Biophysical Characterization of NleD:
An effector protein from *Escherichia coli* O157:H7

Abhishek Chatterjee

A Thesis

in

the Department

of

Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Chemistry) at

Concordia University

Montréal, Quebec, Canada

September 2012

© Abhishek Chatterjee, 2012

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Abhishek Chatterjee

Entitled: Biophysical Characterization of NleD: An effector protein from *Escherichia coli* O157:H7

and submitted in partial fulfillment of the requirements for the degree of

M.Sc. Chemistry

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Paul Joyce Chair

Dr. Peter Pawelek Examiner

Dr. Christine DeWolf Examiner

Dr. Irena Ekiel and Dr. Joanne Turnbull Supervisors

Approved by: Dr. Heidi Muchall
Chair of Department or Graduate Program Director

Dean of Faculty

Date: August 27th, 2012

Abstract

Biophysical Characterization of NleD: An effector protein from *Escherichia coli* O157:H7

Abhishek Chatterjee

Pathogenic *Escherichia coli* strains (O157:H7) are one of the main causes behind the lethal *E. coli* outbreaks in North America, UK and Japan. NleD, a zinc-dependent endopeptidase, is one of the several effector proteins secreted by these *E. coli* strains. After being injected directly into the host cell cytoplasm, NleD regulates the activity of specific MAP kinases (p38α kinase and JNK) leading to the inhibition of the host inflammatory response-signaling network. The current thesis is focused on the biophysical properties of NleD. NleD was identified as a monomer based on the results obtained from size exclusion chromatography, dynamic light scattering and analytical ultracentrifugation. The protein was prone to proteolytic cleavage at its C-terminal region as determined by mass spectrometry. Different approaches to improve the stability of purified NleD were also identified. NleD was found to have enhanced stability at lower temperatures and in the presence of trehalose. Using bioinformatic analysis as a tool, we were able to predict putative similarity between the active sites of botulinum neurotoxin and *E. coli* O157:H7 NleD. Preliminary characterization of the proteolytic activity of wild-type and variant forms of NleD was performed using p38α kinase as a substrate. Residues important for catalysis were identified. Our approach for improving the stability of the enzyme might facilitate its crystallization for structural work, while further mutagenesis studies may help identify the role of active site residues involved in catalysis.

Acknowledgements

I would like to give my deepest and most heartfelt thanks to my supervisor, Dr. Irena Ekiel. The advice and support she gave throughout my project was phenomenal. Her insistence on attention to detail has taught me many lessons that will guide me throughout my career ahead.

I would also like to thank my co-supervisor Dr. Joanne Turnbull, for many interesting discussions, enthusiasm and efforts.

I also wish to thank my committee members, Drs. Christine DeWolf and Peter Pawelek for all their insight, suggestions, help and patience. To my colleague Ms. Laura Mc Donald, I would like to thank for all her help and wish her a successful career ahead.

I would also like to thank all my family members for their guidance throughout my schooling and development.

Last but not the least, I would also like to thank CIHR for financially supporting my research project.

I Dedicate My Thesis to Almighty God, Friends and Family

Table of Contents

List of figures……..…….……viii

List of tables……..…….…….……x

List of abbreviations……...…….……..……....xi

Chapter 1: Introduction……...……1

1.1 Pathogenic *Escherichia coli* (EHEC)……1

1.2 Bacterial effectors…….……...……2

1.3 NleD: An effector secreted from *E. coli.* O157:H7 strain…….....2

1.4 C-Jun-N Terminal Kinase (JNK) and p38α protein kinase……..…6

1.5 Zinc-dependent metalloproteases…….…….....9

1.6 Zinc ion binding site in zincins……..……...……,……..10

1.7 Active site of botulinum neurotoxin and thermolysin……...……..12

Biophysical methods used for protein characterization…….…….……....15

1.8 Characterization using mass spectrometry…….…….....15

1.9 Spectroscopic characterization of proteins…….…….……....16

1.9.1 Circular dichroism (CD) spectroscopy…….……...…..……16

1.9.2 The importance of intrinsic tryptophan fluorescence emission……...……..17

1.10 Characterization of the oligomeric states of proteins….…..……...…….……...…....18

1.10.1 Size exclusion chromatography (SEC)……..……..….……..19

1.10.2 Dynamic light scattering (DLS)…….20

1.10.3 Analytical ultracentrifugation (AUC)……..…… 21

1.11 Characterization of histidine residues in proteins by NMR…….……...…..……22

1.12 Research outline….……..……..…24

Chapter 2: Materials and Methods……..……..26

2.1 Chemicals and reagents…….…...…...26

2.2 Site-directed mutagenesis……..……..……..…...……...26

2.3 Protein expression and purification of wild –type and variant forms of NleD…..……....29

2.4 Determination of protein concentration……...…..30

2.5 Mass spectrometry of purified wild-type NleD……..……..…….……...…..30

2.6 Far-UV CD spectroscopy of NleD…….…….….…..31

2.7 Fluorescence spectroscopy of NleD and the variants…..……...31

2.8 Quaternary structure determination…….32
2.8.1Size exclusion chromatography……..…….……..32
2.8.2 Dynamic light scattering……...……33
2.8.3 Analytical ultracentrifugation…….……..……....33
2.9 1D-^{1}H NMR for NleD…….……....34
2.10 Protein expression and purification of p38α kinase……..34
2.11 Proteolytic assays for wild-type NleD and variants…..……...…….....35
Chapter 3: Results……..……..37
3.1 Expression and purification of NleD and the variants….……..…….....37
3.2 Characterization of NleD using electrospray ionization mass spectrometry…….38
3.3 Analysis of secondary structure using far-UV CD spectroscopy…….……..…43
3.4 Quaternary structure determination of NleD using different techniques….…...…….46
3.5 Characterization of NleD solubility at different temperatures and in the presence of additives…….....54
3.6 Mutagenesis of putative active site residues of NleD…….……..……57
3.7 Thermal stability studies of wild-type NleD and the variants using far-UV CD…...…...59
3.8 Zinc binding studies using intrinsic tryptophan fluorescence as a probe…….…….61
3.9 Characterization of histidines in NleD…….…….68
3.10 Characterization of the proteolytic activity of wild-type NleD and variants…….72
Chapter 4: Discussion……..….……..80
Chapter 5: Summary and Future Directions...…….…….……..87
References…….……...89

List of Figures

Fig. 1.3.1 A scheme illustrating the function of NleD in its anti-inflammatory activity…….4

Fig. 1.3.2 Multiple sequence alignment of NleD and other closely related effectors found in different bacteria……...5

Fig. 1.4.1 Multiple sequence alignment of JNK isoforms and p38α kinase isoforms……....……7

Fig. 1.4.2 Structure of p38α kinase and JNK showing the activation loop and the cleavage site for NleD……...8

Fig. 1.5 Families of zinc metalloproteases and their interrelationship based on the sequence around the zinc binding residues……..……...…...9

Fig. 1.6.1 The active site of thermolysin-like protease……..11

Fig. 1.6.2 Active site of metzincin-like protease……...12

Fig. 1.7 Hydrolytic mechanism of thermolysin and its comparison to botulinum neurotoxin A light chain……...14

Fig. 1.11 1D-^{1}H NMR spectra of NleD. ……...…….…..……...…..22

Fig. 3.1.1 SDS-PAGE analysis of NleD proteins. …….…….…….38

Fig. 3.2.1 Mass spectrum of wild-type NleD……..…….……..29

Fig. 3.2.2 Deconvolved ESI–MS spectrum of purified NleD …….….....…41

Fig. 3.2.3 The primary sequence of purified NleD and the cleavage sites based on the analysis of mass spectrometry data……...……..42

Fig. 3.3.1 Far- UV CD spectrum of wild-type NleD……..……....…...44

Fig. 3.3.2 Secondary structure prediction of NleD using Psipred….……45

Fig. 3.4.1 Logarithmic plot of the standard molecular weight markers and their respective distribution coefficients……...……...……..…..…49

Fig. 3.4.2 Elution profile of wild-type NleD obtained from Superdex 75 chromatography….....50

Fig. 3.4.3 Dynamic light scattering analysis of wild-type NleD……...……....52

Fig. 3.4.4 Sedimentation velocity analysis of wild-type NleD …….…….....…53

Fig. 3.5 Time dependent precipitation of NleD under various experimental conditions.....56

Fig. 3.6 The active site of botulinum neurotoxin……..…...……..58

Fig. 3.7.1 The variation of molar ellipticity at 222 nm as a function of temperature for NleD proteins……60

Fig 3.8.1 Intrinsic fluorescence emission spectra for wild-type NleD and the variants……64
Fig 3.8.2 Changes in fluorescence intensity plotted as a function of increasing zinc chloride concentration…….....66
Fig 3.8.3 A plot of λ_{max} as a function of increasing zinc chloride concentrations…….....67
Fig. 3.9.11D-^{1}H NMR spectra of wild-type NleD……...…….…...……...……70
Fig. 3.9.2 Changes in the ppm shift as a function of increasing pD…….……...…….....71
Fig. 3.10.1SDS-PAGE analysis of the purification of the recombinant hexa-His tagged p38α kinase….....…….……...…73
Fig. 3.10.2 SDS-PAGE based enzyme assays for wild type NleD and the variants…….…….75
Fig. 3.10.3 SDS-PAGE based enzyme assay for wild-type NleD using as a substrate p38α kinase without its hexa-His tag……...…….....…….……...77
Fig. 3.10.4 Quantification of the NleD - catalysed cleavage reaction of p38α kinase…….……..78

List of Tables

Table 2.2.1 List of forward and reverse primers designed to incorporate the desired point mutations in wild-type NleD…….……...27

Table 2.2.2 PCR reaction conditions for site-directed mutagenesis…….……..28

Table 3.2 The comparison of the observed mass obtained from the primary amino acid sequence and the predicted mass obtained from the analysis of the mass spectrometry data…….……..……...……...42

Table 3.5 Solubility of NleD at different temperatures and in the presence of different additives …….…..……...57

Table 3.7 Melting temperatures for wild-type NleD and the variants…….....61

Table 3.8 Change in fluorescence intensity and λ_{max} for wild type NleD and the varaints on the addition of zinc…….…….68

List of Abbreviations

Ala(A)	Alanine
Arg(R)	Arginine
BSA	Bovine serum albumin
CD	Circular dichroism
DLS	Dynamic light scattering
Da	Dalton
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediamine tetra-acetic acid
ESI-Q-ToF	Electron spray ionization Quadrupole Time-of-Flight
Glu(E)	Glutamate
GST	Glutathione S-transferase
His(H)	Histidine
HSQC	Heteronuclear single quantum coherence
IPTG	Isopropyl-β-D-thiogalactopyranoside
LB	Luria broth
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
MW	Molecular weight
O/N	Overnight
PCR	Polymerase chain reaction
Phe(F)	Phenylalanine
PMSF	Phenylmethylsulfonyl fluoride
ppm	Parts per million
RT	Room temperature
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TB	Terrific broth
Trp(W)	Tryptophan
Tyr(Y)	Tyrosine
UV	Ultra violet
WT	Wild-type
λ_{max}	Lambda maximum

Chapter 1: Introduction

The human gastrointestinal tract comprises a range of facultative anaerobes that coexist with human host in a good health [1]. The majority of this micro flora is comprised of *Escherichia coli* [1,2]. Their niche is the mucous layer present in the mammalian colon [1]. Most of the commensal *E. coli* strains are harmless and rarely cause any disease except in immuno-compromised individuals [1,2]. *E. coli* strains, however, can also be virulent in nature [3]. The reason behind their virulence is the acquisition of various genes encoding virulent protein molecules [5]. The mobile genetic elements responsible for encoding these virulent proteins are transferred into the commensal *E. coli* strains by transposons, bacteriophages, plasmids and pathogenicity islands [3]. This gives rise to disease causing *E. coli* strains in humans. The pathogenic *E. coli* include enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregtive *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC) [3].

1.1 Pathogenic *Escherichia coli*

Pathogenic *E. coli* is responsible for a variety of diseases in humans such as urinary tract infection, neonatal meningitis and intestinal diseases (gastroenteritis) [1]. The main culprit behind the majority of *E. coli* outbreaks in North America, the United Kingdom and Japan is enterohaemorrhagic *E. coli* (EHEC) a pathogenic strain of serotype O157:H7 [1,7]. It was first discovered in 1982 and was held responsible for bloody diarrhea (haemorrhagic colitis), non-bloody diarrhea and hemolytic uremic syndrome [7]. The initial source of infection was identified as undercooked hamburgers, but gradually other sources of infection were found, such as sausages, unpasteurized milk, lettuce, cantaloupe and apple juice [7]. Strain O157:H7 possess a 5.5

Mb chromosomal DNA which constitutes the conserved 4.1 Mb backbone sequence and the remaining 1.4 Mb specific to O157:H7 strain [8]. The majority of the 1.4 Mb DNA sequence is horizontally transferred foreign DNA and contains genes encoding effector proteins. In addition to extra chromosomal DNA, O157:H7 also contains a 92 kb pO157 plasmid which houses 100 open reading frames (ORF) encoding pathogenic proteins [8]. *E. coli* O157:H7 follows a complicated process of adhering to the intestinal mucosa and producing a major change in the intestinal cells by altering the actin arrangement in the vicinity of the bacterial infection. Therefore, it is also known as the Attaching and Effacing (A/E) pathogen [9,10].

1.2 Bacterial effectors

Pathogenic *E. coli* produces wide ranges of virulence factors known as effectors. Effectors are a special class of virulent proteins, which are directly injected into the host cell cytoplasm using a needle-like proteinaceous apparatus called Type III (T3SS) [4,5]. They are involved in subverting many different host processes; some have anti-phagocytic activity, while others act on microvilli and inhibit membrane transporter proteins [5]. They are also involved in disrupting tight junctions between cells and in modulating the host's cell signaling pathways, which are involved in inflammation, phosphorylation, ubiquitination and also the MAPK pathways [5].

1.3 NleD: a conserved effector protein in pathogenic *E. coli* strains

Human cells detect the bacterial invasion by the presence of pathogen associated molecular patterns (PAMPS) *e.g.* LPS, bacterial DNA, etc. [9,10]. In response to the invasion, human cells trigger signaling pathways such as the Toll Like receptor (TLR) pathway [9]. These signaling events eventually lead to an inflammatory response via activation of NF–κB and AP-1

transcription factors [9]. Similar signaling cascades are also triggered by EPEC via flagellin dependent pathway. In order to inhibit the inflammatory response the bacterial pathogen injects several effectors into the host cell, which blocks these pathways [9,5]. NleD, one of the many effector proteins encoded by the pathogenic *E. coli* strains, is a zinc-dependent metalloprotease [9,11]. Most of the information available for NleD is focused on gene expression and regulation with functional and *in vivo* studies [4,9,11]. NleD is known to inactivate JNK and p38α kinase by cleaving these proteins in their highly conserved activation loops [9]. The highly conserved TXY sequence (Fig. 1.4.1.) found in the activation loop is the cleavage site of NleD (Fig 1.4.2). This inhibits JNK-dependent apoptosis, which is responsible for the activation of AP-1 transcription factor involved in apoptosis [9]. NleD also cooperates with NleB, NleC and NleE to inhibit interlukin-8 (IL-8) production, which is indirectly related to JNK inhibition [9] (Fig. 1.3.1). NleD is a conserved effector protein among A/E pathogens, for example EPEC, EHEC, *Citrobacter rodentium* (Fig. 1.3.2). The presence of a conserved HEXXH zinc-binding motif classifies NleD under the zincin class of zinc-dependent metalloproteases (Fig. 1.3.2). The NleD protein investigated in the present study is specifically from strain EHEC O157:H7. Although NleD protein is purified *in vitro*[9] there have been no reported biophysical studies on this protein.

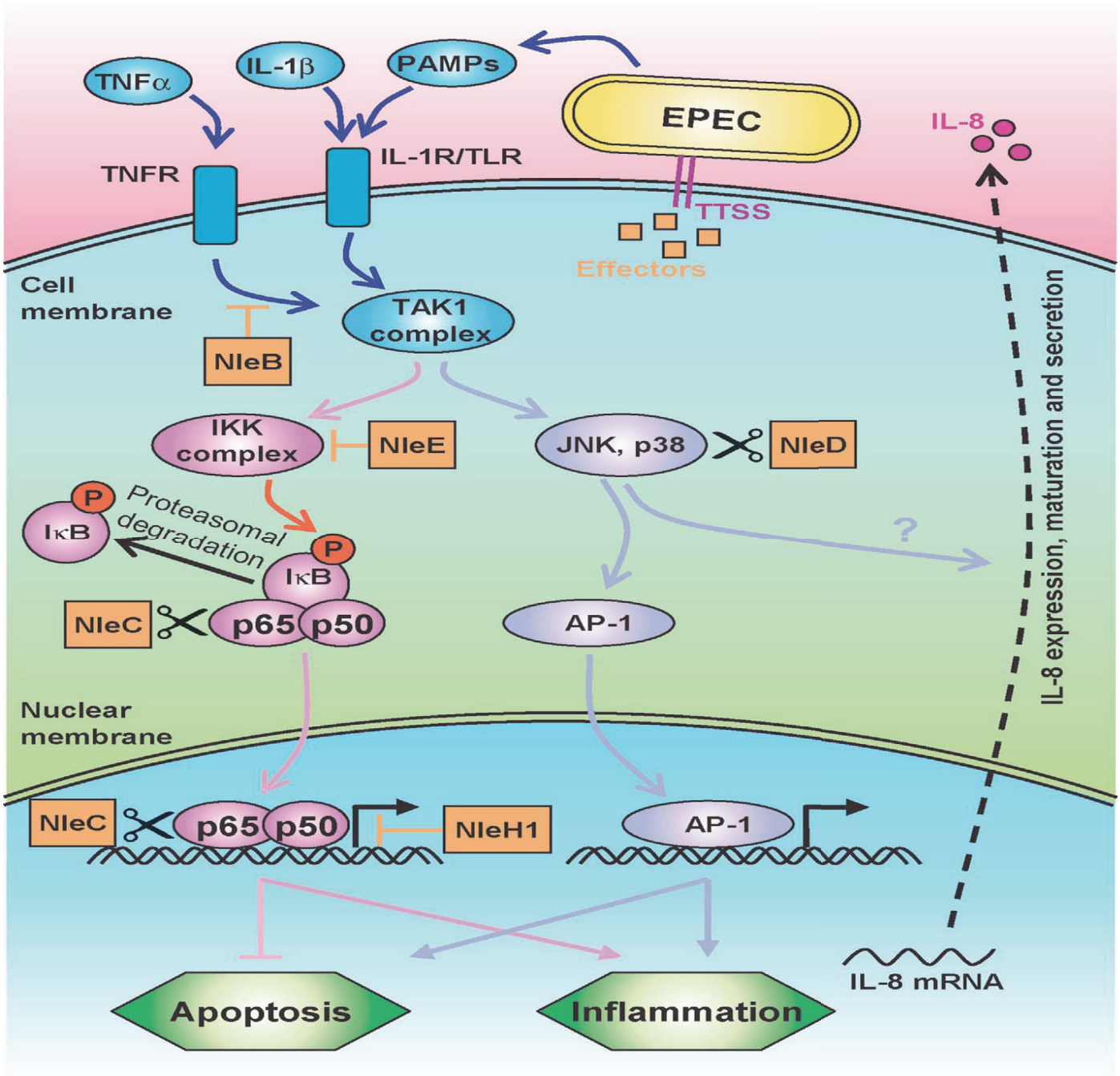


Fig. 1.3.1 A scheme illustrating the function of NleD and its anti-inflammatory activity. NleD cleaves the kinases JNK and p38α kinase inhibiting JNK's proapoptotic activity and indirectly inhibiting the activation of AP-1 transcription factor [9]

```
Xanthomonas_campestris_gi_2896      SNPEESFVVLAHELIHAQHLLAGTSKACKGG---DRYDETSEAGKEELRA 177
Pseudomonas_syringae_gi_660451      MDNDDKITTLAHELVHVRHVLSGSSLADDG----DRYNPRTGSGKEELRA 183
blood_disease_bacterium_gi_344      PSDDDKISQLAHELIHAKHMVAGTWKG---RWG-DDRDPTTPAGKEELRA 187
Ralstonia_solanacearum_gi_2077      PSETDKISQLAHELVHAKHMVAGTWKGGEGEKG-DERDPTTPAGKEELRA 190
Acidovorax_avenae_gi_326318195      ANPANAFITLAHELVHAKHHLAGTMMYGGGPVTREASSSRTDAGREELRA 177
Citrobacter_rodentium_NleD_gi_      SLVDFHACIVFHELLHVFHNLNGERLKVESSQP--ELQTHSPLLLEEART 181
Salmonella_enterica_NleD_gi_16      SLVDFHACIVFHELLHVLHNLNGERLKVESSQP--ESQTHYPFLLEEART 146
E.coli._NleD_gi_320637633           SVVDFHATIVFHELLHVFHNLNGERLKVESSRA--ESQKYSPLLLEEART 178
Candidatus_Hamiltonella_gi_238      EQTELDACIVFHELLHVFHNLNGERLKVAVFQP--EVETYSPFLLEEART 179
Sagittula_stellata_gi_12672927      WHNRDPAVALGHELIHSLHDANGTNDGRNPLPYEDLNGDNRTAPGYEHQA 170
Vibrio_gi_149189313                 GEKAEPYVVLAHELIHSLHCLQGIKMDGR---------------DEELWT 185
Acidiphilium_gi_338980025           -TWLIPHITLGHELIHAWHYLNGAAERDD--------------RSEEHMT 124
                                          : ***:*  *   *                      *  :

Xanthomonas_campestris_gi_2896      VGVGKYEYRKTGQPSENSIRQEHGLPIRKKYKPHGR-------------- 213
Pseudomonas_syringae_gi_660451      VGLDKYSYSLTKEPSENSIRAEHGLPLRMKYRPHQ--------------- 218
blood_disease_bacterium_gi_344      VGLGKYA--ETGEPSENAIRDEHGLPLRRKYY------------------ 217
Ralstonia_solanacearum_gi_2077      IGLGKYAYSETGEPSENSIREENGLALRRKY-------------------- 221
Acidovorax_avenae_gi_326318195      VGLGEYAS--TGEPSENSIRAEHGLPQRTSYSRSGNW------------- 212
Citrobacter_rodentium_NleD_gi_      VGLGAFS---EEVLSENKFREEIGMPRRTFYPHDSSLIHDDNTVTQGFQR 228
Salmonella_enterica_NleD_gi_16      VGLGSFS---EEVLSENKFREEIGVPRRTFYPRDPYLIHDDNTVTQGLQR 193
E.coli._NleD_gi_320637633           VGLGAFS---EEVLSENKFREEIGMPRRTSYPHDSALIHDDNTVSLGFQQ 225
Candidatus_Hamiltonella_gi_238      VGLGSFS---EEVFSENKFREEIGIPRRVSYAHESSLIHDDNTFTMFFEN 226
Sagittula_stellata_gi_12672927      VGLGDYE---DDPFTENKIREEMGEEKRPFY-------------------- 198
Vibrio_gi_149189313                 TGLGKYE---NNPMSENVFRAQFGLAKRQRYF------------------ 214
Acidiphilium_gi_338980025           VGIKGFG---DMPFTENKLRAEAGLPVRTKYFADD--------------- 156
                                     *:  :        :** :* : *   *  *
```

Fig. 1.3.2 Multiple sequence alignment of NleD from different pathogenic *E. coli* strains and other closely related effector proteins found in different bacteria. The alignments were performed using clustal W (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Identical residues are denoted as *, while semi-conserved residues are denoted as **:**. The highly conserved Zn binding motif (HEXXH) found in all the homologs is boxed.

1.4 C-Jun-N terminal kinase (JNK) and p38α protein kinase

MAP kinase pathway is involved in a plethora of signaling events in cellular machinery [16]. It consists of a chain of various protein molecules, which act as intermediates in the relay of information from the surface receptor to the DNA [16]. C-jun N terminal kinases (JNK) and p38α kinase are serine/threonine kinases, which belong to the MAP kinase family [13,14,15]. A variety of different types of cellular stresses and external stimuli lead to activation of JNK and p38α kinase [13,15]. The Epidermal Growth Factor (EGF), Nerve Growth Factor (NGF), Tumor Necrosis Factor-a (TNF-α) and Interleukin-1 are strong activators of these kinases [13]. Activation of both the kinases occurs by dual phosphorylation of threonine and tyrosine residues present in the TXY motif of the activation loop, which leads to phosphorylation of protein targets downstream of the MAP kinase-signaling pathway. c-Jun and ATF-2 are the primary targets of JNK and p38α kinase [14,15]. Activation of these factors leads to heterodimerization among themselves and with other transcription factors *i.e.* CREB, Fos and Fra, giving rise to Activator protein-1 (AP-1) transcription factor. AP-1 is a DNA binding protein and binds to a cis-acting element known as TPA response element [12]. AP-1 is a primary mediator of processes such as apoptosis, inflammation, neurodegeneration, and cytokine production and is dependent on JNK for its activation [12]. Hence, by activation and inhibition of intermediate protein molecules and transcription factors in the MAP kinase pathway, JNK and p38α kinase regulates a variety of processes in mammalian cells.

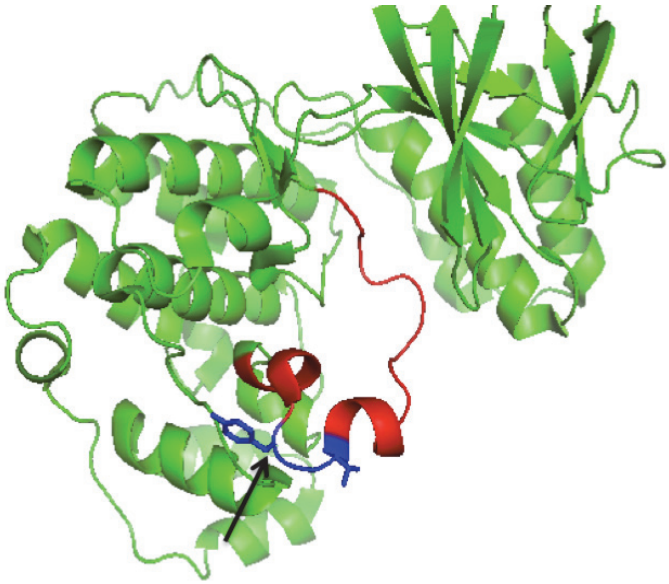
```
p38Alpha     HENVIGLLDVFTPARSLEEFNDVYLVTHLMGADLNNIVKCQKLTDDHVQFLIYQILRGLK 138
p38Beta      HENVIGLLDVFTPATSIEDFSEVYLVTTLMGADLNNIVKCQALSDEHVQFLVYQLLRGLK 139
p38Gamma     HENVIGLLDVFTPDETLDDFTDFYLVMPFMGTDLGKLMKHEKLGEDRIQFLVYQMLKGLR 142
p38Delta     HENVIGLLDVFTPASSLRNFYDFYLVMPFMQTDLQKIMG-MEFSEEKIQYLVYQMLKGLK 139
JNK1         HKNIIGLLNVFTPQKSLEEFQDVYIVMELMDANLCQVIQ-MELDHERMSYLLYQMLCGIK 140
JNK3         HKNIISLLNVFTPQKTLEEFQDVYLVMELMDANLCQVIQ-MELDHERMSYLLYQMLCGIK 178
JNK2         HKNIISLLNVFTPQKTLEEFQDVYLVMELMDANLCQVIH-MELDHERMSYLLYQMLCGIK 140
             *:*:*.**:****  :: :* :.*:*  :* ::* :::    : .:::.:*:**:* *::

p38Alpha     YIHSADIIHRDLKPSNLAVNEDCELKILDFGLARHTDD--EMTGYVATRWYRAPEIMLNW 196
p38Beta      YIHSAGIIHRDLKPSNVAVNEDCELRILDFGLARQADE--EMTGYVATRWYRAPEIMLNW 197
p38Gamma     YIHAAGIIHRDLKPGNLAVNEDCELKILDFGLARQADS--EMTGYVVTRWYRAPEVILNW 200
p38Delta     YIHSAGVVHRDLKPGNLAVNEDCELKILDFGLARHADA--EMTGYVVTRWYRAPEVILSW 197
JNK1         HLHSAGIIHRDLKPSNIVVKSDCTLKILDFGLARTAGTSFMMTPYVVTRYYRAPEVILG- 199
JNK3         HLHSAGIIHRDLKPSNIVVKSDCTLKILDFGLARTAGTSFMMTPYVVTRYYRAPEVILG- 237
JNK2         HLHSAGIIHRDLKPSNIVVKSDCTLKILDFGLARTACTNFMMTPYVVTRYYRAPEVILG- 199
             ::*:*.::******.*:.*:.** *:******* :     ** **.**:*****::*.

p38Alpha     MHYNQTVDIWSVGCIMAELLTGRTLFPGTDHIDQLKLILRLVGTPGAELLKKISSESARN 256
p38Beta      MHYNQTVDIWSVGCIMAELLQGKALFPGSDYIDQLKRIMEVVGTPSPEVLAKISSEHART 257
p38Gamma     MRYTQTVDIWSVGCIMAEMITGKTLFKGSDHLDQLKEIMKVTGTPPAEFVQRLQSDEAKN 260
p38Delta     MHYNQTVDIWSVGCIMAEMLTGKTLFKGKDYLDQLTQILKVTGVPGTEFVQKLNDKAAKS 257
JNK1         MGYKENVDLWSVGCIMGEMVCHKILFPGRDYIDQWNKVIEQLGTPCPEFMKKLQP-TVRT 258
JNK3         MGYKENVDIWSVGCIMGEMVRHKILFPGRDYIDQWNKVIEQLGTPCPEFMKKLQP-TVRN 296
JNK2         MGYKENVDIWSVGCIMGELVKGCVIFQGTDHIDQWNKVIEQLGTPSAEFMKKLQP-TVRN 258
             * *.:.**:*******.*::    :* * *::** . ::.  *.* .*.: ::.   .:.
```

Fig. 1.4.1 Multiple sequence alignment of JNK isoforms and p38α isoforms. The alignments are perfomed using Clustal W (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The conserved 20-22 amino acid loop is boxed in black and the highly conserved TXY site of cleavage by NleD is boxed in red. Identical residues are denoted as *, while semi-conserved and similar residues are denoted as **:** and **.,** respectively.

