

Electronic Thesis and Dissertation Repository

---

8-14-2014 12:00 AM

## Mechanistics of Prothymosin alpha and Nrf2 in the Keap1-Nrf2 mediated Oxidative Stress Response

Halema Khan, *The University of Western Ontario*

Supervisor: Dr. Wing Yiu Choy, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biochemistry

© Halema Khan 2014

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>

 Part of the [Biochemistry Commons](#)

---

### Recommended Citation

Khan, Halema, "Mechanistics of Prothymosin alpha and Nrf2 in the Keap1-Nrf2 mediated Oxidative Stress Response" (2014). *Electronic Thesis and Dissertation Repository*. 2277.  
<https://ir.lib.uwo.ca/etd/2277>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

**MECHANISTICS OF PROTHYMOSIN  $\alpha$  AND NRF2 IN THE  
KEAP1-NRF2 MEDIATED OXIDATIVE STRESS RESPONSE**

(Thesis format: Integrated Article)

by

**Halema Khan**

Graduate Program

in

Biochemistry

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies  
The University of Western Ontario  
London, Ontario, Canada

© Halema Khan, 2014

## Abstract

In an effort to dissect the mechanism of interaction of IDPs, in this thesis we focus on Prothymosin  $\alpha$  (ProT $\alpha$ ) and nuclear factor erythroid 2-related factor 2 (Nrf2), intrinsically disordered proteins, in the Nrf2 mediated oxidative stress response. Kelch-like ECH-associated protein 1 (Keap1) is an inhibitor of Nrf2, a key transcription factor of cytoprotective genes. Under unstressed conditions, Keap1 interacts with Nrf2 in the cytoplasm via its Kelch domain and suppresses Nrf2 activity. During oxidative stress, Nrf2 is released from Keap1 and is shuttled to the nucleus, where it initiates pro cell survival gene transcription. ProT $\alpha$  also interacts with the Kelch domain and mediates the import of Keap1 into the nucleus to inhibit Nrf2 activity.

To gain a molecular basis understanding of the oxidative stress response mechanism, the interaction between ProT $\alpha$  and the Kelch domain of Keap1 has been delineated using nuclear magnetic resonance spectroscopy (NMR), isothermal titration calorimetry (ITC), peptide array analysis, and site-directed mutagenesis. The results revealed that ProT $\alpha$  retains a high level of flexibility, even in the Kelch-bound state. Mutational analysis pinpointed that the region <sup>38</sup>NANEENGE<sup>45</sup> of ProT $\alpha$  is crucial for the interaction with the Kelch domain, while the flanking residues play relatively minor roles in the affinity of binding.

A high yield purification protocol with complete backbone NMR resonance assignment lays the foundation for structural and biophysical studies of the full length-Neh2 domain of the human Nrf2. In this work the full-length Neh2 domain was used to investigate binding to Kelch in the presence of cancer causing somatic mutations.

To understand the mechanistic links between Keap1 mutations and cancer pathogenesis, the molecular effects of a series of mutations (G333C, G364C, G379D, G350S, R413L, R415G, A427V, G430C, and G476R) on the structural and target recognition properties of Keap1 are investigated. These mutations are found to exert differential effects on the protein stability and target binding. Together with the proposed Hinge-and-Latch mechanism of Nrf2/Keap1 binding, these results provide important insight into the molecular impact of different somatic mutations on Keap1's function as an Nrf2 repressor.

**Keywords:** oxidative stress response; Nrf2; Prothymosin  $\alpha$ ; Keap1; intrinsically disordered proteins, nuclear magnetic resonance; protein dynamics; protein-protein interactions; isothermal titration calorimetry

# **Co-Authorship Statement**

## **Chapter 1**

This chapter focuses on literature review of intrinsically disordered proteins and the Nrf2 mediated oxidative stress response. Figure 1.1 presented in this chapter is adapted from the works of Uversky, V.N. Protein Sci. 2002;11(4) p.739-56. Figure 1.2 was adapted and modified from the text book Biochemistry, Genetics and Molecular Biology (2012). Remaining of the figures are made by Halema Khan and cited where necessary.

## **Chapter 2**

Chapter 2 contains work from published manuscript Halema Khan, Elio.A.Cino, Anne Brickenden, Jingsong Fan, Daiwen Yang, Wing-Yiu Choy, J Mol Biol, 2013.425(6):p.1011-27. The manuscript was written by Halema Khan, Elio.A.Cino and Wing-Yiu Choy. The pET15b plasmid for the Kelch cDNA of human Keap1 was kindly gifted by Dr. Mark Hannink at the University of Missouri-Columbia. Jingsong Fan, Daiwen Yang and Wing-Yiu Choy performed the Amide exchange experiments (National University of Singapore) and Wing-Yiu Choy produced Figure 2.3. The peptide arrays were synthesized at The High Throughput Biology Laboratory (Dr. Shawn Li, University of Western Ontario). All other experiments and analyses were carried by Halema Khan.

### **Chapter 3**

The Gateway cloning for the full-length Neh2 and Neh2 $\Delta$ ETGE was performed by Anne Brickenden. Cloning of the Neh2 $\Delta$ DLG was done by Halema Khan and Anne Brickenden. The remainder of the experimental work and analyses were done by Halema Khan.

### **Chapter 4**

Work in this chapter comes from the submitted manuscript Halema Khan, Ryan Killoran, Anne Brickenden, Jingsong Fan, Daiwen Yang and Wing-Yiu Choy *Biochem.J.* (2014). The initial manuscript was written by Halema Khan and revised by Dr. Wing-Yiu Choy. Halema Khan and Ryan Killoran performed the mutagenesis work to generate the Kelch mutants. Halema Khan, Ryan Killoran and Anne Brickenden performed ITC experiments of WT, G350S, A427V and R415G with the Prothymosin  $\alpha$  and Neh2ETGE peptides. Three-dimensional NMR ( $^1\text{H}$ - $^{15}\text{N}$ -HSQC, HN(CO)CACB, HNCACB and  $^{15}\text{N}$ -NOESY-HSQC) data for the  $^2\text{H}$ - $^{15}\text{N}$ - $^{13}\text{C}$  labeled Kelch domain was collected by Daiwen Yang (National University of Singapore). The remaining experimental work and analyses was completed by Halema Khan.

## **Dedication**

*To my mother,  
for always being there for me*

## **Acknowledgments**

I would like to express my deepest appreciation to my supervisor, Dr. James Wing Yiu Choy, for providing me with the opportunity to explore this research avenue. His continuous mentorship, guidance, support and encouragement made it possible to come this far. James, you are a great person and I am really honored to be part of your lab. If I had the choice to start all over again, I would definitely choose the Choy lab. Thank you for everything.

I would also like to thank my advisory committee members, Dr. Gary Shaw, and Dr. Eric Ball for taking interest in my work and for their sincere advice over the years.

Thank you Anne Brickenden. I know you always say that “without you this was possible but it would have taken longer.” Honestly I don’t think I could have survived without your advice and help. You are extremely knowledgeable and talented in what you do. Not only have I found a great colleague, but also made lifelong friend. I truly admire your love for science and the Choy lab. If I could dedicate this thesis to someone else besides my mom, your name would be at the top of the list.

I would like to acknowledge Elio A. Cino, Andrew Maciejewski and Ryan Killoran for their insightful comments and help over the years. You guys made the Choy lab experience a memorable one. Thanks for all the laughter and price-less moments AND pizza times. I know I would never say this otherwise, but I have found caring friends in you guys. Also thanks to Lee-Anne for her continuous help with data analysis in the BICF room.

To my family, without your support and guidance over the years, I would never come this far in my educational pursuit. Mom, your prayers and staying besides me through the



good and bad have given me the strength to keep going. Warzeda, Saboor, Yasir, and Yasmin, thank you for always being there for me. Laiyan, and Eesa, I know you guys are too young right now, but thank you for the smiles during the rough times. Every time I was discouraged, I looked at you guys, and gained strength to continue. I would also like to thank my in-laws for their prayers and continuous support.

Last but not least, Lawangeen, although I do say at times that this educational journey would have ended sooner if you were not around, but, your support and love through the difficult times made it that much more easier. I have mentioned many people without whose help I could not get this far, but you have played a major role in achieving this goal. We have accomplished a lot in the past few years and we have much more ahead of us. I pray that we stay together with our kids through all the ups and downs of life. Thank you for everything and I do love you a lot (now it's written on paper).

# Table of Contents

Abstract.....	ii
Co-Authorship Statement.....	iv
Acknowledgments.....	vii
Table of Contents.....	ix
List of Tables.....	xiv
List of Figures.....	xv
List of Abbreviations.....	xviii
<b>Chapter 1</b> .....	1
1 Introduction.....	1
1.1 Intrinsically Disordered Proteins.....	1
1.1.1 IDP Characterization.....	5
1.1.2 IDP Functions.....	6
1.1.3 IDP Advantages.....	7
1.1.4 IDPs in Disease.....	9
1.1.5 IDPs as potential drug targets.....	10
1.1.6 Functional Related Folding of IDPs.....	11
1.1.7 IDPs and Fuzzy Complexes.....	12
1.1.8 Ligand binding studies of IDPs.....	15
1.2 Oxidative Stress.....	16
1.3 Key players in the Nrf2 mediated Oxidative Stress Response.....	17
1.3.1 Nrf2.....	17
1.3.2 Keap1.....	22
1.3.3 Keap1-Cul3-RBX1 Ligase Complex.....	24
1.3.4 Antioxidant Response Element.....	26

1.3.5	Role of stress mediated cell signaling kinases .....	27
1.3.6	Prothymosin alpha .....	28
1.4	Mechanism of Nrf2 transcription Regulation .....	29
1.5	Positive and Negative Regulators of Nrf2 .....	31
1.6	Dual Nature of Nrf2 in Cancer.....	33
1.7	IDPs in the Nrf2 mediated Oxidative Stress Response.....	34
1.8	Scope of thesis .....	38
1.9	References.....	40
<b>Chapter 2</b>	.....	<b>55</b>
2	Fuzzy complex formation between Prothymosin $\alpha$ and the Kelch domain of Keap1..	55
2.1	Introduction.....	55
2.2	Materials and Methods.....	59
2.2.1	Protein Expression and Purification.....	59
2.2.2	Peptide synthesis.....	60
2.2.3	Site-directed mutagenesis of wild-type ProT $\alpha$ .....	60
2.2.4	Primers used for point mutations. ....	60
2.2.5	NMR spectroscopy.....	61
2.2.6	Analytical ultracentrifugation .....	63
2.2.7	Isothermal titration calorimetry (ITC) .....	64
2.2.8	Peptide array .....	65
2.3	Results.....	65
2.3.1	Mapping the Kelch-binding region on ProT $\alpha$ by NMR spectroscopy .....	65
2.3.2	Disordered ProT $\alpha$ forms a fuzzy complex with the Kelch domain of Keap1 .....	68
2.3.3	Identifying crucial residues of ProT $\alpha$ for Kelch binding .....	72
2.3.4	Quantitative analysis of ProT $\alpha$ and Kelch interaction.....	74

