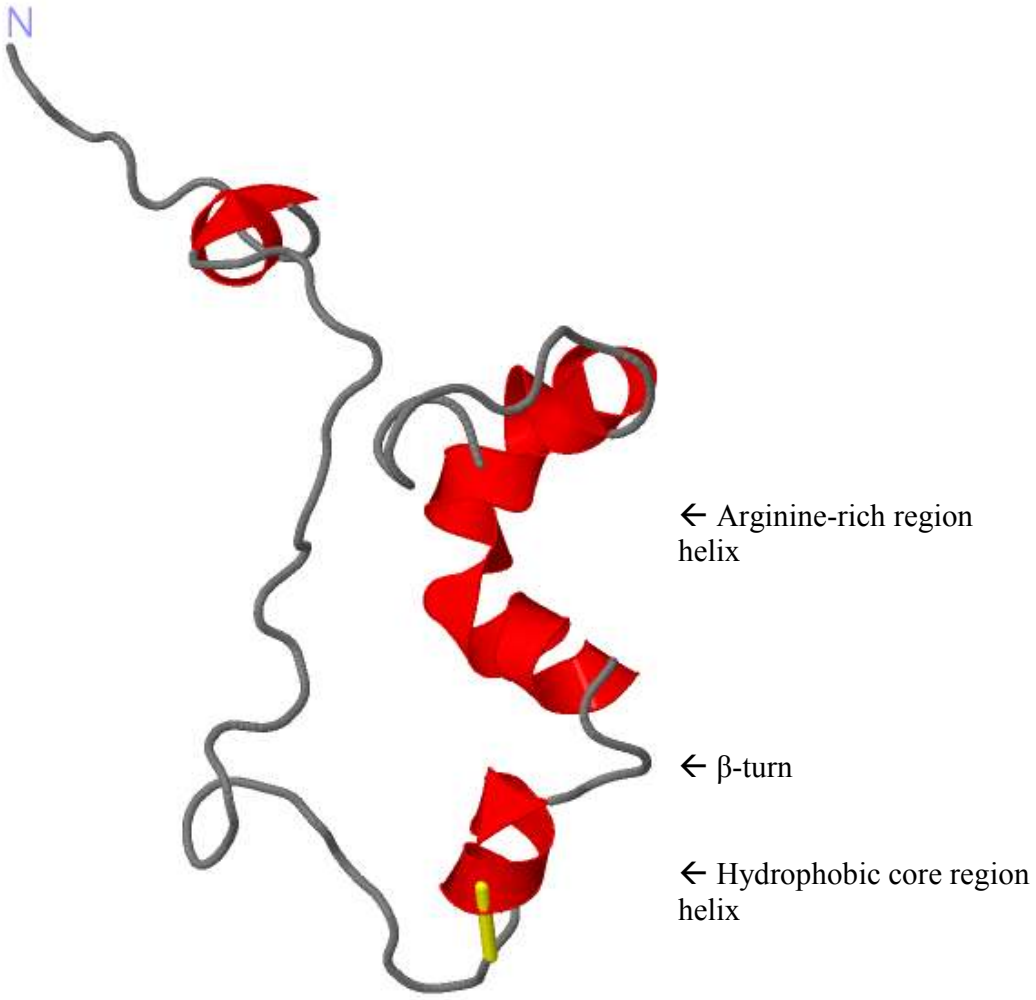

Figure 1.10 The helix-turn-helix structure of residues 38 – 64 of the 75 amino acid EIAV Tat protein. The image was reproduced with permission from H. Sticht[176].

1.5 Intrinsically Disordered Precursors to Amyloid Disease

Amyloid diseases are characterized by the formation of extracellular aggregates called amyloid plaques, composed of amyloid fibril[177]. Amyloid precursors fall into two categories, ones that are disordered in their monomeric state and others that assume a partially unfolded

structure before associating with fibril growth. Examples include Aβ peptide of Alzheimer's disease, atrial natriuretic factor, calcitonin, pro-calcitonin, islet amyloid polypeptide (IAPP), amylin, α-synuclein and the medin polypeptide[178]. A model linking early oligomerization to the adoption of helical structure and helix-helix interaction as a promoter of amyloid growth suggests that key intermediates are responsible for initiating formation of aggregates[179].

β-amyloid 1-42 is the main precursor to amyloid plaques in Alzheimer's Disease[180]. The self-association of this peptide is partly attributed to its hydrophobic nature, enabling it to form extended, ordered oligomers. Increasing evidence about the formation of amyloid plaques supports an equilibrium between intrinsically disordered amyloid precursors with helical conformers (Figure 1.11)[179]. A kinetic study of the growth of amyloid fibril showed a three-step sigmoid profile that is characterized by a lag phase, an exponential growth phase, and a plateau phase[180]. The lag phase indicates a period of time when amyloid growth does not occur, and is correlated with populations of amyloid precursors in a state characterized by an equilibrium between disordered and helical monomers (Figure 1.11(A)). It is usually suggested that small aggregates or seeds form during the lag phase upon which fibril growth can occur rapidly. The conversion of disordered/helical monomers to multimers rich in β-sheet was monitored by circular dichroism, and the rate of appearance of β-sheet secondary structure is exponential, and is referred to as the exponential growth phase (Figure 1.11 (B))[180]. The decrease in the rate of β-sheet structure formation along with a loss of soluble protein signifies insoluble, amyloid growth, and appears as a plateau region after the exponential growth phase, and is referred to as the platuea phase (Figure 1.11 (C)). Factors such as protein hydration, protein dynamics, and the hydrophobic effect are important in the kinetics of fibril growth, but

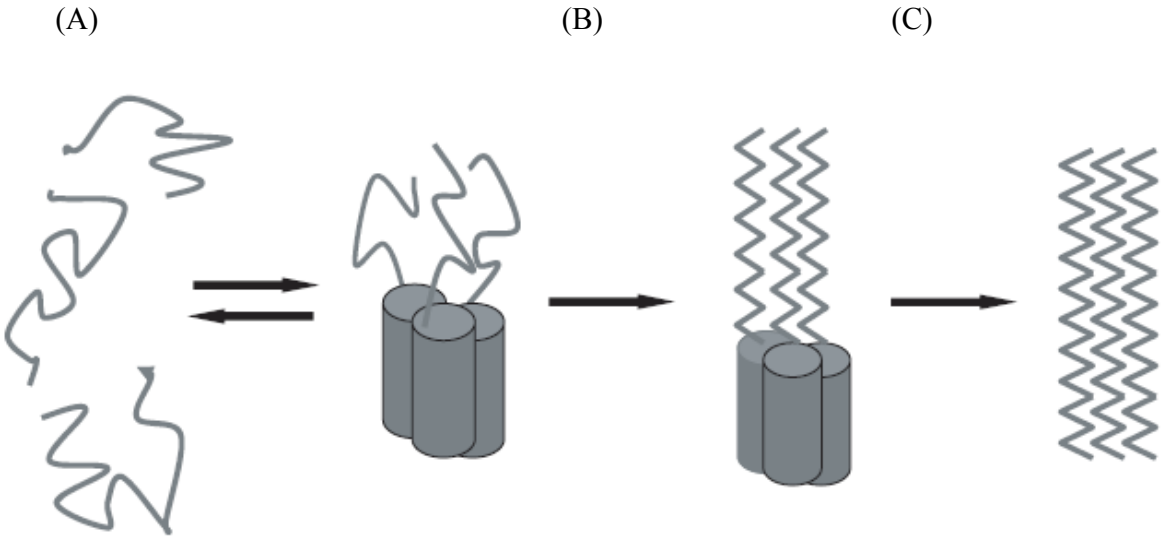

Figure 1.11 A proposed model for the growth of amyloid fibrils[178, 181, 182]. Amyloid growth is initiated by the conversion of intrinsically disordered amyloid precursors and their α helical conformers (A) to β-sheet populated structures (B). This is a slow process. The exponential growth phase in the accumulation of amyloid fibril is represented by the conversion of unordered/helical intermediates to a multimeric, β-sheet populated plaque (C), and the plateau phase consists of extended, insoluble fibril networks.

are also important in describing the populations of precursor monomers. The different phases of amyloid progression can be prolonged by specific inhibitors targeting conformers unique to each phase. The soluble, pre-amyloid disordered/helix populations unique to the lag phase can be targeted by small molecule inhibitors; the end result is an increase in the length of the lag phase, and in some examples, a prevention in the transition to the exponential conversion to β-sheet [183]. Soluble, amyloid precursors are also affected by solvents that alter their hydration, delaying the onset of fibril growth along with a decrease in the rate of fibril formation. One of these that has

been characterized is the polar, aprotic solvent acetonitrile[180]. Certain molecules like Congo Red and tetracycline can specifically decrease the rate of fibril growth, and are believed to be β-sheet specific ligands. It is presumed that the unique surfaces formed by the interaction of β-pleated sheets create cavities where amyloid-binding molecules like Congo Red and Thioflavine T are intercalated[183].

The low dielectric environment found near lipids has been shown to promote α-helical content in disordered amyloid precursors like IAPP, and Electron Paramagnetic Resonance data appear to be consistent with the formation of helices extended over the surface of lipid membranes, and not as transmembrane domains[181, 182]. This observation suggests that intracellular and extracellular lipids may act as sites of concentration of amyloidogenic proteins.

1.5.1 Predicting Amyloid Precursors by Sequence Specific Determinants

Protein aggregation by assembly of short segments into β-sheet structures is the most common mechanism by which proteins aggregate[184]. Although an enrichment in β-sheet secondary structure is commonly observed during aggregation, protein aggregates form various tertiary structures, the most common being amorphous aggregates that do not possess a defined three-dimensional structure[185]. There are many proteins that can be coerced to form ordered fibrillar aggregates high in β-sheet content, either under physiological or non-physiological conditions, however, amyloid formation is also a sequence specific process[186]. Amyloid fibrils are highly ordered aggregates that are formed due to hydrophobic interactions, but their rate of

formation and stabilization are influenced by hydrogen bonding and ionic forces[187]. Amyloidegenesis is side chain dependent, and thus sequence specific[188]. The sequence diversity of more than 200 amyloid hexa-peptides has been explored using transmission electron microscopy, x-ray diffraction, circular dichroism, and fourier transform infrared spectroscopy to better understand the sequence specific determinants of amyloid structure. The results were used to develop the web-based algorithm Waltz that predicts amyloid sequences and identifies amyloid forming regions of proteins[189].

A six residue amyloid nucleating insertion into a protein domain is sufficient to induce amyloid formation, although nucleating cores in full-length proteins are often longer than six residues, and the flanking residues can influence amyloid propensity[190-192]. Most of the experimentally verified amyloid sequences are hexa-peptides and are available in the AmylHex database[193]. The use of short hexa-peptides in developing an amyloid predicting algorithm distinguishes whether or not the extra residues found in many longer amyloid sequences actually participate in the β-sheet. Amino acid identity can be tolerated to various degrees depending on the position within the minimal hexa-peptide amyloid forming unit. Positions 1 and 6 of an amyloid forming hexapeptide sequence can accommodate almost any amino acid; position 5 favours isoleucine, phenylalanine, and tyrosine; positions 2,4, and 5 favour isoleucine; positions 3 and 6 favour valine. Generally, hydrophobic residues are favoured in the middle of the hexapeptide and polar residues are favourable at the ends, but polar residues can sometimes be accommodated in the middle. The presence of the polar residues asparagine and glutamine is actually favourable in the middle and they are present in yeast prion amyloid[191].

1.5.2 Congo Red as an Indicator of Amyloid

Ex vivo amyloid deposits and amyloid samples prepared *in vitro* have been detected by the green-yellow birefringence of Congo Red binding[178]. Congo Red staining has been used for decades in identifying amyloid in tissue, and when viewed under a cross-polarizing microscope, CR displays a green-yellow birefringence. When viewed by unpolarized light, CR-stained amyloid tissue appears red. CR-stained amyloid fibrils can also be detected by the hyperchromic shift in ultraviolet (UV) absorbance of CR from 500 nm (unbound CR) to 540 nm (CR/amyloid fibril) (Figure 1.12) [194]. In the absence of fibril the CR solution is an orange-red colour whereas when bound to amyloid a red-rose colour is observed.

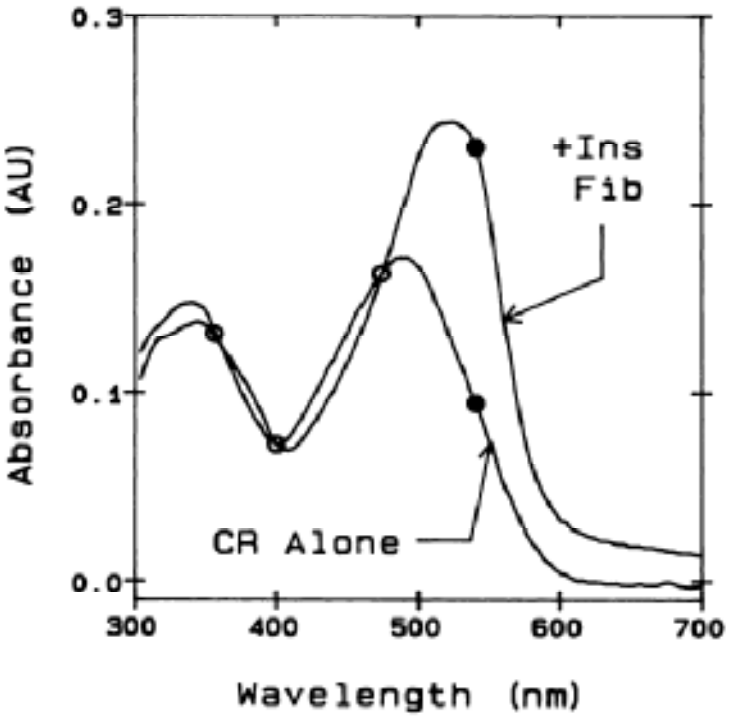

Figure 1.12 The absorbance spectra of 5 μM Congo Red in the presence (+ Ins Fib) and absence (CR Alone) of 11 μM insulin fibrils. This image is reproduced with permission from W. E. Klunk[194].

1.6 The Hofmeister Series and Hydrophobic Effect

In 1888, Franz Hofmeister defined an ion series based on their effect on protein solubility[195]. His early studies observed the effects of ions on the solubility of proteins, and much later it was realized that anions and cations also affect the stability of protein secondary and tertiary structure. On one end of the series, ions increase solvent surface tension[196], decrease the solubility of nonpolar molecules and strengthen the interactions of hydrophobic groups. This is classically referred to as "salting out" and the use of these ions has been used to precipitate folded proteins during protein purification and crystallization, e.g. $(NH_4)_2SO_4$ and Na_2SO_4. The other end of this series describes ions that decrease surface solvent tension and increase the solubility of nonpolar molecules, a term referred to as "salting in".

Hofmeister's ranking for the "salting out" of proteins by anions is:

$$SO_4^{2-} > HPO_4^{2-} > \text{acetate} > \text{citrate} > \text{tartrate} > Cl^- > NO_3^- > ClO_3^- > I^- > ClO_4^- > SCN^-$$

and a ranking for the effectiveness of "salting out" by cations is:

$$NH_4^+ > K^+ > Na^+ > Li^+ > Mg^{2+} > Ca^{2+}$$

The Hofmeister series rationalizes the effects of guanidinium salts as protein denaturants. Guanidinium thiocyanate destabilizes protein structure more than the guanidinium chloride, whereas guanidinium sulfate stabilizes protein structure[197]. An alternative ranking of the effectiveness of ions in "salting out" hydrophobic molecules was determined by McDevit and Long using benzene as their model compound[198]. Their ranking for anions is:

$$SO_4^{2-} > OH^-, F^- > Cl^- > Br^- > NO_3^- > ClO_4^- > I^-$$

and their ranking for the effectiveness of "salting out" by cations is:

$$Ba^{2+} > Na^{+} > K^{+} > Li^{+} > Rb^{+} > NH_4^{+} > Cs^{+} > H^{+}$$

Generally, divalent ions are more effective at salting out than monovalent ions, and ions with small, hydrated radii are more effective than large ions[199].

"Salting in" is a phenomenon in which the presence of a salt favours the solubility of a protein, also known as salting in the peptide group[200]. It is believed that the salting in effect occurs between simple ions and polar molecules, and is dependent on ionic strength. These ions interact much more strongly with unfolded proteins because of the greater solvent accessibility to the peptide backbone, i.e. guanidine thiocyanate is a powerful protein denaturant because of the strong interaction between thiocyanate and the peptide unit[201]. The exact manner in which the Hofmeister series affects protein structure is unclear and several theories have been presented to explain the Hofmeister series effects on proteins. A salting in mechanism has been proposed in which the salt initially interacts with side-chains of the protein, leading to backbone interactions. In a molecular dynamics study of the guanidinium cation (Gdm^{+}), it was observed that at non-denaturing concentrations, the cation associates with planar side chains of Arg, Trp, and Gln in a stacking manner, but also interacts with hydrophobic aliphatic side chains[202]. A similar study correlated the free energy of transfer of amino acid side chains from water to 6 M Gdn-HCl and observed a correlation between the solubility of side chains with their surface areas (regardless of side chain polarity)[203]. It has been argued that "salting in" interactions occur between salts and the peptide backbone by non-specific ion-dipole interactions. In this theory what differentiates an ion's ability to salt in or salt out are opposing interactions with the peptide backbone that compete with solvent interactions that decrease solvent surface tension[200].

The fact that ions at different ends of the Hofmeister series increase or decrease water surface tension led to the idea that salting in and salting out of proteins is mainly caused by the effects of the ions on water structure. Ionic species that are strong "salting out" agents are referred to as kosmotropes, and were believed to be "water structure makers". These ions are strongly hydrated. Chaotropes, or ions that promote "salting in" were believed to be "water structure breakers", and they destabilize folded proteins[199, 204]. However, there have been several experiments that have contradicted the notion that Hofmeister salts affect the bulk properties of water. Water molecule dynamics were characterized in the presence of salt using femtosecond mid-infrared pump-probe spectroscopy[205]. The results showed that the water structure outside the hydration shell of an ion is not influenced by the ion, regardless of whether the ion is a kosmotrope or a chaotrope. The correlation times of the ionic solvation shell were much slower than that of bulk water, regardless of the identity of the ion, but the hydrogen bonding network outside the vicinity of the ion was not affected.

In another study that challenges the understanding that "salting out" ions increase the solvent surface tension, and that this increase in free energy can be countered by an increase in hydrophobic interactions often resulting in protein precipitation, the phase transition of a detergent monolayer and the adjacent water structure were simultaneously observed using vibrational sum frequency spectroscopy (VSFS)[206]. VSFS is sensitive to alkyl chain conformation and interfacial water structure, and from this data, it was ascertained that the ability of anions to disrupt the hydrophobic interactions of the detergent monolayer follows the Hofmeister series. Chaotropic anions were found to be able to penetrate the head groups better

then kosmotropic anions, and by direct interaction, were able to disrupt the ordering of the hydrophobic tail groups.

1.7 Trifluoroethanol Studies

2,2,2-trifluoroethanol (TFE) is often used as a co-solvent to promote the formation of secondary structure in peptides and as a denaturant in protein folding studies. A low concentration of TFE (10 % TFE) stabilized the tertiary structure of folded hen egg white lysozyme as determined by an increase in NMR nuclear overhauser effect contacts , implicating TFE as a stabilizing solvent in some cases[207]. At higher concentrations of TFE, tertiary structures denature, whereas localized secondary structure remains intact[207]. Native-like secondary structure can be induced in peptides with alcohol-based co-solvents like ethanol, or TFE. NMR and far UV CD data have demonstrated that TFE promotes α-helical and β-hairpin formation in peptides, and quite often, the secondary structural elements are reflective of the structure in the native protein. TFE solvation has found applications in the study of partially or entirely disordered proteins. Studies have shown that TFE interacts weakly with non-polar side chains, allowing it to penetrate into hydrophobic cores of proteins[207, 208].

Protein folding is a highly complex process involving numerous non-covalent interactions between the peptide backbone, interactions of the side-chains, covalent cysteine cross-linking, the characteristics of the solvent, the concentration of salts, temperature, the presence of molecular chaperones, or the presence of ligands. The structures of proteins have been shown to be dependent on their solvation states, and the presence of molecules that fold or

unfold proteins. By investigating the properties of solvents and their effects on protein structure, insights into the solvent dependent folding mechanisms can be deduced. The effects of TFE on peptides suggest that solvation states and folding processes are connected.

The effects of alcohol-based co-solvents on protein and peptide structure have been studied for more than 50 years. Early studies of denatured Ribonuclease in 2-chloroethanol and 2-chloroethanol-water mixtures showed localized, folding in the α-helical regions of the protein[209]. Early models suggested that less polar co-solvents could stabilize secondary structure, but partially denature proteins by their ability to penetrate and interact with the hydrophobic cores of proteins. Studies have tried to correlate helical propensity of peptides as determined by structure prediction algorithms to their response in TFE, and have found conflicting results including the appearance of unexpected β-sheet and β-helical structures[210]. It has been claimed that amide-solvent interactions in dilute TFE are disrupted, maximizing amide-amide hydrogen bonding. Similar studies have observed an increase in cis-trans proline isomerization in polypeptides co-solubilised in TFE[211]. It also has been argued that TFE-induced structure is a result of the low dielectric constant of the solvent relative to water and the hydrogen bonding character (strong H-bond donor, poor proton acceptor) of the alcohol moiety. In this model the replacement of the protein's hydration sphere with TFE is entropically driven, as the bifurcated hydrogen bond interaction of the peptide carbonyl group with two water molecules is replaced by one TFE molecule[212]. Although the two waters have been replaced by one TFE molecule, the peptide carbonyl group is still free to form two hydrogen bonds. One hydrogen bond forms with TFE while the other hydrogen bond can form with an amide proton.

The result is a net decrease in the protein's hydration sphere and the promotion of inter-residue interactions by hydrogen bonding[212].

Water behaves as a proton donor and acceptor, and because of its high polarity (dielectric constant ε = 78.54 at 25 °C) [213], it might destabilize conformationally unstable proteins by interacting with ionic amino acid side chains that may otherwise be involved in intra-residue ionic bonding. The promotion of secondary and tertiary structure has been attributed to the lower dielectric constant of TFE (ε = 26.69 at 25 °C) [213]. As the solution polarity is lowered, the equilibrium between peptide-solvent and intra-peptide interactions is shifted. Intra-residue ionic interactions and hydrogen bonding between amide hydrogens and carbonyl groups of the backbone become favoured, although this correlation in TFE is not complete[214]. Although TFE has been implicated to favour peptide backbone hydrogen bonding, direct interaction of TFE with aliphatic side chains has been shown by proton and fluorine NMR studies. TFE penetrates the outer hydration layer and associates with hydrophobic regions of proteins[215].

Chemokine-derived antimicrobial peptides are disordered in their truncated forms, and do not reflect the secondary structures of their native states when they are a part of the full length protein[216]. In the presence of lipids, or TFE, these disordered peptides are induced to adopt their native helical character. TFE studies may enhance our understanding of lipid-associating proteins, and it has often been called a "membrane-mimetic" solvent. Attempts to study transmembrane or membrane associated proteins in TFE have come to mixed conclusions, and the applicability of organic solvents in this field of study is questionable. Transmembrane

proteins rely on complex tertiary interactions that have been shown to be disrupted in the presence of TFE.

A study displaying the effectiveness of TFE to support native protein folding was reported[217]. Proper folding of a C-terminal helix of thioredoxin (TRX) was shown to be dependent on proper hydrophobic packing with surfaces of the remainder of the protein. In its native state, the C-terminal residues 94-108 adopt an α-helix as part of the 108 residue TRX oxidoreductase enzyme. As a peptide, TRX94-108 is disordered. TRX1-93 assumes a partially folded state, but when combined with TRX94-108, the near UV circular dichroism of tryptophan and tyrosine and the fluorescence emission spectra match those of the wild type protein. This, along with the re-establishment of wild-type enzymatic activity when TRX1-93 and TRX94-108 were combined, verified that the enzyme properly re-folded. By alanine mutagenesis of various residues of TRX94-108, it was determined that proper folding and enzymatic activity was dependent on the hydrophobic interactions of specific aliphatic residues with the larger TRX1-93 fragment. Co-solubilization of TRX94-108 in a TFE/aqueous mixture re-established the peptide's native-like helical secondary structure. Further screening of TRX94-108 mutants alone in TFE correlated the lack of helicity with the same mutated residues that abolished proper folding and activity in aqueous solution when in the presence of TRX1-93.

TFE has been used to test the IDP/ligand interaction "fly-casting" model. The "fly-casting" model for IDP folding transitions proposes that the disordered protein makes an initial contact with its ligand that does not initially result in binding-induced order. The ligand provides complementary interactions that guide the protein folding[218]. An alternative IDP folding model

referred to as "conformational selection" postulates that IDPs exist as rapidly inter-converting conformers. The solution state equilibrium populations consist of the extended conformer, a small, transiently populated folded, active conformer, and any subpopulation of conformers in between. The folded, active subpopulation is able to rapidly bind to its ligand[219]. The kinetics of folding and binding of the endogenous inhibitor IA_3 to yeast aspartic proteinase A (YPrA) of *Saccharomyces cerevisiae* was studied to determine if the inhibitor followed a "fly-casting" model or participated in "conformational selection"[220]. IA_3 contains an N-terminal disordered region that adopts an α-helix by binding above the active site cleft and inhibits activity by blocking substrate entry[221]. Considerably faster folding rates were observed for enzyme induced inhibitor folding, compared to the TFE induced folding rate, highlighting the efficient coupling of recognition and folding in this case, thus supporting the fly-casting model.

1.7.1 Random Coil-to-Helix Two-State Transitions

Protein folding, at least for small, single domain proteins, is known to involve a two-state, cooperative process. At equilibrium, the protein exists in only the native folded state or the unfolded state; no intermediate states are thermodynamically stable. The transition between states can be induced with pH, denaturant, salt, or temperature. A two-state transition implies the existence of two energy minima, separated by an energy barrier[222]. The folding of intrinsically disordered peptides to their helical conformers in the presence of TFE has been modeled by two-state transition[220]. Helical propensities of 10 to 14 residue model polypeptides in TFE have been quantified from the understanding of two-state transition models, suggesting

that TFE influences a pre-existing random-coil/helix equilibrium[223]. The application of two-state transition theory to larger peptides or proteins in TFE may be questionable due to the possibility that the protein may adopt intermediate conformers, in other words, display non-cooperative folding.

1.8 Light Scattering as an Indicator of Protein Aggregation

Elastic light scattering in the near UV range has been used to measure the multimerization of proteins. In an early study of blood clotting, Ferry and Morrison characterized the activity of thrombin by observing the enzymatic polymerization of monomeric fibrinogen to cross-linked, insoluble, and multimeric fibrin[224]. The near-ultraviolet absorption spectrum of soluble, solution state fibrinogen is similar to the absorption spectrum of most soluble proteins containing aromatic residues. Upon an elapsed exposure to thrombin, fibrinogen is converted to the multimeric, insoluble "aggregate" form termed fibrin, and the conversion is quantified by an increase in light scattering above 300 nm, where the protein does not absorb light.

In a study of the regulation of Heat shock protein 27 (Hsp27) by phosphorylation, the thermal aggregation of citrate synthase was monitored by ultraviolet (UV) spectroscopy[225]. The phosphorylated, inactive form of Hsp27 was unable to prevent, or retard thermal aggregation of citrate synthase at various incubation times at 43 °C. The temperature induced aggregation was measured by an increase in light scattering at 400 nm.

When the dimensions of a protein are considerably smaller than the wavelength of incident light, changes in particle size have no effect on the absorbance of the particle. However, when the particle is large compared to the wavelength of the incident light, the radiation scattered from different points on the particle varies inversely as the result that the turbidity of a solution of scattering particles varies inversely as the square of their radii and their volumes[226]. Thus, in this regime increased particle size results in increased turbidity.

1.9 Circular Dichroism

Measurements of protein backbone circular dichroism (CD) allow for the classification of global protein secondary structure in the solution state. Circular dichroism is the phenomenon of the differential absorption of left circularly polarized light (LCPL) and right circularly polarized light (RCPL), resulting in an elliptical polarization of light and is defined as:

$$\tan(\theta) = \frac{E_L - E_R}{E_L + E_R}$$

where E_L and E_R describe electric field vectors of LCPL and RCPL. Ellipticity is defined as the arctangent of the ratio of the difference ($E_L - E_R$) and sum ($E_L + E_R$) of the two electric field vectors and is reported in degrees of ellipticity (θ). $E_L - E_R$ defines the minor axis of an ellipse and $E_L + E_R$ defines the major axis of an ellipse (Figure 1.13).

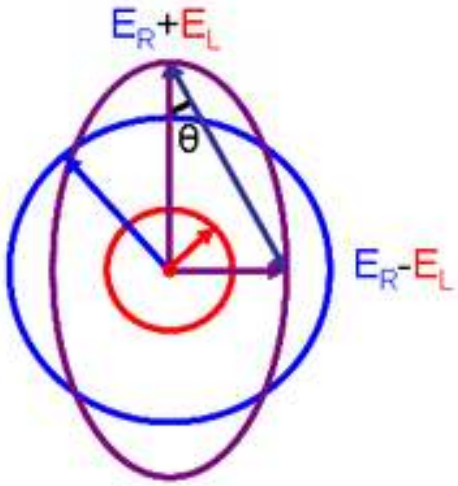

Figure 1.13 The combination of circularly polarized electric field vectors (E_R and E_L) describing an elliptically polarized electric field vector (purple). E_R (blue) represents right circularly polarized light, and E_L represents left circularly polarized light. The addition of E_R and E_L defines the major axis of an ellipse; the subtraction of E_L from E_R defines the minor axis of the ellipse. This image image reproduced with permission from Y. H. Eric[227].

In the far UV region (180 nm – 250 nm), the peptide backbone is chromophoric. The absorption of light by backbone amide electrons is accompanied by the promotion of a lone pair electron to the pi bonding or anti-bonding orbitals ($n\rightarrow\pi$, or $n\rightarrow\pi^*$ transition). CD of the peptide amides is dependent on the asymmetry of the peptide backbone environment, allowing ensemble secondary structural characterization of soluble protein. As illustrated in Figure 1.14, secondary structural elements are easily distinguished by their characteristic CD spectra: helical proteins have large negative bands at 222 nm and 208 nm and a large positive band at 193 nm. Proteins high in β-sheet conformation have a negative band at 210-220 nm and a positive band at 195 nm. Disordered proteins show a negative band below 210 nm with little ellipticity above 210 nm[228].

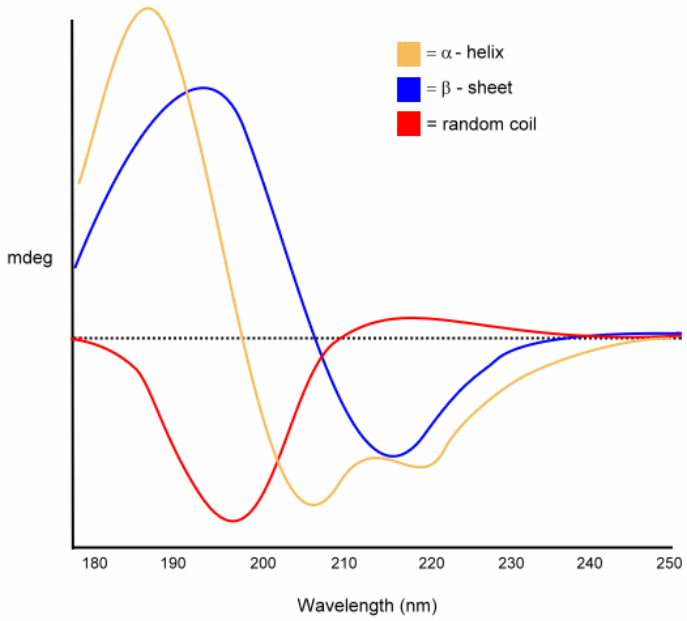

Figure 1.14 The characteristic CD spectra of α-helical peptide backbone conformation (yellow), β-sheet peptide backbone conformation (blue), and random coil peptide backbone conformation (red). This image reproduced with permission from K. A. Thomas[229].

Disordered protein CD spectra show much similarity to collagen proteins. Collagens are unique proteins consisting of three polypeptide chains forming a triple helix. Each strand of collagen resembles a poly-L-proline helix, in which the peptide bonds are in a *trans* arrangement. CD is a useful tool for characterizing the secondary structure of unknown proteins. CD is also a quick and effective method for monitoring secondary structural changes due to heat, chemical denaturants, and ligand interactions.

There are many different methods to analyze CD spectra to estimate secondary structure. All methods represent the CD spectrum of a protein as a linear combination of the spectra of its

secondary structural elements, inlcuding the contribution from such elements as aromatic chromophores or prosthetic groups[228]:

$$\Theta_\lambda = \Theta_{\lambda h} + \Theta_{\lambda s} + \Theta_{\lambda l} + \Theta_{\lambda u} + \Psi$$

where Θ_λ is the ellipticity of a protein as a function of wavelength, $\Theta_{\lambda h}$ is the ellipticity contribution from helical secondary structure, $\Theta_{\lambda s}$ is the contribution from β-sheet secondary structure, $\Theta_{\lambda l}$ is the contribution from loop or turn secondary structure, $\Theta_{\lambda u}$ is the contribution of secondary structure from peptides that do not adopt dihedral angles specified by any of the previous structures, and Ψ is the contribution from elements such as aromatic chromophores or prosthetic groups[230]. In some algorithms the estimation of secondary structure is based on the measured ellipticities of protein standards with defined conformations determined by infrared spectroscopy (IR), x-ray scattering or x-ray crystallography[231]. The spectra of known proteins are compared to those of an unknown composition using least squares analysis, ridge regression, singular value decomposition, a self-consistent method, or a neural network analysis[228]. Secondary structure determination by spectral deconvolution is accurate to within about 10 % for every population of secondary structure[232].

In the near UV region (250 nm – 320 nm), the aromatic residues tryptophan, tyrosine and phenylalanine are chromophoric and their circular dichroism is sensitive to the asymmetry of their environments. Asymmetrical environments occur when aromatic residues become buried during hydrophobic packing of folded proteins, or upon exposure to a ligand. Near UV CD studies are suited to observing changes in tertiary structure.

A two-state transition influenced by pH, denaturant, salt, or temperature can be revealed by CD, provided that the protein solution is homogeneous. A two-state transition is evident from a point in the overlap of multiple spectra acquired over a stimulus range where all the species exhibit equal molar ellipticity and where the stimulus can be thermal or chemical. The overlap is referred to as the iso-dichroic point.

1.10 Nuclear Magnetic Resonance Spectroscopy

NMR allows researchers to determine the structure and dynamics of proteins in solution. As a direct result of the sophistication of computers, the development of Fourier transformation, specialized pulse sequences, stronger magnetic fields and cryoprobes, modern NMR spectrometers are capable of characterizing the structure and dynamics of many biomolecules, including proteins up to 100 kDa[149]. NMR is uniquely applicable to the study of conformational exchange of proteins because it can obtain data that span timescales from seconds to picoseconds[148]. For slow conformational exchange, direct structural characterization is possible from the appearance of multiple resonance peaks in t he NMR spectrum and relative conformer populations can be determined by peak integration. Protein dynamics occuring on the millisecond to microsecond time scale can be observed by measuring the transverse relaxation rate during the exchange of nuclei between different conformers. Exchanging nuclei cause their coherent magnetization to be dephased, and can be refocused by a repetition of 180° radio-frequency (RF) pulses as in the Carr-Purcell-Meiboom-Gill (CPMG) experiment, or with a continuous-wave 'spin-lock' RF field as in the rotating-frame relaxation experiment[233]. The dependence of the transverse relaxation rate on the strength of the refocusing RF field defines a

relaxation dispersion profile, from which the conformer conversion rates, populations, and chemical shift differences can be extracted[234].

Nuclei such as ^{1}H, ^{15}N, ^{13}C, and ^{31}P have magnetic moments, or spin. The chemical environment of these nuclei can be probed by NMR, and this data can be manipulated to give information about the 3-dimensional structure and dynamics of proteins. When proteins containing the aforementioned isotopes are placed in the strong magnetic fields of NMR spectrometers, the spins of these nuclei precess about the field. The fundamental equation of NMR tells us that the rate of a nucleus precessing in a magnetic field is dependent on its gyromagnetic ratio:

$$\omega = -\gamma B_0$$

where ω is the precession frequency (or Larmor frequency) of the nucleus of interest, γ is its gyromagnetic ratio (constant for every nucleus), and B_0 is the applied magnetic field. The precession frequency is influenced by local magnetic fields, including currents from electrons situated in nearby molecular orbitals and solvent interactions like H-bonding, in addition to the magnetic field induced by the NMR spectrometer. The equation defining the Larmor frequency can be re-written to inlcude a term referred to as the shielding constant (σ) which is an averaged value describing local magnetic influences (or chemical shift anisotropy) of a nucleus:

$$\omega_\sigma = \gamma (B_0 - \sigma B_0)$$

The new precession frequency of the nucleus (ω_σ) is defined in terms of its chemical shielding constant, and converted to a chemical shift value (δ) in terms of the operating frequency of the NMR magnet (F). The different frequencies of nuclei that are detected by an NMR spectrometer

are obtained relative to a reference frequency from a nucleus of a known standard, in some cases 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). The frequency of the magnet is usually expressed in megahertz, and the difference between the precession of the nucleus of interest (ω_σ) and the precession of a nucleus of a standard material (i.e. DSS), ω is usually expressed in hertz, therefore the chemical shift of a nucleus in terms of the operating frequency of the NMR magnet is expressed in parts per million (ppm). The chemical shift of a nucleus of interest is expressed as:

$$\delta = \frac{\omega_\sigma - \omega}{F}$$

Chemical shift dispersion describes the spread of chemical shifts of a particular nucleus in an NMR spectrum. One of the main characteristics of intrinsically disordered or partially unfolded proteins is the low chemical shift dispersion of resonances, especially those of backbone amide hydrogens (Figure 1.15). Multi-dimensional NMR experiments combined with isotopic enrichment of nuclei of interest in protein samples (^{15}N and ^{13}C) helps to resolve resonance overlap in spectra characteristic of disordered proteins. Assignment strategies utilizing ^{15}N, $^{1}H^{N}$, and ^{13}CO nuclei in disordered proteins is particularly useful because they show better chemical shift dispersion than $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ nuclei[235].

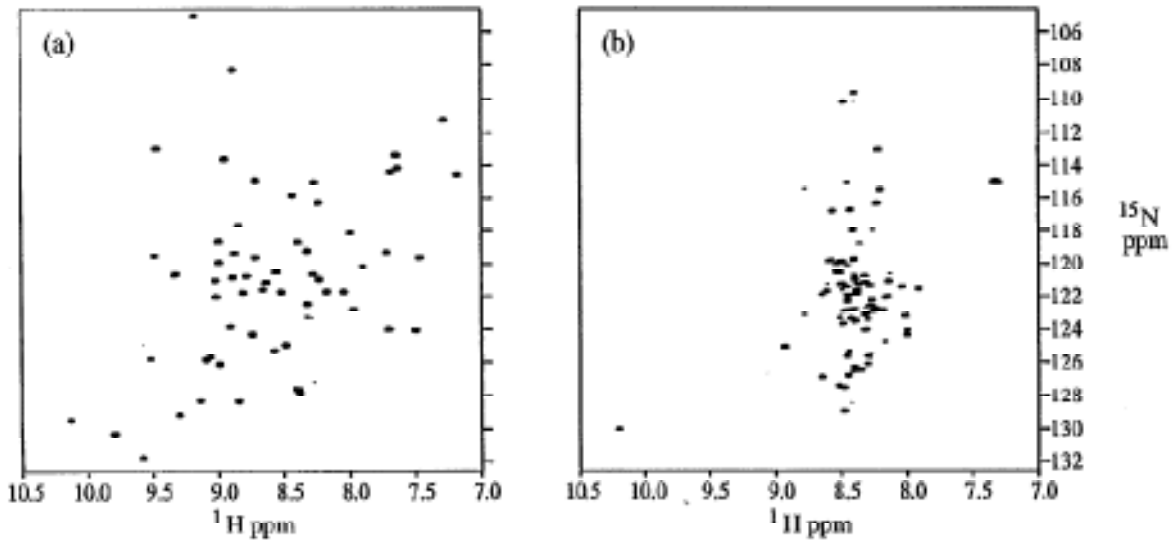

Figure 1.15 The ^{1}H-^{15}N NMR HSQC spectra of: (a) native drkN SH3; (b) the guanidium chloride unfolded state of drkN SH3[1]. Spectra were recorded at 14 °C on a 600 MHz spectrometer. Note the low chemical shift dispersion of unfolded drkN SH3 in 2 M Gdm^+ (b). Reproduced with permission from L. E. Kay[1].

The preparation of intrinsically disordered proteins at concentrations suitable for NMR studies is difficult. Inherent characteristics such as high net charge, regions of hydrophobicity, poorly defined structures and a high degree of flexibility contribute to the precipitation of intrinsically disordered proteins at concentrations desirable for NMR study (0.5-1 mM). Uniform ^{15}N and ^{13}C isotopic enrichment of the Tat protein has increased the sensitivity for NMR studies.

NMR spectroscopy is a powerful tool for the structural and dynamic characterization of intrinsically disordered proteins as their flexibility and high dynamic nature makes their

crystallization difficult for X-ray crystallography. Disordered proteins are not amenable to crystallization, although some have been crystallized with binding partners. These structures often lack complete electron density of the disordered protein because, upon binding only regions associated with the interaction attain a fixed conformation, leaving the remainder of the protein in a fluid-like state. An example of this is found in the study of the crystallization complex of Tat and P-TEFb, for which residues 58-86 of Tat could not be defined, suggesting that this region does not undergo binding-induced folding, and that this region of Tat does not interact with P-TEFb[70].

1.10.1 ^{1}H-^{15}N Heteronuclear Single Quantum Coherence Spectroscopy

One of the simplest and most common multi-dimensional heteronuclear NMR techniques that is the foundation for more complicated multi-dimensional NMR experiments is heteronuclear single quantum coherence (HSQC) spectroscopy. The ^{1}H-^{15}N HSQC experiment is sometimes used as a quick test to determine if the protein sample is adequate for higher-dimensional NMR studies. Once the assignment of backbone amide hydrogens has been accomplished by three-dimensional NMR, resonance peaks of an ^{1}H-^{15}N HSQC spectrum can be easily observed for structural changes. In theory, the number of resonance peaks of a ^{1}H-^{15}N HSQC spectrum corresponds to the number of backbone amide hydrogens with the exception of proline residues, but in practice this may not be the case. Conformational sampling on slower timescales can lead to multiple peaks for a single resonance, as seen in *cis-trans* proline

isomerization. Also, resonances may broaden and eventually disappear as conformational sampling approaches timescales in the microseconds and millisecond time domains.

Chemical shift indexing (CSI) is a method of comparison of the experimentally determined chemical shifts of a protein to experimentally determined shifts of residues in random coil conformation for the purposes of secondary structure determination[236]. The presence of several consecutive residue chemical shift differences which are similarily biased by their differences from random coil chemical shifts suggests the existence of secondary structure. $^{1}H^{N}$ and ^{15}N peptide backbone shifts upfield of the random coil shifts are representative of α-helices, whereas shifts downfield are typical of residues part of a β-sheet conformation. The magnitude of the chemical shift differences also reflects the certainty of the existence of secondary structure. According to Wishart, the amide proton, C^{α} proton, and the carbonyl carbon are the most sensitive nuclei to conformation[236]. Once resonance peaks have been assigned, variations in the chemical shifts from the standard "random" coil values provide insight into the presence of any secondary structure, or the propensity to form elements of secondary structure[237].

1.11 Purpose of the Research

The purpose of this work is to study the folding of the intrinsically disordered exon-1 Tat protein to gain a better understanding of Tat's role as a viral transcription activator *in vivo*. In order to study Tat, a set of solution state conditions will be explored in which the disordered protein remains in its monomerically reduced form. The effect of pH, counter ions, detergents, a

binding partner, and solvent on the secondary structure of the intrinsically disordered protein Tat are studied. The structural variability of Tat is an important characteristic of intrinsically disordered proteins, endowing them with a wide functional repertoire. By gaining an understanding of the forces that allow an intrinsically disordered protein to fold, we can begin to understand the mechanism by which intrinsically disordered proteins interact with binding partners, and the structural implications of these interactions. Also, the identification of molecular recognition features and understanding the conformers of the intrinsically disordered Tat_{1-72} protein, may contribute to the development of therapeutic agents in the treatment of HIV-1.

The research presented in this thesis explores conditions in which Tat is soluble to develop an understanding of the protein for study at neutral pH. Although first exon Tat and full length Tat have been shown to be disordered in acidic conditions, it is not clear if the protein remains disordered at neutral pH. Secondary structure characteristics of Tat are explored by circular dichroism. The helical inducing effects of TFE are investigated by ^{1}H-^{15}N Heteronuclear Single Quantum Coherence Spectroscopy in order to determine which regions display helical propensity.

Because HIV-1 Tat has been implicated in neural dysfunction in AIDS patients and because Tat is notoriously insoluble, the nature of insoluble Tat and conditions that lead to its precipitation will be explored. A specific aim is to determine if Tat_{1-72} aggregates in a β-amyloid conformation as this might suggest a new mechanism for Tat neurotoxicity.

Chapter 2

2. Materials and Methods

2.1 Plasmid Construction

The *E. coli* codon-optimized exon 1 Tat_{1-72} gene of the HIV-1 BH10 isolate contained in the pSV2tat72 expression vector was obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH from Dr. Alan Frankel[61].

2.1.1 Tat_{1-72} Sequence Properties

The version of Tat in this study is the 72 residue exon-1 product that is fused to a 20 residue His-tag. The sequence of the protein is as follows:

MGSSHHHHHH SSGLVPRGSH **MEPVDPRLEP WKHPGSQPKT ACTNCYCKKC CFHCQVCFIT KALGISYGRK KRRQRRRPPQ GSQTHQVSLS KQ**

Tat_{1-72} residues are designated in bold. It has been shown by spectral density mapping and hydrogen-deuterium exchange that the His-tag itself remains disordered at pH 4[238]. Previous attempts at exploiting the thrombin cleavage site to remove the tag have resulted in significant reduction of yield, possibly because of an internal thrombin cleavage site between Lys-61 and Ala-62 or the existence of a thrombin inhibitory segment Arg-76 - Pro-77 - Pro-78[239, 240]. The use of a His-tag for proteins to be studied by NMR spectroscopy is advantageous because it

permits rapid protein purification. Removal of the tag is not mandatory, as in x-ray crystallography studies, because it has been noted for a large number of proteins that the presence of the His-tag does not affect the structure or function of the fusion protein, having little effect on NMR spectra[241].

The estimated charge of the His-tagged Tat_{1-72} protein over a range of pH 1 -14 is calculated using Protein Calculator v3.3 courtesy of Chris Putnam of The Scripps Research Institute, and is shown in Table 2.1.

Table 2.1 Estimate charge of His-tagged Tat_{1-72} over a pH range of 1 -14.

pH	Charge
1	27
2	26.9
3	26.4
4	25.2
5	23.3
6	20.6
7	15.1
8	11
9	5.9
10	0.1
11	-4.8
12	-8.9
13	-12.3
14	-12.9

2.2 Expression of Unlabelled His-tagged Tat_{1-72}

50 mL of pre-sterilized Terrific Broth (TB) containing 30 µg/mL kanamycin, 34 µg/mL chloramphenicol, and 0.8% v/v glycerol in a 200 mL baffled flask were inoculated with 100-200 µL of glycerol stock of *E. coli* BL21DE3pLysS cells transformed with the pSV2tat72 expression

vector. Glycerol stocks containing bacterial culture in liquid medium combined with 60% glycerol were stored at -80°C. The 50 mL culture was incubated overnight in a rotary shaker at 37°C along with 1 L of pre-sterilized TB in a 2 L baffled flask containing 34 μg/mL chloramphenicol, 30 μg/mL kanamycin, and 0.8% glycerol. Approximately 10 mL of overnight culture was aseptically transferred to the 1 L broth pre-incubated at 37°C. Tat_{1-72} expression was induced with the addition of 60 mg isopropyl-β-D-thiogalactopyranoside (IPTG) pre-dissolved in 2 mL of double deinonized H_2O (ddH_2O) to a final concentration of 250 μM when cell growth achieved an optical density of 0.6 at 600 nm. After 5 hours, Tat_{1-72} expression was halted by cooling the culture-containing flask in an ice bath for 15 minutes. Cells were harvested by centrifugation for 20 minutes at 1700 x g at 4°C, the supernatant was removed, and the cells were re-suspended in 170 mL of extraction buffer (100 mM sodium phosphate, 10 mM Tris-HCl, 10 mM tris-2-carboxyethyl phosphine (TCEP), pH 7.2). To the cell suspension 0.1 mg each of deoxyribonuclease (DNase), ribonuclease (RNase), and lysozyme was added. The cell suspension was purged with argon for 15 minutes, sealed, and stored at -80°C until needed.

2.3 Expression of ^{15}N/ ^{13}C-labelled His-tagged Tat_{1-72}

An overnight culture of pSV2tat72 containing *E. coli* cells was grown in the same manner as for the unlabelled expression (50 mL TB containing 34 μg/mL chloramphenicol, 30 μg/mL kanamycin, and 0.8% glycerol) and incubated at 37°C in a rotary shaker. A 10 mL aliquot of overnight culture was transferred to four 2 L baffled flasks pre-incubated at 37°C, each containing 1 L of TB containing antibiotics and 0.8% glycerol. Cell growth was monitored by optical density measurements at 600 nm until the absorbance reading was 0.6. Cell growth was

halted by submerging the flasks in an ice bath for fifteen minutes. The culture was centrifuged for 20 minutes at 1700 x g at 4°C, the supernatant was removed and the cell pellet re-suspended in 50 mL of M9 minimal medium[242] to wash away residual TB medium. The re-suspension was centrifuged again for 20 minutes at 1700 x g at 4°C and the supernatant was removed. The cell pellet was resuspended in 1 L of M9 minimal medium supplemented with vitamins and nutrients and pre-incubated at 37°C. The contents of the M9 minimal medium are listed in Table 2.2. If ^{15}N isotopically labelled Tat_{1-72} protein was desired, 0.7 g of $^{15}NH_4Cl$ and 2 g of D-glucose that were pre-dissolved in 10 mL of M9 minimal medium were added to 1 L of M9 medium. If ^{15}N and ^{13}C isotopically labelled Tat_{1-72} protein was desired, 0.7 g of $^{15}NH_4Cl$ and 2 g of $^{13}C_6$ D-glucose pre-dissolved in 10 mL of M9 minimal medium were added to 1 L of M9 minimal medium. The cells were allowed to adjust to the medium for 15 minutes, and overexpression of Tat_{1-72} was induced with 240 mg of IPTG, at a final concentration of approximately 1 mM. Tat expression was allowed to proceed for 5 hours, and expression was halted by cooling the 2 L baffled flask in an ice bath for 15 minutes. Cells were harvested by centrifugation at 1700 x g at 4°C for 20 minutes then the supernatant was removed. The cell pellet was re-suspended in 170 mL of extraction buffer. To the lysis buffer, 0.1 mg each of DNase, RNase, and lysozyme were added, the suspension was purged with argon for fifteen minutes, sealed, and stored at -80°C until needed.