A) p38α kinase (PDB: 1R39)



B) JNK2 (PDB: 3E70)

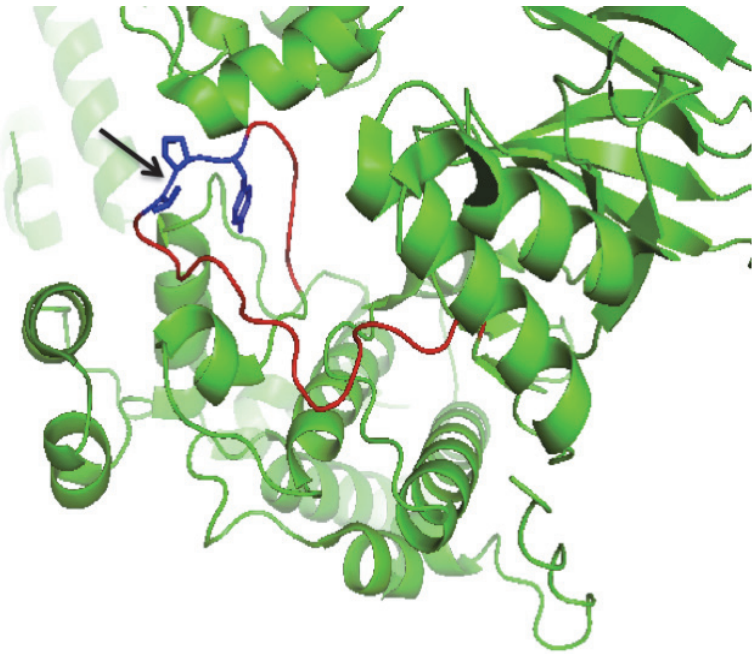


Fig. 1.4.2 Structure of p38α kinase and JNK showing the activation loop and the cleavage site for NleD. The activation loop is shown in red and the cleavage site in the activation loop is marked with an arrow. The cleavage site (TXY) is shown in blue in both the kinases. The figures were generated using PyMol (http://www.pymol.org/).

1.5 Zinc-dependent metalloproteases

Zinc metalloprotease is a large super family of enzymes with a variety of functions; for example they can participate in embryonic development, can act as virulent toxins, and can cause arthritis and sometimes cancer. These metalloproteases are usually divided into various subclasses based on the unique signature sequence around the zinc-binding site [17]. Fig. 1.5.1 explains briefly the different families of zinc metalloproteases and their features, which distinguish them from the other metalloprotease in the same class.

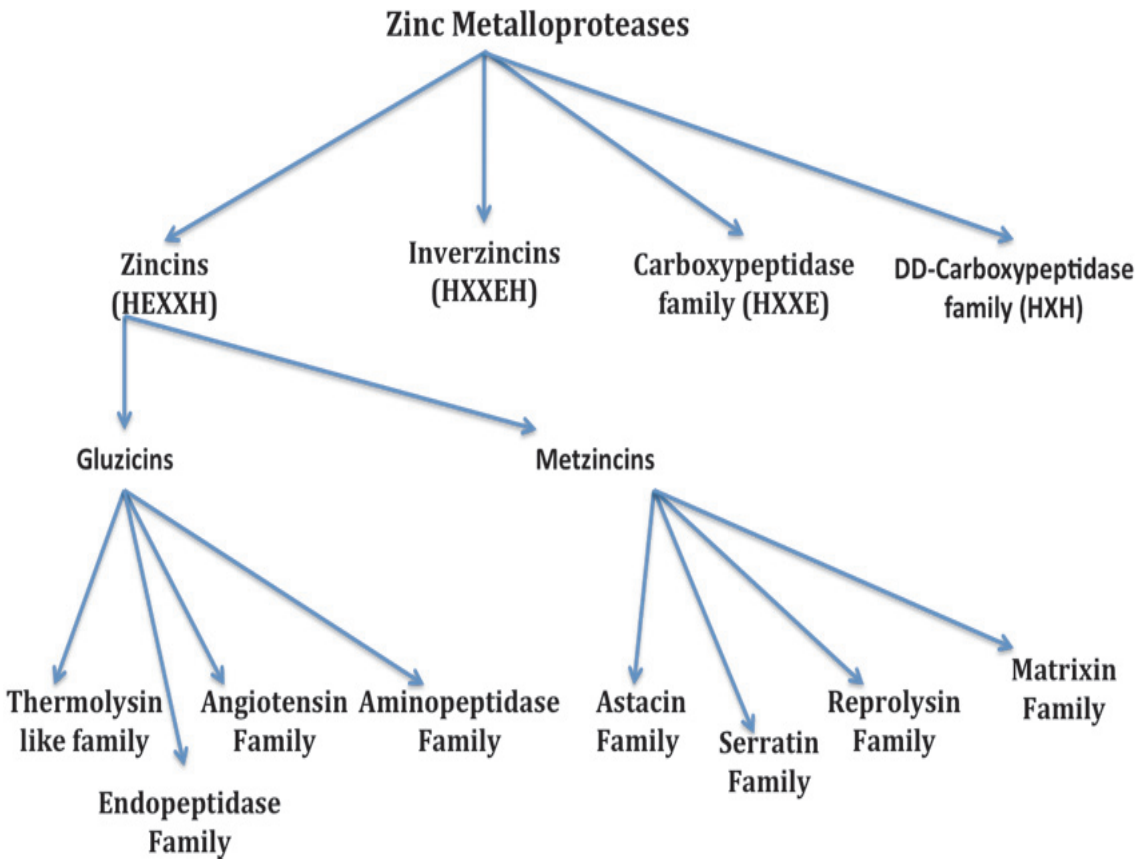


Fig. 1.5 Families of zinc metalloproteases and their interrelationship based on the sequence surrounding the zinc binding residues. In addition to the HEXXH motif, gluzincins posses a glutamic acid while metzincins possess an additional histidine and methionine residue to satisfy the zinc valency [17].

Zincins have a conserved HEXXH motif, similar to that found in NleD [18]. Zincins is further divided into two broad categories, gluzincins and metzincins, based on the type of the additional residue(s) needed to satisfy the valency of zinc and to stabilize the binding between the metal ion and the protease [17]. Gluzincins require a glutamate while metzincins need a histidine and a methionine in their active site. NleD most likely belongs to one of the subfamilies characterized under zincins, however, this has yet to be determined experimentally. Additionally, it is not known which residues other than the conserved HEXXH motif might be involved in stabilizing the metal ion in the active site of NleD.

1.6 Zinc ion binding site in zincins

Most of the zinc metalloproteases posses a three coordinated Zn^{2+} binding site with tetrahedral or distorted tetrahedral geometry [19]. In such cases usually an additional bound water molecule to the metal ion satisfies the valency of the zinc ion [22]. The metal-bound water acts as a nucleophilic hydroxyl group at neutral pH in order to attack the substrate for proteolysis [22]. The thermolysin family of enzymes is one example of such a protease. In this case, the third zinc coordinating residue is a glutamic acid (E) usually located C-terminal to the HEXXH motif (Fig. 1.8.1) [20]. Accordingly, they are also known as gluzincins [19].

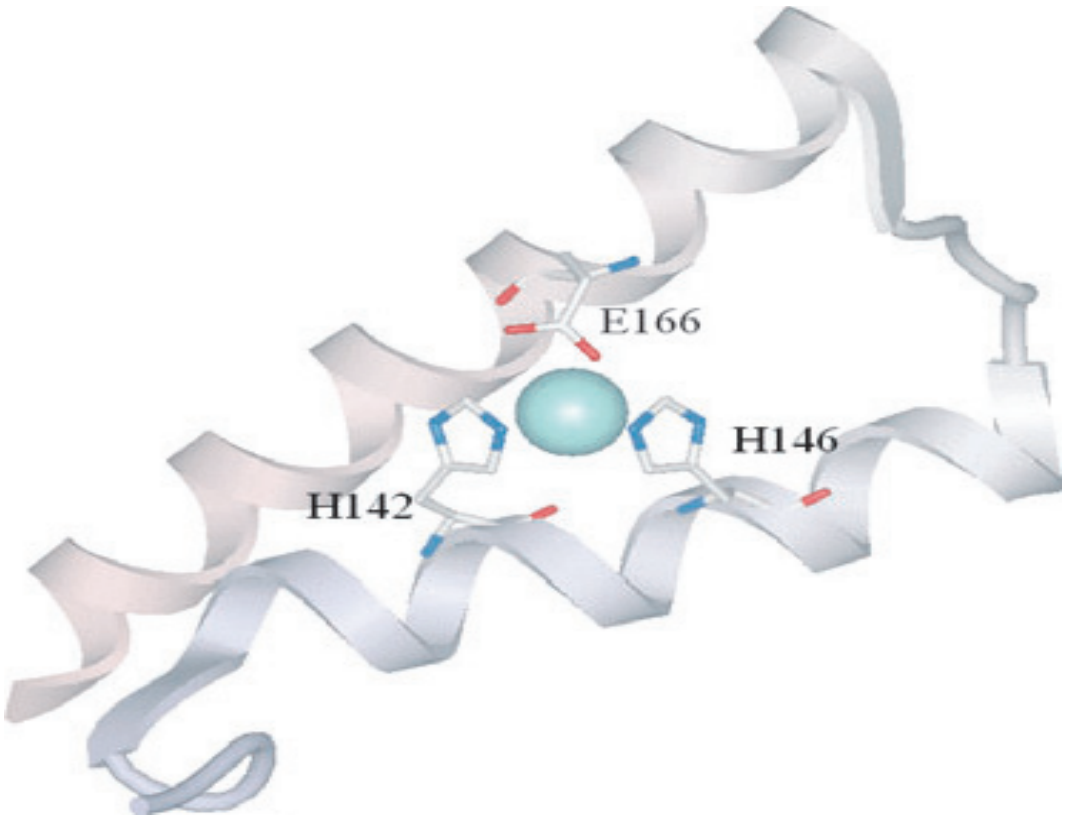


Fig. 1.6.1 The active site of a thermolysin-like protease. Its structure is derived from PDB-1KEI.[19,23]

In metzincins the third zinc binding ligand is a histidine or an aspartate located in the extended motif HEXXHXXGXX(H/D) [19,21]. The region C terminal to the third histidine residue forms a unique loop that houses the conserved methionine and accordingly, is designated as the Met turn [21] (Fig. 1.6.2). The three different histidines are involved in binding to the zinc ion.[21]

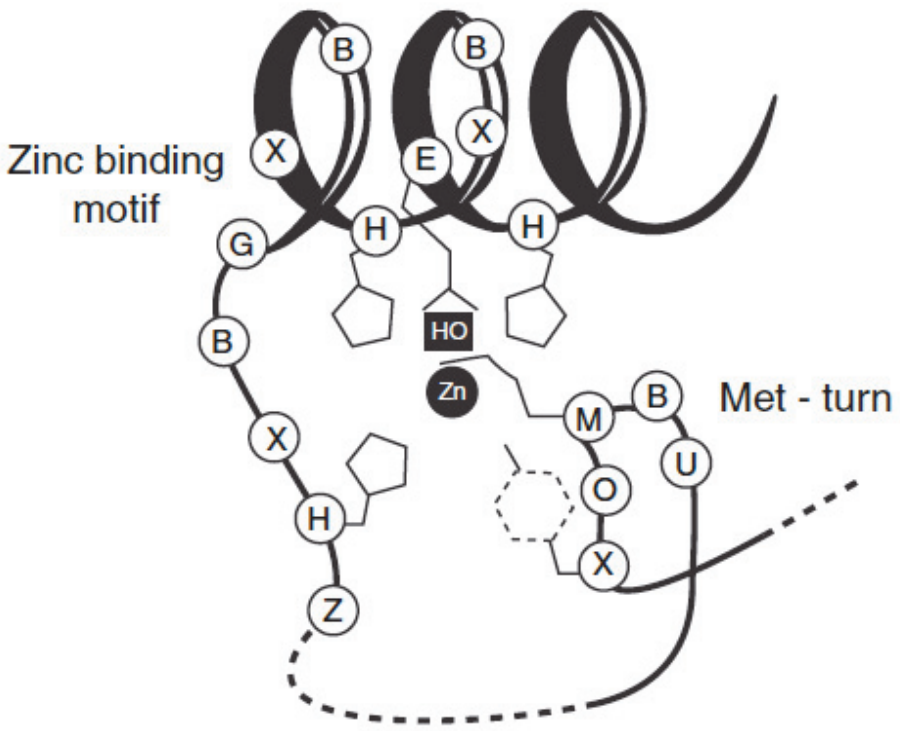


Fig. 1.6.2 Active site of metzincin-like protease showing the Met turn and the amino acid environment around the zinc ion. Highlighted in this diagram is the critical methionine residue, the three zinc ligating histidines groups and the sequence motifs common for the metzincins. Z is a family-specific conserved residue, which could be glutamic acid, proline or aspartic acid. U and O are also conserved family-specific residues and B is a bulky hydrophobic residue anchored in the hydrophobic core of the protein.[19,21]

1.7 Active site of botulinum neurotoxin and thermolysin

Based on our bioinformatic analysis a structural similarity was predicted between NleD and botulinum neurotoxin, a thermolysin like protease (Section 3.6). A brief overview is provided in this section comparing the hydrolytic mechanisms of thermolysin and botulinum neurotoxin A light chain (BoNT/A). Both the proteases fall under gluzincin family of metalloproteases and require zinc for their catalytic activity. In thermolysin, the catalytic glutamate present in the zinc ion-binding motif (HEXXH) Glu143 (Glu223 in BoNT/A), is used to hold the polarized water molecule in place, which together with the zinc ion is involved in the nucleophilic attack of the carbonyl carbon of the scissile peptide bond of the substrate and leading to the formation of

oxyanion intermediate (Fig. 1.7.1) [61,78]. The attacking water molecule donates a proton, which travels via the carboxyl group of Glu143 (Glu223 in BoNT/A), which in turn, may help stabilize the tetrahedral intermediate forming a salt bridge with the positively charged amide nitrogen (Fig. 1.7.1) [61,78]. The negative charge developing on the carbonyl oxygen in the transition state is stabilized by interaction between the protonated side chain of His 231[59](Arg362 in BoNT/A) and the hydroxyl group of Tyr157 (Tyr365 in BoNT/A). Asp226 (Glu350 in BoNta) is speculated to properly orient His 231 in the transition state (Fig. 1.7.1) [61,78]. The protonated amide nitrogen of the scissile bond in the substrate causes the disruption of the C-N bond and is proposed to receive a proton from the attacking water, which is possibly again mediated by Glu143 (Glu223 in BoNT/A) (Fig. 1.7.1) [61,78].

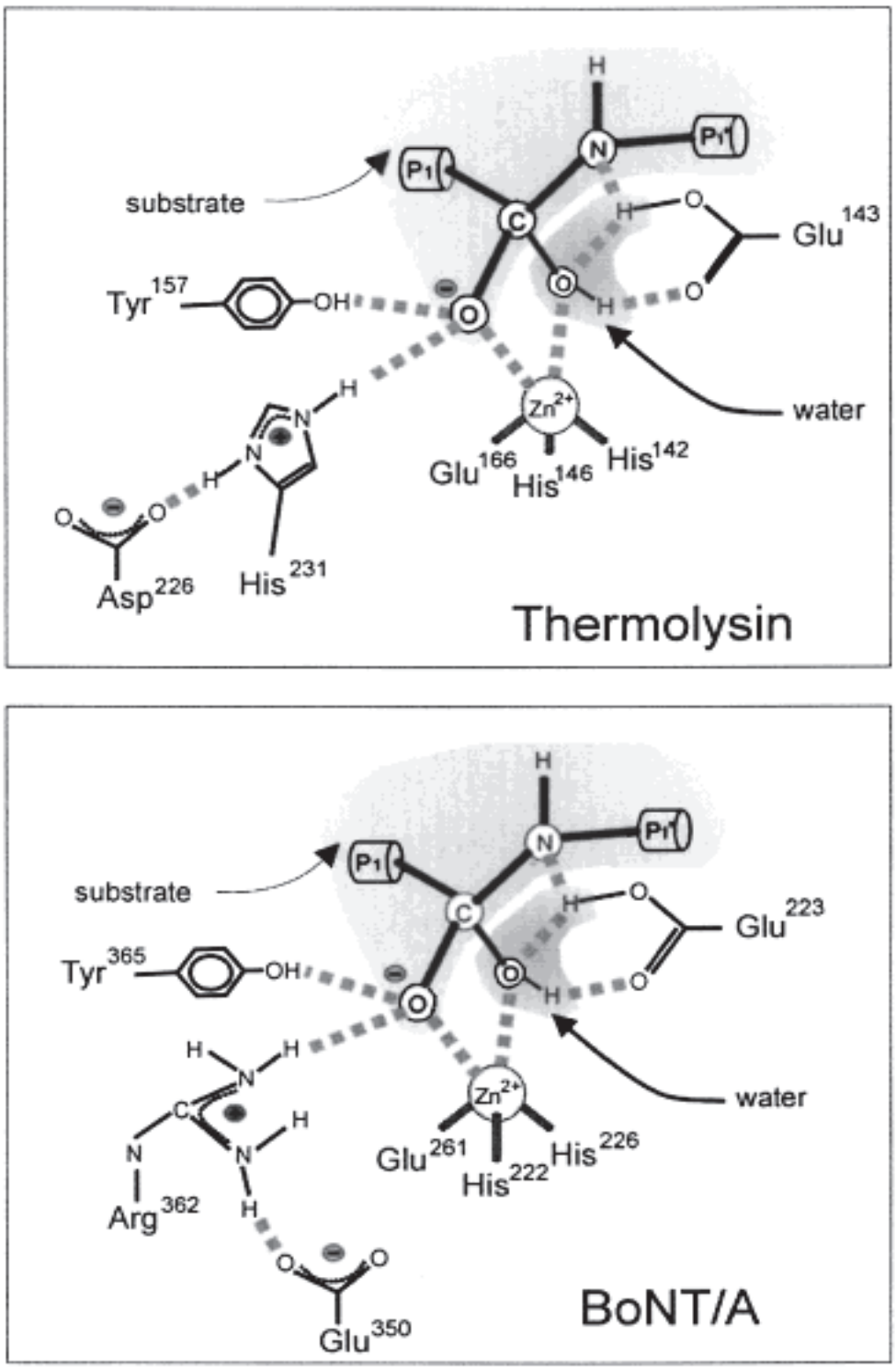


Fig 1.7 Hydrolytic mechanisms of thermolysin and its comparison to the thermolysin-like protease botulinum neurotoxin A light chain [61].

Biophysical methods used for protein characterization

1.8 Characterization using mass spectrometry

Mass spectrometry is a technique, which measures the relative mass to charge ratio (m/z) of ionized molecules [26]. Hence, ionization of molecules is a critical step in mass spectrometry [24]. The high energy applied in the ionization source causes fragmentation of molecular ions prior to detection [26]. Ionization of large molecules such as proteins and carbohydrates is a requirement for their detection [25]. These molecules are usually prone to fragmentation on ionization and hence need to be gently ionized. Electrospray ionization (ESI) is routinely used for the ionization of large molecules. A high electric potential of 2.5 – 5 kV is applied to the sample solution infused in through a capillary [27]. The electric potential causes a net positive charge to be formed in the sample droplet. The droplet is sprayed and evaporated (desolvated) by the action of a nebulizing gas such as nitrogen and is often assisted by the application of heat [26]. The positive ions on the sprayed droplets experience coulombic repulsion, as they are forced closer together. Eventually, this repulsion exceeds the surface tension of the droplets and the positive ions are pulled away from the droplet-giving rise to the Taylor cone [26].

A critical step in mass spectrometry is sample preparation, which includes the use of volatile buffers and acids. These solvents are easily removed during the desolvation process [26]. Protonation of the sample is necessary by the addition of a positive molecular ion such as a proton. Multiple basic residues such as arginines and lysines present in a protein sample or a polypeptide can be protonated, giving rise to multiple peaks in the mass spectrum [26].

1.9 Spectroscopic characterization of proteins

The primary amino acid sequence of a protein can adopt distinct secondary and tertiary structures. Secondary structural features such as α-helices and β-sheets describe the orientations of the polypeptide backbone, while tertiary structure refers to the 3D arrangement of the amino acid residues within a protein. Both secondary and tertiary structures define the protein's pattern of folding and its native structure.

NMR and X-ray crystallography are used to determine the high-resolution structure of a protein. Characterization of the global structure of the protein in solution, however, can be performed using simpler and lower resolution spectroscopic techniques such as Fourier transformed infrared (FTIR), UV-VIS absorption, and circular dichroism (CD); two of these techniques will be discussed further below. These techniques can also be useful to monitor perturbations in global secondary and tertiary protein structure upon changes in temperature and pH, and in the presence of organic solvents or different chemical denaturants such as urea and guanidine hydrochloride.

1.9.1 Circular dichroism (CD) spectroscopy

CD spectroscopy is an easy and rapid solution technique to determine the secondary structure and folding characteristics of a recombinantly expressed, purified protein [28,29]. It is widely used to determine if the purified protein is well folded or if alterations of a protein's amino acid sequence affect the conformational stability of the protein [28,29].

CD measures the difference in absorption of left-handed and right-handed circularly polarized light in a sample in solution, producing elliptically polarized light. The 3D conformation of the amide backbone, which relates to the protein's secondary structure, is well suited to CD analysis since it detects optically active protein molecules [31,32]. Aromatic amino

acids such as tryptophan, phenylalanine and tyrosine residues, and disulfide bridges also exhibit CD absorption in the near UV wavelength range (250 nm – 350 nm) [33].

Peptide bonds found abundantly in protein molecules acts as chromophores for far-UV CD absorption (180 nm – 250 nm). The magnitude and direction of electronic transitions is controlled by the bond angle and the rotation around the individual peptide bonds [30]. The CD signal from the protein backbone results from the two types of electronic transitions $\pi \rightarrow \pi^*$ (from the 3π orbital of carbon, nitrogen, oxygen) which lies near 190 nm, and $\eta \rightarrow \pi^*$ (from a lone pair of oxygen) the lower energy transition at 220 nm [30]. The CD spectra assignments for proteins are based on the observed spectra of homo-polypeptides, which can form a α-helix, a β-strand or random structure depending on the solution pH [33]. The CD spectrum showing a large negative band at 222 nm ($\eta \rightarrow \pi^*$ transition), a smaller negative band at 208 nm and a large positive band at 190 nm are very characteristic of a right-handed α-helix. The double minimum obtained at 208 nm and 222 nm is particularly striking and can be used to estimate the percent helix composition of the protein molecule [32]. Anti-parallel β- sheets show a small negative band at 215 nm ($\eta \rightarrow \pi^*$ transition) and a larger positive band near 198 nm ($\pi \rightarrow \pi^*$ transition) [33]. Unordered structures are characterized by a large negative absorption at 200 nm.

Therefore, following changes in ellipticity at specific wavelengths can be used to monitor protein unfolding and to estimate the protein's secondary structure composition.

1.9.2 The importance of intrinsic tryptophan fluorescence emission

The mechanics of fluorescence emission can be divided into three main stages:

Excitation: An external source such as an incandescent lamp or a laser supplies a photon of energy, which creates an excited singlet state [37]. **Excited state:** The excited fluorophore has a finite lifetime, in which it undergoes multiple conformational changes and interactions with the

molecular environment. Gradually the excited fluorophore dissipates energy and reverts to a relaxed singlet state, which is responsible for fluorescence emission [37]. **Fluorescence Emission:** Finally, the photon energy is emitted on returning the fluorophore from a relaxed singlet state to the ground state, because of the energy dissipation from the excited state. The difference in the excitation energy and emission energy is known as the Stokes shift [37].

The fluorescence emission intensity of tryptophan depends on the nature of the environment surrounding the amino acid sequence. For example, in a hydrophobic environment the fluorescence emission intensity of tryptophan can be elevated whereas the presence of a nearby protonated residue (protonated histidine group, cluster of glutamatic acid residues) can quench tryptophan emission. The maximum emission wavelength of tryptophan is also strongly dependent on its environment due to the presence of strong Strokes shift, which correlates with the polarity of the solvent [34].

Monitoring the change in tryptophan fluorescence emission can provide valuable information about structural changes that occur with such processes as ligand binding or unfolding of native protein structure [36,35]. Based on its position with respect to the active site, tryptophan emission can also be used to quantify the strength of ligand binding by determination of the dissociation constant (K_d.) for the ligand.

1.10 Characterization of the oligomeric state of proteins

A significant fraction of cellular machinery is comprised of oligomeric proteins [38]. Protein oligomerization can be necessary for the functional control, allosteric regulation and also for higher order complexity. Oligomerization can involve the association of identical monomers (homo-oligomeric) or a mixture of distinct monomers (hetero-oligomeric) depending upon

function of the protein molecule. Homo-oligomeric proteins are more commonly found in a cell than hetero-oligomeric proteins [38].

Different types of oligomers may have different strengths of association between their subunits. Proteins, which retain their oligomeric state, may have dissociation constants in nanomolar range [38,39]. Others may have an environmentally dependent oligomerization, which is usually dependent on the concentration of protein or salt, pH, and temperature. Such proteins have a higher dissociation constant usually in micromolar or the millimolar range [38,39].

Oligomerization can also occur in response to a stimulus, for example upon binding to a nucleotide or upon phosphorylation, which may lead to a change in the affinity of the subunits by orders of magnitude. Hydrophobic interactions also play an important role in protein oligomerization [39]. Usually amino acid residues present on the oligomeric interfaces are hydrophobic in nature [38,39]. Hydrogen bonds and salt bridges are also important in stabilization of the oligomeric interfaces [38].

1.10.1 Size exclusion chromatography (SEC)

SEC is a convenient technique to determine the average molecular weight of a protein molecule [42]. The basic principle behind separation is based on the hydrodynamic volume or size [41,42]. The protein sample is dissolved in an appropriate solvent and is injected in to a column packed with porous particles with a fairly defined pore size. The mobile phase used is the same as the solvent used to dissolve the sample [40]. Assuming the protein sample injected in the column is heterogeneous, protein molecules that are too large to penetrate the pores in the porous beads elute in the void or interstitial volume of the column. As the molecular size of the protein molecules present in the sample continues to decrease in comparison to the average pores size of the beads, protein molecules penetrate the pores [41]. They access a greater pore volume and finally

elute at later times. The smallest particles, having complete access to the total pore volume, elute at the total pore elution volume of the column. Thus, the higher molecular weight protein molecules elute first followed by lower molecular weight protein molecules [42]. SEC can only be used to determine the relative molecular weight of the protein and requires calibration by standard protein samples of known molecular weight.

1.10.2 Dynamic light scattering (DLS)

DLS exploits the property of the monochromatic or coherent beam of light incident on a solution of macromolecules (such as proteins, colloidal particles) being scattered by the solute particles in all directions [43]. The refractive index of the solvent used is distinct from that of the solute [43]. The light waves scattered from different macromolecules finally combine at a distant photomultiplier tube and produce a non-uniform scattered intensity. In an ideal case, if the particles are stationary, the scattered light intensity is constant in all directions. In reality however, particles are in a constant Brownian motion leading to fluctuations in the scattered light intensity on the detection plane [43]. This difference in the intensity of fluctuation is measured by DLS. The detected frequency of the scattered light depends on the Doppler effect, since it can be higher or lower depending on the movement of particles away from or towards the detector. Hence, the frequency distribution of the scattered light is slightly broader than the incident light. This light scattering can be used to determine the particle size and the sphere size distribution, and thus describes the motion of the particle in the medium. The diffusion coefficient of the particle in the medium is measured using the autocorrelation function [43].

DLS has multiple advantages, which includes smaller experimental duration, it is relatively easy to use and requires a modest development cost. It measures a variety of parameters of interest, for example the radius of gyration, the translational diffusion constant and

the percent polydispersity [45]. Its most common use is to analyze samples with broad distribution of species with varying molecular masses [46]. As every technique has its own advantage and disadvantage some of the drawbacks of DLS include: 1) it is difficult to distinguish between different oligomeric states of a protein; 2) one cannot distinguish between very small-sized species; 3) it is not used for the analysis of non-rigid molecules [44].

1.10.3 Analytical ultracentrifugation (AUC)

AUC measures the changes in the sedimentation behavior of proteins under an extremely high gravitational field [47,48]. It is based on the principle of redistribution of mass under the applied gravitational field to an extent when the gravitational potential energy is exactly balanced by the chemical potential energy at every radial position. The key forces involved in regulating the sedimentation process are the buoyant force, gravitational force and the translational frictional force [48]. First, expression for gravitational force is $F_{sed}=m\omega^2 r$ and is directly proportional to the square of the rotor speed [48]. As a result, a wide range of particle size can be studied by adjusting the speed of the rotor. Second, Archimedes principle governing the buoyant force and given by $Fb = -mv\mathrm{r}\omega^2 r$, acts against the sedimentation forces and is proportional to the amount of solvent displaced. Hence, the type of solvent used may be an important consideration. Third, the frictional force is dependent on the translational frictional coefficient (f) and the sedimentation velocity -- $F_f = s(kT/D)\omega^2$ [47]. The Svedberg equation is derived from the sum of all the three equations mentioned above giving a final equation: $s/D = M(1-v\mathrm{r})/\mathrm{RT}$ [47,48].

Usually two types of experiments are conducted by AUC: sedimentation velocity and sedimentation equilibrium [50]. Sedimentation velocity experiments measures the time course of sedimentation process under the influence of high centrifugal force. Different types of protein molecules have varying rates of sedimentation. The hydrodynamic properties of similar size

protein molecules have a unique sedimentation boundary, which makes the separation strongly size-dependent. A typical AUC run can be used to characterize protein sample comprising a range of molar mass varied by a 1000-fold. A high rotor speed is usually chosen for sedimentation velocity experiments, to minimize the effect of diffusion and separate based on hydrodynamic properties of the protein [50].

Sedimentation equilibrium measures the molecular mass of the protein and is widely used to study protein – protein interactions [49]. It is usually used to determine the oligomeric state of the protein in its native form, the equilibrium constant (K_d) for reversible and non-reversible protein interactions, and to measure the stoichiometery of binding within the protein complexes [49,50].

1.11 Characterization of histidine residues in proteins by NMR

Histidine is one of the 20 amino acids found within proteins. The functional group within histine's side chain is an imidazole ring. The pKa of the imidazole ring is normally reported to be in the range of 6.0 – 7.0. Histidine can act as a useful probe for one-dimensional (1D) NMR experiments. Two sharp signals in the proton spectrum given by $C^{\delta}H$ and $C^{\varepsilon}H$ present in the histidine side chain, originate from the ring resonance of the imidazole ring [51](Fig. 1.10.). The signal observed is distinctive from the multitude of signals obtained from other protons present in a protein [51]. Protonation of the imidazole ring leads to shifts in the resonance signals, which helps in determining the pKa value of the individual histidines present in a protein [51]. This information can be used to generate useful information about the electrostatic and structural environment of the protein [51,52]. It can also be used to assess hydrogen bonds in various regions of the protein.