2.4 Discussion.....	82
2.5 References.....	91
<b>Chapter 3</b> .....	<b>98</b>
3 Structural characterization of the full-length Neh2 domain of human Nrf2 and its interaction with the Kelch domain of human Keap1 .....	98
3.1 Introduction.....	98
3.2 Materials and Methods.....	101
3.2.1 Construction of Neh2, Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG in E. coli Expression Vectors .....	101
3.2.2 Expression of the Neh2, Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG protein.....	102
3.2.3 Purification of the Soluble Neh2, Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG protein .....	103
3.2.4 Purification of Neh2 and Neh2 $\Delta$ ETGE protein from Inclusion bodies .	104
3.2.5 Purification of Neh2, Neh2 $\Delta$ ETGE, and Neh2 $\Delta$ DLG protein with 8 M Urea.....	105
3.2.6 Analytical Ultracentrifugation Experiments .....	106
3.2.7 Circular Dichroism.....	106
3.2.8 NMR Spectroscopy.....	107
3.3 Results.....	108
3.3.1 Purification of the Neh2 domain.....	108
3.3.2 Backbone Assignment of the human Neh2 domain of Nrf2.....	117
3.3.3 Neh2 binding to the Kelch domain of the human Keap1.....	122
3.3.4 Two site substrate recognition of the human Nrf2-Keap1 system.....	125
3.3.5 Secondary structure analysis of the Neh2 domain, Neh2 $\Delta$ DLG and Neh2 $\Delta$ ETGE .....	128
3.3.6 Hydrodynamic radius of Neh2 protein .....	130
3.3.7 Backbone Dynamics of the Neh2 protein .....	133
3.4 Discussion.....	135

3.5	References.....	141
<b>Chapter 4</b>	.....	<b>144</b>
4	Molecular effects of cancer-associated somatic mutations on the structural and target recognition properties of Keap1.....	144
4.1	Introduction.....	144
4.2	Materials and Methods.....	151
4.2.1	Cloning, expression and purification of Keap1-Kelch variants.....	151
4.2.2	Neh2 constructs, expression and purification.....	152
4.2.3	Peptides.....	153
4.2.4	Nuclear magnetic resonance (NMR) experiments.....	153
4.2.5	Circular Dichroism (CD) Spectropolarimetry.....	154
4.2.6	Isothermal Titration Calorimetry.....	154
4.3	Results.....	155
4.3.1	Mutations exert differential effects on the solubility of the Kelch domain of Keap1.....	155
4.3.2	Effects of mutations on the target recognition of the Kelch domain.....	157
4.3.3	Somatic mutations in the Kelch domain affect its interaction with the full-length Neh2 domain.....	163
4.3.4	Probing structural changes of Kelch and its mutational variants by CD and NMR.....	170
4.3.5	NMR analysis of the interactions of Kelch with Neh2 and ProT $\alpha$ peptides.....	177
4.4	DISCUSSION.....	184
4.5	References.....	190
<b>Chapter 5</b>	.....	<b>194</b>
5	Summary.....	194
5.1	Introduction.....	194
5.2	Previous Work.....	195

5.3 Fuzzy complex between the flexible ProT $\alpha$ and Kelch domain of human Keap1 .....	198
5.4 Binding interface of human ProT $\alpha$ with the Kelch domain of human Keap1 ....	199
5.5 The Neh2 domain of human Nrf2.....	200
5.6 Interaction between the Neh2 and Kelch domain.....	201
5.7 Order in disorder and its implication in target binding.....	202
5.8 Somatic mutations in the Kelch domain and its impact on target recognition ...	203
5.9 Conclusion .....	206
5.10Future Work.....	207
5.11References.....	209
Curriculum Vitae .....	213

## List of Tables

Table 1.1. A list of somatic mutations identified within the Kelch domain of Keap1 .....	35
Table 2.1. Thermodynamic parameters for binding of the full-length ProT $\alpha$ , ProT $\alpha$ peptide, and ProT $\alpha$ mutants to Kelch domain of Keap1 at 25 °C (first set) .....	75
Table 2.2. Thermodynamic parameters for binding of the ProT $\alpha$ peptide, full-length ProT $\alpha$ and its mutants to Kelch domain of Keap1 at 25 °C (second set) .....	76
Table 2.3. Molecular weight values of Kelch calculated from the sedimentation equilibrium experiment. ....	79
Table 3.1. Protein yield in mg per liter of culture media from different purification methods .....	116
Table 3.2. $^{15}\text{N}$ , $^{13}\text{C}$ and $^1\text{H}$ resonance assignments for the human Neh2 domain .....	119
Table 4.1. Somatic mutations included in this study.....	148
Table 4.2. Thermodynamic parameters for the binding of the Neh2 and ProT $\alpha$ peptides to the Kelch domain of human Keap1 at 25 °C.....	159
Table 4.3. Thermodynamic parameters for the binding of the Neh2 and ProT $\alpha$ peptides to the human Kelch domain of Keap1 at 25 °C (duplicate run).....	160
Table 4.4. Thermodynamic parameters for the binding of the Kelch domain of Keap1 to Neh2 $\Delta$ ETGE at different temperatures. ....	165
Table 4.5. Thermodynamic parameters for the binding of the wild-type and mutational variants of Kelch to the full length Neh2 at 5° C.....	167
Table 4.6. Thermodynamic parameters for the binding of the Kelch and Kelch mutational variants to full length Neh2 at 5° C (duplicate runs) .....	168
Table 4.7. Deconvoluted CD data for the WT Kelch and mutational variants .....	172
Table 5.1. A table listing structural differences between Neh2 and ProT $\alpha$ . . ....	204

## List of Figures

Figure 1.1. The Protein Quartet Model. ....	4
Figure 1.2. Implications of oxidative stress in pathology. ....	18
Figure 1.3. Structural topology of the Neh2 domain of human Nrf2.....	21
Figure 1.4. Structural topology of the Keap1 and a schematic diagram of the Keap1 bound to Nrf2.. ....	23
Figure 1.5. Keap1 as a substrate adaptor for the cullin3-RBX1 ligase complex. ....	25
Figure 1.6. Schematic diagram of the Keap1-Nrf2 mediated gene transcription.....	30
Figure 1.7. Crystal structures of the ProT $\alpha$ and Neh2 peptides bound to the Kelch domain of Keap1. ....	37
Figure 2.1. NMR titration experiments of ProT $\alpha$ with Kelch.....	67
Figure 2.2. Backbone $^{15}\text{N}$ relaxation experiments of ProT $\alpha$ in the free (blue) and bound state with 2 equivalents of Kelch. ....	69
Figure 2.3. Protection factor ( $P$ ) of backbone amides in ProT $\alpha$ in the absence (blue bars) and presence of Kelch (red *). ....	71
Figure 2.4. Permutation peptide array experiments. ....	73
Figure 2.5. ITC profiles of Keap1 with full length ProT $\alpha$ and its mutant variants. ....	77
Figure 2.6. Sedimentation equilibrium analysis of Kelch and ProT $\alpha$ binding. ....	78
Figure 3.1. Sequence alignment of the mouse and human Neh2 domains of Nrf2.....	99
Figure 3.2. Disorder prediction for the human and mouse Neh2.....	100
Figure 3.3. SDS-PAGE (12%) analysis of the overexpression and purification of the human Neh2 domain. ....	111
Figure 3.4. $^1\text{H}$ - $^{15}\text{N}$ HSQC of the human Neh2 domain of Nrf2.....	112
Figure 3.5. Purification of Neh2 by denaturation in 8M Urea. ....	113
Figure 3.6. $^1\text{H}$ - $^{15}\text{N}$ HSQC of the Neh2 domain of human Nrf2.....	114
Figure 3.7. Neh2 Mass spectrum and Stoichiometry. ....	115
Figure 3.8. Backbone resonance assignment of the Neh2 domain of the human Nrf2 protein.....	118



Figure 3.9. Secondary structure propensity of the Neh2 domain of Nrf2.....	121
Figure 3.10. An overlay of $^1\text{H}$ - $^{15}\text{N}$ HSQC spectrum of Neh2 domain (black) and in complex with the Kelch domain of Keap1 (red).....	123
Figure 3.11. $^1\text{H}$ - $^{15}\text{N}$ HSQC of Neh2 variants.....	124
Figure 3.12. $^1\text{H}$ - $^{15}\text{N}$ HSQC of Neh2 variants in free and Kelch bound states at 25 °C and 30°C.....	126
Figure 3.13. Secondary structure of the Neh2 protein and its variants measured by far-UV CD spectropolarimetry .....	129
Figure 3.14. Thermal melting of Neh2 FL and its variants Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG at 222 nm.....	131
Figure 3.15. Analytical ultracentrifugation experiment of.....	132
Figure 3.16. Backbone $^{15}\text{N}$ NMR relaxation data for Neh2 full-length ( <i>red</i> ) and Neh2 $\Delta$ DLG ( <i>blue</i> ).. ..	134
Figure 4.1. Mutation sites located in the Kelch domain of Keap1 (PDB: 1U6D (Li, Zhang et al. 2004)). .....	150
Figure 4.2. Coomassie-stained SDS-PAGE depicting the purification of Kelch and its mutational variants.....	156
Figure 4.3. $^1\text{H}$ - $^{15}\text{N}$ HSQC NMR spectra of G333C (a), G379D (b), and R413L (c) mutants of the Kelch domain. ....	158
Figure 4.4. ITC profiles of titrating ProT $\alpha$ -ENGE peptide to wild-type Kelch.....	162
Figure 4.5. ITC profiles of titrating wild-type Kelch to Neh2 $\Delta$ ETGE at 25 °C (a), 20 °C (b), 15 °C (c), 10 °C (d), and 5 °C, respectively. ....	164
Figure 4.6. ITC profiles of titrating wild-type Kelch and mutational variants to full length Neh2 5 °C .....	169
Figure 4.7. CD spectra of Kelch and mutational variants .....	171
Figure 4.8. Backbone resonance assignments of the Kelch domain of human Keap1. ..	174
Figure 4.9. Overlay of the $^1\text{H}$ - $^{15}\text{N}$ TROSY-HSQC NMR spectra of wild-type Kelch (black) and mutants .....	175
Figure 4.10. $^1\text{H}_\text{N}$ and $^{15}\text{N}$ chemical shift changes of the Kelch mutants. ....	176
Figure 4.11. $^1\text{H}$ - $^{15}\text{N}$ HSQC NMR spectra of the wild-type Kelch domain in the absence (black) and presence of five molar equivalents of (a) Neh2-ETGE peptide, (b) Neh2-DLG peptide, and (c) ProT $\alpha$ peptide.....	179

Figure 4.12. Overlay of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectra of the (a) G364C, (b) R415G, (c) G350S, and (d) A427V Kelch mutations in the absence (black) and presence (orange) of a two molar equivalent of the Neh2-ETGE peptide. .... 180

Figure 4.13. Overlay of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectra of the Kelch (a) G364C, (b) R415G, (c) G350S, and (d) A427V mutations in the absence (black) and presence (orange) of a five molar equivalent of the Neh2-DLG peptide. .... 182

Figure 4.14. Composite  $^1\text{H}_\text{N}$  and  $^{15}\text{N}$  chemical shift perturbation ( $\Delta\delta=[(\Delta\delta_{\text{HN}})^2+(\Delta\delta_{\text{N}}/5)^2]^{1/2}$ ) analysis of the wild type and the four mutants. .... 183