Table 2.2 M9 Minimal Medium used for the expression of ^{15}N or $^{15}N/^{13}C$ enriched Tat_{1-72}[242].

Component	Concentration (mM)
KH_2PO_4	22
Na_2HPO_4	42
$^{15}NH_4Cl$	12.8
$MgSO_4$	2
$CaCl_2$	0.01
NaCl	8.5
$FeSO_4$	0.01
U-^{13}C6-glucose	10.7
$(NH_4)_6(MoO_7)_{24}$	3×10^{-6}
H_3BO_3	4×10^{-4}
$CoCl_2$	3×10^{-5}
$CuSO_4$	1×10^{-5}
$MnCl_2$	8×10^{-5}
$ZnSO_4$	1×10^{-5}
Choline chloride	2.9×10^{-3}
Folic acid	1.1×10^{-3}
Pantothenic acid	2.1×10^{-3}
Nicotinamide	4.1×10^{-3}
Myo-inositol	5.5×10^{-3}
Pyridoxal hydrochloride	2.4×10^{-3}
Thiamine hydrochloride	1.5×10^{-3}
Riboflavin	1.4×10^{-4}
Biotin	4.1×10^{-3}

2.4 Purification of His-tagged Tat_{1-72}

The frozen cell lysate was subjected to 3 freeze-thaw cycles. The lysate was thawed until it was completely melted, but not allowed to warm to room temperature. After the last thaw, the lysate was put onto ice and sonicated by three 30 s bursts at medium intensity, each period of sonication separated by a 30 s rest interval. Guanidine-HCl was added to the lysate to a final concentration of 5 M. The lysate was centrifuged at 12 000 x g at 4°C for 40 minutes. The supernatant containing soluble Tat protein was removed, equilibrated to room temperature, and adjusted to pH 7.2 with aqueous sodium hydroxide.

The His-tagged Tat_{1-72} protein was purified by cobalt affinity resin chromatography. The metal affinity column was prepared using 4 mL of TALON Superflow^{TM} Metal Affinity Resin (Clontech) stored in 20% ethanol. The resin was prepared following the recommendations in the TALON^{TM} Metal Affinity Resins User Manual. The resin was charged with 10 column volumes (CV) of 50 mM cobalt (II) chloride, washed with 7 CV of ddH_2O, 3 CV of 0.3 M sodium chloride, and 3 CV of ddH_2O. The resin was equilibrated with 30 mL of extraction buffer containing 5 M guanidine-hydrochloride (Gdn-HCl). The lysate was poured onto the column. The column was washed with 20 mL of extraction buffer containing 5 M Gdn-HCl, then with 30 mL of wash buffer (50 mM sodium phosphate, 10 mM TCEP, pH 6.6) containing 5 M Gdn-HCl. The protein was eluted with 10 mL of 0.22 μm filtered elution buffer (50 mM acetate, 10 mM TCEP, pH 4) containing 5 M Gdn-HCl. The first 3 mL were discarded and seven 1 mL fractions were collected. The first three fractions of darkest intensity were presumed to be of highest Tat_{1-72} concentration, and were each diluted five-fold with elution buffer and dialyzed in separate dialysis tubing. Fractions 4 and 5 were combined, fractions 6 and 7 were combined, and each combination was diluted two-fold with elution buffer and placed into separate dialysis tubing. The fractions were serially dialyzed against three 1 L dialysis buffers of 0.1 M, 0.05 M, 0.01 M glacial acetic acid pH 3.5 and finally against 1 L of ddH_2O pH 3.5. Each dialysis buffer was prepared by pH adjustment with hydrochloric acid, degassing by sonication under a vacuum for 15 minutes, and purging with argon for 15 minutes. The duration of dialysis was as follows: 4 hour dialysis in 0.1 M acetic acid, overnight dialysis in 0.05 M acetic acid, 4 hour dialysis in 0.01 M acetic acid, and 3 hour dialysis in ddH_2O. A 50 uL sample was removed from each bag, was diluted ten-fold by the dialysate and the ultraviolet absorbance at 280 nm was measured; the dialysate was used as a blank. The samples in the five dialysis bags were individually

transferred to five 50 mL polypropylene conical tubes, were shell frozen, and lyophilized for 48 hours. The freeze-dried protein was stored at -20°C until needed.

2.5 Ultraviolet Absorbance Spectroscopy

The calculated molar extinction co-efficient (ε) of Tat_{1-72} at 280 nm is 9090 cm^{-1} M^{-1}. All UV spectroscopy samples were placed into a 1 cm quartz cuvette. All spectra were collected using a dual beam reference UV spectrophotometer. All Tat_{1-72} samples were referenced to solutions containing the appropriate buffer blanks, all spectra were baseline corrected, and control spectra were subtracted when necessary. The concentration of Tat_{1-72} protein was determined by the application of the Beer-Lambert law: Absorbance = ε · concentration · cell path.

2.5.1 Tat_{1-72} Solubility Studies

Tat_{1-72} solubility studies were conducted with the aim of finding conditions in which Tat_{1-72} is stable, monomeric, reduced, and soluble approaching physiological pH. Various criteria were taken into consideration when preparing solutions for these studies: presence of reducing agent, buffer identity, ionic additives and ionic strength, presence of detergent, organic co-solvent, biologically relevant co-factors.

2.5.1.1 Detergent Assays

Detergent assays with reducing agent were prepared above critical micellar concentrations[243]. A stock Tat_{1-72} solution of about 1 mg/mL was prepared with an initial concentration between 100-110 μM by solvating lyophilised protein with 20 mM sodium acetate, 20 mM MES, 2.5 mM TCEP, at pH 4. Stock Tat_{1-72} was diluted in half with various detergent solutions to a final concentration of 50 – 55 μM. Detergent solutions were prepared so that their final concentration after dilution with stock Tat_{1-72} would be about 10 times their critical micellar concentration (CMC), and in a similar buffer (20 mM sodium acetate/MES, 2.5 mM TCEP, pH 4). Following a 30 minute equilibration period, a UV spectrum was collected. The sample was titrated with 5 μL aliquots of dilute ammonium hydroxide, and allowed to equilibrate for 30 minutes. UV spectra were collected at the end of each equilibration period. A bulk absorption of reference solutions of the same condition and pH was subtracted from every sample, and all dilution corrections were accounted for (maximum dilution factor of 1.03). Figures 2.1 – 2.6 show the structures of the detergents, Figures 2.7 – 2.9 show the structures of the buffering agents, Figure 2.10 shows the structure of the reducing agent TCEP, and Figure 2.11 shows the structure of the organic solvent TFE.

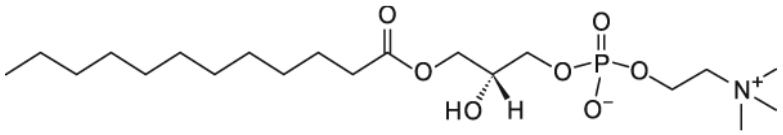

Figure 2.1 1-Lauroyl-2-hydroxy-*sn*-glycero-3-phosphocholine (LHGP)

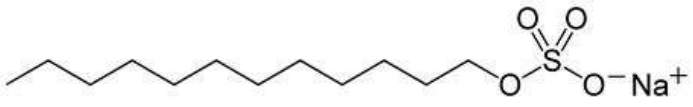

Figure 2.2 Sodium lauryl sulfate (SDS)

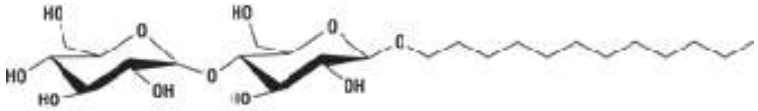

Figure2.3 n-Dodecyl-beta-D-maltoside (NDBM)

Figure2.4 Dodecyl trimethyl ammonium bromide (DTAB)

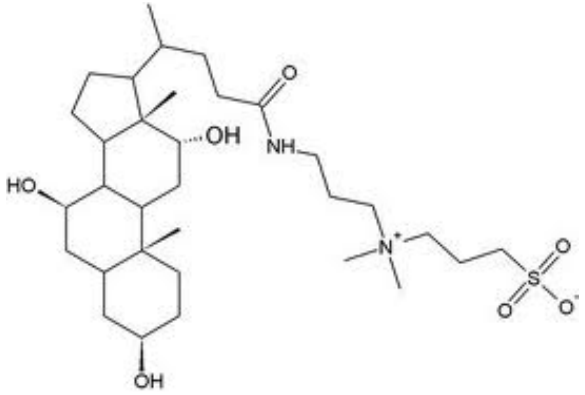

Figure 2.5 3-[(3-Cholamidopropyl) dimethylammonio] propanesulfonic acid (CHAPS)

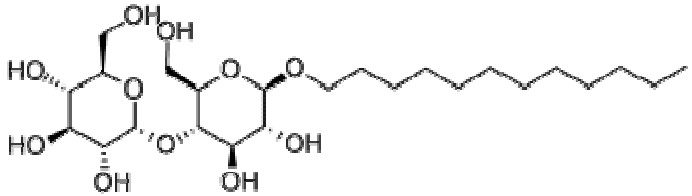

Figure 2.6 N-dodecyl beta-D-maltoside (DDM)

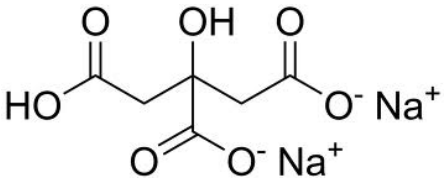

Figure 2.7 Disodium citrate

Figure 2.8 2-(*N*-morpholino)ethanesulfonic acid (MES)

Figure 2.9 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)

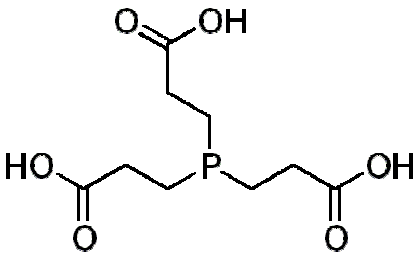

Figure 2.10 Tris(2-carboxyethyl)phosphine (TCEP)

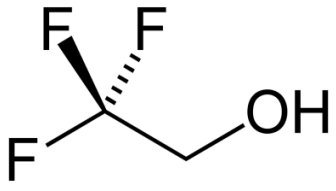

Figure 2.11 2,2,2-Tri-fluoroethanol (TFE)

Detergent assays without reducing agent were done at concentrations above the CMC and buffering agents included buffers chosen for their minimal "salt" effects in water, and their compatibility with Tat_{1-72} at pH 4 as determined by initial solvation trials. It was observed that for detergent solutions without reducing agent (in the absence of Tat_{1-72}) no considerable change in the bulk absorption of solution occurs when varying the pH. Samples were titrated with 3 μL of dilute ammonium hydroxide and equilibrated for 24 hours between readings. All spectra were referenced to appropriate blank solutions.

2.5.1.2 Trifluoroethanol Solubility

Tat_{1-72} was dissolved in 10 mM acetate, 10 mM MES, pH 4. The sample was centrifuged at 13 000 X g until no visible protein pellet formed, it was spin filtered through a 0.22 μm filter, and diluted with trifluoroethanol to a final solution composition of 30% TFE, 7 mM acetate, 7 mM MES, apparent pH 4.7. The concentration of Tat_{1-72} was 110 μM. Concentrated zinc (II) acetate was added to the Tat_{1-72}/30% TFE solution to a final concentration of 220 μM (two molar equivalents of zinc to Tat_{1-72}). The sample was incrementally titrated to pH 7.2 with dilute ammonium hydroxide over a period of 6 hours (secondary structure analysis by circular dichroism was done between pH changes). At pH 7.2, a UV spectrum was acquired between 240-320 nm. An additional spectrum was collected after a three day equilibration. All spectra were referenced to appropriate blank solutions.

2.6 Circular Dichroism Spectropolarimetry

CD spectra were collected using a Jasco J-810 spectropolarimeter. Samples were equilibrated to 25°C (unless noted otherwise) by an external Julabo 25-HE water bath. The basic acquisition parameters were (unless specifically indicated elsewhere):

Cell Length: 0.01 cm, Sensitivity: Standard (100 mdeg), Scan Rate: 5 nm/min,
Response Time: 8 s, Accumulations: 1, Bandwidth: 1 nm, Range: 180 nm – 260 nm

The measurements in units of millidegrees were converted to mean residue ellipticity (MRE) by the formula:

$$[\theta_\lambda]_{MRE} = \frac{[\theta_\lambda] \bullet 0.001 \bullet M}{n \bullet C \bullet l \bullet 10}$$

MRE is reported in deg•cm^2•dmol^{-1}. $[\theta_\lambda]$ is the ellipticity in units of millidegrees, M is the atomic mass of the protein, n is the number of amino acids in the protein, C is the mg/ml concentration of protein, l is the cuvette pathlength in centimetres. It is desirable to report ellipticity as MRE because it expresses the ellipticity of a protein as contributed by the molar concentration of peptide units.

The spectropolarimeter was calibrated with (+)-10-camphorsulphonic acid. Wavelength is calibrated in reference to a negative band at 192.5 nm, and a positive band at 290.5 nm. Intensity is calibrated by the absorbance at these two wavelengths, where the ellipticity (ε) is dependent on the concentration of the calibrant. At 192.5 nm, the $\Delta\varepsilon = -4.9$ L · mol^{-1} cm^{-1};

ΔAbsorbance (A) = –1.05 X 10^{-3}. At 290.5 nm, Δε = 2.36 L · mol^{-1} · cm^{-1}; ΔA = 5.08 X 10^{-4}. The ratio of intensities should be 2, which is an indicator of the accuracy of the instrument at lower wavelengths[244].

2.6.1 Deconvolution of CD Spectra

Methods have been developed to deconvolute the various contributing secondary structural elements present in a protein solution from the measured CD spectrum. These algorithms rely on reference CD spectra of proteins of known structure, solved by IR, x-ray diffraction patterns, and crystal structure determination. DICHROWEB is an analysis web server hosted by the Department of Crystallography at Birkbeck College at the University of London[245] and was accessed for structural prediction using three different analysis programs: CONTINLL[232, 246] Selcon3 (self-consistent method)[247, 248], and CDSSTR[249]. Deconvolution is performed on the far UV spectrum between 185 nm – 260 nm.

2.6.2 The Effect of Sodium Chloride on Tat_{1-72} Secondary Structure

Two milligrams of lyophilized Tat_{1-72} was dissolved in 50 mM citrate, 100 mM NaCl, pH 5.5. The sample was centrifuged at 13 000 X g on a benchtop Fisher micro-centrifuge model 235A until no visible pellet was formed and then it was equilibrated for 24 hours. The final concentration of Tat_{1-72} was 190 μM and a CD spectrum was acquired. The spectrum was baseline corrected.

For a separate trial, about two milligrams of lyophilized Tat_{1-72} was dissolved in 50 mM citrate, 100 mM NaCl, pH 5.10. The sample was centrifuged at 13 000 X g on a benchtop Fisher micro-centrifuge until no visible pellet was formed, titrated to pH 5.35 and then equilibrated for 24 hours. The final concentration of Tat was 220 μM and a CD spectrum was acquired and corrected with a baseline.

2.6.3 The Effect of Trifluoroethanol on Tat_{1-72} Secondary Structure

2.6.3.1 Tat_{1-72} Solvation in 0% - 90% Trifluoroethanol

A stock Tat_{1-72} solution was prepared in 10 mM acetate, pH 4. The sample was centrifuged until no visible pellet formed. The concentration of Tat_{1-72} was determined to be 550 μM. A series of solutions ranging from 0-90% TFE, in 10 mM acetate at pH 4 was prepared. An aliquot of Tat_{1-72} was added to each TFE solution so that the final concentration of Tat_{1-72} was 100 μM, except for the 85% and 90% TFE samples, which had a Tat_{1-72} concentration of 83 μM and 55 μM, respectively. The solutions were equilibrated for 30 minutes prior to the measurement of pH and acquisition of a CD spectrum. The spectra were baseline corrected and dilution factor corrections were applied to the 85% and 90% TFE samples.

2.6.3.2 The Effect of Temperature on Tat_{1-72} in 20% Trifluoroethanol

A Tat_{1-72} solution was prepared in 10 mM acetate, 10 mM TCEP, pH 4. It was centrifuged until no visible pellet was formed, and diluted with TFE to a final TFE concentration of 20% and apparent pH of 4.2. The concentration of Tat_{1-72} was 100 μM. The sample was

allowed to equilibrate for 30 minutes at 25°C prior to the acquisition of a CD spectrum. CD spectra were then acquired at 5, 25, 35, 50, 65, and 80°C after a 30 minute equilibration period at each temperature. The spectra were baseline corrected.

2.6.3.3 Zinc Titration of Tat_{1-72} in 30% Trifluoroethanol

Tat_{1-72} was prepared in 10 mM acetate, pH 4.0, equilibrated overnight and diluted with TFE to a final TFE concentration of 30% and apparent pH 4.28. The final concentration of Tat_{1-72} was 140 μM. The solution was titrated with 0.5, 1, 2, 3, and 4 molar equivalents (MEQ) of zinc (II) acetate. Prior to the acquisition of a CD spectrum, the sample was allowed to equilibrate to the new condition for 15 minutes. The spectra were baseline corrected.

2.6.3.4 pH Titration of Tat_{1-72} in 30% Trifluoroethanol with 2 Molar Equivalents of Zinc(II)

Tat_{1-72} was dissolved in 10 mM acetate, 10 mM MES, pH 4 and allowed to equilibrate overnight. The solution was diluted with TFE to a final concentration of 30% TFE. The concentration of Tat_{1-72} was 100 μM and 2 MEQ of Zn(II) were added to the solution. The sample was titrated from pH 4.69 to 7.02 with aqueous ammonium hydroxide. After the addition of titrant, the sample was allowed to equilibrate for 15 minutes, followed by the acquisition of a CD spectrum. The spectra were baseline corrected and a dilution correction was applied (maximum dilution factor of 1.0035).

2.6.4 The Effect of SDS on Tat_{1-72} Secondary Structure

A 95 μM stock solution of Tat_{1-72} in 10 mM acetate, pH 4.0 was diluted to 15 μM in a solution of 2 mM SDS. The SDS was 2 mM in 10 mM acetate at pH 4. The sample was equilibrated for 15 minutes followed by the acquisition of a far UV CD spectrum. The spectrum was baseline corrected.

2.6.5 Tat_{1-72} and TAR in the presence of DDM and 2 MEQ Zn(II)

TAR was provided by Professor Hashim M. Al-Hashimi of the Department of Chemistry & Biophysics Research Division at the University of Michigan in Ann Arbor, Michigan. The ribonucleotide was prepared by Ms. Liz Dethoff. The TAR ribonucleotide in this study consists of 29 nucleotides. The stem-loop RNA is the shortened version of the 5' viral mRNA beginning at guanosine-17 and ending at cytidine-45. The secondary structure of the truncated version of HIV-1 TAR is shown in Figure 2.12.

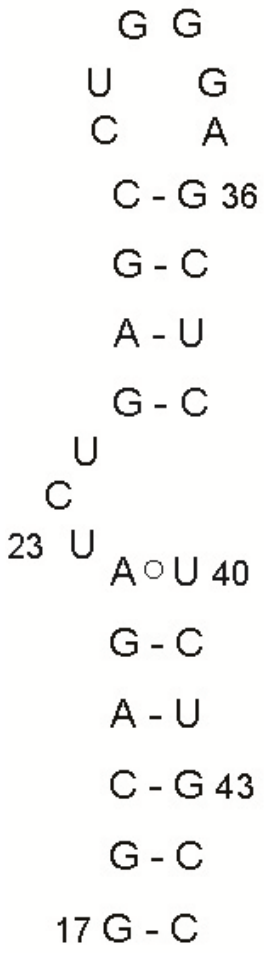

Figure 2.12 The 29 nucleotide truncated version of HIV-1 TAR.

A 95 μM stock solution of Tat_{1-72} in 10 mM acetate, pH 4.0 was diluted with 10 mM acetate, pH 4.0 to a 50 μM working solution. To the Tat_{1-72} solution, 2 μl of concentrated zinc (II) acetate was added to a final concentration of 100 μM. 200 μl of 20 μM TAR in 10 mM acetate at pH 4 containing 15 mM DDM was prepared. 80 μl of the Tat_{1-72} solution was added to the TAR/DDM solution. The final concentrations were: 14 μM Tat_{1-72}, 14 μM TAR, 10 mM DDM, 28 μM zinc(II) acetate, 10 mM acetate. The sample was equilibrated for 15 minutes followed by the acquisition of a far UV CD spectrum. The spectrum was baseline corrected. The sample precipitated after a 24 hour equilibration at pH 4.0.

2.7 NMR Sample Preparation

For NMR samples without TFE, lyophilized, uniformly ^{15}N enriched Tat_{1-72} was dissolved in 700 µl of 10 mM acetate aqueous buffer pH 4 containing 10 mM TCEP, 5% D_2O, and a trace of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), to a final Tat_{1-72} concentration of 0.5 mM – 0.7 mM. The appropriate volumes of TFE were added to the sample for final TFE concentrations of 7% and 15% TFE. After the addition of TFE, the protein sample was purged with argon for 5 minutes and sealed prior to NMR acquisition.

The $^{13}C/^{15}N$-labelled Tat_{1-72} in 30% TFE for two and three dimensional NMR analysis was prepared by addition of lyophilized $^{13}C/^{15}N$-labelled Tat_{1-72} in 500 µl of 10 mM acetate buffer pH 4, 10 mM TCEP, 5% D_2O, and trace amounts of DSS, to a concentration of 0.7 mM – 0.8 mM protein. The sample was diluted with TFE to 30% and purged with argon for five minutes prior to NMR acquisition.

2.8 Nuclear Magnetic Resonance Spectroscopy

All NMR spectra of ^{15}N and $^{15}N/^{13}C$ labelled Tat_{1-72} were acquired using a 600 MHz Varian INOVA spectrometer with a 14.1 Tesla field strength. The spectrometer was equipped with a triple resonance probehead at 25 °C. The pulse sequences for ^{1}H-^{15}N HSQC[250], HNCO[251], and HN(CA)CO[252] were standard gradient sensitivity enhanced Varian BioPack pulse sequences[253]. Proton shifts were referenced to water resonating at 4.82 ppm from the internal standard DSS; ^{15}N and ^{13}C were referenced indirectly[236].

The sweep widths for the 2-dimensional ^{1}H-^{15}N HSQC spectra of ^{15}N enriched Tat_{1-72} in 0%, 7%, and 15% TFE were 16 parts per million (ppm) in the direct dimension (centre frequency of 599.68 Hz), and 32 ppm in the indirect dimension (centre frequency of 60.76 Hz). An excitation pulse width of 6.7 μs was employed. 1024 X 128 complex points were collected, with an acquisition time of 0.107 seconds. A total of 32 scans were collected. Data were Fourier transformed using NMRPipe[*(254)*] and Spinworks. Prior to transformation spectra were zero filled to twice the data set, apodized using a squared cosine bell function and Zhu-Bax linear predicted. Spectra were analyzed using Spin Works 3.1.7 (courtesy of Dr. Kirk Marat at the University of Manitoba) and Sparky[*(255)*].

The sweepwidths for the 3-dimensional HNCA and HN(CO)CA for partial assignment of $^{15}N/^{13}C$ Tat_{1-72} in 30% TFE was 8 ppm in the ^{1}H dimension (centre frequency of 60.76 Hz), 17 ppm in the ^{13}C dimension (centre frequency of 150.82 Hz), and 32 ppm in the ^{15}N dimension (centre frequency of 60.77 Hz). An excitation pulsewidth of 6.7 μs was employed. 1024 X 128 X 32 complex points were collected with an acquisition time of 0.054 seconds. A total of 8 scans were obtained. Data were Fourier transformed using NMRPipe. Prior to transformation, spectra were zero filled to twice the data set, apodized using a squared cosine bell function, and linear predicted. Spectra were analyzed using Sparky[*(255)*].

2.8.1 Chemical Shift Indexing

2.8.1.1 ^{1}H-^{15}N HSQC Spectra of ^{15}N enriched Tat_{1-72} in 0 %, 7 %, 15 %, 30 % TFE

Chemical shifts can be useful indicators of secondary structure. The chemical shifts of peptide backbone nuclei like $^{1}H^{N}$ and ^{15}N depend on their local orientation and the electronic shielding effects of local nuclei. The chemical shift dispersion of $^{1}H^{N}$ and ^{15}N in denatured or intrinsically disordered proteins make them viable candidates for study by NMR. Chemical shift indexing of ^{15}N and $^{15}N/^{13}C$ enriched Tat_{1-72} was done by determining the difference in chemical shifts between Tat_{1-72} residues and those of experimentally determined residues in the random coil conformation. A reference set of random coil values were determined by Schwarzinger et al.[256]. The chemical shifts of residues in the random coil state were determined from a series of model pentapeptides of structure Ac-G-G-X-G-G-NH_2 in 8 M urea. Due to the sensitivity of amide proton and amide nitrogen chemical shifts on neighbouring residues, the reference random coil shifts were corrected for the identity of the preceding and proceeding residues[2] by the correction factor calculation:

$$\delta_R(\text{corrected}) = \delta_{random}(R) + \Delta\delta(R_{-1}) + \Delta\delta(R_{+1})$$

where δ_R(corrected) is the corrected random coil shift value for residue 'R', δ_{random} is the uncorrected random coil shift value for 'R', $\Delta\delta(R_{-1})$ is the chemical shift correction based on the identity of the preceding residue, and $\Delta\delta(R_{+1})$ is the chemical shift correction based on the identity of the proceeding residue.

Chemical shift indexing has been useful in showing the existence of secondary structure in well folded and partially unfolded proteins[257, 258]. The shift differences in wholly disordered proteins are smaller, but have indicated transient structure or structural propensity [259].

2.9 Congo Red Binding Assay

Congo Red is the sodium salt of benzidinediazo-bis-1-naphthylamine-4-sulfonic acid, formula: $C_{32}H_{22}N_6Na_2O_6S_2$, molecular weight: 696.66 g/mol (Figure 2.13). It is a derivative of benzidine and naphthionic acid. The molecule is referred to as an azo dye, containing two azo groups. These dyes are intensely coloured due to n-electron delocalization. When dissolved in aqueous buffers of alkaline or weakly acidic pH, CR solutions are red and absorb maximally at 500 nm. In unpolarized light, the dye appears red in stained amyloids, due to a maximal absorption at 540 nm. The exact mechanism by which CR interacts with amyloid fibrils is unclear, but it is believed that it forms non-covalent interactions by hydrogen bonding with β-sheet structure. The molecule contains two amine and two sulfate groups assumed to intercalate with β-sheet structures, as typically seen in amyloid.

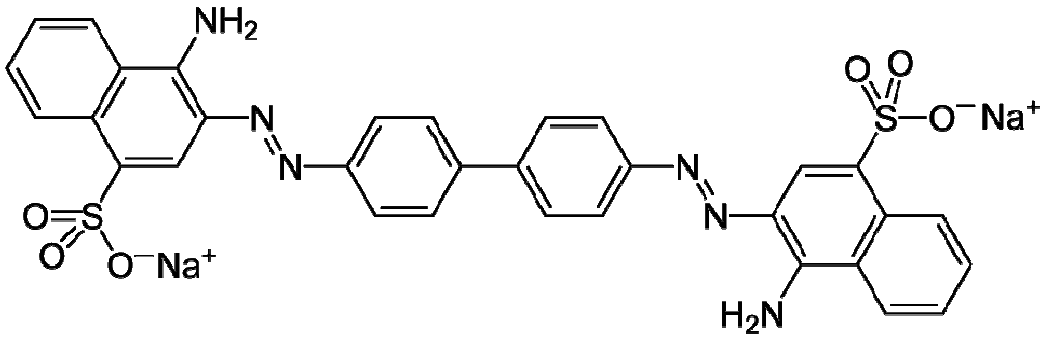

Figure 2.13 Di-sodium salt of benzidinediazo-bis-1-naphthylamine-4-sulfonic acid.

A stock CR solution was prepared by 0.2 μm filtration of a 0.4 mM CR solution in 5 mM potassium phosphate, 150 mM sodium chloride, pH 7.2. The final concentration of CR was 0.375 mM. A 20 μM solution of Tat_{1-72} was prepared by solubilising lyophilized protein in 50 mM citrate, 100 mM NaCl, pH 4. The 1 mL solution was centrifuged for 1 minute at 13 000 X g until there was no visible pellet using a benchtop Fisher micro-centrifuge model 235A. The Tat_{1-72} solution was adjusted to pH 5.35 with aqueous ammonium hydroxide. 10 μL of CR stock solution was diluted into 490 μL of Tat_{1-72} solution (50 mM citrate, 100 mM NaCl, pH 5.35) to a final concentration of CR of approximately 7.5 μM. After a 40 minute equilibration period, a UV spectrum was acquired in the range 400 nm – 600 nm. The addition of 10 μL of 0.375 mM CR to 50 mM citrate, 100 mM NaCl, pH 5.45 was repeated in the absence of protein and a spectrum was acquired in the range 400 nm – 600 nm. An increase in absorbance at 540 nm of the CR + Tat_{1-72} spectrum compared to CR alone is indicative of an association between CR and an amyloid-like structure[194, 260].

2.9.1 Polarized Light Microscopy

A 1 mL sample of 200 μM Tat_{1-72} in 100 mM NaCl, 10 mM acetate, pH 4 was combined with 5 μL of CR from a CR stock solution consisting of 10 mM CR, 5 mM potassium phosphate, 150 mM sodium chloride, pH 7.2, giving a final CR concentration of 50 μM. A 5 μL aliquot was placed on a glass slide, and observed through an Olympus BX51 polarizing optical microscope for green-yellow birefringence.

2.9.2 Amyloid Specific Sequence Determination Using Web-based Algorithm "WALTZ"

The sequence of the 92 residue His-tagged Tat_{1-72} protein was submitted to the web-based algorithm "WALTZ" provided at the internet address http://waltz.vub.ac.be/submit.cgi, accessed Oct. 13, 2010[*189*].

Chapter 3

3. Results

3.1 Expression of His-tagged Tat_{1-72}

Typically, a 1 L TB culture of BL21(DE3)pLysS cells grown to an optical density of 0.6 – 0.9 absorbance units at 600 nm and expressing unlabelled Tat_{1-72} would yield a wet mass of cells of 10-12 g after centrifugation. A 4 L TB culture of BL21(DE3)pLysS cells grown to the early stages of exponential growth (~ 2 hours) for expression of ^{15}N or $^{15}N/^{13}C$ Tat_{1-72} would also yield 10-12 g of cells. Cells were resuspended in a large volume of re-suspension buffer for numerous reasons. Tat_{1-72} forms multimers possibly by ionic interactions of its acidic, proline-rich region and basic, arginine-rich region, and irreversibly via cysteine linkages between thiol groups. Dilution is thought to reduce the likelihood of multimerization. In smaller lysis buffer volumes, the cell extract was particularly viscous, possibly slowing the diffusion of each of DNase, RNase, and lysozyme throughout the solution. Lysate viscosity was decreased by the addition of 0.1 mg of DNase, RNase, and lysozyme. Tat_{1-72} is basic, and at pH 7.2 the estimate net charge is +25, thus it adheres to anionic molecules. Cell lysis is achieved by freeze-thaw cycles in preference to the French press. Once the plastic bottle containing the cells is purged and sealed under argon to prevent oxidative cross-linking of the protein, the cells can be lysed by freeze-thawing with minimal exposure to atmosphere. A French press is not applied because it would be more difficult to keep the solution reduced[261].

3.2 Purification of His-tagged Tat_{1-72} and $^{13}C/^{15}N$-enriched His-tagged Tat_{1-72}

Purification of the recombinant His-Tat_{1-72} fusion protein is achieved by making use of the affinity of the poly-His tag for immobilized cobalt ions of a resin. Cobalt resin is employed over nickel resin for its selective affinity of the His-tag, even in the presence of high concentrations of denaturant. The Talon Superflow Metal Affinity Resin is part structural sepharose bead, and part functional chelator of cobalt. It co-ordinatets to cobalt ion through 4 of 6 metal co-ordination sites leaving two sites available for His. When the resin is equilibrated with cobalt ion, it is pink in colour. The imidazole group of Histidine has a pKa of approximately 6 depending on the effects of neighbouring amino acids. The protonation of either tautomeric form of Histidine creates a resonance stabilized positive charge over the imidazole ring that is repulsed from the cationic cobalt. Thus, affinity for the cobalt-resin can be controlled by pH. His-Tat_{1-72} is retained on the column when the lysate is poured over the column at pH 7.2. Tat_{1-72} retention on the column was observed by a brown discoloration of the pink resin. The brown band was eluted at pH 4, returning the pink colour of the cobalt-resin. Denaturant and reducing agent are removed by serial dialysis, and the protein appears as a fine, fluffy lyophilisate. Protein expression of unlabelled, ^{15}N, and $^{15}N/^{13}C$ enriched Tat_{1-72} usually yielded 10-15 mg of unaggregated, pH 4 soluble, lyophilized protein after purification.

3.3 Ultraviolet Absorption Spectroscopy

Routine determination of Tat_{1-72} concentrations were done by ultraviolet absorption spectroscopy. The concentrations of Tat_{1-72} solutions were determined by absorption at 280 nm, using the theoretical Tat_{1-72} molar extinction coefficient of $\varepsilon_{280} = 9090\ M^{-1}\ cm^{-1}$. The precise determination of Tat_{1-72} at higher concentrations is difficult due to Tat_{1-72} self-association that causes light scattering, therefore Tat_{1-72} was always measured at concentrations below 100 µM. The absorption spectrum of Tat_{1-72} in this range (Figure 3.1) is dominated by a characteristic tryptophan absorption at 280 nm overlapping with a tyrosine absorption, and a characteristic trough at 250 nm. In Figure 3.1 the absorption decay above 300 nm is evidence of a lack of light scattering indicative of a well solubilized protein. Figure 3.2 shows the absorption of 0.033 mM tryptophan in 10 mM acetate at pH 4. Note the incomplete lack of absorption above 300 nm. Tat_{1-72} protein does not decay completely to zero which may be a result of a very small population of protein that is aggregated possibly because the protein self-associates non-covalently.

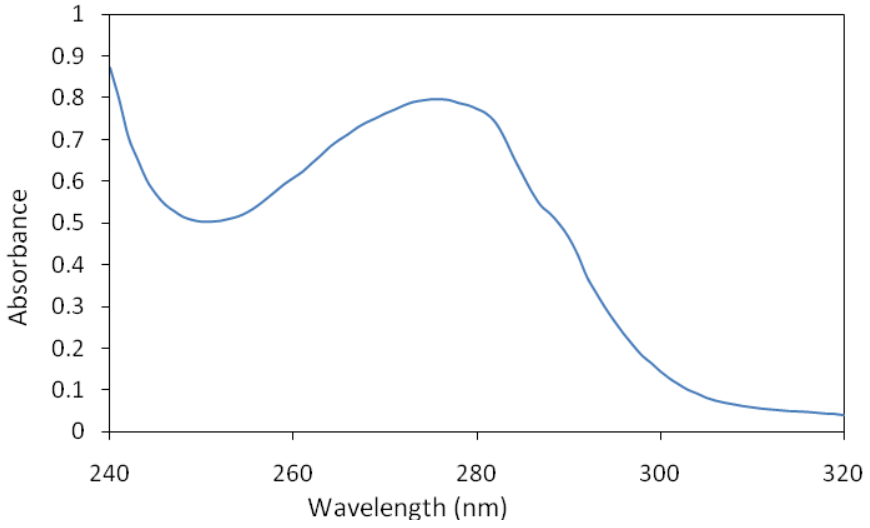

Figure 3.1 UV absorbance of 85 µM Tat_{1-72} in 10 mM acetate, pH 4.

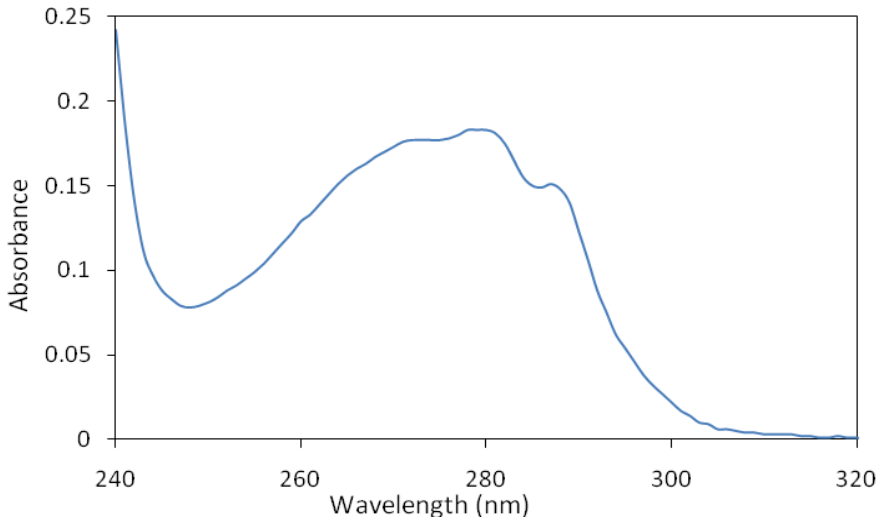

Figure 3.2 UV absorbance of 33 μM tryptophan in 10 mM acetate, pH 4.

3.4 Circular Dichroism Spectropolarimetry

Figure 3.3 shows the far UV CD spectrum of Tat_{1-72} in 10 mM TCEP, 10 mM acetate, pH 4. The spectrum contains a negative CD band at 200 nm characteristic of a disordered random coil. Table 3.1 shows the results of the deconvolution of the spectrum by the Selcon3, CONTINLL, and CDSSTR methods, and they all show that the protein is mainly disordered[230]. This is in agreement with previously published results of multinuclear NMR data by Dr. Shaheen Shojania showing that HIV-1 Tat_{1-72} is intrinsically disordered at pH 4[20]. The average of the three algorithms shows that the highest amounts of secondary structure occupies the disordered and turn conformations. There is some disagreement between the Selcon3/CONTINLL methods and CDSSTR method between the proportion of structure defined as turn and disordered. Selcon3 and CONTINLL respectively predict 44% and 48 % disordered and 27% and 29 % turn,

whereas CDSSTR predicts 78% disordered and 8% turn. A turn is defined as the close approach of two C^{α} atoms (< 7 Å) when the corresponding residues are not involved in an α-helix or β-sheet structure[262]. The turn may or may not be stabilized by a hydrogen bond between the backbone amide proton and the carbonyl group. In the case where the turn is stabilised by an H-bond, it is referred to as a γ-turn when the H-bond is separated by one residue; it is called a β-turn when the H-bond is separated by two residues; it is an α-turn when the H-bond is separated by three residues; and it is a π-turn when the H-bond is separated by four residues[262]. The CD spectrum of a turn is similar to that of a β-sheet. The differences in secondary structure determination between Selcon3/CONTINLL and CDSSTR may be due to differences in the data sets upon which the deconvolutions are based.

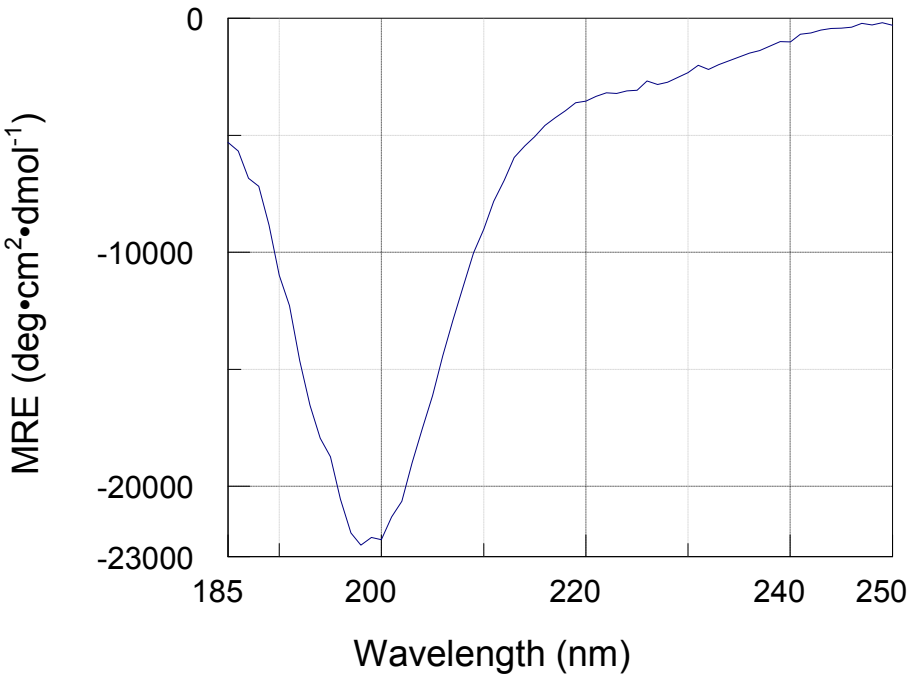

Figure 3.3 CD spectrum of 190 μM Tat_{1-72} in 10 mM TCEP, 10 mM acetate, pH 4.

Table 3.1 Secondary structure determination by deconvolution of the CD spectrum in Figure 3.3.

Algorithm	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Disordered	Total
Selcon3	0.042	0.092	0.077	0.095	0.266	0.438	1.01
CONTINLL	0.037	0.036	0.068	0.083	0.293	0.484	1.001
CDSSTR	0.01	0.03	0.05	0.04	0.08	0.78	0.99
Average	0.030	0.053	0.065	0.073	0.213	0.567	1.000

3.5 Tat_{1-72} Solubility Studies

The aim of these studies was to select for conditions that could solubilize Tat_{1-72} in conditions approaching neutral pH. Due to the flexibility and intrinsic disorder of the protein, its high net positive charge and the tendency to form cysteine linked oligomers, Tat_{1-72} tends to form insoluble oligomers and this may be influenced by the chemical nature of its solvent. Finding conditions at neutral pH that are compatible with Tat_{1-72} is the goal. Presumably, the Tat/TAR/P-TEFb affinity is optimal at the neutral pH conditions of the eukaryotic cytosol of infected cells, therefore the need to mimic this pH *in vitro* is clear. Table 3.2 shows all of the conditions assayed for Tat_{1-72} solubility. Conditions of buffers, detergents, reducing agents, salt, the organic solvent TFE, and zinc were used to increase the solubility of Tat_{1-72}. The table also indicates if the protein was soluble in certain conditions and, where known, the secondary structure of the protein.

Table 3.2 Solubility assays for His-tagged Tat_{1-72}. Highlighted in pink is the Tat/TFE study over a temperature range of 5 °C – 80 °C.

	Buffer	pH	Detergent	Reducing Agent	NaCl	TFE	Zinc	Soluble	Structure
1	Acetate	4.2	—	—	—	—	—	Yes	Random Coil
2	Acetate	4.0	—	TCEP	—	—	—	Yes	Random Coil
3	HEPES	4.0	—	—	—	—	—	Yes	Random Coil
4	HEPES	5.8	—	—	—	—	—	Yes	—
5	HEPES	6.5	—	—	—	—	—	No	—
6	HEPES	4.0	—	TCEP	—	—	—	No	—
7	HEPES/Disodium Citrate	4.0	—	—	—	—	—	Yes	—
8	MES	4.0	—	—	—	—	—	Yes	Random Coil
9	MES	5.0	—	—	—	—	—	No	—
10	MES/Disodium Citrate	4.0	—	—	—	—	—	Yes	—
11	Citrate	5.0	—	—	—	—	—	No	—
12	Disodium Citrate	4.0	—	—	—	—	—	Yes	—
13	Citrate	5.4	—	—	Yes	—	—	Yes	β-sheet
14	Citrate	6.2	—	—	Yes	—	—	Yes	—
15	Acetate	4.1	SDS	—	—	—	—	Yes	α-helix
16	Acetate/HEPES	4.5	SDS	—	—	—	—	Yes	—
17	Acetate/HEPES	6.7	SDS	—	—	—	—	Yes	—
18	HEPES	4.0	CHAPS	—	—	—	—	Yes	—
19	HEPES	5.7	CHAPS	—	—	—	—	Yes	—
20	HEPES	3.9	DTAB	—	—	—	—	Yes	—
21	HEPES	5.0	DTAB	—	—	—	—	Yes	—
22	HEPES/Disodium Citrate	4.0	CHAPS	—	—	—	—	No	—
23	HEPES/Disodium Citrate	4.0	DTAB	—	—	—	—	No	—
24	MES	4.0	CHAPS	—	—	—	—	No	—
25	MES	4.0	DTAB	—	—	—	—	No	—
26	MES/Disodium Citrate	4.0	CHAPS	—	—	—	—	No	—

27	MES/Disodium Citrate	4.0	DTAB	—	—	—	—	No	—
28	Citrate	4.0	CHAPS	—	—	—	—	No	—
29	Citrate	4.0-5.3	DTAB	—	—	—	—	Yes	—
30	Disodium Citrate	4.0	CHAPS	—	—	—	—	No	—
31	Disodium Citrate	4.0	DTAB	—	—	—	—	No	—
32	Acetate/Mes	4.0-7.0	SDS	TCEP	—	—	—	Yes	α-helix
33	Acetate/Mes	4.0-6.4	LHGP	TCEP	—	—	—	—	—
34	Acetate	5.0	NDMB	TCEP	—	—	—	No	—
35	Acetate	4.6	DTAB	TCEP	—	—	—	No	—
36	Acetate	4.7	CHAPS	TCEP	—	—	—	No	—
37	Acetate	4.0	—	—	—	0-90 %	—	Yes	Random Coil/α-helical
38	Acetate	4.2	—	TCEP	—	20%	—	Yes	Random Coil
39	Acetate	4.3	—	—	—	30%	0-4 MEQ	Yes	α-helix
40	Acetate/MES	4.7-7.2	—	TCEP	—	30%	2 MEQ	Yes	α-helix
41	Acetate	4.0	—	—	—	0-30 %	—	Yes	Random Coil/α-helical

3.5.1 Buffers

The starting point for these studies is the well-established solubility of Tat_{1-72} in acetate buffer at pH 4 (see Table 3.2, first entry). Long-term stability of Tat_{1-72} at pH values approaching pH 7 in the absence of reducing agent was explored in MES, HEPES, and citrate. Tat_{1-72} precipitated in both MES and citrate when equilibrated for 24 hours at pH 5 (Table 3.2, entry 9 and 11).

Tat_{1-72} was soluble in HEPES at pH 4 for at least 24 hours, and upon titration to pH 5.8, showed some stability at this pH. Further titration to neutral pH resulted in a precipitate. The UV spectrum of Figure 3.4 shows evidence of aggregation by an increase in the absorption of the trough region (250 nm) and an increase in the absorbance above 300 nm, and a decrease in the peak absorbance at 280 nm. The decrease in UV absorbance at 280 nm suggests a loss of protein, which could have occurred as a small amount of Tat_{1-72} precipitated out of solution, or by adhering to the glass surface of the absorbance cuvette. Tat_{1-72} precipitated during the 24 hour equilibration period after the pH was raised to about 6.5 (data not shown, Table 3.2, entry 5). Tat_{1-72} was not stable in 20 mM HEPES at pH 4 in the presence of 10 mM TCEP over a period of 24 hours as indicated by protein precipitation (data not shown, Table 3.2, entry 6).

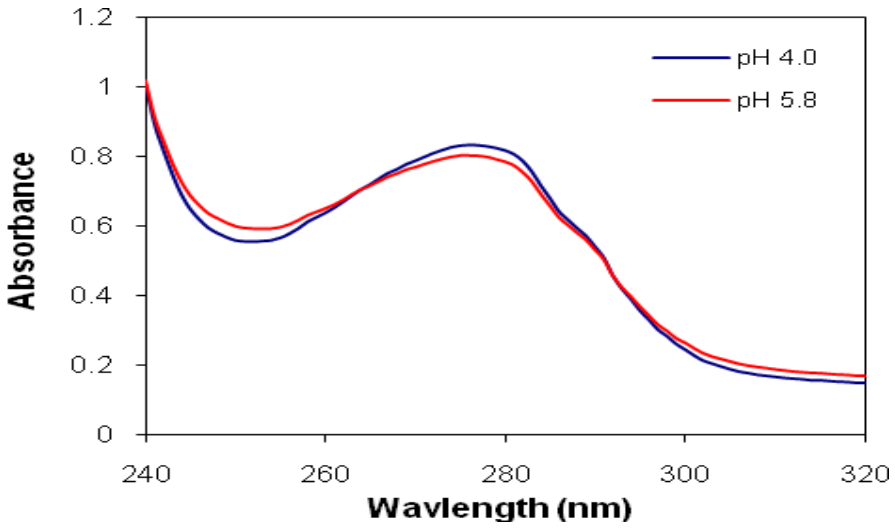

Figure 3.4 The absorbance spectra of 90 μM Tat_{1-72} in 20 mM HEPES at pH values of 4.0 and 5.8 collected after 24 hour equilibration at each pH.