Often 1D-^{1}H NMR experiments are conducted in 99.9% D_2O. Under such conditions most of the hydrogen atoms bound to amides are exchanged with deuterium [51]. The pH titration experiment is commenced at either higher or a lower pH values, and over a range at which the

protein is stable [51]. A 1D-^{1}H or a Heteronuclear Single Quantum Correlation (HSQC) spectrum is used to monitor the effect of pH on the histidine peaks observed on the spectrum.

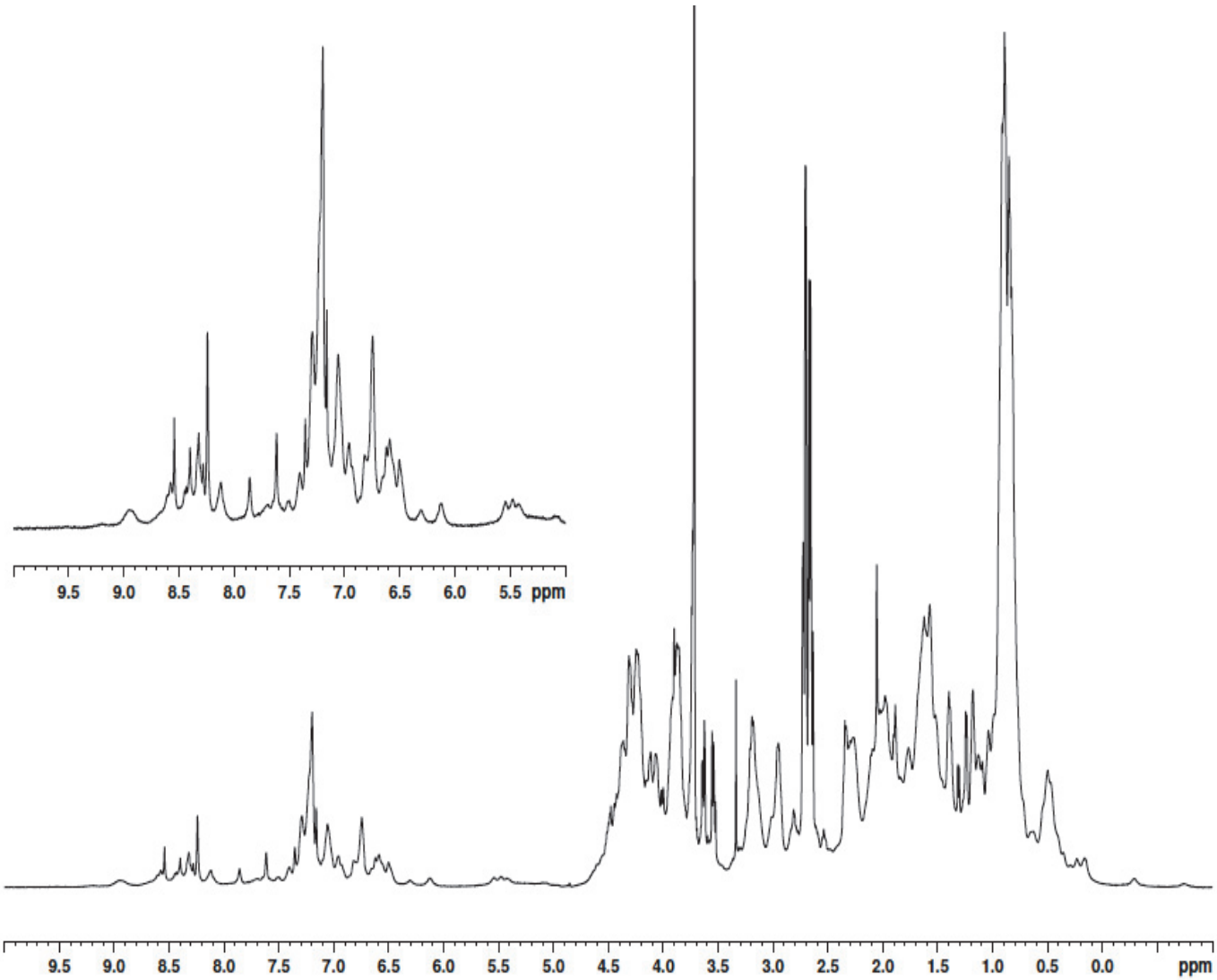


Fig. 1.11 1D-^{1}H NMR spectra of NleD. The experiment was conducted in 99.9% D_2O. The spectrum from 5.5 ppm to 9.5 ppm shows signals originating from the protons present in aromatic side chains, $C^{\delta}H$ of imidazole ring present in the histidines side chain, and residues with no exchanged amides. The region of interest for monitoring protons in the histidine side chains approximately lay from 7.5 ppm to 9.5 ppm. The region from 0.0 ppm to 4.5 ppm show signals originated from protons present in the aliphatic side chains.

1.12 Research outline

As previously mentioned there have been very few studies conducted on NleD. Its 3D structure is not known, nor has NleD been assigned to any particular family of the metalloproteases. Additionally there is no information regarding its biophysical properties.

The overall goal of this project was to determine the oligomeric state of NleD *in vitro*, and to identify different experimental conditions that could enhance the stability of NleD. This information could greatly assist in the determination of a high-resolution 3D structure of the enzyme. We also sought to identify the amino acid residues present in the active site of NleD involved in catalysis, as steps towards assigning this enzyme to a particular family within the metalloprotease super family.

To achieve our goals NleD was expressed and purified using gluthione S-transferase (GST)-based affinity chromatography and verified its characterized by mass spectrometry. The oligomeric state of NleD was predicted using size exclusion chromatography, dynamic light scattering and analytical ultracentrifugation. The stability of NleD was characterized at different temperatures and also in the presence of various protein stabilizers. Six different NleD variants were generated based on structural predictions made from extensive bioinformatic analysis. The thermal stability of NleD variants was characterized using far-UV CD. Additionally, the proteolytic activity of NleD and its variants was characterized using SDS-polyacrylamide gel-based assays using p38α kinase as a substrate.

Our results indicate that NleD is likely a monomer *in vitro* but during purification is prone to proteolytic digestion at its C-terminal region. NleD was found to have a low thermal stability but a significant improvement in its stability was observed at 4°C. Additionally, trehalose was found to be the most promising all the chemical protein additives tested. Bioinformatic analysis suggests that NleD belongs to the thermolysin-like family of metalloproteases. The results from

site-directed mutagenesis studies indicated residues His149, Glu174, Glu175, Arg203 and Tyr206 are important for the catalytic activity of NleD and are likely part of NleD's active site.

Chapter 2: Materials and Methods

2.1 Chemical and reagents

Phenyl-methyl-sulfonyl fluoride (PMSF) was from Sigma and was prepared in methanol at a concentration of 100 mM and stored at -20ºC. Methanol, acetonitrile (ACN) chloroform, trifluoroacetic acid (TFA) (HPLC grade) and C_{18} resin were all purchased from Fisher Scientific. Oligonucleotides of HPLC purity were obtained from Alpha DNA (Montréal, QC). The QuickChange XL site-directed mutagenesis kit was purchased from Stratagene, while the Plasmid Miniprep Kit was purchased from Sigma. Gel filtration molecular weight standards were obtained from Bio-Rad. DNA grade Superdex 75 resin was from Amersham Biosciences, while glutathione sepharose and nickel-nitrilotriacetic acid (Ni-NTA) resins were from Amersham Biosciences.

Dialysis membrane (6-8 kDa cutoff) was from Fisher. High fidelity *Pfu* DNA polymerase and 6x DNA loading dye were purchased from Bio Labs. EDTA-free protease inhibitor cocktail was obtained from Roche Applied Science. Isopropyl β-D-1-thio galactopyranoside (IPTG), dithiolthreitol (DTT) was purchased from Bioshop. Bio Labs supplied the mixture of deoxynucleotide triphosphates (dNTPs). All other reagents used were from Sigma and Bioshop and were of the highest grade available.

2.2 Site-directed mutagenesis

Plasmid DNA containing the NleD insert was obtained from the CIHR-funded Central Cloning Facility at Biotechnology Research Institute (BRI) and was used as a template for site-directed mutagenesis. The DNA construct contained a N-terminal Glutathione S- Transferase (GST) tag and ampicillin resistance as a selectable marker.

Primer	Sequence	T_m(°C)
H149A Forward	5'-CATGAGTTGCTCCATGTTTTCGCCAATTTAAATGGGGAGCGTTT-3'	78.3
H149A Reverse	5'-AAACGCTCCCCATTTAAATTGGCGAAAACATGGAGCAACTCATG-3'	78.3
E174A Forward	5'-CTCTCCACTTTTACTCGCAGAAGCCAGGACTGTTG-3'	79.2
E174A Reverse	5'-CAACAGTCCTGGCTTCTGCGAGTAAAAGTGGAGAG-3'	79.2
E175A Forward	5'-CCACTTTTACTCGAAGCAGCCAGGACTGTTGGG-3'	79.1
E175A Reverse	5'-CCCAACAGTCCTGGCTGCTTCGAGTAAAAGTGG-3'	79.1
E191Q Forward	5'-CTTTTTCAGAGGAGGTGCTTTCACAGAATAAATTCCACGAAGAGATTGG-3'	79.5
E191Q Reverse	5'-CCAATCTCTTCGTGGAATTTATTCTGTGAAAGCACCTCCTCTGAAAAAG-3'	79.5
R203K Forward	5'-TTGGGATGCCCCGTAAAACCTCCTACCCG-3'	78.8
R203K Reverse	5'-CGGGTAGGAGGTTTTACGGGGCATCCCAA-3'	78.8
Y206F Forward	5'-GCCCCGTAGAACCTCCTTCCCGCACGA-3'	80.1
Y206 Reverse	5'-TCGTGCGGGAAGGAGGTTCTACGGGGC-3'	80.1

Table 2.2.1 List of forward and reverse primers designed to incorporate the desired point mutations in wild-type NleD. Changed bases, which encode the desired mutation, are underlined. The online softwaare Primer X was used for designing the primers and estimating the T_m values.

The DNA template was isolated from a 10 mL of *E. coli* DH5α cell culture using Sigma Aldrich Plasmid Miniprep Kit. The concentration of the isolated plasmid was determined spectrophotometrically at 260 nm.

The lyophilized samples of the synthesized oligonucleotides obtained from Alpha DNA were re-suspended in autoclaved Milli-Q water giving a final oligonucleotide concentration of 100 μM (Table 2.2.1.).

Site-directed mutagenesis was conducted using the Quick-change site-directed mutagenesis kit from Strategene. Reaction mixtures contained 5 ng of double-stranded plasmid DNA template, 20 ng of both forward and reverse oligonucleotide primers, 5 μL of 10X reaction buffer supplied with the mutagenesis kit, 5 μL of dNTP mix, 1 μL of the *Pfu* DNA polymerase (2.5 U/μL) and the final reaction volume was made to 50 μL by adding autoclaved ddH_20. The reaction mix was subjected to the PCR conditions shown in Table 2.2.2

No. of Cycles	**Cycle Name**	**Temperature (°C)**	**Time (min)**
1	**Denaturation**	**95**	**5**
	Denaturation	**95**	**0.5**
18	**Annealing**	**55**	**1**
	Extension	**68**	**7**
1	**Final Extension**	**68**	**12**
1	**Storage**	**4**	**Overnight**

Table 2.2.2 PCR reaction conditions for site-directed mutagenesis

After completion of the PCR reaction, the reaction mixture was incubated with 10 U of *Dpn*1 at 37 °C for 1 h to digest methylated and hemi-methylated DNA. The *Dpn*1-treated reaction mixture

(10 μL) was then transformed into competent *E. coli* DH5α cells and plated on LB-agar plate supplemented with 100 μg/mL of ampicillin (Amp). Selected colonies were grown in 5 mL of LB-Amp broth. The amplified plasmid was isolated using a Sigma Aldrich Plasmid Miniprep kit and was sent for sequencing at the Genome Quebec Sequencing platform at McGill University (NANUQ). The sequencing results obtained were aligned with the wild-type DNA sequence of NleD using BLAST, to verify the incorporation of the correct point mutations.

2.3 Protein expression and purification of wild-type and variant forms of NleD

The recombinant NleD and the variants were expressed in *E coli* strain BL-21 Gold. The BL-21 Gold competent cells transformed with recombinant NleD were grown at 37°C in 1 L of Terrific Broth (TB) supplemented with Amp and induced with 500 μM of isopropyl β- D-thiogalactopyranoside (IPTG) after reaching an $O.D._{600}$ between 0.6–0.8. The induction was continued at 20°C for 16 h in an incubator shaker. After 16 h cells were harvested by centrifugation at 8000Xg for 30 min at 4°C and the pellets were stored at -20°C.

Cell pellet was resuspended in a lysis buffer of 50 mM Tris (pH 7.2), 100 mM NaCl, 1 mM DTT, 2mM EDTA, supplemented with 0.2 mM PMSF and protease inhibitor cocktail (prepared according to manufacturer instructions). Cells were lysed by sonication for 3 min in bursts of 10 s and placed on ice between bursts. Insoluble debris were removed from the lysate by centrifugation at 28,384Xg for 1 h at 4°C using a Beckman ultracentrifuge. Affinity chromatography was performed using glutathione sepharose resin. The supernatant obtained from the 1 L of culture was incubated with 5 mL glutathione sepharose resin pre–equilibrated with 50 mM Tris (pH 7.2), 100 mM NaCl, 1 mM DTT and secured to a shaker at 4°C for 1 h in order to ensure effectively binding of protein to the resin. The glutathione sepharose resin was washed with 20 column volumes of wash buffer to remove non-specifically bound impurities.

NleD protein was cleaved from its affinity tag while bound to the glutathione resin by adding recombinant TEV protein (His-tag bound) in a ratio, 1 mg of TEV to 20 mg of NleD, and then incubating the mixture at 4°C for 3 h. The cleaved protein was eluted from the glutathione sepharose resin by washing the column with 50 mM Tris (pH 7.2), 100 mM NaCl, 1 mM DTT. The protein was then passed through 0.5 mL of Ni^{2+} NTA column equilibrated with the same buffer in order to remove His-tag bound TEV enzyme. A final yield of 8 mg/L of purified NleD was obtained. The purified protein was concentrated using a Centricon with a molecular weight cut-off of 10 kDa. The concentrated protein fraction (500 uL aliquots at about 5 mg/mL) was passed through Superdex 75 gel filtration column (bed volume of 27.7 mL) pre-equilibrated with 50 mM Tris (pH 7.2), 100 mM NaCl, 1 mM DTT, in order to obtain a homogenous protein fraction.

2.4 Determination of protein concentration

Protein concentration was determined using three different methods: 1) a Biorad protein assay based on the Bradford method; 2) a Bicinchoninic acid assay (BCA) method; 3) absorbance measurement of the protein at 280 nm. The standard curves for the protein kit based assays (Bradford and BCA) were constructed using BSA (43824 $M^{-1}cm^{-1}$) [81], and for $O.D._{280}$ an extinction coefficient for NleD of 11460 M^{-1} cm^{-1} was determined from its primary amino acid sequence using Expasy-protparam (http://web.expasy.org/protparam/).

2.5 Mass spectrometry analysis of purified wild-type NleD

The quality of purified NleD was further verified using electrospray ionization mass spectrometry (ESI-MS). The protein samples were desalted using C_{18} resin equilibrated with 0.1% trifluoroacetic acid (TFA). After passing the protein through the resin, it was further

washed with 0.1% TFA to remove any unbound impurities. Finally, the bound protein is eluted with 80% acetonitrile and 0.1% formic acid. The samples were sent for analysis to the Mass Spectrometry Facility (Micromass Q-TOF Ultima) at the National Research Council. Data analysis and deconvolution of the charged envelope were performed using Mass Lynx 4.0 software supplied with the instrument.

2.6 Far-UV CD spectroscopy of NleD

Far-UV CD spectra of NleD were collected using Jasco-815 spectropolarimeter equipped with a Peltier heating/cooling temperature control system. All the spectrum for wild-type NleD was recorded at 20 °C in a 0.1 cm path-length rectangular cell (250 μL) from 260 nm - 190 nm with the following parameters: 20 nm/min scan rate, 0.2 nm resolution, 0.25 sec response time, 1 nm bandwidth and a sensitivity of 100 mdeg. The protein was dialyzed against 50 mM Tris (pH 7.2), 100 mM NaF overnight. For the final spectrum five accumulations were averaged and the contribution of the dialysis buffer was subtracted. Protein concentration was determined by absorbance readings at 280 nm (see section 2.4). The web based software Dichroweb (http://dichroweb.cryst.bbk.ac.uk/html/process.shtml) was used for the calculation of secondary structure content from the CD data obtained.

For variable temperature experiments, changes in ellipticity at 222 nm for wild-type NleD and the variants were recorded from 25°C to 65°C. Temperature was controlled by a Peltier system and was increased at a rate of 20°C/hr, with a resolution of 0.2°C. The protein samples were prepared in the same buffer as described in the previous paragraph. Smoothing of the thermal denaturation curves was performed using JASCO software supplied with the instrument (default settings) and T_m values were calculated using the first derivative of the curves.

2.7 Fluorescence spectroscopy for NleD and the variants

The fluorescence emission spectra for the wild-type NleD and the variants were recorded at room temperature using Varian Eclipse Spectrofluorimeter at the National Research Council. A Varian quartz cuvette was used for the readings with pathlength 1 cm X 1 cm. Protein purified in 50 mM Tris (pH 7.2), 100 mM NaCl, 1 mM DTT was diluted to a concentration of 8 μM. The excitation wavelength used was 295 nm and the emission spectra were recorded from 310 nm to 400 nm. The fluorescence emission scans were measured at a "slow scan rate". Each spectrum shown is an average of 15 scans and the spectra were corrected for the buffer. Excitation and emission bandwidth used were 5 nm. The λ_{max} for every emission spectrum was calculated from the first derivative of the curve and the entire data was exported to Excel for making the plots.

2.8 Quaternary structure determination

2.8.1 Size exclusion chromatography

The native molecular weight of wild-type NleD was determined by SEC at room temperature using a Superdex 75 column (column volume = 27.7 mL). Experiments were conducted at five different concentrations of protein and in the presence and absence of zinc. The mobile phase consists of 50 mM Tris (pH 7.2), 100 mM NaCl, 1 mM DTT. The molar concentration of zinc chloride ($ZnCl_2$) added to the buffer was same as the molar concentration of the protein injected into the column, 13μM. Buffers were filtered using 0.44-micron pore size membrane from Millipore and were degassed prior to equilibration of the column. The volume of protein injected was 500 μL and the flow rate was set to 400 μL/min. Elution was monitored at O.D. $_{280}$ .The molecular weight standard curve is shown in Fig. 3.4.1. Calibration curve was generated as a plot between the distribution coefficient (Kd) against the log of molecular mass of the protein standard. The relationship used to calculate the distribution coefficient was Kd = (V_e-

V_o) / V_i where $V_i = V_t - V_o$ and V_o is the column void volume (Blue Dextran), V_t is the total volume of the column, V_i is the internal volume of the column and V_e is the elution volume.

2.8.2 Dynamic light scattering

A Wyatt Dynapro apparatus was used for the analysis of wild-type NleD. For DLS experiments the buffer conditions, the temperature and the protein concentrations were the same or very similar as those outlined for SEC. The experiments were also performed in the presence and absence of zinc chloride (see Fig 3.4.3 for details). Prior to collection of data the samples were centrifuged at 21100Xg for 5 min in order to remove insoluble aggregates. For each sample measurement, data were collected for 100 acquisitions with 10 s averaging time per acquisition.

2.8.3 Analytical ultracentrifugation

Sedimentation velocity data were collected on a Beckman XL-1 analytical ultracentrifuge present at Concordia University. The instrument was equilibrated at the desired temperatures *i.e.* 4°C and 15°C, for two different runs. NleD was extensively dialyzed against 50 mM Tris (pH 7.2), 100 mM NaCl, 0.5 mM TCEP, and then 500 μL of the protein sample was then loaded into the cell. The sample was centrifuged at 42000 rpm for 12 h with the same buffer used as blank in the balancing cell. A total of 200 scans were collected per run in which the optical density at 280 nm was measured.

The quaternary structure of NleD as a function of the sedimentation coefficient (Fig. 3.4.4) was analyzed using Sedfit (http://www.analyticalultracentrifugation.com/download.htm). Sedenterp (http://www.jphilo.mailway.com/download.htm#SEDNTERP) was used for estimation of the appropriate parameters such as solution viscosity, solution density, etc.

2.9 1D-^{1}H NMR for NleD

NMR sample was prepared by dialyzing the purified protein against 50 mM phosphate (pH 7.2), 100 mM NaCl, 1 mM DTT in 99.9% D_2O at 4°C overnight. The dialysis buffer used was 10 times the volume of the protein sample. The dialyzed sample was transferred to a Shigemi tube and the 1D-1H spectra were collected room temperature on Bruker 600 MHz instrument at the National Research Council. The experiment was conducted at room temperature. A 0.1 N NaOD solution was used for pD titration of the protein sample. The pD of the protein was measured before and after the collection of the 1D-^{1}H spectra. A total of nine spectra were collected with a pD interval of ~ 0.2 units. The operational pH of the buffer solution in 99.9% D_2O (measured with a glass electrode) is converted to pD value by adding 0.4 [79].

2.10 Protein expression and purification of p38α kinase

The cDNA for p38α kinase (MAPK14) was obtained from Harvard Medical School, PlasmID collection (Source: MGC) consortium and was cloned into pMCSG7 (modified pET21a) by Ms. Laura McDonald. The construct contained an N-terminal hexa-His tag preceded the gene of interest and ampicillin resistance as a selectable marker. The recombinant p38α kinase was expressed in *E. coli* strain BL-21 Gold. The BL-21 Gold competent cells transformed with recombinant p38α kinase were grown at 37°C in LB medium and protein expression was induced with 500 μM of IPTG when an $O.D._{600}$ between 0.6 – 0.8 was reached. The induction was performed at 20°C for 16 h in an incubator shaker. Cells were harvested by centrifugation at 8000 Xg for 30 min at 4°C and the cell pellets were stored at -20°C.

Cell pellet was resuspended in lysis buffer of 50 mM Tris (pH 7.2), 100 mM NaCl, 1mM DTT, supplemented with 0.2 mM PMSF, 20 mM imidazole, and a protease inhibitor cocktail (prepared according to manufacturer instructions). Cells were lysed by sonication for 3min in

bursts of 10 s and placed on ice between bursts. Insoluble debris was removed from the lysate by centrifugation at 28,384Xg for 1 h at 4°C using a Beckman centrifuge. Affinity purification was performed using Ni-NTA resin. Briefly, the supernatant obtained was incubated with 5 mL Ni-NTA resin pre–equilibrated with a buffer of 50 mM Tris (pH 7.2), 100 mM NaCl, 1 mM DTT, and 20 mM imidazole, and the mixture was gently shaken at 4°C for 1 hr for effective binding of protein to the resin. The Ni-NTA resin was washed with 20 column volumes of wash buffer to remove non-specifically bound impurities. Finally, the protein of interest was eluted using a buffer of 50 mM Tris (pH 7.2), 100 mM NaCl, 1 mM DTT, and 250 mM imidazole. The eluted protein was then dialyzed at 4°C against same buffer but without imidazole. A final yield of 35 mg/L of purified hexa-His p38α kinase was obtained.

As outlined for the purification of NleD, the concentrated protein fraction was further passed through Superdex75 gel filtration column pre-equilibrated with 50 mM Tris (pH 7.2), 100 mM NaCl, and 1 mM DTT, in order to obtain homogenous hexa-His p38α kinase.

To obtain preparations of p38α kinase without the N-terminal hexa-His tag, the purified protein was incubated with recombinant TEV protein (His-tag bound) in a ratio, 1 mg of TEV to 20 mg of NleD, and the mixture was incubated at 4°C for 2 h. The protein was then passed through 0.5 mL of Ni^{2+} NTA column equilibrated with the same buffer in order to remove His-tag bound TEV enzyme.

2.11 Proteolytic activity assays for NleD and the variants

The enzymatic reaction was performed at 4°C for 10 days in a buffer of 50 mM Tris (pH 7.2), 100 mM NaCl, and 1 mM DTT. The reaction mix was composed of 10 μM of purified NleD as the protease, 10 μM of purified affinity tagged p38α kinase as substrate and 10 μM of zinc chloride. After 24 h intervals, 5 μL of the reaction mixture were removed and centrifuged at

21100Xg for 5 min to remove any precipitated aggregates. To stop the reaction, the supernatant was mixed with 5 μL of 2X gel loading buffer (1.5 M Tris-HCl, 4% SDS, 20% glycerol, 0.2 M DTT and a pinch of bromophenol blue) and the mixture was boiled at 100°C for 5 min followed by immediate cooling on ice. The aliquot was then stored at -20°C before loading onto an SDS-polyacrylamide gel.

SDS-PAGE was performed based on the standard protocol from Sambrook and Russel [80]. The stacking gel (pH 6.8) and resolving gel (pH 8.3) contained 4% and 12% acrylamide, respectively, and were made in Tris-buffer. Pre-made 30% acrylamide mix from Sigma was used. The Tris buffers were made in distilled water and were stored at room temperature. A 10% SDS solution was made in distilled water and stored at room temperature. A 10% ammonium persulfate solution prepared in water was stored at -20°C. Electrophoresis buffer was composed of 25 mM Tris base, 250 mM glycine and 0.1% SDS, pH 8.3. Samples loaded onto the SDS polyacrylamide gel were electrophoresed at 200 V using a Biorad power supply.

Silver staining was used to visualize the protein bands. Gels were soaked in fixer composed of 50% ethanol and 5% acetic acid for 30 min. After fixing, washing was performed in 50% ethanol for 10 min followed by 2x10 min washes in distilled water. Sensitization was accomplished in 0.02% sodium thiosulfate for 1-2 min followed by immediate washing with distilled water for 10 min. Gels were then stained in a 0.1% silver nitrate solution for 30 min with gentle agitation. Lastly, the gels were soaked in a developer solution composed of 2% sodium carbonate and 0.03% formaline for 3-5 min or until the protein bands were clearly visible.

Chapter 3: Results

3.1 Expression and purification of NleD and the variants

E. coli Bl-21 Gold cells transformed with the expression vector containing the NleD gene were grown in Terrific Broth with ampicillin as a selectable marker. Protein expression was induced with 0.5 mM IPTG and soluble NleD proteins were purified using glutathione affinity chromatography. The purification protocol used is outlined in section 3.1.1. All the NleD variants (H149A, D174A, D175A, E191Q, R203K, Y206F) contained single amino acid replacements of residues, which were predicted to be in the enzyme's active site and/or were highly conserved residues.

Denaturing polyacrylamide gels in Fig 3.1.1 show the successful expression and purification of wild-type NleD (panel A) and a comparison of the purity of wild-type protein and the variants (panel B). Coomassie brilliant blue was used for staining and visualization of protein bands. The major band of ~26 kDa in panel A corresponded to the mass predicted for monomeric NleD and indicated that the wild-type protein was purified to near homogeneity. Although insufficient NleD protein was analyzed in panel B, the results suggested that the variants were purified to the same extent as the wild-type protein.

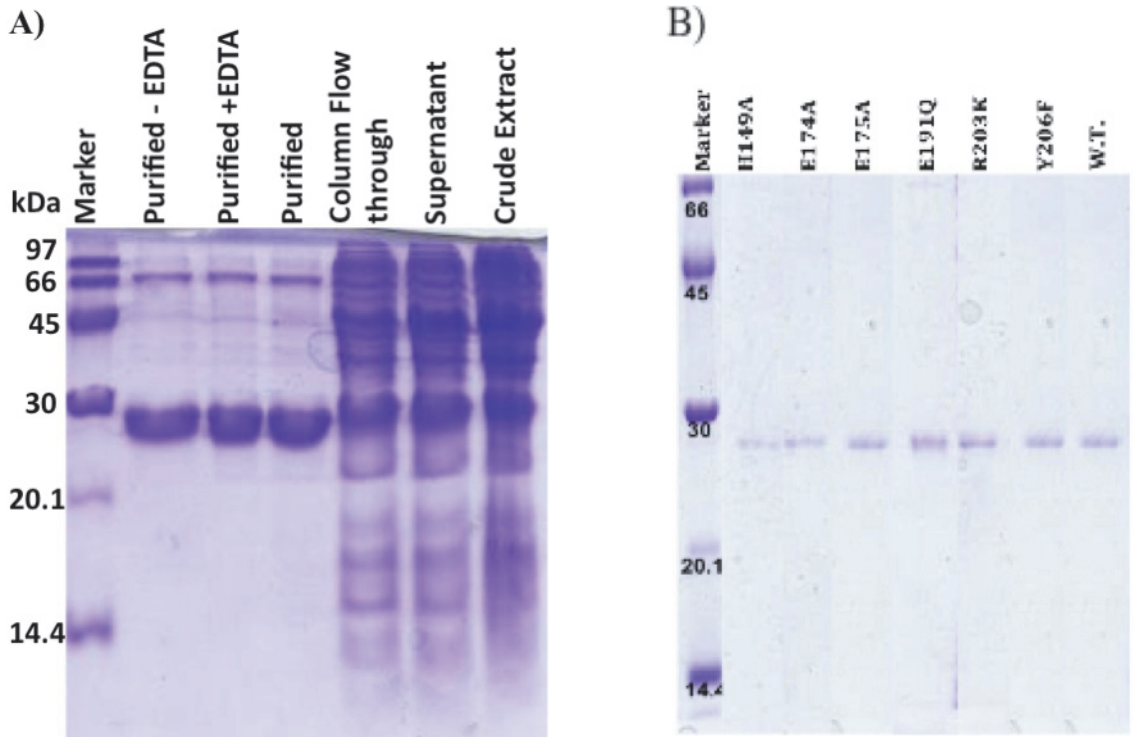


Fig. 3.1 SDS-PAGE analysis of NleD proteins. A) SDS-PAGE analysis (15% acrylamide) of the purification procedure of wild-type NleD by glutathione sepharose affinity chromatography. B) SDS-PAGE analysis (12% acrylamide) of wild-type NleD and the variants purified by affinity chromatography prior to SEC.