## List of Abbreviations

$\delta$	chemical shift
$\varphi$	psi angle
$\psi$	phi angle
ADC	adenocarcinoma
BTB	broad complex, tramtrack and bric-a-brac
ARE	antioxidant response element
Cul3	cullin 3
CSI	chemical shift index
CD	circular dichroism
D <sub>2</sub> O	deuterium oxide
Da	daltons
DGR	double glycine repeat
DSS	2,2'-dimethyl-2-silapentane-5-sulfonate
DTT	dithiothreitol
<i>E. coli</i>	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionization
GST	glutathione S-transferase
HSQC	heteronuclear single quantum coherence
Hz	hertz
IDP	intrinsically disordered protein
IDPRs	intrinsically disordered protein regions
IPTG	isopropyl $\beta$ -D-thiogalactopyranoside
IVR	intervening region
Keap1	kelch-like ECH-associated protein 1
$K_d$	dissociation constant
kDa	kilo daltons
LCC	large cell carcinoma
MW	molecular weight
NaCl	sodium chloride
Nrf2	nuclear factor erythroid 2-related factor 2
NH <sub>4</sub> Cl	ammonium chloride
NMR	nuclear magnetic resonance
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
ProT $\alpha$	prothymosin alpha
RBX1	Ring finger domain
PCR	polymerase chain reaction
RNase 1	human pancreatic ribonuclease
ROS	Reactive oxygen species
ppm	parts per million
RyR	ryanodine receptor
SAXS	x-ray scattering

SCC	small cell carcinoma
SDS	sodium dodecyl sulfate
SSP	secondary structure propensity
TEV	tobacco etch virus
Tris	tris(hydroxymethyl) aminomethane
Ub	ubiquitin
WT	wild type
Ala (A)	alanine
Arg (R)	arginine
Asn (N)	asparagine
Asp (D)	aspartic acid
Cys (C)	cysteine
Gln (Q)	glutamine
Glu (E)	glutamic acid
Gly (G)	glycine
His (H)	histidine
Ile (I)	isoleucine
Leu (L)	leucine
Lys (K)	lysine
Met (M)	methionine
Phe (F)	phenylalanine
Pro (P)	proline
Ser (S)	serine
Thr (T)	threonine
Trp (W)	tryptophan
Tyr (Y)	tyrosine
Val (V)	valine

# Chapter 1

## Introduction

### 1.1 Intrinsically Disordered Proteins

Intrinsically disordered proteins (IDPs) are functional proteins that lack a well-defined three dimensional structure (Dunker, Lawson et al. 2001; Garza, Ahmad et al. 2009). The discovery of these unstructured proteins challenged the long accepted protein structure-function paradigm, which states that the three dimensional structure is a prerequisite for a protein to perform its biological activity. While there is no doubt that for some proteins structure and function are related, it is now widely accepted that many biologically active proteins do not conform to a well folded structure. These disordered proteins generally lack sufficient hydrophobic residues required for a protein to form a three-dimensional organized entity. In fact, it has been reported that on average typical disordered proteins have a lower composition of Phe, Ile, Trp, Tyr, Cys, Val and Leu residues (Romero, Obradovic et al. 2001; Dunker, Brown et al. 2002a; Vucetic, Brown et al. 2003). These proteins are found to be rich in prolines and charged residues, especially Lys, Glu, and Arg amino acids (Romero, Obradovic et al. 2001; Dunker, Brown et al. 2002a; Vucetic, Brown et al. 2003). A few examples of disordered proteins with these properties include Prothymosin  $\alpha$ ,  $\alpha$ -synuclein, non-A $\beta$  component of AD amyloid precursor protein (NACP), and Apo-cytochrome c (Stellwagen, Rysavy et al. 1972; Gast, Damaschun et al. 1995; Weinreb, Zhen et al. 1996; Uversky, Gillespie et al. 2000a). Because of differential amino acid composition of ordered and disordered proteins, many predictors have been developed that allow determining unstructured proteins/regions with

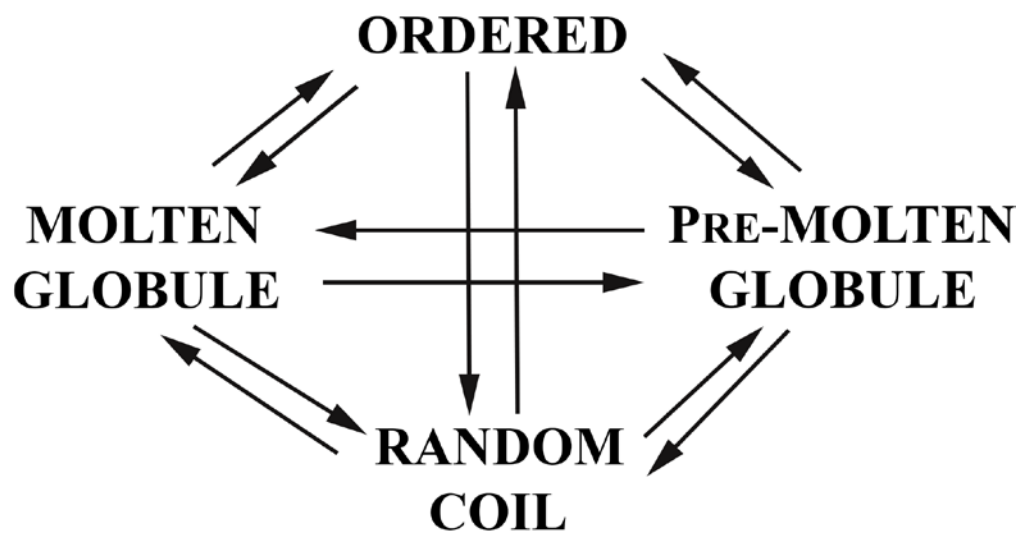
high accuracy. Some of these bioinformatics programs used for protein structural analysis are PONDR, GLOBPLOT, DIOPRED2 and DisEMBLE (Romero, Obradovic et al. 1997; Uversky, Gillespie et al. 2000a; Dunker, Brown et al. 2002b; Linding, Russell et al. 2003; Ward, Sodhi et al. 2004).

Structural analyses of disordered proteins/domains have revealed that many IDPs do not behave as true random coils, rather these proteins span several conformations that interconvert and contain local elements of secondary structure (Dyson and Wright 2002; Dyson and Wright 2005). The backbone bonds and angles of such proteins fluctuate over time, with no definite values, thus, resulting in ensembles of various structures, referred to as “protein clouds” (Dunker and Uversky 2010). Extended regions, regions with residual secondary structure, domains with secondary structure, and domains with poorly packed side chains have all been observed in many proteins lacking structure (Uversky 2011; Uversky 2013a). The SH3 domain of the *Drosophila* signal transduction drk protein illustrates this phenomenon best as it exists between folded and unfolded states in aqueous buffer and experiences multiple conformations in the unfolded state (Zhang and Forman-Kay 1995; Choy and Forman-Kay 2001). Similarly, FG-nucleoporins contain long disordered regions that adopt different conformations (from a molten globule to a fully extended-coil) under physiological conditions (Milles and Lemke 2011).

Intrinsic disorder is not just restricted to full-length proteins; in fact, unstructured regions are part of many well folded proteins and are associated with crucial biological functions (Wright and Dyson 1999). Analyses of genome sequences using disorder predictors have identified about 35-51% of proteins in eukaryote that contain disordered

regions of 40 or more consecutive residues (Dunker, Obradovic et al. 2000). 41% of human plasma membrane proteins are reported to contain ID regions of 30 or more residues (Minezaki, Homma et al. 2007). An example of a protein with ordered and disordered regions is p53. It is a structured homo-tetramer, the core domains of which are linked by the disordered N and C-terminal domains, responsible for transcriptional activation and regulation, respectively, (Bell, Klein et al. 2002; Dawson, Muller et al. 2003).

The prevalence of IDPs and intrinsically disordered regions (IDPRs) and their significance in cellular tasks necessitated revision of the structure-function paradigm to include these proteins and regions (Uversky 2002). The new view describing protein structure and function includes proteins with well folded structure, pre-molten globules, molten globules, and native coils; and is referred by the “The Protein Quartet model,” (Figure 1.1) (Uversky 2002; Uversky 2013a). According to this new model, biological functions can be performed by proteins in any of the four conformations and under the course of their biological activity these proteins can transition between various structural rearrangements (Uversky 2013a). For example, the 50S ribosomal proteins, L22 and L27, and 30S ribosomal protein S19 are unfolded in solution; however, they exist as well-folded rigid structures in the functional ribosomes (Venyaminov, Gudkov et al. 1981; Yusupov, Yusupova et al. 2001; Uversky 2002). Additionally, Prothymosin  $\alpha$  is a highly flexible protein but transforms into a compact structure in the presence of zinc ions; this transition in structure facilitates its interaction with targets (Uversky, Gillespie et al. 2000b; Yi, Boys et al. 2007).



**Figure 1.1. The Protein Quartet Model.** This new model takes into account protein function from any of the four conformations (Ordered, pre-molten globule, molten globule, and random coil) and proteins can switch conformations between any of these four states (figure adapted from Uversky 2002).



The discovery of IDPs raises the question as how these proteins avoid the cellular degradation machinery? Hochstrasser explains that proteolytic degradation is tightly regulated in eukaryotic cells and is only initiated in when a protein undergoes ubiquitination (Hochstrasser 1996). In this case, unfolded proteins can carry out their specific roles but experience rapid turnover (Wright and Dyson 1999). Comprehensive analyses of IDP regulation inside the cell has shown that mRNA encoding IDPs have higher decay rates, and these proteins are modulated by post-translational modifications to have shorter half-lives (Gspöner, Futschik et al. 2008; Uversky 2011). It has also been suggested that evolution in sequence allows disordered proteins to perhaps avoid the unfolded protein response altogether (Uversky and Dunker 2013b).

### ***1.1.1 IDP Characterization***

The lack of a well-defined structure complicates structural studies of IDPs. In fact, since IDPs do not conform to a true random coil, a stable unique structure is not attainable for an IDP of interest. The aim of IDP structural studies is to determine conformational propensities, transient long range interactions and regions of restricted or high mobility. Such information can always be beneficial in determining the biological function of the protein under study (Eliezer 2007). Several methods such as X-ray crystallography, Circular Dichroism (CD) spectropolarimetry, protease digestion, small angle X-ray scattering (SAXS), disorder predictors (listed above), and Nuclear Magnetic Resonance (NMR) spectroscopy can be used in characterizing intrinsic disorder in a protein (Dunker, Lawson et al. 2001; Uversky and Longhi 2011). Crystallography, CD, and other methods only indicate the presence of disorder in a protein. For example, in crystallography, a disordered region is indicated by missing electron density. A

downside to identifying disorder by this method is that whether the missing region is due to intrinsic disorder in a protein or a result of experimental error cannot be known without performing further experiments (Dunker, Lawson et al. 2001). Similarly, CD provides semi-quantitative data allowing to distinguish between ordered, molten globule and random coil structures (Dunker, Lawson et al. 2001). It cannot provide residue specific information with regards to what regions of the protein are ordered and disordered. NMR on the other hand can detect structure, provides motional information per residue, and can be used to map the binding interface of IDPs with targets (Choy and Forman-Kay 2001; Paci, Vendruscolo et al. 2002; Juneja and Udgaonkar 2003; Dyson and Wright 2004; Kay 2005).

### ***1.1.2 IDP Functions***

The structural flexibility that arises due to intrinsic disorder in proteins plays a significant functional role in biological processes. The vast functions of IDPs include regulation of transcription and translation, cellular signal transduction, the storage of small molecules, and membrane transport (Dunker, Lawson et al. 2001; Dunker, Cortese et al. 2005; Garza, Ahmad et al. 2009). A well characterized example of disordered transcriptional activation domain is the kinase-inducible activation (KID) domain of CREB, which interacts with the KIX domain of CBP (Radhakrishnan, Perez-Alvarado et al. 1998; Dunker, Oldfield et al. 2008). Similarly, in the eukaryotic translation initiation pathway the eIF4G in the large multi subunit protein-RNA complex, required for recruitment of the ribosomes, is devoid of structure in the absence of its target (Hershey, McWhirter et al. 1999; Uversky 2002).