The circular dichroism spectrum of 90 μM Tat_{1-72} in 20 mM HEPES at pH 4.0 is shown in Figure 3.5. The spectrum is dominated by a negative CD band at 200 nm that is characteristic

of a disordered random coil, with a weaker negative CD band in the region between 220 nm – 240 nm. Table 3.3 shows the deconvolution results. The secondary structure content of Tat_{1-72} in 20 mM HEPES at pH 4.0 is: 11.7 % helical, 19.6 % β-strand, 20.9 % turn, 47.6 % disordered. According to the deconvolution results, Tat_{1-72} is mainly disordered in HEPES at pH 4.0.

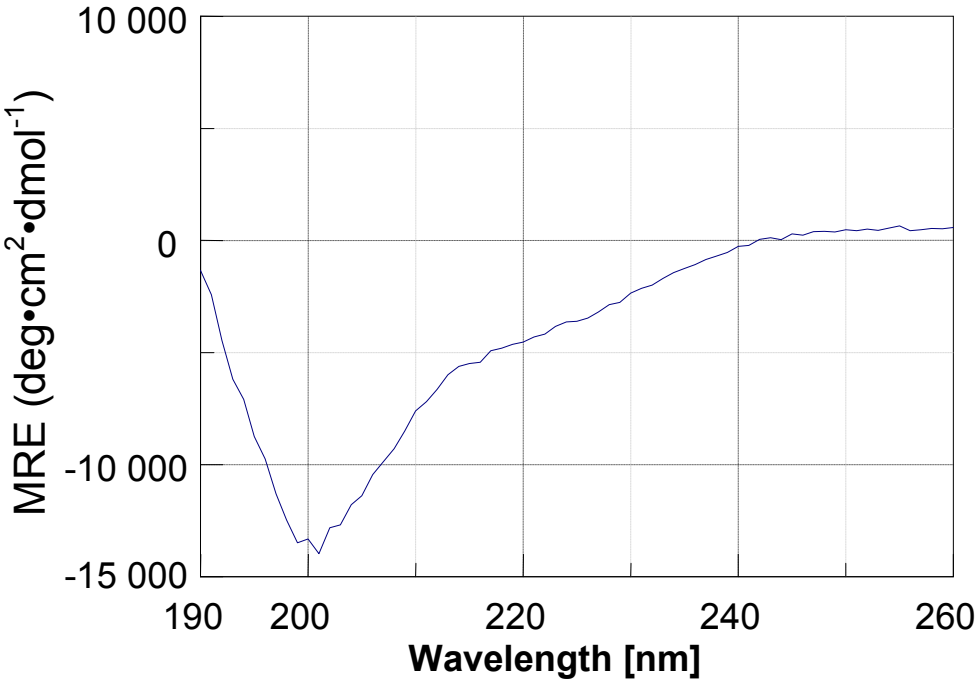

Figure 3.5 The CD spectrum of 90 μM Tat in 20 mM HEPES, pH 4.0.

Table 3.3 Secondary structure determination by deconvolution of the CD spectrum of 90 μM Tat_{1-72} in 20 mM HEPES, pH 4.

Algorithm	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Disordered	Total
Selcon 3	0.064	0.096	0.117	0.098	0.251	0.372	0.997
CONTINLL	0.035	0.087	0.123	0.111	0.266	0.377	0.999
CDSSTR	0.04	0.03	0.09	0.05	0.11	0.68	1
Average	0.046	0.071	0.110	0.086	0.209	0.476	0.999

3.5.2 Buffer and Detergent Combinations

Tat_{1-72} was screened for solubility in different buffer and detergent combinations in the absence of reducing agent. Tat_{1-72} showed good solubility in various detergents at about pH 4 (lines 15 – 36), Table 3.2). Tat_{1-72} was shown to have good solubility in each of HEPES, MES, and citrate at pH = 4.

Figure 3.6 shows the UV spectra of Tat_{1-72} in 20 mM HEPES and 20 mM of the zwitterionic CHAPS detergent at pH 4.0 and 5.7. This is above the CMC of CHAPS (CMC of CHAPS is 8mM)[263]. The initial UV absorbance spectrum at pH 4.0 differs little from the spectrum at pH 5.7. After a 24 hour equilibration at pH 5.7, there is a slight increase in absorbance between 245 nm – 270 nm indicating that some protein may have aggregated. The absorbance above 300 nm does not change between the pH values, but both do not decay to zero, indicating that there might be a Tat_{1-72} population that is a soluble multimer. A Tat_{1-72} precipitate formed during further titration to about pH 6.3.

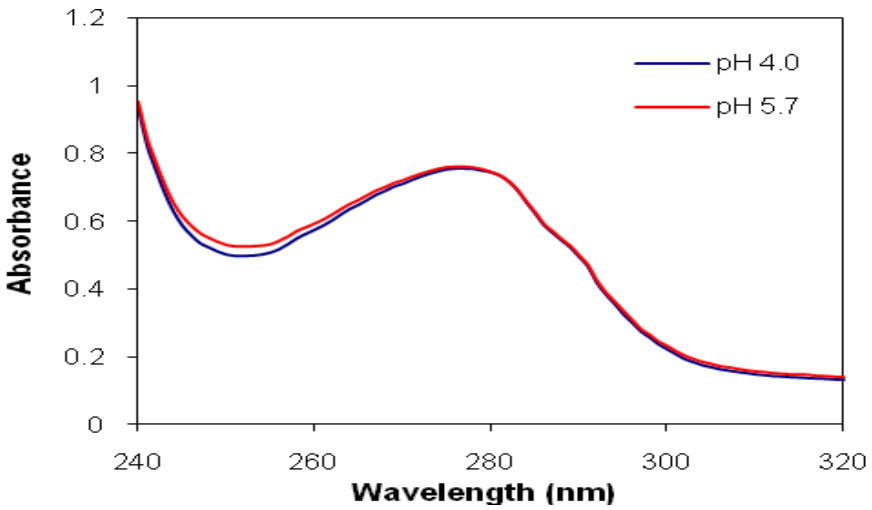

Figure 3.6 The 24 hour absorbance spectrum of 83 μM Tat_{1-72} in 20 mM HEPES and 20 mM CHAPS at pH values of 4.0 and 5.7.

Figure 3.7 shows the absorption spectra of Tat_{1-72} in 20 mM HEPES and 20 mM of the cationic DTAB detergent at pH 3.9 and 5.0. This is above the CMC of DTAB (CMC of DTAB is 14.1 mM)[264]. The spectra differ very little from each other, indicating that Tat_{1-72} was solubilized at both pH values for at least 24 hours. There is a slight increase in the absorption at 245 nm – 270 nm between pH 3.9 – pH 5.0 indicating that a small amount of Tat_{1-72} might have aggregated upon titration to pH 5.0. The absorbance above 300 nm increases slightly from pH 3.9 to 5.0, and does not decay to zero in both cases. The increase suggests protein aggregation upon titration to pH 5.0. The fact that both do not decay to zero above 300 nm suggests the existence of a small population of aggregated Tat_{1-72}. Further titration to pH 5.5 resulted formed a Tat_{1-72} precipitate.

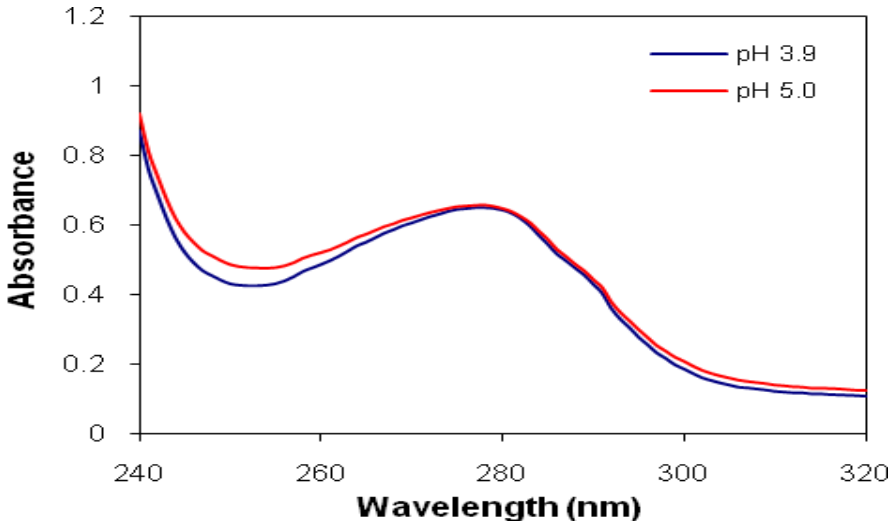

Figure 3.7 The 24 hour absorbance spectra of 75 μM Tat_{1-72} in 20 mM HEPES and 20 mM DTAB at pH values of 3.9 and 5.0.

Figure 3.8 shows the absorption spectra of Tat_{1-72} in 20 mM citrate and 20 mM of the cationic DTAB detergent at pH values 4.0 and 5.3, equilibrated for 24 hours at each pH. DTAB is above the CMC. When comparing the spectra at pH 4.0 and pH 5.3, there is an increase in absorbance between 245 nm – 270 nm, and an increase above 300 nm suggesting that a small amount of protein aggregated upon titration to pH 5.3. The absorption above 300 nm does not decay to zero, probably due to the existence of a small population of aggregated protein. Further titration to pH 6 caused Tat_{1-72} to precipitate (data not shown).

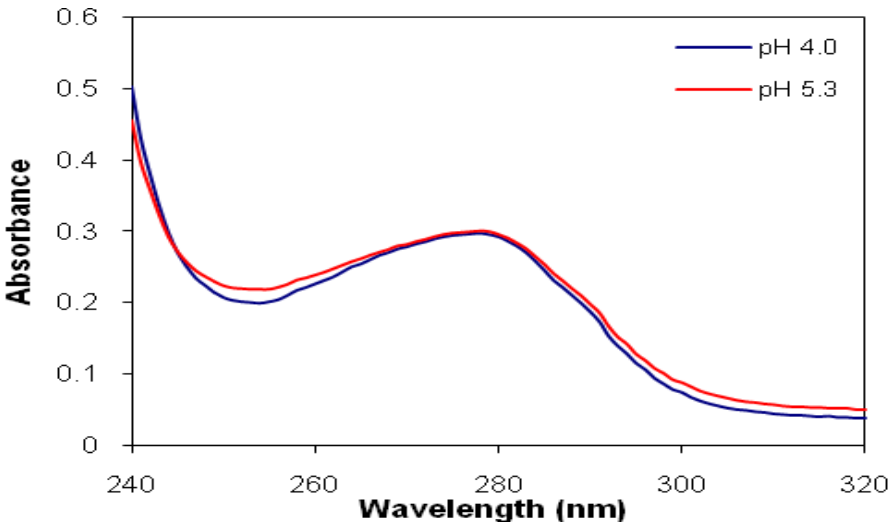

Figure 3.8 The 24 hour absorbance spectra of 33 μM Tat_{1-72} in 20 mM citrate and 20 mM DTAB at pH values 4.0 and 5.3.

Figure 3.9 shows the spectra of Tat_{1-72} in 10 mM acetate, 10 mM HEPES, and 20 mM SDS at pH values 4.47 to 6.67. Tat_{1-72} was equilibrated at each pH for 24 hours prior to the acquisition of spectra. The overlayed spectra at each pH increment are nearly identical to each other, suggesting that Tat_{1-72} is well solubilised in the anionic detergent SDS in the absence of reducing agent. Light scattering above 300 nm indicates that a small population of Tat_{1-72} is aggregated. Tat_{1-72} precipitated instantly after the pH of the solution was raised to about 7. The spectra for which Tat_{1-72} was not soluble at pH 4 for at least 24 hours are not shown (conditions shown in Table 3.2).

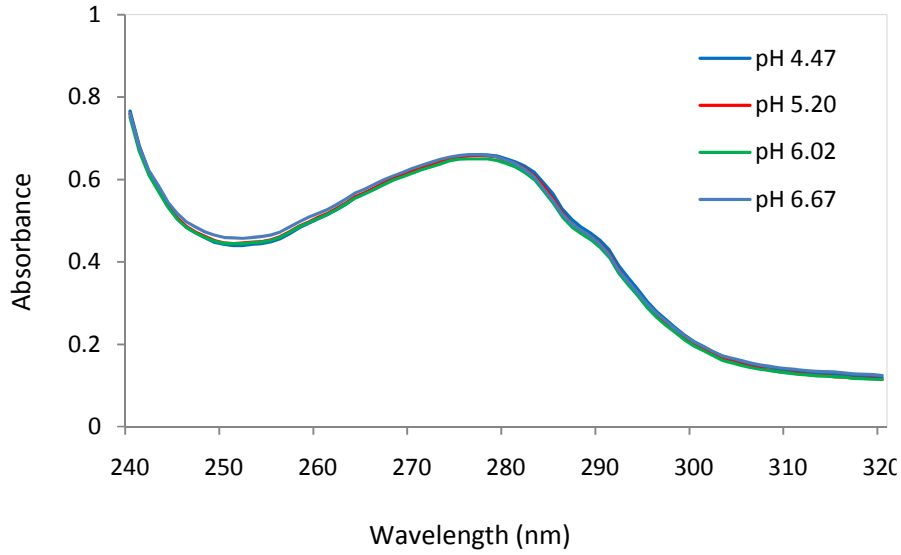

Figure 3.9 The 24 hour absorbance spectra of 72 μM Tat_{1-72} in 10 mM acetate, 10 mM HEPES, and 20 mM SDS at pH values 4.47 – 6.67.

Figure 3.10 shows that the CD spectrum of 14 μM Tat_{1-72} in 10 mM acetate at pH 4.05 resembles that of an intrinsically disordered protein as described earlier (Figure 3.3). The addition of SDS below the CMC causes a major change giving rise to a spectrum that has characterstic α-helical negative helical bands at 208 nm and 222 nm and a positive shoulder at 190 nm. Deconvolution shows that the helical content of Tat_{1-72} increases from 9 % to 23 %, while the content of all other secondary structures decreases slightly. β-strand decreases from 27 % to 22 %, turn content decreases from 24 % to 20 %, and disordered content decreases from 40 % to 45 % (Table 3.4).

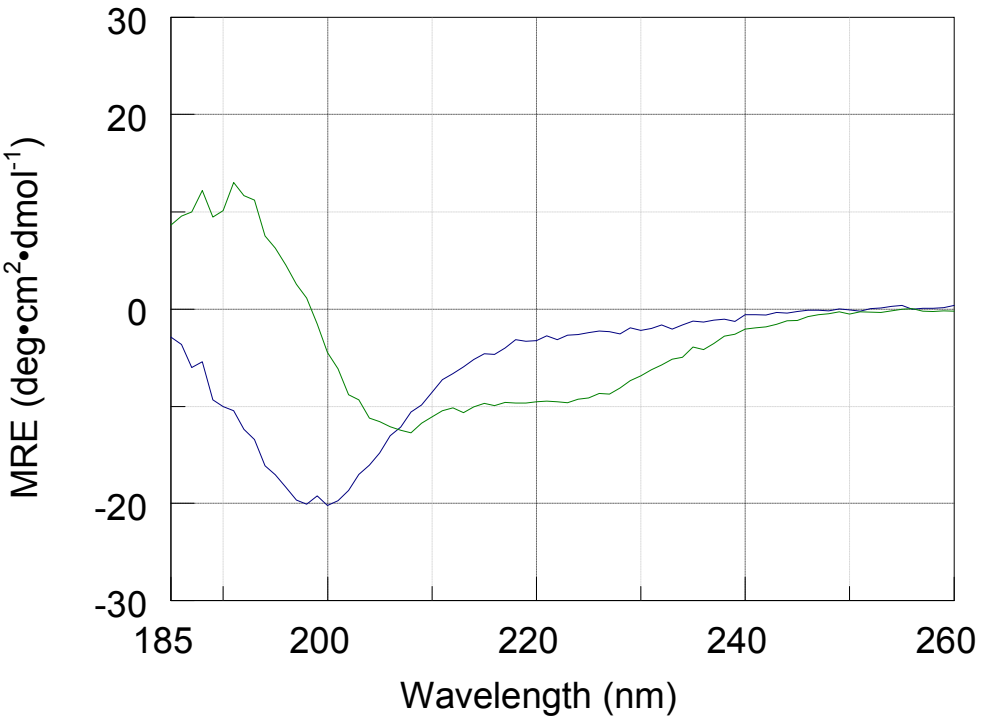

Figure 3.10 CD spectrum of 14 uM Tat_{1-72} in 10 mM acetate, pH 4.05 (blue) and 14 uM Tat_{1-72} in 10 mM acetate, pH 4.07, and 1.75 mM SDS (green).

Table 3.4 The average of the CD spectrum deconvolution of 14 μM Tat_{1-72} in the absence and presence of 1.75 mM SDS by the Selcon3, CONTINLL, and CDSSTR methods. See A5. of Appendix A for the complete set of deconvolution data.

	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
15 uM Tat_{1-72}	0.023	0.062	0.160	0.110	0.241	0.397	1.000
15 uM Tat_{1-72}, 1.75 mM SDS	0.124	0.104	0.132	0.084	0.197	0.348	0.988

To summarize the results thus far, in the absence of reducing agent, Tat_{1-72} is most stable in 20 mM HEPES, 20 mM CHAPS as pH 7 is approached, as judging from changes in UV absorption. The addition of detergent, regardless of ionic character, was shown to be compatible with Tat_{1-72} at pH = 4. Tat_{1-72} precipitation near neutral conditions presumably occurred on account of cysteine cross-linking. In the absence of reducing agent, detergents were unable to prevent self-association of Tat_{1-72}. Near pH 4, below the pKa of cysteine, most of the protein is monomerically reduced, and soluble in an array of conditions.

3.5.3 The Effects of Detergents with Reducing Agent

Figure 3.11 shows a series of spectra of Tat_{1-72} in 100 mM SDS, 20 mM sodium acetate, 20 mM MES, and 2.5 mM TCEP, between pH 4.03 – 7.03. SDS is above the CMC (CMC of SDS is 8 mM)[265]. The samples were incubated for 1 hour after elevation to each pH value. The spectra display the same UV absorbance characteristics as monomeric Tat_{1-72} at pH 4 in acetate buffer (Fig. 3.1). Furthermore, the characteristic trough at 250 nm, peak at 280 nm, and the shoulder at 290 nm were clear at all pH values between 4.03 and 7.03 and over a period of 24 hours at neutral pH. Twenty-four hour stability was deduced from a comparison of spectra acquired initially at pH 7.2, and after a period of 24 hours at the same pH (Figure 3.12). The absorbance does not decay to zero above 300 nm probably because of the existence of a small population of soluble, aggregated Tat_{1-72} protein. Note that the absorbance for 24 hour Tat_{1-72} in SDS does not change above 300 nm, but does increase slightly in the peak area, between 260 – 285 nm.

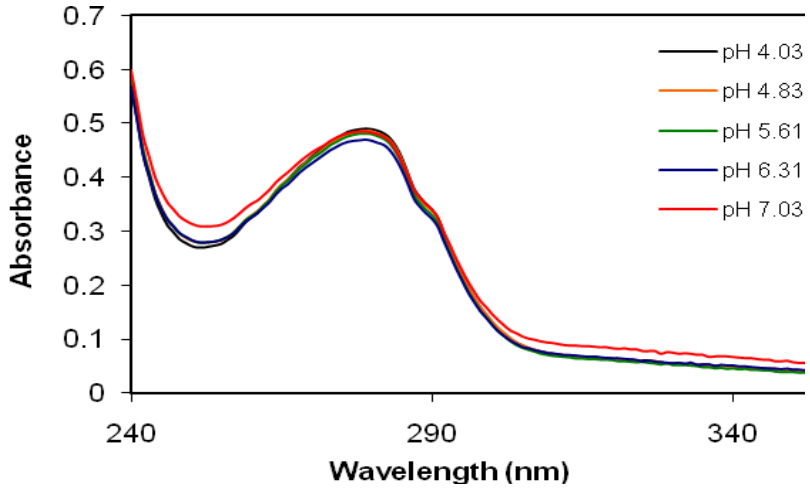

Figure 3.11 48 µM The absorption spectra of Tat_{1-72} solubilised by 100 mM SDS in 20 mM acetate, 20 mM MES, with 2.5 mM TCEP at various pH values. SDS is above the CMC.

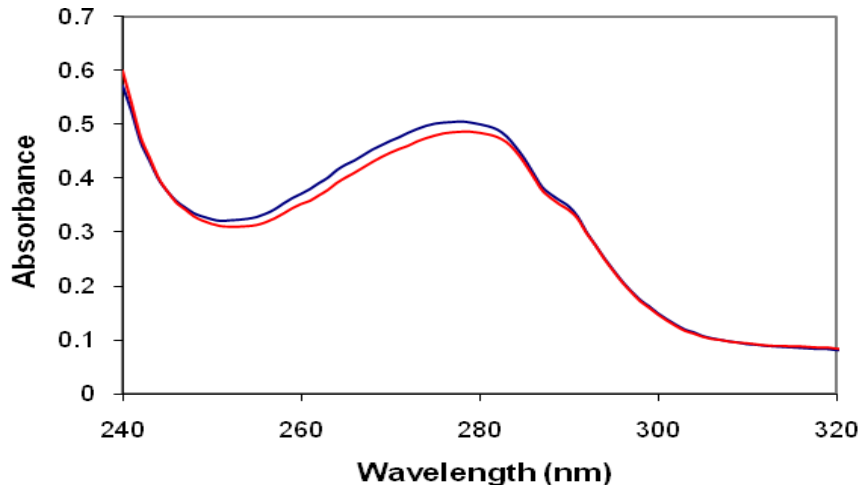

Figure 3.12 Absorption of twenty-four hour 48 µM Tat_{1-72} in 100 mM SDS, 2.5 mM TCEP, 50 mM acetate, and 50 mM MES, pH 7. Red: initial; Blue: 24 hours.

Tat_{1-72} solubilized in 10 mM LHGP, 20 mM sodium acetate, 20 mM MES, and 2.5 mM TCEP over a pH range of 4.04 – 6.37 showed the appearance of an unusual absorption band at

330 nm at pH 5.87 and pH 6.37 (Figure 3.13). The spectra of Tat_{1-72} in LHGP were identical between pH 4.03 – 4.88. Above pH 4.88, and continually until pH 6.37, the peak at 280 nm decreases and the absorption in the trough region at 250 nm increases. Tat_{1-72} absorption is still recognizable in the signal minimum at 250 nm, a peak intensity at 280 nm with a characteristic shoulder at 290 nm.

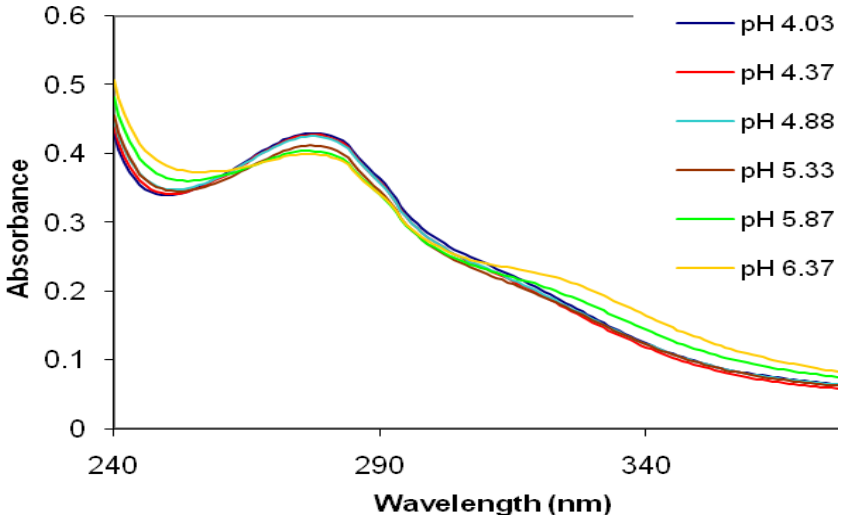

Figure 3.13 UV absorbance spectrum of 48 μM Tat_{1-72} in 10 mM LHGP, 20 mM sodium acetate, 20 mM MES, and 2.5 mM TCEP at various pH values.

Tat_{1-72} precipitated at pH 5 in the presence of the non-ionic NDBM detergent and TCEP, and at pH 4.7 in the presence of the zwitterionic CHAPS detergent and TCEP. This was observed by the appearance of a milky precipitate (data not shown). In the presence of the cationic detergent DTAB and TCEP, Tat_{1-72} precipitated at pH 4.6 (data not shown).

3.6 The Effect of Sodium Chloride

Figure 3.14 shows that Tat_{1-72} is partially soluble in 50 mM citrate, 100 mM sodium chloride at an initial pH value of 5.4 for at least 24 hours, and is stable for another 24 hours after titration to pH 6.2. Further pH elevation resulted in Tat_{1-72} precipitating. Figure 3.14 shows a significant increase in UV absorbance above 300 nm compared to Tat_{1-72} in 10 mM acetate, 10 mM TCEP, pH 4 (Figure 3.1) suggesting that there is some aggregated protein present. The circular dichroism spectrum of Tat_{1-72} in this condition shows a dramatic increase in β-sheet structure (Figure 3.15). One possibility is that the increase in β-sheet in this condition is due to interaction between the hydrophobic groups of Tat_{1-72}, creating a non-covalent self-association. After titration to pH 6.2 and equilibration for 24 hours, there was a further slight increase in UV absorbance between 240 nm – 250 nm, but no change in any other part of the UV spectrum, suggesting that an additional small proportion of protein aggregated at pH 6.2.

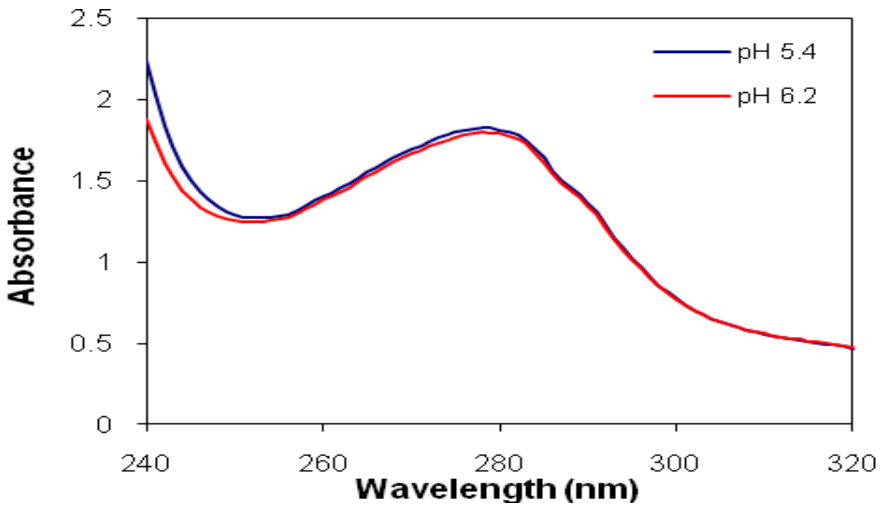

Figure 3.14 The absorbance spectra of 190 μM Tat_{1-72} in 50 mM citrate, 100 mM sodium chloride at pH values of 5.4 and 6.2.

The CD band shifts from 200 nm (Figure 3.3) to 220 nm (Figure 3.15) and the deconvolution of the CD spectrum shows that Tat_{1-72} secondary structure shifts from predominantly disordered to predominantly β-strand. The secondary structure composition is 7.3 % helical, 37.4 % β-strand, 22.2 % turn, and 32.8 % disordered (Table 3.5). A similar spectrum was obtained when this experiment was repeated twice with protein from the same Tat_{1-72} preparation. However, attempts to repeat this spectrum with protein from a different Tat_{1-72} preparation met with mixed success.

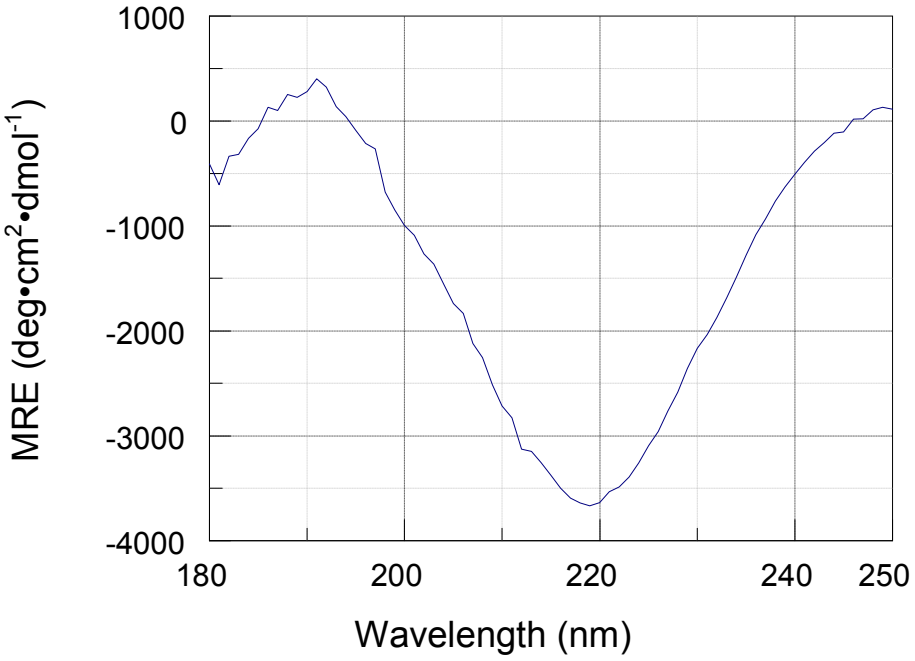

Figure 3.15 CD spectrum of 190 μM Tat_{1-72} in 50 mM citrate, 100 mM NaCl at pH 5.5.

Table 3.5 Secondary structure determination by deconvolution of the CD spectrum of 190 μM Tat_{1-72} in 50 mM citrate, 100 mM NaCl, pH 5.5.

Algorithm	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Disordered	Total
Selcon3	0.027	0.079	0.229	0.118	0.221	0.325	0.999
CONTINLL	0.022	0.051	0.254	0.131	0.214	0.329	1.001
CDSSTR	0	0.04	0.26	0.13	0.23	0.33	0.99
Average	0.016	0.057	0.248	0.126	0.222	0.328	0.997

When the experiment was repeated with 220 μM Tat_{1-72} from a different preparation in 50 mM citrate, 100 mM NaCl, pH 5.35 the CD spectrum shown in Figure 3.16 was obtained and the deconvolution results are shown in Table 3.6. In Figure 3.16, the CD signal between 190 nm – 205 nm is noise due to the intense absorbance of the sample. The negative CD at 206 nm suggests the existence of a disordered conformation. The region between 220 nm - 240 nm is unique because it shows a relatively weak, negative signal compared to the signal between 205 nm – 220 nm, but one that is considerably more intense than the signal of Tat with reducing agent in acetate at pH 4 (Figure 3.3). The secondary structure analysis suggests the protein contains mostly a mixture of β-strand, turns and disordered structure (Table 3.5 and 3.6).

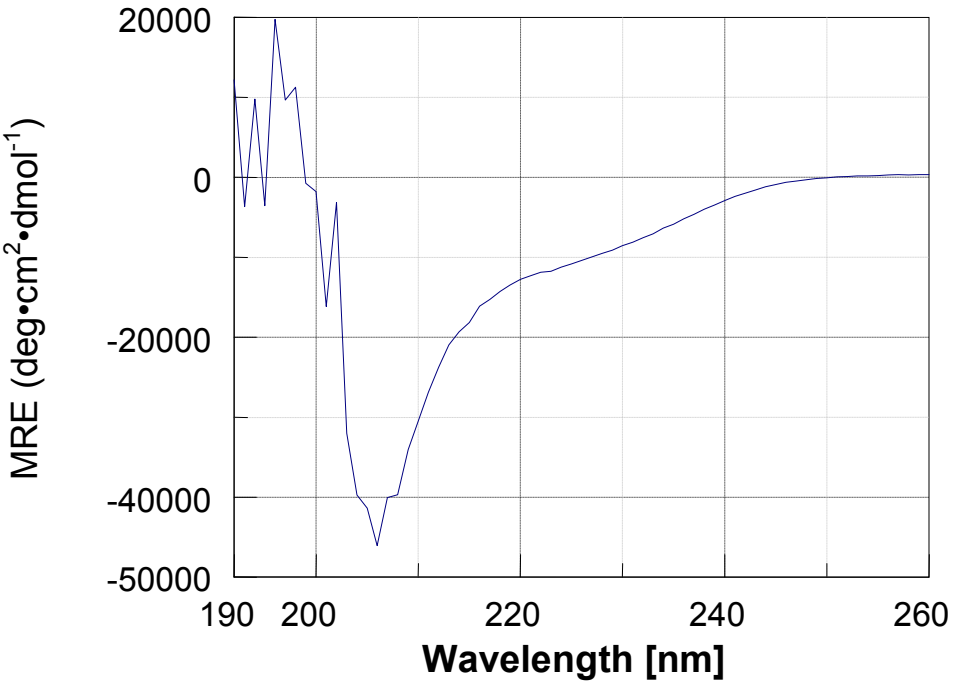

Figure 3.16 CD spectrum of 220 μM Tat_{1-72} in 50 mM citrate, 100 mM NaCl at pH 5.35.

Table 3.6 Secondary structure determination by deconvolution of the CD spectrum of 220 μM Tat_{1-72} in 50 mM disodium citrate, 100 mM NaCl, pH 5.35.

Algorithm	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	-0.026	0.034	0.265	0.071	0.256	0.342	0.942
Provencher & Glockner	0.032	0.067	0.256	0.115	0.22	0.31	1
CDSSTR	0.04	0.1	0.23	0.1	0.25	0.28	1
Average	0.015	0.067	0.250	0.095	0.242	0.311	0.981

The deconvolution data of 220 μM Tat_{1-72} in 50 mM citrate, 100 mM NaCl pH 5.35 shown in Table 3.6 is very similar to the deconvolution of 190 μM Tat in the same conditions at pH 5.55 (Table 3.5). The deconvolution averages for each secondary structure motif is (in brackets, the comparison to the results from Table 3.5): 8.2 % helical (7.3 %), 34.5 % β-sheet (37.5 %), 24.2 % turn (22.2 %), and 31.1 % disordered (32.8 %).

3.7 The Effects of TFE

The presence of the organic solvent TFE in aqueous mixtures has been shown to induce a helical structure in disordered fragments of proteins that adopt helical folds in native conformations, along with model peptides designed for α-helical propensity[211]. Studying the effects of TFE on the His-tagged Tat_{1-72} protein may help identify regions with α-helical propensity in the disordered state, or the presence of inducible α-MoRFs. Regions of the Tat_{1-72} protein that adopt

structure in the presence of TFE may be the same regions that upon favourable ligand interactions fold to form an α-helix *in vivo*.

In all Tat_{1-72}/TFE studies, an aqueous protein solution at pH 4 consisting of mainly monomeric Tat_{1-72} was prepared and then filtered through a 0.2 μm filter and centrifuged. No precipitation was observed upon dilution of aqueous Tat_{1-72} with TFE. Tat_{1-72} is soluble in aqueous/TFE mixtures buffered with acetate at pH 4.05 (Figure 3.17).

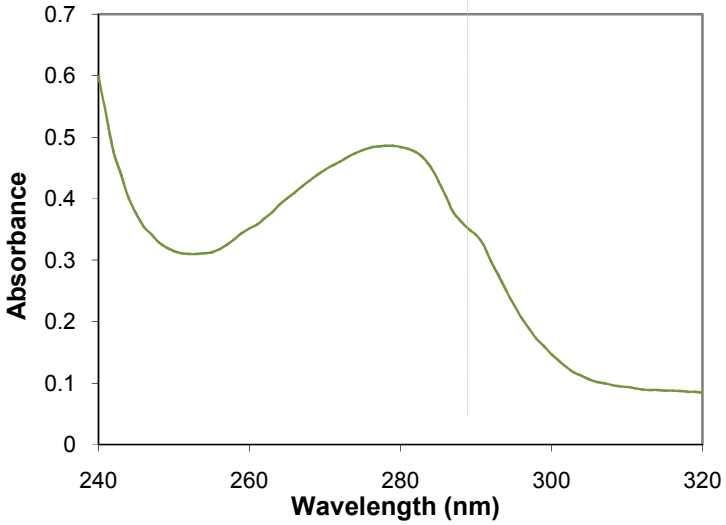

Figure 3.17 The absorption spectrum of 54 μM Tat in 10 mM acetate, pH 4.05, 30 % TFE.

In the absence of TFE, Tat_{1-72} showed the characteristic negative CD intensity at 200 nm of a disordered protein (Figure 3.3). As the TFE concentration is increased, the band at 200 nm decreases in intensity, and shifts to lower wavelength. A new band forms at 208 nm above 20% TFE accompanied by a broad shoulder at 220 nm and a positive band at 190 nm (Figure 3.18). This line-shape is representative of an α-helical signal. The CD signal at 222 nm is characteristic of proteins composed mainly of α-helices and has often been used to monitor changes in α-

helical content[266]. The apparent pH of the solution increased with the addition of TFE and began to plateau at 60% TFE at pH 4.37 (Table 3.7, Figure 3.19).

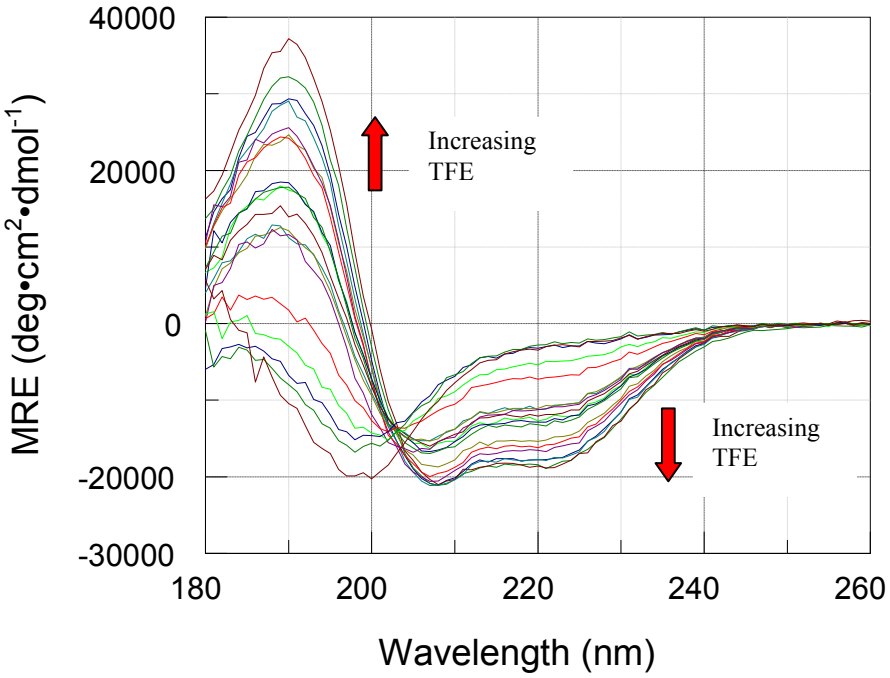

Figure 3.18 An overlay of CD spectra of Tat_{1-72} in 10 mM acetate at an intitial of pH 4, diluted to final TFE concentrations of 0% - 90% TFE. The final concentration of Tat_{1-72} is 100 µM except for the 85% TFE and 90% TFE solutions (83 µM and 55 µM Tat_{1-72}, respectively). Signal intensity in the 85% and 90% TFE Tat_{1-72} solutions were corrected by a factor of 1.176 and 1.818, respectively. Solution apparent pH varies with increasing addition of TFE as described in the text and shown in Figure 3.19.

Table 3.7 Apparent pH's of Tat_{1-72}/TFE solutions upon dilution of aqueous Tat_{1-72} solutions at an initial pH of 4.

% TFE	0	5	10	15	20	25	30	35	40	45
pH	3.98	3.98	4.02	4.06	4.11	4.13	4.17	4.19	4.24	4.29
% TFE	50	55	60	65	70	75	80	85	90	
pH	4.32	4.35	4.37	4.37	4.38	4.39	4.39	4.4	4.4	

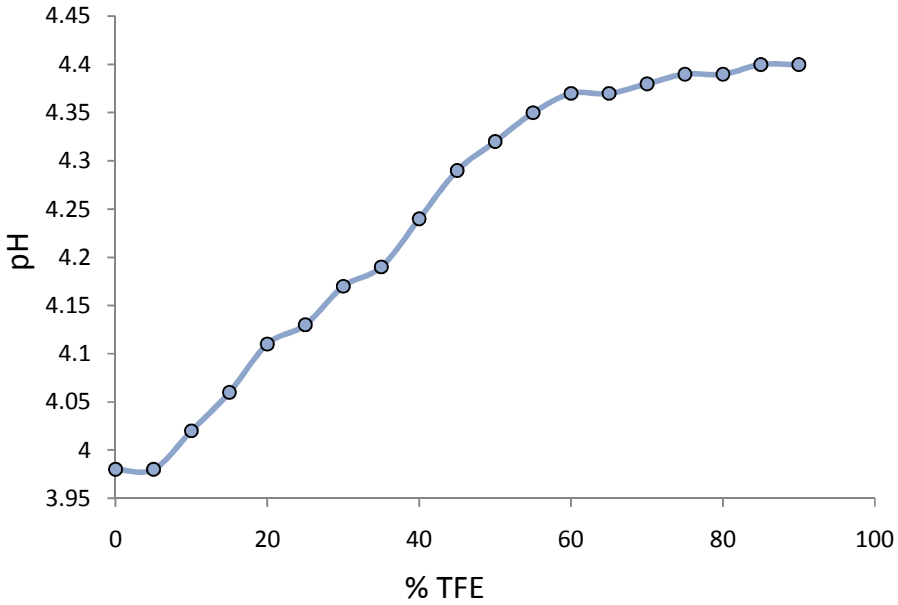

Figure 3.19 The change in apparent pH with the addition of TFE to 10 mM acetate at an initial pH of 4. The closed circles are the measured pH values at increasing 5 % increments of TFE, the lines between the measured values are interpolated.

The CD spectra of 100 μM Tat_{1-72} in 0%-90% TFE shown in Figure 3.18 contain an iso-dichroic point at 203 nm, suggestive of a two-state transition. In this case it appears that incremented additions of TFE favour the helical conformation in the random coil/α-helix equilibrium. Proteins that contain high helical content display negative bands at 208 nm and 222 nm, along with a positive band at 190 nm[266]. By tracking the changes in ellipticity at these wavelengths, the changes in helical content as a function of TFE concentrations can be observed. Figures 3.20 and 3.21 show the dependency of ellipticity at 190 nm and 222 nm, respectively. Table 3.8 shows the changes in MRE as a function of TFE concentration for 0 % - 25 % TFE and

30 % - 90 % TFE. The ellipticities at 190 nm and 222 nm appear to be most sensitive to TFE in the 0 % - 25 % range, as the change in ellipticity versus concentration of TFE is greatest there.

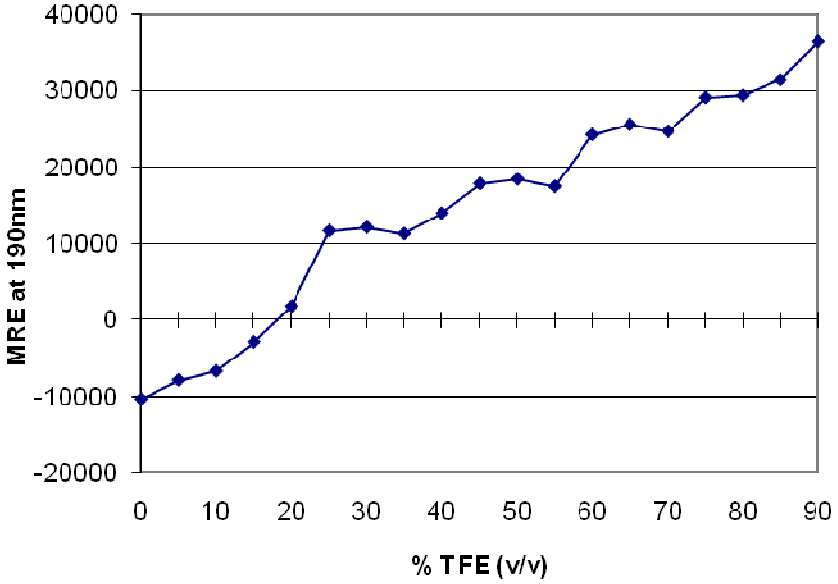

Figure 3.20 The mean residue ellipticity at 190 nm of 100 μM Tat_{1-72} with increasing TFE.

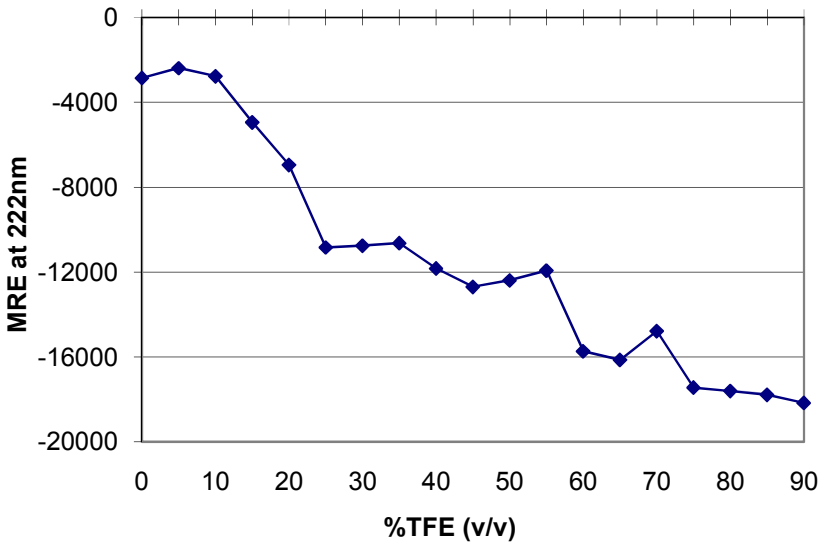

Figure 3.21 The mean residue ellipticity at 222 nm of 100 μM Tat_{1-72} with increasing TFE.

Above 25 % TFE, Tat_{1-72} is less sensitive to TFE, although the protein helical content increases with higher concentrations of TFE. Figure 3.22 is a graphical representation of the secondary

structure changes by spectral deconvolution of Tat_{1-72} in 0 % - 90 % TFE. The deconvolution shows that with the addition of TFE, disordered content of Tat_{1-72} decreases, with the concomitant increase in helicity. From the deconvolutions there is little effect of TFE on the β-strand content, and minor effects on the turn content. The deconvolution data suggest that the disordered regions of Tat_{1-72} are transitioning to a helical conformation, with little effect on the other segments. At 90 % TFE about 65 % of the protein is in a helical conformation.

Table 3.8 MRE changes as a function of changes in TFE.

	190 nm	222 nm
0% - 25% TFE	881	-319
25% - 90% TFE	381	-113

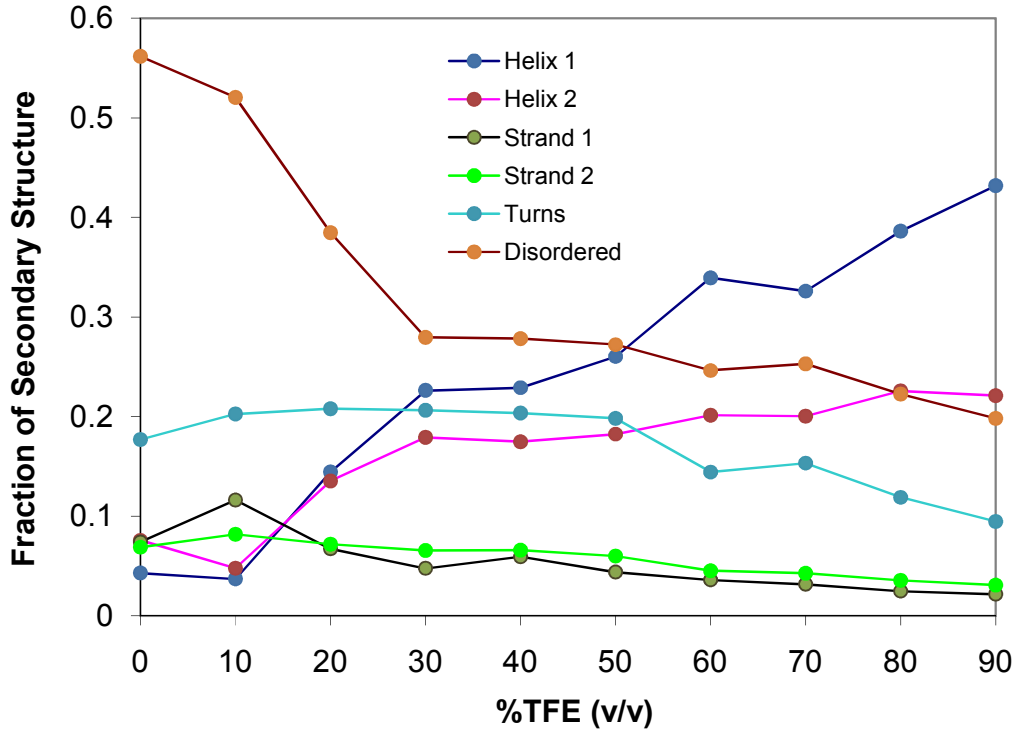

Figure 3.22 Average secondary structure deconvolution of Tat_{1-72} as a function of solution TFE composition. The results are an average from the CONTINLL method, Selcon3 method, and the CDSSTR method deconvolution programs (See A1. of Appendix A for complete set of

deconvolution data). Spectral deconvolutions were processed for every 10% increment of TFE volume.

3.7.1 The Effect of Temperature on Tat_{1-72} in 20% Trifluoroethanol

The conformational stability of Tat_{1-72} in 20 % TFE was probed. Figure 3.23 shows the overlay of CD spectra of Tat_{1-72} in 20 % TFE over a temperature range of 5 °C – 80 °C. In this concentration of TFE at 25 °C, Tat_{1-72} protein was determined to be 39 % disordered, 20 % turns, 14% β-strand, and 27 % helical (Table 3.9). In general, secondary structural characteristics induced by 20 % TFE are resistant to changes in thermal energy of the solution as deduced by the lack of change in CD signal over the range of temperatures from 5 °C – 80 °C. A slight increase in the disordered composition and decrease in the helical content of Tat_{1-72} in 20 % TFE is observed at 5°C, shown by a small shift in the negative band at 205 nm to lower wavelengths. Table 3.9 shows the secondary structure deconvolutions of Tat_{1-72} at 5 °C, 25 °C and 80 °C.