3.2 Characterization of NleD using electrospray ionization mass spectrometry

Purified NleD was also characterized using ESI-Q-TOF mass spectrometry. The charge envelope of the purified wild-type protein is shown in Fig. 3.2.1 and its deconvolved spectrum is shown in Fig. 3.2.2. The deconvolved mass spectrum showed the presence of one major peak at 25947.0 Da, which is in excellent agreement with the theoretical mass of the wild-type protein based on its primary amino acid sequence (25947.0 Da). Worth noting, the N-terminal methionine (the first amino acid residue in the sequence) is replaced by glycine and serine in recombinant NleD, a modification that is derived from the expression construct.

In addition to the peak corresponding to the full-length protein, four other peaks of significantly lower intensity were observed in the deconvolved spectrum and the masses of these

species are listed in Table 3.2.1 Interestingly, careful examination of the charge envelop in Fig. 3.2.1 also reveals these minor species, although one additional peak of 17315.7 Da observed in the deconvolved spectrum (Fig. 3.2.2) had no identifiable peaks in the charge envelope of purified NleD (Fig 3.2.1). The number of amino acid residues missing from the C or N terminus of partially degraded NleD was inferred by comparison of the measured mass to the mass calculated from protein sequences using PAWS proteolysis software (http://bioinformatics.genomicsolutions.com/paws.html#). The analysis indicates that proteolytic cleavage occurs at three different positions in the C-terminal region of the full-length protein (Fig. 3.2.3).

Many careful measures were taken to avoid proteolysis during the expression and purification of the enzyme, such as the inclusion of a protease inhibitor cocktail, the addition of 5 mM EDTA and of 0.2 mM PMSF to the cell lysis buffer before sonication, performing all purification steps in the cold room (5°C), and conducting the induction of protein expression at 15°C. Nevertheless, from the results of mass spectrometry analysis we estimate that ~ 20% of purified NleD contained truncations in the C-terminal region of the polypeptide. In order to eliminate and or reduce proteolysis, a C-terminal truncated construct (amino acids 2-212) should be considered for future studies, as it may yield a homogenous protein sample suitable for structure determination by X-ray crystallography or NMR. The current preparation of purified NleD has not yielded diffraction quality crystals.

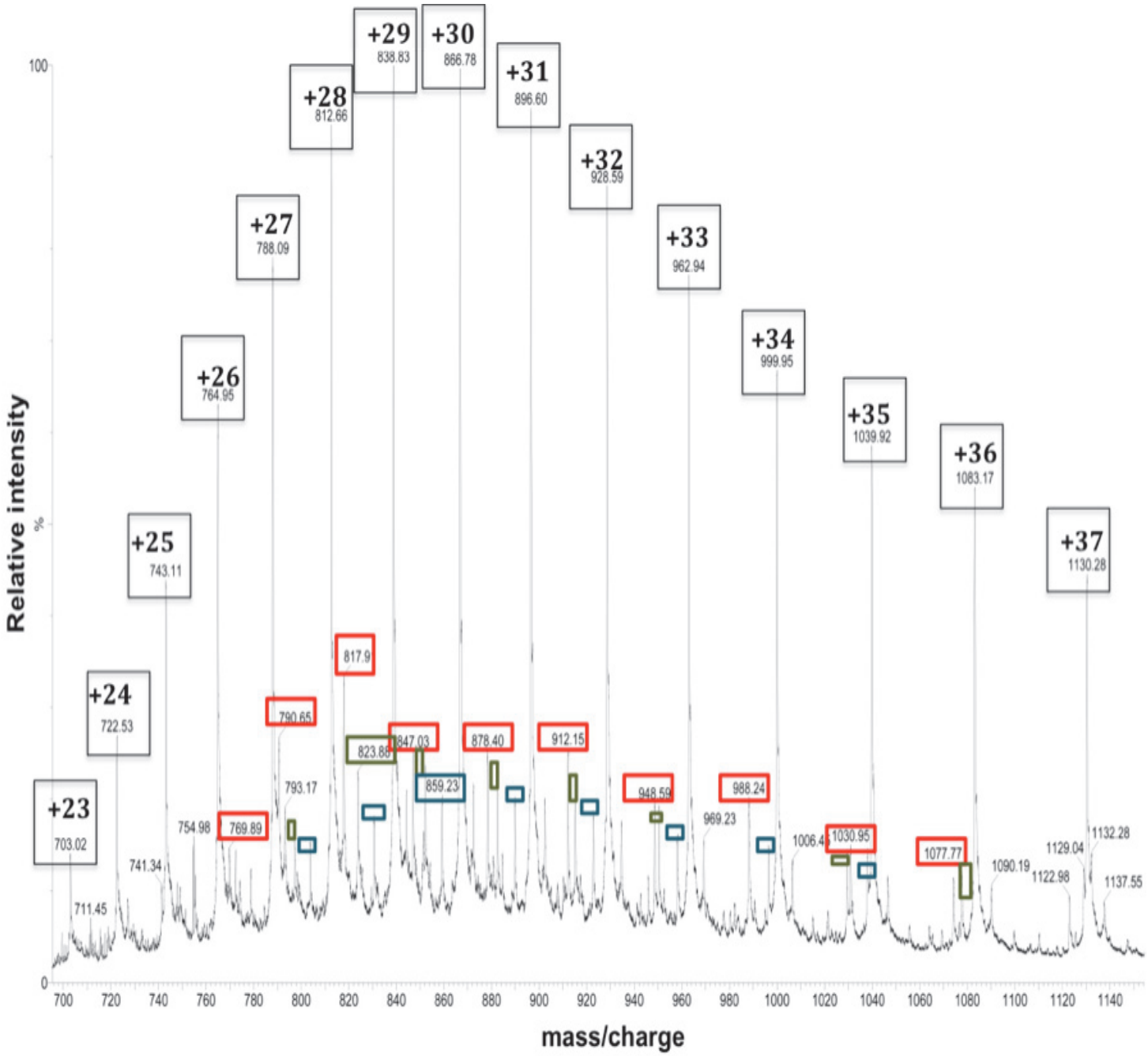


Fig. 3.2.1 Mass spectrum of wild-type NleD showing the major charge envelope of the purified NleD. The major peaks correspond to full-length NleD while some smaller peaks likely originate from the truncated versions of the full-length NleD due to uncontrolled proteolysis during purification. Species in the charge envelope of one fragment (represented as colour-coded boxes) represent the number of protonation states of that fragment (an example is shown for the full-length protein fragment). Other fragments are represented as: blue box, fragment 1; green box, fragment 2; red box, fragment 3 (see Table 3.2.1 for details).

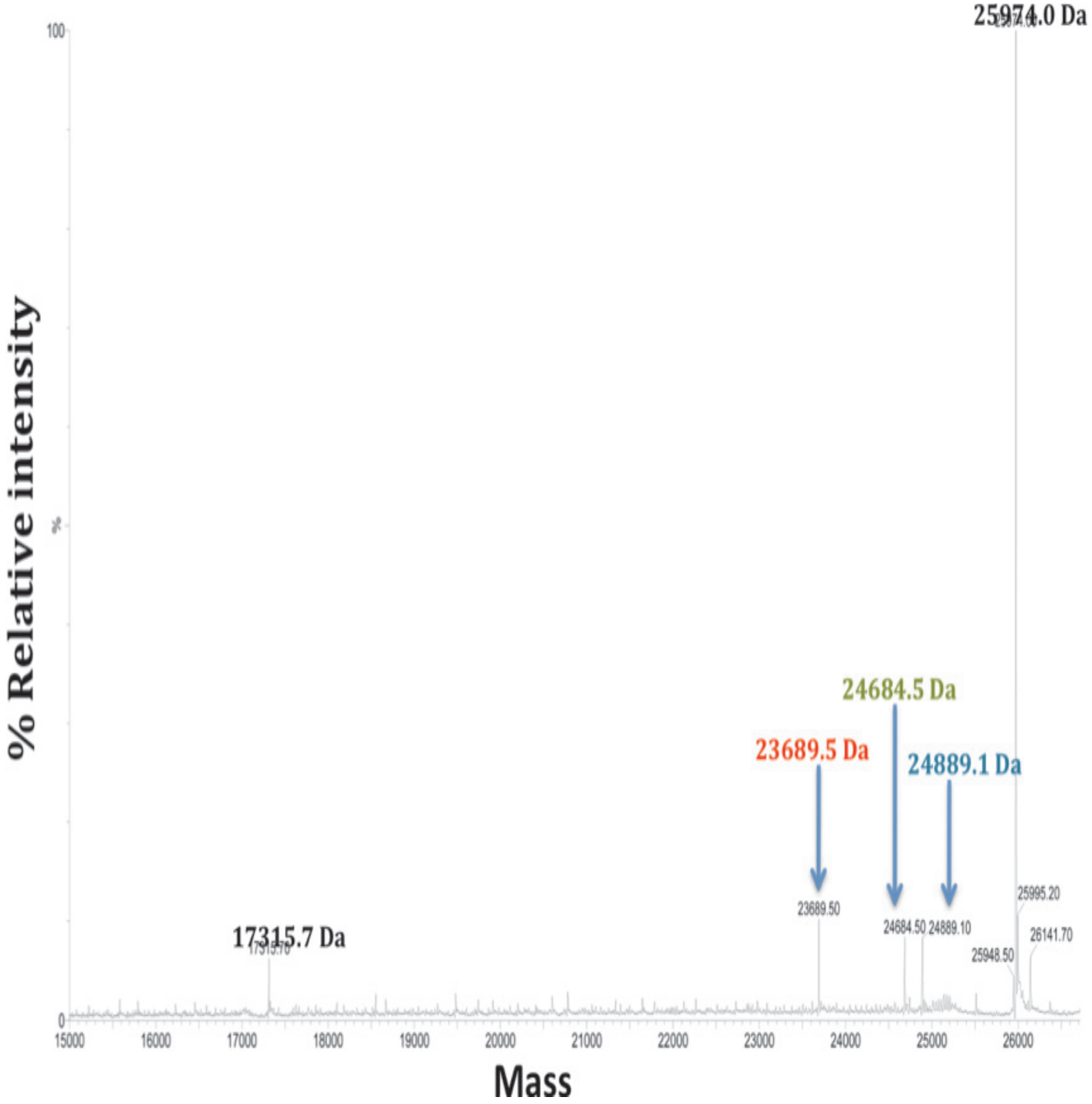


Fig. 3.2.2 Deconvolved ESI –MS spectrum of purified NleD. A major peak of 25974.0 Da represents the full-length protein. The smaller peaks of 24889.1 Da, 24684.5 Da and 23689.5 Da represents fragments of NleD. Software used for the analysis was Mass lynx 4.0.

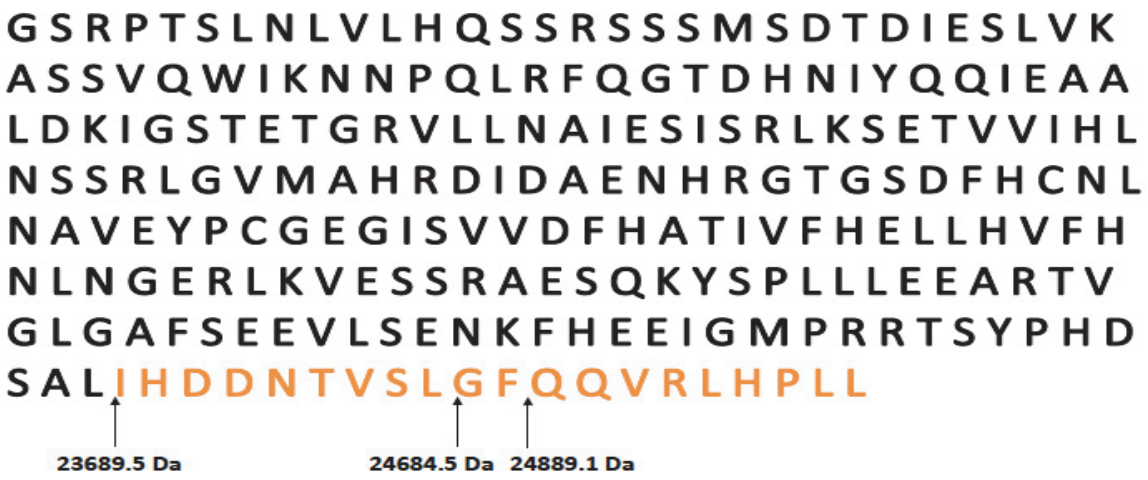


Fig. 3.2.3 The primary sequence (residues 1 to 233) of purified NleD and the cleavage sites based on the analysis of mass spectrometry data. The N-terminal methionine is replaced by glycine and serine, which is a cloning artifact derived from the cloning vector.

Peptide Fragment	A.A. Length	Observed Mass (Da)	Predicted Mass (Da)	Difference (Da)
Full-length	1 - 233	25974.0	25974.0	0
1	1 - 224	24889.1	24888.7	0.4
2	1 - 222	24684.5	24684.4	0.1
3	1 - 213	23689.5	23689.4	0.1

Table 3.2 The comparison of the observed mass obtained from the primary amino acid sequence and the predicted mass obtained from the analysis of the mass spectrometry data. A difference of less than 1 Da was obtained which was sufficient to accurately identify the cleaved fragments.

3.3 Analysis of secondary structure using far-UV CD spectroscopy

Far-UV CD was used to determine the secondary structure content of the wild-type NleD protein. The amide and carbonyls present in protein backbone absorbs the circularly polarized light in the range of 170 nm to 250 nm. Hence, the spectra were scanned from 190 nm – 260 nm. The CD spectrum of NleD showed a double minimum at 208 nm and 222 nm (Fig. 3.3.1) characteristic of α-helical structure within the folded protein. The secondary structure analysis of the CD data was performed using Dichroweb (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml). Secondary structure analysis predicted NleD to be composed of approximately 47% α-helices, 13% β-pleated sheets and 37% unordered structure and turns.

The secondary structure content was also analyzed based on the primary amino acid sequence of NleD using Psipred (http://bioinf.cs.ucl.ac.uk/psipred/). The bioinformatic analysis predicted NleD to contain 31% α-helices, 11% β-sheets and the rest (59%) unordered structures and turns (Fig: 3.3.2). The comparisons of the results from the two predictive programs are in moderate agreement.

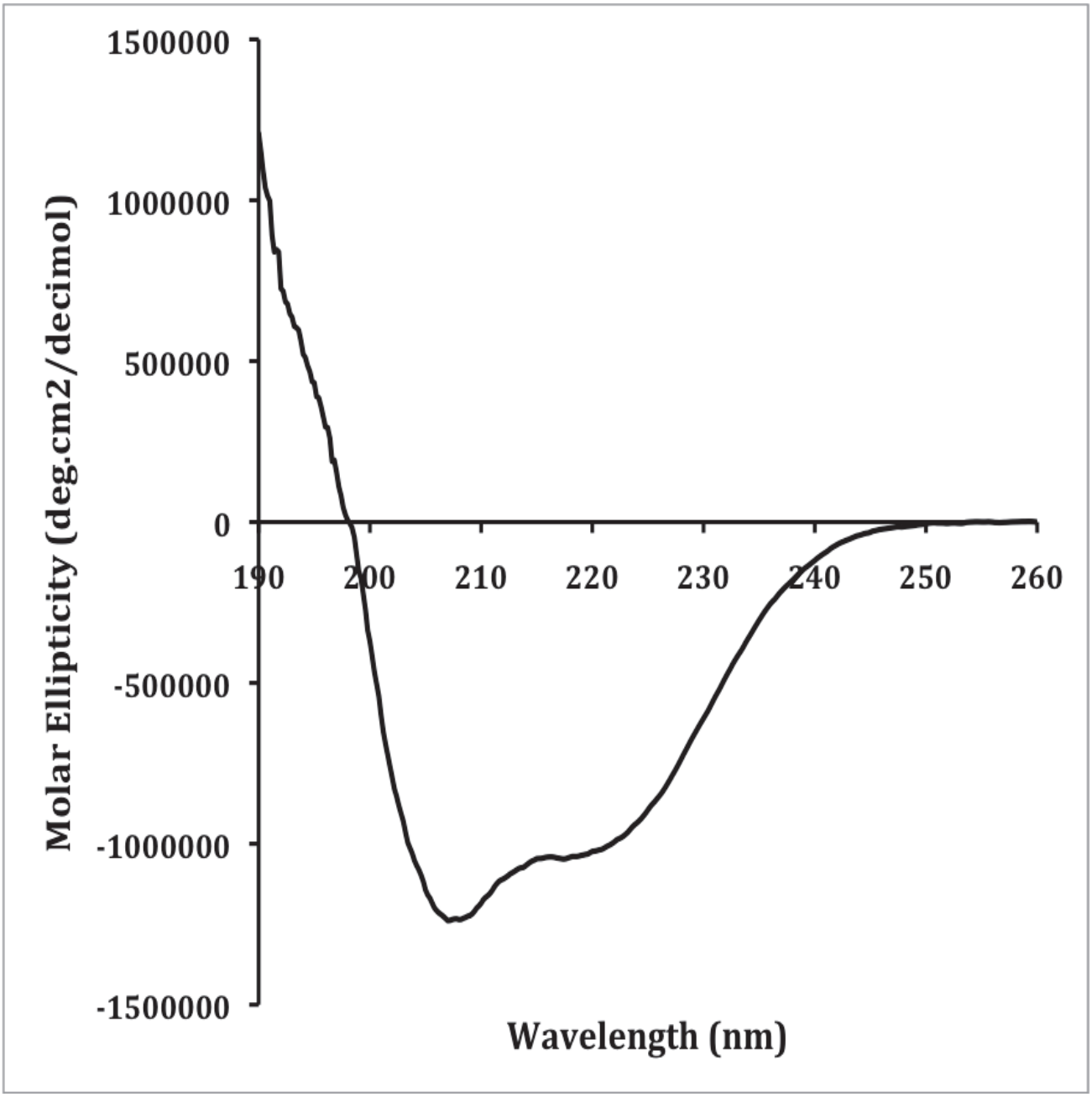


Fig. 3.3.1 Far-UV CD spectrum of wild-type NleD. The data are expressed as molar ellipticity. The curves represent an average of 5 scans obtained at a speed of 20 nm/min. The CD signal in millideg was converted to molar ellipticity using the relationship $\theta_m = 100. \theta/m.l$, where θ is the observed ellipticity, m is the molar concentration of the protein and l is the path length of the cell in cm.

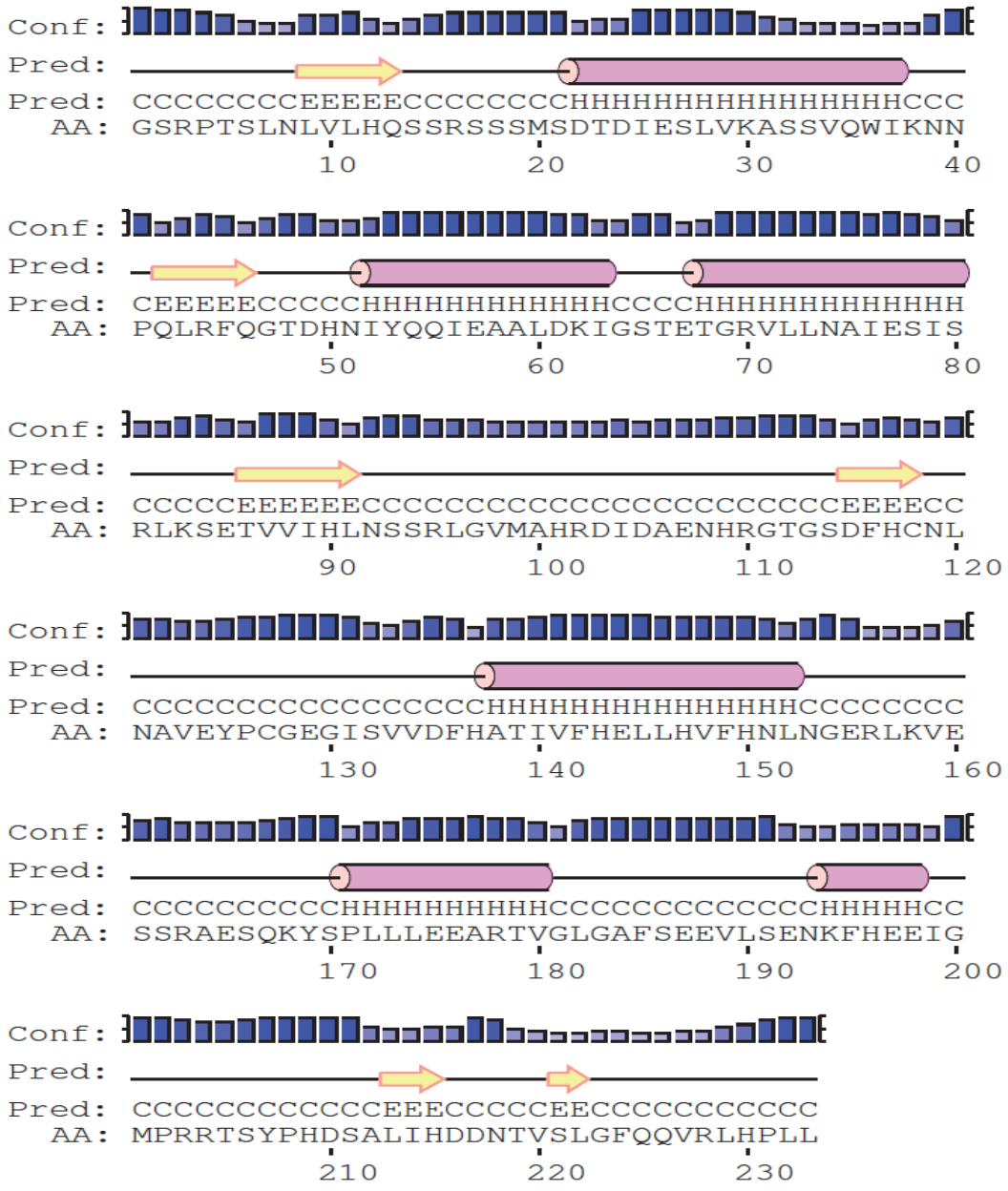


Fig. 3.3.2 Secondary structure prediction of NleD using Psipred.. The blue bars on top of the sequence represents the confidence level of the secondary structure prediction by Psipred [77] (http://bioinf.cs.ucl.ac.uk/psipred/). Yellow arrows represents predicted β-sheets, pink cylinders represents predicted α-helices, the straight line represents unordered regions. C represents random coil, E represents β-sheets and H represents α-helical regions in the protein, based on the Psipred prediction.

3.4 Quaternary structure determination of NleD using different techniques

The purpose of the present study is to determine the quaternary structure of NleD *in vitro*. Size exclusion chromatography (SEC) exploits the hydrodynamic properties of the analyte, which does not interact with the stationary phase [65]. Since the hydrodynamic properties of a protein molecule in solution may not be fixed, the technique will predict arbitrary molecular weight. NleD was characterized by SEC at room temperature using a Superdex 75 column. Characterization was performed using six different concentrations of NleD (from 0.78 μM to 108 μM) in a Tris buffer (50 mM Tris (pH 7.2), 100 mM NaCl and 1 mM DTT) as the mobile phase (Fig. 3.4.2). In order to determine the molecular weight in the presence of zinc, NleD was dialyzed overnight against same buffer containing an equimolar concentration of zinc chloride. The dialyzed protein sample was characterized by SEC, using as a mobile phase, the same buffer components as indicated for dialysis. Protein standards used to calibrate the molecular weight based on their elution volumes are shown in Fig: 3.4.1. The standards were analyzed using the same conditions described for NleD. The elution profiles for NleD acquired at different concentrations of protein were superimposable with respect to the elution volumes (Fig. 3.4.2). Over the concentration range investigated in this study, NleD eluted as a single peak at ~ 15.57 mL suggesting that the protein's quaternary state is independent of concentration. A similar elution profile for the protein was obtained in the presence of a mobile phase containing zinc chloride $ZnCl_2$, which shows that NleD's quaternary structure was unaffected by zinc binding (Fig. 3.4.2). Worth noting, $ZnCl_2$ is very soluble in the mobile phase containing Tris buffer. The equation obtained from the logarithmic plot of molecular weights and the distribution coefficients of the standard molecular weight markers was used to estimate the molecular weight of NleD. The value of ~ 38 kDa lay between that predicted for the monomer (25.9 kDa) and the dimer (51.8 k Da).

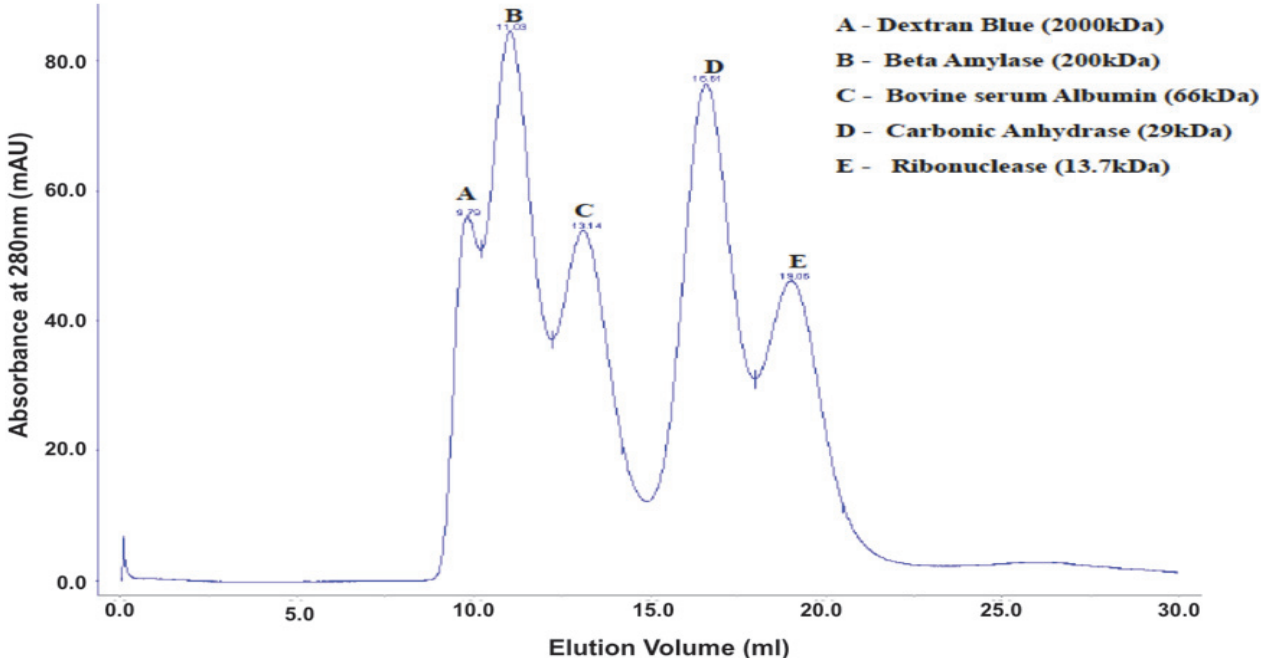
NleD's quaternary structure was also characterized by dynamic light scattering (DLS). The data (Fig. 3.4.3) were collected at the same buffer conditions and over a similar range of protein concentrations as performed for SEC. At all concentrations, NleD is found to have an average hydrodynamic radius of ~ 2.9 nm. The average percent polydispersity was ~ 14% suggesting the purified NleD solution was likely monodispersed. The experiment was also conducted in the presence of equimolar concentrations of zinc chloride. In all the cases the average molar mass was found to be ~ 40 kDa, which was consistent with the data obtained from SEC. Again the predicted molar mass was in between the molecular weight of the monomeric or dimeric species.

We also exploited a third technique, analytical ultracentrifugation (AUC) [47,48], in addition to SEC and DLS to determine the hydrodynamic properties of NleD.

Sedimentation velocity run at 42000 rpm was conducted on the purified NleD. Sedfit was used for data analysis. Analyzed data was obtained in the form of sedimentation coefficient distribution C(s) as a function of sedimentation coefficient. The C(s) analysis describes the sedimentation boundary of the species present in the solution with respect to their s-values. It results in separate peaks for multiple sedimenting species in solution. It is a useful technique to predict the molar mass of the unknown protein and to determine its hydrodynamic shape.

Initially the sedimentation velocity run was performed at 15°C. The sample used had a protein concentration of 0.45 mg/mL (18 μM). On analyzing the data it was observed that NleD sedimented as a monomer with a sedimentation coefficient of 1.83 and root mean square deviation (RMSD) of 0.00537. The frictional ratio of 1.09 showed that the sedimented protein was most likely globular in nature. In addition to a major peak of the monomer, a second broad peak (Fig. 3.4.4) with a sedimentation coefficient of 3.25 was observed with a frictional ratio of ~ 4.5. One possibility might be that the protein solution at 15°C contains well-folded monomeric

NleD along with some tetrameric NleD protein or some partially unfolded monomeric species. Hence, the sedimentation behavior of NleD as a function of temperature was analyzed. An identical experiment was performed at 4°C as at 15°C -- under the same buffer conditions and at similar protein concentration (0.5 mg/mL). For this second centrifugation experiment, one single protein peak was observed which yielded a sedimentation coefficient ~ 1.9 and a RMSD of 0.00641 as derived from the Sedfit analysis. The frictional ratio obtained was 1.23 suggesting that the sedimented species (M.W. from AUC ~ 25.5 kDa) possessed a relatively compact globular structure. Interestingly, the second broad peak with a high sedimentation coefficient was not observed at 4°C in comparison to at 15°C (Fig. 3.4.4). This showed that most likely the existence of this second peak was dependent on temperature. Over findings promoted us to conduct additional experiments in order to determine the thermal stability of the protein (discussed later in section 3.5).