Intrinsic disorder is also found within molecular chaperones, proteins that assist RNA or other misfolded proteins in attaining their ordered structure (Todd, Lorimer et al. 1996; Tompa and Csermely 2004). The disorder segments in chaperones are believed to help adopt well folded structures by either loosening the unfolded/misfolded structure or function as recognition elements (Tompa and Csermely 2004). The A1 protein of heteronuclear ribonucleoprotein (hnRNP) has a disordered Gly-rich C-terminal domain that is involved in renaturation and facilitates assembly of the RNP complex (Pontius and Berg 1990). Another chaperone with structural disorder is the p23, a co-chaperone of Hsp90 (Weikl, Abelmann et al. 1999).

Structural adaptability of IDPs allows them to interact with multiple partners and perform various functions, thus classifying IDPs as moonlighting proteins (Tompa, Szasz et al. 2005). Moonlighting is the ability of a protein to perform a number of unrelated tasks (Jeffery 1999; Jeffery 2004). Examples of IDPs involved in moonlighting are the, p21<sup>Cip1/WAF1</sup> and p27, IDPs that regulate cell proliferation through activating or inhibiting cyclin dependent kinases (Cheng, Olivier et al. 1999; Olashaw, Bagui et al. 2004; Dunker, Oldfield et al. 2008). Likewise, p53, which is a tumor suppressor, is involved in gene expression, apoptosis induction, DNA repair, and response to cellular stress (Cheng, LeGall et al. 2006).

### ***1.1.3 IDP Advantages***

Although flexibility is observed in all states of protein conformation, the extreme dynamic nature of IDPs is more advantageous than folded and molten globular proteins in carrying out certain cellular functions (Mittag, Kay et al. 2009). A detailed analysis of

IDP functional advantages is recently published by Uversky (Uversky 2013a). Some functional advantages of IDPs include the presence of a large interactive surface, promoting fast and easy recognition of a binding partner (Dunker, Lawson et al. 2001; Dyson and Wright 2005). The large surface area facilitates specific binding and allows for easy dissociation from partners when required (Dunker, Lawson et al. 2001; Dyson and Wright 2005). This type of behavior is observed in regulatory signaling proteins which form protein complexes rapidly to target specific gene expression and have high turnover rate ensuring the proteins are not present for a long period of time (Sugase, Dyson et al. 2007).

Due to lower steric hindrance compared to natively structured proteins, IDPs often behave as hub proteins, proteins that are involved in a large number of distinct interactions (Dunker, Cortese et al. 2005). This property of IDPs allows them to take part in multiple biological processes. For example, p53 uses the flexibility in its disordered regions to bind to multiple targets (Oldfield, Meng et al. 2008). More so, it adopts different conformations upon interaction with different partners, allowing the protein to take part in many cellular functions some of which are stated earlier (Cheng, LeGall et al. 2006; Oldfield, Meng et al. 2008).

IDPs are solvent exposed, thus, posttranslational modification sites (PTMs) are a common occurrence in IDP sequences (Dunker, Brown et al. 2002a; Dunker, Brown et al. 2002b; Xie, Vucetic et al. 2007). Several types of post-translational modifications that can include acetylation, ADP-ribosylation, ubiquitination, prenylation, carboxylation, glycosylation, methylation, ubiquitination, and SUMOylation are found within disordered regions (Iakoucheva, Radivojac et al. 2004; Radivojac, Iakoucheva et al. 2007; Xie,

Vucetic et al. 2007). These modifications are usually reversible and are associated with signaling and regulation (Xie, Vucetic et al. 2007). An illustrative example is the p53 protein, in which most known PTM sites occur within the disordered N and C terminal domains, and modifications of these sites determine the function, localization and turnover rate of p53 (Bode and Dong 2004).

#### ***1.1.4 IDPs in Disease***

IDPs are involved in many diverse biological roles, therefore, it is apparent that malfunction of these proteins can play a vital role in various health disorders. Causes leading to aberrant function of these proteins rises due to misfolding of partially folded regions, point mutations, exposure to internal or external toxins, impaired post-translational modifications, impaired trafficking, loss of binding partner, oxidative damage or an increased chance of degradation (Dunker, Oldfield et al. 2008). IDPs are found in several pathological conditions, such as, cardiovascular diseases, diabetes, cancer, and neurodegenerative disorders (Dunker, Oldfield et al. 2008). Sequence analyses of proteins by disorder predictors have shown approximately 79% and 61% of cancer and cardiovascular associated proteins, respectively, to contain regions of disorder (Iakoucheva, Brown et al. 2002; Uversky, Oldfield et al. 2008). This high prevalence of IDPs in human diseases has given rise to the ‘disorder in disorder’ concept, referred to as the D2 concept (Uversky, Oldfield et al. 2008). Some examples illustrating the D2 concept include the A $\beta$ -peptides in Alzheimer’s disease and  $\alpha$ -synuclein in Parkinson’s disease, which are in a partially folded states pre-amyloid formation (Tompa 2009; Frimpong, Abzalimov et al. 2010), and myelin basic protein (MBP), dysregulations of

which are linked to human autoimmune demyelinating disease multiple sclerosis (Homchaudhuri, De Avila et al. 2010).

### ***1.1.5 IDPs as potential drug targets***

The abundance of IDPs in cell signaling and regulatory roles and their frequency in human pathogenesis makes them unique drug targets. However traditional drug design, which relies on well-defined structures of proteins/regions, is an ineffective means to target IDPs in health disorders (Cheng, LeGall et al. 2006). In order to target IDPs, new approaches, accounting for the lack of structure in IDPs need to be considered (Cheng, LeGall et al. 2006). Two approaches have been put forward to target IDPs in disease. The first method takes into account the ability of IDPs to fold upon target binding (Cheng, LeGall et al. 2006). The drug target mimics the bound state ordered structure of the targeted IDP and competes for interaction with the binding partner. The successful design of ‘Nutlins’ targeting the p53-Mdm2 interaction is a great example of targeting disordered regions with small molecules in cancer prevention (Vassilev 2004; Vassilev, Vu et al. 2004; Cheng, LeGall et al. 2006; Uversky 2013a). The ‘Nutlins’ mimic the bound state structure of the disordered N-terminal region of p53, which forms a helix upon binding to Mdm2, and binds to target with higher affinity. Oral treatment of mice with ‘nutlins’ resulted in 90% inhibition of tumor growth (Klein and Vassilev 2004; Vassilev 2004; Vassilev, Vu et al. 2004).

An alternative practical method to block IDP protein-protein interaction is to target IDP binding sites with small molecules (Cheng, LeGall et al. 2006). This method has been effectively applied to the transcription factor c-Myc, which forms a complex

with its heterodimerization partner Max, and is associated with various types of cancers (Hammoudeh, Follis et al. 2009; Uversky 2013a). The small molecule upon complex formation induces conformational changes in the binding site of c-Myc and prevents its interaction with Max (Hammoudeh, Follis et al. 2009). This technique can also be applied to disordered proteins that experience “functional misfolding.” That is, some disordered proteins adopt non-native conformations that result in protein-protein interactions and consequently unwanted cellular activity (Uversky 2012). Binding of small molecules to these misfolded regions can stabilize these states in IDPs, therefore, preventing the occurrence of non-native interactions (Uversky 2012; Uversky 2013a).

### ***1.1.6 Functional Related Folding of IDPs***

Structural studies of IDPs in protein-protein interaction has established that a coupled folding process occurs when an IDP interacts with a globular protein or another IDP (Wright and Dyson 2009). As stated formerly, some disordered proteins are completely devoid of structure under physiological conditions, while others contain residual structure or short amphipathic motifs called ‘molecular recognition features’, more commonly known as MoRFs (Oldfield, Cheng et al. 2005; Mohan, Oldfield et al. 2006). These MoRFs consist of approximately 20 residues and are proposed to function in signaling and recognition of protein or nucleic acid partners (Oldfield, Cheng et al. 2005). Alpha-helix,  $\beta$ , irregular, and complex MoRFs are recognized and these short disordered segments are observed to undergo a disorder-to-order transition upon target binding (Oldfield, Cheng et al. 2005; Vacic, Oldfield et al. 2007). The coupled folding results in stable formation of helices,  $\beta$ -strands or other irregular structures in target bound state (Vacic, Oldfield et al. 2007). The C-terminal domain of the p21 protein

contains a  $\beta$ -MoRF, which interacts with the proliferating cell nuclear antigen (PCNA) and blocks DNA replication (Gulbis, Kelman et al. 1996; Vacic, Oldfield et al. 2007). Another well characterized example of a protein that undergoes transition to an ordered state is that of the kinase-inducible transcriptional-activation domain (KID) of CREB mentioned earlier. The KID domain exists as a disordered region both in the full length and in isolation (Dyson and Wright 2005). Upon binding to its target, the KIX domain of CREB-binding protein (CBP), adopts a helix-turn-helix conformation (Radhakrishnan, Perez-Alvarado et al. 1998; Dyson and Wright 2005; Sugase, Dyson et al. 2007). Another example that best illustrates a protein that can bind to multiple partners and adopt different structure is of p53, which can form a helix, a sheet, and two different irregular structures when binding to four different partners (Dunker, Oldfield et al. 2008).

### ***1.1.7 IDPs and Fuzzy Complexes***

The couple folding of IDPs revived the classical structure-function paradigm, suggesting that upon binding IDPs behaved as “traditional” folded proteins (Dunker, Oldfield et al. 2008). Detailed analyses of IDPs in complex states recognized that coupled folding is not applicable to all IDPs. Several examples of unstructured proteins in ligand-bound state have made it evident that disorder can exist even in complex states and has important implications in cellular functions (Tompa and Fuxreiter 2008). The term “fuzziness” has been suggested by Tompa *et al.* to emphasize the importance of this structural obscurity that rises in protein-protein interactions. Structural fuzziness in the ligand bound state has been categorized into five distinct models; polymorphic, clamp, flanking, random coil, or the sequence independent model (Tompa and Fuxreiter 2008).



In polymorphic complexes, single or multiple well-defined conformations are observed in the target-bound states and these structures can be resolved (Tompa and Fuxreiter 2008). Interestingly, these multiple conformations are able to perform discrete biological functions. For example, the human pancreatic ribonuclease (RNase 1), which is responsible for many cellular functions including cleavage of single stranded RNA, double stranded RNA and double stranded RNA-DNA hybrids. Several conformations are observed for the loop regions, involved in various enzymatic activities, and for side chains of critical residues on the surface, implicated in membrane binding, translocation, and RNA hydrolysis (Kover, Bruix et al. 2008). Similarly, depending on the binding affinity, disordered loop of dihydropyridine receptor (DHPR) binds ryanodine receptor (RyR) and can result in activation or inhibition of the RyR channel opening (Haarmann, Green et al. 2003; Fuxreiter and Tompa 2012). It has been implied that these different activities result from different interactions between the disordered loop and gating residues of the RyR channel (Fuxreiter and Tompa 2012).