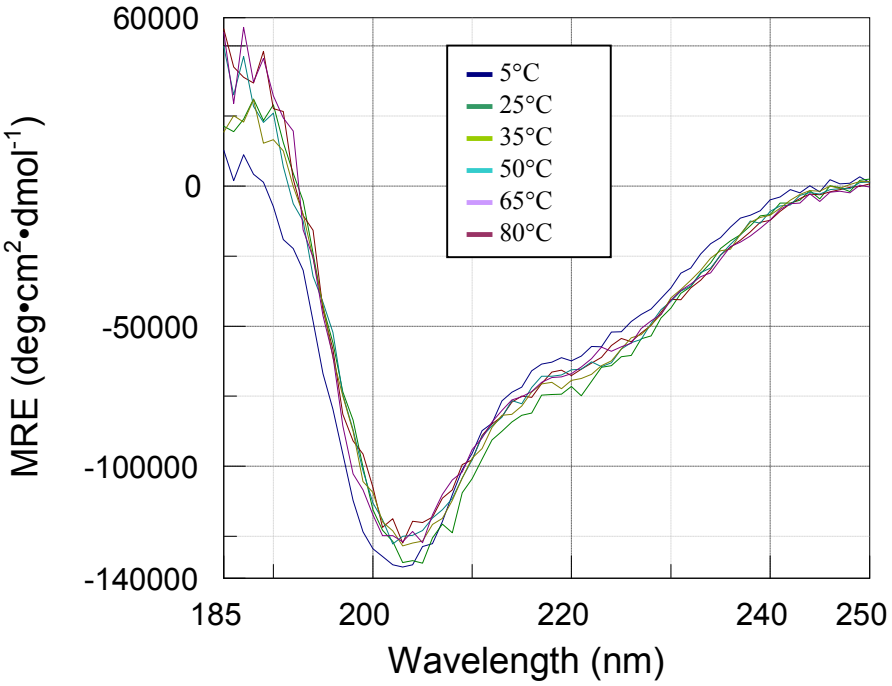

Figure 3.23 CD spectra of 100 µM Tat_{1-72} in 20% TFE, 8 mM TCEP, 8 mM acetate, apparent pH 4.2 from 50 °C – 80 °C.

Table 3.9 Secondary structure determination by deconvolution of the CD spectrum of 100 µM Tat_{1-72} at 25°C in 20 % TFE, 8 mM TCEP, apparent pH 4.2. The average of secondary structures by the Selcon3, CONTINLL, and CDSSTR methods are reported. See A2. of Appendix A for the complete data set.

	Helix1	Helix2	Strand1	Strand2	Turns	Disordered	Total
5 °C	0.075	0.081	0.103	0.079	0.192	0.472	1.002
25 °C	0.144	0.135	0.067	0.072	0.208	0.385	1.012
80 °C	0.094	0.095	0.11	0.083	0.191	0.429	1.002

3.7.2 Zinc Titration of Tat_{1-72} in 30% Trifluorethanol

It has been shown by plasma desorption mass spectrometry that the Tat_{1-72} peptide corresponding to the cysteine rich region is capable of binding two cadmium ions[116]. Although there are no biological roles for cadmium, it shares similar chemical characteristics to zinc, making Cd^{+2} suitable for protein-metal ion interaction studies because of its optical properties. Furthermore, in many instances TCEP precipitates Tat_{1-72} at elevated pH and zinc was explored as a possible substitute for preventing disulphide formation. A zinc titration of Tat_{1-72} in TFE was done to elucidate any zinc-dependent changes in secondary structure. Changes in CD signal upon addition of successive molar equivalents of zinc(II) are a result of zinc-induced conformational changes of the Tat_{1-72} protein, and are suggestive of an interaction between Tat_{1-72} and zinc.

Figure 3.24 shows the CD spectra of Tat_{1-72} in 0 – 4 MEQ of Zn^{+2} at apparent pH 4.28. The intensity of the signals at 190 nm, 208 nm, and 222 nm decrease slightly with the addition of zinc(II) with the greatest decreases occuring with the addition of 0.5 MEQ of Zn^{+2}. From the spectral deconvolution data shown in Table 3.10, there is minimal overall change in structure with the addition of Zn(II). The greatest change in secondary structure occurred with the initial addition of 0.5 MEQ Zn^{+2}. The mean residue ellipticities of 140 μM Tat_{1-72} in 10 mM acetate, pH 4.28, diluted to 30% TFE (v/v), in the presence of increasing concentrations of zinc(II) is observed in Figure 3.25. From Tabled 3.10 it appears that the addition of 0.5 MEQ of Zn^{2+} reduces the amount of disordered structure and increases the amount of turn structure, a result that might be expected if Zn^{2+} is binding to Tat_{1-72}.

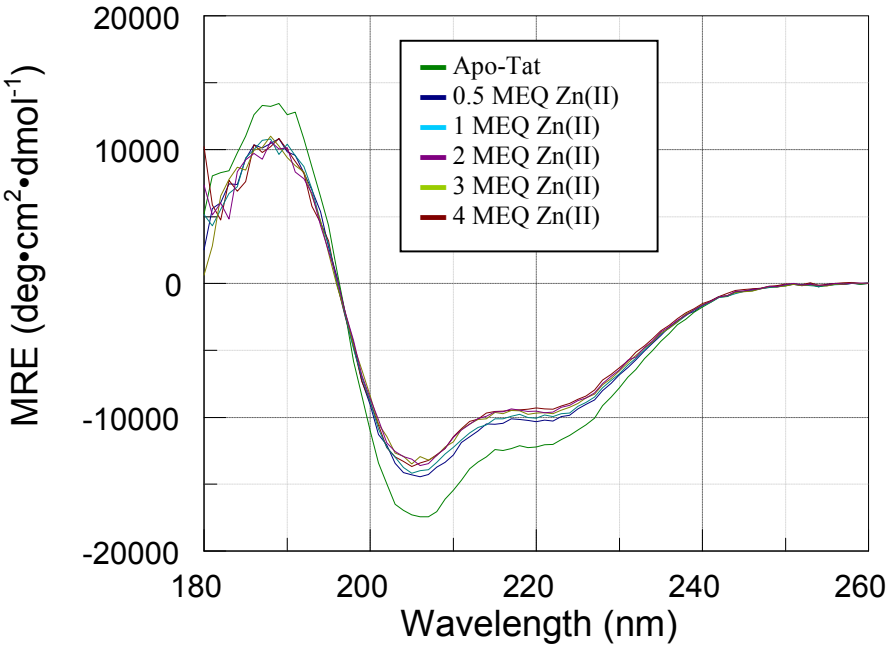

Figure 3.24 CD Spectra of 140 µM Tat_{1-72} in 10 mM acetate, apparent pH 4.28, diluted to 30% TFE (v/v), in the presence of increasing concentrations of zinc(II). The CD signal was corrected for the dilution with increasing volumes of zinc(II) acetate.

Table 3.10 Secondary Structure determination by the average spectral deconvolution using Selcon3, CONTINLL, and CDSSTR methods of 140 µM Tat_{1-72} in 30% TFE, 7 mM acetate, apparent pH 4.28, with increasing amounts of zinc(II) acetate. See A3. of the Appendix A for the complete set of deconvolution data.

	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Disordered	Total
Apo-Tat_{1-72}	0.166	0.141	0.067	0.065	0.197	0.377	1.000
0.5 MEQ Zn^{2+}	0.164	0.141	0.071	0.068	0.206	0.365	1.008
1 MEQ Zn^{2+}	0.165	0.141	0.069	0.067	0.201	0.371	1.004
2 MEQ Zn^{2+}	0.165	0.141	0.069	0.067	0.201	0.371	1.004
3 MEQ Zn^{2+}	0.165	0.141	0.070	0.067	0.203	0.369	1.005
4 MEQ Zn^{2+}	0.165	0.141	0.069	0.067	0.202	0.370	1.004

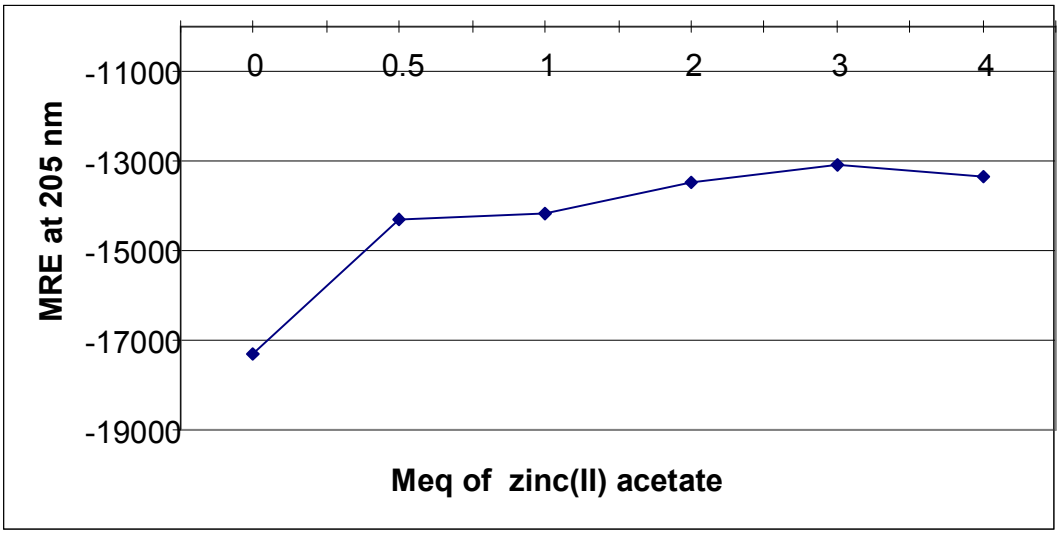

Figure 3.25 The mean residue ellipticity at 205 nm of 140 μM Tat_{1-72} in 10 mM acetate, apparent pH 4.28, diluted to 30% TFE (v/v), in the presence of 0.5, 1, 2, 3, and 4 MEQ's of zinc (II). The ellipticity changes maximally with the addition of 0.5 MEQ of zinc (II). The ellipticity does not change with a successive titration to 1 MEQ's of zinc (II) following which smaller changes in ellipticity occur.

3.7.3 pH Titration of Tat_{1-72} in 30% Trifluoroethanol with 2 Molar Equivalents of Zinc(II)

Tat_{1-72} was not able to remain in a 30 % TFE solution when neutral pH was approached. In the absence and presence of the reducing agent TCEP, Tat_{1-72} precipitated at pH values above 5 (data not shown). To test the hypothesis that the cysteine residues of Tat_{1-72} can remain in their reduced state in the presence of zinc near physiological pH conditions, a 100 μM Tat_{1-72} in 30% TFE, and 2 MEQ Zn^{+2} solution buffered with 7 mM acetate and 7 mM MES was titrated with 1-3 μl of aqueous NH_4OH from pH 4.69 to pH 7.2 (Figure 3.26). Generally, with the increase in pH, the CD signal lineshape at 206 nm and 222 nm decreases in intensity, accompanied by a slight decrease at 190 nm. The deconvolution data suggest minor structural changes upon

increase in pH. Table 3.11 shows that from pH 4.69 to pH 7.2 the helical content decreases approximately 6% from 41% to 35%, whereas the disordered content increases approximately 6% from 49% to 55%. β-strand secondary structure changes about 1 % throughout the pH titration, and the turn content changes less than 1 %. A decrease in ellipticity of the Tat_{1-72} protein in the presence of TFE shows that the protein becomes slightly more disordered as the pH approaches neutrality.

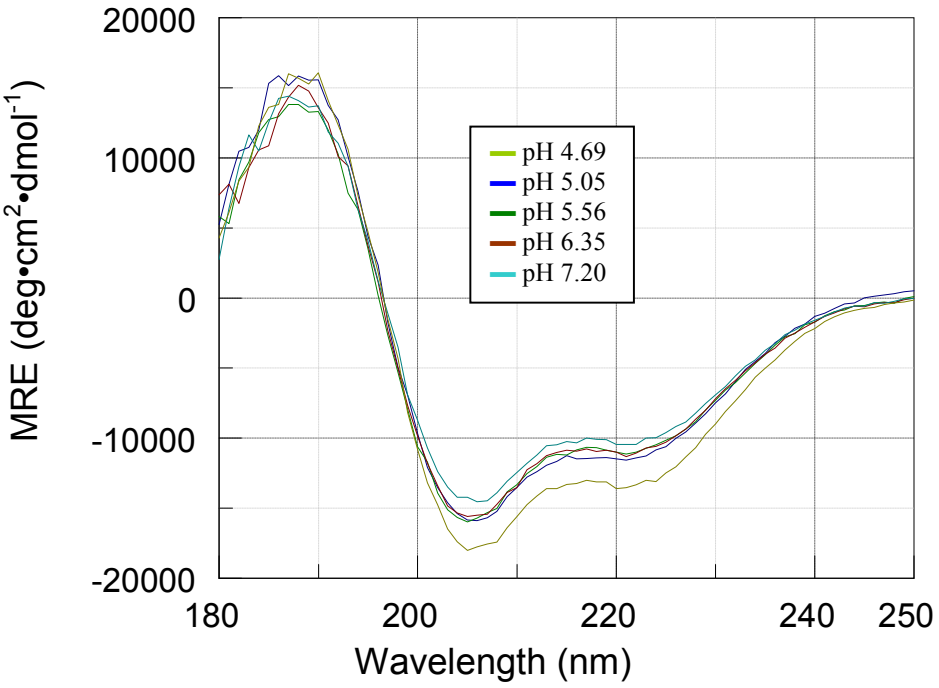

Figure 3.26 CD Spectra of 100 μM Tat_{1-72} in the presence of 7 mM acetate, 7 mM MES, 30% TFE and 2 MEQ's of zinc(II) over a range of apparent pH from 4.69 to 7.20.

Table 3.11 Secondary structure spectral deconvolution of the pH titration of 100 μM Tat_{1-72} in 30% TFE and 2 MEQ's of zinc(II) acetate. See A4. of Appendix A for the complete set of deconvolution data.

	Helix1	Helix2	Strand1	Strand2	Turns	Disordered	Total
pH 4.69	0.222	0.183	0.040	0.058	0.209	0.295	1.006
pH 5.05	0.210	0.169	0.035	0.052	0.193	0.342	1.002
pH 5.56	0.201	0.168	0.039	0.054	0.202	0.344	1.008
pH 6.35	0.192	0.162	0.054	0.062	0.197	0.335	1.001
pH 7.20	0.194	0.156	0.052	0.055	0.197	0.351	1.005

The absorption spectra shown in Figure 3.27 suggest that Tat_{1-72} can remain soluble in a solution of neutral pH in 30 % TFE in the presence of 2 MEQ's of Zn^{2+}. Figure 3.27 shows absorption spectra of Tat_{1-72} in 30 % TFE at pH 7.2 (red) and the same sample after 72 hours at apparent pH 6.9 (green). The interaction of zinc with the cysteine region appears to prevent oxidation, or at least reduces the rate of oxidation compared to Tat_{1-72} in the absence of TFE and Zn^{2+}. Both spectra of 110 μM Tat_{1-72} in 30% TFE at pH 7.2 and 6.9 show little evidence of light scattering suggesting that the Tat_{1-72} protein is soluble, monomeric, and reduced. A decrease in absorption between the initial and 72 hour scans did occur, but was not accompanied by an increase in light scattering, suggesting that the concentration of soluble protein decreased between trials, but not as a result of aggregation, but possibly a loss in protein. The sample may have been slightly diluted or lost as a result of sticking to the quartz cuvette. Note that the net charge of Tat_{1-72} decreases from +25 to +15 upon increasing the pH from 4 to 7 (see Table 2.1 of Materials and Methods section).

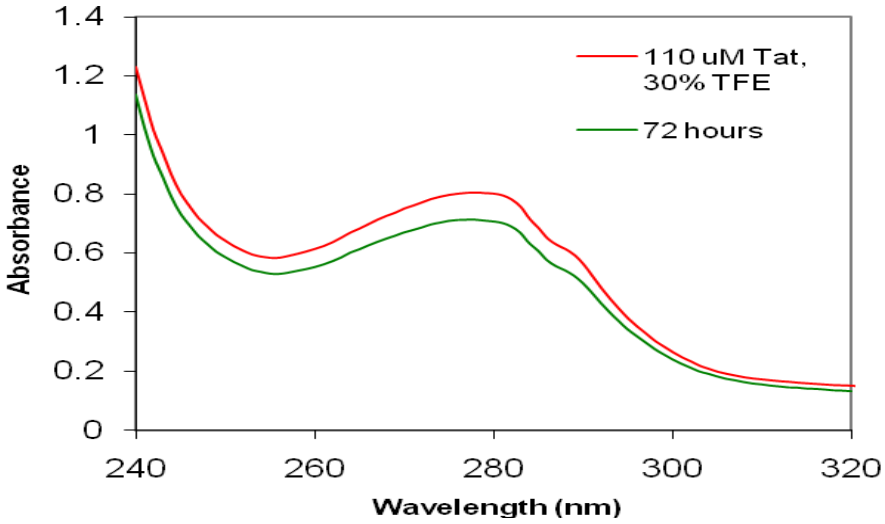

Figure 3.27 Absorption spectra of 110 µM Tat_{1-72} in 30% TFE, 7 mM acetate, 7 mM MES, 2 MEQ's Zn(II), apparent pH 7.2, and the same sample at pH 6.9 after 72 hours.

3.8 NMR Spectroscopy of Tat_{1-72}

Dissolution of ^{15}N and/or ^{13}C enriched Tat_{1-72} in 10 mM TCEP and 10 mM acetate, pH 4 resulted in an aqueous solution that was visibly precipitate-free. After repeated centrifugation, no pellet was formed. The sample was precipitate free before and after NMR acquisition. The ^{1}H-^{15}N HSQC spectrum of ^{15}N isotopically enriched His-tagged, exon-1 Tat_{1-72} is shown in Figures 3.28 and 3.29 where 72 of the 83 non-proline and non-N-terminal residues of the fusion protein are assigned in good agreement with previously published assignments in similar conditions at pH 4 by Dr. Shajeen Shojania[20]. The ^{1}H-^{15}N HSQC spectra of Tat_{1-72} in Figures 3.28 and 3.29 show amide backbone resonances falling within 1 ppm in the direct dimension and 20 ppm in the indirect dimension. The backbone resonances of the ^{1}H-^{15}N HSQC resonances show typical groupings of glycine residues and the serine/threonine residues that have been observed for proteins in the presence of high concentrations of denaturants[258]. There is residue

overlap in some regions of the ^{1}H-^{15}N HSQC spectrum of Tat_{1-72}, mainly corresponding to residues of the arginine-rich region.

Several resonances including Gly-35, Ser-36, and the indole amine resonance of the tryptophan-31 side chain show multiple resonances in the ^{1}H-^{15}N HSQC spectrum of Tat_{1-72}. The three residues are situated in the proline-rich region. Trp-31 is flanked by Pro-30 and Pro-34, and Gly-35 and Ser-36 are flanked by Pro-34 and Pro-38, respectively. The conformer populations of Gly-35 and Ser-36 are 92% and 8% for both conformers of both residues. From the integration of the Trp side chain cross-peaks, the conformer populations are 89%, 6%, and 6%. It is likely that the multiple resonances appear as a result of *trans-cis* isomerisation of proline residues[267].

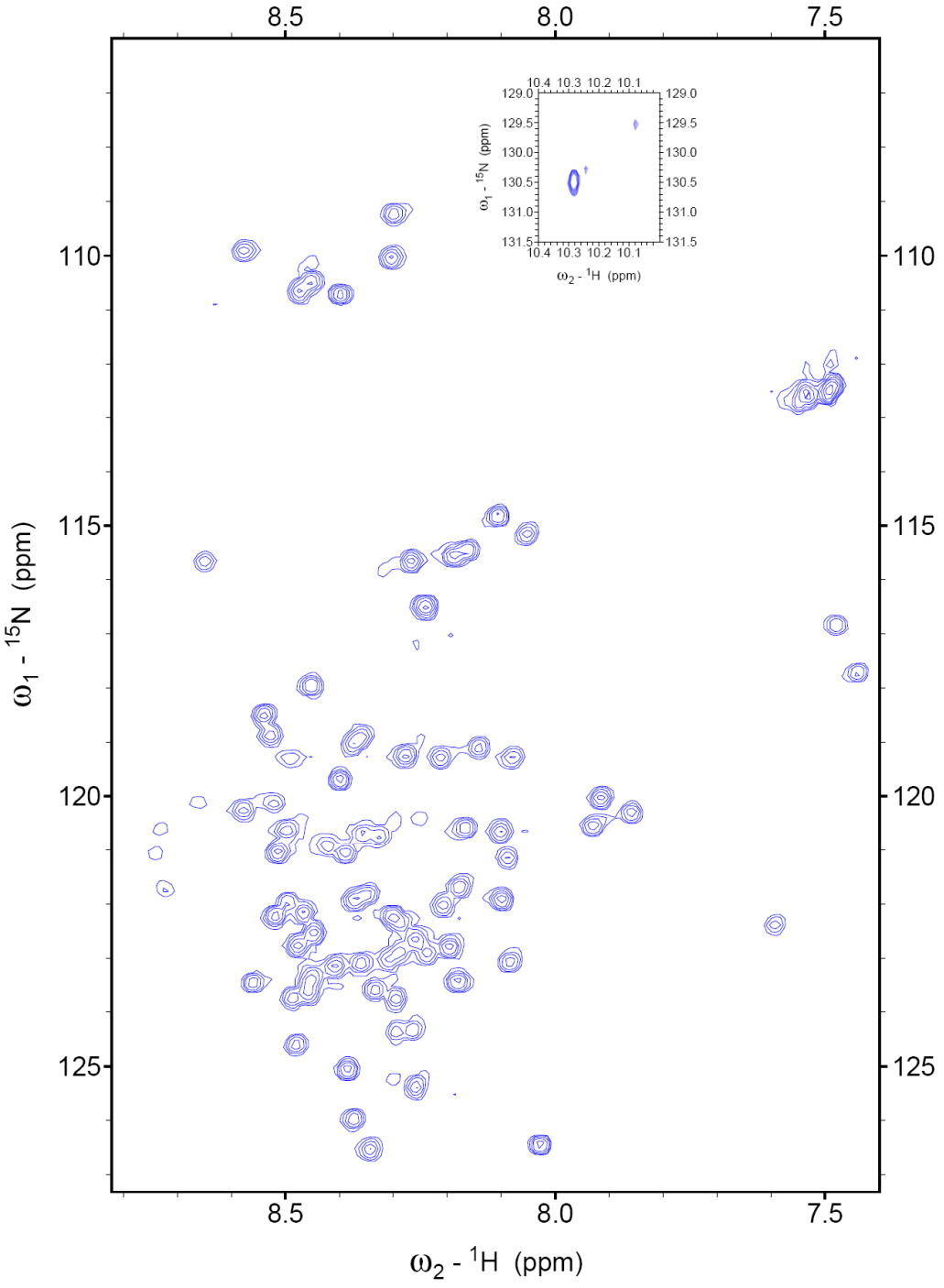

Figure 3.28 ^{1}H-^{15}N HSQC spectrum at 298 K of ^{15}N enriched, His-tagged $Tat_{1\text{-}72}$ in 10 mM acetate at pH 4 in the presence of 10 mM TCEP reducing agent. The inset box shows the indole amine resonances of Trp-31 (Figure 3.29 shows the assignments of this spectrum).

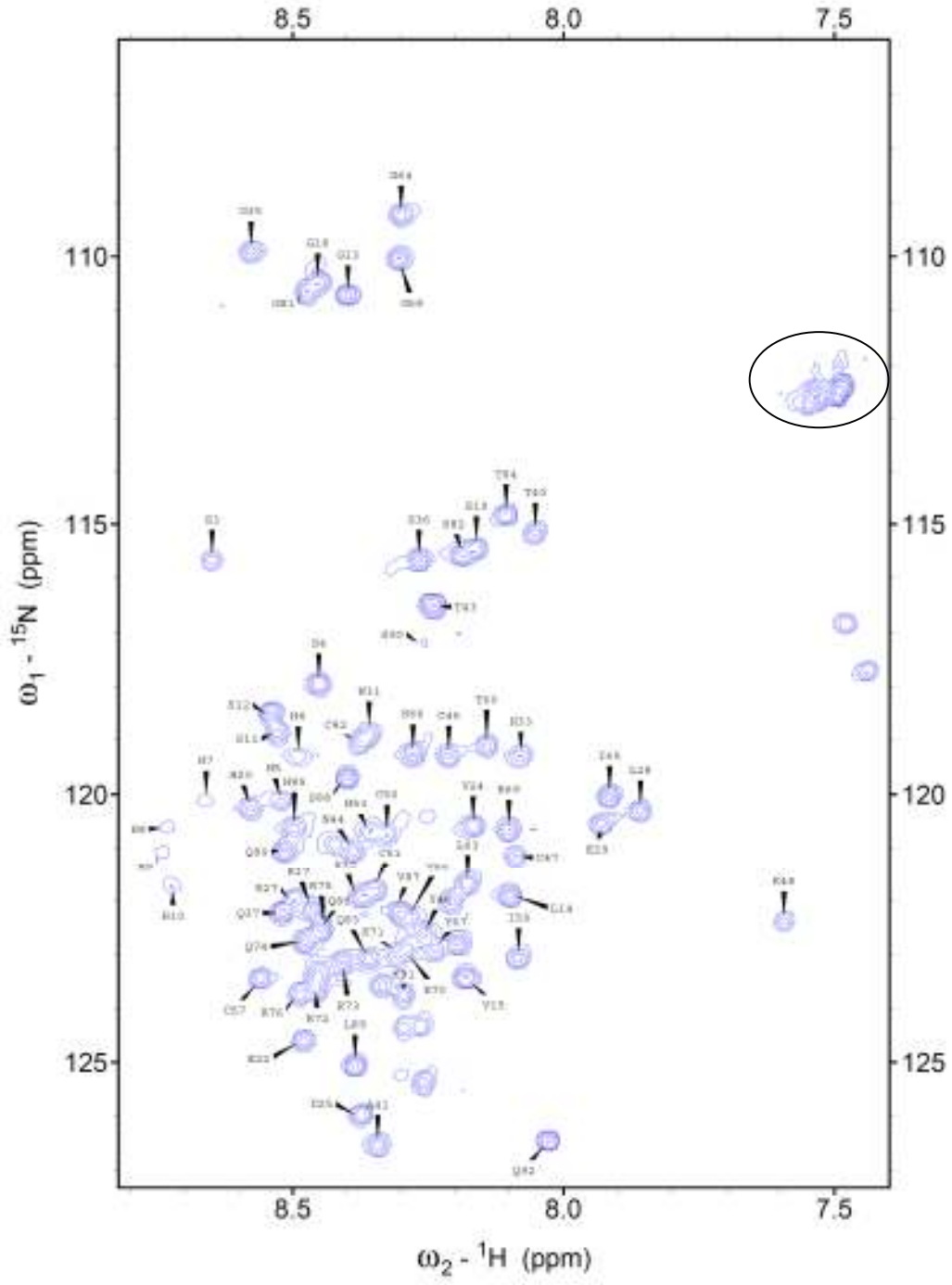

Figure 3.29 ^{1}H-^{15}N HSQC spectrum at 298 K of ^{15}N enriched, His-tagged Tat_{1-72} in 10 mM acetate at pH 4 in the presence of 10 mM TCEP with resonance assignments. The circled resonances are those of the amide groups of Asn and Gln sidechains.

Figure 3.30 shows that the ^{1}H-^{15}N HSQC peak intensities relative to the most intense resonance of Thr-43 vary widely over the length of the protein. In general, cross-peak intensity loss can be attributed to an overall decrease in the molecular tumbling rate of a protein as a result of an increase in molecular weight. Here, Tat_{1-72} self-association caould decrease the overall rotational correlation times. Another possible explanation is an increase in conformational dynamics on the ms – μs timescale or an increase in H-exchange rates. Most peaks are similar in intensity and fall within ± 0.1 intensity units of the average relative intensity of 0.488. A few peaks are absent, likely reflecting conformational exchange in those regions on the ms – μs timescale. The peak intensities of residues 5-10 (His-tag) are below average, suggesting conformational exchange. Above average peak intensity was observed for residues 13-20 of the His-tag, residue 43 of the hydrophobic core region, residues 69, 72, and 73 of the arginine-rich regions, and residues 80-92 of the glutamine-rich region. The most intense peaks might reflect local structure that slows down the H-exchange rates.

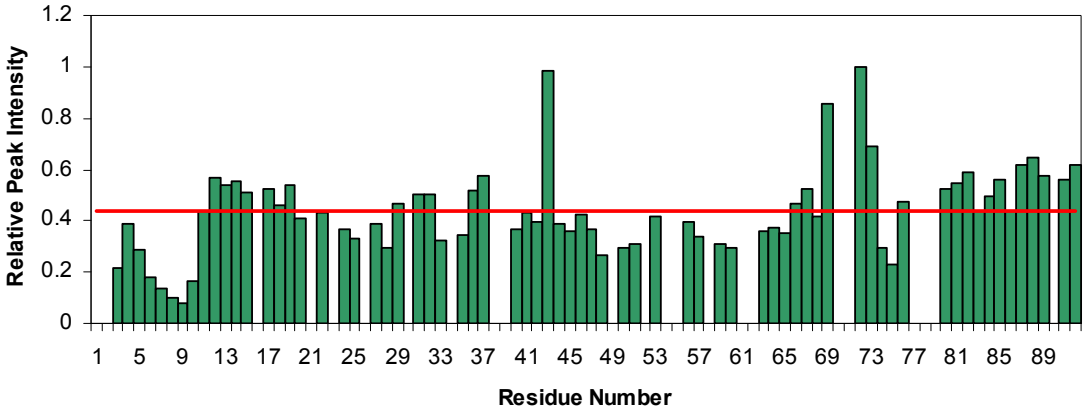

Figure 3.30 ^{1}H-^{15}N HSQC relative peak intensities of Tat_{1-72} in the presence of reducing agent at pH 4. Peak intensity is shown relative to the most intense residue, Thr-43. The red line indicates the average relative intensity, 0.488.

Chemical shift indexing (CSI) of the ^{15}N resonances (Figure 3.31) and ^{1}H resonances (Figure 3.32) shows that in comparison to the theoretical random coil shifts[2] corrected for sequence dependent effects[3], Tat_{1-72} does not display any well defined regions of secondary structure. This is evident by the alternating chemical shift indices between β-sheet and α-helical values and the lack of three or more adjacent residues with similar chemical shift differences of significant deviation from random coil values to suggest local secondary folding as suggested by Wishart[268].

Table 3.12 The theoretical $^{1}H^{N}$ and ^{15}N random coil NMR chemical shifts for each amino acid[2] corrected for sequence dependent effects[3].

Amino Acid	$^{1}H^{N}$	^{15}N
Ala	8.37	124.1
Cys-ox	8.54	118.7
Cys-red	8.62	121.61
Asp	8.67	119.71
Glu	8.52	121.51
Phe	8.29	122.99
Gly	8.41	107.5
His	8.72	119.23
Ile	8.28	125.13
Lys	8.47	122.97
Leu	8.39	123.31
Met	8.55	121.67
Asn	8.61	119.61
Gln	8.57	121.98
Arg	8.52	122.68
Ser	8.56	118.02
Thr	8.39	114.75
Val	8.28	123.5
Trp	8.13	125.03
Tyr	8.24	123.48

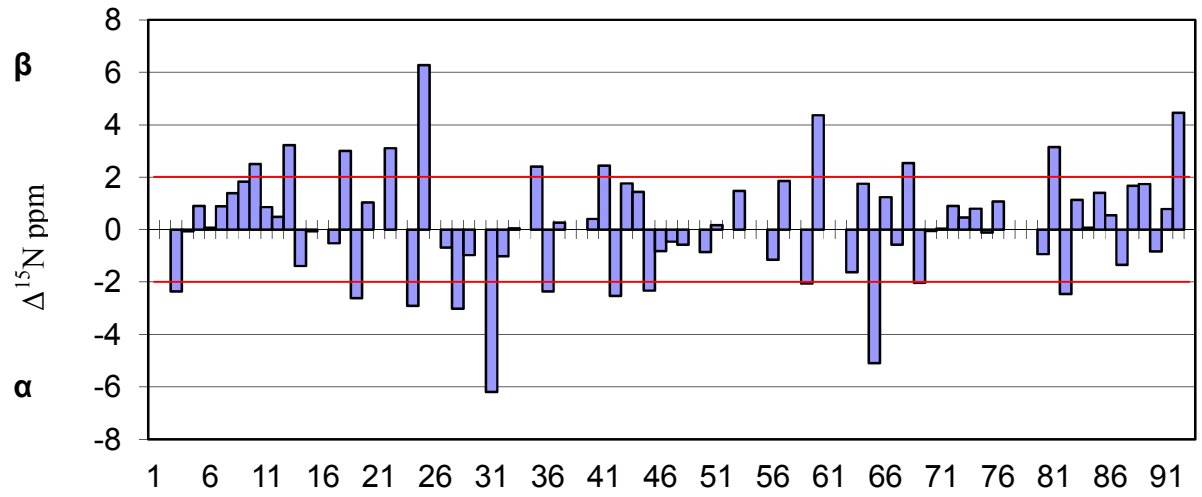

Figure 3.31 ^{15}N chemical shift differences of Tat_{1-72} HSQC amide backbone cross-peaks (Figure 3.29) from random coil values corrected for local sequence effects. Reference lines (red, ± 2 ppm) correspond to thresholds where differences begin to reflect regions of possible secondary structure formation. The plot range corresponds to two standard deviations (S.D.) from the mean value determined from the chemical shift tables in the Biological Magnetic Resonance Bank (http://www.bmrb.wisc.edu/).

Chemical shift indexing of the $^{1}H^{N}$ resonances of Tat_{1-72} at pH 4 in the presence of reducing agent shows α-helical propensity prominent in the cysteine (Δ ppm = -0.877 – 0.019) and glutamine-rich regions (Δ ppm = -0.542 – 0.069), and to lesser extent, parts of the hydrophobic core (Δ ppm = -0.365 − -0.003) and basic regions (Δ ppm = -0.419 − -0.033) (Figure 3.32). This is evident in several locations where 3 or more contiguous residues exceed the threshold (± 0.15 ppm) for α-helix structures. A key difference in the CSI of $^{1}H^{N}$ compared to ^{15}N resonances is the general trend towards helicity throughout the whole protein. From the CD spectrum of Tat_{1-72} in Figure 3.3, it is clear that the peptide backbone is mainly disordered,

but the $^{1}H^{N}$ chemical shift differences in the cysteine, core, arginine, and glutamine regions are show α-helical tendency, and this agrees with the results published by Shojania for $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C^{O}$, and H^{α} chemical shift indexing[20].

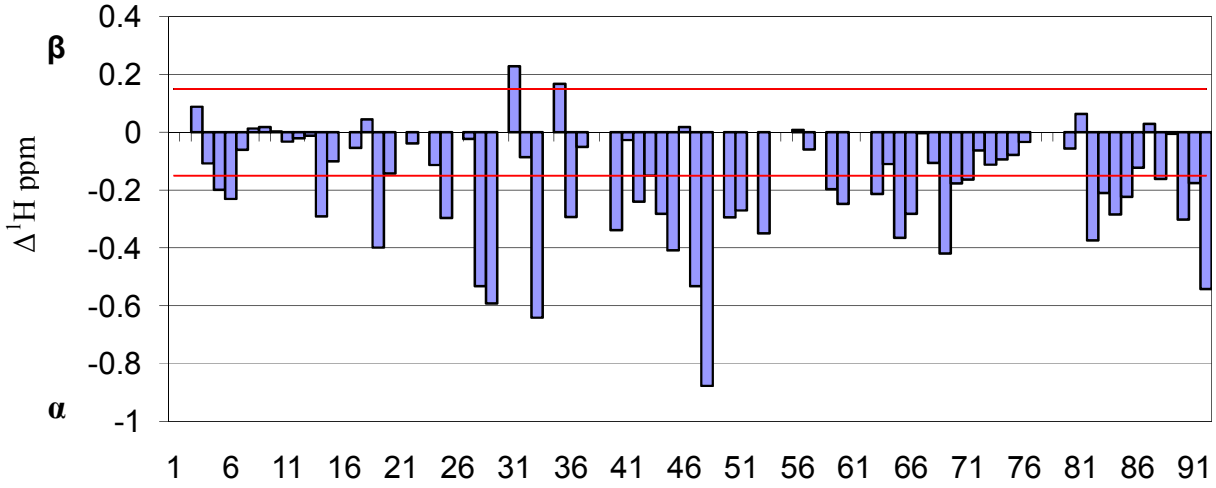

Figure 3.32 $^{1}H^{N}$ chemical shift differences of $Tat_{1\text{-}72}$ HSQC amide backbone cross-peaks from random coil values corrected for local sequence effects. Reference lines (red, ± 0.15 ppm) correspond to thresholds where differences begin to reflect regions of possible secondary structure formation. The plot range corresponds to two S.D. from the mean value determined from the chemical shift tables in the Biological Magnetic Resonance bank (http://www.bmrb.wisc.edu/).

3.9 Two-Dimensional ^{1}H-^{15}N Heteronuclear Single Quantum Coherence Experiments of ^{15}N enriched Tat_{1-72} in the presence of TFE

3.9.1 7% TFE

At 7% TFE, the ^{1}H-^{15}N HSQC Tat_{1-72} amide cross-peaks experience an upfield shift in both dimensions as seen by the overlay of HSQC spectra of Tat_{1-72} in 0 % and 7 % TFE shown in Figure 3.33. Seventy of the 83 non-proline residues could be assigned by comparison to the assigned (Figure 3.34) ^{1}H-^{15}N chemical shifts in the presence of 0% TFE (Figure 3.29). Two peaks, Lys-70 and Lys-71 of the arginine-rich region, were not assigned due to ambiguity from spectral overlap. Resonances in the arginine rich region (72 – 76) experienced significant spectral overlap. In general, the backbone amide resonances were more intense and are better resolved with the addition of 7 % TFE, with the exception of residues 72 – 76, in the presence of 7% TFE compared to 0% TFE. A possible explanation is that H-exchange rates are decreased in 7 % TFE by two possible mechanisms. In one, OH^- concentrations are lower because K_W is suppressed in the organic solvent. A second possibility is that in high concentrations of organic solvent amide water solvation may be reduced that could contribute to lower H-exchange rates[269]. The addition of TFE to NMR samples at low concentrations increased peak resolution of a 42-residue polypeptide designed to adopt a tight helix-turn-helix motif in the presence of TFE[270] . Studies of a 14-residue IDP that forms an α-helix in the presence of TFE also showed an increase in peak resolution at all concentrations of TFE[259].

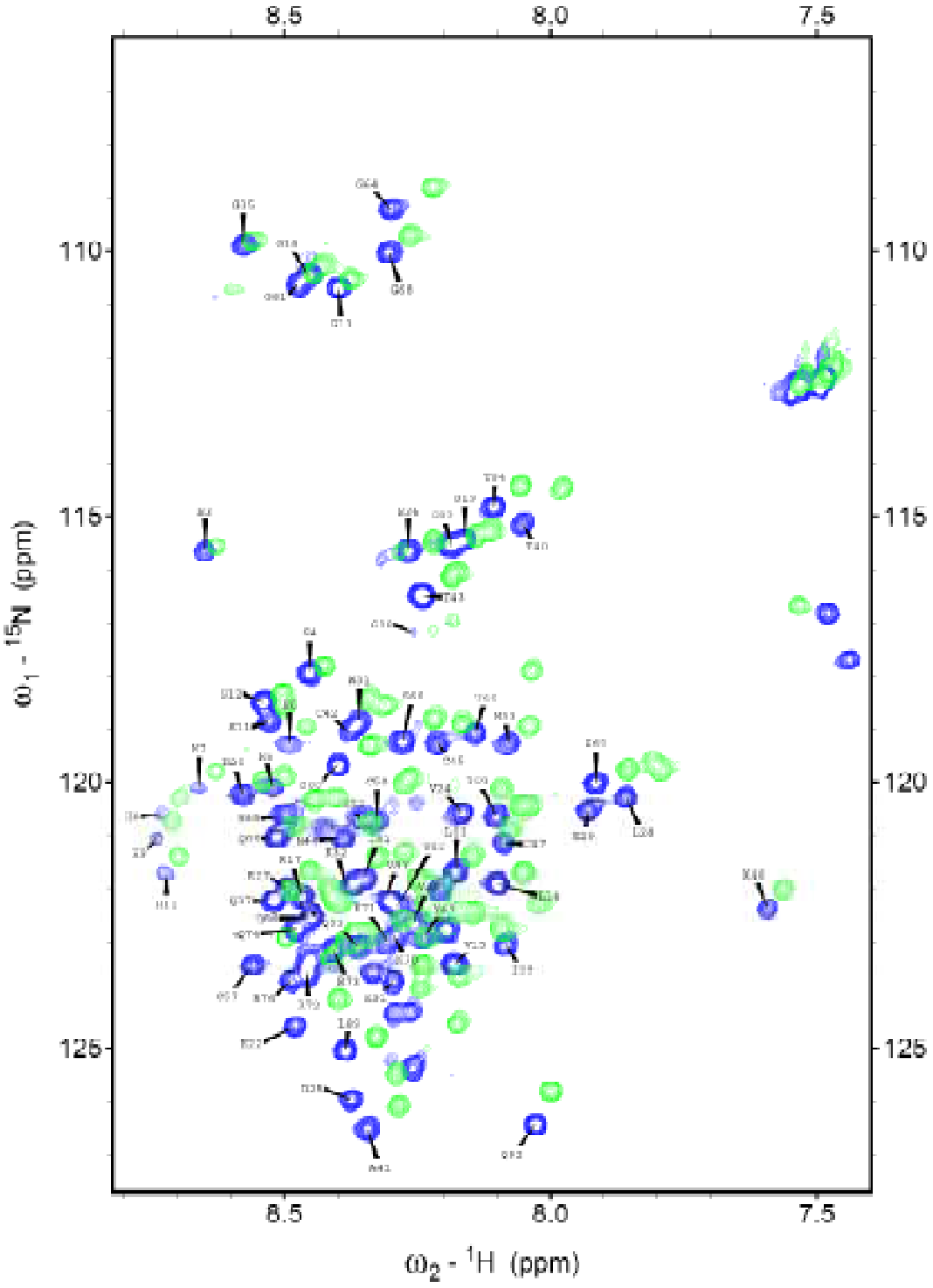

Figure 3.33 ^{1}H-^{15}N HSQC spectra of backbone amide resonances for ^{15}N enriched $Tat_{1\text{-}72}$ in 10 mM TCEP, 0% TFE (blue, pH 4.0) and 7% TFE (green, apparent pH 4.01) overlayed at 298 K.

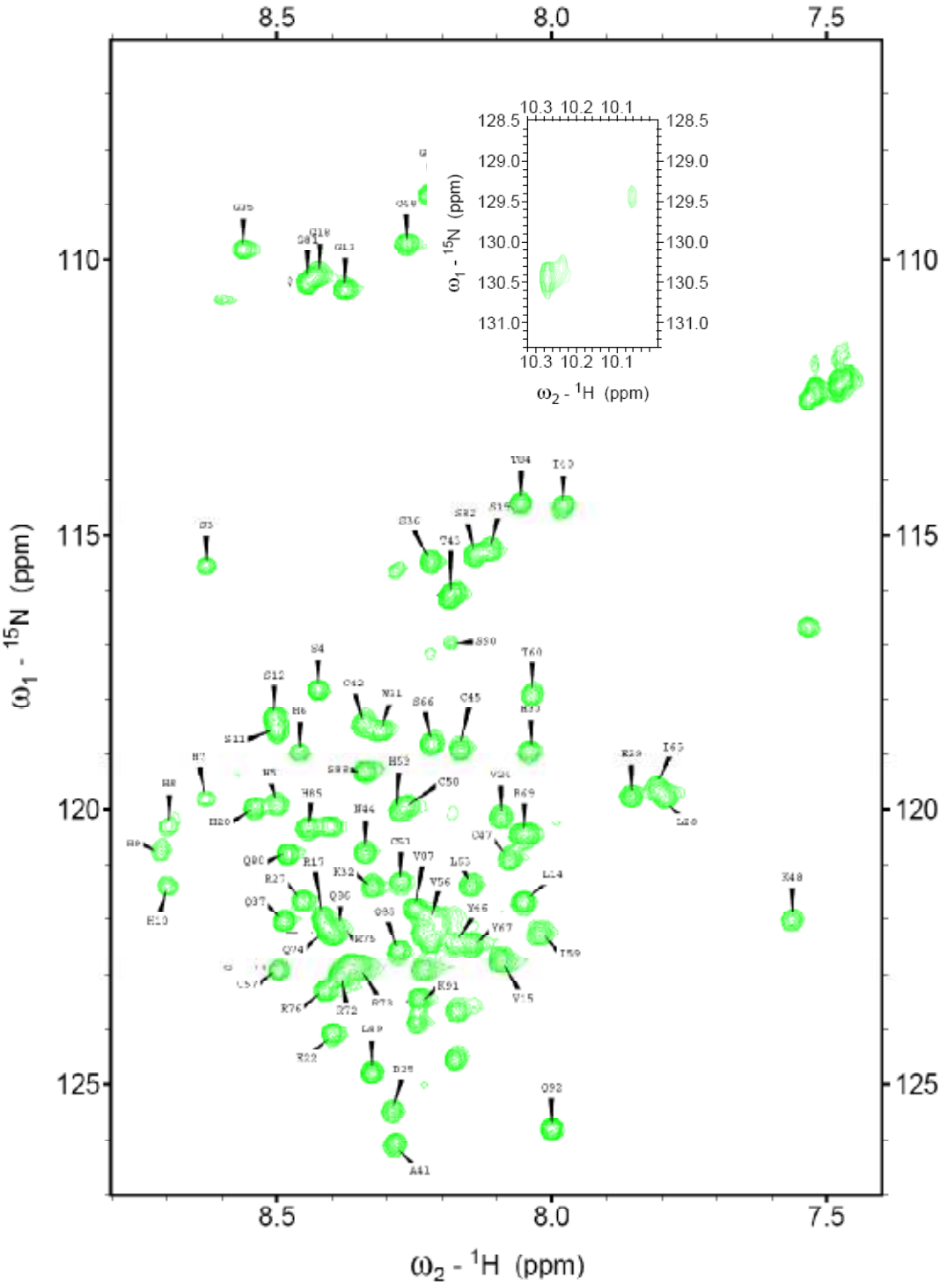

Figure 3.34 ^{1}H-^{15}N-HSQC amide backbone resonance assignments at 298 K for ^{15}N enriched, His-tagged $Tat_{1\text{-}72}$ in 9 mM TCEP with 7% TFE, apparent pH 4.01. The inset box shows the indole amine resonances of Trp-31.

The peak intensities relative to Thr-43 throughout the sequence show a similar pattern to those of Tat_{1-72} without TFE. The average relative peak intensity decreases from 0.488 to 0.437 relative to Thr-43 and a decrease in the relative intensities of the His-tag is observed (Figure 3.35).

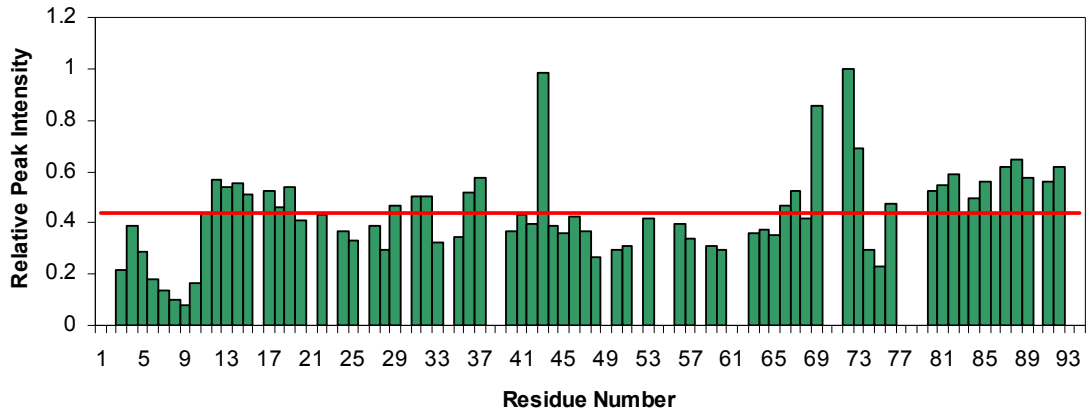

Figure 3.35 Relative ^{1}H-^{15}N HSQC peak intensities of His-tagged Tat_{1-72} in 7% (v/v) TFE, 9 mM acetate, 9 mM TCEP, apparent pH 4.01. The average relative intensity denoted by the red line is 0.437.

There are three resonances that appear for the indole amine of Trp-31. Upon integration, the relative populations are 88 % for the most intense resonance, and 6 % for each of the weaker resonances and these are unchanged from their values in water. The relative populations for the Gly-35 and Ser-36 resonances are 75 % and 25 %. In 0 % TFE, Gly-35 and Ser-36 showed a population distribution of 92 % and 8 %. In 7 % TFE, the minor conformers of Gly-35 and Ser-36 populations are equal, and are 3-fold greater than in the absence of TFE, suggesting that proline-isomerism may be influenced by TFE.

Chemical shift indexing of the ^{15}N resonance peaks of Tat_{1-72} in 7% TFE shown in Figure 3.36 indicate that Tat_{1-72} is disordered with little preference for any secondary structure. A short region between residues 27-33 of the proline-rich region is suggestive of an α-helical propensity, but the small chemical shift deviations from random coil values still implies that the regions are mostly disordered. The residues of the proline-rich region in the presence of 7 % TFE may indicate a location that is adopting helical structure, or evidence for a region where kinetic interconversion between disordered and helical structure can occur. Recall that the CD spectra of 5 % and 10 % TFE are characterized by a predominantly random-coil spectrum (Figure 3.18).