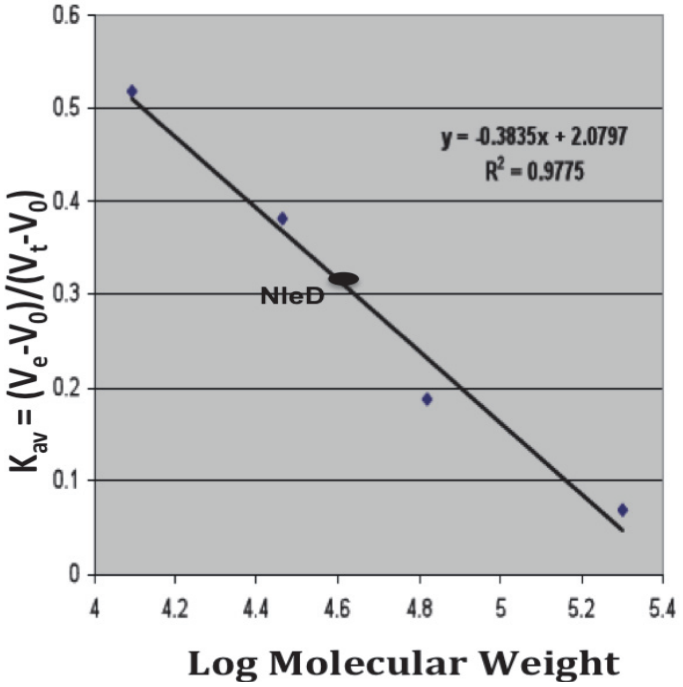
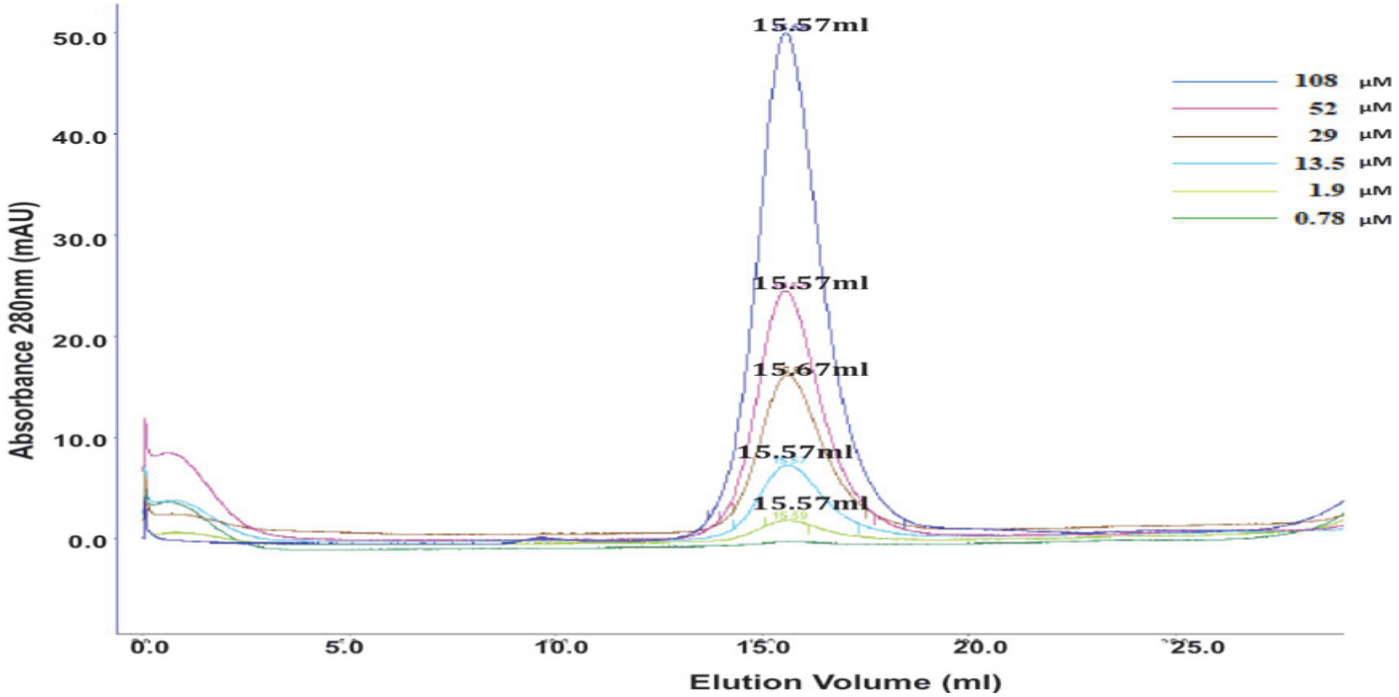


Fig. 3.4.1 Logarithmic plot of the standard molecular weight markers and their respective distribution coefficients. The plot was used as a calibration curve to estimate the molecular weight of wild-type NleD.

A)



B)

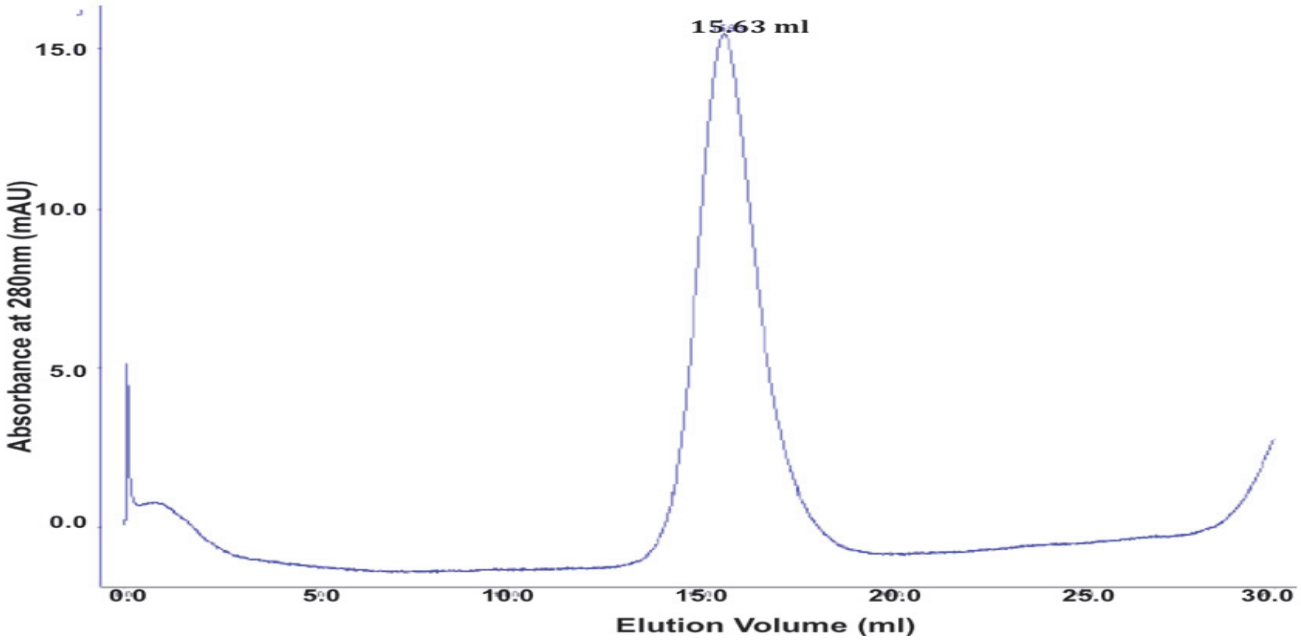
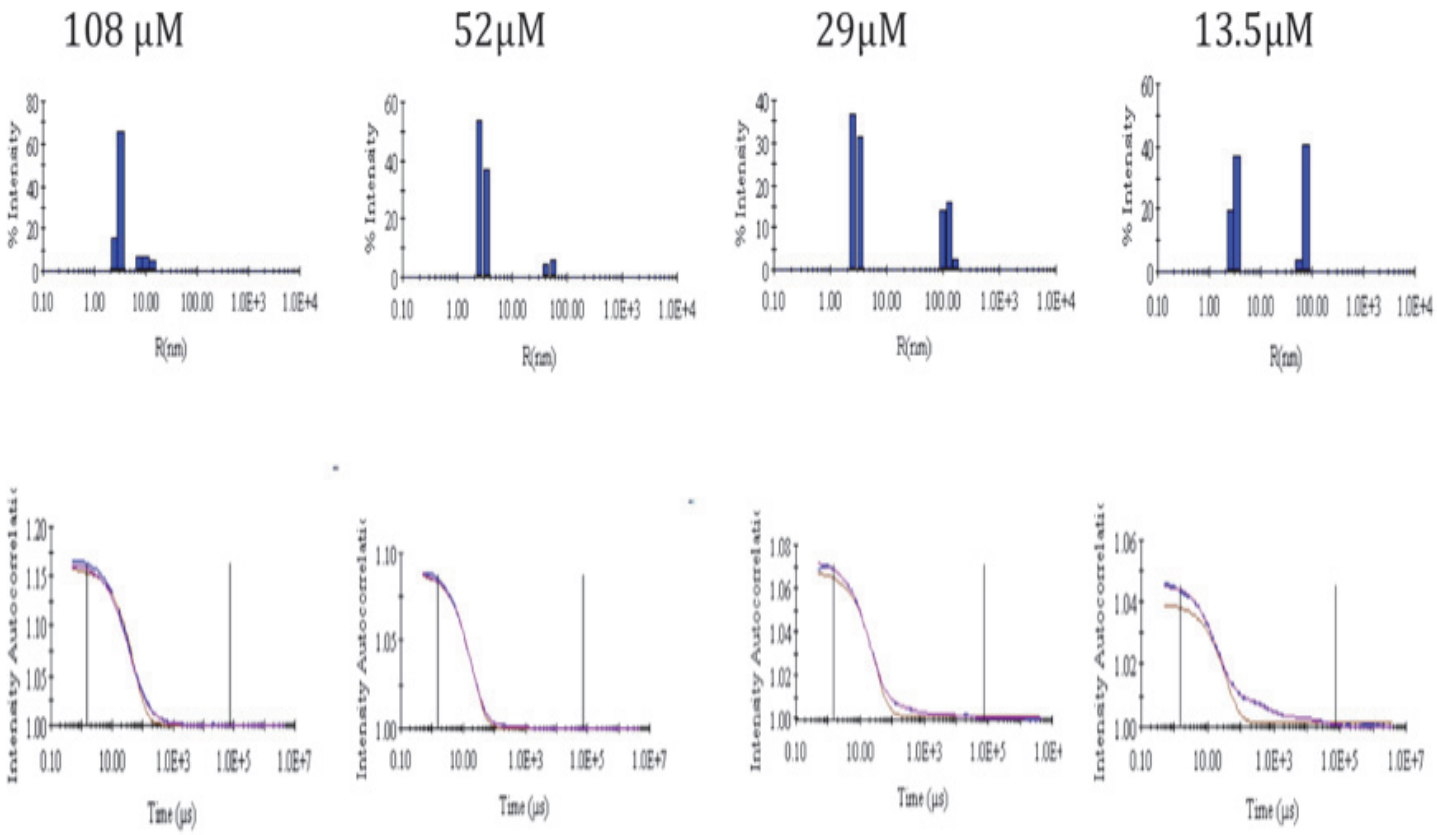


Fig. 3.4.2 Elution profile of wild-type NleD obtained from Superdex 75 chromatography. Panel A: shows the elution profile of wild-type NleD at different concentrations of protein. Panel B: shows the elution profile of wild-type NleD (13 µM) in the presence of $ZnCl_2$ (13 µM). Experiments were performed at room temperature.

A)



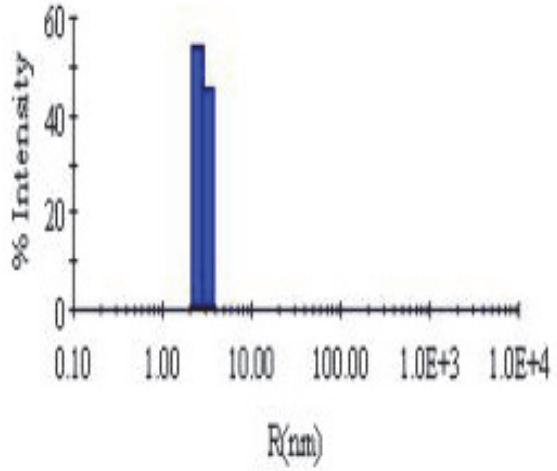
	Peak 1	Peak 2
Radius(nm)	2.9	50.2
% Pd	14	13.3
M.Wt.(kDa)	41	32116
%Intensity	90.4	9.9
%Mass	100	0

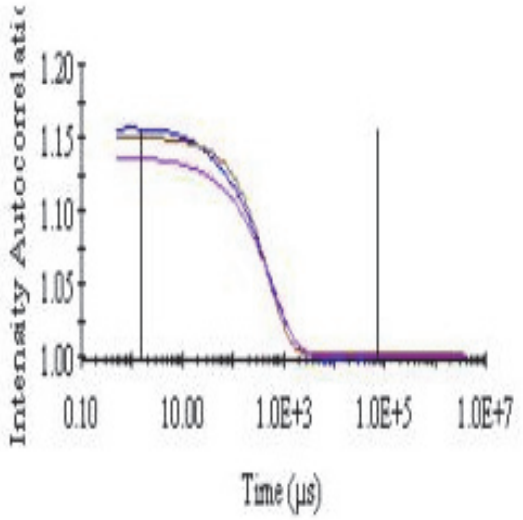
Peak 1	Peak 2
2.9	41.8
14	0
41	20854
90.1	9.6
100	0

Peak 1	Peak 2
3	119.5
14	17.9
42	244130
67.9	32.1
100	0

Peak 1	Peak 2
2.9	3.4
13.9	0
40	61
41.6	33.2
84.3	15.7

B)

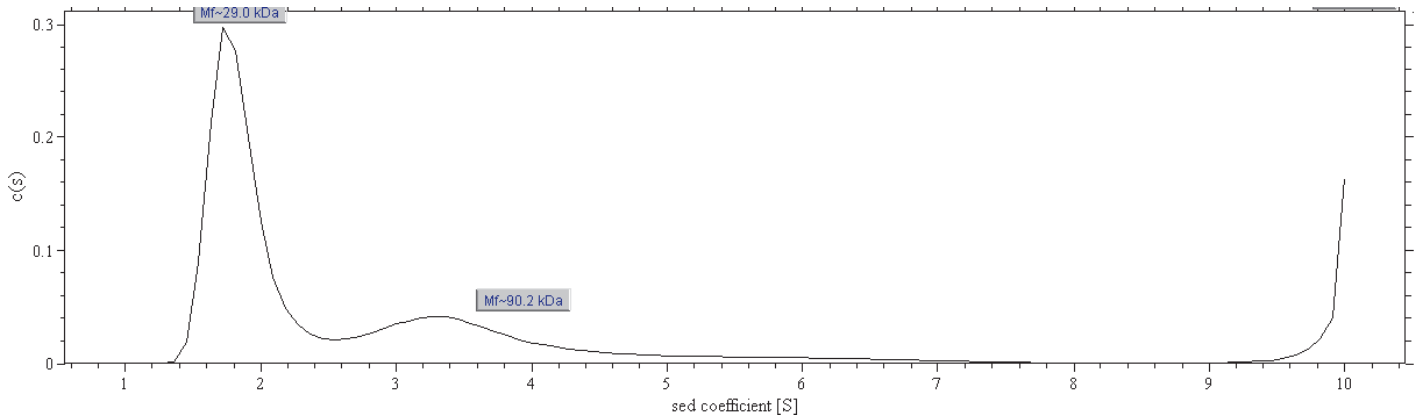




Radius(nm)	2.9
% Pd	14
M.Wt.(kDa)	40
%Intensity	100
%Mass	100

Fig. 3.4.3 Dynamic light scattering analysis of wild-type NleD. A) DLS data for apo - NleD at different concentrations. B) DLS data for NleD in the presence of equimolar concentrations of Zn Cl_2 (13 μM). Proper superposition of the autocorrelation curves (intensity of autocorrelation as a function of time (μs)) is evident for majority of the readings. The average molecular weight obtained from all the curves is ~ 41 kDa. For all the samples the data were collected at room temperature.

A) 15°C



B) 4°C

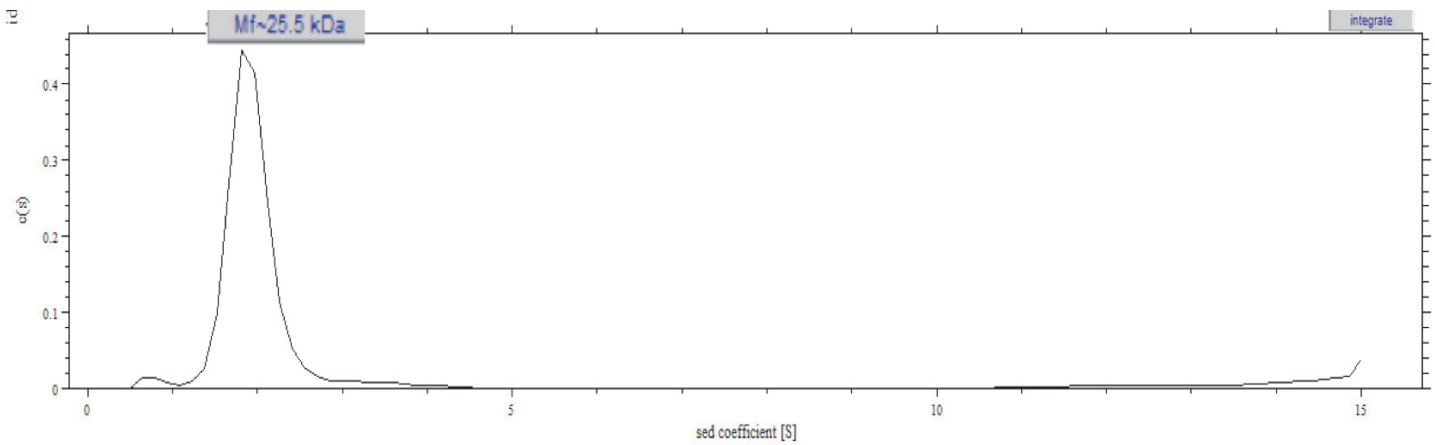


Fig. 3.4.4 Sedimentation velocity analysis of NleD. Analytical ultracentrifugation experiments were performed at A) 15°C and B) 4°C. C(s) distribution fit to sedimentation coefficient[s] was obtained using Sedfit . One major peak with a sedimentation coefficient between 1.83 and 1.9 (M.W. ~ 25.5 – 29 kDa) is observed at both the temperatures. An additional peak with a sedimentation coefficient of 3.25 (M.W. ~ 90.2 kDa) present at 15°C is absent in the experiment conducted at 4°C.

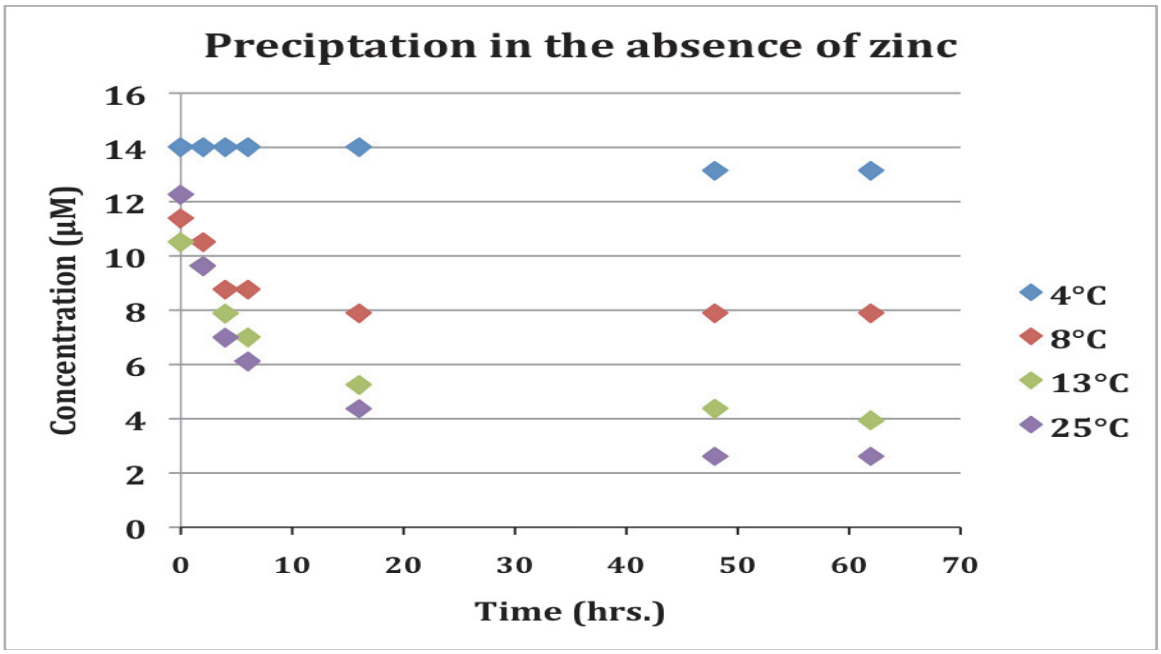
3.5 Characterization of NleD's solubility at different temperatures and in the presence of additives

The thermal stability of NleD was characterized by incubating 10-15 μM of purified NleD at 4°C, 8°C, 13°C and 25°C. Fractions were collected after specific time intervals, centrifuged at 21100Xg for 5 min, and then the $O.D._{280}$ measurements were recorded on the supernatant fractions. From these measurements the amount of soluble protein in the fractions was quantified and compared to amount of protein added initially to the reaction. The results are shown in Fig. 3.5.1 A and tabulated in Table 3.5.1. From all the temperatures tested, NleD remained the most soluble (precipitated the least) at 4°C. However, the rate of precipitation was higher in the presence of zinc chloride (Fig. 3.5.1 B).

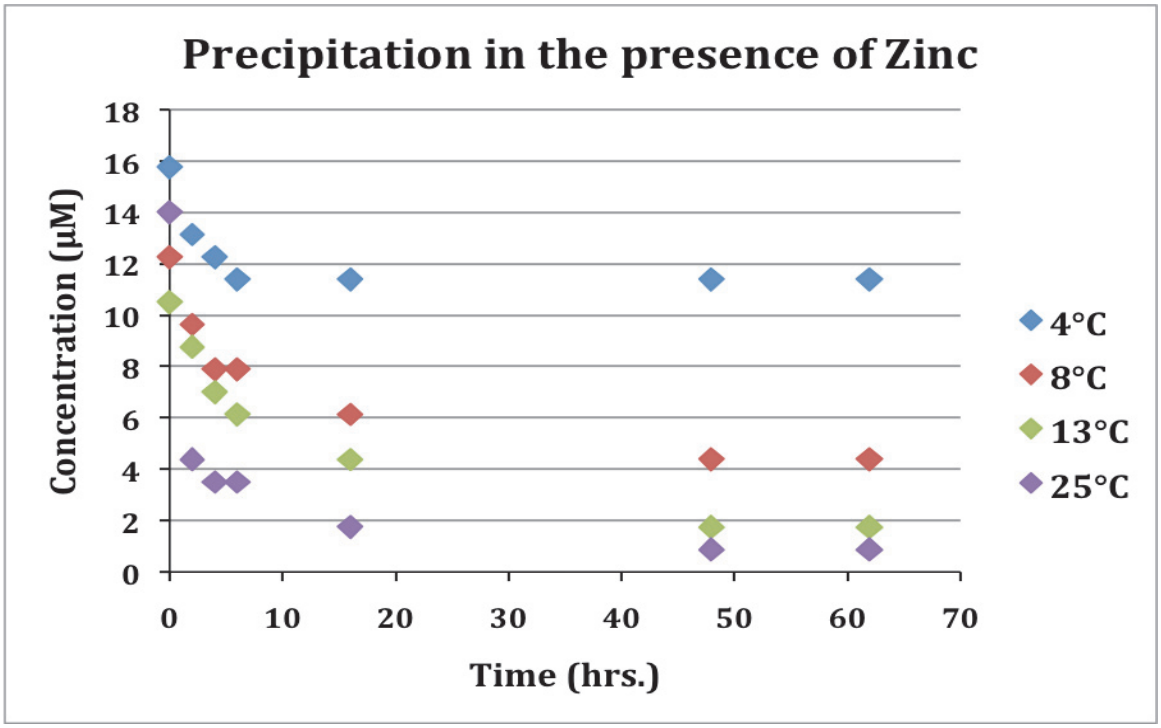
Polyols, trehalose, and sucrose are reported in the literature to be effective protein stabilizers[72,73,74]. Accordingly, in addition to temperature, different chemical additives, such as trehalose, ammonium sulfate and sucrose, were also added to the protein sample in order to record their impact on protein precipitation. Shown in Fig. 3.5.1 is a graphical representation of the concentration of protein remaining in solution (measured by calculating the molar concentration of NleD from protein absorbance at 280 nm) as a function of increasing incubation times at 15°C and in the presence of the specified additive. The results are summarized in Table 3.5.1. Trehalose appeared to be the most promising additive in helping to maintain protein solubility in the presence of zinc chloride at 15°C. However, there was no improvement in the stability of NleD with zinc at room temperature in the presence of trehalose (data not shown). Hence, extensive trials using different concentrations of trehalose or using combinations of different additives with trehalose needs to be performed in order to improve NleD's solubility significantly *in vitro* at room temperature. Interestingly, the most significant impact on protein

solubility was observed by reducing the incubation temperature to 4°C in the absence of any additives.

A)



B)



C)

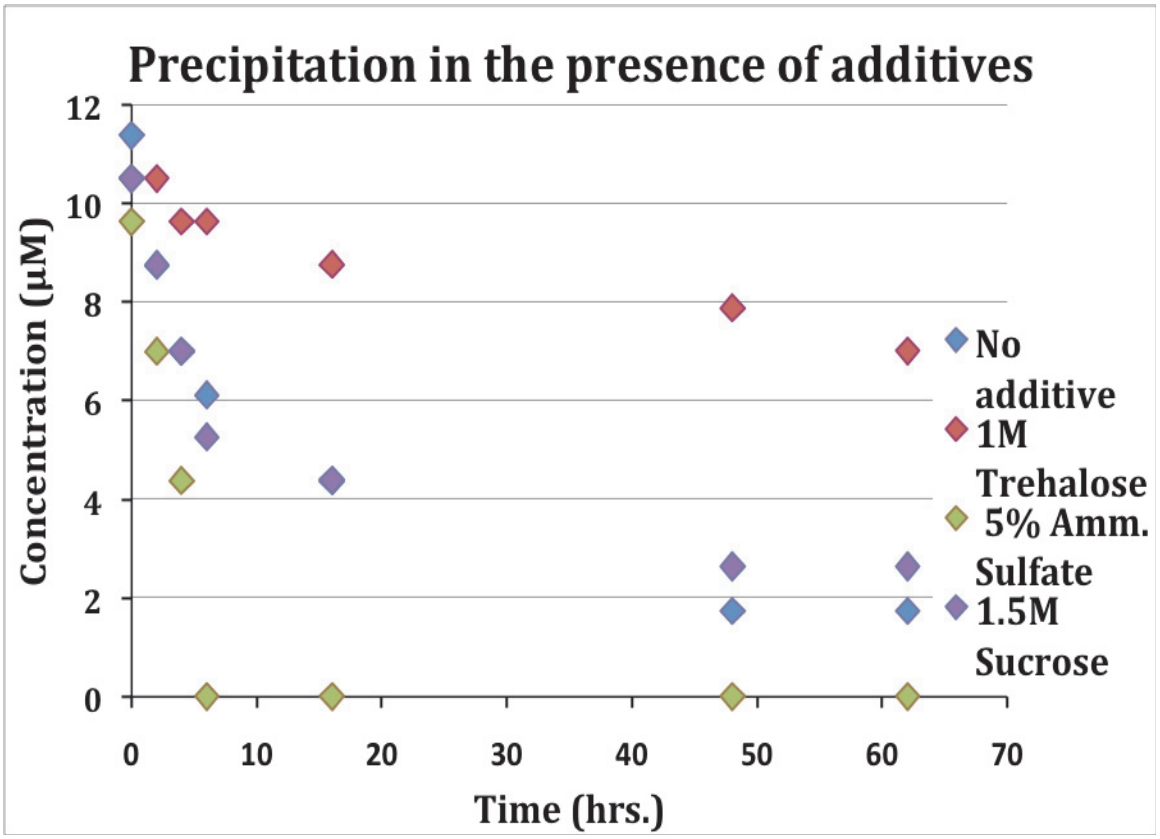


Fig. 3.5 Time-dependent precipitation of NleD under various experimental conditions. Purified NleD (10 µM- 16 µM) was incubated for increasing lengths of time in buffer 50mM Tris, 100mM NaCl, 1mM DTT. The concentration of soluble protein was determined by $O.D._{280}$ readings after centrifugation. When testing the effects of zinc chloride, an equimolar concentration of zinc chloride to protein was added. A) The decrease in protein concentration is higher at room temperature. In the absence of zinc chloride, the least amount of precipitation (highest amount of soluble protein) is observed at 4°C. B) Precipitation appears to be higher in the presence of zinc chloride. C) Protein solubility in the presence of zinc was also recorded at 15°C in the presence of 5% ammonium sulfate, 1.5 M sucrose and 1 M trehalose. Trehalose was the most promising additive.

	Conditions	% Precipitation after 62 h
	4°C	28%
With $ZnCl_2$	8°C	64%
10-16µM	13°C	83%
	25°C	94%
	4°C	6%
Without $ZnCl_2$	8°C	30%
10-16µM	13°C	62%
	25°C	78%
	No additive	91%
With $ZnCl_2$ at 15°C	1 M trehalose	32%
10-12µM	5% amm. sulfate	100%
	1.5 M sucrose	75%

Table 3.5 Solubility of NleD at different temperatures and in the presence of different additives. Values for the percent precipitation are calculated from the molar concentrations of soluble NleD under the condition indicated when compared to the value at time 0 hr. Soluble protein content for each sample was calculated from the absorbance readings at 280 nm. Values represent the results from single experiments.

3.6 Mutagenesis of putative active site residues of NleD

Our results from extensive bioinformatic analysis using approaches such as Threading, BLAST, and multiple sequence alignment, suggested a structural similarity between *E. coli* O157:H7 NleD and botulinum neurotoxin. NleD is predicted to fall under the zincins family of the zinc-dependent metalloproteases, due to the presence of a highly conserved HEXXH zinc-binding motif also found in botulinum neurotoxin (Fig. 3.6.1) [58]. This observation facilitated our prediction of putative active site residues in NleD. Accordingly, we have engineered site-specific amino acid replacements at six positions and have characterized the proteolytic activity of the variants. Arg363 and Tyr366, two of the most conserved residues in the neurotoxin and reported to be involved in catalysis[58,61], likely correspond to Arg203 and Tyr206 in NleD. Additionally,

Glu262 is involved in stabilizing the penta coordination state of zinc and Glu350 is involved in stabilizing Arg363 during catalysis in the neurotoxin-catalyzed reaction [61]. Of the three glutamic acid residues of NleD (Glu174, Glu175 and Glu191) two are thought to perform similar functions as Glu262 and Glu350 in neurotoxin. Our bioinformatic analysis also suggested the presence of His149 in NleD's active site based on the sequence alignments; H149 is a highly conserved residue (Fig. 1.3.2) adjacent to the HEXXH (142-146) in the primary sequence. Putative active site residues in NleD were replaced by site-directed mutagenesis to yield the following variants: H149A, E174A, E175A, E191Q, R203K and Y206F. Amino acid substitutions were selected that we felt would least affect the stability and conformation of the native structure of NleD. The effect of mutations on the stability of NleD is discussed in section 3.8 and proteolytic activity of the wild-type and variant NleD proteins is discussed in section 3.10.