The clamp model describes a protein with two ordered regions connected by a segment or linker that remains disordered in the complex state (Tompa and Fuxreiter 2008; Fuxreiter and Tompa 2012). The ordered regions serve as a clamp and are conformationally restricted by the linker, which is not involved in direct protein-protein interactions (Fuxreiter and Tompa 2012). A great example illustrating the clamp model is the myosin VI protein, which interacts with actin polymer via its two heads that are separated by an 80 residue disordered region (Rock, Ramamurthy et al. 2005; Fuxreiter and Tompa 2012).

The flanking model takes into consideration the disordered regions around MoRFs or other recognition elements such as linear motifs (Tompa and Fuxreiter 2008; Fuxreiter and Tompa 2012). Deleting these flanking regions can impact target binding (Zor, Mayr et al. 2002; Fuxreiter and Tompa 2012). As observed in the case of KID binding to the KIX domain of CBP, in which only a segment of 29 residues adopts helical structure, deletion or extension in the surrounding disordered region can decrease the  $K_d$  of the complex formation. A recently reported example of flanking model is a small chloroplast protein CP-12 in the complex state with the glyceraldehyde-3-phosphate dehydrogenase, GAPDH, involved in light/dark regulation of Calvin cycle (Mileo, Lorenzi et al. 2013). The binding site of CP-12 forms an  $\alpha$ -helix in GAPDH bound state, while the remaining flanking regions remain unstructured.

When the entire or most of the IDP remains unstructured in the bound state, a random model complex is formed (Fuxreiter and Tompa 2012). Biophysical studies of these IDPs suggest that such proteins behave similarly in the free and target-bound states, thus they show absence of ordered structure in all states (Fuxreiter and Tompa 2012). The single-stranded DNA binding protein samples several unstructured conformations in the bound state with the DNA (Savvides, Raghunathan et al. 2004). Sic1, a disordered protein, inhibits cyclin-dependent kinase (CDK) activity by binding to CDK via multiple sites. Interestingly, it retains its dynamic structure in the target bound state (Mittag, Orlicky et al. 2008).

The fifth group classifying fuzzy complexes are proteins that interact in a sequence independent manner (Fuxreiter and Tompa 2012). Structural differences in the bound state are ascribed to the variation in amino acid contacts. Interchanges of residues

in the binding site do not affect target recognition, as witnessed for Ewing's sarcoma fusion proteins (Ng, Potikyan et al. 2007; Fuxreiter and Tompa 2012). The dynamic interaction between the acidic transcription activator Gcn4 and one activator-binding domain of the mediator subunit Gal11/Med15 (Brzovic, Heikaus et al. 2011) is another great example for sequence independent model. In the bound state the Gcn4 forms a helix, which adopts multiple orientations in the binding cleft of Gal11 (Brzovic, Heikaus et al. 2011).

### ***1.1.8 Ligand binding studies of IDPs***

Several techniques are available for studying interactions of intrinsically disordered proteins with targets. Some of these include crystallography, CD, surface plasmon resonance (Gerothanassis., Anastassios et al. 2002), isothermal titration calorimetry (ITC), and NMR. NMR spectroscopy is by far the most suitable tool for investigating details of IDPs in their target bound state. Many parameters in NMR, such as chemical shifts, line widths and relaxation rates are atom specific and contain information about the local conformation and dynamics of IDPs (Uversky and Longhi 2011). Like NMR, Isothermal titration calorimetry is another unique method that allows the study of molecular interactions of IDPs with proteins, nucleic acids, small organic molecules, metals, and ions (Pierce, Raman et al. 1999). Both these methods are the main experimental techniques used to study IDP interactions with targets in the work presented here.

## 1.2 Oxidative Stress

Reactive oxygen species (ROS) and free radicals are present in all aerobic organisms (Martindale and Holbrook 2002; Fulda, Gorman et al. 2010). These ROS species are either generated within the cell through signal transduction pathways, in the process of metabolism, or by exposure to an external insult (Martindale and Holbrook 2002). Normally, the cell maintains equilibrium between the production of ROS and the antioxidant species, and under these normal conditions the ROS can play important role in regulatory functions (Martindale and Holbrook 2002). However, when this balance is disrupted, the cell experiences oxidative stress. The increase in ROS can jeopardize cell survival by causing damage to nucleic acids, proteins, carbohydrates and lipids (Finkel and Holbrook 2000). Hence, the cell initiates a defense response in order to return to homeostasis by combating the ROS. Alternatively, if the severity of the insult is high, signaling pathways for cell death are activated (Finkel and Holbrook 2000; Martindale and Holbrook 2002).

The signaling pathways activated to combat ROS include the extracellular signal-regulated kinase (ERK), c-Jun amino-terminal kinase (JNK) and p38 mitogen-activated protein kinase (MAPK) signaling cascades, the phosphoinositide 3-kinase (PI(3)K)/Akt pathway, the nuclear factor (NF)- $\kappa$ B signaling system, p53 activation, and the heat shock response pathways (Finkel and Holbrook 2000; Martindale and Holbrook 2002). The heat shock response, ERK, PI(3)K/Akt and NF- $\kappa$ B signaling pathways encourage cell survival, whereas activation of p53, JNK and p38 are associated with cell death (Finkel and Holbrook 2000).

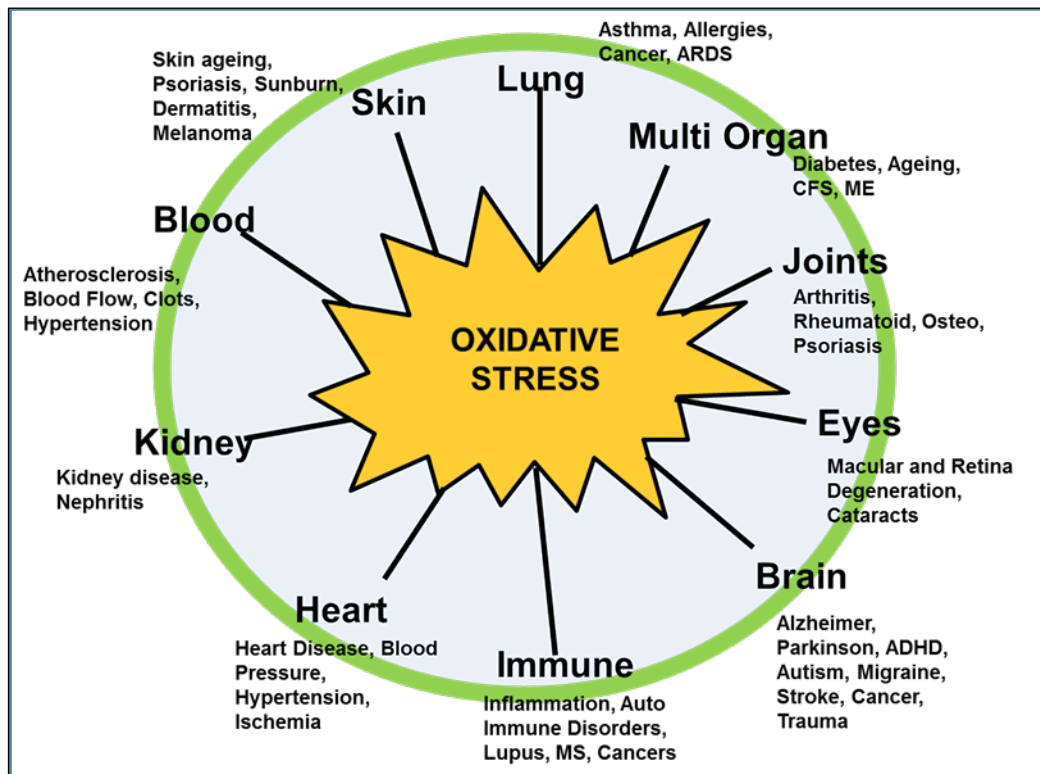
Oxidative stress is the most accepted etiopathological factor in many diseases including cancers, atherosclerosis, diabetes mellitus, rheumatic diseases, autoimmune disorders and neurodegenerative disorders (Figure 1.2) (Calabrese, Lodi et al. 2005; Mihaela Ilie 2012). Because of its important role in major diseases, the oxidative stress has gained significant consideration in research with over 3000 publications in the last five years in PubMed alone. One major pathway activated in response to oxidative stress is the Nuclear (erythroid 2-like) related factor 2 (Nrf2) pathway. The Nrf2 mediated oxidative stress response plays a major role in xenobiotic detoxification and is the main focus of the work presented in this thesis.

### **1.3 Key players in the Nrf2 mediated Oxidative Stress Response**

The signalling and regulatory pathways in the cell are highly complex processes with many proteins/enzymes taking part to perform distinct functions. Correspondingly, the Nrf2 mediated response to oxidative stress involves many players. For the purpose of this work, following is a brief description of a few of the key players involved in this pathway.

#### **1.3.1 Nrf2**

The nuclear factor erythroid 2 (NFE-2)-related factor 2, more commonly known as Nrf2, was first discovered during studies of the  $\beta$ -globins. In these works, targets for the DNase I hypersensitive site 2, also known as AP1-NFE2 (activating protein 1 and nuclear factor erythroid 2) motif, in  $\beta$ -globins were being cloned (Moi, Chan et al. 1994). The hunt for transcription factors that bind to the AP1-NFE2 site led to the discovery of



**Figure 1.2. Implications of oxidative stress in pathology.** Oxidative stress is associated with many different diseases involving different organs of the human body as illustrated in this diagram. This wheel diagram was adapted and modified from (Mihaela Ilie 2012)

several targets from the cap 'n' collar (CNC) subfamily of the basic leucine zipper (bZIP) transcription factors which include the p45-NFE2 (Andrews, Erdjument-Bromage et al. 1993), Nrf1 (Chan, Han et al. 1993), Nrf2 (Moi, Chan et al. 1994) Nrf3 (Kobayashi, Ito et al. 1999) as well as BACH1(Oyake, Itoh et al. 1996) and BACH2 (Muto, Hoshino et al. 1998).