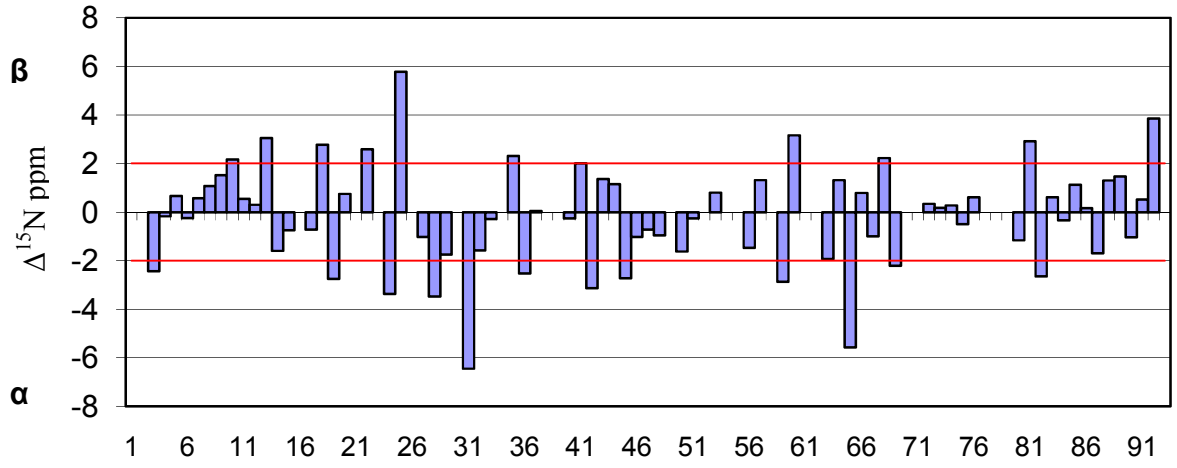

Figure 3.36 ^{15}N chemical shift differences from random coil values corrected for local sequence effects of Tat_{1-72} in 7% TFE. Reference lines (red, ± 2) correspond to thresholds where differences begin to reflect regions of possible secondary structure formation. The plot range corresponds to two S.D. from the mean value determined from the chemical shift tables of the Biological Magnetic Resonance Bank (http://www.bmrb.wisc.edu/).

In Figure 3.37, the $^{1}H^{N}$ chemical shift differences of Tat_{1-72} in 7% TFE show preferred structural tendencies that were suggested by shift differences of Tat_{1-72} in the absence of TFE. Helical propensity, suggested by chemical shift indexing, is evident between residues 40 – 53 encompassing most of the cysteine-rich region and residues 59-76 encompassing the latter half of the hydrophobic core region and most of the arginine-rich region. The glutamine-rich region from residue 82-92 also shows helical propensity. The regions of helical propensity are centered around residues with significant deviations from random coil values with the exception of Leu-28 and Glu-29, and His-33. Residues of regions of α-helical propensity that show a prominent bias towards helicity in their chemical shift indexing, which may identify local regions of helical propensity are Cys-45, Cys-47, Lys-48 of the cysteine region, Ile-65 and Arg-69 of the basic region. Loop regions, or regions of disorder are suggested by alternating chemical shift indices, and indices of insignificant variation from random coil values. These regions are centered on residues 7-13 of the His-tag (Δ ppm = -0.061 − -0.009), residues 56-57 between the cysteine-rich and hydrophobic core regions (Δ ppm = -0.123 − -0.059), and residues 80-81 of the glutamine region (Δ ppm = -0.089 – 0.034).

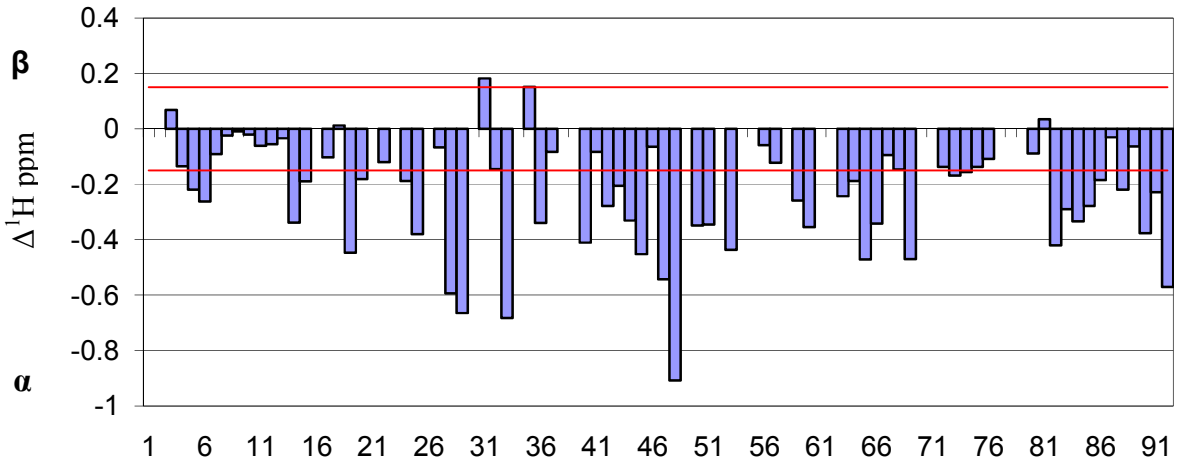

Figure 3.37 $^{1}H^{N}$ chemical shift differences from random coil values corrected for local sequence effects of Tat_{1-72} in 7% TFE. Reference lines (red, ± 0.15) correspond to thresholds where differences begin to reflect regions of possible secondary structure formation. The plot range corresponds to two S.D. from the mean value determined from the chemical shift tables of the Biological Magnetic Resonance Bank (http://www.bmrb.wisc.edu/).

TFE at 7 % (v/v) causes a shift of all $^{1}H^{N}$ resonances toward helicity, but the extent of the effect differs among the regions of Tat_{1-72}. From 0 % - 7 % TFE, the core region shows the largest average change in chemical shift values, 0.075 ppm, and in decreasing order, the arginine region change is 0.060 ppm, the proline region is 0.056 ppm, the glutamine region is 0.053 ppm, and the His-tag is 0.034 ppm.

3.9.2 15% TFE

Figure 3.39 shows that sixty-one of the possible 80 non-proline residues of ^{15}N enriched Tat_{1-72} dissolved in 15 % TFE, 10 mM acetate, 10 mM TCEP could be assigned by good

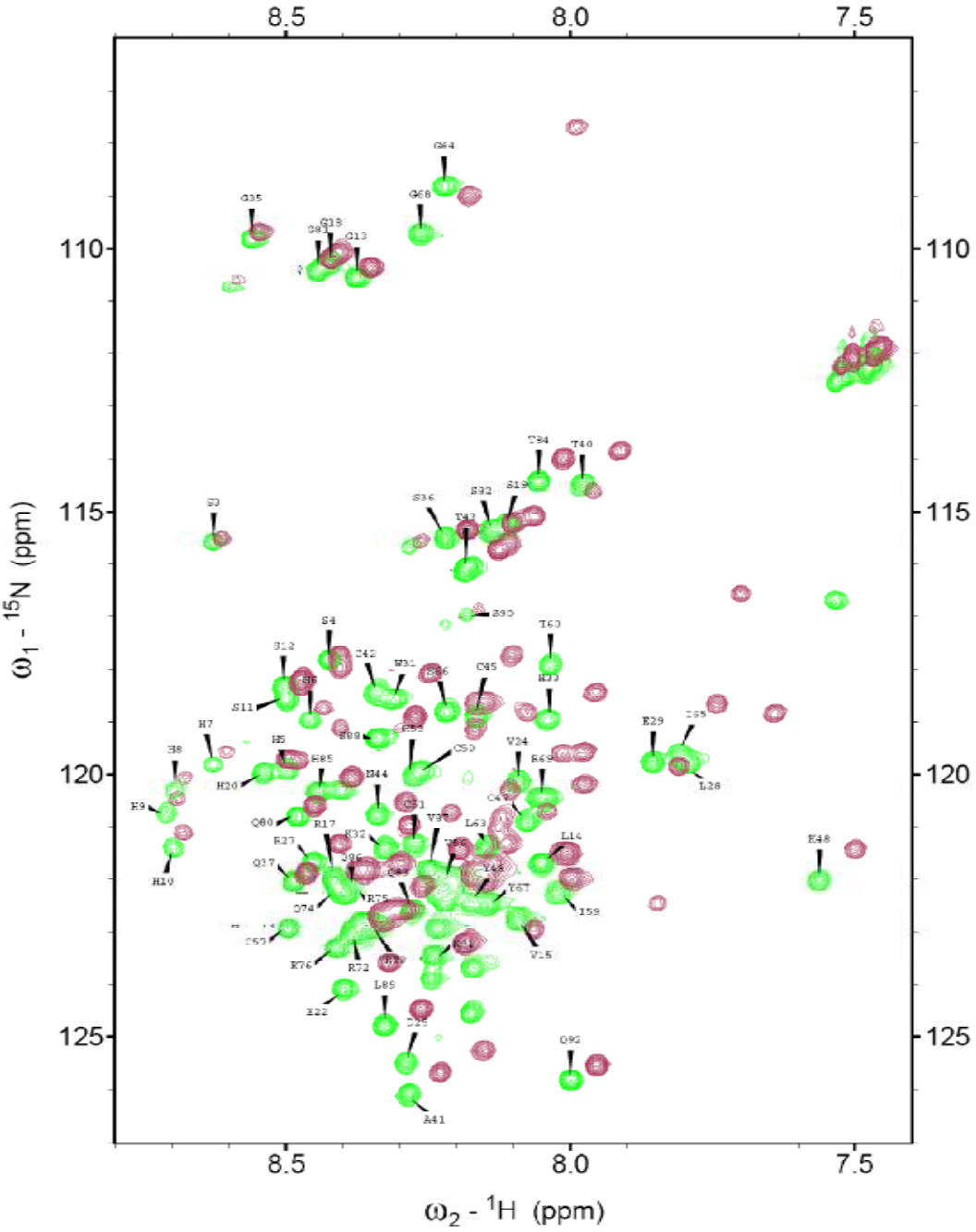

Figure 3.38 ^{1}H-^{15}N-HSQC spectra of ^{15}N enriched Tat_{1-72} in 7% and 15% TFE at 298 K. Green peaks with residue labels are the backbone amide resonances of ^{15}N enriched Tat_{1-72} with 9 mM TCEP in 7% TFE at apparent pH 4.01. Maroon peaks are overlay resonances of ^{15}N enriched Tat_{1-72} with 8.5 mM TCEP in 15% TFE at apparent pH 4.05.

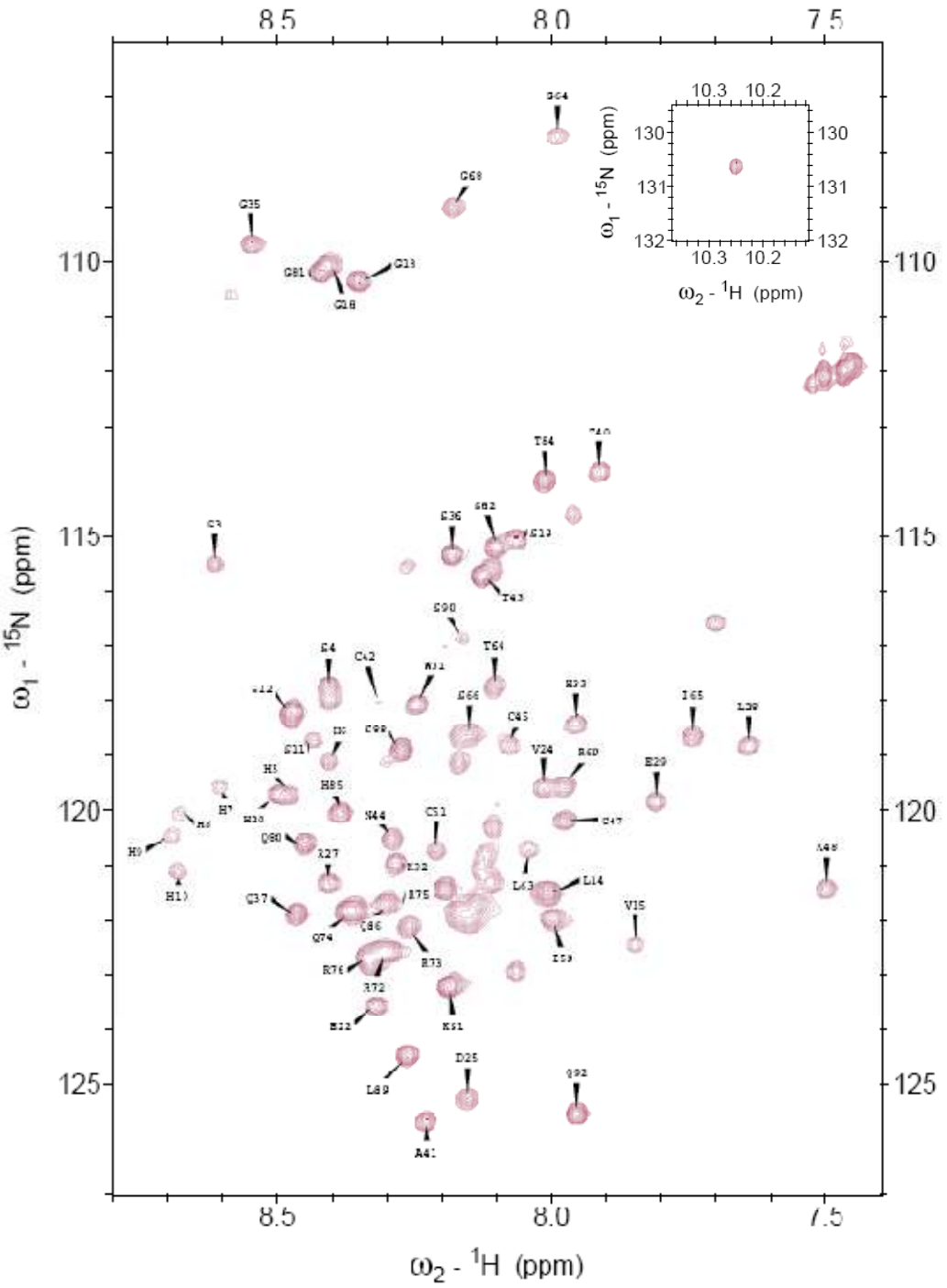

Figure 3.39 ^{1}H-^{15}N HSQC backbone amide resonance assignments at 298 K for ^{15}N enriched, His-tagged Tat_{1-72} with 8.5 mM TCEP in 15% TFE at apparent pH 4.05. The inset box shows the resonance for Trp-31.

agreement with ^{15}N Tat_{1-72} in 7% TFE (Figure 3.38) and by cross-referencing to a partially assigned ^{1}H-^{15}N HSQC spectrum of ^{15}N-Tat_{1-72} in 30 % TFE (Figure 3.44). The ^{1}H-^{15}N HSQC cross-peaks that could not be assigned or have disappeared from the 15 % (v/v) TFE solution are Arg-17 of the His-tag, Tyr-46, Cys-50, Phe-52, Val-56 and Cys-57 of the cysteine region, Tyr-67 of the core region, and Glu-83 and Val-87 of the glutamine-rich region. Residues may not be assigned either because they overlap with another resonance in the new conditions or because their dynamics have moved into the ms – μs timescale range, broadening out the resonance.

Relative to Gln-74 of the assigned peaks of Figure 3.40 is 0.312, which is lower than the previous relative intensity averages in 7% and 0% TFE conditions. The most intense peak in 0 % and 7 % TFE was Thr-43. Since slower H-exchange cannot explain the lower relative intensity to Thr-43, this suggests that conformational dynamics explains the intensity change. There is not a consistent variation of the peak intensities from the average, with the exception of residues 6 – 11 of the His-tag where the intensitites are decreased. The sequence of five amino acids in the His-tag could be undergoing conformational averaging that reduces their intensitites. Similar to the addition of 7 % TFE (v/v), the overlay spectra of 7 % TFE and 15 % TFE of Figure 3.38 show a general upfield shift in the resonances in both the proton and ^{15}N dimensions, and this is expected if TFE is inducing helical structure in the protein.

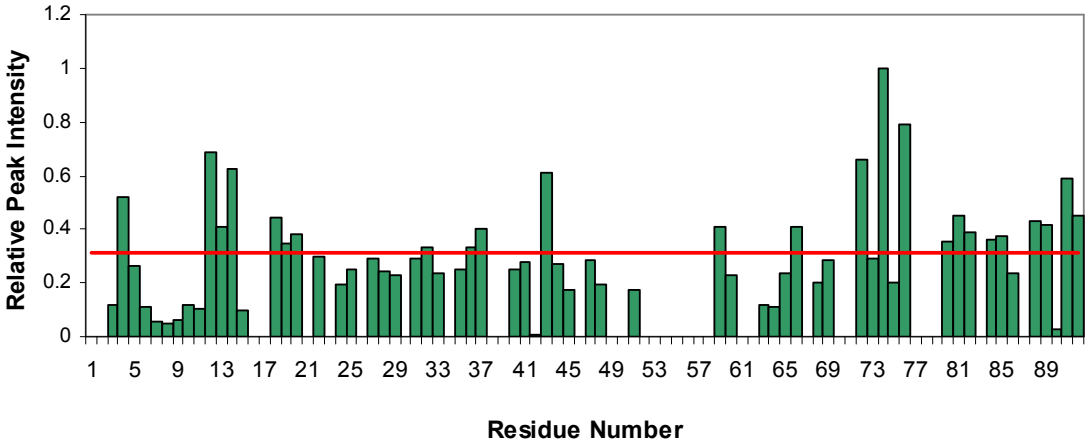

Figure 3.40 Relative ^{1}H-^{15}N peak intensities of His-tagged Tat_{1-72} in 15% (v/v) TFE, 8.5 mM acetate, 8.5 mM TCEP, at apparent pH 4.05. The average relative intensity indicated by the red line is 0.312.

In the presence of 15 % TFE, the side chain resonances of the Trp-31 indole amine are not much different from what was observed in 7 % TFE, showing three resonances with relative populations of 90 %, 5 %, and 5 %. Gly-35 exists in two conformers with populations of 73 % and 27 % and Ser-36 exists in conformers with relative populations of 78 % and 22 %. This is not very different from the conformers existing in 7 % TFE (75 % and 25 %).

From the chemical shift indexing of ^{15}N resonance differences from random coil values shown in Figure 3.41, there is little secondary structure preference as observed by shift difference bias toward α-helix or β-sheet. The shift differences are closely similar to those observed for Tat_{1-72} in 7% TFE. There is some indication of β-sheet propensity in the His-tag, with little or no preference for either helical, or β-sheet in any other region of the protein, as would be indicated by a sequence of at least three or more residues with the same propensity and

significant deviation from the random coil values. Recall that the CD spectrum of Tat_{1-72} in 15 % TFE suggests a predominantly random-coil conformation (Figure 3.18).

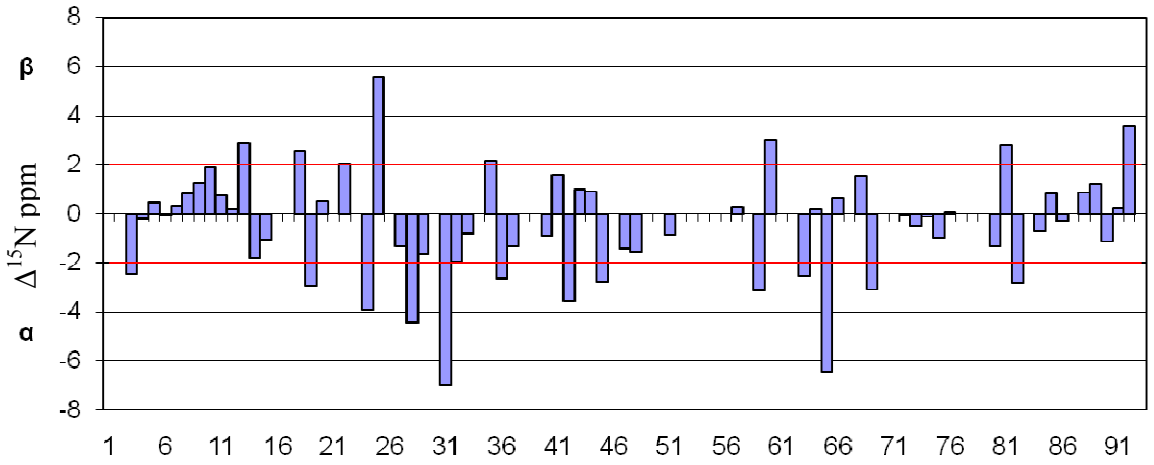

Figure 3.41 ^{15}N chemical shift differences of His-tagged Tat_{1-72} in 15% TFE from random coil values corrected for local sequence effects. Reference lines (red, ± 2) correspond to thresholds where differences begin to reflect regions of possible secondary structure formation. The plot range corresponds to two S.D. from the mean value determined from the chemical shift tables of the Biological Magnetic Resonance Bank (http://www.bmrb.wisc.edu/).

The $^{1}H^{N}$ chemical shift indexing of Tat_{1-72} in 15% TFE is shown in Figure 3.42. Helical propensity in the cysteine-rich region between residues 40-53 and in the core and basic regions of residues 59-76 are flanked by regions that could not be assigned due to a lack of resonances, as is the case for the residues separating residues 63-76 from residues 82-92. Residues Glu-25, Leu-28 and Glu-29, His-33 and Thr-40 of the proline-rich region show strong helical propensity, although they cannot be assigned to a helical conformation due to a lack of three residues of similar deviation, and the fact that this region has interspersed proline residues (Pro-xxx-Pro motif that is highly conserved among HIV-1 Tat variants). Similar residues in 7% TFE and 15%

TFE show the strongest helical propensity and are centred in regions of helical propensity, which may suggest localized regions of helical adoption. These residues are Cys-45, Cys-47, Lys-48 of the cysteine region and Ile-65 and Arg-69 of the basic region.

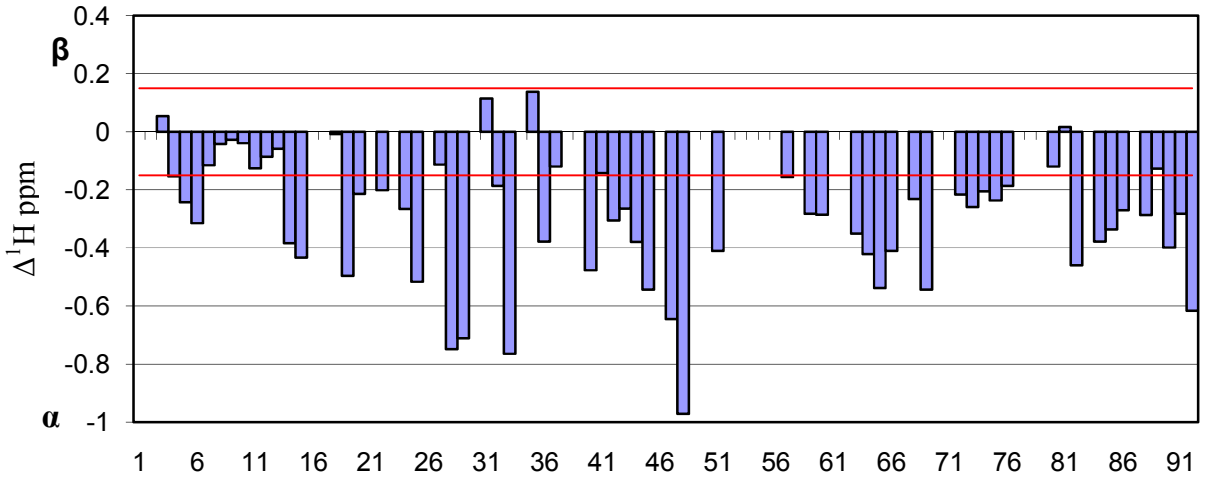

Figure 3.42 $^{1}H^{N}$ chemical shift differences of Tat_{1-72} in 15% TFE from random coil values corrected for local sequence effects. Reference lines (red, ± 0.15) correspond to thresholds where differences begin to reflect regions of possible secondary structure formation. The plot range corresponds to two S.D. from the mean value determined from the chemical shift tables of the Biological Magnetic Resonance Bank (http://www.bmrb.wisc.edu/).

From 0 % TFE to 15 % TFE, the region that experiences the greatest, α-helical biased average change in chemical shift is in the hydrophobic core region (Δ ppm = -0.538 − -0.156) with an average chemical shift change of 0.146 ppm, and in decreasing order, the arginine-rich region (Δ ppm = -0.544 − -0.119) shifts 0.139 ppm, the proline-rich region (Δ ppm = -0.749 − 0.138) changes 0.121 ppm, the cysteine-rich region (Δ ppm = -0.972 − -0.142) changes 0.114 ppm, the glutamine-rich region (Δ ppm = -0.616 − 0.017) changes 0.107 ppm, and the His-tag (Δ

ppm = -0.496 − -0.008) changes 0.079 ppm. All of these $^{1}H^{N}$ shift changes are upfield and in the direction of helix formation.

3.9.3 30% TFE

Fifty-seven of the 80 non-proline residues of His-tagged Tat_{1-72} could be assigned by the 3-D heteronuclear triple resonance experiments HNCA and HN(CO)CA and by comparison to previously published results[20]. Forty-seven residues could be assigned by their $^{1}H^{N}$, ^{15}N, and $^{13}C^{\alpha}$ resonances, while 10 were assigned by comparison to previously published results. 3-D heteronuclear triple resonance assignments exploit one and two bond scalar couplings to connect atoms of the protein backbone. The HNCA experiment correlates intraresidue (i) $^{1}H^{N}$, ^{15}N, and $^{13}C^{\alpha}$ resonances and also shows connectivity between the $^{1}H^{N}$(i) and ^{15}N(i) resonances with the $^{13}C^{\alpha}$ resonance of the preceding residues (i-1)[271]. In the complementary HN(CO)CA experiment only the $^{1}H^{N}$(i), ^{15}N(i). and $^{13}C^{\alpha}$(i-1) peaks appear. By comparison of the complementary spectra, the C^{α} shift of every preceding residue is determined, allowing sequential residue assignment. Figure 3.43 shows an example of the strip plots used for sequential assignment of Tat_{1-72} in 30 % TFE.

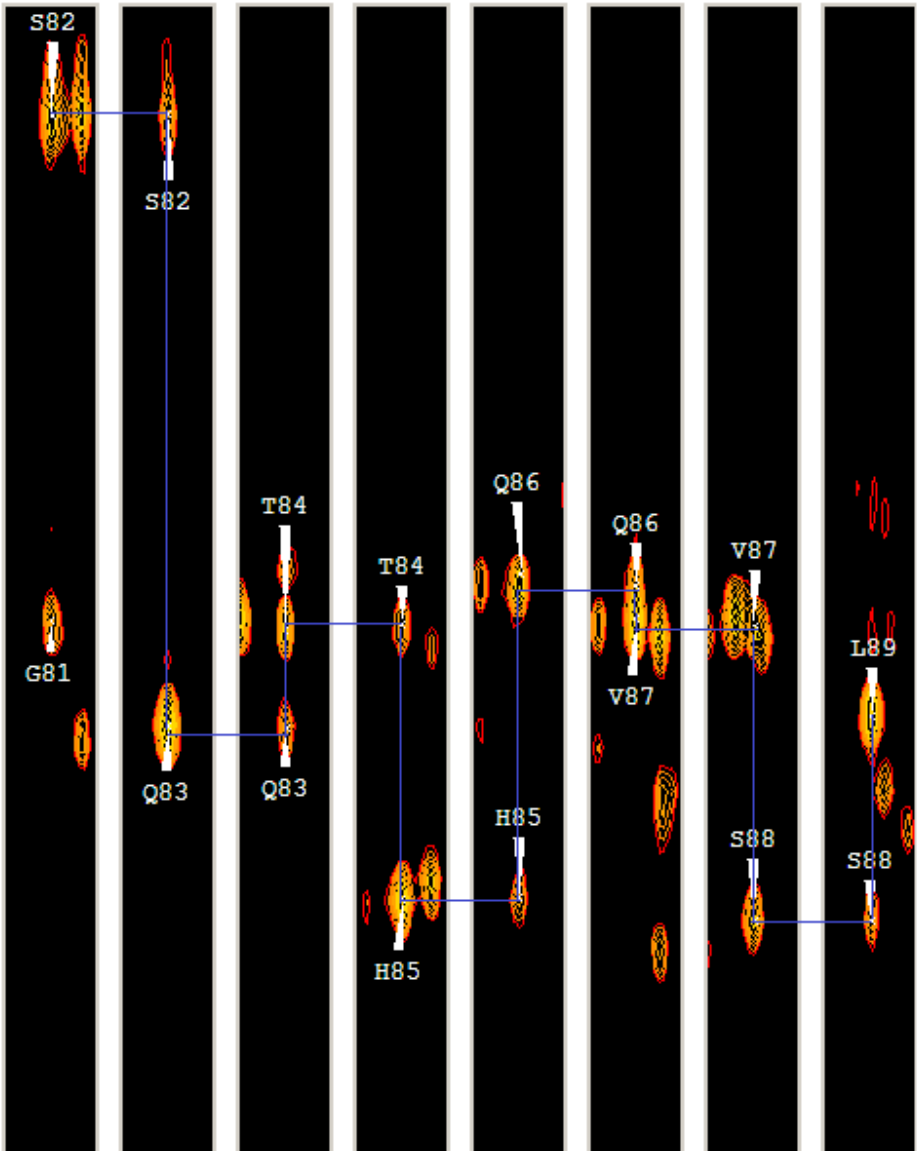

Figure 3.43 Strip plots correlating the $^{1}H^{N}$(i) and N(i) resonances with the C^{α} resonance of the preceding residues (i-1) by HNCA NMR spectroscopy for ^{15}N-, ^{13}C-labelled Tat_{1-72} dissolved in 30 % TFE, 7 mM TCEP, 7 mM acetate, apparent pH of 4.2.

Figure 3.45 shows a ^{1}H-^{15}N HSQC spectrum of ^{15}N and ^{13}C-enriched Tat_{1-72} dissolved in 30 % TFE, 7 mM acetate, 7 mM TCEP, at apparent pH 4.2. Peak broadening, along with the disappearance of peaks made resonance assignment difficult for residues His-20 of the His-tag, Glu-22, Val-24, Arg-27, Trp-31, Lys-32, and His-33 of the proline rich region, Cys-47 of the cysteine rich region, Ile-59, Thr-60, Leu-63, and Tyr-67 of the core region, Arg-75 and Arg-76 of the arginine rich region, and Gln-80 and 92 of the glutamine rich region. 3-D NMR allowed the assignment of resonances that were ambiguous in lower concentrations of TFE. These are: Tyr-46, Lys-49, and Cys-50 of the cysteine-rich region, and Gln-83 of the glutamine-rich region.

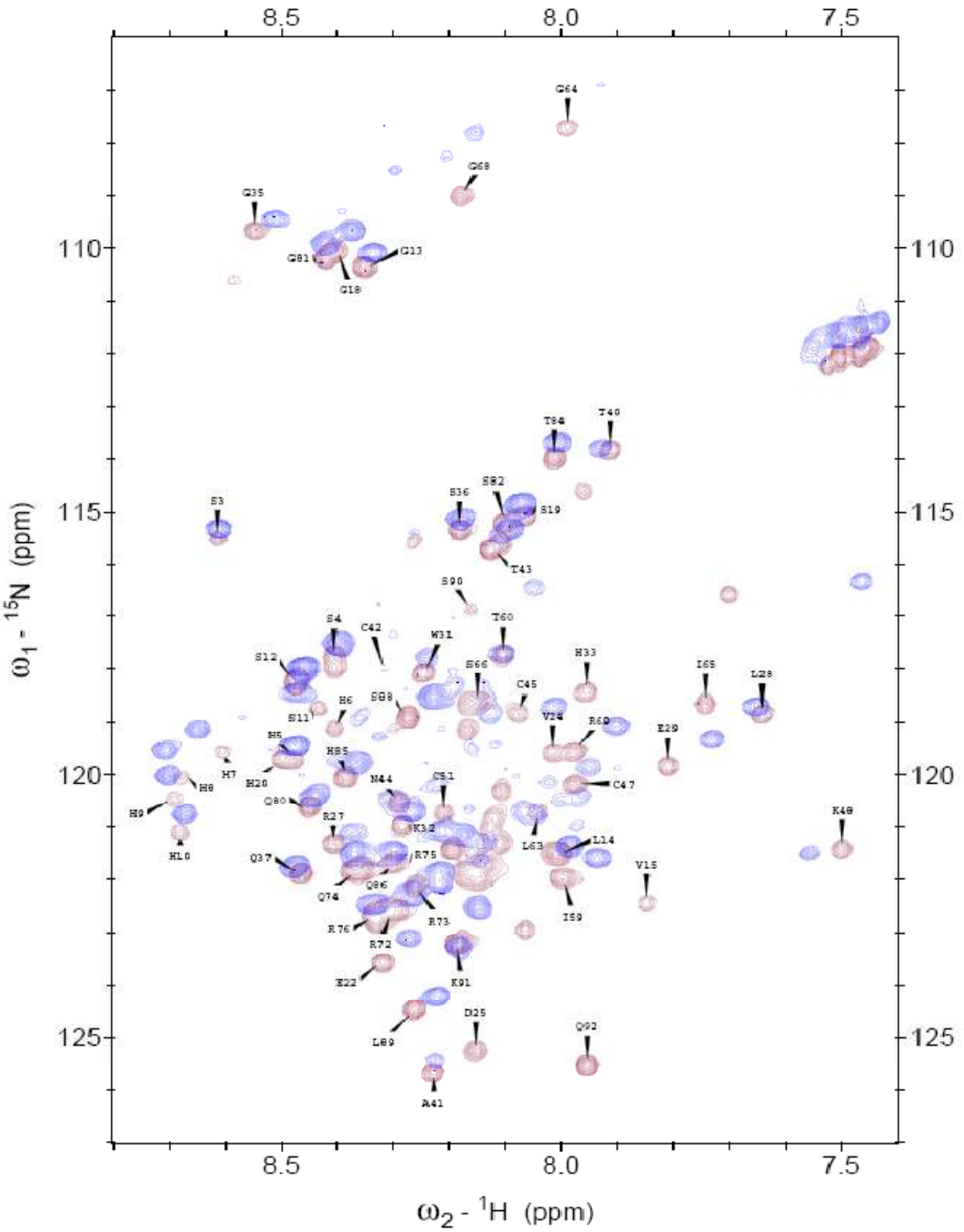

Figure 3.44 ^{1}H-^{15}N-HSQC spectra of ^{15}N enriched $Tat_{1\text{-}72}$ in 15% and 30% TFE at 298 K. Purple peaks with residue labels are the backbone amide resonances of ^{15}N enriched $Tat_{1\text{-}72}$ with 8.5 mM TCEP in 15% TFE at apparent pH 4.05. Blue peaks are resonances of ^{15}N enriched $Tat_{1\text{-}72}$ with 7 mM TCEP in 30% TFE at apparent pH 4.2.

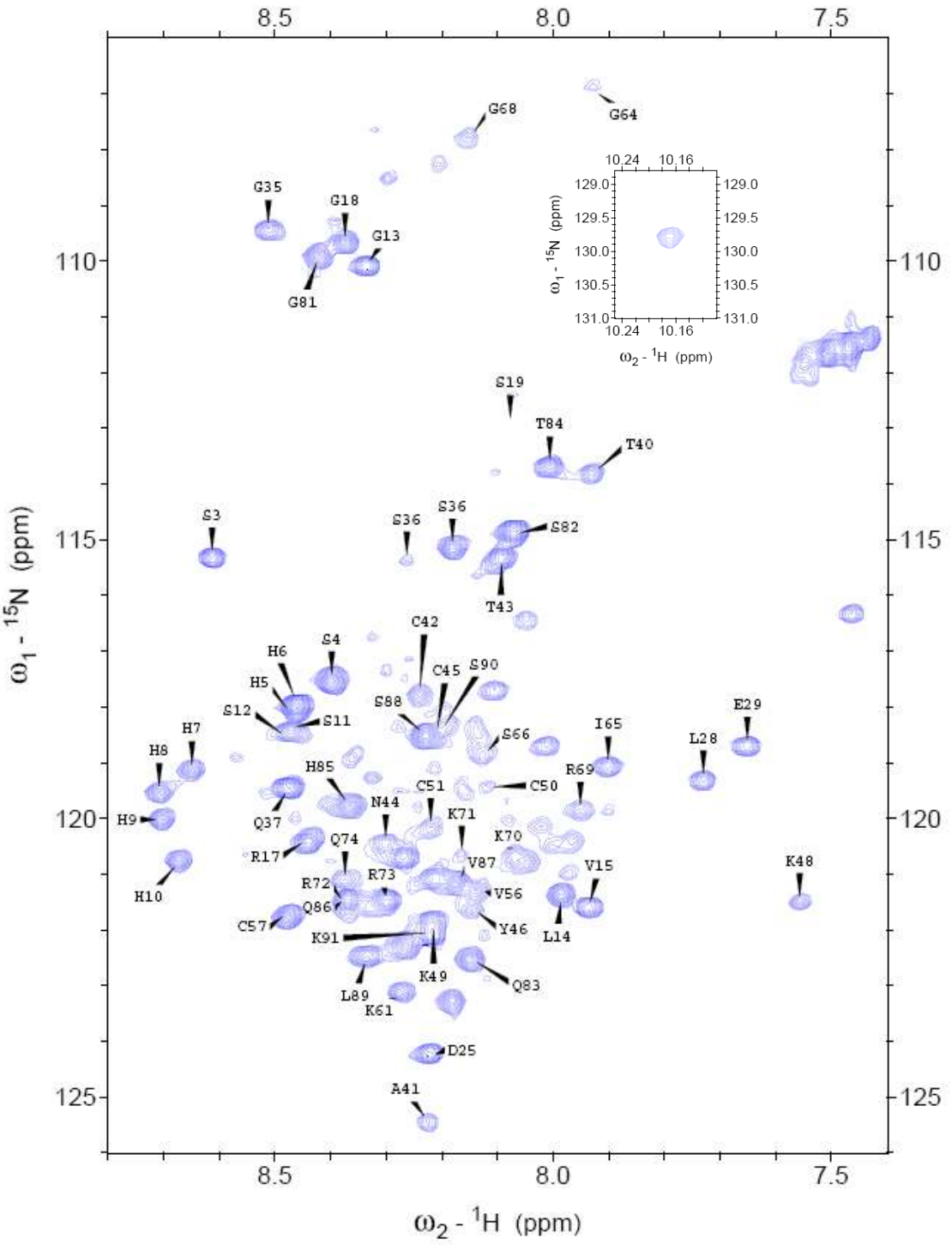

Figure 3.45 ^{1}H-^{15}N HSQC backbone amide resonance assignments at 298 K for ^{15}N enriched, His-tagged Tat_{1-72} with 7 mM TCEP in 30 % TFE at apparent pH 4.2. The inset box shows the resonance for the indole amine of Trp-31.

Figure 3.46 shows the relative peak intensities of Tat_{1-72} in 30 % TFE, 7 mM acetate, 7 mM TCEP, at apparent pH of 4.2. The intensities of Tat_{1-72} are reported relative to the most intense assigned residue, His-6. The only major change to residue intensities compared to the relative intensities of Tat_{1-72} in 15 % TFE is an increase in the relative intensities of the His-tag region, between residues 3 – 15 suggesting a major change in conformational dynamics or H-exchange in that region of Tat_{1-72}. The average peak intensity is 0.409, and is higher than Tat_{1-72} in 15 % TFE, but lower than the average peak intensity of Tat_{1-72} in 0% and 7 % TFE.

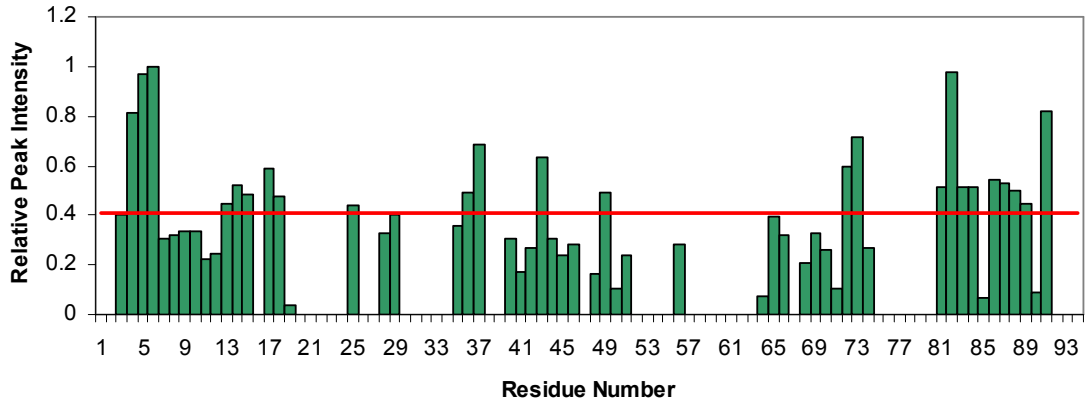

Figure 3.46 Relative ^{1}H-^{15}N peak intensities of His-tagged Tat_{1-72} in 30 % (v/v) TFE, 7 mM acetate, 7 mM TCEP, at apparent pH 4.2. The average, relative intensity denoted by the red line is 0.409.

In 30 % TFE, multiple resonance peaks appear for Ser-36, but the multiple resonances disappear for Gly-35 and the indole amine of Trp-31. From integrating the resonance peaks of Ser-36, the relative populations of the conformers are 90 % and 10 %. The minor conformer of Ser-36 has decreased from 25 % to 10 % with an increase in TFE fraction from 15 % to 30 % (v/v) possibly reflecting changes in conformational dynamics such as cis-trans proline isomerisation in the region of the protein.

The chemical shift indexing of the ^{15}N resonances of the ^{1}H-^{15}N HSQC spectrum of Tat_{1-72} in 30 % TFE shows minimal secondary structural homogeneity in terms of helicity or β-sheet (Figure 3.47). Residues 45 – 51 of the cysteine rich region show a helical tendency by their shift difference from random coil values, with the exception of Cys-47 for which no assignment could be made. Residues 69 – 74 also show a tendency for helicity. The three residues 69 – 72 of the arginine-rich region show a shift difference towards helicity that is greater than the threshold reference line plot, denoting a region of possible helical formation. There are numerous single peaks or pairs of peaks that show strong helical character: His-19, Leu-28, Glu-29, Ser-36, Gln-37, Cys-42, Ser-65, and Ser-82. The peaks that show strong β-sheet character are Ser-12, Arg-17, Asp-25, and Gly-81.

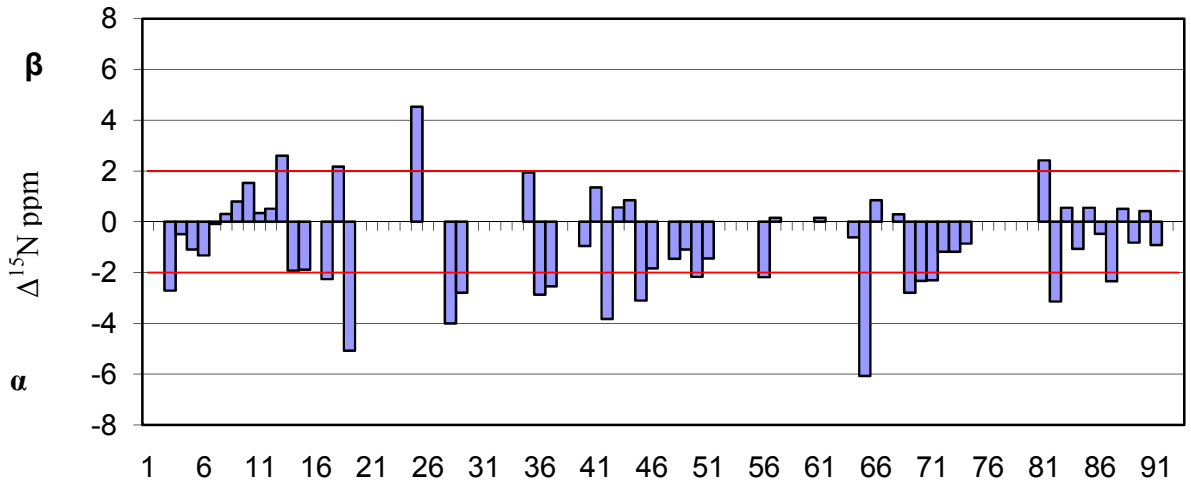

Figure 3.47 ^{15}N chemical shift differences from random coil values of Tat_{1-72} in 30 % TFE corrected for local sequence effects. Reference lines (red, ± 2) correspond to thresholds where differences begin to reflect regions of possible secondary structure formation. The plot range corresponds to two S.D. values from the mean value determined from the chemical shift tables of the Biological Magnetic Resonance Bank (http://www.bmrb.wisc.edu/).

The chemical shift indexing of the $^{1}H^{N}$ resonances of Tat_{1-72} in 30 % TFE suggests the presence of helical regions in the same regions observed to have helical propensity in lower concentrations of TFE (Figure 3.48). These regions are defined by consecutive sequential bias of the chemical shift differences from random coil values towards α-helical structure. These regions include residues 40 – 51, 64 – 74, and 82 – 91. The residues 40 – 51 begin in the latter portion of the proline-rich region and include Thr-40 and Ala-41, and traverse the cysteine region up to residue Cys-51. The resonance with the strongest α-helical character is Lys-48. Cys-47 is unassigned. Tyr-46 shows minimal deviation from a random coil structure, and the existence of an α-helical structure within this region could be interrupted depending on the character of the unassigned Cys-47. If both residues 46 and 47 showed minimal deviation from random coil values, then it could be assumed that the helix structure within residues 40 – 51 are interrupted by a flexible linker, but it is unclear. The argument is opposite if Cys-47 were assigned and showed strong helical character. Residues 64 – 74 begin at Gly-64 of the core domain and end at Gln-74 of the arginine-rich region. Within this series, Tyr-67 is unassigned. The final series of residues showing helical character begin at Ser-82 and end at Lys-91. These residues compose the majority of the glutamine-rich region. Val-87 could not be assigned, and depending on its character, may create a flexible region disrupting the helix in this region. Gln-92 is unassigned in 30 % TFE, but probably shows helical character, based on the chemical shift indexing of this residue in all conditions of this study (0 % - 15 % TFE). In all experiments, the terminal residue showed helical propensity. The $^{1}H^{N}$ CSI also revealed strong helical character in residues of the proline rich region, at Asp-25, Leu-28, and Glu-29.

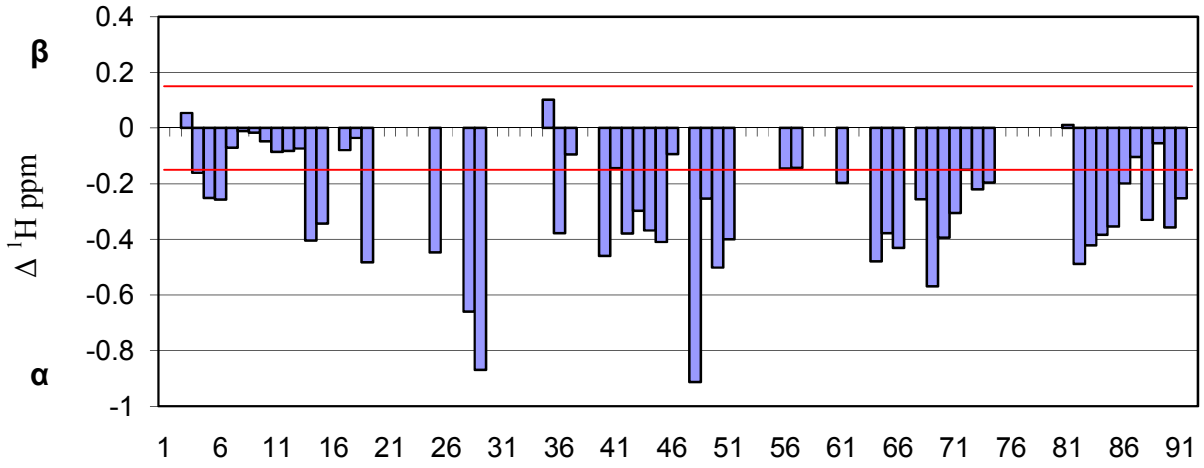

Figure 3.48 $^{1}H^{N}$ chemical shift differences from random coil values of His-tagged Tat_{1-72} with 30 % TFE corrected for local sequence effects. Reference lines (red, ± 0.15) correspond to thresholds where differences begin to reflect regions of possible secondary structure formation. The plot range corresponds to two S.D. from the mean value determined from the chemical shift tables of the Biological Magnetic Resonance Bank (http://www.bmrb.wisc.edu/).

3.9.4 Summary of the Effects of the TFE Titration on Tat_{1-72}

The effects of TFE on the chemical shifts of Tat_{1-72} are not uniform. Table 3.13 shows the average change in $^{1}H^{N}$ chemical shift per region of Tat_{1-72} with the addition of 7 %, 15 %, and 30 % TFE. The region of Tat_{1-72} which is most sensitive to TFE, and shows the greatest change in chemical shift compared to the ^{1}H-^{15}N cross-peaks of Tat_{1-72} in the absence of TFE is the hydrophobic core region. This is apparent by the greatest change in $^{1}H^{N}$ shift compared to all other regions at all intervals of TFE in this study. The next most sensitive region at all intervals of TFE is the arginine-rich region, and listed by decreasing sensitivity: the proline-rich region, the cysteine-rich region, the glutamine-rich region, and finally, the His-tag.

Table 3.13 Average $^{1}H^{N}$ chemical shift differences of the regions of Tat_{1-72} at the intervals of TFE used in the NMR studies.

	7 % - 0 %	15 % - 0 %	30 % - 0 %
His-tag region	-0.0341	-0.0791	-0.0638
Proline-rich region	-0.0555	-0.1209	-0.1250
Cysteine-rich region	-0.0547	-0.1093	-0.1137
Hydrophobic core region	-0.0749	-0.1455	-0.1548
Arginine-rich region	-0.0599	-0.1390	-0.1371
Glutamine-rich region	-0.0532	-0.0979	-0.1068

The $^{1}H^{N}$ chemical shift referencing of Tat_{1-72} shown in Figure 3.48 suggests that TFE-induced helical propensity exists in residues 40 – 51 of the cysteine-rich region, residues 64 – 74 of the latter portion of the hydrophobic core region and continuously into the arginine-rich region, and residues 82 – 91 of the terminal glutamine-rich region.