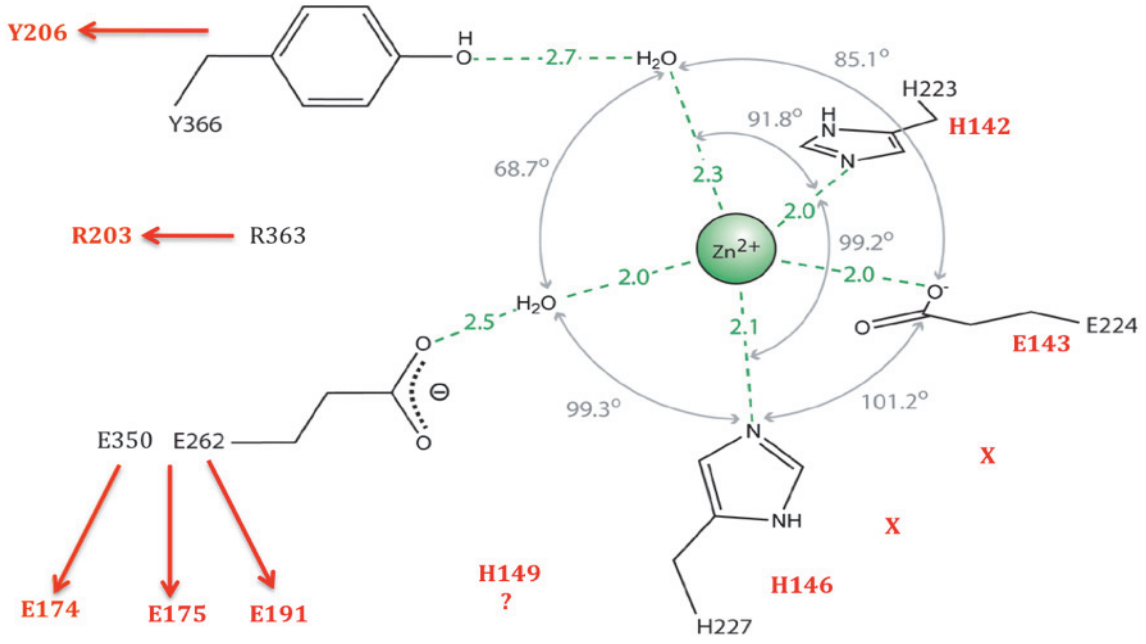


Fig. 3.6 The active site of botulinum neurotoxin [58]. Residues predicted to be in the active site of NleD are shown in red and the corresponding residues in the neurotoxin's active site are shown in black [58]. Note: of the three glutamic acid residues found in NleD (Glu174, Glu175 and Glu191), two are believed to correspond to Glu350 and Glu262 in the active site of botulinum neurotoxin.

3.7 Thermal stability studies of wild-type NleD and the variants using far – UV CD

The effect of amino acid replacements on the thermal stability of the wild-type enzyme was studied using variable temperature far-UV CD. The mean residue ellipticities of proteins were monitored at 222 nm, following the loss of α-helical content as temperature was increased gradually from 25 to 65°C. A common buffer (50 mM Tris (pH 7.2) 100 mM NaCl, 1 mM DTT) was used for all proteins. A single cooperative transition in the unfolding of the secondary structure of wild-type NleD and variants was observed between 37°C to 52°C (Fig. 3.7.1). T_m values (temperature where half of the ellipticity is lost) were calculated using first derivatives of the melting curves. The values are listed in Table 3.7.1 and show that the thermal stability of the variants was very similar to the wild-type protein.

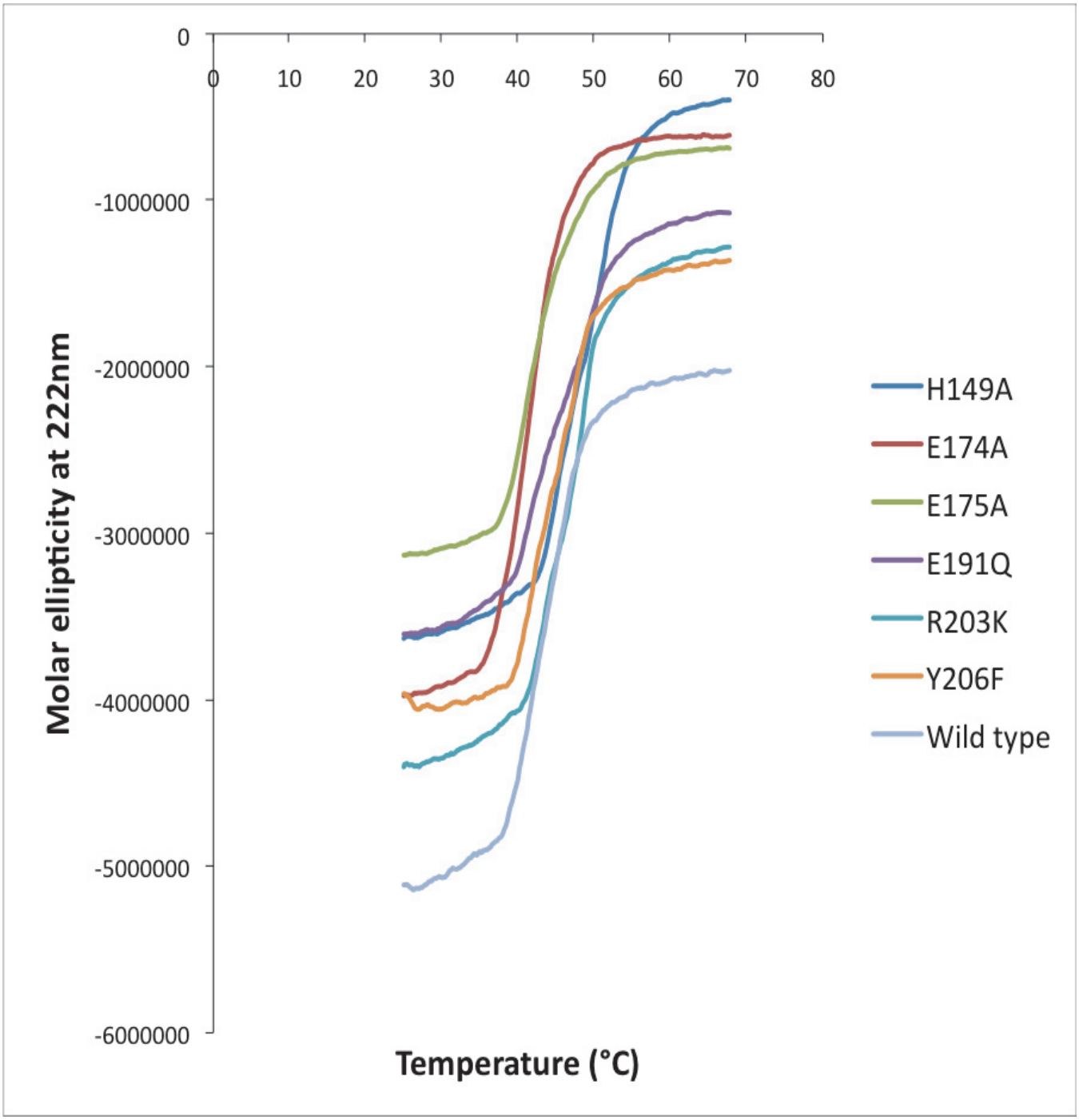


Fig. 3.7 The variation of molar ellipticity at 222 nm as a function of temperature for NleD proteins. (Protein concentrations used ranged from ~ 10-15 μM). Data were recorded in a 2 mm pathlength cuvette at a ramping speed of 20 °C/hr. CD units in millidegrees were converted to molar ellipticity according to the equation given in the legend of Fig. 3.3.1. The T_m value for each curve was predicted based on the inflection point calculated using the first derivative of the curve.

NleD protein	T_m (°C)
H149A	45.6
E174A	41.4
E175A	41.2
E191Q	43.0
R203K	44.0
Y206F	43.8
Wild-type	43.2

Table 3.7 Melting temperatures for wild-type NleD and the variants. All T_m values were calculated as described in the text and represent the results from a single determination for each protein.

3.8 Zinc binding studies using intrinsic tryptophan as a probe

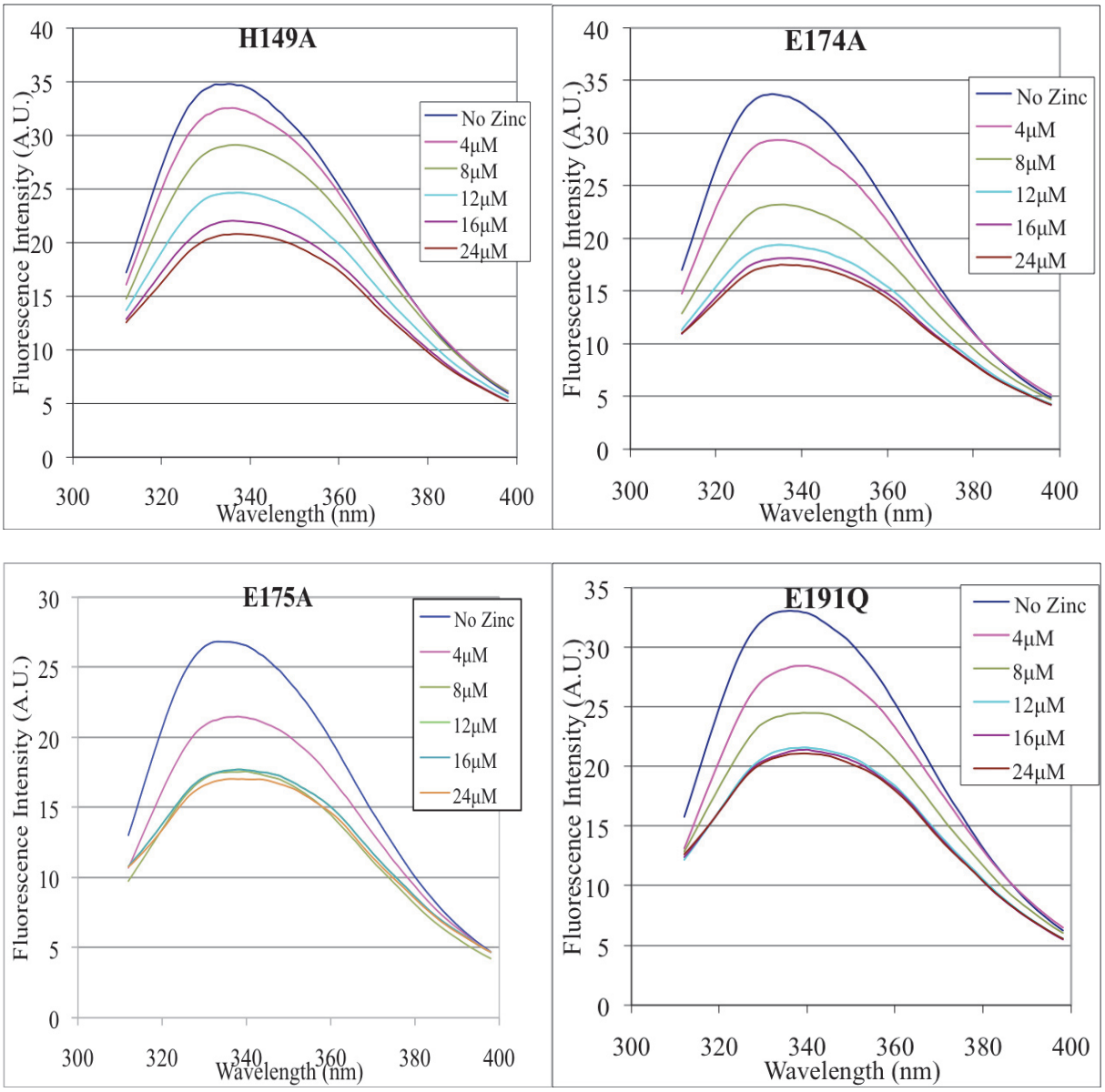
Perturbations in the environment of tryptophan can be easily monitored using steady-state fluorescence spectroscopy, thus, tryptophan fluorescence emission can act as a useful probe of structural changes in a protein [37,75,76]. For example tryptophan emission intensity may be higher if a tryptophan residue is buried versus solvent exposed within the protein structure, or if the residue has shifted away from intrinsic quenching agents such as disulfide bonds and protonated histidines [37]. Additionally the λ_{max} of the emission intensity can be shifted depending upon the polarity of the environment of tryptophan [37]. The primary sequence of NleD contains just one tryptophan residue at position 35. This can likely serve a valuable chromophore for protein ligand binding depending upon the position of tryptophan relative to the location of the active site. [75,76]

The protein fluorescence emission wild-type NleD and the six variant proteins were scanned from 310 nm to 400 nm following excitation at 295 nm. Spectra were also recorded after incubating protein with increasing concentrations of zinc chloride. The results are shown in Fig. 3.8.1. The first derivative of the curves was used to calculate the values for $\lambda_{max.}$ of the emission (Table 3.8.1).

For all proteins in the absence of zinc, chloride $\lambda_{max.}$ of the emission intensity was ~333 nm to 335 nm suggesting that NleD's single tryptophan residue lay in a partially buried environment and this environment was not markedly perturbed by the amino acid replacement. Tryptophan fluorescence intensity decreased with increasing zinc chloride concentrations (Fig. 3.8.1), a trend consistently observed for the wild-type NleD and the variants. Additionally, a prominent red shift in the λ_{max} accompanied quenching of the emission intensity in all the cases. The plots of both the changes in tryptophan fluorescence intensity at 333 nm and in λ_{max} with increasing zinc chloride concentrations are shown in Fig. 3.8.2 and Fig. 3.8.3, respectively. Interestingly, both types of plots exhibited saturation behavior when the molar concentration of zinc chloride reached three times that of the protein. Worth noting, wild-type NleD and all variants yielded comparable results.

A likely explanation for the changes in fluorescence emission intensity could be precipitation of NleD with increasing amounts of zinc chloride, reflecting a decrease in the amount of the chromophore in solution. (In section 3.5 precipitation of NleD in the presence and absence of zinc at different temperatures is described in detail). By comparison, any shifts in λ_{max} emission would be independent of protein concentration. Accordingly, the observed red shift reflects a change in the environment of the protein's single tryptophan from a less polar to a more polar environment, most likely due to a conformational change of NleD upon zinc binding. Surprisingly, the magnitude of the red shifts was similar for wild-type enzyme and all the variants, even though the site-specific replacements were directed to a region where zinc is expected to bind. This finding could be interpreted as follows: 1) as the spectral changes upon the addition of zinc chloride appears similar for all of the proteins, the observed red shift could be due to nonspecific binding of zinc to some other region in NleD distant from the active site but in the vicinity of the single tryptophan residue; 2) the behavior of tryptophan probed by steady-state

fluorescence is independent of zinc binding. Further studies on zinc binding to wild-type NleD and the variants using additional techniques such as isothermal titrating microcalorimetry (ITC), surface plasmon resonance (SPR) or NMR might provide more insight.



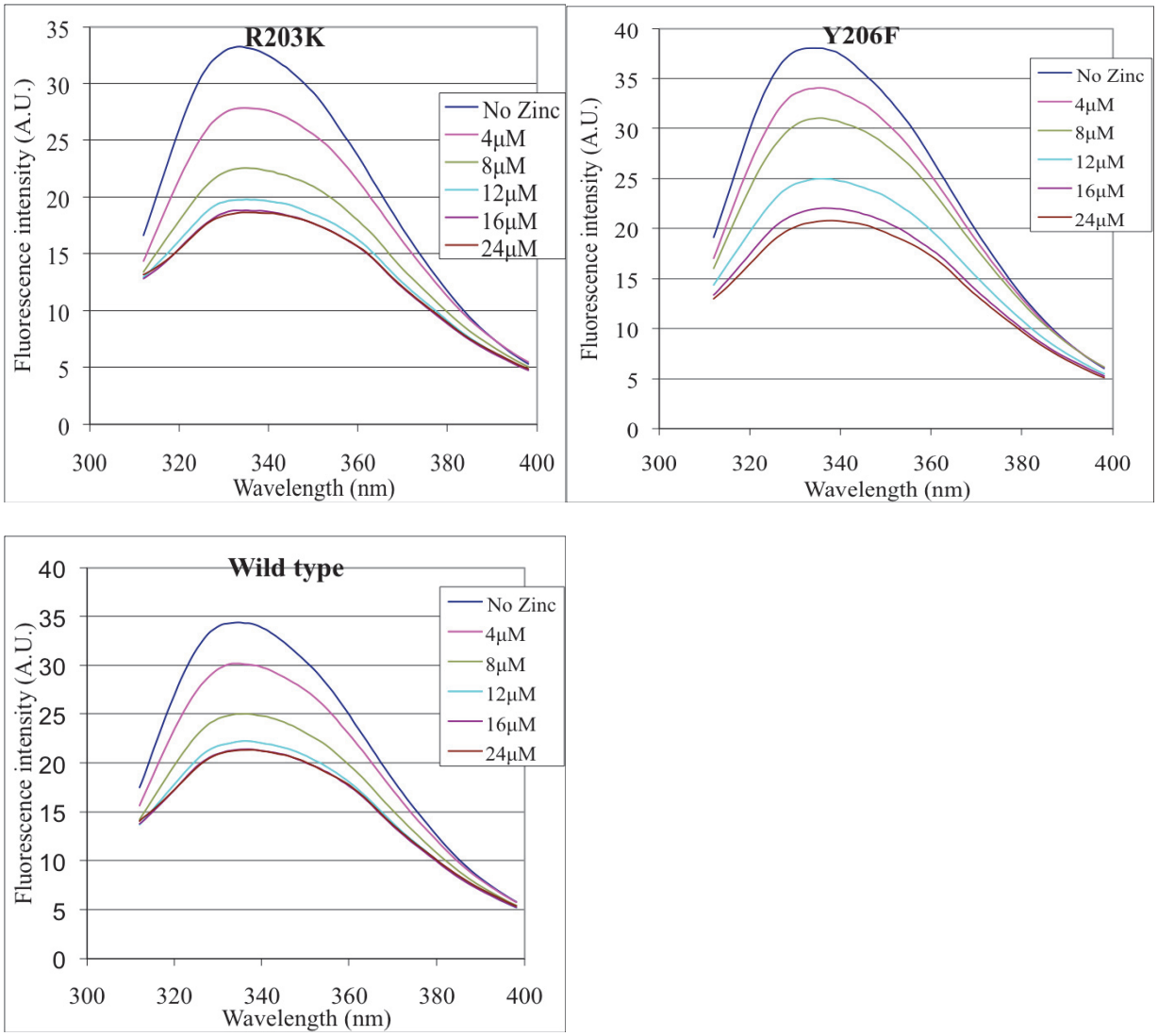
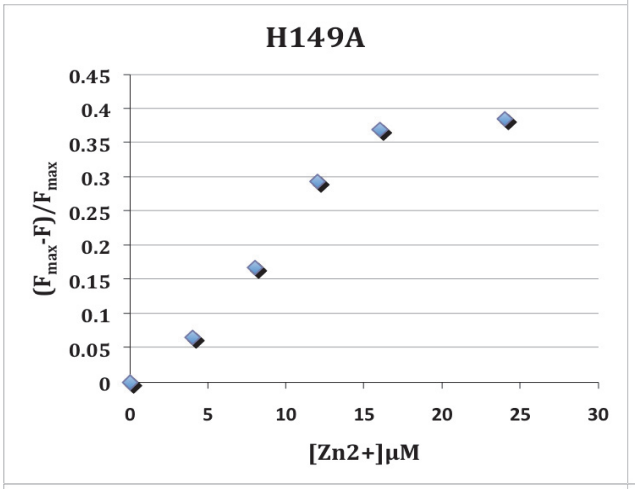
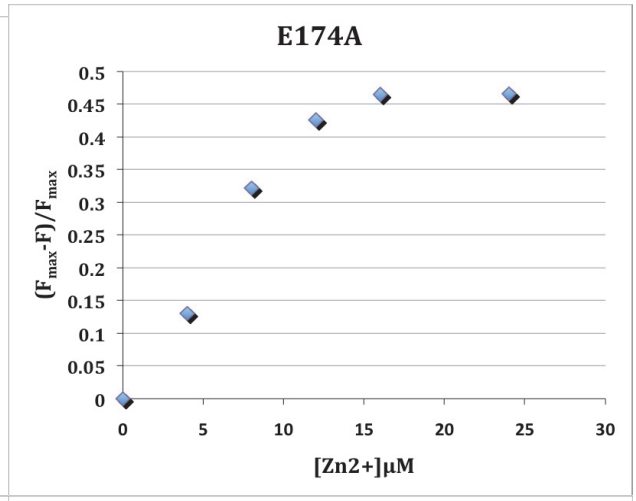
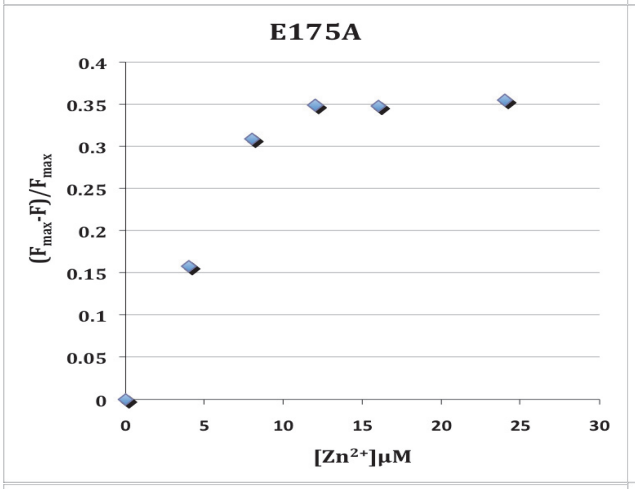
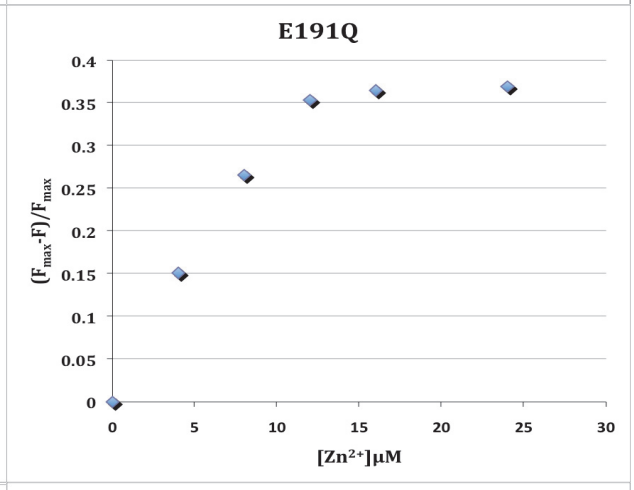


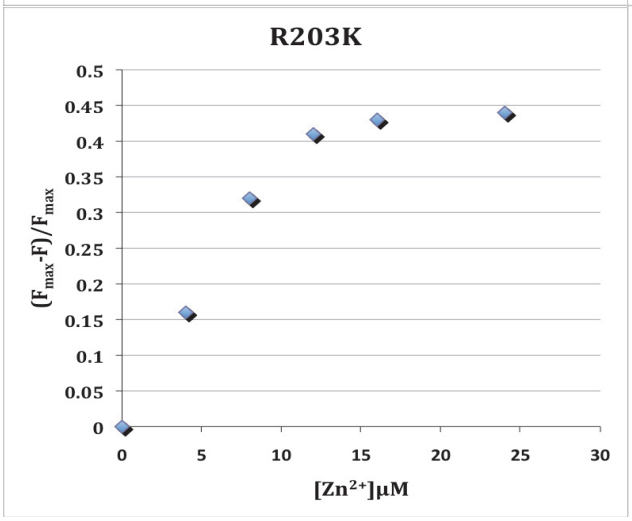
Fig. 3.8.1 Intrinsic fluorescence emission spectra for wild-type NleD and the variants. The scans were recorded at room temperature after excitation at 295 nm and for proteins incubtaed with increasing concentrations of zinc chloride, shown to the right of the emission spectra.

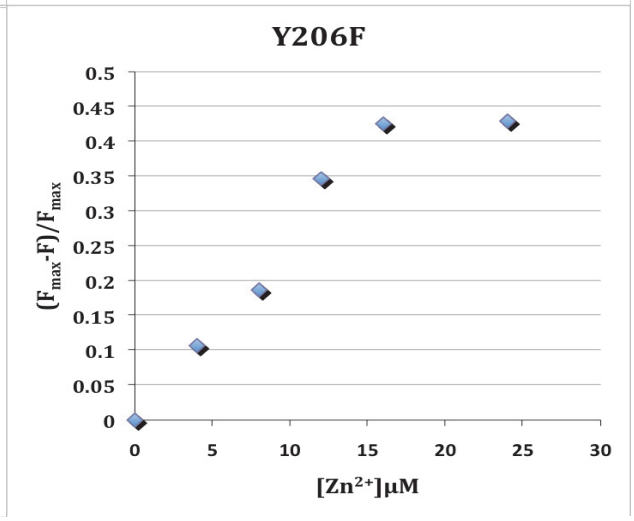

H149A
(Fmax-F)/Fmax
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 5 10 15 20 25 30
[Zn2+]µM


E174A
(Fmax-F)/Fmax
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 5 10 15 20 25 30
[Zn2+]µM


E175A
(Fmax-F)/Fmax
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 5 10 15 20 25 30
[Zn2+]µM


E191Q
(Fmax-F)/Fmax
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 5 10 15 20 25 30
[Zn2+]µM


R203K
(Fmax-F)/Fmax
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 5 10 15 20 25 30
[Zn2+]µM


Y206F
(Fmax-F)/Fmax
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 5 10 15 20 25 30
[Zn2+]µM

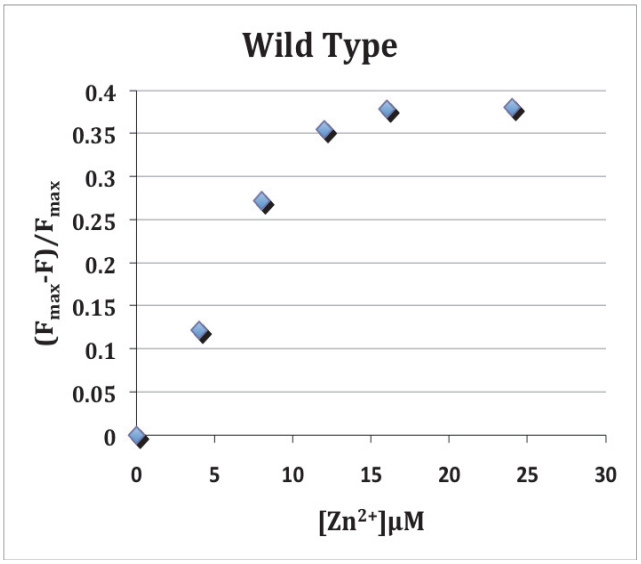
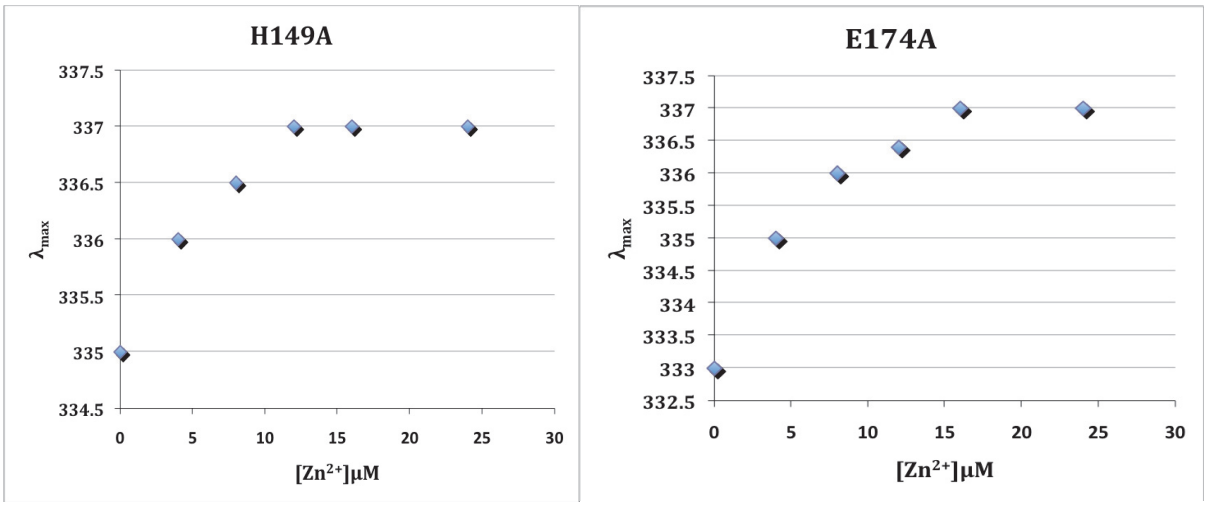


Fig. 3.8.2 Changes in fluorescence intensity plotted as a function of increasing zinc chloride concentration. A change in emission intensity at 333 nm following excitation at 295 nm is reported as $(F_{max}-F)/F_{max}$ where F_{max} is the maximal fluoresence emission intensity in the absence of zinc and F is the intensity observed in the presence of zinc. The curves show saturation behaviour at 16 µM zinc chloride, which is twice the concentration of NleD.