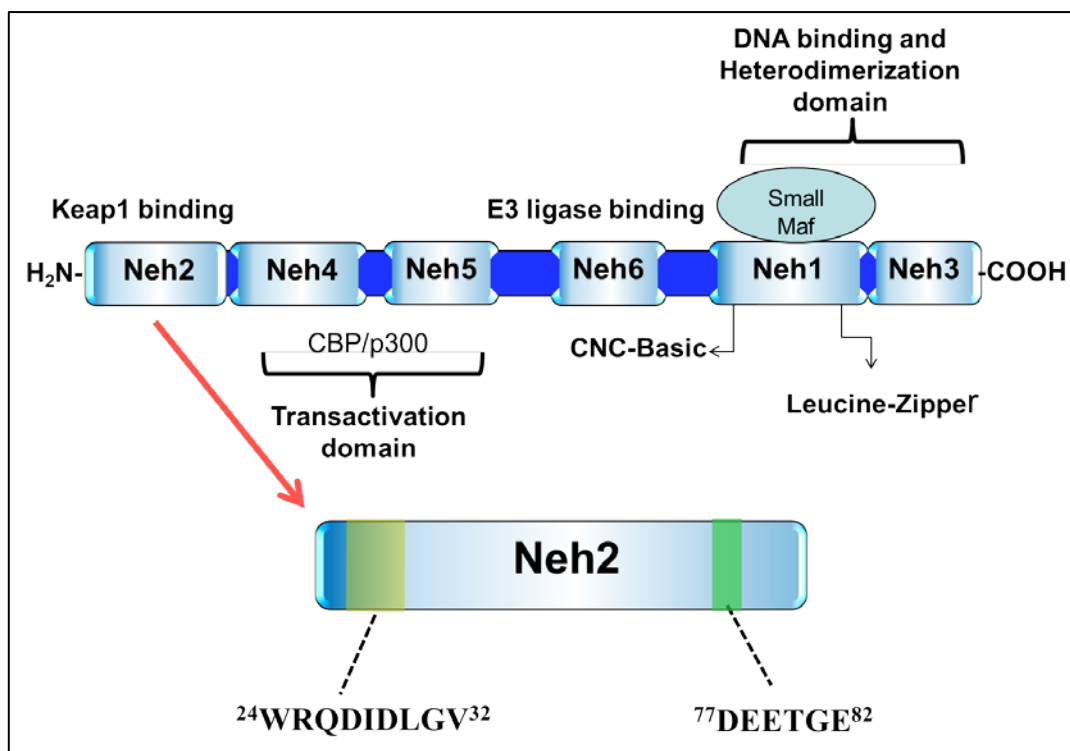
The CNC family of transcription factors are characterized by the presence of the CNC domain composed of a 43 amino acid region that is conserved amongst mammals, birds, fish and insects and is located at the N-terminus of the DNA binding domain (Sykiotis and Bohmann 2010). The DNA-binding domain includes a nuclear localization signal, which has been shown to be functional in p45, Nrf2 and BACH2 (Hoshino, Kobayashi et al. 2000; Theodore, Kawai et al. 2008; Perdomo, Fock et al. 2010). BACH1 and BACH2 proteins are defined by the presence of a BTB (Broad complex, Tramtrack, Bric-a-brac) domain required for protein–protein interaction (Oyake, Itoh et al. 1996). Additionally, most CNC factors are known as transcriptional activators; however, BACH1 and BACH2 function as transcription repressors (Sykiotis and Bohmann 2010). Interestingly, none of the CNC based transcriptional factors bind to DNA independently, rather each first forms a complex with a small MAF protein, another bZIP protein (Igarashi, Kataoka et al. 1994). The resulting obligate heterodimeric complexes bind to NFE2 (Nuclear Factor-Erythroid 2), MARE (Maf recognition element), ARE (antioxidant response element) and StreB (stress-response element)/EpRE (electrophile response element) type DNA binding sites (Chevillard and Blank 2011).

Of the many CNC proteins, the role and regulation of the Nrf2 has been most extensively studied. The important role for the Nrf2 in the protection from xenobiotic

and oxidative stresses was first shown in the analyses of Nrf2-null mice (Itoh, Chiba et al. 1997). Although, these mice were viable and displayed no phenotypic defects, they were sensitive to extrinsic and intrinsic oxidative assaults. For instance, Nrf2-null mice exhibit an increased formation of DNA adducts in the lung when exposed to diesel exhaust (Aoki, Sato et al. 2001), severe form of liver toxicity is developed after administration of acetaminophen (Enomoto, Itoh et al. 2001), an increased susceptibility to cigarette smoke-induced emphysema (Iizuka, Ishii et al. 2005) and an aggravated bleomycin-induced pulmonary fibrosis is developed (Kikuchi, Ishii et al. 2010). Additionally, Nrf2 knockout mice unexpectedly develop various inflammatory disorders, including glomerulonephritis and immune-mediated hemolytic anemia, multi-organ autoimmune inflammation (Lee, Chan et al. 2004; Ma, Battelli et al. 2006; Yoh, Itoh et al. 2001).

Structural analysis has revealed six evolutionary conserved domains in the Nrf2 protein that are named Neh1 through Neh6 (Itoh, Wakabayashi et al. 1999). Figure 1.3 depicts these six domains including their roles and functions. Neh1, in the c-terminal half of Nrf2, constitutes the basic DNA binding domain and the leucine zipper for dimerization with targets (Itoh, Wakabayashi et al. 1999; Sun, Chin et al. 2009). Acetylation of the Neh1 domain by p300/CBP results in promoter-specific DNA binding of Nrf2 (Sun, Chin et al. 2009). The carboxy-terminal Neh3 domain of Nrf2 is important for the transcriptional activity of Nrf2 and may play role in recruiting components of the transcriptional complex (Nioi, Nguyen et al. 2005). The Neh4 and Neh5 regions are transactivation domains which synergistically regulate transcriptional activation of cytoprotective genes (Itoh, Wakabayashi et al. 1999; Zhang, Hosoya et al. 2007). The Neh2 and Neh6 are involved in mediating Nrf2 degradation (McMahon, Thomas et al.





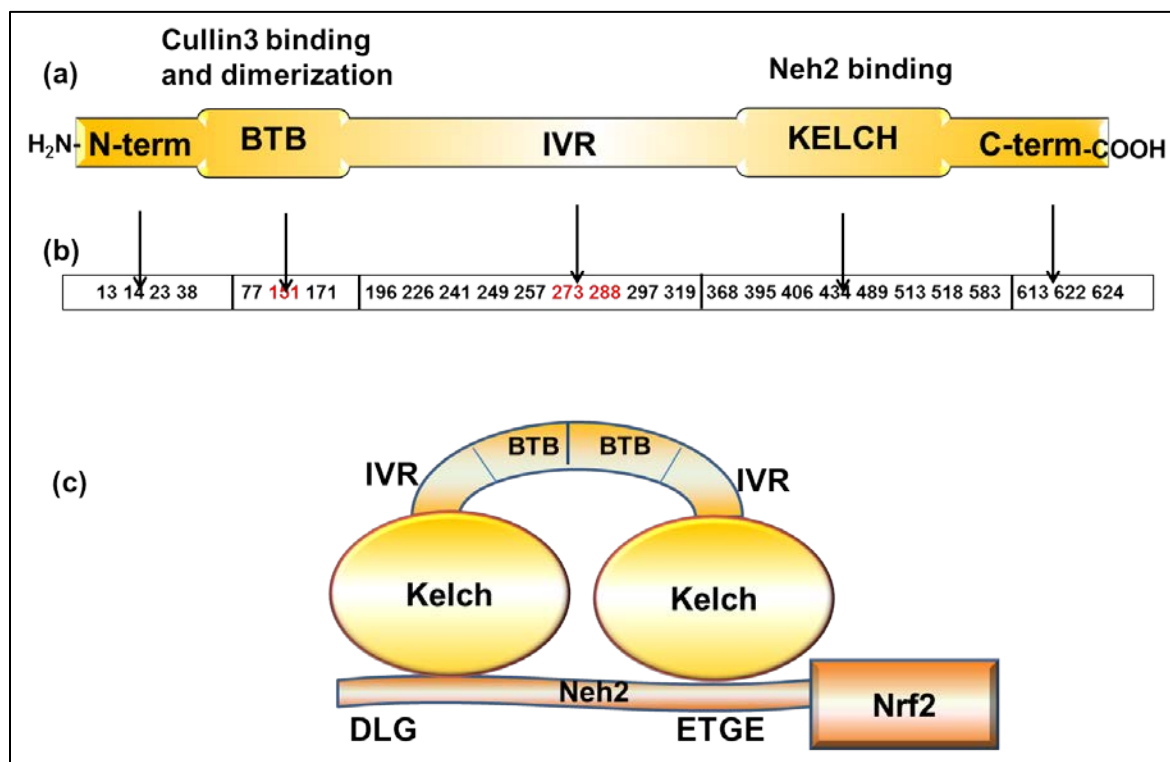
**Figure 1.3. Structural topology of the Neh2 domain of human Nrf2.**

There are six highly conserved domains (Neh1 to Neh6) in the Nrf2 transcription factor. Neh2 is located in the N-terminus and is found to interact with the Keap1 to negatively regulate the Nrf2 protein. Neh1 and Neh3 are the C-terminal basic leucine zipper containing domains that dimerize with the small Mafs and bind to ARE sequences in DNA. Neh4 and Neh5 are transactivation domains that are identified to interact with the co-activator CBP/p300 during transcriptional activation in the nucleus. The Neh6 interacts with E3 ligases thus mediates Nrf2 degradation. Diagram based on the review by Kansanen et al. 2013 (Kansanen., Kuosmanen et al. 2013).

2004). The Neh2 domain of the Nrf2 protein has been structurally characterized and has been shown to be intrinsically disordered containing little secondary structural content (Tong, Katoh et al. 2006). This N-terminal domain is involved in the negative regulation of Nrf2. It has been shown that the absence of this domain leads to increase in the transcription activity of the Nrf2 protein (Itoh, Wakabayashi et al. 1999).

### ***1.3.2 Keap1***

The Keap1 (Kelch-like ECH-associated protein1) protein was identified as an Nrf2 regulating protein using a yeast two-hybrid system (Itoh, Wakabayashi et al. 1999). The human Keap1 is composed of 624 amino acids containing high density of cysteine residues and has five distinct domains: the N-terminal region, the Bric-a-Brac, tramtrack, broad-complex (BTB), a cysteine rich intervening region (IVR), Kelch domain or the double glycine repeat region (DGR), and the C-terminal domain. A structural topology of the Keap1 domains and their biological roles is provided in Figure 1.4a along with a box identifying the cysteine rich regions (Figure 1.4b). Based on Figure 1.4b it is clear that Keap1 is a cysteine rich protein with the mouse and human Keap1 containing 25 and 27 cysteine residues respectively (Dinkova-Kostova, Holtzclaw et al. 2002). The presence of high number of cysteine residues suggests that these cysteines may be involved in sensing changes in the cellular redox status (Dinkova-Kostova, Holtzclaw et al. 2002). Various studies have demonstrated that the sulfhydryl groups of various Keap1 cysteine residues can be directly altered by oxidation, reduction and alkylation. Of these 151, 273 and 288 (Figure 1.4b, residues in red) appear to be critical for the dissociation of Nrf2 from the Keap1-Nrf2 complex (Dinkova-Kostova, Holtzclaw et al. 2002;



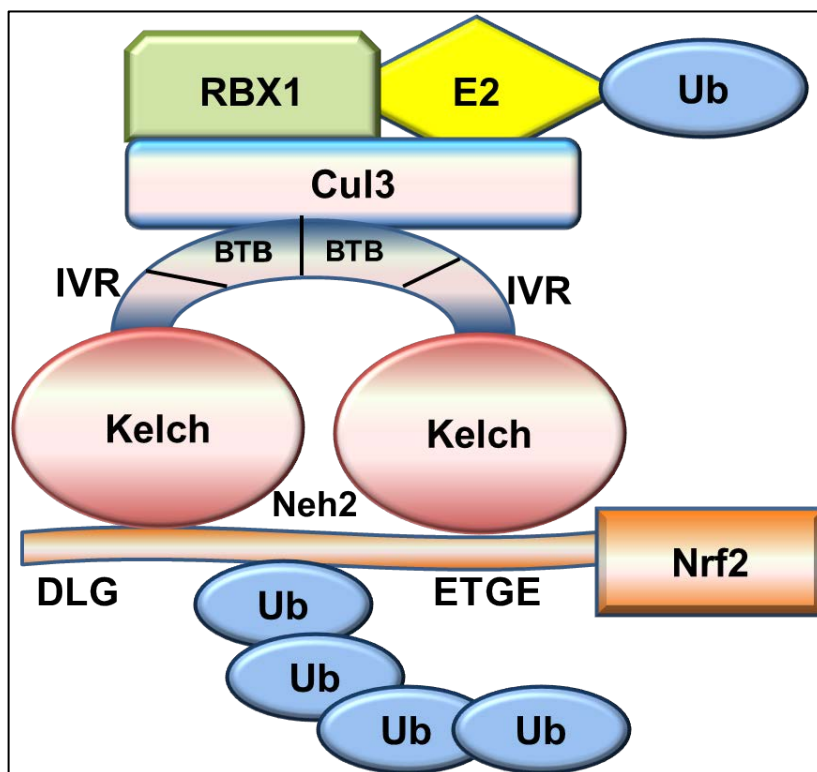
**Figure 1.4. Structural topology of the Keap1 and a schematic diagram of the Keap1 bound to Nrf2.** (a) Structural topology of the Keap1 highlighting the five distinct domains of the protein. The BTB domain is responsible for the homodimerization of the Keap1 protein. The BTB and the IVR regions are responsible for proteasome dependent degradation of Nrf2. The Kelch domain interacts with the Neh2 domain anchoring Nrf2 to the actin cytoskeleton. (b) Cysteine residues localized in the five distinct domains of the Keap1 protein. Cys residues that play crucial role in the regulation of Nrf2 are highlighted in red. (c) Schematic diagram of the Keap1-Nrf2 complex displaying the hinge and latch mechanism of interaction between the Neh2 domain of Nrf2 and the Kelch domain of Keap1. Figure is adapted and modified from (Kansanen., Kuosmanen et al. 2013).