Multiple Tat_{1-72} conformers are suggested by the observation of multiple cross-peaks per residue. Both Gly-35 and Ser-36 populate a minor conformer (92 % and 8 % population conformers). The addition of 7 % TFE increases the minor conformer three-fold to 25 %. Successive titration to 15 % TFE has no effect. At 30 % TFE, Ser-36 cross-peaks show conformers of 90 % and 10 %. 30 % TFE causes the disappearance of minor conformers and that it is homogenizing the multiple conformers observed in aqueous buffer solutions. From 0 % to 30 % TFE, the increase in helical ellipticities at 190 nm and 222 nm is greatest (see Figure 3.20, Figure 3.21 and Table 3.8). After 30 % TFE, the rate of change of the helical ellipticites decreases 3-fold, therefore, the 7 % - 30 % TFE range is referred to as intermediate concentrations of TFE, and in this range the effect on Tat_{1-72} is most significant. At 30 % TFE

and higher, the ellipticities are not as sensitive, and from this one can conclude that the major structural changes in the form of helical influence by 30 % TFE onto the secondary structure of Tat_{1-72} have become evident. At intermediate concentrations of TFE (7 % - 30 %), the populations of the Gly-35 and Ser-36 minor conformers increase from 8 % in the absence of TFE to 25 % with the addition of 15 % TFE. It seems that TFE is beginning to favour the minor conformer, and at 30 % TFE, the minor conformer decreases to 10 %. This means that above intermediate concentrations of TFE, one of the conformers is favoured, and possibly, the effect of 30 % TFE may be the inversion of major and minor conformer populations.

The average chemical shift of each region is calculated simply by determining the average chemical shift from each residue within that region. The average chemical shift for each amino acid was calculated and is shown in Table 3.14. There does not seem to be any clear correlation between residue identity and sensitivity to 7 %, 15 %, and 30 % TFE solution composition.

Table 3.14 The average chemical shift difference by amino acid of ^{1}H-^{15}N enriched, His-tagged Tat_{1-72} in 7 % TFE and 0 % TFE, 15 % TFE and 0 % TFE, and 30 % TFE and 0 % TFE in decreasing order, with the exception of unassignable residues.

7 % - 0 % TFE		15 % - 0 % TFE		30 % – 0 % TFE	
Tyrosine	-0.0920	Valine	-0.2425	Glutamate	-0.2780
Aspartate	-0.0840	Aspartate	-0.2210	Valine	-0.1773
Isoleucine	-0.0840	Leucine	-0.1420	Aspartate	-0.1510
Glutamate	-0.0780	Glutamate	-0.1410	Glycine	-0.1302
Valine	-0.0730	Arginine	-0.1378	Threonine	-0.1237
Threonine	-0.0715	Isoleucine	-0.1295	Alanine	-0.1180
Arginine	-0.0597	Alanine	-0.1150	Lysine	-0.2425
Alanine	-0.0570	Tryptophan	-0.1140	Tyrosine	-0.1130
Glutamine	-0.0500	Cysteine	-0.1133	Cysteine	-0.1130
Leucine	-0.0495	Glycine	-0.1017	Glutamine	-0.1088
Lysine	-0.0480	Lysine	-0.1015	Leucine	-0.0973
Cysteine	-0.0480	Threonine	-0.0962	Arginine	-0.0933
Tryptophan	-0.0470	Glutamine	-0.0926	Serine	-0.0869
Serine	-0.0448	Serine	-0.0860	Histidine	-0.0477
Histidine	-0.0396	Histidine	-0.0760	Isoleucine	-0.0140
Glycine	-0.0363				
Average	-0.0602	Average	-0.1273	Average	-0.1180

3.10 Tat_{1-72} Interaction with TAR in the presence of DDM and 2 MEQ Zn(II)

The solubility of Tat_{1-72} in 20 mM SDS at pH 6.67 and in 100 mM SDS with reducing agent suggested that these conditions might be suitable for studying the conformation of Tat_{1-72} together with TAR. However, when Tat_{1-72} and TAR were combined in the presence of SDS with and without reducing agent at pH 4, a precipitate formed. Substituting zinc for the reducing

agent TCEP in the presence of SDS could not prevent precipitation of Tat_{1-72} and TAR when they were combined.

When Tat_{1-72} and TAR were combined in the presence of the non-ionic detergent n-dodecyl beta-D-maltoside (DDM) and TCEP, a precipitate formed. However, in the presence of DDM and 2 MEQ Zn(II) Tat_{1-72} and TAR are soluble. The Tat_{1-72} CD spectrum in the presence of DDM and 2 MEQ Zn(II) is dominated by a large, negative band at 200 nm indicating a mainly disordered structure as depicted in Figure 3.49 (green). The CD spectrum of TAR in the presence of DDM and 2 MEQ Zn(II) displays a positive band at 186 nm, a weak negative band at 210 nm, and a slightly positive ellipticity in the region of 220 nm – 260 nm as depicted in Figure 3.49 (blue). The CD signal of combined Tat_{1-72} and TAR together in the presence of DDM and 2 MEQ Zn(II) shows a positive ellipticity at 187 nm and a broad negative ellipticity with a peak at 212 nm as shown in Figure 3.49 (brown). This shift in CD signal when the two molecules are combined in the presence of DDM and 2 MEQ Zn(II) is evidence of a binding interaction. Furthermore, Figure 3.50 shows that the spectrum of Tat_{1-72} and TAR in solution (Figure 3.49 – brown and Figure 3.50 - brown) is clearly quite different from the spectrum obtained by adding the spectrum of Tat_{1-72} in DDM to the spectrum of TAR in DDM (Figure 3.50 – green) and that a significant conformational change has occurred as a result of the interaction. From the peak positions in the spectrum it appears as if TAR has induced a helical conformation in Tat_{1-72}, however the spectrum cannot be deconvoluted because of the TAR signals.

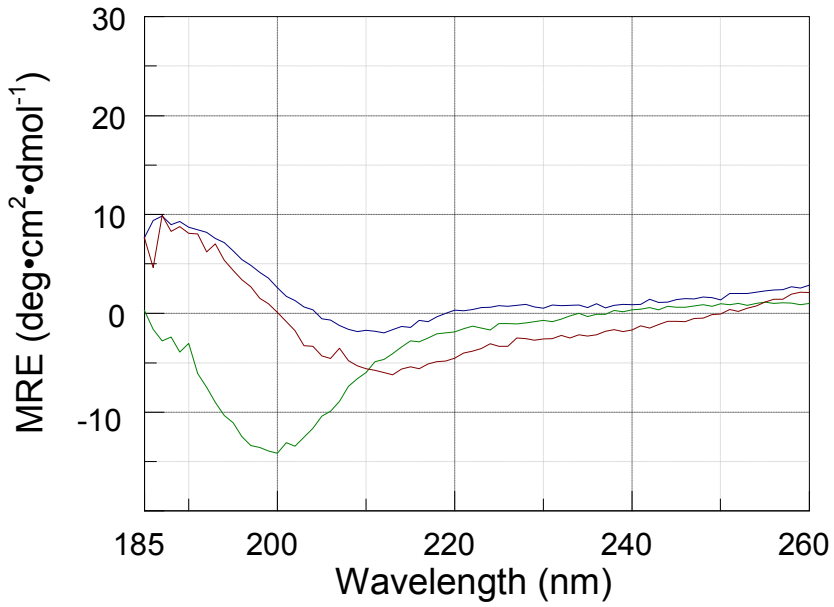

Figure 3.49 CD spectra of 13 µM Tat_{1-72}, 10 mM DDM, 2 MEQ Zn (II) at pH 4.0 (green) and 13 µM TAR, 10 mM DDM, 2 MEQ Zn (II) at pH 4.0 (blue); 13 µM Tat_{1-72} + 13 µM TAR, 10 mM DDM, 2 MEQ Zn (II) at pH 4.0 (brown).

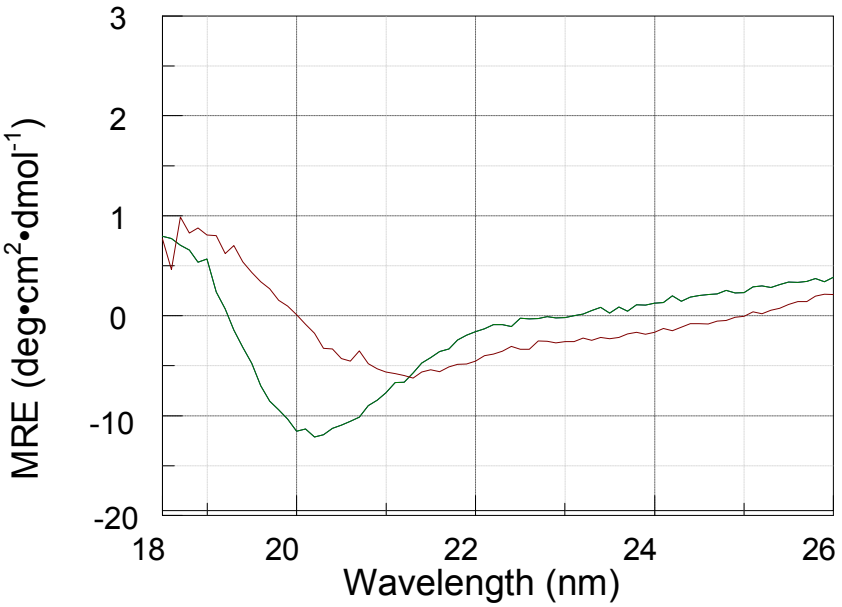

Figure 3.50 The additive spectra of 13 µM Tat_{1-72} alone in 10 mM DDM 2 MEQ Zn(II) and 13 µM TAR alone in DDM, 2 MEQ Zn(II) (green) compared to the actual spectrum of 13 µM Tat_{1-72} and 13 µM TAR in 10 mM DDM, 2 MEQ Zn(II) (brown).

3.11 Congo Red Binding Assay

To determine if Tat_{1-72} aggregates can form the cross-beta amyloid fibril structure, Congo Red Binding assays were performed on the protein in conditions where it is predominantly a random coil, a mixture of α-helix and random coil, or a mixture of β-sheet and random coil. The most successful experiment involved a mixture of β-sheet and random coil conformations, and were induced by pre-equilibrating Tat_{1-72} for 24 hours in 50 mM citrate, 100 mM NaCl, at pH 5.35 followed by the addition and 40 minute equilibration of 10 μL of 375 μM CR to 20 μM Tat_{1-72}. The shift in the absorbance of CR to longer wavelengths is shown in Figure 3.51. Compare this figure to Figure 1.12 in the Introduction. There is clearly an interaction of Tat_{1-72} with CR as can be seen from the hyperchromic shift in CR absorbance. After a 40 minute equilibration period of CR with Tat_{1-72}, the characteristic red colour of the dye was observed in solution, but no visible precipitate was observed. After a period of 24 hours, there was visible evidence of a red precipitate settling out of solution.

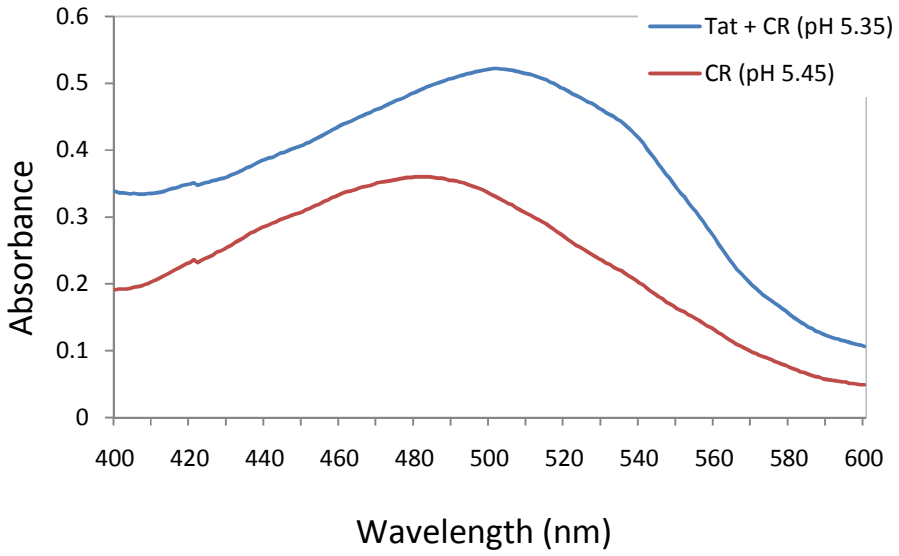

Figure 3.51 Congo red interaction assay in 50 mM citrate, 100 mM NaCl, pH 5.35. The ability of Tat_{1-72} to form cross-beta amyloid-like fibrils was probed by affinity for Congo Red at pH 5.35. Blue: 0.14 µM CR + 20 µM Tat_{1-72} (40 minute equilibration), Red: 0.14 µM CR (initial absorbance). A hyperchromic shift is observed when CR is combined with Tat_{1-72} that is characteristic of amyloid fibril.

3.12 Polarized Light Microscopy

To determine if Tat_{1-72} aggregates contain amyloid structure, polarized light microscopy was applied. Figure 3.52 shows a Tat_{1-72} aggregate stained with CR observed under a polarizing microscope without cross-polarization. This precipitate was prepared with Tat_{1-72} predominantly in a β-sheet/random coil conformation. The centre of the aggregate appears dark, while the edges of the aggregate are stained red. Figure 3.53 is the same preparation of Tat_{1-72} and CR observed with a 90° cross-polarizer in place. Around the edges of the precipitate a yellow-green birefringence is observed in Figure 3.53, but this is not observed in the centre of the precipitate.

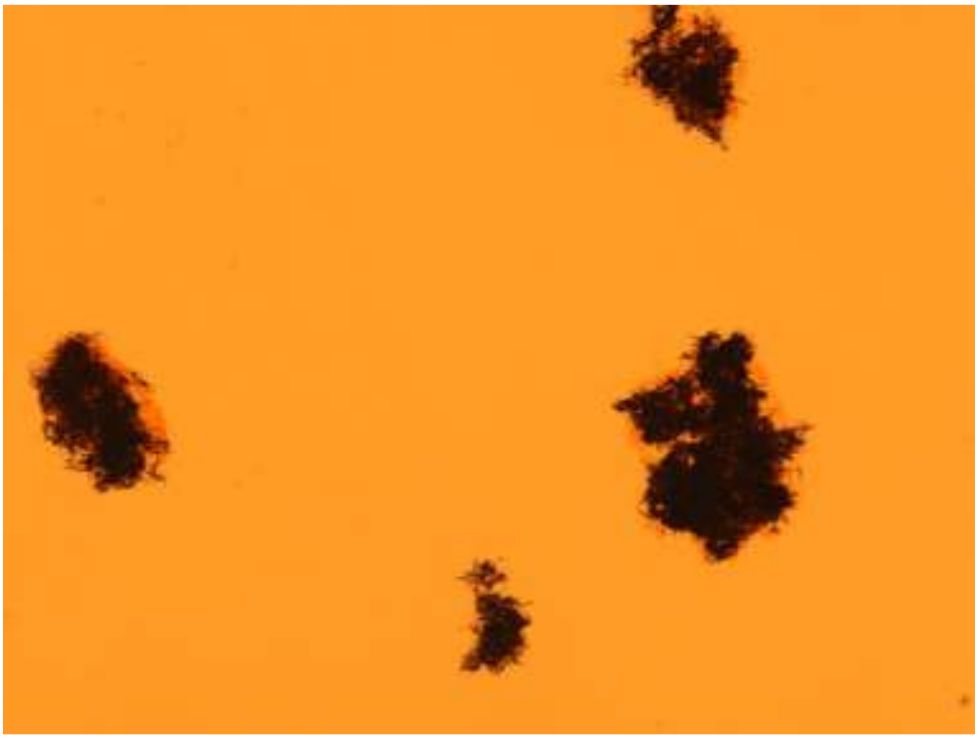

Figure 3.52 $Tat_{1\text{-}72}$ precipitate stained with Congo Red observed by unpolarized light.

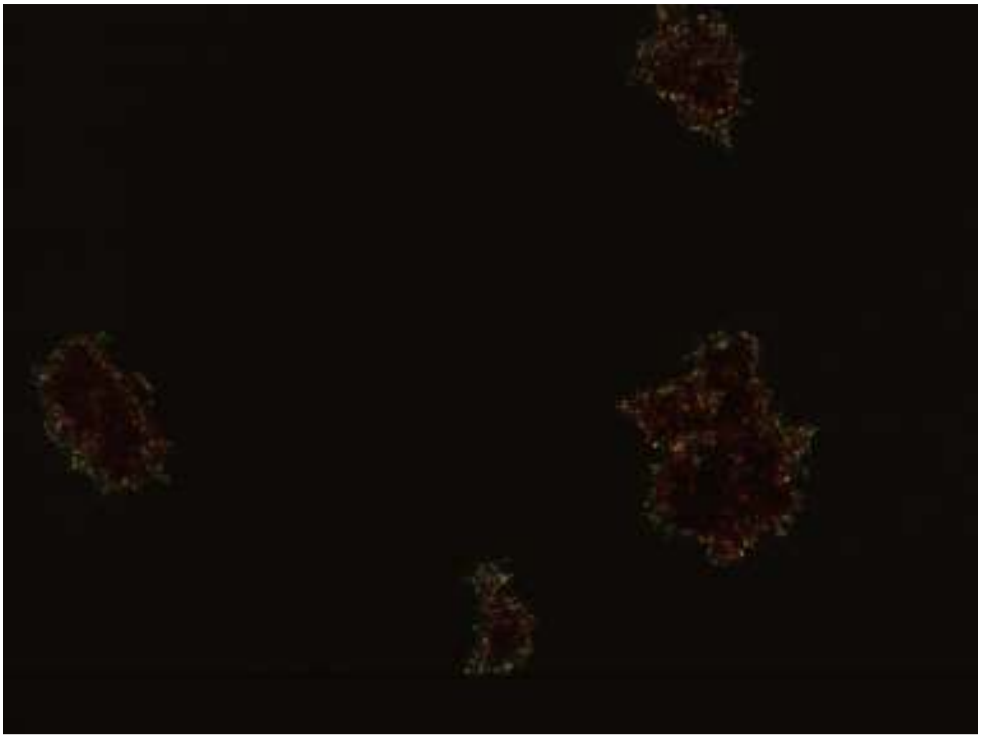

Figure 3.53 $Tat_{1\text{-}72}$ precipitate stained with Cong Red observed by cross-polarized (90°) light.

The edges of the aggregates display green birefringence.

3.13 Amyloid Specific Sequence Determination Using Web-based Algorithm "WALTZ"

The results of the sequence submission to "WALTZ" (http://waltz.vub.ac.be/submit.cgi) of the 92 residue, His-tagged Tat_{1-72} protein indicate that an amyloidegenic region exists between residues 55 – 64. Figure 3.54 indicates that the amyloidegenic includes the last three residues of the cysteine-rich region and seven residues of the hydrophobic core region.

His-Tag
MGSSHHHHHHSSGLVPRGSH

Proline-rich Region
MEPVDPRLEPWKHPGSQPKTA

Cysteine-rich Region
CTNCYCKKCCFHCQVC

Hydrophobic core Region
FITKALGISYG

Arginine-rich Region
RKKRRQRRRPP

Glutamine-rich Region
QGSQTHQVSLSKQ

Figure 3.54 The amyloidegenic region of His-tagged, Tat_{1-72} determined by the "WALTZ" algorithm (red).

Chapter 4

4. Discussion

4.1 Protein Expression and Purification

Tat_{1-72} purification was done in the presence of a high concentration of denaturant (5 M guanidine) and in strongly reducing conditions (10 mM TCEP). The combination of guanidine and TCEP has been shown to be effective at increasing yields of monomeric, soluble Tat_{1-72} protein and shows good compatibility with cobalt metal affinity resin[272]. The strong reducing agent TCEP effectively prevents intermolecular covalent cross-linking by disulfide bond formation[273]. Guanidine promotes soluble, monomeric Tat_{1-72} by disrupting the interactions that lead to precipitation. Tat_{1-72} can self-associate possibly by interactions between its acidic and basic segments. Guanidine also disrupts interactions between the positively charged protein and anionic cellular components such as other proteins, negatively charged phospholipid headgroups, DNA, and RNA. A high concentration of guanidine does not prevent the interaction of de-protonated histidine residues of the His-tag with the cobalt resin near neutral pH, and does not prevent the protonation of histidine at pH values below 6. In the absence of guanidine the protein elutes slowly in a large volume of elutant[272]. This suggests that guanidine can also disrupt the interactions between Tat_{1-72} and components of the metal affinity resin. Hydrophobic amino acid solubility is promoted in the presence of Gdn-HCl and likely prevents Tat_{1-72} self-association by hydrophobic interactions[197]. Presumably, the denaturing conditions would have little effect on Tat_{1-72} structure because Tat_{1-72} is an intrinsically disordered protein.

Tat was eluted by pouring the pH 4 elution buffer over the column while collecting 8 X 1 mL fractions. Because Tat_{1-72} was found to precipitate when the undiluted column fractions are combined into one dialysis bag; the dialysis protocol was modified from the original published procedure[20]. The first 3 fractions were usually the most concentrated and were diluted 4 times with elution buffer to a total volume of 4 ml each. The remaining 5 fractions were pooled together. The pooled fractions were added to 4 dialysis bags of approximately 6 mL each. When four dialysis bags of 6 mL each are prepared, the concentration of Tat_{1-72} remains low enough to prevent precipitation as guanidine and TCEP are gradually removed and replaced by acetate buffer of pH 3.5.

4.2 Ultraviolet Absorption Spectroscopy

A UV absorption spectrum of Tat_{1-72} showing little evidence of light scattering is obtained in 10 mM acetate buffer at pH 4 prepared from glacial acetic acid, with no additional salts, for a number of reasons (Figure 3.1). At pH 4 it was determined that Tat_{1-72} is monomeric and reduced without the need to add reducing agent to solution[238]. Oxidation of cysteine requires a deprotonated thiol and at pH 4 the thiols are fully protonated[20, 274]. At a 10 mM concentration of acetate buffer prepared from glacial acetic acid, a low concentration of buffer is present in solution while still maintaining an efficient buffering capacity. The addition of salts and high ionic strength have been shown to influence the conformations of IDPs[219], and in the case of Tat_{1-72}, may promote its aggregation. Observing UV far spectra of Tat_{1-72} in 10 mM acetate at pH 4 is a good method for measuring the concentration of Tat_{1-72} in solution.

Comparing the UV spectrum of Tat_{1-72} in Figure 3.1 to the UV spectrum of tryptophan in Figure 3.2 we note some differences, specifically in the trough region between 240 – 260 nm and in the region above 300 nm. In the spectrum of Tat_{1-72}, the trough region absorbs slightly more than the trough region of Trp relative to the most intense absorbance at 280 nm, and the UV absorbance of Tat_{1-72} does not decay to zero above 300 nm. Both effects may be due to aggregation of a very small population of Tat_{1-72} that leads to light scattering. However, some of the absorption in the protein spectrum near 260 nm is due to the presence of two phenylalanine and two tyrosine residues in Tat_{1-72}.

4.3 Circular Dichroism Spectroscopy

In proteins, the main chromophoric contributor to the CD signal in the far UV region is the peptide bond, and the signal is influenced by asymmetry in its environment. For example, left and right-handed α-helices have opposite ellipticity. The CD signal in a protein is the ensemble average over time of the signals from amide bonds along the protein backbone. Peptide bonds that adopt specific phi-psi angles give rise to unique CD signals for each major secondary structure including α-helices, β-sheet, hydrogen bond stabilized turns, and random coils (Figure 1.14). Intrinsically disordered, or natively unfolded proteins give rise to unique CD signals not because of their limited dihedral angles, but because their flexibility allows them to sample a wide range of dihedral angles. It is for this reason that structure determination by CD spectroscopy gives insights into secondary structure on a global, population averaged scale. Global secondary structure determination by CD spectral deconvolution will be informative by providing the relative proportions of each secondary structure in solution, averaged over time.

At pH 4, Tat_{1-72} is intrinsically disordered[20]. This has been re-confirmed by the agreement between the present CD and NMR studies and the previous work by Dr. Shaheen Shojania[238] and Thach Vo[172]. In the present work spectral deconvolutions using the Selcon3, CONTINLL, and CDSSTR algorithms were reported as an average (henceforth referred to as the spectral deconvolution average). The spectral deconvolution average determined that the secondary structure of Tat_{1-72} in 10 mM acetate, with 10 mM TCEP at pH 4 is 57 % disordered, 21 % turn, 14 % β-strand, and 8 % α-helical. Tat_{1-72} displays a strong negative band at 190 nm and a weaker negative band at 220 nm, characteristic of disordered proteins[230]. This result is in close agreement with other studies of HIV-1 Tat protein variants (86-101 residues) [124, 169, 171, 275, 276]. Structural determination of the Tat_{1-86} Mal variant by spectral deconvolution showed that the protein is composed of 43 % disordered, 30 % β-turn, 22 % β-sheet, and 5 % α-helix at pH 4.5 and agrees reasonably well with the present results[124]. The presence of exon-2 residues may be the explanation for the greater proportion of β-sheet and turn structures in the Tat Mal variant.

There may be ambiguity regarding the distribution of structure within Tat between turn and β-strand structure as their representative CD spectra are quite similar[277], with a minor negative band at 215 nm and a major positive band at 195 nm. The distribution of these two structures in secondary structure determination by spectral deconvolution is dependent on the quality of the data sets of structures on which the deconvolutions are based. In either case, the CD spectrum of Tat in 10 mM acetate, 10 mM TCEP at pH 4 is mainly disordered.

4.4 Tat Solution Stability

As we have seen, Tat_{1-72} was determined to be a random coil at pH 4. It is also important to determine if the protein is intrinsically disordered at pH 7. The goal of these studies was to find compatible buffers and solution additives to solubilize Tat_{1-72} in conditions approaching neutral pH. A physiological condition in which Tat_{1-72} stability is promoted would allow for an environment in which the high affinity interaction between Tat and TAR could be studied in solution. This, among other Tat interactions that have been described in the literature, must, at least in part, be sensitive to the same factors that are known to affect intrinsically disordered proteins *in vitro*, such as temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding[219]. The conformational changes of His-tagged Tat_{1-72} induced by the presence of organic solvents, counter-ions, and detergent may present routes to conformational changes that imitate those involved in ligand recognition and thereby provide insight into them.

4.4.1 Buffers

Tat solubility was investigated in the absence of reducing agent with the hydrogen ion buffers MES, HEPES[278] and citrate. Tat_{1-72} showed better stability in HEPES compared to MES (stable in HEPES at pH 5.8 compared to pH 4.0 in MES). It is difficult to explain why Tat_{1-72} is more soluble in HEPES as opposed to MES. Both are zwitterionic at neutral pH and positively charged at pH 4. HEPES contains an additional hydrogen bonding hydroxyl group that may promote Tat_{1-72} stability. The CD data shown in Figure 3.5 and Table 3.3 indicate that Tat_{1-72} in

HEPES at pH 4.0 exists in a mainly disordered conformation that is quite similar to Tat_{1-72} in acetate with TCEP at pH 4 (Figure 3.3, Table 3.1). The secondary structure motifs of Tat_{1-72} in HEPES at pH 4.0 and Tat_{1-72} in acetate and TCEP at pH 4.0, respectively are: 48 %, and 57 % disordered, 21 % and 21 % turn, 20 % and 14 % β-strand, 12 % and 8 % α-helical.

Citric acid was chosen as a buffering agent due to its functional group similarity to acetate, and because of its role as a natural reducing agent, often used in foods as a preservative. Citric acid is an important molecule in the metabolism of most cells, and is a vital Krebs cycle intermediate. Biologically and commercially, citric acid is known as an anti-oxidant and preservative, and this quality was hypothesized to impart stability to Tat in solution, possibly preventing disulphide bridging in the absence of TCEP. The molecule contains three ionizable groups, making it a useful buffer over an extended pH range. However, citric acid was not able to solubilize Tat_{1-72} at pH 4 for a period of 24 hours. At pH 4.0, citric acid exists as a mixture of triprotic, diprotic, monoprotic and unprotonated acids owing to pKa's of 3.15, 4.75 and 5.40[279]. One possibility is that multiply-charged buffering species might cause the precipitation of the basic Tat_{1-72} protein. Another possibility is that long-term solubility may be influenced by trace amounts of Tat_{1-72} oligomer present in different amounts in different protein preparations. This is the case with β-amyloid[180, 185].

4.4.2 Buffer, Detergent, and TCEP Combinations

In the absence of reducing agent Tat_{1-72} was soluble up to about pH 5.7 for 24 hours in HEPES buffer with CHAPS above the CMC. The presence of the zwitterionic CHAPS detergent

does not increase Tat_{1-72} solubility in HEPES because Tat_{1-72} was stable at pH 5.8 in HEPES without CHAPS for 24 hours (Figure 3.4). However, this experiment demonstrated that Tat_{1-72} is compatible with CHAPS at pH 5.7.

Tat_{1-72} showed compatibility with the cationic detergent DTAB in HEPES at pH 5.0 and in citrate at pH 5.3 for 24 hours each. Tat_{1-72} also showed better solubility in the anionic detergent SDS when buffered in HEPES (Figure 3.9); Tat_{1-72} was stable for 24 hours at pH 6.67 in HEPES-buffered SDS. This suggests that the ionic properties of the detergent head group can affect Tat_{1-72} aggregation. Tat_{1-72} is a basic protein with a net charge of +15 at neutral pH. The anionic sulfate head group of SDS may interact with the basic Tat_{1-72} protein, and according to Figure 3.10, elicits a shift to α-helical conformation. The increased solubility might also be due to the change in conformation. It is uncertain from this study which region of Tat_{1-72} is undergoing conformational change but the results from the TFE study suggest that the formation of an α-helix in the arginine-rich region may occur by the interaction of basic side chains with the anionic head groups.

Detergents can have varying effects on protein structure. Packing of the hydrophobic cores of globular proteins can be disturbed by side-chain association with long, non-polar detergent tails. Ionic interactions between detergent headgroups and protein surfaces may disrupt protein hydration spheres and potentially disrupt intrapeptide interactions, which occur by hydrogen bonding and Van der Waals forces. In other cases, low concentrations of detergent may have little or no effect on protein structure.

The 24 hour stability of Tat_{1-72} with SDS in the presence of reducing agent is not surprising (Figure 3.11 and Figure 3.12). The negatively charged sulfate head groups might interact with Tat_{1-72} owing to the net positive charge of the protein. It might also be proposed that the long alkyl groups of the detergent molecules interact with the hydrophobic side-chains of Tat_{1-72}, offering better solubilization of the protein by hindering inter-protein interactions. However, the poor solubility of Tat_{1-72} in other detergents in the presence of reducing agent suggests that Tat_{1-72} aggregation involves not only hydrophobic interactions, but also ionic, polar and cysteine-cysteine interactions. Precipitation in the non-ionic detergent NDBM in the presence of reducing agent also shows that inter-protein hydrophobic interactions are not solely responsible for aggregation. The poor solubility of Tat_{1-72} in cationic DTAB, but good solubility in anionic SDS suggests that the mean net charge on the protein plays a key role in self-associative pathways that lead to aggregation. The sequence-specific properties like high-net charge and short regions of hydrophobicity that favour protein disorder *in vivo* hinders the study of Tat *in vitro*.

We consider several possible interactions between Tat_{1-72} and SDS. $Tat1_{-72}$ may be largely inserted into the core of the micelle, partially inserted into the micelle (tethered by the core domain), interacting with the hydrophilic groups of the micelle exposed to solution, possibly extended over the surface of the micelle or likely, in an equilibrium with all of these states. ^{1}H-^{15}N HSQC spectra of Tat_{1-72} in concentrations of SDS above the CMC (not shown) displayed extensive line broadening and peak loss, potentially caused by association with the large, slow tumbling SDS micelles, or conformational exchange.

Figure 3.10 shows the CD spectrum of Tat_{1-72} in SDS. The characteristic 190 nm, 208 nm, and 222 nm peaks in the CD spectrum, indicate a protein with substantial α-helical content. Secondary structure determination by spectral deconvolution (Table 3.4) shows that with the addition of SDS, disordered content decreases from 40 % to 35 %, turn content decreases from 24 % to 20.0 %, β-strand content decreases from 27 % to 22 %, while α-helical content increases from 9 % to 23 %. It appears as if secondary structure is being converted from disordered, turn, and β-sheet to α-helical structure. In a study of HIV-1 Tat peptides corresponding to the different regions of Tat, the presence of SDS could not induce α-helical tendency in any of the Tat fragments[280] implying that the SDS-induced structure of Tat_{1-72} is a process that requires cooperation among different regions of the protein. This insight, together with the spectral deconvolution implies that α-helical conversion seems to be occurring non-specifically and possibly over the entire span of the protein and may be a clue to the mechanism for the cellular membrane translocation activity of Tat.

SDS is known to induce helical structure on disordered or partially disordered proteins and this is also the case with Tat_{1-72} based on the Tat_{1-72}/SDS CD spectrum (Figure 3.10)[281]. This effect is in contrast to the effect of SDS on well-folded proteins; SDS is commonly used as a denaturant during SDS-PAGE, although local regions of α-helicity may be promoted in unfolded proteins. In SDS-PAGE, SDS is known to denature proteins and weaken tertiary contacts. The opposite effect on Tat_{1-72} implies that disorder is intrinsically coded in the Tat_{1-72} sequence and is important in the binding induced folding during interactions with partner proteins and TAR RNA. Similar types of folding-induced interactions have been observed with phospholipids and the effect of SDS on Tat_{1-72} suggests that cellular membranes composed of

anionic phospholipids could induce structure in Tat_{1-72}. In support of this, it has been shown that cationic peptides interact preferentially with anionic over zwitterionic mammalian membranes[282]. In addition, models have been developed for membrane dependent processes in the disease progression of Alzheimer's and Parkinson's Diseases that involve helical intermediates of unfolded proteins[181, 182]. These cases support the idea that protein interactions with the polar head groups, and the hydrophilic surface of the bilayer are essential to the process of inducing structure in unfolded proteins.

The conformations that are induced in Tat_{1-72} at concentrations of SDS below the CMC should not rule out the existence of a micelle, composed of detergent and multiple Tat_{1-72} monomers. It was shown by isothermal titration calorimetry that in the presence of α-Synuclein, the concentrations of SDS needed to form a micelle were greatly decreased [283]. To explain this, detergent interactions at concentrations below their CMC that would support micelle formation would need to be supplemented by numerous protein detergent interactions. The assembly is not a homogenous detergent micelle but rather a protein-detergent aggregate.

4.4.3 Summary of Tat Aggregation

Protein aggregation is a non-specific term that can be used to describe several different types of interactions. Aggregates may be soluble or insoluble, covalent or noncovalent, reversible or irreversible, and native or denatured. Tat_{1-72} is intrinsically disordered, therefore Tat_{1-72} aggregates likely involve more or less disordered polypeptides. For all studies, Tat_{1-72} was deemed soluble if it remained within solution after repeated centrifugation at 13 000 X g and

filtration through a 0.2 micron cellulose membrane and displayed a UV absorption spectrum relatively free of light scattering.

Ionic interactions between the acidic, proline-rich region and the basic, arginine-rich region of Tat_{1-72} theoretically can occur in a soluble state. Therefore, ionic interactions between soluble Tat_{1-72} monomers or the interaction between basic Tat_{1-72} and anionic TAR could be the initial events that lead to rapid, irreversible, insoluble aggregation. Detergents can help solubilize Tat_{1-72}, but depending on their head group charge, vary in their efficiency at keeping Tat_{1-72} soluble at pH values near 7. Tat_{1-72} in the anionic detergent SDS without TCEP is soluble at pH values closer to 7 than other detergents in this study.

The presence of anionic counter-ions with reducing agent did not prevent aggregation of Tat_{1-72}, suggesting multimerization mediated by the hydrophobic core region. The solubility of Tat_{1-72} at neutral pH occurred when cysteine oxidation was controlled by reducing agent or interaction with Zn^{2+}, and when ionic and hydrophobic Tat_{1-72} self-associative interactions were limited, as was the case for Tat_{1-72} in SDS with TCEP. For Tat_{1-72} in SDS with TCEP the sulfate head groups likely prevented ionic interactions between Tat_{1-72} monomers, cysteine cross-linking was prevented by TCEP, and hydrophobic interactions between Tat_{1-72} monomers were prevented by the aliphatic group of the detergent.

4.4.4 Tat Interaction with TAR

Figure 3.50 shows that the CD spectrum of Tat_{1-72} in DDM added to the spectrum of TAR in DDM (green) and the spectrum of Tat_{1-72} and TAR together in DDM (brown) are clearly different from each other. The positive ellipticity at 190 nm and the negative ellipticity at 222 nm of Tat_{1-72} and TAR together in DDM suggests that a helix is induced in Tat_{1-72} upon binding to TAR. This is an example of ligand binding causing an intrinsically disordered domain to fold, and supports the coupling of a binding induced structural transition; the entropic loss associated with the partial folding of an intrinsically disordered protein may be offset by ligand binding and/or solvent release.

A peptide corresponding to the basic region of Tat has been shown to interact with TAR with a nanomolar affinity, $K_D = 6 \text{ X } 10^{-9}$, indicating the strength of this interaction[*(284)*]. Interestingly, the TAR affinity of a larger fragment of Tat encompassiong two cysteine residues of the cysteine-rich region, the hydrophobic core region, and the arginine-rich region, the Tat/TAR affinity is a whole order of magnitude less[*(285)*]. Mutations to the cysteines did not result in a change to the interaction affinity, suggesting that they are not involved in the binding. The nanomolar affinity of the arginine-rich region of Tat for TAR does not ensure a soluble complex between Tat_{1-72} and TAR, but rather, causes rapid precipitation, even at micromolar concentrations. Titrating Tat_{1-72} into a dilute solution of TAR or *vice versa* also resulted in precipitation. In a study of the interaction of a Tat peptide consisting of the arginine- and glutamine-rich regions (the carboxy-terminal of the first exon) with the 27 base stem-loop TAR, complexation above a concentration of 0.6 mM resulted in precipitation[*(286)*]. Non-specific ionic

interactions of the highly negatively charged ribonucleic acid and the basic Tat peptide are likely the explanation. In this study of exon-1 Tat_{1-72}, precipitation is likely enhanced by inter-protein hydrophobic interactions between the core domains and an oxidation prone cysteine region. Non-ionic DDM was employed to solubilize Tat_{1-72} in the presence of TAR, but a non-ionic detergent was not sufficient to solubilize Tat_{1-72} with reducing agent in the absence of TAR, suggesting that Tat_{1-72} and Tat_{1-72}/TAR solubility is ionically influenced. Covalent cross-linking between cysteine residues appears to have been prevented by the presence of 2 MEQ's of Zn(II), whereas the anionic reducing agent TCEP resulted in precipitation of Tat_{1-72}/TAR with DDM. Anionic SDS was sufficient to solubilize Tat_{1-72} with reducing agent in the absence of TAR, but not sufficient to prevent precipitation of Tat_{1-72} in the presence of TAR. This is also strong evidence that a charge interaction occurs between Tat_{1-72} and TAR, and the stability of this is disrupted by the presence of anionic SDS. Thus, it is likely that SDS competes with TAR for an interaction with Tat_{1-72}. Due to the charge complementary of Tat_{1-72} and TAR, a non-ionic detergent is better suited to prevent hydrophobic interactions between Tat_{1-72} monomers but permit the ionic interaction between Tat_{1-72} and TAR. An early conformational study of TAR showed that arginine (and guanidine) binds specifically to TAR, suggesting that the interaction between the arginine-rich region of Tat_{1-72} and TAR is specific, as other basic residues showed no specific affinity[287]. Although the arginine residues have been implicated in specificity, other basic residues must enhance affinity between Tat and the negatively charged phosphate groups of TAR by complementary charge interactions.

4.4.5 Structural Implications

Tat_{1-72} solubility studies have shown that in the absence of TAR, Tat_{1-72} solubility is influenced by the hydrophobic region of Tat_{1-72}, the net charge of the protein, and the cysteine-rich region. The types of self-associative interactions that need to be accounted for when solubilizing Tat_{1-72} *in vitro* reflect the manner in which Tat interacts with P-TEFb and TAR *in vivo*, and may also give insight into the mechanisms of Tat_{1-72} pathogenicity.

The hydrophobic region of Tat_{1-72} has been suggested to cause self-association and *in vitro* it has been shown that this region becomes closely associated with P-TEFb and has been implicated in Tat related pathogenicity. In a crystallization study, the hydrophobic region of HIV-1 Tat packs against a hydrophobic patch of Cyclin T1 residues Phe176, Leu184, Thr248 and Pro249[70]. The total surface buried during the interaction between HIV-1 Tat and P-TEFb is 3,499 $Å^2$, 88 % of which is on Cyclin T1. This is about twice the average value for most stable protein-protein interactions[288]. Related to pathogenicity, the Tat_{1-86} Mal variant binds to tubulin/microtubules through a four-amino-acid subdomain consisting of residues 36 - 39 of its conserved hydrophobic core region, leading to the alteration of microtubule dynamics and activation of a mitochondria-dependent apoptotic pathway[121].

The induction of helical structure by SDS is related to the helical propensity of Tat, and has been shown to occur *in vitro*. The formation of an α-helix in the basic, arginine-rich region of EIAV Tat in the presence of TAR has been shown by crystallization studies[67]. The basic, RNA recognition motif of EIAV Tat that recognizes TAR consists of residues 49 – 57 of the

arginine-rich region, and residues 50 – 58 form an α-helix when interacting with TAR. Helical structure has also been suggested in the glutamine-rich region of the Tat_{1-86} Mal variant[275], and a short 3_{10}-helix in the cysteine-rich region was observed in a crystal structure formed between HIV-1 Tat and P-TEFb[70].

The oxidation of the cysteine-rich region near neutral pH was shown to be prevented in some cases by the use of TCEP with SDS. The reduced state of the cysteine residues of Tat is important for this region's role as a zinc finger. Referring to the aforementioned crystallization study, the cysteine-rich region forms a zinc bridged interaction with Cys261 of Cyclin T1 in which all of the cysteine residues are reduced[70].

4.5 Tat_{1-72} Can Exist in a β-strand Conformation

In 50 mM citrate solution and 100 mM NaCl, Tat_{1-72} remained soluble at pH 6.2 for 24 hours in the absence of reducing agent (Figure 3.14). The CD spectral deconvolution results in Table 3.5 indicate that the secondary structure of Tat_{1-72} in 50 mM citrate and 100 mM NaCl at pH 5.5 is significantly different than Tat_{1-72} in 10 mM acetate, 10 mM TCEP at pH 4. In the presence of 100 mM sodium chloride, disordered content decreases from 57 % to 33 %, β-strand content increases from 14 % to 37 %, turn content is similar (21 % and 22 %), and α-helical content is similar (8 % and 7 %). Thus, it appears that disordered polypeptide is being converted to β-strand. The UV absorption spectra of Tat_{1-72} in 50 mM citrate and 100 mM NaCl at pH 5.4 and 6.2 (Figure 3.14) show significant light scattering above 300 nm which is an indication of aggregation. The sample was micro-centrfuged with no visible pellet forming, suggesting that

the Tat_{1-72} exists in a multimeric, soluble form in these conditions. The only change in structure is the conversion of disordered poly-peptide to β-strand. Perhaps high salt encourages hydrophobic interactions and β-sheet formation. Sodium has been shown to be a promoter of hydrophobic interactions when in solution as an ion, referred to as a "salting out" effect[289].

4.5.1 Amyloid properties of Tat_{1-72}

The results of applying the web-based algorithm "WALTZ" to the amino acid sequence of Tat_{1-72} indicate that His-tagged Tat_{1-72} does have an amyloidegenic region at pH 7. Residues 55 – 64 of the His-tagged Tat_{1-72} does have an amyloidegenic region at pH 7. Residues 55 – 64 of the Tat_{1-72} protein, beginning at Val-55 and Cys-56 of the cysteine-rich region and ending at Gly-64 of the hydrophobic core region were determined to be amyloid-prone. With the exception of the first two residues, this sequence spans more than half of the hydrophobic core region.

The UV spectrum of Tat_{1-72} in the presence of sodium chloride shown in Figure 3.14 suggests that the protein exists in an oligomerized state as evident by light scattering above 300 nm, but Tat_{1-72} in 20 mM citrate and 100 mM NaCl does not show any evidence of time dependent precipitation. CD spectra suggest that β-strand structure is increased in the presence of sodium chloride (Figure 3.15, Figure 3.16, Table 3.5, and Table 3.6). The UV spectrum of Congo Red shown in Figure 3.51 shows a redshift at 540 nm upon addition of Tat_{1-72} typical of CR-stained amyloid. Figure 3.53 shows a Tat_{1-72} aggregate stained with CR under a cross-polarizer. The yellow-green birefringence is evidence of amyloid-like interactions at the edges

of the aggregates that is the trademark of Congo Red interaction with amyloid tissue. Elongated fibrils of amyloidosis are formed by the arrangement of disordered precursors that are stabilized by inter-protein hydrogen bonding, forming an extended network of β-sheets[180]. The mechanism of Congo Red interacting with amyloid is likely an intercalation between the β-sheets, where a specific, regular orientation with the fibril results in yellow-green birefringence of the dye[290]. Tat_{1-72} association with Congo Red resulted in precipitation of all Tat_{1-72} samples, regardless of the presence of sodium chloride. As part of the Hofmeister series, sodium affects protein solubility by promoting hydrophobic interactions[204, 289]. Only samples containing sodium in the buffer showed birefringence when stained with Congo Red, although the precipitation of Tat_{1-72} by CR may have occurred due to non-specific charge attractions between the dye and Tat_{1-72}. The β-sheet secondary structure determined by spectral deconvolution, the red shift and birefringence observed with binding to CR, and the significant UV light scattering above 300 nm suggests that Tat_{1-72} can form an amyloid fibril with the association possibly mediated by its hydrophobic region.

4.6 The Effects of TFE on Tat_{1-72}

4.6.1 The Solubility of Tat_{1-72} in TFE

Tat_{1-72} showed good solubility in TFE in the absence of reducing agent at pH 4.05 (Figure 3.17). However, raising the pH of Tat_{1-72} in 30% TFE in the presence and absence of TCEP resulted in protein precipitation. In the absence of TCEP it is likely that cysteine oxidation contributes to irreversible aggregation and precipitation of the protein. In the presence of TCEP, it seems likely that anionic TCEP can precipitate the basic Tat protein by directly interacting

with it. Only in the presence of SDS was Tat soluble at neutral pH in the presence of TCEP and TFE is unable to prevent the interaction between TCEP and Tat as SDS appears to be able to do. Precipitation of Tat_{1-72} could be avoided however, by adding 2 MEQ's of Zn^{2+} to Tat_{1-72} dissolved in 30 % TFE and then titrating the solution from pH 4.69 to pH 7.2 (Figure 3.26). Presumably Tat_{1-72} is soluble in 30 % TFE in the presence of zinc because zinc prevents oxidative disulphide bond formation and 30 % TFE provides a strong solvation environment for the protein. The nature of the solvent TFE and its influence on the structure of Tat will be discussed in the next sections.

4.6.2 The Structure of Tat_{1-72} in TFE – CD Studies

Addition of TFE induces α-helix formation in Tat_{1-72} according to both the appearance of the CD spectra in Figure 3.18, CD spectral deconvolutions (Figure 3.22), and analysis of NMR chemical shifts (Figure 3.48). Superimposing the CD spectra of Tat dissolved in 0 % - 90 % TFE results in the appearance of an iso-dichroic point at about 203 nm (Figure 3.18). The iso-dichroic point (or isosbestic point) is the wavelength at which the molar absorptivity is the same for all of the protein CD spectra[*(291)*]. The iso-dichroic point in the CD spectra of Tat_{1-72} is evidence for a two-state transition[*(292)*]. The two states are represented by the predominantly random coil and α-helical CD spectra that are observed at 0 % TFE and 90 % TFE, respectively. The simplest interpretation of this result is that Tat_{1-72} exists at all TFE concentrations in an equilibrium between two conformations and that TFE promotes the formation of an α-helical conformation and water promotes the formation of a random-coil conformation[*(293)*].

CD spectral deconvolution confirms that the effect of TFE on Tat_{1-72} secondary structure is a decrease in disordered content, an increase in α-helical content, a relatively smaller decrease of turn content, and a minimal effect on β-strand content (Figure 3.22, Appendix A, *A1*). The majority of the changes in Tat_{1-72} secondary structure occur between 0 % and 30 % TFE (Figure 3.20, Figure 3.21, Table 3.8). The disordered structure decreases 30 % up to 30 % TFE, and only decreases about 10 % throughout the rest of the titration. Helical content increases 30 % up to the addition of 30 % TFE, and only 15 % for the remainder of the titration.

Deconvolution of the CD spectrum in Figure 3.22 suggests that at 30 % TFE, about 40 % of Tat_{1-72} exists as an α-helix. CD analysis could be implying that 40 % of Tat_{1-72} is occupying a well-defined α-helix, or that different regions of Tat_{1-72} in 30 % TFE are undergoing dynamic exchange between random coil and α-helix conformations, and the α-helix conformation is being occupied on average 40 % of the time. This is a limitation of CD studies of protein backbone conformation; CD spectral deconvolution determines a global, population-averaged conformation. However, the minimal effect of TFE on turn and β-strand composition during the TFE titration implies that these regions are not prone to adopting a helical conformation, and therefore must be separating regions adopting helicity. Also, the large increase in helical conformation and the correspondingly large decrease in disordered conformation between 0 % and 30 % TFE, supports the view that certain regions of the protein are prone to conversion from a random coil conformation to an α-helix.