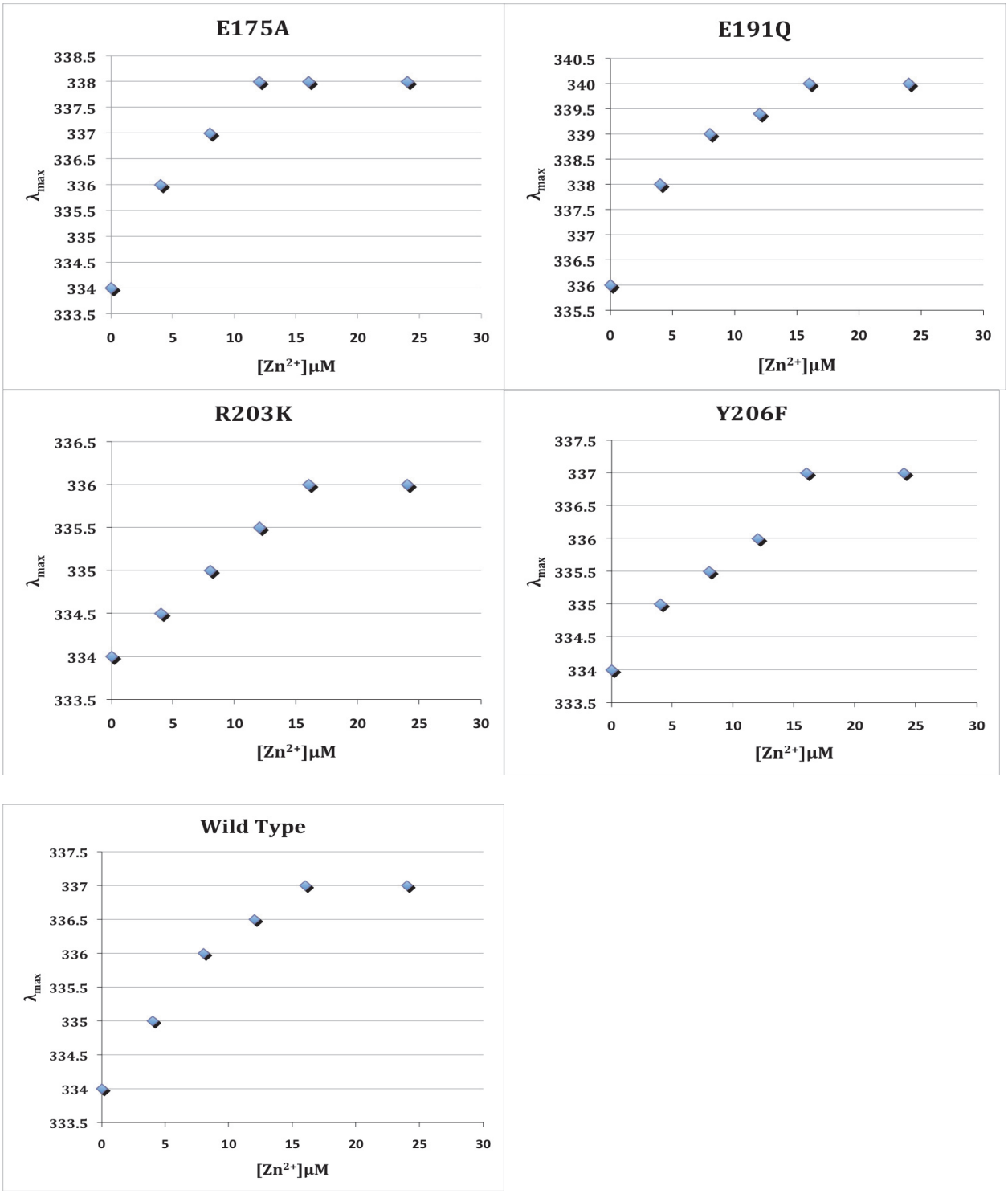


Fig. 3.8.3 A plot of λ_{max} emission as a function of increasing zinc chloride concentration. Changes in λ_{max} emission were reported after excitation at 295 nm and after incubating protein with increasing concentration of zinc chloride. All plots show saturation behaviour as the the zinc chloride concentration reaches 16 μM, which is twice the concentration of NleD.

Protein	$(F_{max}-F)/F_{max}$	λ_{max} shift (nm)	λ_{max} (nm)
H149A	0.38±0.08	2±0.3	335
E174A	0.46±0.1	4±0.4	333
E175A	0.35±0.09	4±0.5	333
E191Q	0.36±0.06	4±0.3	334
R203K	0.44±0.12	2±0.4	333
Y206F	0.42±0.09	3±0.5	334
Wild-type	0.38±0.07	3±0.4	334

Table 3.8 Changes in fluorescence emission intensity and λ_{max} for wild-type NleD and the variants on the addition of zinc. Values were recorded when zinc chloride was present at three-fold the molar concentration of protein. The value represents means ± standard deviation from three different measurements.

3.9 Characterization of histidines in NleD

As the active site of NleD contains histidine residues involved in zinc binding and catalysis [9], studying the behavior of histidines in solution is of interest. NMR was used as an experimental approach [51]. Histidine can act as a valuable probe for NMR studies due to a sharp distinct signal given by the $C^{\delta}H$ and $C^{\varepsilon}H$ hydrogens of the imidazole ring within the histidine side chain [51,52]. This property of histidine makes it easier to detect any change in its NMR signal with alterations in pH due to protonation or deprotonation of the side chain.

The experiments are conducted in 99.9% D_2O, which provides appropriate conditions for the protons on the protein's amide groups to be exchanged by deuterium [51]. Hence, the signal from protons present on histidine side chains can be distinctly observed in a 1D-^{1}H NMR spectrum.

The amino acid sequence of NleD contains a total of fourteen histidines. At pD 6.88 fifteen different individual peaks were observed from 7.5 ppm to 8.6 ppm (Fig. 3.9.1). Every peak obtained is a representation of the average population of protonated and deprotonated states

of protons either from -NH or $C^{\delta}H$ from histidines. With the increase in pD, a gradual shift was observed in the peaks along with line broadening. One interpretation of this result is that the histidine residues giving the signal in 1D-^{1}H spectrum are titrable. Changes in the ppm shift plotted as a function of pD is illustrated in Fig 3.9.2. In order to have a more precise characterization of the peaks observed in the 1D-^{1}H NMR spectra either mutagenesis of every individual histidine in NleD or complete peak assignment of all the histidines present in the protein is necessary. These types of experiments would help to verify if the peaks observed in the 1D-^{1}H NMR spectrum (from 7.5 ppm to 8.6 ppm) belong to histidine residues, specifically to the two histidines within the HEXXH motif present in the active site of NleD. Following the pH-dependence of the behavior of these important histidine residues would allow the estimation of their pKa values.

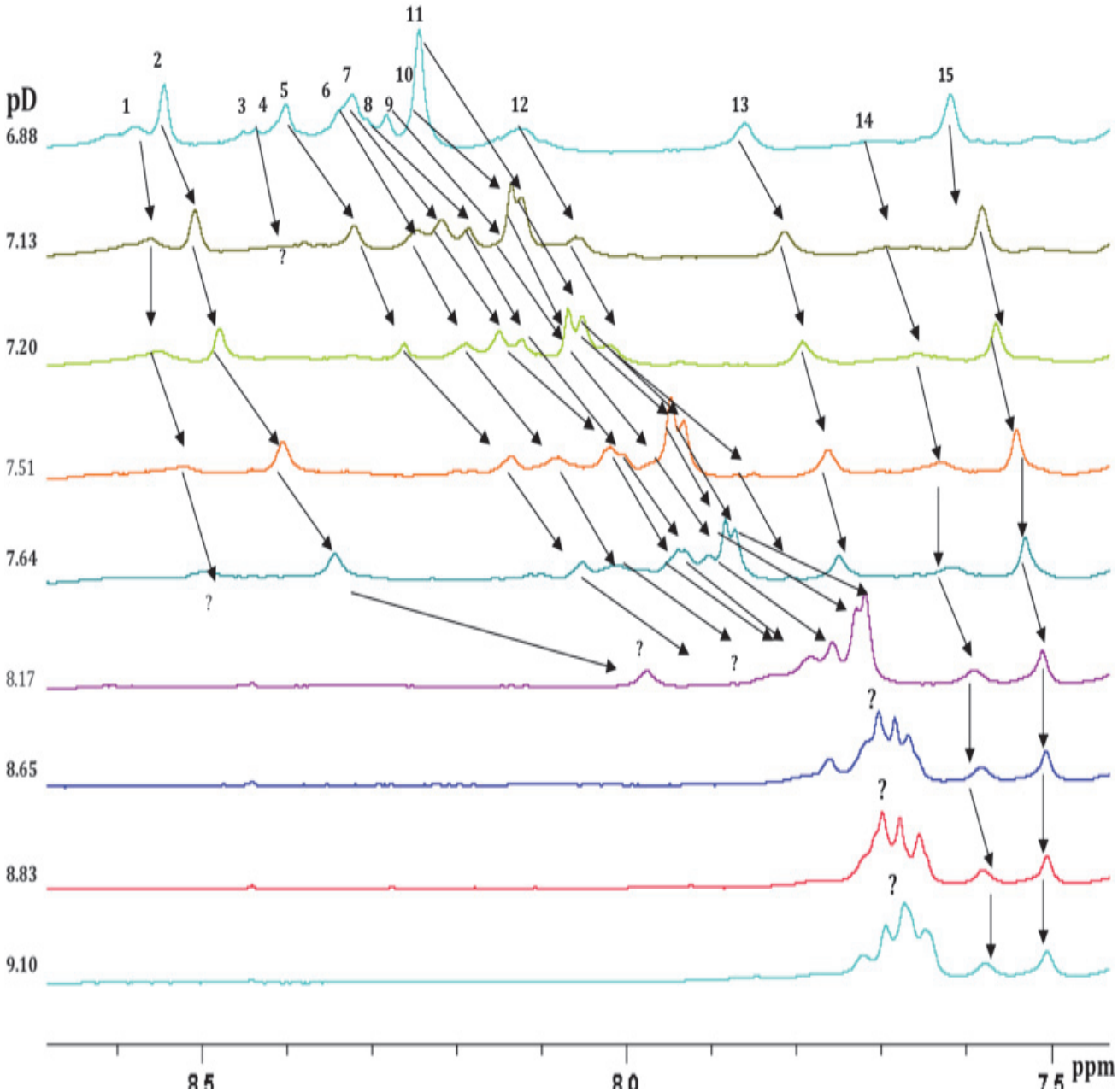


Fig. 3.9.1 1D-^{1}H NMR spectra of wild-type NleD. Spectra were collected at room temperature between the pD 6.88 to pD 9.10 at increments of ~ pD 0.2 units. The operational pH of the buffer solution in 99.9% D_2O (measured with a glass electrode) is converted to pD value by adding 0.4 units [79].

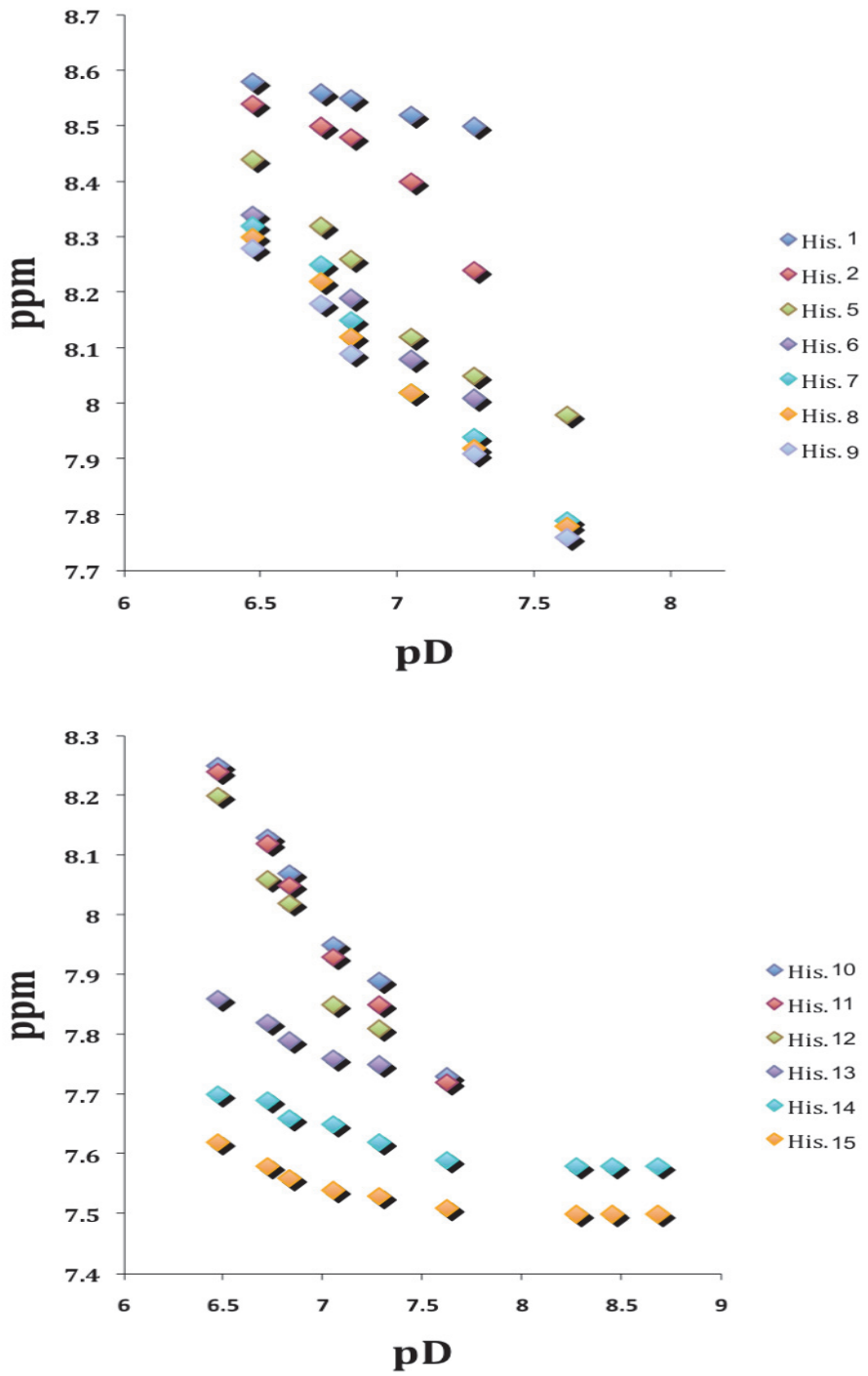


Fig. 3.9.2 Changes in the chemical shift (ppm) as a function of increasing pD. All the peaks observed had a gradual change in their ppm values with the increasing pD. (Note: each peak is arbitrarily assigned as a histidine residue).

3.10 Characterization of the proteolytic activity of wild-type NleD and variants

Based on bioinformatic analysis, residues His149, Glu174, Glu175, Glu191, Arg203 and Tyr206 were predicted to be near or within the active site of NleD site (section 3.7). Accordingly, variant proteins were characterized carrying amino acid substitutions at these positions; these included H149A, E174A, E175A, E191Q, R203K, Y206F. If these residues were indeed active site residues then the proteolytic activity of NleD should be affected by the amino acid replacements.

SDS polyacrylamide gel-based assays were used to assess the activity of the wild-type enzyme and its variants. The substrate used for the proteolysis assay was p38α kinase. This protein substrate was recombinantly expressed in *E. coli* to carry an N-terminal hexa-histidine tag, and purified to near homogeneity using Ni-NTA affinity chromatography as described in section 2.10. An SDS polyacrylamide gel of the purification of p38α kinase is shown in Fig. 3.10.1.

To conduct the assay, NleD protein (protease catalyst) and hexa-His p38α kinase (substrate) were mixed in equimolar concentrations in the presence of zinc chloride and the mixtures were incubated at 4°C for up to10 days.

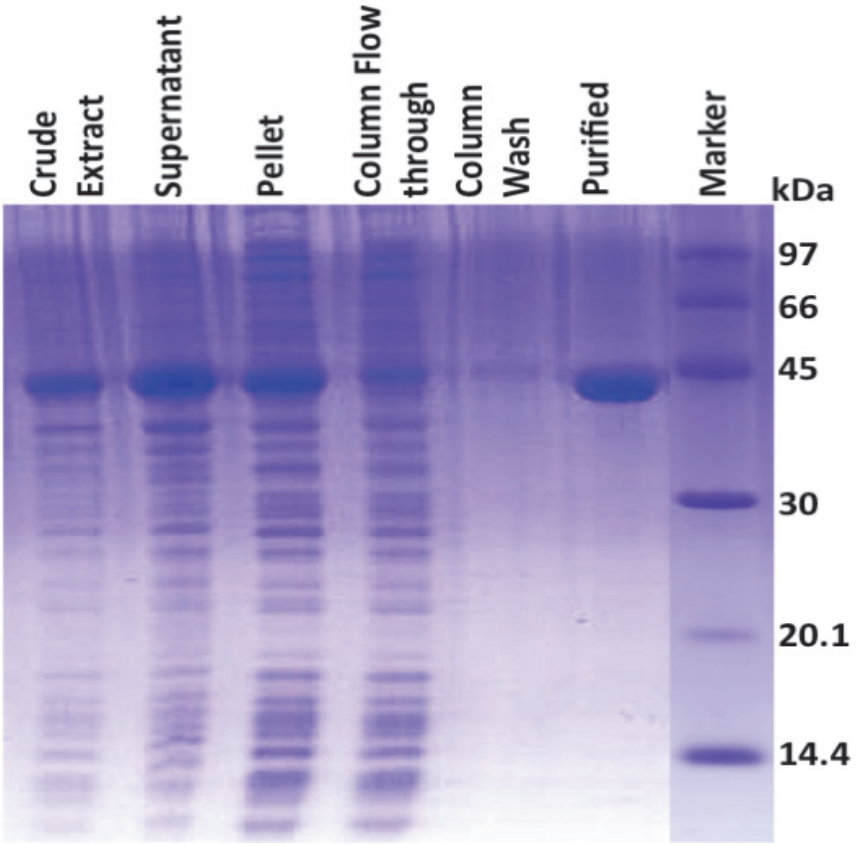
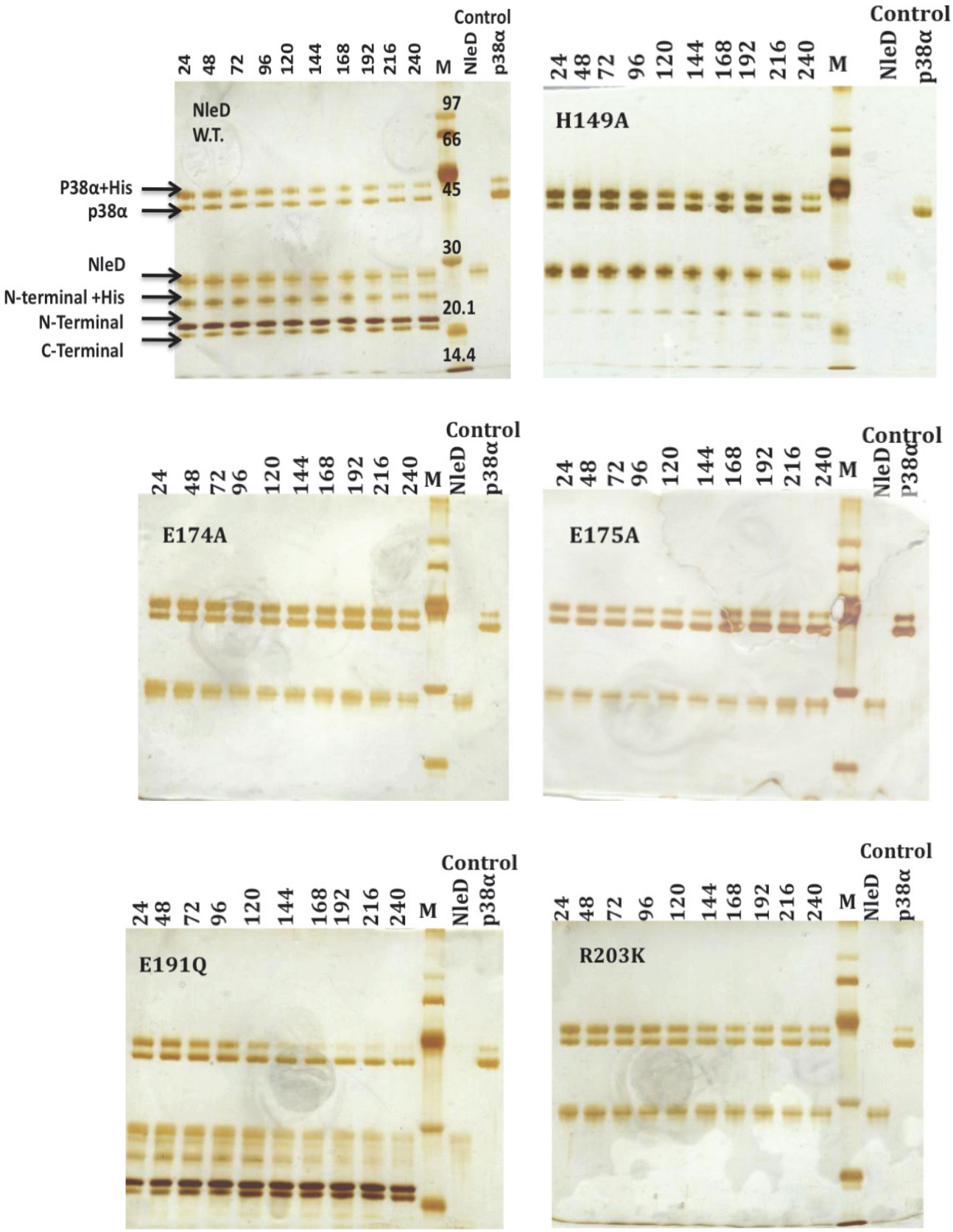


Fig. 3.10.1 SDS-PAGE analysis of the purification of recombinant hexa-His tagged p38α kinase. His-tagged p38α kinase was purified using nickel-affinity chromatography; fractions at each step of the procedure were analyzed. "Purified" refers to protein eluted after washing the column with 250 mM imidazole. A 12% acrylamide gel was used in the analysis. Proteins bands were visualized by staining with Coomassie brilliant blue.


24 48 72 96 120 144 168 192 216 240 M NleD p38α Control
NleD
W.T.
97
66
45
30
20.1
14.4
P38α+His
p38α
NleD
N-terminal +His
N-Terminal
C-Terminal
H149A
E174A
E175A
E191Q
R203K

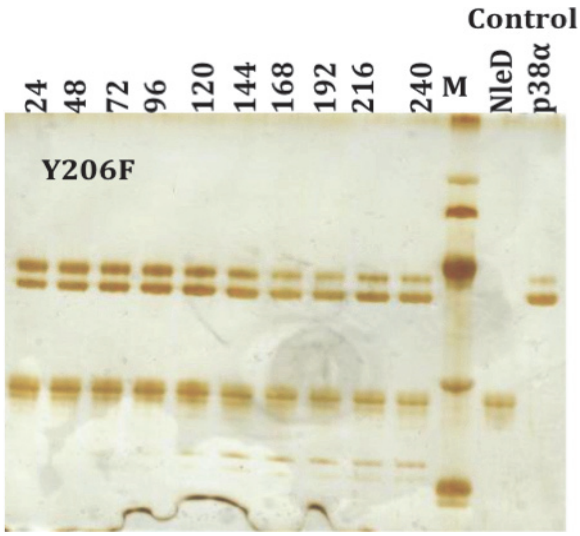


Fig. 3.10.2 SDS-PAGE-based enzyme assays for wild-type NleD and the variants. Assays were conducted over a time period of 10 days of 4 °C. All the enzyme assays were performed in parallel under same buffer conditions (50 mM Tris (pH 7.2), 100 mM NaCl, and 1 mM DTT) and equimolar concentrations of zinc. Specific times (in h) at which the reactions were monitored are shown on top of the gels. M refers to the protein molecular weight makers. The controls include purified NleD and hexa-His tagged p38α kinase. Protein bands were visualized by silver staining.

NleD is reported to cleave at a single site in p38α kinase -- before the tyrosine in the TXY motif present in the 22 amino acid activation loop of the kinase. Two different cleavage products were expected: an N-terminal peptide fragment of 23.4 kDa and a C-terminal fragment of 20.4 kDa, their masses determined from the primary amino acid sequence of hexa-His p38α kinase. It is worth noting that without a hex-His tag, the N-terminal fragment would have yielded an expected mass of 20.9 kDa, presumably too similar in mass to the C-terminal fragment for effective resolution of the products by SDS-PAGE. Hence, the p38α kinase used as a substrate for the NleD-catalyzed reaction, possessed an intact N-terminal hexa-His tag.

The extent of cleavage was monitored by SDS-PAGE analysis and the results of the assays for wild-type enzyme and each of the variants is shown in Fig. 3.10.1. Unexpectedly, a complex hydrolytic profile was obtained (shown in the first panel for wild-type enzyme). The

two predicted cleavage products were obtained, one at ~ 23 kDa (denoted N-terminal + His) and the other at ~ 20 kDa (denoted C-terminal), along with the bands corresponding to the catalyst NleD (~ 26 kDa) and the substrate His-tagged p38α kinase (~ 43.9kDa). Additional cleavage products were also observed, likely corresponding to p38α kinase (~ 41.1kDa) and the N-terminal product (~ 21 kDa) which are both missing the hexa-His tag. To verify if these additional bands were due to non-specific cleavage of the hexa-His tags, the experiments were repeated using p38α kinase in which the affinity tag had been removed. As shown in Fig. 3.10.3 and as expected, the analysis yielded two bands corresponding to the two cleavage products, as well as a band for the catalyst and the substrate.

To further verify the identity of the cleavage products, the assay for wild-type enzyme was repeated to include more data points at shorter incubation times (Fig. 3.10.3) and band intensities were analyzed using software image J (http://rsbweb.nih.gov/ij/download.html). The amount of substrate (NleD), catalyst (p38α kinase) and the cleavage products (the sum of the N- and C- terminal fragments of p38α kinase) in the reaction mixture were calculated from the band intensities using image J software and plotted as a function of time (Fig. 3.10.4).

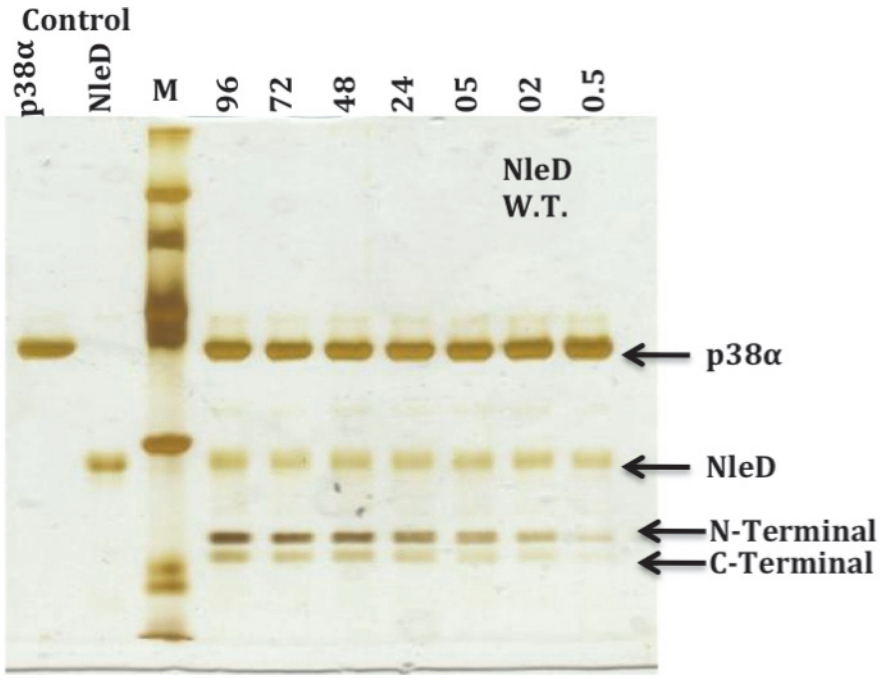


Fig. 3.10.3 SDS-PAGE based enzyme assay for wild-type NleD using as a substrate p38α kinase without its hex-His tag. M refers to the protein molecular weight makers. The controls include purified NleD, and p38α kinase with its hexa-His tag cleaved off. Protein bands were visualized by silver staining

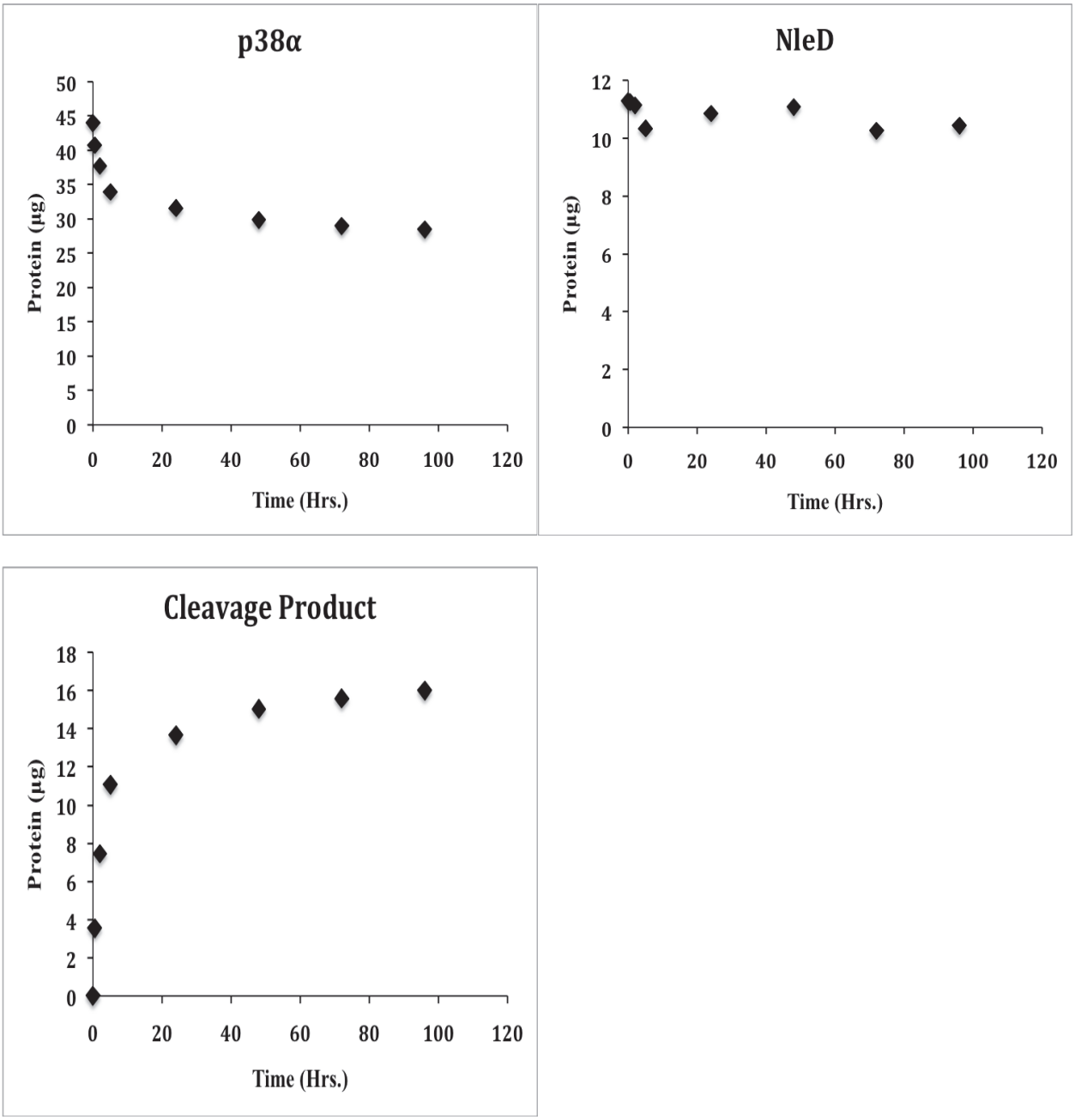


Fig. 3.10.4 Quantification of the NleD-catalyzed cleavage reaction of p38α kinase. Graphical representation of the changes in the amount of substrate (p38α kinase) hydrolyzed by enzyme (NleD) and the cleavage products (obtained from combining the integrated band intensities of both the N-terminal and C-terminal fragments of p38α kinase) as a function of incubation time. The amount of protein was calculated from the integrated intensities of the relevent bands from the silver stained gel using image J.