Wakabayashi, Dinkova-Kostova et al. 2004; Kobayashi and Yamamoto 2006). Additionally, the involvement of BTB and IVR domains (where these critical cysteine residues reside) of Keap1 in Nrf2 degradation has been shown to be crucial in recruiting ubiquitin-proteasome factors (Kobayashi, Kang et al. 2004).

The Kelch and C-terminal regions of the Keap1 are responsible for anchoring the Nrf2 in place, allowing its ubiquitination (Kang, Kobayashi et al. 2004). The crystal structure of the Kelch domain has revealed a six bladed  $\beta$ -propeller conformation with several conserved amino acids that are important for maintaining the hydrogen bond network, linking the propeller blades and forming the hydrophobic core of the protein (Li, Zhang et al. 2004). In particular, there is a conserved glycine doublet found in each blade of the Kelch domain (Li, Zhang et al. 2004). These two glycine residues are involved in an intrablade hydrogen bond network that may be instrumental in the folding of individual blade structures. As demonstrated in Figure 1.4c, two motifs, namely the DLG (latch) and the ETGE (hinge), in the Neh2 domain of Nrf2 independently associate with the Kelch domain of Keap1 (Padmanabhan, Tong et al. 2006; Tong, Katoh et al. 2006).

### ***1.3.3 Keap1-Cul3-RBX1 Ligase Complex***

Keap1 serves as a substrate linker protein for Cul3-RBX1 ligase complex to ubiquitinate proteins and subsequently target them for degradation by the 26S proteasome (Furukawa and Xiong 2005). An illustrative diagram of the Keap1 in complex with the Cul3-RBX1 ligase complex is presented in Figure 1.5. This figure 1.5 demonstrates how Keap1 anchors the Nrf2 for ubiquitination. The ubiquitination of a substrate protein such as Nrf2 is normally accomplished by consecutive reactions initiated by ubiquitin



**Figure 1.5. Keap1 as a substrate adaptor for the cullin3-RBX1 ligase complex.**

Nrf2 interacts with two Keap1 proteins via its DLG and EETGE motifs to the Kelch domains. The Keap1 protein forms a homodimer via the BTB domain and interacts with the Cul3-RBX1 complex. The RBX1-E2 synthesizes a polyubiquitin chain that is transferred to the Nrf2 by the Ub-E3. Figure is made based on the works of Taguchi, 2011 (Taguchi, Motohashi et al. 2011).

activating enzymes and involves three steps. First step is the ubiquitin activation, performed by E1; second step involves conjugation of the ubiquitin to the protein and this is completed by the E2 enzyme; the third step, performed by the E3 enzyme, is the ubiquitin ligation (Hochstrasser 1996). The E3 consists of scaffold Cullin protein that is bound to the Ring box protein-1 (RBX1) (Deshaies 1999) (Figure 1.5). The RBX1 is a small 108 residue protein with a RING finger domain (Ohta, Michel et al. 1999). The RING proteins are capable of activating E2 enzymes to synthesize polyubiquitin chains in the presence of E1 (Furukawa, Ohta et al. 2002).

The Cullin proteins are evolutionary conserved scaffold proteins that play a critical role in post-translation modification of proteins involving ubiquitin. There are eight Cullin proteins that have been identified to date, Cul1 to Cul7 and PARC (p53-associated, parkin-like cytoplasmic protein) (Sarikas, Hartmann et al. 2011). Keap1 has been found to associate favorably with the Cul3 protein (Figure 1.5) (Kobayashi, Kang et al. 2004; Zhang, Lo et al. 2004). Cul3 assemble E3 ligases by bringing the RING-E2 and substrates together (Furukawa, Ohta et al. 2002; Furukawa and Xiong 2005).

#### ***1.3.4 Antioxidant Response Element***

The antioxidant response element (ARE) was discovered by Rushmore et al as a cis-acting regulatory enhancer sequence in the upstream region of the rat glutathione S-transferase Ya (GST Ya) subunit (Rushmore and Pickett 1990). Characterization of the ARE sequence in the rat GST Ya by mutational analysis delineated the ARE core sequence to 5'-TGACnnnGC-3', which is essential for both basal and inducible activity of the GST Ya (Rushmore, Morton et al. 1991). In addition to the rat GST Ya gene, the

ARE is responsible for activation and induction of several other genes, including the rat and human NAD(P)H:quinone oxidoreductase 1 (NQO-1) protein,  $\gamma$ -glutamylcysteine catalytic ( $\gamma$ -GCSH) and regulatory ( $\gamma$ -GCSI) subunits, and heme oxygenase (HO-1) (Favreau and Pickett 1991; Jaiswal 1991; Inamdar, Ahn et al. 1996; Mulcahy, Wartman et al. 1997; Moinova and Mulcahy 1998; Wild, Gipp et al. 1998).

Venugopal et al established Nrf2 as a critical regulator of the ARE sequences (Venugopal and Jaiswal 1996). In their works, Venugopal *et al* discovered increased CAT activity from an ARE-CAT construct upon expression of transfected Nrf2. The CAT activity was found to diminish in the presence of a mutant ARE-CAT construct. *In vivo* studies using gene knockout mice demonstrated Nrf2 as a regulator for the transcription activation of ARE-responsive genes (Itoh, Chiba et al. 1997). Expression levels of  $\gamma$ -GCSH and NQO-1 were reduced in Nrf2-null mice when compared with the heterozygous mice (Itoh, Chiba et al. 1997; Chanas, Jiang et al. 2002).

### ***1.3.5 Role of stress mediated cell signaling kinases***

Stress-related signaling kinases, mitogen-activated protein kinase (MAPK), p38 MAPK, phosphatidylinositol 3-kinase (PI3K), c-Jun-N-terminal kinase (JNK), endoplasmic reticulum (ER)-resident kinase (PERK), and protein kinase C (PKC) have been reported to mediate Nrf2 regulation of ARE-responsive genes (Alam, Wicks et al. 2000; Yu, Chen et al. 2000; Zipper and Mulcahy 2000; Huang and Ingber 2002; Balogun, Hoque et al. 2003; Cullinan, Zhang et al. 2003; Keum, Yu et al. 2006; Xu, Yuan et al. 2006). Many of these kinases directly phosphorylate Nrf2 and activate the transcription of ARE-regulated genes. Interestingly, both positive and negative regulation of Nrf2 has been associated with the p38 and the JNK kinases (Alam, Wicks et al. 2000; Yu, Chen et

al. 2000; Zipper and Mulcahy 2000; Huang and Ingber 2002; Balogun, Hoque et al. 2003; Cullinan, Zhang et al. 2003; Keum, Yu et al. 2006; Xu, Yuan et al. 2006).

### ***1.3.6 Prothymosin alpha***

Prothymosin-alpha (ProT $\alpha$ ), a precursor to thymosin  $\alpha$ , is a well-studied protein with diverse roles in the cells ranging from protecting cells from apoptosis to cell proliferation (Gomez-Marquez, Segade et al. 1989; Evstafieva, Belov et al. 2003). The natural abundance of ProT $\alpha$  and its evolutionary conservancy implies its importance to biological activity (Haritos, Tsolas et al. 1984; Hannappel and Huff 2003). Due to the nature of tasks it participates in, a dual role has been associated with ProT $\alpha$  (Ioannou, Samara et al. 2012). Intracellularly, ProT $\alpha$  controls the cell cycle by inhibiting apoptosis, regulates DNA remodeling during cell proliferation, and controls gene expression by interacting with the histone H1 (Szabo, Ehleiter et al. 1992; Wu, Shiau et al. 1997; Karetsou, Kretsovali et al. 2002). Extracellularly, ProT $\alpha$  has been shown to stimulate immune response (Ioannou, Samara et al. 2012). For instance, treatment of mice with ProT $\alpha$  resulted in their protection from infections associated with *Candida albicans* (Pan, Haritos et al. 1986). In lymphocytes, it has been demonstrated that ProT $\alpha$  increased antigen induced T cell proliferation, production of interleukins and T cell receptors (Baxevanis, Frillingos et al. 1990; Ioannou, Samara et al. 2012). Amongst its vast functions, in 2005 the Karapetian group ascribed an additional important role for ProT $\alpha$  in the Nrf2-Keap1 signaling pathway (Karapetian, Evstafieva et al. 2005). This group discovered that the Keap1-Cul3-RBX1 ligase complex, required for the ubiquitination of Nrf2, is deviant of a nuclear localization signal (Karapetian, Evstafieva et al. 2005). The Keap1-Cul3-RBX1 ligase complex binds to ProT $\alpha$  and passes the nuclear membrane