4.6.3 The Structure of Tat_{1-72} in TFE – NMR Studies

To determine if certain regions of Tat_{1-72} are more or less prone to forming an α-helix HSQC NMR spectroscopy was used because it permits the observation of structural changes at individual residues as long as they are observable in the NMR spectrum and assigned. One interpretation of the CD spectral deconvolution of Tat in 30 % TFE is that 40 % of the protein or 37 of the 92 residues adopt a helical secondary structure. From the chemical shift indexing of $^{1}H^{N}$ resonances relative to random coil values, 32 residues are identified that constitute regions of helicity in reasonable agreement with the CD data. Parenthetically, it is clear that the $^{1}H^{N}$ resonances of Tat_{1-72} are more sensitive to changes in secondary structure than the ^{15}N resonances as has been shown in the literature[257].

The three regions of Tat_{1-72} that show the greatest upfield change in chemical shift with the addition of TFE are the hydrophobic core (-0.155), the Arg-rich region (-0.137), and the Pro-rich region (-0.125) (Table 3.13). It is not surprising that the hydrophobic core forms an α-helix in the presence of TFE. Not only do polypeptide backbone hydrogen bonds become favoured as water is excluded by the strong hydrogen bond donor TFE, but the aliphatic side chains could form favourable interactions with less polar TFE further encouraging intramolecular H-bonding. In the crystal structure of Tat complexed with human pTEFb a short α-helix is observed to extend from the end of the final zinc co-ordination site in the cysteine-rich region into the core region (Figure 1.9, B). In the EIAV Tat – Cyclin T1 – TAR crystal structure a helix extends from the last residue in the core region at Leu-48 to the end of the Arginine-rich basic segment at Ile-59 (Figure 1.9, A). Since these two helices extend over two sequence regions of Tat this suggests

another way of analyzing the chemical shift perturbation data which is to identify segments of high helical propensity in 30% TFE (Figure 3.48) separated by regions of low helical propensity. This contrasts with the analysis above that focused on the changes in chemical shift of the sequence regions of Tat induced by TFE. Analyzed this way, residues 40-51 of the cysteine-rich region have the highest mean upfield shift from random coil values (0.384) indicating that this region has the highest helical propensity. The crystallographic study of HIV-1 Tat in complex with P-TEFb showed that a short 3_{10}-helix is formed in the homologous region by the residues separating the two zinc co-ordination sites[70] (Figure 1.9, B). In our His-tagged, HIV-1 Tat_{1-72}, the location of this helix would be at Lys-48 and Lys-49 of the cysteine-rich region and it is interesting to note that Lys-48 has the greatest $^{1}H^{N}$ upfield deviation from random coil values. In summary, although the Cys-rich region is not entirely helical in the only known crystal structure it does appear to have the highest helical propensity in the protein.

The region that shows the next highest mean upfield shift from random coil values (0.339) is between residues 64 – 74 (Figure 3.48). This corresponds to the latter 5 residues of the hydrophobic core region and 6 of the 8 non-proline residues of the arginine-rich region. The high helical propensity of the Arg-rich segment is in good accord with the observed helix in the EIAV-Tat- Cyclin T1-TAR crystal structure[67] (Figure 1.9, B). The basic helix is a common motif in nucleic acid/protein interactions, which allows the presentation of basic side-chains to interact within the major grooves of double helix DNA, or in this case, the double-helical stem-loop of TAR.

Finally, the present results are in good agreement with a conformational study of EIAV Tat in 40 % TFE that identified 23 residues that adopted a helical conformation; residues 38 to 43 of the hydrophobic core region and residues 48 to 64 of the arginine-rich region formed helices[176]. In the presence of 40 % TFE, it was shown by NOESY and C^{α} proton shifts that the entire core and basic regions of EIAV Tat form a helix, interrupted by a break at a glycine residue in the core region[300]. Thus, although there are some differences in sequence between the human and equine proteins they share the same helix-forming segments.

4.6.4 The Effect of Zinc on Tat_{1-72} in 30 % TFE

Addition of 0.5 mole equivalents zinc to Tat_{1-72} in 30% TFE at pH 4.28 resulted in a minor loss of ellipticity at all wavelengths suggesting that if zinc binds to Tat in these conditions it causes only a minor rearrangement of the backbone at this pH (Figure 3.24 and Figure 3.25). The spectral deconvolution data for the titration of Tat_{1-72} in 30 % TFE with Zn^{2+} does not indicate any major changes in structure (less than a 1 % change for any motif) (Table 3.10). It is not surprising that interactions between Tat and zinc are weak at pH 4.28 because the Cys thiols are expected to be protonated at this pH. The fact that most of the CD changes occur upon addition of 0.5 mole equivalents of zinc supports earlier studies that suggested that zinc induces the formation of a Tat dimer in water[116]. Previous work by Thach Vo probed the Tat_{1-72}/zinc interaction by ^{1}H-^{15}N HSQC-NMR spectroscopy at pH 4.02 and noted an increase in dynamic exchange of the cysteine residues believed to be interacting with zinc with the addition of 0.5 MEQ Zn^{2+} [274]. It is interesting that a cationic metal can induce Tat self-association since we have noted above that anionic TCEP can do the same. The crystal structure of HIV-1 Tat in

complex with pTEFb shows two well-defined zinc binding sites where one zinc is coordinated by Cys-22, His-33, Cys-34 and Cys-37 and the other is coordinated by Cys-25, Cys-27 and Cys-30 of Tat and Cys 261 of cyclin T1[70] (Figure 1.9, B and C).

As noted above, addition of 2 MEQ of zinc to Tat_{1-72} dissolved in 30% TFE permitted the pH of the solution to be raised from 4.69 to 7.20. The CD spectra shown in Figure 3.26 indicate small losses in ellipticity as the pH is raised and spectral deconvolution suggests a small loss in α-helical structure and a small increase in disordered structure. Since Tat_{1-72} is insoluble at pH 7.2 in 30% TFE in the absence of zinc it is impossible to determine which of the observed structural changes are caused by the pH increase and which may be caused specifically by increased affinity of Tat for zinc at higher pH values. As observed above, the Cys-rich region has one of the highest helix propensities in Tat as indicated by chemical shift changes yet in the crystal structure only a small 3_{10} helix forms connecting the two binding sites[70]. One might speculate here that TFE induces helicity in the Cys-rich region in the absence of zinc but that as the pH rises and zinc binds more tightly the helix in the Cys-rich region unwinds leaving behind only the connecting helix.

4.6.5 The Effect of Temperature on Tat_{1-72} in 20 % TFE

The effect of temperature on the secondary structure of Tat_{1-72} in 20% TFE was measured to determine the stability of the TFE-induced helical structure (Figure 3.23). Interestingly, the greatest disorder is observed for Tat_{1-72} at 5 °C and the greatest order is observed at 25 °C (Table 3.9). Tat_{1-72} is 47 % disordered at 5 °C, compared to 39 % disordered at 25 °C. When the

temperature is raised from 5 °C to 25 °C the helical content increases from 16 % to 28 %. Heat-induced folding has been observed for IDPs, where structure was induced over temperatures from 3 °C to 30-50 °C, and further heating was accompanied by less pronounced effects[165, 296, 297]. The induction of structure with an increase in temperature could occur because of the increase in the strength of hydrophobic interactions and just such an explanation has been used to explain the partial folding of IDPs as temperature is elevated[219]. At 80 °C, Tat_{1-72} is more disordered and less helical than at 25°C suggesting that the thermal energy overcomes the weak interactions favouring secondary structure formation at 25 °C.

4.6.6 Structure-Inducing Effects of TFE

Certain characteristics of TFE such as a lower dielectric constant relative to water and its preferential role as a hydrogen bond donor are pertinent to explanations of the structure-inducing effects of the co-solvent. One of the best supported explanations of its effects on proteins is based on the magnetic field dependence of the spin-lattice relaxation (magnetic relaxation dispersion) of ^{2}H- and ^{17}O-labeled water, and ^{19}F labelled TFE in a study of β-Lactoglobulin[294]. The authors obtain clear-cut evidence of the binding of TFE both on the surface of the protein and in the interior. They propose that the hydration sphere of β-Lactoglobulin changes in the presence of TFE. Upon increasing the TFE concentration, TFE displaces water on the surface of the protein, decreasing the time of association between water and the protein on the surface and within the interior of the protein. The permanent replacement of water with TFE molecules within the interior of the protein showed a strong correlation to a structural transition occurring

at 30% TFE. These effects caused both a breakdown of the tertiary structure of the protein and an increase in intramolecular backbone H-bonding that stabilized helical secondary structure. When water, penetrating and surrounding the protein, becomes replaced with the larger, less polar TFE molecules that are capable of donating H-bonds but not accepting H-bonds and when TFE interacts with non-polar regions of the protein by Van der Waals dispersion forces, intra-peptidyl hydrogen bonds become more favourable and tertiary interactions are discouraged.

The interactions of TFE with peptides, IDP's and folded proteins are not identical because peptides and IDP's may not form a tertiary structure. However, in all cases TFE promotes secondary structure formation and this has not been completely explained in the literature. In a quantitative study of the α-helical propensities of peptides by titration with TFE, the free energy of α-helix formation exhibited a linear dependence on the mole ratio of TFE to water[223] suggesting a gradual replacement of the peptide's hydration sphere with TFE and supporting the idea of a general solvent effect. Thus, it seems likely that non-polar TFE promotes intraresidue H-bonding by decreasing the dielectric strength of the solvent near the protein. In addition, by replacing two water molecules H-bonded to the amide carbonyl group by a single TFE molecule the entropy gain may help overcome the entropy loss associated with the formation of secondary structure from a disordered random coil[223]. It has also been suggested that stabilization of secondary structure results from weak interactions between TFE and non-polar side-chains. However, we could find no evidence for preferential interactions between TFE and hydrophobic side-chains in Tat_{1-72} based on chemical shift changes induced by TFE on Tat_{1-72} (Table 3.14). That is, there is no evidence that TFE induces greater chemical shift changes at residues containing hydrophobic side-chains. However, the region of the protein that

is most sensitive to TFE is the so-called hydrophobic core of Tat_{1-72} (Table 3.13) and it seems possible that interactions between non-polar TFE and the hydrophobic side-chains are important in driving this region to adopt an β-helical conformation. Clearly more research is needed to understand the structure-inducing effects of TFE.

The effect of TFE on the folding of Tat_{1-72} is similar to the reported effects of TFE on peptides that have been designed for helical propensity, or similar to disordered peptides that are helical when in their native form. One study proved that only peptides designed for helical tendency formed helices at 15 – 40 % TFE[(295)]. In this study, β-hairpin-designed peptides precipitated in the presence of TFE. The peptides represented regions from two structurally related proteins: ubiquitin and the B1 domain of protein L. Common to the peptides of both proteins is that only peptides with helical "tendencies" folded in the presence of TFE where the helical "tendency" is based on whether or not helical structure is observed in the native structure. These results indicate that the induction of secondary structure by TFE is not dependent on native-like tertiary contacts, but instead depend on local, non-random interactions that avoid non-native structures and favour structural propensities by local interactions[(296)]. Local, co-operative folding was also demonstrated in denaturing concentrations of urea by the Outer Membrane Protein X (OMP X) so this observation is not only restricted to water-soluble proteins[(258)].

The hydrophobic core domain is conserved among Tat variants, and the behaviour of this region in TFE provides insight into a possible mechanism of interaction with lipid membranes that may occur during the process of membrane translocation. Although the cell penetrating property of Tat is attributed to the basic region of the protein, and any deletion within this region

resulted in a reduction of membrane translocation activity[106], the basic arginine-rich region may only be responsible for initial events in cell penetration, possibly the association of this cationic region with anionic membrane surfaces of cells. Folding of the hydrophobic core region into an α-helix may be a succeeding event that initiates membrane translocation by endocytosis[297]. Folding of the nearby hydrophobic core region as it comes within contact of the hydrophobic lipid portion of the membrane could be crucial in the uptake process.

4.7 Future Research

Future research of Tat should be directed at understanding the solution state interactions of Tat and TAR, as well as developing a better understanding of how Tat recruits P-TEFb from the 7SK snRNP to the transcriptional complex via the tripartite interaction between P-TEFb, Tat, and TAR. The elucidation of the MoRFs of Tat_{1-72} can initiate research into developing specific small molecule inhibitors that bind to the regions of Tat that undergo disordered-to-ordered transitions, preventing *in vivo* interactions that are associated with HIV disease progression and pathogenicity. Designing small molecule inhibitors is feasible and has been applied to the Myc oncoprotein by identifying the region of the Myc oncoprotein which undergoes binding induced conformational change, and inhibiting its interaction with Max by small molecule inhibition[161].

Tat can cross the blood-brain-barrier[101], is found at elevated amounts in the brain tissue of people suffering from HIV-related dementia[87], and has been theorized to be involved in the progression of HIV-related dementia. This study has shown that Tat_{1-72} shares many of the qualities of amyloid plaque forming precursors, including: the existence of transient, helical

intermediates, the ability to form multimeric associations, association with Congo Red causing the characteristic yellow-green birefringence and a red shift in absorbance at 540 nm, as well as a region predicted to be amyloidegenic. It is not conclusive that neural Tat pathogenicity is similar in mechanism to amyloid disease, but further research into Tat induced neuropathy in light of these recent findings may show similarities to diseases such as Alzheimer's and Parkinson's Disease.

Future studies of Tat_{1-72} should include the full length 101 residue protein because it is conserved in all natural HIV isolates. Although the first exon product is successful in transcriptional activation, the second exon product may be responsible for pathogenetic effects that are unrelated to transcriptional transactivation.

Bibliography

1. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1997) Characterization of the Backbone Dynamics of Folded and Denatured States of an SH3 Domain, *Biochemistry 36*, 2390–2402.
2. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-Dependent Correction of Random Coil NMR Chemical Shifts, *J. Am. Chem. Soc. 123*, 2970-2978.
3. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) 1H, 13C and random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *Journal of Biomolecular NMR 5*, 67-81.
4. (2010) UNAIDS Join United Nations Programme on HIV/AIDS.
5. Lerner, B. H., Gulick, R. M., and Dubler, N. N. (1998) Rethinking Nonadherence: Historical Perspectives on Triple-Drug Therapy for HIV Disease, *Annals of Internal Medicine 129*, 573-578.
6. Cofen, J. M. (1995) HIV population dynamics in vivo: implications for genetic variation, pathogenesis and therapy, *Science 267*, 483-489.
7. Popovic, M., Sarngadharan, M. G., Read, E., and Gallo, R. C. (1984) Detection, Isolation, and Continuous Production of Cytopathic Retroviruses (HTLV-III) from Patients with AIDS and Pre-AIDS, *Science 224*, 497-500.
8. Barré-Sinoussi, F., Chermann, J., Rey, F., Nugeyre, M., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983) Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS), *Science 220*, 868-871.
9. Popovic, M., Sarin, P. S., Robert-Gurroff, M., Kalyanaraman, V. S., Mann, D., Minowada, J., and Gallo, R. C. (1983) Isolation and Transmission of Human Retrovirus (Human T-Cell Leukemia Virus) *Science 219*, 856-859.
10. Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Bijan Safai, White, G., Foster, P., and Markham, P. D. (1984) Frequent Detection and Isolation of Cytopathic Retroviruses (HTLV-III) from Patients with AIDS and at Risk for AIDS, *Science 224*, 500-503.
11. Reya, M. A., Spirea, B., Dormontb, D., Barre-Sinoussia, F., Montagniera, L., and Chermann, J. C. (1984) Characterization of the RNA dependent DNA polymerase of a new human T lymphotropic retrovirus (lymphadenopathy associated virus), *Biochemical and Biophysical Research Communications 121*, 126-133.
12. Chiu, I., Yaniv, A., Dahlberg, J. E., Gazit, A., Skuntz, S. F., Tronick, S. R., and Aaronson, S. A. (1985) Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses, *Nature 317*, 366-368.
13. Sonigosmall, P., Alizon, M., Staskusshort, K., Klatzmann, D., Cole, S., Danos, O., Retzelshort, E., Tiollaissmall, P., Haaseshort, A., and Wain-Hobsonsmall, S. (1985) Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus, *Cell 42*, 369-382.
14. Gonda, M. A., Wong-Staal, F., Gallo, R. C., Clements, J. E., Narayan, O., and Gilden, R. V. (1985) Sequence Homology and Morphologic Similarity of HTLV-III and Visna Virus, a Pathogenic Lentivirus, *Science 227*, 173-177.
15. Levy, J. A. (2003) The search for the CD81 cell anti-HIV factor (CAF), *TRENDS in Immunology 24*, 628-632.

16. Barré-Sinoussi, F. (1996) HIV as the cause of AIDS, *The Lancet 348*, 31-35.
17. Earl, P. L., Doms, R. W., and Moss, B. (1990) Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein, *Proc. Natl. Acad. Sci 87*, 648-652.
18. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998) Structure of an HIVgp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody, *Nature 393*, 648-659.
19. Turner, B. G., and Summers, M. F. (1998) Structure Biology of HIV, *J. Mol. Biol. 285*, 1-32.
20. Shojania, S., and O'Neil, J. D. (2006) HIV-1 Tat is a Natively Unfolded Protein, *The Journal of Biological Chemistry 281*, 8347-8356.
21. Speck, R. F., Wehrly, K., Platt, E. J., Atchison, R. E., Charo, I. F., Kabat, D., and Chesebro, B. (1997) Selective Employment of Chemokine Receptors as Human Immunodeficiency Virus Type 1 Coreceptors Determined by Individual Amino Acids within the Envelope V3 Loop, *Journal of Virology 71*, 7136-7139.
22. Catasti, P., Fontenot, J. D., Bradbury, E. M., and Gupta, G. (1995) Local and Global Structural Properties of the HIV-MN V3 Loop, *The Journal of Biological Chemistry 270*, 2224-2232.
23. Zwick, M. B., Saphire, E. O., and Burton, D. R. (2004) gp41: HIV's shy protein, *Nature Medicine 10*, 133-134.
24. Garzon, M. T., Lidon-Moya, M. C., Barrera, F. N., Prieto, A., Gomez, J., Mateau, M. G., and Neira, J. L. (2009) The dimerization domain of the HIV-1 capsid protein binds a capsid protein-derived peptide: A biophysical characterization, *Protein Science 13*, 1512-1523.
25. Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C.-T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P. (1989) Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1, ***Nature** 337*, 615-620.
26. Wlodawer, A. (1993) Structure-Based Inhibitors of HIV-1 Protease, *Annu. Rev. Biochem. 62*, 543-585.
27. Goff, S. P. (1990) Retroviral reverse transcriptase: Synthesis, structure, and function, *Journal of Acquired Immune Deficiency Syndromes 3*, 817-831.
28. Lafemina, R. L., Schneider, C. L., Robbins, H. L., Callahan, P. L., Legrow, K., Roth, E., Schleif, W. A., and Emini, E. A. (1992) Requirement of Active Human Immunodeficiency Virus Type 1 Integrase Enzyme for Productive Infection of Human T-Lymphoid Cells, *Journal of Virology 66*, 7414-7419.
29. Mangasarian, A., and Trono, D. (1997) The multifaceted role of HIV Nef, *Research in Virology 148*, 30-33.
30. Gallay, P., Swingler, S., Song, J., Bushman, F., and Trono, D. (1995) HIV Nuclear Import Is Governed by the Phosphotyrosine-Mediated Binding of Matrix to the Core Domain of lntegrase, *Cell 83*, 569-576.
31. Miller, M. D., Farnet, C. M., and Bushman, F. D. (1997) Human immunodeficiency virus type 1 preintegration complexes: Studies of organization and composition *Journal of Virology 71*, 5382-5390.
32. Miller, J. H., Presnyak, V., and Smith, H. C. (2001) The dimerization domain of HIV-1 viral infectivity factor Vif is required to block virion incorporation of APOBEC3G, *Retrovirology 4*.

33. Frankel, A. d., and Young, J. A. Y. (1998) HIV-1: Fifteen Proteins and an RNA, *Annu. Rev. Biochem. 67*, 1-25.
34. Salgado, G. F., Vogel, A., Marquant, R., Feller, S. E., Bouaziz, S., and Alves, I. D. (2009) The Role of Membranes in the Organization of HIV-1 Gag p6 and Vpr: p6 Shows High Affinity for Membrane Bilayers Which Substantially Increases the Interaction between p6 and Vpr., *Journal of Medicinal Chemistry 52*, 7157-7162.
35. Veronese, F. D., DeVico, A. L., Copeland, T. D., Oroszlan, S., and Gallo, R. C. (1985) Characterization of gp41 as the Transmembrane Protein Coded by the HTLV-III/LAV Envelope Gene *Science 229*, 1402-1405.
36. Watson, K., and Edwards, R. J. (1999) HIV-1 trans-Activating (Tat) Protein, *Biochemical Pharmacology 58*, 1521-1528.
37. Rimsky, L., Hauber, J., Dukovich, M., Malim, M. H., Langlois, A., Cullen, B. R., and Greene, W. C. (1988) Functional replacement of the HIV-1 rev protein by the HTLV-1 rex protein, *Nature 335*, 738-740.
38. Zapp, M. L., and Green, M. R. (1989) Sequence-specific RNA binding by the HIV-1 Rev protein, *Nature 342*, 714-716.
39. Meyer, B. E., and Malim, M. H. (1994) The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm, *Genes & Development 8*, 1538-1547.
40. Vincent, M. J., Raja, N. U., and Jabbar, M. A. (1993) Human immunodeficiency virus type 1 Vpu protein induces degradation of chimeric envelope glycoproteins bearing the cytoplasmic and anchor domains of CD4: Role of the cytoplasmic domain in Vpu-induced degradation in the endoplasmic reticulum *Journal of Virology 67*, 5538-5549.
41. Kim, S., Byrn, R., Groopman, J., and Baltimore, D. (1989) Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: Evidence for differential gene expression *Journal of Virology 63*, 3708-3713.
42. Cullen, B. R. (1991) Regulation of HIV-1 gene expression, *The FASEB Journal 5*, 2361-2368.
43. Harrich, D., Ulich, C., F.Garcia-Martinez, L., and B.Gaynor, R. (1997) Tat is required for efficient HIV-1 reverse transcription, *The EMBO Journal 16*, 1224-1235.
44. Malim, M. H., Hauber, J., Fenrick, R., and Cullen, B. R. (1988) Immunodeficiency virus rev trans-activator modulates the expression of the viral regulatory genes, *Nature 335*, 181-183.
45. Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M.-A., Santos-Ferreira, M., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J. L., and Montagnier, L. (1986) Isolation of a new human retrovirus from West African patients with AIDS, *Science 233*, 343-346.
46. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L., and Alizon, M. (1987) Genome organization and transactivation of the human immunodeficiency virus type 2, *Nature 326*, 662-669.
47. Silva, T. I. d., Cotten, M., and Rowland-Jones, S. L. (2008) HIV-2: the forgotten AIDS virus, *Trends in Microbiology 16*, 588-595.
48. Hamel, D. J., Sankalé, J.-L., Eisen, G., Meloni, S. T., Mullins, C., Gueye-Ndiaye, A., Mboup, S., and Kanki, P. J. (2007) Twenty years of prospective molecular epidemiology in Senegal: Changes in HIV diversity, *AIDS Res Hum Retroviruses 23*, 1189-1196.
49. O'Donovan, D., Ariyoshi, K., Milligan, P., Ota, M., Yamuah, L., Sarge-Njie, R., and Whittle, H. (2000) Maternal plasma viral RNA levels determine marked differences in

mother-to-child transmission rates of HIV-1 and HIV-2 in The Gambia *AIDS 14*, 441-448.
50. Gottlieb, G. S., Hawes, S. E., Agne, H. D., Stern, J. E., Critchlow, C. W., Kiviat, N. B., and Sow, P. S. (2006) Lower levels of HIV RNA in semen in HIV-2 compared with HIV-1 infection: Implications for differences in transmission *AIDS 20*, 895-900.
51. Jaffar, S. (2004) The natural history of HIV-1 and HIV-2 infections in adults in Africa: a literature review, *Bull. World Health Organ. 82*, 462-469.
52. Berry, N., Ariyoshi, K., Jaffar, S., Sabally, S., Corrah, T., Tedder, R., and Whittle, H. (1998) Low peripheral blood viral HIV-2 RNA in individuals with high CD4 percentage differentiates HIV-2 from HIV-1 infection *Journal of Human Virology 1*, 457-468.
53. Rowland-Jones, S. L. (2006) Protective immunity against HIV infection: lessons from HIV-2 infection, *Future Microbiol. 1*, 427-433.
54. Popper, S. J., Sarr, A. D., Travers, K. U., Guèye-Ndiaye, A., Mboup, S., Essex, M. E., and Kanki, P. J. (1999) Lower human immunodeficiency virus (HIV) type 2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2 *Journal of Infectious Diseases 180*, 1116-1121.
55. Popper, S. J., Dieng Sarr, A., Guèye-Ndiaye, A., Mboup, S., Essex, M. E., and Kanki, P. J. (2000) Low plasma human immunodeficiency virus type 2 viral load is independent of proviral load: Low virus production in vivo *Journal of Virology 74*, 1554-1557.
56. Jeang, K. T., and Gatignol, A. (1994) Comparison of regulatory features among primate lentiviruses in *Current Topics in Microbiology and Immunology*, pp 123-144.
57. Myers, G., Korber, B., and Foley, B. *Human Retroviruses and AIDS: A Compilation and Analysis of Nucleic Acid and Amino Sequences, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory. Los Alamos, NM. pp. III-11 to III-26*.
58. Neuveut, C., and Jeang, K.-T. (1996) Recombinant Human Immunodeficiency Virus Type 1 Genomes with tat Unconstrained by Overlapping Reading Frames Reveal Residues in Tat Important for Replication in Tissue Culture, *Journal of Virology 70*, 5572-5581.
59. Schwartz, S., Felber, B. K., Benko, D. M., Fenyo, E.-M., and Pavlakis, G. N. (1990) Cloning and Functional Analysis of Multiply Spliced mRNA Species of Human Immunodeficiency Virus Type 1, *Journal of Virology 64*, 2519-2529.
60. West, M. J., and Karn, J. (1999) Stimulation of Tat-associated kinase-independent transcriptional elongation from the human immunodeficiency virus type-1 long terminal repeat by a cellular enhancer, *The EMBO Journal 18*, 1378-1386.
61. Frankel, A. D., and Pabo, C. (1988) Cellular Uptake of the Tat Protein from Human Immunodeficiency Virus, *Cell 55*, 1189-1193.
62. A.Garcia, J., Harrich, D., Pearson, L., Mitsuyasu, R., and B.Gaynor, R. (1988) Functional domains required for tat-induced transcriptional activation of the HIV-1 long terminal repeat, *The EMBO Journal 7*, 3143-3147.
63. Mahlknecht, U., Dichamp, I., Varin, A., Lint, C. V., and Herbein, G. (2008) NF-K-B-dependent control of HIV-1 transcription by the second coding exon of Tat in T cells, *Journal of Leukocyte Biology 83*, 718-727.
64. Li, C. J., Ueda, Y., Shi, B., Borodyansky, L., Huang, L., Li, Y. Z., and Pardee, A. B. Tat protein induces self-perpetuating permissivity for productive HIV-1 infection, *Proc. Natl. Acad. Sci. USA 94*.

65. Campbell, G. R., Watkins, J. D., Esquieu, D., Pasquier, E., Loret, E. P., and Spector, S. A. (2005) The C terminus of HIV-1 Tat modulates the extent of CD178-mediated apoptosis of T cells, *Journal of Biological Chemistry 280*, 38376-38382.
66. Richter, S., Cao, H., and Rana, T. M. (2002) Specific HIV-1 TAR RNA Loop Sequence and Functional Groups Are Required for Human Cyclin T1-Tat-TAR Ternary Complex Formation, *Biochemistry 41*, 6391-6397.
67. Anand, K., Schulte, A., Vogel-Bachmayr, K., Scheffzek, K., and Geyer, M. (2008) Structural insights into the Cyclin T1–Tat–TAR RNA transcription activation complex from EIAV, *Nature Structural & Molecular Biology 15*, 1287-1292.
68. Karn, J. (1999) Tackling Tat, *J. Mol. Biol. 293*, 235-254.
69. Core, L. J., Waterfall, J. J., and Lis, J. T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters, *Science 322*, 1845-1848.
70. Tahirov, T. H., Babayeva, N. D., Varzavand, K., Cooper, J. J., Sedore, S. C., and Price, D. H. (2010) Crystal structure of HIV-1 Tat complexed with human P-TEFb, *Nature 465*.
71. Rana, T. M., and Jeang, K.-T. (1999) Biochemical and Functional Interactions between HIV-1 Tat Protein and TAR RNA, *Archives of Biochemistry and Biophysics 365*, 175-185.
72. Barboric, M., and Lenasi, T. (2010) Kick-sTARting HIV-1 transcription elongation by 7SK snRNP deporTATion, *Nature 17*, 928-931.
73. Mujtaba, S., He, Y., Zeng, L., Farooq, A., Carlson, J. E., Ott, M., Verdin, E., and Zhou, M.-M. (2002) Structural basis of lysine-acetylated HIV-1 Tat recognition by PCAF bromodomain *Molecular Cell 9*, 575-586.
74. D'Orso, I., and Frankel, A. D. (2009) Tat acetylation modulates assembly of a viral-host RNA–protein transcription complex, *Proc. Natl. Acad. Sci 106*, 3101-3106.
75. Schulte, A., Czudnochowski, N., Barboric, M., Schönichen, A., Blazek, D., Peterlin, B. M., and Geyer, M. (2005) Identification of a Cyclin T-Binding Domain in Hexim1 and Biochemical Analysis of Its Binding Competition with HIV-1 Tat, *The Journal of Biological Chemistry 280*, 24968-24977.
76. D'Orso, I., and Frankel, A. D. (2010) RNA-mediated displacement of an inhibitory snRNP complex activates transcription elongation, *Nature Structural & Molecular Biology 17*, 815-823.
77. Sobhian, B., Laguette, N., Yatim, A., Nakamura, M., Levy, Y., Kiernan, R., and Benkirane, M. (2010) HIV-1 Tat Assembles a Multifunctional Transcription Elongation Complex and Stably Associates with the 7SK snRNP, *Molecular Cell 38*, 439-451.
78. Diribarne, G., and Bensaude, O. (2009) 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor *RNA Biology 6*, 122-128.
79. Shojania, S., and O'Neil, J. (2011) *Order from disorder: Structure, function and dynamics of the HIV-1 Transactivator of transcription*, John Wiley & Sons (in press).
80. Dow, E. C., Liu, H., and Rice, A. P. (2010) T-Loop Phosphorylated Cdk9 Localizes to Nuclear Speckle Domains Which May Serve as Sites of Active P-TEFb Function and Exchange Between the Brd4 and 7SK/HEXIM1 Regulatory Complexes, *Journal of Cellular Physiology 22*, 84-93.
81. Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999) NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation, *Cell 97*, 41-51.

82. Gautier, V. W., Gu, L., O'Donoghue, N., Pennington, S., Sheehy, N., and Hall, W. W. (2009) In vitro nuclear interactome of the HIV-1 Tat protein, *Retrovirology 6*, 39-57.
83. Nuovo, G. J., Becker, J., Burk, M. W., Margiotta, M., Fuhrer, J., and Steigbigel, R. T. (1994) In situ detection of PCR-amplified HIV-1 nucleic acids in lymph nodes and peripheral blood in patients with asymptomatic HIV-1 infection and advanced-stage AIDS, *J. Acquired Immune Defic. Syndr. 7*, 916-923.
84. Ensoli, B., Buonaguro, L., Barillari, G., Fiorelli, V., R. Gendelman, Morgan, R. A., Wingfield, P., and Gallo, R. C. (1993) Release, uptake, and effects of extracellular human immunodeficiency virus type 1 Tat protein on cell growth and viral transactivation, *J. Virol. 67*, 277-287.
85. Schwarze, S. R., Ho, A., Vocero-Akbani, A., and Dowdy, S. F. (1999) In vivoprotein transduction: delivery of a biologically active protein into the mouse, *Science 285*, 1569-1572.
86. Toborek, M., Lee, Y. W., Pu, H., Malecki, A., Flora, G., Garrido, R., Hennig, B., Bauer, H. C., and Nath, A. (2003) HIV-Tat protein induces oxidative and inflammatory pathways in brain endothelium, *J. Neurochem. 84*, 169-179.
87. Ziegler, A., and Seelig, J. (2004) Interaction of the protein transduction domain of HIV-1 TAT with heparan sulfate: binding mechanism and thermodynamic parameters, *Biophys. J. 86*, 254-263.
88. Cohen, S. S., Li, C., Ding, L., Cao, Y., Pardee, A. B., E. M. Shevach, and Cohen, D. I. (1999) Pronounced acute immunosuppression in vivo mediated by HIV Tat challenge, *Proc. Natl. Acad. Sci 96*, 10842-10847.
89. Haughey, N. J., and Mattson, M. P. (2002) Calcium dysregulation and neuronal apoptosis by the HIV-1 proteins Tat and gp120, *J. Acquired Immune Defic. Syndr. 31*, 55-61.
90. Hudson, L., Liu, J., Nath, A., Jones, M., Raghavan, R., Narayan, O., Male, D., and Everall, I. (2000) Detection of the human immunodeficiency virus regulatory protein tat in CNS tissues, *J. Neurovirol. 6*, 145–155.
91. Orsini, M. J., Debouck, C. M., Webb, C. L., and Lysko, P. G. (1996) Extracellular Human lmmunodeficiency Virus Type 1 Tat Protein Promotes Aggregation and Adhesion of Cerebellar Neurons, *The Journal of Neuroscience 16*, 2546-2552.
92. Hayman, M., Arbuthnott, G., Harkiss, G., Brace, H., Filippi, P., Philippon, V., Thomson, D., Vigne, R., and Wright, A. (1993) Neurotoxicity of peptide analogues of the transactivating protein tat from Maedi–Visna virus and human immunodeficiency virus, *Neuroscience 53*, 1-6.
93. Maragos, W. F., Tillman, P., Jones, M., Bruce-Keller, A. J., Roth, S., Bell, J. E., and Nath, A. (2003) Neuronal injury in hippocampus with human immunodeficiency virus transactivating protein, Tat, *Neuroscience 117*, 43-53.
94. Magnuson, D. S., Knudsen, B. E., Geiger, J. D., Brownstone, R. M., and Nath, A. (1995) Human immunodeficiency virus type 1 tat activates non-Nmethyl-d-aspartate excitatory amino acid receptors and causes neurotoxicity, *Ann. Neurol. 37*, 373-380.
95. Nath, A., Psooy, K., Martin, C., Knudsen, B., Magnuson, D. S., Haughey, N., and Geiger, J. D. (1996) Identification of a human immunodeficiency virus type 1 Tat epitope that is neuroexcitatory and neurotoxic, *J. Virol. 70*, 1475-1480.
96. Tardieu, M., Hery, C., Peudenier, S., Boespflug, O., and L. Montagnier. (1992) Human immunodeficiency virus type 1-infected monocytic cells can destroy human neural cells after cell-to-cell adhesion, *Ann. Neurol. 32*, 11-17.

97. Yankner, B. A., Duffy, L. K., and Kirschner, D. A. (1990) Neurotrophic and neurotoxic effects of amyloid beta protein: reversal by tachykinin neuropeptides, *Science 250*, 279-282.
98. Narayan, O., Wolinsky, J. S., Clements, J. E., Strandberg, J. D., Griffin, D. E., and Cork, L. C. (1982) Slow virus replication: the role of macrophages in the persistence and expression of visna viruses of sheep and goats, *J. Gen. Virol. 59*, 345-356.
99. Davis, L. E., Hjelle, B. L., Miller, V. E., Palmer, D. L., Llewellyn, A. L., Merlin, T. L., Young , S. A., Mills, R. G., Wachsman, W., and Wiley, C. A. (1992) Early viral brain invasion in iatrogenic human immunodeficiency virus infection, *Neurology 42*, 1736-1739.
100. Persidsky, Y., Ghorpade, A., Rasmussen, J., Limoges, J., Liu, X. J., Stins, M., Fiala, M., Way, D., Kim, K. S., Witte, M. H., Weinand, M., Carhart, L., and Gendelman, H. E. (1999) Microglial and astrocyte chemokines regulate monocyte migration through the blood-brain barrier in human immunodeficiency virus-1 encephalitis, *Am. J. Pathol. 155*, 1599-1611.
101. Banks, W. A., Robinson, S. M., and Nath, A. (2005) Permeability of the blood–brain barrier to HIV-1 Tat, *Experimental Neurology 193*, 218-227.
102. Avraham, H. K., Jiang, S., Lee, T. H., Prakash, O., and Avraham, S. (2004) HIV-1 Tat-mediated effects on focal adhesion assembly and permeability in brain microvascular endothelial cells, *J. Immunol. 173*, 6228-6233.
103. Cheng, J., Nath, A., Knudsen, B., Hochman, S., Geiger, J. D., Ma, M., and Magnuson, D. S. (1998) Neuronal excitatory properties of human immunodeficiency virus type 1 Tat protein, *Neuroscience 82*, 97–106.
104. Liu, Y., Jones, M., Hingtgen, C. M., Bu, G., Laribee, N., Moir, R. E. T. R. D., Nath, A., and He, J. J. (2000) Uptake of HIV-1 tat protein mediated by low-density lipoprotein receptor-related protein disrupts the neuronal metabolic balance of the receptor ligands, *Nature Medicine 6*, 1380– 1387.
105. Hallbrink, M., Floren, A., Elmquist, A., Pooga, M., Bartfai, T., and Langel, U. (2001) Cargo delivery kinetics of cell-penetrating peptides, *Biochim. Biophys. Acta 1515*, 101-109.
106. Park, J., Ryu, J., Kim, K. A., Lee, H. J., Bahn, J. H., Han, K., Choi, E. Y., Lee, K. S., Kwon, H. Y., and Choi, S. Y. (2002) MuTATional analysis of a human immunodeficiency virus type 1 TAT protein transduction domain which is required for delivery of an exogenous protein into mammalian cells, *J. Gen. Virol. 83*, 1173-1181.
107. Gupta, B., Levchenko, T. S., and Torchilin, V. P. (2005) Intracellular delivery of large molecules and small particles by cell-penetrating proteins and peptides, *Advanced Drug Delivery Reviews 57*, 637-651.
108. Vendel, A. C., and Lumb, K. J. (2003) Molecular Recognition of the Human Coactivator CBP by the HIV-1 Transcriptional Activator Tat, *Biochemistry 42*, 910-916.
109. Mayr, B., and Montminy, M. (2001) Transcriptional regulation by the phosphorylation-dependent factor CREB, *Nat. Rev. Mol. Cell. Biol. 8*, 599-609.
110. Janknecht, R., and Hunter, T. (1996) A growing coactivator network, *Nature 383*, 22-23.
111. Chan, H. M., and La Thangue, N. B. (2001) p300/CBP proteins: HATs for transcriptional bridges and scaffolds *Journal of Cell Science 114*, 2363-2373.

112. Marzio, G., Tyagi, M., Gutierrez, M. I., and Giacca, M. (1998) HIV-1 Tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter, *Proc. Natl. Acad. Sci USA 95*, 13519-13524.
113. III, H. J. C., Ciofi-Baffoni, S., Banci, L., Bertini, I., Zhang, L., George, G. N., and Winge, D. R. (2007) The Characterization and Role of Zinc Binding in Yeast Cox4, *The Journal of Biological Chemistry 282*, 8926-8934.
114. Ferraroni, M., Tilli, S., Briganti, F., Chegwidden, W. R., Supuran, C. T., Wiebauer, K. E., Tashian, R. E., and Scozzafava, A. (2002) Crystal Structure of a Zinc-Activated Variant of Human Carbonic Anhydrase I, CA I Michigan 1: Evidence for a Second Zinc Binding Site Involving Arginine Coordination, *Biochemistry 41*, 6237-6244.
115. Calzi, M. L., and Poole, L. B. (1997) Requirement for the Two AhpF Cystine Disulfide Centers in Catalysis of Peroxide Reduction by Alkyl Hydroperoxide Reductase, *Biochemistry 36*, 13357-13364.
116. Frankel, A. D., Chen, L., Cotter, R. J., and Pabo, C. (1988) Dimerization of the tat protein from human immunodeficiency virus: A cysteine-rich peptide mimics the normal metal-linked dimer interface, *Proc. Natl. Acad. Sci 85*, 6297-6300.
117. Garber, M. E., Wei, P., KewalRamani, V. N., Mayall, T. P., Herrmann, C. H., Rice, A. P., Littman, D. R., and Jones, K. A. (1998) The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein, *Genes & Development 12*, 3512-3572.
118. Jeang, K.-T., Xiao, H., and Rich, E. A. (1999) Multifaceted Activities of the HIV-1 Transactivator of Transcription, Tat, *The Journal of Biological Chemistry 274*, 28837-28840.
119. Garber, M. E., Wei, P., KewalRamani, V. N., Mayall, T. P., Herrmann, C. H., Rice, A. P., Littman, D. R., and Jones, K. A. (1998) The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein, *Genes and Dev. 12*, 3512-3527.
120. Egelé, C., Barbier, P., Didier, P., Piémont, E., Allegro, D., Chaloin, O., Muller, S., Peyrot, V., and Mély, Y. (2008) Modulation of microtubule assembly by the HIV-1 Tat protein is strongly dependent on zinc binding to Tat, *Retrovirology 5*.
121. Chen, D., Wang, M., Zhou, S., and Zhou, Q. (2002) HIV-1 Tat targets microtubules to induce apoptosis, a process promoted by the pro-apoptotic Bcl-2 relative Bim, *The EMBO Journal 21*, 6801-6810.
122. Misumi, S., Takamune, N., Ohtsubo, Y., Waniguchi, K., and Shoji, S. (2004) Zn2+ binding to cysteine-rich domain of extracellular human immunodeficiency virus type 1 Tat protein is associated with Tat protein-induced apoptosis., *AIDS Res Hum Retroviruses 20*, 297-304.
123. Laidlera, P., Koniecznya, L., and Bobrzecka, K. (1979) The controlled disulphide crosslinking of different proteins as a method for synthesis of combined protein antigens and copolymers of other design, *Molecular Immunology 16*, 141-144.
124. Campbell, G. R., Pasquier, E., Watkins, J., Bourgarel-Rey, V., Peyrot, V., Esquieu, D., Barbier, P., Mareuil, J. d., Braguer, D., Kaleebu, P., Yirrell, D. L., and Loret, E. P. (2004) The Glutamine-rich Region of the HIV-1 Tat Protein Is Involved in T-cell Apoptosis, *The Journal of Biological Chemistry 279*, 48197-48204.
125. Bayer, P., Kraft, M., Ejchart, A., Westendorp, M., Frank, R., and Rosch, P. (1995) Structural Studies of HIV-1 Tat Protein, *J. Mol. Biol. 247*, 529-535.

126. Campbell, G. R., Pasquier, E., Watkins, J., Peyrot, V. B.-R. V., Esquieu, D., Barbier, P., Mareuil, J. d., Braguer, D., Kaleebu, P., Yirrell, D. L., and Loret, E. P. (2004) The Glutamine-rich Region of the HIV-1 Tat Protein Is Involved in T-cell Apoptosis, *The Journal of Biological Chemistry 279*, 48197-48204.
127. Wright, P. E., and Dyson, H. J. (1999) Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm, *J. Mol. Biol. 293*, 321-331.
128. Le Gall, T., Romero, P. R., Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2007) Intrinsic disorder in the Protein Data Bank, *J. Biomol. Struct. Dyn. 24*, 325-342.
129. Garner, E., Cannon, P., Romero, P., Obradovic, Z., and Dunker, A. K. (1998) Predicting disordered regions from amino acid sequence: common themes despite differing structural characterization., *Genome Inform. 9*, 201-214.
130. Uversky, V. N., and Dunker, A. K. (2010) Understanding protein non-folding, *Biochimica et Biophysica Acta 1804*, 1231-1264.
131. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. (1998) Thousands of proteins likely to have long disordered regions., *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 437-448.
132. Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content., *Bioinformatics 21*, 3433-3434.
133. Obradovic, Z., Kissinger, C., and Villafranca, J. E. (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations, *Pac. Symp. Biocomput. 3*, 471-482.
134. Romero, P., Obradovic, Z., and Dunker, A. K. (1997) Sequence data analysis for long disordered regions prediction in the calcineurin family, *Genome Inform. 8*, 110-124.
135. Ward, J. J., Sodhia, J. S., McGuffina, L. J., Buxtona, B. F., and Jones, D. T. (2004) Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life *Journal of Molecular Biology 337*, 635-645.
136. Chen, J. W., Romero, P., Uversky, V. N., and Dunker, A. K. (2006) Conservation of Intrinsic Disorder in Protein Domains and Families: I. A Database of Conserved Predicted Disordered Regions, *Journal of Proteome Research 5*, 879-887.
137. Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics, *Protein Science 11*, 739-756.
138. Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins, *FEBS Letters 579*, 3346-3354.
139. Meador, W. E., Means, A. R., and Quiocho, F. A. (1992) Target enzyme recognition by calmodulin: 2.4 angstrom structure of a calmodulin-peptide complex, *Science 257*, 1251-1255.
140. Liu, J., Faederb, J. R., and Camacho, C. J. (2009) Toward a quantitative theory of intrinsically disordered proteins and their function, *PNAS 106*, 19819-19823.
141. Huang, Y., and Liu, Z. (2009) Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding-Binding Process: A Critical Assessment of the "Fly Casting" Mechanism *Journal of Molecular Biology 393*, 1143-1159.
142. Parsegian, V. A., Rand, R. P., and Rau, D. C. (2000) Osmotic stress, crowding, preferential hydration, and binding: A comparison of perspectives, *PNAS 97*, 3987-3992.

143. Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C., and Struhl, K. (1990) Folding transition in the DMA-binding domain of GCN4 on specific binding to DNA, *Nature 347*, 575-578.
144. O'Neil, K. T., Hoess, R. H., and DeGrado, W. F. (1990) Design of DNA-Binding Peptides Based on the Leucine Zipper Motif, *Science 249*, 774-778.
145. Bracken, C., Carr, P. A., Cavanagh, J., and Palmer III, A. G. (1999) Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: Implications for the entropy of association with DNA *Journal of Molecular Biology 285*, 2133-2146.
146. Bienkiewicz, E. A., Adkins, J. N., and Lumb, K. J. (2002) Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1), *Biochemistry 41*, 752-759.
147. Chi, S. W., Kim, D. H., Lee, S. H., Chang, I., and Han, K. H. (2007) Pre-structured motifs in the natively unstructured preS1 surface antigen of hepatitis B virus, *Protein Science 16*, 2108-2117.
148. Baldwin, A. J., and Kay, L. E. (2009) NMR spectroscopy brings invisible protein states into focus, *Nature Chemical Biology 5*, 808-814.
149. Kay, L. E. (2005) NMR studies of protein structure and dynamics, *Journal of Magnetic Resonance 173*, 193-207.
150. Craig, T. A., Veenstra, T. D., Naylor, S., Tomlinson, A. J., Johnson, K. L., Macura, S., Juranic, N., and Kumar, R. (1997) Zinc Binding Properties of the DNA Binding Domain of the 1,25-Dihydroxyvitamin D3 Receptor, *Biochemistry 36*, 10482-10491.
151. Freund, J., Vertesy, L., Koller, K. P., Wolber, V., Heintz, D., and Kalbitzer, H. R. (1995) Complete 1H nuclear magnetic resonance assignments and structural characterization of a fusion protein of the alpha-amylase inhibitor tendamistat with the activation domain of the human immunodeficiency virus type 1 Tat protein, *J. Mol. Biol. 250*, 672-688.
152. Lacy, E. R., Filippov, I., Lewis, W. S., Otieno, S., Xiao, L., Weiss, S., Hengst, L., and Kriwacki, R. W. (2004) p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding., *Nat. Struct. Mol. Biol. 11*, 358-364.
153. Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between a- and g-synuclein: Implications for fibrillation, *Protein Science 15*, 2795-2804.
154. Sanchez, R., and Sali, A. (1997) Advances in comparative protein-structure modelling, *Curr. Opin. Struct. Biol. 7*, 206–214.
155. Mok, Y. K., Elisseeva, E. L., Davidson, A. R., and Forman-Kay, J. D. (2001) Dramatic stabilization of an SH3 domain by a single substitution: Roles of the folded and unfolded states, *J. Mol. Biol. 307*, 913–928.
156. Bussell, R., and Eliezer, E. (2001) Residual structure and dynamics in Parkinson's disease-associated mutants of a-synuclein, *J. Biol. Chem. 276*, 45996–46003.
157. Uversky, V. N., Li, J., Souillac, P., Millett, I. S., Doniach, S., Jakes, R., Goedert, M., and Fink, A. L. (2002) Biophysical properties of the synucleins and their propensities to fibrillate, *J. Biol. Chem. 277*, 11970–11978.
158. Imming, P., Sinning, C., and Meyer, A. (2006) Drugs, their targets and the nature and number of drug targets, *Nat. Rev. Drug Discov. 5*, 821-834.

159. Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept, *Annu. Rev. Biophys. 37*, 215-246.
160. Vavouri, T., Semple, J. I., Garcia-Verdugo, R., and Lehner, B. (2009) Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity, *Cell 138*, 198-208.
161. Metallo, S. J. (2010) Intrinsically disordered proteins are potential drug targets, *Current Opinion in Chemical Biology 14*, 481-488.
162. Vacic, V., Oldfield, C. J., Mohan, A., Radivojac, P., Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2007) Characterization of Molecular Recognition Features, MoRFs, and Their Binding Partners, *Journal of Proteome Research 6*, 2351-2366.
163. Follis, A. V., Hammoudeh, D. I., Wang, H. B., Prochownik, E. V., and Metallo, S. J. (2008) Structural rationale for the coupled binding and unfolding of the c-Myc oncoprotein by small molecules, *Chem. Biol. 15*, 1149-1155.
164. Dang, C. V. (1999) c-Myc target genes involved in cell growth, apoptosis, and metabolism, *Molecular and Cellular Biology 19*, 1-11.
165. Uversky, V. N., Li, J., Bower, K., and Fink, A. L. (2002) *Neurotoxicology 23*, 527–536.
166. Ptitsyn, O. B., and Uversky, V. N. (1994) The molten globule is a third thermodinamical state of protein molecules, *FEBS Letters 341*, 15-18.
167. Cavanagh, J. F., Wayne, J., Palmer III, A. G., and Skelton, N. J. (1996) *Protein NMR Spectroscopy: Principle and Practice*, Academic Press, Inc., San Diego, CA.
168. Hu, Y., Macinnis, J. M., Cherayil, B. J., Fleming, G. R., Freed, K. F., and Perico, A. (1990) Polypeptide Dynamics - Experimental Tests of an Optimized Rouse-Zimm Type Model, *J. Chem. Phys. 93*, 822-836.
169. Watkins, J. D., Campbell, G. R., Halimi, H., and Loret, E. P. (2008) Homonuclear ^{1}H NMR and circular dichroism study of the HIV-1 Tat Eli variant, *Retrovirology 5*.
170. Slice, L. W., Codner, E., Antelman, D., Holly, M., Wegrzynski, B., Wang, J., Toome, V., Hsu, M.-C., and Nalin, C. M. (1992) Characterization of Recombinant HIV-1 Tat and Its Interaction with TAR RNA, *Biochemistry 31*, 12062-12068.
171. Foucault, M., Mayol, K., Receveur-Brechot, V., Bussat, M.-C., Klinguer-Hamour, C., Verrier, B., Beck, A., Haser, R., Gouet, P., and Guillon, C. (2009) UV and X-ray structural studies of a 101-residue long Tat protein from a HIV-1 primary isolate and of its mutated, detoxified, vaccine candidate, *Proteins: Structure, Function, and Bioinformatics 78*, 1441-1456.
172. Vo, T. N. (2008) Spectroscopic Characterization of HIV-1 Transactivator of Transcription (Tat) Interactions with Divalent Metals in *Chemistry*, p 197, University of Manitoba, Winnipeg.
173. Anand, K., Schulte, A., Fujinaga, K., Scheffzek, K., and Geyer, M. (2007) Cyclin Box Structure of the P-TEFb Subunit Cyclin T1 Derived from a Fusion Complex with EIAV Tat, *J. Mol. Biol. 370*, 826-836.
174. Hetzer, C., Dormeyer, W., Schnolzer, M., and Ott, M. (2005) Decoding Tat: the biology of HIV Tat posttranslational modifications, *Microbes Infect. 7*, 1364–1369.
175. Puglisi, J. D., Chen, L., Blanchard, S., and Frankel, A. D. (1995) Solution Structure of a Bovine Immunodeficiency Virus Tat-TAR Peptide-RNA Complex, *Science 270*, 1200-1203.

176. Sticht, H., Willbold, D., Ejchart, A., Rosin-Arbesfeld, R., Yaniv, A., Gazit, A., and Rosch, P. (1994) Trifluoroethanol stabilizes a helix-turn-helix motif in equine infectious-anemia-virus trans-activator protein, *Eur. J. Biochem. 225*, 855-861.
177. Hawkins, P. N., and Pepys, M. B. (1995) *Eur. J. Nucl. Med. 22*, 595–599.
178. R. Kisilevsky, P. E. F. (1997) Aβ amyloidogenesis: Unique, or variation on a systemic theme?, *Critical Reviews in Biochemistry and Molecular Biology 32*, 361-404.
179. Abedini, A., and Raleigh, D. P. (2009) A role for helical intermediates in amyloid formation by natively unfolded polypeptides?, *Physical Biology 6*.
180. Bartolini, M., Bertucci, C., Bolognesi, M. L., Cavalli, A., Melchiorre, C., and Andrisano, V. (2007) Insight Into the Kinetic of Amyloid (1-42) Peptide Self-Aggregation: Elucidation of Inhibitors Mechanism of Action, *ChemBioChem 8*, 2152-2161.
181. Apostolidou, M., Jayasinghe, S. A., and Langen, R. (2008) Structure of Alpha-Helical Membrane-bound Human Islet Amyloid Polypeptide and Its Implications for Membrane-mediated Misfolding *The Journal of Biological Chemistry 283*, 17205-17210.
182. Teplow, D. B., Lazo, N. D., Bitan, G., Bernstein, S., Wyttenbach, T., Bowers, M. T., Baumketner, A., Shea, J.-E., Urbanc, B., Cruz, L., Borreguero, J., and Stanley, H. E. (2006) Elucidating Amyloid B-Protein Folding and Assembly: A Multidisciplinary Approach, *Acc. Chem. Res. 39*, 635-645.
183. Kuner, P., Bohrmann, B., Tjernberg, L. O., Naslund, J., Huber, G., Celenk, S., Gruninger-Leitch, F., Richards, J. G., Jakob-Roetne, R., Kemp, J. A., and Nordstedt, C. (2000) Controlling Polymerization of b-Amyloid and Prion-derived Peptides with Synthetic Small Molecule Ligands, *The Journal of Biological Chemistry 275*, 1673-1678.
184. Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G., and Dobson, C. (2002) Kinetic partitioning of protein folding and aggregation, *Nature Structural Biology 9*, 137-133.
185. Rousseau, F., Schymkowitz, J., and Serrano, L. (2006) Protein aggregation and amyloidosis: confusion of the kinds?, *Curr. Opin. Struct. Biol. 16*, 118-126.
186. Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C. M. (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates, *Nature 424*, 805-808.
187. Chiti, F., and Dobson, C. M. (2006) Protein misfolding, functional amyloid, and human disease, *Annu. Rev. Biochem. 75*, 333-366.
188. Wang, X., and Chapman, M. R. (2008) Sequence Determinants of Bacterial Amyloid Formation, *J. Mol. Biol. 380*, 570-580.
189. Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Paz, M. L. d. l., Martins, I. C., Reumers, J., Morris, K. L., Copland, A., Serpell, L., Serrano, L., Schymkowitz, J. W. H., and Rousseau, F. (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices, *Nature 7*, 237-245.
190. Zurdo, S. V. J., Narayanan, S., Parreno, M., Mangues, R., Reif, B., Chiti, F., Giannoni, E., Dobson, C. M., Aviles, F. X., and Serrano, L. (2004) Short amino acid stretches can mediate amyloid formation in globular proteins: The Src homology 3 (SH3) case, *PNAS 101*, 7258-7263.
191. Osherovich, L. Z., Cox, B. S., Tuite, M. F., and Weissman, J. S. (2004) Dissection and design of yeast prions, *PLoS Biol. 2*, E86.

192. Tartaglia, G. G., Pawar, A. P., Campioni, S., Dobson, C. M., Chiti, F., and Vendruscolo, M. (2008) Prediction of Aggregation-Prone Regions in Structured Proteins, *Journal of Molecular Biology 380*, 425-436.
193. Thompson, M. J. (2006) The 3D profile method for identifying fibril-forming segments of proteins, *Proc. Natl. Acad. Sci. USA 103*, 4074–4078.
194. Klunk, W. E., Pettegrew, J. W., and Abraham, D. J. (1989) Two Simple Methods for Quantifying Low-affinity Dye-Substrate Binding, *The Journal of Histochemistry and Cytochemistry 37*, 1293-1297.
195. Hofmeister, F. (1888) *Arch. Exp. Pathol. Pharmocol. 24*, 247-260.
196. Jarvis, N. L., and Scheiman, M. A. (1968) Surface Potential of Aqueous Electrolyte Solutions, *The Journal of Physical Chemistry 72*, 74-78.
197. Creighton, T. E. (1984) *Proteins Structures and Molecular Properties*, W. H. Freeman and Company, New York.
198. McDevit, W. F., and Long, F. A. (1952) The Activity Coefficient of Benzene in Aqueous Salt Solutions, *J. Am. Chem. Soc. 74*, 1773-1777.
199. Collins, K. D., and Washabaugh, M. W. (1985) The Hofmeister effect and the behaviour of water at interfaces, *Quarterly Reviews of Biophysics 18*, 323-422.
200. Baldwin, R. L. (1996) How Hofmeister Ion Interactions Affect Protein Stability, *Biophysical Journal 71*, 2056-2063.
201. Nandi, P. K., and Robinson, D. R. (1972) The effects of salts on the free energy of the peptide group, *The Journal of the American Chemical Society 94*, 1299-1308.
202. Mason, P. E., Brady, J. W., Neilson, G. W., and Dempsey, C. E. (2007) The Interaction of Guanidinium Ions with a Model Peptide, *Biophys. J. 93*, L04-L06.
203. Nozaki, Y., and Tanford, C. (1970) *J. Biol. Chem. 245*, 1648-1652.
204. Zhang, Y., and Cremer, P. S. (2006) Interactions between macromolecules and ions: the Hofmeister series, *Current Opinion in Chemical Biology 10*, 658-663.
205. Omta, A. W., Kropman, M. F., Woutersen, S., and Bakker, H. J. (2003) Negligible effect of ions on the hydrogen-bond structure in liquid water, *Science 301*, 347-349.
206. Gurau, M. C., S.M., L., Castellana, E. T., Albertorio, F., Kataoka, S., and Cremer, P. S. (2004) On the mechanism of the Hofmeister effect, *J. Am. Chem. Soc. 126*, 10522-10523.
207. Povey, J. F., Smales, C. M., Hassard, S. J., and Howard, M. J. (2007) Comparison of the effects of 2,2,2-trifluoroethanol on peptide and protein structure and function, *Journal of Structural Biology 157*, 329-338.
208. Buck, M. (1998) Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins, *Quarterly Reviews of Biophysics 31*.
209. Weber, R. E., and Tanford, C. (1958) The Configuration of Ribonuclease at Low pH in 2-Chlorothanol and in 2-Chloroethanol-Water Mixtures, *J. Am. Chem. Soc. 81*, 3255-3260.
210. Luidens, M. K., Figge, J., Breese, K., and Vajda, S. (1996) Predicted and trifluoroethanol-induced α-helicity of polypeptides *Biopolymers 39*, 367-376.
211. Cammers-Goodwin, A., Allen, T. J., Oslick, S. L., McClure, K. F., Lee, J. H., and Kemp, D. S. (1996) Mechanism of stabilization of helical conformations of polypeptides by water containing trifluoroethanol *Journal of the American Chemical Society 118*, 3082-3090.
212. Rajan, R., and Balaram, P. (1996) A model for the interaction of trifluoroethanol with peptides and proteins *International Journal of Peptide and Protein Research 48*, 328-336.

213. Gente, G., and Mesa, C. L. (2000) Water–Trifluoroethanol Mixtures: Some Physicochemical Properties, *Journal of Solution Chemistry 29*, 1159-1172.
214. Llinas, M., and Klein, M. P. (1975) Charge Relay at the Peptide Bond. A Proton Magnetic Resonance Study of Solvation Effects on the Amide Electron Density Distribution, *Journal of the American Chemical Society 97*, 4731-4737.
215. Yang, Y., Barker, S., Chen, M. J., and Mayo, K. H. (1993) Effect of low molecular weight aliphatic alcohols and related compounds on platelet factor 4 subunit association, *J. Biol. Chem. 268*, 9223-9229.
216. Nguyen, L. T., Chan, D. I., Boszhard, L., Zaat, S. A. J., and Vogel, H. J. (2010) Structure–function studies of chemokine-derived carboxy-terminal antimicrobial peptides, *Biochimica et Biophysica Acta 1798*, 1062-1072.
217. Santos, J., Sica, M. P., Buslje, C. M., Garrote, A. M., Ermacora, M. R., and Delfino, J. M. (2009) Structural Selection of a Native Fold by Peptide Recognition. Insights into the Thioredoxin Folding Mechanism, *Biochemistry 48*, 595-607.
218. Shoemaker, B. A., Portman, J. J., and Wolynes, P. G. (200) Proc. Natl. Acad. Sci. U.S.A., *97*.
219. Uversky, V. N. (2009) Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding, *Protein J. 28*, 305-325.
220. Narayanan, R., Ganesh, O. K., Edison, A. S., and Hagen, S. J. (2008) Kinetics of Folding and Binding of an Intrinsically Disordered Protein: The Inhibitor of Yeast Aspartic Proteinase YPrA, *J. Am. Chem. Soc. 130*, 11477-11485.
221. Li, M., Phylip, L. H., Lees, W. E., Winther, J. R., Dunn, B. M., Wlodaawer, A., Kay, J., and Gustchina, A. (2000) *Nat. Struct. Biol. 7*, 113–117.
222. Chan, H. S., Bromberg, S., and Dill, K. A. (1995) *Models of cooperativity in protein folding*, Vol. 348, Press Syndicate of the University of Cambridge, Cambridge.
223. Jasanofft, A., and Fersht, A. R. (1994) Quantitative Determination of Helical Propensities from Trifluoroethanol Titration Curves, *Biochemistry 33*, 2129-2135.
224. Ferry, J. D., and Morrison, P. R. (1947) Preparation and Properties of Serum and Plasma Proteins. VIII. The Conversion of Human Fibrinogen to Fibrin under Various Conditions, *Contribution from the Department of Physical Chemistry, Harvard Medical School 69*, 388-400.
225. Rogalla, T., Ehrnsperger, M., Preville, X., Kotlyarov, A., Lutsch, G., Ducasse, C., Paul, C., Wieske, M., Arrigo, A.-P., Buchner, J., and Gaestel, M. (1999) Regulation of Hsp27 Oligomerization, Chaperone Function, and Protective Activity against Oxidative Stress/Tumor Necrosis Factor a by Phosphorylation, *The Journal of Biological Chemistry 274*, 18947-18956.
226. Koch, A. L. (1961) Some calculations on the turbidity of mitochondria and bacteria, *BBA - Biochimica et Biophysica Acta 51*, 429-441.
227. Eric, Y. H. (2010) Electric Vectors
228. Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure, *Nature Protocols 1*, 2876-2890.
229. Thomas, K. A. (1996) *Journal of Biochemical Chemistry 2*, 603-606.
230. Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, *Analytical Biochemistry 287*, 252-260.

231. Greenfield, N., and Fasman, G. D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation, *Biochemistry 8*, 4108–4116.
232. Stokkum, I. H. M. v., Spoelder, H. J. W., Bloemendal, M., Grondelle, R. v., and Groen, F. C. A. (1990) Estimation of Protein Secondary Structure and Error Analysis from Circular Dichroism Spectra, *Analytical Biochemistry 191*, 110-118.
233. Akke, M. (2002) NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis, *Current Opinion in Structural Biology 12*, 642-647.
234. Palmer, A. G., Kroenke, C. D., and Loria, J. P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules, *Methods in Enzymology 339*, 204-238.
235. Yao, J., Dyson, H. J., and Wright, P. E. (1997) Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins, *FEBS Letters 419*, 285-289.
236. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR, *Journal of Biomolecular NMR 6*, 135-140.
237. Yao, J., Chung, J., Eliezer, D., Wright, P. E., and Dyson, H. J. (2001) NMR Structural and Dynamic Characterization of the Acid-Unfolded State of Apomyoglobin Provides Insights into the Early Events in Protein Folding, *Biochemistry 40*, 3561-3571.
238. Shojania, S. (2007) Nuclear Magnetic Resonance and Dynamic Characterization of the Intrinsically Disordered HIV-1 Tat Protein, in *Chemistry*, p 277, University of Manitoba, Winnipeg.
239. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries, *Proc. Natl. Acad. Sci 97*, 7754-7759.
240. Hasan, A. A. K., Amenta, S., and Schmaier, A. H. (1996) Bradykinin and its metabolite, Arg-Pro-Pro-Gly-Phe, are selective inhibitors of α-thrombin-induced platelet activation *Circulation 94*, 517-528.
241. Edwards, A. M., Arrowsmith, C. H., Christendat, D., Dharamsi, A., Friesen, J. D., Greenblatt, J. F., and Vedadi, M. (2000) Protein production: feeding the crystallographers and the NMR spectroscopists, *Nature structural biology 11*, 970-972.
242. Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974) *J. Bacteriology 119*, 736–747.
243. Mukerjee, P., and Mysels, K. J. (1971) Critical Micelle Concentration of Aqueous Surfactant Systems, *US Government Printing Office NSRDS-NMS 36*.
244. Manley, D., and O'Neil, J. D. (2003) *Methods in Enzymology 228*, 89-101.
245. Whitmore, L., and Wallace, B. A. (2008) Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases, *Biopolymers 89*, 392-400.
246. Provencher, S. W., and Glockner, J. (1981) Estimation of Globular Protein Secondary Structure from Circular Dichroism, *Biochemistry 20*, 33-37.
247. Lees, J. G., Miles, A. J., Wien, F., and Wallace, B. A. (2006) A reference database for circular dichroism spectroscopy covering fold and secondary structure space, *Bioinformatics 22*, 1955-1962.
248. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (1999) Estimation of the number of a-helical and b-strand segments in proteins using circular dichroism spectroscopy, *Protein Science 8*, 370-380.

249. Compton, L. A., and W. Curtis Johnson, J. (1986) Analysis of Protein Circular Dichroism Spectra for Secondary Structure Using a Simple Matrix Multiplication, *Analytical Biochemistry 155*, 155-167.
250. Kay, L. E., Keifer, P., and Saarinen, T. J. (1992) *J. Am. Chem. Soc. 114*, 10663–10665.
251. Ikura, M., Kay, L. E., and Bax, A. (1990) *Biochemistry 29*, 4659–4667.
252. Yamazaki, T., Lee, W., Revington, M., Mattiello, D. L., Dahlquist, F. W., Arrowsmith, C. H., and Kay, L. E. (1994) *J. Am. Chem. Soc. 116*, 6464–6465.
253. Kay, L. E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, *J. Am. Chem. Soc. 114*, 10663-10665.
254. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277–293.
255. Goddard, T. D., and Kneller, D. G. Sparky 3, Univeristy of California, San Francisco.
256. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Wright, P. E., and Dyson, H. J. (2000) Random coil chemical shifts in acidic 8 M urea: Implementations of random coil shift data in NMRView, *Journal of Biomolecular NMR 18*, 43-48.
257. Wishart, D. S., and Nip, A. M. (1998) Protein chemical shift analysis: a practical guide, *Biochem. Cell Biol. 76*, 153-163.
258. Tafer, H., Hiller, S., Hilty, C., Fernandez, C., and Wuthrich, K. (2004) Nonrandom Structure in the Urea-Unfolded Escherichia coli Outer Membrane Protein X (OmpX), *Biochemistry 43*, 860-869.
259. Najbar, L. V., Craik, D. J., Wade, J. D., and McLeish, M. J. (2000) Identification of Initiation Sites for T4 Lysozyme Folding Using CD and NMR Spectroscopy of Peptide Fragments, *Biochemistry 39*, 5911-5920.
260. Nilsson, M. R. (2004) Techniques to study amyloid fibril formation in vitro, *Methods 34*, 151-160.
261. Imamura, K., Mimura, T., Okamoto, M., Sakiyama, T., and K., N. (2000) Adsorption Behavior of Amino Acids on a Stainless Steel Surface, *J Colloid Interface Sci. 229*, 237-246.
262. Rose, G. D., Gierasch, L. M., and Smith, J. A. (1985) Turns in peptides and proteins, *Adv. Protein Chem. 37*, 1-109.
263. Hjelmeland, L. M. (1980) *Proc. Natl. Acad. Sci. USA 77*.
264. Wydro, P., and Paluch, M. (2005) A study of the interaction of dodecyl sulfobetaine with cationic and anionic surfactant in mixed micelles and monolayers at the air/water interface, *Journal of Colloid and Interface Science 286*, 387-391.
265. Mukerjee, P., and Mysels, K. J. (1971) Critical Micelle Concentration of Aqueous Surfactant Systems, *NSRDS-NBS 36, US. Government Printing Office, Washington,.D.C.*
266. Matsumura, Y., Shinjo, M., Mahajan, A., Tsai, M.-D., and Kihara, H. (2010) α-Helical burst on the folding pathway of FHA domains from Rad53 and Ki67 *Biochimie 92*, 1031-1039.
267. Sarkar, P., Reichman, C., Saleh, T., Birge, R. B., and Kalodimos, C. G. (2007) Proline cis-trans Isomerization Controls Autoinhibition of a Signaling Protein, *Molecular Cell 27*, 413-426.
268. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The Chemical Shift Index: A Fast and Simple Method for the Assignment of Protein Secondary Structure through NMR Spectroscopy, *Biochemistry 31*, 1647-1651.

269. Woodward, C. K., Ellis, L. M., and Rosenberg, A. (1975) The Solvent Dependence of Hydrogen Exchange Kinetics of Folded Proteins, *The Journal of Biological Chemistry 250*, 440-444.
270. Baltzer, O. (1996) Structure and dynamics of a designed helix-loop-helix dimer in dilute aqueous trifluoroethanol solution. a strategy for NMR spectroscopic structure determination of molten globules in the rational design of native-like proteins *Folding and Design 1*, 347-356.
271. Bax, A., and Ikura, M. (1991) An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins, *Journal of Biomolecular NMR 1*, 99-104.
272. Shojania, S., Henry, G. D., Chen, V. C., Vo, T. N., Perreault, H., and O'Neil, J. D. (2009) High yield expression and purification of HIV-1 Tat1−72 for structural studies, *Journal of Virological Methods doi:10.1016/j.jviromet.2009.11.021*.
273. Burns, J. A., Butler, J. C., Moran, J., and Whitesides, G. M. (1991) Selective Reduction of Disulfides by Tris(2-carboxyethy1)phosphine, *J. Org. Chem. 56*, 2648-2650.
274. Vo, T. (2008) Spectroscopic Characterization of HIV-1 Transactivator of Transcription (Tat) Interactions with Divalent Metals *Master of Science Thesis, Department of Chemistry, University of Manitoba*.
275. Gregoire, C., Jr., J.-M. P., Esquieu, D., Opi, S., Campbell, G., Solomiac, M., Lebrun, E., Lebreton, J., and Loret, E. P. (2001) Homonuclear ^{1}H-NMR Assignment and Structural Characterization of Human Immunodeficiency Virus Type 1 Tat Mal Protein, *Biopolymers 62*, 324-335.
276. Campbell, G. R., Senkaali, D., Watkins, J., Esquieu, D., Opi, S., Yirrell, D. L., Kaleebu, P., and Loret, E. P. (2007) Tat mutations in an African cohort that do not prevent transactivation but change its immunogenic properties, *Vaccine 25*, 8441-8447.
277. Brahmachari, S. K., Ananthanarayanan, V. S., Brahms, S., Brahms, J., Rapaka, R. S., and Bhatnagar, R. S. (1979) Vacuum Ultraviolet Circular Dichroism Spectrum of B-turn in Solution, *Biochemical and Biophysical Research Communications 86*, 605-612.
278. Good, N. E., Winget, G. D., Winter, W., Connolly, T. N., Izawa, S., and Singh, R. M. M. (1966) Hydrogen Ion Buffers For Biological Research, *Biochemistry 5*, 467-477.
279. Dawson, R. M. C. (1959) Data for Biochemical Research, *Oxford Clarendon Press*.
280. Gregoire, C. J., and Loret, E. P. (1996) Conformational Heterogeneity in Two Regions of TAT Results in Structural Variations of This Protein as a Function of HIV-1 Isolates, *The Journal of Biological Chemistry 271*, 22641-22646.
281. Anthony, J. S., and Moscarello, M. A. (1971) A Conformation Change Induced in the Basic Encephalitogen by Lipids, *Biochimica et Biophysica Acta 243*, 429-433.
282. Lohner, K., and Prenner, E. J. (1999) Differential scanning calorimetry and X-ray diffraction studies of the specificity of the interaction of antimicrobial peptides with membrane-mimetic systems, *Biochim. Biophys. Acta 1462*, 141–156.
283. Giehm, L., Oliveira, C. L. P., Christiansen, G., Pedersen, J. S., and Otzen, D. E. (2010) SDS-Induced Fibrillation of α-Synuclein: An Alternative Fibrillation Pathway, *J. Mol. Biol. 401*, 115-133.
284. Calnan, B. J., Biancalana, S., Hudson, D., and Frankel, A. D. (1991) Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition, *Genes & Development 5*, 201-210.

285. Metzger, A. U., Bayer, P., Willbold, D., Hoffmann, S., Frank, R. W., Goody, R. S., and Rosch, P. (1997) The Interaction of HIV-1 Tat(32-72) with its Target RNA: A Fluorescence and Nuclear Magnetic Resonance Study, *Biochemical and Biophysical Research Communications 241*, 31-36.
286. Long, K. S., and Crothers, D. M. (1999) Characterization of the Solution Conformations of Unbound and Tat Peptide-Bound Forms of HIV-1 TAR RNA, *Biochemistry 38*, 10059-10069.
287. Tan, R., and Frankel, A. D. (1992) Circular Dichroism Studies Suggest That TAR RNA Changes Conformation upon Specific Binding of Arginine or Guanidine, *Biochemistry 31*, 10288-10294.
288. Janin, J. (1997) Specific versus non-specific contacts in protein crystals, *Nature Struct. Biol. 4*, 973–974.
289. Thomas, A. S., and Elcock, A. H. (2007) Molecular Dynamics Simulations of Hydrophobic Associations in Aqueous Salt Solutions Indicate a Connection between Water Hydrogen Bonding and the Hofmeister Effect, *J. Am. Chem. Soc. 129*, 14887-14898.
290. Howie, A. J., and Brewer, D. B. (2009) Optical properties of amyloid stained by Congo red: History and mechanisms, *Micron 40*, 285-301.
291. Johnson Jr., W. C. (1988) Secondary structure of proteins through circular dichroism spectroscopy, *Annu. Rev. Biophys. Bio. 17*, 145-166.
292. Corrêa, D. H. A., and Ramos, C. H. I. (2009) The use of circular dichroism spectroscopy to study protein folding, form and function, *African Journal of Biochemistry Research 3*, 164-173.
293. Gans, P. J., Lyu, P. C., Manning, M. C., Woody, R. W., and Kallenbach, N. R. (1991) The helix-coil transition in heterogeneous peptides with specific sidechain interactions: theory and comparison with CD spectral data, *Biopolymers 31*, 1605-1614.
294. Kumar, S., Modig, K., and Halle, B. (2003) Trifluoroethanol-Induced β → α Transition in β-Lactoglobulin: Hydration and Cosolvent Binding Studied by 2H, 17O, and 19F Magnetic Relaxation Dispersion, *Biochemistry 42*, 13708-13716.
295. Ramirez-Alvarado, M., Serrano, L., and Blanco, F. J. (1997) Conformational analysis of peptides corresponding to all the secondary structure elements of protein L Bl domain: Secondary structure propensities are not conserved in proteins with the same fold, *Protein Science 6*, 162-174.
296. Ramirez-Alvarado, M., Serrano, L., and Blanco, F. J. (1996) Conformational analysis of peptides corresponding to all the secondary structure elements of protein L Bl domain: Secondary structure propensities are not conserved in proteins with the same fold, *Protein Science 6*, 162-174.
297. Lundberg, M., Wikstrom, S., and Johansson, M. (2003) Cell Surface Adherence and Endocytosis of Protein Transduction Domains, *Molecular Therapy 8*, 143-150.

Appendix A

CD Deconvolution Data

A1. Deconvolution data for Tat in 0% - 90% TFE

tat0%	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.017	0.022	0.008	0.019	0.038	0.908	1.013
Provencher & Glockner	0.062	0.105	0.064	0.088	0.253	0.427	0.999
CDSSTR	0.05	0.1	0.15	0.1	0.24	0.35	0.99
Average	0.043	0.076	0.074	0.069	0.177	0.562	1.001

Tat10%	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.034	0.044	0.061	0.048	0.1	0.732	1.019
Provencher & Glockner	0.047	0.049	0.108	0.098	0.268	0.429	0.999
CDSSTR	0.03	0.05	0.18	0.1	0.24	0.4	1
Average	0.037	0.048	0.116	0.082	0.203	0.520	1.006

Tat20%	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.103	0.093	0.091	0.067	0.162	0.519	1.035
Provencher & Glockner	0.15	0.143	0.021	0.069	0.252	0.365	1
CDSSTR	0.18	0.17	0.09	0.08	0.21	0.27	1
Average	0.144	0.135	0.067	0.072	0.208	0.385	1.012

Tat 30%	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.185	0.158	0.069	0.072	0.214	0.307	1.005
Provencher & Glockner	0.224	0.189	0.004	0.055	0.235	0.292	0.999
CDSSTR	0.27	0.19	0.07	0.07	0.17	0.24	1.01
Average	0.226	0.179	0.048	0.066	0.206	0.280	1.005

Tat40%	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.172	0.135	0.128	0.089	0.224	0.284	1.032
Provencher & Glockner	0.235	0.199	0	0.049	0.217	0.301	1.001
CDSSTR	0.28	0.19	0.05	0.06	0.17	0.25	1
Average	0.229	0.175	0.059	0.066	0.204	0.278	1.011

tat50%	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.223	0.157	0.091	0.07	0.23	0.271	1.041

Provencher & Glockner	0.248	0.19	0.001	0.05	0.215	0.296	1
CDSSTR	0.31	0.2	0.04	0.06	0.15	0.25	1.01
Average	0.260	0.182	0.044	0.060	0.198	0.272	1.017

Tat 60% TFE	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.302	0.175	0.068	0.056	0.181	0.257	1.039
Provencher & Glockner	0.326	0.199	0	0.04	0.162	0.272	0.999
CDSSTR	0.39	0.23	0.04	0.04	0.09	0.21	1
Average	0.339	0.201	0.036	0.045	0.144	0.246	1.013

tat 70% TFE	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.297	0.173	0.069	0.057	0.181	0.256	1.032
Provencher & Glockner	0.311	0.198	0.006	0.042	0.169	0.273	0.999
CDSSTR	0.37	0.23	0.02	0.03	0.11	0.23	0.99
Average	0.326	0.200	0.032	0.043	0.153	0.253	1.007

Tat 80%	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.358	0.195	0.044	0.044	0.156	0.235	1.033
Provencher & Glockner	0.371	0.212	0	0.033	0.141	0.243	1
CDSSTR	0.43	0.27	0.03	0.03	0.06	0.19	1.01
Average	0.386	0.226	0.025	0.036	0.119	0.223	1.014

Tat 90%	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.403	0.197	0.032	0.031	0.122	0.211	0.996
Provencher & Glockner	0.393	0.206	0.003	0.032	0.122	0.244	1
CDSSTR	0.5	0.26	0.03	0.03	0.04	0.14	1
Average	0.432	0.221	0.022	0.031	0.095	0.198	0.999

A2. Deconvolution Data for Tat in 20% TFE at 5℃, 25℃ and 80℃

Tat 5°C, 20% TFE	Helix1	Helix2	Strand1	Strand2	Turns	Disordered	Total
Self-Consistent Method	0.086	0.102	0.11	0.093	0.24	0.375	1.005
Provencher & Glockner	0.089	0.102	0.108	0.093	0.237	0.371	1
CDSSTR	0.05	0.04	0.09	0.05	0.1	0.67	1
Average	0.075	0.081	0.103	0.079	0.192	0.472	1.002

Tat 25°C, 20% TFE	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Disordered	Total
Selcon3	0.103	0.093	0.091	0.067	0.162	0.519	1.035
Provencher & Glockner	0.15	0.143	0.021	0.069	0.252	0.365	1
CDSSTR	0.18	0.17	0.09	0.08	0.21	0.27	1
Average	0.144	0.135	0.067	0.072	0.208	0.385	1.012

Tat 80°C, 20% TFE	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Disordered	Total
Selcon3	0.094	0.105	0.126	0.094	0.227	0.351	0.997
Provencher & Glockner	0.099	0.109	0.124	0.096	0.226	0.345	0.999
CDSSTR	0.09	0.07	0.08	0.06	0.12	0.59	1.01
Average	0.0943	0.0947	0.11	0.08333	0.191	0.4286667	1.002

A3. Deconvolution data for Zn^{+2} titration of Tat in 30% TFE

Apo-Tat 30%TFE	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.185	0.168	0.052	0.062	0.217	0.307	
Provencher & Glockner	0.223	0.193	0.006	0.052	0.238	0.288	1.000
CDSSTR	0.240	0.180	0.070	0.070	0.180	0.280	1.020
Average	0.216	0.180	0.043	0.061	0.212	0.292	1.000

0.5 MEQ Zn Tat 30%TFE	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.167	0.148	0.086	0.072	0.225	0.325	1.024
Provencher & Glockner	0.189	0.164	0.035	0.061	0.239	0.312	1.000
CDSSTR	0.190	0.150	0.060	0.060	0.170	0.390	1.020
Average	0.182	0.154	0.060	0.064	0.211	0.342	1.015

1 Meq Zn Tat 30%TFE	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.162	0.146	0.088	0.074	0.224	0.325	1.018
Provencher & Glockner	0.180	0.157	0.048	0.065	0.235	0.316	1.001
CDSSTR	0.170	0.150	0.050	0.060	0.180	0.390	1.000
Average	0.171	0.151	0.062	0.066	0.213	0.344	1.006

2 Meq Zn Tat 30%TFE	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.152	0.139	0.100	0.079	0.225	0.322	1.018
Provencher & Glockner	0.170	0.149	0.060	0.068	0.235	0.319	1.001
CDSSTR	0.170	0.130	0.060	0.060	0.170	0.410	1.000
Average	0.164	0.139	0.073	0.069	0.210	0.350	1.006

3 Meq Tat 30%TFE	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.158	0.145	0.083	0.069	0.227	0.324	1.006
Provencher & Glockner	0.168	0.151	0.062	0.067	0.234	0.317	0.999
CDSSTR	0.170	0.120	0.060	0.050	0.150	0.450	1.000
Average	0.165	0.139	0.068	0.062	0.204	0.364	1.002

4 Meq Tat 30%TFE	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.152	0.142	0.095	0.081	0.232	0.335	1.037
Provencher & Glockner	0.164	0.141	0.068	0.071	0.231	0.325	1.000
CDSSTR	0.170	0.140	0.060	0.060	0.180	0.400	1.010
Average	0.162	0.141	0.074	0.071	0.214	0.353	1.016

A4. Deconvolution data for pH titration of Tat in 30% TFE with 2 Meq's of zinc(II) acetate

pH 4.69	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
Self-consistent Method	0.189	0.159	0.069	0.071	0.224	0.306	1.019
Provencher & Glockner	0.227	0.199	0.000	0.043	0.233	0.298	1.000
CDSSTR	0.250	0.190	0.050	0.060	0.170	0.280	1.000
Average	0.222	0.183	0.040	0.058	0.209	0.295	1.006

pH 5.05	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
Self-consistent Method	0.195	0.164	0.039	0.056	0.216	0.325	0.995
Provencher & Glockner	0.216	0.182	0.016	0.050	0.214	0.322	1.000
CDSSTR	0.220	0.160	0.050	0.050	0.150	0.380	1.010
Average	0.210	0.169	0.035	0.052	0.193	0.342	1.002

pH 5.56	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
Self-consistent Method	0.185	0.164	0.053	0.058	0.223	0.332	1.014
Provencher & Glockner	0.208	0.180	0.015	0.053	0.224	0.320	1.000
CDSSTR	0.210	0.160	0.050	0.050	0.160	0.380	1.010
Average	0.201	0.168	0.039	0.054	0.202	0.344	1.008

pH 6.35	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
Self-consistent Method	0.169	0.155	0.084	0.069	0.206	0.309	0.992
Provencher & Glockner	0.197	0.170	0.028	0.056	0.224	0.325	1.000
CDSSTR	0.210	0.160	0.050	0.060	0.160	0.370	1.010
Average	0.192	0.162	0.054	0.062	0.197	0.335	1.001

pH 7.20	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
Self-consistent Method	0.189	0.161	0.046	0.058	0.227	0.331	1.013
Provencher & Glockner	0.192	0.168	0.039	0.057	0.223	0.322	1.001
CDSSTR	0.200	0.140	0.070	0.050	0.140	0.400	1.000
Average	0.194	0.156	0.052	0.055	0.197	0.351	1.005

A5. Deconvolution data for Tat in SDS.

14 uM Tat	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.036	0.069	0.13	0.1	0.237	0.404	0.976
Provencher & Glockner	0.023	0.066	0.149	0.11	0.246	0.407	1.001
CDSSTR	0.01	0.05	0.2	0.12	0.24	0.38	1
Average	0.023	0.062	0.160	0.110	0.241	0.397	1.000

S

14 uM Tat, 1.75 mM SDS	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unordered	Total
Self-consistent Method	0.123	0.096	0.155	0.094	0.204	0.303	0.975
Provencher & Glockner	0.129	0.126	0.131	0.087	0.227	0.3	1
CDSSTR	0.12	0.09	0.11	0.07	0.16	0.44	0.99
Average	0.124	0.104	0.132	0.084	0.197	0.348	0.988

Appendix B

Resonance Assignments for His-tagged Tat_{1-72}

B1. Resonance Assignments for His-tagged Tat1-72 at pH 4 and 293 K

Position	Residue	H^N (ppm)	N (ppm)
1	Methionine	—	—
2	Glycine	—	—
3	Serine	8.649	115.657
4	Serine	8.452	117.948
5	Histidine	8.521	120.132
6	Histidine	8.49	119.296
7	Histidine	8.66	120.118
8	Histidine	8.733	120.617
9	Histidine	8.739	121.062
10	Histidine	8.722	121.725
11	Serine	8.528	118.878
12	Serine	8.54	118.507
13	Glycine	8.398	110.719
14	Leucine	8.099	121.914
15	Valine	8.18	123.43
16	Proline	N/A	N/A
17	Arginine	8.466	122.162
18	Glycine	8.454	110.494
19	Serine	8.161	115.402
20	Histidine	8.578	120.259
21	Methionine	—	—
22	Glutamate	8.481	124.612
23	Proline	N/A	N/A
24	Valine	8.167	120.583
25	Aspartate	8.374	125.984
26	Proline	N/A	N/A
27	Arginine	8.497	121.994
28	Leucine	7.858	120.295
29	Glutamate	7.929	120.538
30	Proline	N/A	N/A
31	Tryptophan	8.359	118.833
32	Lysine	8.384	121.949
33	Histidine	8.079	119.276
34	Proline	N/A	N/A
35	Glycine	8.577	109.906
36	Serine	8.267	115.657
37	Glutamine	8.519	122.236
38	Proline	N/A	N/A
39	Lysine	—	—
40	Threonine	8.052	115.158

41	Alanine	8.343	126.531
42	Cysteine	8.38	119.078
43	Threonine	8.24	116.508
44	Asparagine	8.388	121.047
45	Cysteine	8.212	119.283
46	Tyrosine	8.259	122.66
47	Cysteine	8.088	121.149
48	Lysine	7.593	122.39
49	Lysine	—	—
50	Cysteine	8.326	120.749
51	Cysteine	8.35	121.774
52	Phenylalanine	—	—
53	Histidine	8.371	120.706
54	Cysteine	—	—
55	Glutamine	—	—
56	Valine	8.289	122.352
57	Cysteine	8.561	123.461
58	Phenylalanine	—	—
59	Isoleucine	8.083	123.07
60	Threonine	8.142	119.108
61	Lysine	—	—
62	Alanine	—	—
63	Leucine	8.177	121.682
64	Glycine	8.3	109.244
65	Isoleucine	7.915	120.026
66	Serine	8.278	119.258
67	Tyrosine	8.237	122.899
68	Glycine	8.304	110.035
69	Arginine	8.101	120.653
70	Lysine	8.293	122.912
71	Lysine	8.306	123.003
72	Arginine	8.458	123.589
73	Arginine	8.408	123.147
74	Glutamine	8.476	122.771
75	Arginine	8.441	122.562
76	Arginine	8.487	123.745
77	Arginine	—	—
78	Proline	N/A	N/A
79	Proline	N/A	N/A
80	Glutamine	8.514	121.035
81	Glycine	8.474	110.652
82	Serine	8.186	115.565
83	Glutamine	8.36	123.106
84	Threonine	8.106	114.816
85	Histidine	8.497	120.634
86	Glutamine	8.447	122.531
87	Valine	8.309	122.162
88	Serine	8.399	119.682
89	Leucine	8.385	125.048
90	Serine	8.258	117.193

91	Lysine	8.295	123.759
92	Glutamine	8.028	126.443

B2. Resonance Assignments of His-tagged Tat1-72 in 7% TFE at pH 4.01 and 298 K

Position	Residue	H^N (ppm)	N (ppm)
1	Methionine	—	—
2	Glycine	—	—
3	Serine	8.628	115.591
4	Serine	8.425	117.843
5	Histidine	8.5	119.893
6	Histidine	8.458	118.982
7	Histidine	8.629	119.802
8	Histidine	8.696	120.299
9	Histidine	8.711	120.749
10	Histidine	8.699	121.393
11	Serine	8.499	118.563
12	Serine	8.505	118.323
13	Glycine	8.376	110.543
14	Leucine	8.051	121.711
15	Valine	8.091	122.752
16	Proline	N/A	N/A
17	Arginine	8.417	121.952
18	Glycine	8.422	110.276
19	Serine	8.113	115.266
20	Histidine	8.539	119.985
21	Methionine	—	—
22	Glutamate	8.399	124.097
23	Proline	N/A	N/A
24	Valine	8.092	120.132
25	Aspartate	8.29	125.494
26	Proline	N/A	N/A
27	Arginine	8.453	121.664
28	Leucine	7.796	119.839
29	Glutamate	7.855	119.758
30	Proline	N/A	N/A
31	Tryptophan	8.312	118.577
32	Lysine	8.325	121.395
33	Histidine	8.037	118.942
34	Proline	N/A	N/A
35	Glycine	8.561	109.814
36	Serine	8.22	115.495
37	Glutamine	8.486	122.024
38	Proline	N/A	N/A
39	Lysine	—	—
40	Threonine	7.979	114.493
41	Alanine	8.286	126.109
42	Cysteine	8.342	118.471

43	Threonine	8.184	116.118
44	Asparagine	8.339	120.764
45	Cysteine	8.168	118.878
46	Tyrosine	8.176	122.458
47	Cysteine	8.076	120.89
48	Lysine	7.562	122.009
49	Lysine	—	—
50	Cysteine	8.271	119.985
51	Cysteine	8.275	121.351
52	Phenylalanine	—	—
53	Histidine	8.283	120.028
54	Cysteine	—	—
55	Glutamine	—	—
56	Valine	8.221	122.037
57	Cysteine	8.497	122.927
58	Phenylalanine	—	—
59	Isoleucine	8.021	122.256
60	Threonine	8.035	117.909
61	Lysine	—	—
62	Alanine	—	—
63	Leucine	8.147	121.384
64	Glycine	8.222	108.821
65	Isoleucine	7.809	119.561
66	Serine	8.218	118.808
67	Tyrosine	8.145	122.48
68	Glycine	8.264	109.73
69	Arginine	8.05	120.468
70	Lysine	—	—
71	Lysine	—	—
72	Arginine	8.383	123.014
73	Arginine	8.351	122.848
74	Glutamine	8.413	122.251
75	Arginine	8.383	122.187
76	Arginine	8.412	123.287
77	Arginine	—	—
78	Proline	N/A	N/A
79	Proline	N/A	N/A
80	Glutamine	8.481	120.817
81	Glycine	8.444	110.414
82	Serine	8.139	115.372
83	Glutamine	8.279	122.595
84	Threonine	8.056	114.407
85	Histidine	8.442	120.35
86	Glutamine	8.385	122.144
87	Valine	8.249	121.802
88	Serine	8.34	119.323
89	Leucine	8.327	124.778
90	Serine	8.183	116.987
91	Lysine	8.241	123.484
92	Glutamine	8	125.833

B3. Resonance Assignments of His-tagged Tat1-72 in 15% TFE at pH 4.05 and 298 K

Position	Residue	H-N (ppm)	N (ppm)
1	Methionine	—	—
2	Glycine	—	—
3	Serine	8.614	115.518
4	Serine	8.406	117.788
5	Histidine	8.477	119.666
6	Histidine	8.405	119.128
7	Histidine	8.605	119.565
8	Histidine	8.678	120.051
9	Histidine	8.693	120.488
10	Histidine	8.681	121.126
11	Serine	8.434	118.748
12	Serine	8.474	118.23
13	Glycine	8.352	110.369
14	Leucine	8.007	121.51
15	Valine	7.847	122.456
16	Proline	N/A	N/A
17	Arginine	—	—
18	Glycine	8.402	110.021
19	Serine	8.064	115.067
20	Histidine	8.506	119.725
21	Methionine	—	—
22	Glutamate	8.319	123.565
23	Proline	N/A	N/A
24	Valine	8.015	119.607
25	Aspartate	8.153	125.262
26	Proline	N/A	N/A
27	Arginine	8.407	121.318
28	Leucine	7.641	118.839
29	Glutamate	7.809	119.848
30	Proline	N/A	N/A
31	Tryptophan	8.245	118.073
32	Lysine	8.284	120.985
33	Histidine	7.955	118.424
34	Proline	N/A	N/A
35	Glycine	8.548	109.667
36	Serine	8.182	115.356
37	Glutamine	8.451	120.628
38	Proline	N/A	N/A
39	Lysine	—	—
40	Threonine	7.913	113.846
41	Alanine	8.228	125.661
42	Cysteine	8.315	118.054
43	Threonine	8.126	115.72

44	Asparagine	8.291	120.526
45	Cysteine	8.077	118.818
46	Tyrosine	—	—
47	Cysteine	7.975	120.173
48	Lysine	7.498	121.414
49	Lysine	—	—
50	Cysteine	—	—
51	Cysteine	8.21	120.755
52	Phenylalanine	—	—
53	Histidine	—	—
54	Cysteine	—	—
55	Glutamine	—	—
56	Valine	—	—
57	Cysteine	8.464	121.892
58	Phenylalanine	—	—
59	Isoleucine	7.997	121.993
60	Threonine	8.104	117.761
61	Lysine	—	—
62	Alanine	—	—
63	Leucine	8.04	120.725
64	Glycine	7.989	107.691
65	Isoleucine	7.742	118.653
66	Serine	8.15	118.634
67	Tyrosine	—	—
68	Glycine	8.179	109.012
69	Arginine	7.976	119.547
70	Lysine	—	—
71	Lysine	—	—
72	Arginine	8.304	122.583
73	Arginine	8.261	122.141
74	Glutamine	8.365	121.822
75	Arginine	8.283	121.7
76	Arginine	8.334	122.745
77	Arginine	—	—
78	Proline	N/A	N/A
79	Proline	N/A	N/A
80	Glutamine	8.451	120.628
81	Glycine	8.427	110.275
82	Serine	8.1	115.204
83	Glutamine	—	—
84	Threonine	8.012	114.006
85	Histidine	8.384	120.038
86	Glutamine	8.3	121.68
87	Valine	—	—
88	Serine	8.273	118.901
89	Leucine	8.263	124.492
90	Serine	8.161	116.875
91	Lysine	8.187	123.22
92	Glutamine	7.954	125.53

B3. Resonance Assignments of His-tagged Tat1-72 in 30% TFE at pH 4.2 and 298 K

Position	Residue	H^N (ppm)	N (ppm)	C^α (ppm)
1	Methionine	—	—	—
2	Glycine	—	—	43.392
3	Serine	8.613	115.307	58.194
4	Serine	8.398	117.531	55.417
5	Histidine	8.468	118.13	58.245
6	Histidine	8.462	117.897	55.205
7	Histidine	8.649	119.14	55.384
8	Histidine	8.708	119.531	55.402
9	Histidine	8.703	120.02	55.499
10	Histidine	8.672	120.764	54.876
11	Serine	8.473	118.355	55.284
12	Serine	8.477	118.531	58.557
13	Glycine	8.336	110.093	45.21
14	Leucine	7.985	121.385	55.051
15	Valine	7.936	121.604	—
16	Proline	N/A	N/A	63.07
17	Arginine	8.44	120.416	56.413
18	Glycine	8.374	109.665	65.57
19	Serine	8.077	112.933	—
20	Histidine	—	—	—
21	Methionine	—	—	—
22	Glutamate	—	—	—
23	Proline	N/A	N/A	—
24	Valine	—	—	58.033
25	Aspartate	8.223	124.235	—
26	Proline	N/A	N/A	—
27	Arginine	—	—	56.859
28	Leucine	7.73	119.307	54.769
29	Glutamate	7.651	118.705	—
30	Proline	N/A	N/A	—
31	Tryptophan	—	—	—
32	Lysine	—	—	—
33	Histidine	—	—	—
34	Proline	N/A	N/A	63.564
35	Glycine	8.512	109.43	45.394
36	Serine	8.181	115.144	58.367
37	Glutamine	8.475	119.439	—
38	Proline	N/A	N/A	—
39	Lysine	—	—	56.704
40	Threonine	7.929	113.793	61.672
41	Alanine	8.225	125.45	52.821
42	Cysteine	8.24	117.778	58.872
43	Threonine	8.092	115.309	62.521
44	Asparagine	8.301	120.455	54.014
45	Cysteine	8.21	118.502	56.39

46	Tyrosine	8.146	121.64	—
47	Cysteine	—	—	57.298
48	Lysine	7.557	121.511	55.255
49	Lysine	8.215	121.869	56.39
50	Cysteine	8.118	119.431	56.571
51	Cysteine	8.22	120.161	—
52	Phenylalanine	—	—	—
53	Histidine	—	—	—
54	Cysteine	—	—	—
55	Glutamine	—	—	—
56	Valine	8.134	121.308	—
57	Cysteine	8.476	121.767	—
58	Phenylalanine	—	—	—
59	Isoleucine	—	—	—
60	Threonine	—	—	—
61	Lysine	8.272	123.127	—
62	Alanine	—	—	—
63	Leucine	—	—	—
64	Glycine	7.93	106.884	63.018
65	Isoleucine	7.901	119.056	—
66	Serine	8.128	118.859	—
67	Tyrosine	—	—	59.837
68	Glycine	8.153	107.785	46.176
69	Arginine	7.951	119.872	57.399
70	Lysine	8.075	120.638	—
71	Lysine	8.164	120.659	—
72	Arginine	8.37	121.5	—
73	Arginine	8.299	121.495	—
74	Glutamine	8.374	121.116	—
75	Arginine	—	—	—
76	Arginine	—	—	—
77	Arginine	—	—	—
78	Proline	N/A	N/A	—
79	Proline	N/A	N/A	—
80	Glutamine	—	—	56.066
81	Glycine	8.421	109.909	45.214
82	Serine	8.071	114.876	58.363
83	Glutamine	8.148	122.529	56.147
84	Threonine	8.006	113.674	62.04
85	Histidine	8.366	119.774	55.321
86	Glutamine	8.37	121.5	56.065
87	Valine	8.176	121.15	62.49
88	Serine	8.229	118.522	58.124
89	Leucine	8.335	122.486	—
90	Serine	8.202	118.438	59.738
91	Lysine	8.217	122.051	—
92	Glutamine	—	—	—