As shown in Fig. 3.10.4, the intensity of the band coresponding to untagged p38α kinase gradually decreased concommitant with an increase in the bands corresponding to the cleavage products. The band intensity assigned to NleD did not appear to decrease over the course of reaction. Hence, the results suggest that the lower bands observed in the denaturing gel were cleavage products from p38α kinase.

The activities of the variants were compared to wild-type NleD (Fig. 3.10.1). The variants H149A and Y206F showed markedly reduced protease activity compared to wild-type enzyme as reflected by the appearance of extremely faint bands corresponding to the cleavage products, over the entire time period of the enzymatic reaction (the band corresponding to the N-terminal fragment is the most visible). By comparison, E174A, E175A and R203K showed no activity even after an incubation period of 10 days (no bands observed corresponding to the expected cleavage products) The results showed that the afore-mentioned amino acids are likely involved in catalysis. One of the variants, E191Q, was found to be as active as the wild-type enzyme as all expected cleavage products are clearly resolved. In summary, the results suggest that residues Glu174, Glu175, Arg203, His149 and Tyr203 are likely present in NleD's active site.

Chapter 4: Discussion

The current study is the first reported biophysical characterization of recombinantly expressed NleD from the pathogenic *E. coli* strain O157:H7. We have used a battery of biophysical tools to obtain information regarding the structural features of the protein in solution as well as its stability. These include: mass spectrometry, size exclusion chromatography, dynamic light scattering and analytical ultracentrifugation, NMR, steady-state fluorescence and circular dichroism spectroscopies and extensive bioinformatic analysis. We have also utilized an *in vitro* functional assay, based on the ability of NleD to hydrolyze a protein substrate p38α kinase, to assess the importance of putative active site residues in the catalytic mechanism of the NleD. To facilitate these studies NleD and six variant proteins carrying site-specific replacements directed against putative active site residues were recombinantly expressed and purified in addition to the substrate (p38α kinase) for our functional assay. Finally, we also identified conditions that help stabilize NleD in solution. Our findings may augment and facilitate high-resolution structural studies on this protein by X-ray crystallography and NMR spectroscopy in the future.

4.1 Characterization of the oligomeric state of NleD

Techniques that exploit the hydrodynamic properties of a macromolecule can be used to help estimate its molecular mass, which in turn, can provide clues as to its quaternary structure in solution. Three different hydrodynamic methods used for the analysis of NleD were size exclusion chromatography (SEC), dynamic light scattering (DLS) and analytical ultracentrifugation (AUC). The molecular weight from SEC was predicted based on the elution volume of the purified NleD, which is directly proportional to the average hydrodynamic volume

occupied by the protein molecule in solution [65]. Interestingly, our results from SEC analysis indicated that NleD, either the apo or zinc-bound form, eluted earlier than expected for a monomer but later than expected for a dimer (Figs. 3.4.2). Thus, if NleD is monomeric in solution under our experimental conditions, then its elution behavior suggests that NleD occupies a larger hydrodynamic volume than expected for a perfectly spherical protein. This might occur if the protein is partially unfolded under the experimental conditions or if it possesses an elongated rod-like structure, which increases its hydrodynamic volume [63,64,65].

The results are also consistent with the presence of a rapid and dynamic equilibrium between the monomeric and dimeric protein species. If this were the case, then the equilibrium should be dependent on protein concentration. SEC experiments conducted over a 100-fold range of protein concentrations, however, did not result in a shift of the elution profile of the protein (Fig. 3.4.2) and thus did not support this hypothesis.

A second tool, dynamic light scattering (DLS) was used to verify the molecular weight of the protein under the same buffer conditions, the same temperature (room temperature) and over a similar range in protein concentration as used in SEC. DLS is reported to be a technique complementary to SEC [43,66,67,68]. In all the cases the molecular weight predicted from DLS (~ 40kDa) and based on hydrodynamic radius measurements, was similar to the molecular weight obtained from SEC (~ 38kDa) (Fig. 3.4.3). This suggested, most likely, that under the experimental conditions used, NleD was partially unfolded or possessed an elongated rod-like or ellipsoidal shape.

The idea that native NleD is not perfectly spherical in shape at room temperature was supported by the findings from ESI-MS analysis, which showed that the C-terminal region of the polypeptide contained a considerable amount of random structure and/or was loosely folded, and was prone to proteolytic cleavage (Fig. 3.2.3). Additionally, the results obtained by monitoring

the degree of protein precipitation as a function of temperature (Fig. 3.5.1) suggested that NleD was less thermally stable at room temperature. If NleD were partially unfolded at room temperature as a result of a lower thermal stability, then its hydrodynamic properties would be affected, supporting the measurements from SEC and DLS.

AUC using sedimentation velocity analysis was performed which measures the velocity of the macromolecular solute in a specific solvent under the applied centrifugal field [69,70]. These measurements can also be used to predict the molar mass and hydrodynamic shape of the protein species. Experiments performed at temperatures significantly lower than ambient temperature, *i.e.* 15°C and 4°C (Fig. 3.4.4) revealed that the majority of protein species sedimented as a monomer with a molecular weight 25.5 kDa. Additionally, NleD was likely globular in nature based on the frictional ratios. Interestingly, a small additional peak was observed when experiments were performed at 15°C but not 4°C, This species possesses a higher sedimentation coefficient, in keeping with a more loosely folded form of the protein such as a small fraction of partially unfolded protein species which might accumulate at 15°C during the overnight centrifugation. Together these finds suggested that NleD is monomeric in solution and the majority of the protein retains its globular-like structure at lower temperatures (4°C). At higher temperatures, however, NleD may be more loosely folded and/or partially unfolded and prone to precipitation. We further speculate that maybe at the higher temps, the folded and partially folded forms are in dynamic equilibrium

4.2 Effect of temperature and trehalose on NleD stabilization

The results from several independent experiments (Fig. 3.8.1 of fluorescence emission, Table 3.5.1) point to the fact that NleD is prone to precipitation at room temperature and this precipitation is exacerbated in the presence of zinc chloride. Thus, preliminary studies were

undertaken in this thesis to determine conditions that could help stabilize NleD at room temperature, a temperature that more easily facilitates structural studies using NMR.

Many additives have been examined for their ability to stabilize proteins *in vitro*[71]. They help prevent protein unfolding and aggregation by stabilizing the folded state of the protein and/or by inhibiting protein-protein interactions [53]. Some of these commonly used additives are polyols, osmolytes, polyamine and salts [72,73,74]. In addition to additives, reducing experimental temperatures has also been reported to stabilize more thermally labile proteins *in vitro* [55,56,57]. In this thesis the stability of NleD is characterized in the presence of three of the more commonly used protein stabilizers, specifically the osmolytes trehalose and sucrose, and the salt ammonium sulfate. Stability studies were also performed at different temperatures as we speculated that NleD might be thermally labile based on the results obtained from AUC experiments.

Our results clearly indicate that the time-dependent precipitation of the protein, either in the presence or absence of zinc chloride was very dependent on temperature and that NleD was the most stable at 4 °C (Fig 3.7). Interestingly, the melting temperature (T_m) of NleD of 42 °C, as deduced though variable temperature far-UV CD experiments, suggested that partially unfolding could occur at temperatures much lower than the T_m. Recording CD spectra from 190-260 nm at different temperatures could help resolve this issue.

NleD's stability is temperature dependent. Thus, different additives were tested for their ability to improve the protein's stability at the higher temperatures. Of those additives tested in the presence of zinc, trehalose appeared to be the most promising; the degree of protein precipitation (at 32%) observed in these experiments performed at 15 °C in the presence of 1M trehalose was markedly reduced (from 83%) observed in the absence of trehalose at the same temperature (Table 3.5.1). In fact the apparent stability of NleD at 15 °C in the presence of

trehalose was the same as at 4 °C in the absence of trehalose (compare values of percent precipitation of 32% and 28% in Table 3.5.1). Unfortunately trehalose did not significantly reduce NleD's precipitation at room temperature in the presence of zinc chloride (data not shown).

It is well documented that trehalose is an effective protein stabilizer [73]. In fact, this sugar is a much better stabilizer than other disaccharides such as sucrose and maltose. Trehalose possesses the highest number of polymorphs in both crystalline and amorphous states, a property not exhibited by any other dissacharide [54]. Trehalose can undergo a reversible transition between its polymorphic forms without losing the integrity of their crystalline structure, enabling trehalose to entrap water molecules within its glassy matrix and act as a potent protein stabilizer [54]. Trehalose is also a kosmotroph and can form stronger interactions with water molecules better than water can with itself, hence lowering the freezing temperature [54]. Accordingly, trehalose is also an efficient cryoprotactant.

In summary, our findings identify the importance of temperature in stabilizing NleD, but also suggest this effect might be enhanced in the presence of trehalose. Further studies should include a more extensive characterization of NleD at higher temperatures using different concentrations of trehalose alone and in combination with other additives.

4.3 Characterization of active site residues in NleD

Baruch and colleagues had previously reported the cleavage of JNK and p38 kinase by NleD *in vivo* [9]. They postulated the presence of a single cleavage site present in the activation loop of these kinases [9]. In the present study, the proteolytic activity of NleD was also monitored using p38α kinase as a substrate for the cleavage reaction. The functional assay was extended,

however, to include putative active site variants, as steps towards identifying the importance of specific amino acid residues in the catalytic mechanism of NleD. This will be discussed later.

SDS-PAGE-based enzyme assays were used to study the proteolytic digestion of p38α kinase by NleD. Due to the reduced stability of NleD at room temperature (section 3.5), the enzyme assays were conducted at 4°C. It was speculated that at this low temperature, NleD would exhibit markedly reduced proteolytic activity. Accordingly, the assays were conducted over a time period of 10 days to ensure that the proteolytic activity of any less active variants could be detected. Due to the presence of a single cleavage site in p38α kinase two bands as cleavage products (23.4 kDa and 20.4 kDa) were expected by electrophoretic analysis of the reaction. Instead, however, a more complex pattern involving an additional (third) band (~ 21 kDa) was observed (Fig. 3.10.2). We speculate that this stems from uncontrolled proteolysis of the N-terminal hexa-His-tag of p38α kinase (Section 3.10) rather than from the degradation of NleD, even though mass spectrometry analysis shows that NleD is susceptible to proteolysis at its C-terminal end (Section 3.2). Our hunch was confirmed by two independent experiments: 1) performing the time-dependent assay using a different preparation of p38α kinase that lacked the N-terminal hexa-His tag (only two bands were observed, Fig. 3.10.3) and 2) by quantifying the amount of reactants and products present during the course of the reaction using image J (the cleavage of p38α kinase correlated well with the increase in expected products, while the levels of full-length NleD remained constant, Fig. 3.10.4). Thus, the two bands observed between 20 - 24 kDa were likely the cleavage products of p38α kinase.

Extensive bioinformatic analysis (section 3.6) led to the hypothesis suggesting a structural similarity between the active site of botulinum neurotoxin and NleD. Six putative active site residues were identified in NleD (Fig. 3.6.1), then six variants were prepared carrying a single amino acid replacement. The results from variable temperature far-UV CD measurements (Table

3.7.1) and tryptophan fluorescence emission measurements (Fig. 3.8.1) suggested that the amino acid replacements did not significantly alter the thermal stability of NleD, nor its global tertiary structure. Comparison of the wild-type and variant proteins using a functional assay of the proteolytic activity of NleD, however, identified five of the six targeted residues (His149, Glu174, Glu175, Arg203 and Tyr206) likely as active site residues important for the catalytic activity of the enzyme. As shown in Fig. 3.10.2, the proteolytic activities of E174A, E175A and R203K were abolished by the substitutions while the activities of H149A and Y206F were greatly reduced. By contrast E191Q was as active as wild type enzyme and accordingly this residue was considered not important for catalysis. These preliminary experiments support our hypothesis that NleD is a thermolysin-like protease with an active site geometry similar to botulinum neurotoxin (Fig 3.6.1). In order to probe further the catalytic mechanism of NleD and to determine if NleD's mechanism of substrate hydrolysis is identical to that of neurotoxin, the kinetic parameters k_{cat} and K_m for the reactions catalyzed by wild-type enzyme and the variants must be determined, and these values compared to those reported in the literature for neurotoxin [61,62] Additionally, an NMR or crystal structure of NleD would be an invaluable tool for these mechanistic studies.

Chapter 5 – Summary and Future Directions

The aims of this thesis was to express and purify recombinant NleD from pathogenic *E. coli* O157:H7, to characterize various biophysical properties of this protein and to identify residues which may be present in its active site. The work represents the first biophysical findings reported for NleD.

NleD is likely a monomer *in vitro*, based on the aggregate findings by size exclusion chromatography, dynamic light scattering and analytical ultracentrifugation. Purified NleD is susceptible to proteolytic cleavage at its C-terminal region as determined by mass spectrometry. The protein also exhibited reduced stability at room temperature, particularly in the presence of zinc chloride, but its stability could be enhanced in the presence of trehalose, and at lower temperatures. These findings combined with results from mass spectrometry provide clues as to conditions that may help stabilize NleD for crystallization. Bioinformatic analysis predicted a structural similarity between the active sites of botulinum neurotoxin and NleD. Preliminary characterization of the proteolytic activity of wild-type NleD and six variant forms carrying single-site amino acid replacements was performed using p38α kinase as a substrate. The results of our functional assay show that His149, Glu174, Glu175, Arg203 and Tyr206 are important for catalysis and are likely active site residues. This thesis reports the first studies directed towards the identification of active site residues of NleD,

Future efforts should focus on obtaining high-resolution structural information of NleD. Conditions identified in this thesis to stabilize NleD may help in obtaining diffraction quality crystals. Based on the data from mass spectrometry, a shorter construct (residues 2-213), which carries C-terminal deletions should also be used for crystal screens. As putative active site residues have been identified, it might be possible to select a specific variant that will allow the

interaction of NleD and p38α kinase but prevent the proteolytic cleavage of the substrate. If the appropriate conditions are identified to promote a stable complex between an inactive NleD variant and p38α kinase, this could facilitate crystallization of the complex between the protease and its substrate.

Additionally, a shorter NleD construct (residues 2-213) may prove to be valuable for structural studies using the high-resolution solution technique of NMR. NMR could also be used to study the binding between p38α kinase and active site variants of NleD (2-213).

References

1) Kapler, J.B., Nataro, J.P. and Mobley, H.L.T. Pathogenic *Escherichia coli.* *Nature Reviews Microbiology* (2004) 2: 123-140

2) Bhavasar, A.P., Guttman, J.A., and Finlay, B. Manipulation of host cell pathways by bacterial pathogens. *Nature Reviews* (2007) 449: 827-834

3) Croxen, M.A. and Finaly, B. Molecular mechanisms of *Escherichia coli* Pathogencity. *Nature Reviews* (2010) 8: 26-38

4) Tobe,T., Beatson, S.A., Taniguchi, H., Abe, H., Bailey, C.M., Fivian, A., Younis R., Matthews, S., Marches, O., Frankel, G., Hayashi, T. and Pallen, M.J. An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *PNAS* (2006) 103: 14941-14946

5) Dean, P. and Kenny, B.The effector repertoire of enteropathogenic *E. coli*: ganging up *on* the host cell. *Curr. Opin. Micro*. (2009) 12: 101–109.

6) Tree, J.J., Wolfson,E.B., Wang, D., Roe, A.J. and Gally, D.L. Controlling injection: regulation of type III secretion in enterohaemorrhagic *Escherichia coli*. *Trends Micro.* (2009) 17: 361-70.

7) Alexander, R. C. W., Jaclyn, S. P., Michael, D. B., Munera, D., Robinson, K.S., Lee,S.F., Frankel, G. and Hartland, E.L. Enteropathogenic and enterohaemorrhagic *Escherichia coli*: even more subversive elements. *Mol. Micro.* (2011) 80: 1420–143

8) Lim, J.Y.J, Yoon, J.J. and Hovde, C.J.C.J A Brief Overview of *Escherichia coli* O157:H7 and its Plasmid O157. *J. Micro. Biotechnol*. (2010) 20: 1–10

9) Baruch, K., Arie, L.A., Nadler, C., Koby, S., Yerushalmi, G., Neriah, Y.B., Yogev, O., Shaulian, E., Guttman, C., Zarivach, R. and Rosenshine, I. Metalloprotease type III effectors that specifically cleave JNK and NF-κB. *EMBO J.* (2011) 30: 221 – 23

10) Agin, T.S. and Wolf, M.K. Identification of a family of intimins common to *Escherichia coli* causing attaching and effacing lesions in rabbits, humans and swine. *Infect. Immun.* (1997) 65: 320-326

11) Marches, O., Wiles, S., Dziva, F., La Ragione, R.M., Schuller, S., Best, A., Phillips, A.D., Hartland, E.L., Woodward, M.J., Stevens, M.P., Frankel, G. Characterization of two non-locus of enterocyte effacement-encoded type III- translocated effectors, NleC and NleD, in attaching and effacing pathogens. *Infect. Immun.* (2005) 73: 8411-8417

12) Shaulian, E., Karim, M. AP-1 in cell proliferation and survival. *Oncogene* (2001) 20: 2390- 2400

13) Minden, A. and Karin, M. Regulation and function of the JNK subgroup of MAP kinases. *Biochim. Biophys. Acta.* (1997) 1333: F85–F104

14) Bergami, P.L., Lau, E. and Ronai, Z. Emerging roles of ATF2 and the dynamic AP1 network in cancer. *Nat. Rev.,Can.* (2010) 10: 65-76

15) Sudgen, P.H. and Clerk, A. Stress responsive mitogen-activated protein kinase (c-Jun N-terminal kinase and p38 mitogen activated protein kinases) in the myocardium. *Circ. Res.* (1998) 83: 345-352

16) Pearson, G., Robinson, F., Gibson, T.B., Xu, B., Karandikar, M., Berman, K. and Cobb, M.H. Mitogen-activated protein kinase pathways: regulation and physiological functions. *Endocr. Rev.* (2001) 22: 153-158

17) Hooper, N.M. Families of zinc metalloproteases. *FEBS Lett*. (1994) 34: 1-6

18) Jongeneel, C.V., Bouvier, J. and Bairoch, A. A unique signature identifies a family of zinc-dependent metallopeptidases. *FEBS Lett.* (1989) 242: 211-214

19) Adekoya, O.A. and Sylte, I. The Thermolysin Family (M4) of enzymes: therapeutic and biotechnological potential. *Chem. Biol. Drug.* (2009) 73:7–16

20) de Kreij, A., Venema, G. and Van den Burg, B. Substrate specificity in the highly heterogeneous M4 peptidase family is determined by a small subset of amino acids. *J. Biol. Chem.* (2000) 275: 31115–31120

21) Stocker, W. and Bode, W. Structural features of a superfamily of zinc-endopeptidases: the metzincins. *Curr. Opin. Struct. Biol.* (1995) 5: 383–390

22) Matthews, B.W. Structural Basis of the action of thermolysin and related zinc peptidases. *Acc. Chem. Res*. (1988) 21: 333–340

23) Matthews, B.W., Colman, P.M., Jansonius, J.N., Titani , K., Walsh, K.A. and Neurath, H. Structure of thermolysin. *Nat. New Biol.* (1972) 238: 41–43

24) Domon, B. and Aebersold, R. Mass spectrometry and protein analysis. *Science* (2006) 312: 212-217

25) Jonsson, A.P. Mass Spectrometry for protein and peptide characterization. *Cell.Mol.Life Sci.* (2001) 58: 868-884

26) Kebarle, P. A brief overview of the present status of the mechanism involved in electro spray mass spectrometry. *J. Mass Spec.* (2000) 35: 804-817

27) Huang, E.C., Pramanik, B.N., Tsarbopoulos, A., Reichert, P., Ganguly, A.K., Trotta, P.P. and Nagabhushan, T.L. Application of electrospray mass spectrometry in probing protein-protein and protein-ligand noncovalent interactions. *Am. Soc. Mass Spec.* (1993)

4: 624-630

28) Kelly, S.M., Jess, T.J. and Price, N.C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta,* (2005) 1751: 119-39.

29) Creighton, T.E. (1993) *Proteins: structure and molecular properties.* (2nd Edition) New york, WH Freeman and company

30) Gratzer, W.B. and Cowburn, D.A. Optical Activity of Biopolymers. *Nature* (1969) 222: 426-431

31) Correa, D.H.A. and Ramos, C.H.I. The use of circular dichorism spectroscopy in protein folding, form and function. *African J. Biochem. Res.* (2009) 3:164-173

32) Sarver, R.W. and Krueger, W.C. An infrared and circular dichorism combined approach to the analysis of protein secondary structure. *Anal. Biochem.* (1991) 199: 61-67

33) Greenfield, N.J. and Fasman, G.D. Computed circular dichorism spectra of protein Conformation. *Biochem.* (1969) 8: 4108-4116

34) Vivian, J.T. and Callis, P.R. Mechanism of tryptophan fluorescence shifts in proteins. *Biophys. J.* (2001) 80: 2093-2109

35) Chen, Y. and Barkley, M.D. Toward understanding tryptophan fluorescence in Proteins. *Biochem.* (1998) 37: 9976-9982

36) Royer, C.A. Fluorescence Spectroscopy. *Methods Mol. Biol*. (1995) 40: 65-89

37) Lakowicz, J.R. (2006) *Principles of fluorescence spectroscopy.* (3 rd Edition) Singapore, Springer

38) Ali, M.H. and Imperiali, B. Protein oligomerization: how and why? *Bioorg. Med. Chem.* (2005) 15: 5013-5020

39) Heuberger, E.H.M.L., Veenhoff, L.M., Durkens, R.H., Friesen, R.H.E. and Poolman, B. Oligomeric state of membrane transport proteins analyzed with blue native electrophoresis and analytical ultracentrifugation. *J. Mol. Biol.* (2002) 317: 591-600

40) Barth, H.G., Jackson, C. and Boyes, B.E. Size exclusion chromatography. *Anal. Chem.* (1994) 66: 595R-620R

41) Mori, S. and Barth, H.G. (1999) *Size exclusion chromatography.* Germany Springer

42) Barth, H.G., Boyes, B.E. and Jackson, C. Size excluison chromatography and other

related techniques. *Anal. Chem.* (1998) 70: 251R-278R

43) Pecora, R. (1985) *Dynamic Light Scattering: Applications of photon correlation Spectroscopy.* New York, Plenum press

44) Jachimska, B., Wasilewska, M. and Adamczyk, Z. Characterization of globular protein solutions by dynamic light scattering, electrophoretic mobility and viscosity measurment. *Langmuir* (2008) 24: 6866-6872

45) Gast, K., Zirwer, D., Frohne, M.M. and Damaschun, G. Compactness of the kinetic globule of bovine α-lactalbumin: a dynamic light scattering study. *Prot. Sci.* (1998) 7:2004-2011

46) Murphy, R.M. Static and dynamic light scattering study of biological macromolecules: what can we learn? *Curr. Opinio. Biotech.* (1997) 8: 25-30

47) Lebowitz, J. , Lewis, M.S., Schuck, P. Modern analytical ultracentrifugation in protein science: a tutorial review. *Prot. Sci.* (2002) 11: 2067-2079

48) Scott, D.J., Harding, S.E. and Rowe, A.J. (2005) *Analytical ultracentrifugation techiques and methods.* Cambridge, The Royal Society of Chemistry

49) Ebel, C. Analytical ultracentrifugation for sudy of biological macromolecules. *Progr. Colloid. Polym. Sci.* (2004) 127: 73-82

50) Cole, J.L., Lary, J.W., Moody, T.P. and Laue, T.M. Analytical ultracentrifugation : sedimentation velocity and sedimentaion equilibrium. *Methods Cell Biol.* (2008) 84: 143-179

51) Markley, J.L. Observation of histidine residues in proteins by means of nuclear magnetic resonance spectroscopy. *Accts. Chem. Res*. (1975) 8: 7080

52) Edgcomb, S.P., Murphy, K.P. Variability in the pKa of histidine side chains correlated with burial within proteins. *Proteins* (2002) 49: 1-6

53) Singer, M.A. and Lindquist, S. Multiple effects of trehalose on protein folding *in vitro* and *in vivo.* *Mol. Cell* (1998) 1: 639-648

54) Jain, N.A. and Roy, I. Effect of trehalose on protein structure. *Prot. Sci.* (2009) 18: 24-36

55) Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L. and Engel, P.C. Protein thermostability in extremopiles. *Biochimie* (1998) 80: 933-941

56) Vogt, G. and Argos, P. Protein thermal stability: hydrogen bonds or internal packing? *Fold Des.* (1997) 2: S40-S46

57) Bischof , J.C. and He, X. Thermal stability of proteins. *Ann. N.Y. Acad. Sci.* (2005) 1066: 12–33

58) Silvaggi, N.R., Wilson, D.,Tzipori, S. and Allen, K.N. Catlytic features of botulinum neurotoxin light chain revealed by high resolution structure of an inhibitory peptide complex. *Biochem.* (2008) 47: 5736-5745

59) Beaumont, A., O'Donohue, M.J., Paredes, N., Rousselet, N., Assicot, M., Bohuon, C., Fournie-Zaluski, M.C. and Roques, B.P. The role of His 231 in thermolysin like enzymes. *J. of Biol. Chem.* (1995) 270: 16803-16808

60) Pelmenschikov,V., Blomberg, M.R., Siegbahn, P.E. A theoretical study of the mechanism for peptide hydrolysis by thermolysin. *J. Biol. Inorg. Chem.* (2002) 7: 284-298

61) Binz, T., Bade, S., Rummel, A., Kollewe, A. and Alves, J. Arg 362 and Tyr 365 of the botulinum neurotoxin type A light chain are involved in transition state stabilization. *Biochem.* (2002) 41: 1717-1723

62) Ahmed, S.A., Oslon, M.A., Ludivico, M.L.,Gilsdorf , J., Smith, L.A. Identification of residues surrounding the active site of type A Botulinum neurotoxin important for substrate recognition and catalytic activity. *Protein J.* (2008) 27: 151-162

63) Uversky, V.N. Use of fast protein size exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochem.* (1993) 32: 13288-13298

64) Corbett, R.J.T. and Roche, R.S. Use of high speed size exclusion chromatography for study of protein folding and stability. *Biochem.* (1984) 23:1888-1894

65) Batas, B., Jones, H.R. and Chaudhuri, J.B. Study of hydrodynamic volume changes that occur during refolding of lysozyme using size-exclusion chromatogarphy. *J. Chromatogr. A.* (1997) 766: 109-119

66) Gast, K., Damaschun, G., Misselwitz, R. and Zirwer, D. Application of dynamic light scattering to studies of protein folding kinetics. *Eur. Biophys. J.* (1992) 21: 357-362

67) Gast K., Zirwer D., Damaschun H.,Hahn U., Frohne M.M., Wirth M. and Damaschun G. Ribonuclease T1 has different dimensions in the thermally and chemically denatured states: a dynamic light scattering study. *FEBS Lett.* (1997) 403: 245-248

68) Adel, A., Nadia, M., Mohamed, O. and Abdelhafidh, G. Study of thermally and chemically unfolded conformations of bovine serum albumin by means of dynamic light scattering. *Mat. Sci. Eng.* (2008) 28: 594–600

69) Laue, T.M. Analytical ultracentrifugation. *Curr. Proto. Prot. Sci.* (2001) 7.5.1-7.5.9

70) Behlke, J. and Ristau, O. Molecular mass determination by sedimentation velocity

experiments and direct fitting of the concentration profiles. *Biophys. J.* (1997) 72: 428-434

71) Hamada, H., Arakawa, T. and Shiraki, K. Effect of additives on protein aggregation. *Curr. Pharam. Biotechnol.* (2009) 10: 400-407

72) Sousa, R. Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. *Acta. Cryst,* (1995) D51: 271-277

73) Kaushik, J.K. and Bhat, R. Why is trehalose an exceptional protein stabilizer? *J. Biol. Chem.* (2003) 278: 26458-26465

74) Lee, J.C. and Timasheff, S.N. The stabilization of protein by sucrose. *J. Biol. Chem.* (1981) 256: 7193-7201

75) Stryer,L. Fluorescence spectroscopy of proteins. *Science* (1968) 162: 526-533

76) Chen, R.F., Edelhoch, H. and Steiner, R.F. Fluorescence of proteins (1969) Leach S.J. Ed. *Physical principles and techniques of protein chemistry Part A.* New York, Academic press

77) Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* (1999) 292: 195-202

78) Hangauer, D. G., Monzingo, A. F., and Matthews, B. W. An interactive computer graphics study of thermolysin-catalyzed peptide cleavage and inhibition by N-carboxymethyl dipeptides. *Biochem.* (1984) *23*: 5730-5741

79) Covington, A.K., Paabo, M., Robinson, R.A. and Bates, R.G. Use of the glass electrode in deuterium oxide and the relation between the standardized pD (pa_o) scale and the operational pH in heavy water. *Anal. Chem. 40: 700-706*

80) Sambrook, J. and Russel, D.W. (2001) *Molecular cloning and lab manual* (3rd edition) New York, Cold Spring Harbour Press

81) Peter, T. (1975). Putman, F.W. Ed. *The plasma proteins.* Academic press. pp. 133-181