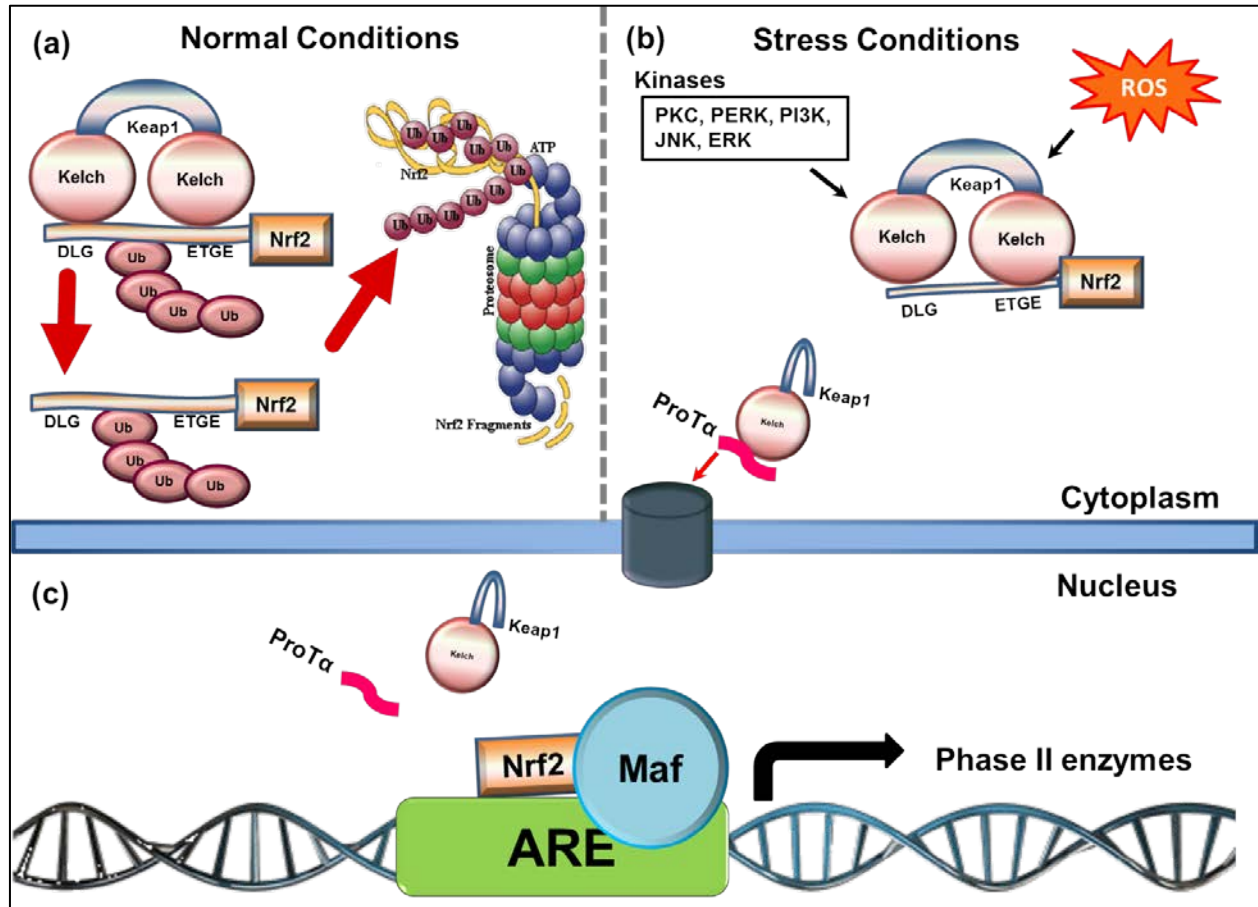


using ProTα nuclear localization signal (Niture and Jaiswal 2009). Since ProTα is one of the major focuses of this study, it is discussed in further detail in Chapter 2.

#### **1.4 Mechanism of Nrf2 transcription Regulation**

Over the last decade scientific research has answered many questions relating to the mechanism by which Nrf2 mediated gene transcription is regulated (Taguchi, Motohashi et al. 2011; Tkachev, Menshchikova et al. 2011; Baird, Lleres et al. 2013). Based on scientific reports, Keap1 appears to be a key negative regulator of the Nrf2 protein involving physical entrapment of the inactive Nrf2 to the cytoplasmic actin filaments under unstressed conditions (Itoh, Wakabayashi et al. 1999; Kang, Kobayashi et al. 2004). During oxidative or electrophilic stress, the Nrf2 liberates from the Keap1 anchorage complex and newly synthesized Nrf2 accumulates in the nucleus leading to Nrf2-ARE transcription of cytoprotective genes (Kobayashi, Kang et al. 2004; Jain and Jaiswal 2006; Kobayashi and Yamamoto 2006). A detailed schematic diagram of the pathway is presented in Figure 1.6.

The current mode of Keap1-Nrf2 interaction is described as a “hinge and latch” model, where two distinct motifs in the Neh2 domain of Nrf2, the DLG and ETGE motifs, are shown to interact with the Kelch domain of Keap1 (Tong, Katoh et al. 2006; Tong, Kobayashi et al. 2006; Tong, Padmanabhan et al. 2007). It has been determined that the immobilization of the Nrf2 by Keap1 occurs in a sequential manner where the high affinity site, ETGE motif, binds first, followed by the docking of the DLG motif onto a second Kelch domain of the homodimer Keap1 (Li, Zhang et al. 2004; Padmanabhan, Tong et al. 2006; Tong, Kobayashi et al. 2006; Tong, Padmanabhan et al. 2007). The ETGE motif locks the Nrf2 in place, whereas, the DLG motif is responsible



**Figure 1.6. Schematic diagram of the Keap1-Nrf2 mediated gene transcription.**

(a) In a regular cell environment, two Keap1 molecules bind to the N-terminal Neh2 domain of the Nrf2 via the DLG and ETGE motif. Nrf2 is polyubiquitinated by the Cul3-based E3 ligase complex and targeted for degradation by the 26S proteasome. (b) During stress conditions, the Nrf2-Keap1 complex is disrupted, leading to accumulation of Nrf2 in the nucleus. (c) The Nrf2 forms a complex with the small Maf proteins and binds to the ARE sequences of target DNA, initiating transcription of Phase II enzymes. Once the cellular environment returns to normal conditions, the nuclear Nrf2 is sequestered by the Keap1 to return Nrf2 concentrations also to normal levels. The Keap1 lacks a nuclear localization signal, and thus is exported to the nucleus in complex with the ProTa. The schematic diagram is based on the work presented in (Taguchi, Motohashi et al. 2011)

for orienting the N-terminal lysine residues of Nrf2 for ubiquitination (Kobayashi, Kang et al. 2004). The presence of stressors triggers conformational changes in the Keap1 (Dinkova-Kostova, Holtzclaw et al. 2002). These structural changes in the Keap1 cause release of the DLG motif; however, the ETGE remains bound. The Nrf2 binds to the Keap1 in this manner until Keap1 is saturated. This results in free cytoplasmic Nrf2 which is transported to the nucleus for the transcription of cytoprotective gene expression (Tong, Katoh et al. 2006).

An alternative manner of Nrf2 regulation by Keap1 has been recently proposed and is called the cyclic model. Based on this model the Nrf2 is not released from the Keap1-Cul3-RBX1 ligase complex; rather, the orientation of the DLG motif is shifted such that the lysine residues required for ubiquitination are no longer accessible by the E2-ubiquitin conjugating enzyme (Baird, Leres et al. 2013). Thus Nrf2 proteosomal degradation is disrupted and newly synthesized Nrf2 escapes into nucleus and transcription of enzymes that allow restoration of cellular environment is initiated (Itoh, Chiba et al. 1997; Itoh, Wakabayashi et al. 1999).

Once the cellular environment is restored, the Nrf2 concentrations in the nucleus need to be regulated (Niture and Jaiswal 2009). Interestingly, the Keap1 lacks a nuclear localization signal and is shuttled to the nucleus in complex with ProTα, where, ProTα dissociates and Nrf2 degradation is initiated again to reestablish homeostatic conditions (Niture and Jaiswal 2009).

## **1.5 Positive and Negative Regulators of Nrf2**

Many cellular mechanisms are in place to regulate Keap1 or Nrf2 in a manner that increases or abolishes the expression of ARE-dependent genes (Tkachev, Menshchikova

et al. 2011). Several proteins stabilizing the Nrf2 transcriptional activity have been reported in literature, including the DJ-1, p21<sup>Cip1/WAF1</sup>, p62 (sequestosome 1), CR6-interacting factor 1 (CRIF1), and PALB2 (Clements, McNally et al. 2006; Chen, Sun et al. 2009; Kang, Hong et al. 2010; Ma, Cai et al. 2012).

The disordered p62 is an adapter protein that carries multiple functions within the cell. It is involved in carrying ubiquitinated proteins to the proteasome, and by products of degradation to the lysosomes (Seibenhener, Geetha et al. 2007). The protein is also required for carrying aggregated, damaged or unfolded/misfolded proteins to the lysosome (Tkachev, Menshchikova et al. 2011; Seibenhener, Geetha et al. 2007). Intriguingly, p62 contains a sequence similar to the ETGE motif of Neh2 and competes for Keap1 binding with the DLG site of Neh2 (Jain and Jaiswal 2006; Cino, Killoran et al. 2013). The p62 is known to target Keap1 for degradation (Taguchi, Fujikawa et al. 2012). As a result, the Nrf2 degradation is halted, and an increase in cytoprotective enzyme expression is observed (Jain and Jaiswal 2006). The Kelch domain is a hub for many other disordered proteins and its interaction with other targets liberates Nrf2 to carry transcriptional activity.

Another alternative mode of Nrf2 stability is adopted by the disordered p21<sup>Cip1/WAF1</sup>. However this disordered protein competes with the Keap1 for binding to the DLG motif of Neh2 domain (Chen, Sun et al. 2009; Tkachev, Menshchikova et al. 2011). By inhibiting Nrf2 ubiquitination, p21 increases the activation of antioxidant defense pathway. p21 is also known to increase cell survival by its involvement in cell cycle arrest and cellular repair processes (Chen, Sun et al. 2009; Taguchi, Motohashi et al. 2011).

Many other mechanisms are in place that regulate the transcriptional activity of Nrf2. Few of these methods include phosphorylation of the Y568 residue of Nrf2 by Tyrosine kinase Fyn in the nucleus resulting translocation of Nrf2 from nucleus to the cytoplasm where it is degraded by Keap1 (Jain and Jaiswal 2007; Tkachev, Menshchikova et al. 2011). Other negative regulators of the pathway compete for binding to the DNA instead. For example, the association of BACH1 with ARE sequences prevents Nrf2 mediated gene expression (Sun, Brand et al. 2004; MacLeod, McMahon et al. 2009).

## **1.6 Dual Nature of Nrf2 in Cancer**

The Keap1-Nrf2 pathway is considered to behave as a “double-edged sword” (Hayes and McMahon 2006). On one side, the expression of detoxifying enzymes helps eliminate ROS, prevents DNA damage and can repress cancer metastasis (Satoh, Moriguchi et al. 2010; Taguchi, Motohashi et al. 2011). On the other hand, constitutive Nrf2 activity in cancer cells provides advantages for tumour cell proliferation (Taguchi, Motohashi et al. 2011). The constitutive cellular activity of Nrf2 arises when it escapes Keap1 facilitated degradation. Many somatic mutations in Nrf2 and Keap1 have been reported in tumor derived cell lines or cancer tissues that results in the disruption of Nrf2 regulation (Padmanabhan, Tong et al. 2006; Hayes and McMahon 2009; Hast, Cloer et al. 2014). Mutations in the Neh2 domain of Nrf2 have been mostly discovered within the ETGE and DLG motifs of the Neh2 domain, resulting in the disruption of Keap1-Nrf2 complex formation (Shibata, Kokubu et al. 2008; Kim, Oh et al. 2010).

Mutations within the Keap1, although reported in all three domains, are predominant in the Kelch domain of Keap1 (Singh, Misra et al. 2006; Ohta, Iijima et al.

2008; Shibata, Kokubu et al. 2008; Network 2012; Yoo, Kim et al. 2012; Hast, Cloer et al. 2014). Unlike the Nrf2, many of the mutations reported in the Kelch domain of Keap1 are away from the binding interface with targets, yet are shown to inhibit Nrf2 regulation. Thus somatic mutations in Keap1 are linked to conformational changes that compromise its ability to ubiquitinate Nrf2 (Lo, Li et al. 2006; Hayes and McMahon 2009). Table 1.1 lists few mutations, the type of cancers they associated with and their location in the Kelch domain.

### **1.7 IDPs in the Nrf2 mediated Oxidative Stress Response**

ProT $\alpha$  and Nrf2 are both intrinsically disordered proteins that play critical roles in the oxidative stress response (Itoh, Chiba et al. 1997; Uversky, Gillespie et al. 1999; Karapetian, Evstafieva et al. 2005; Tong, Katoh et al. 2006; Niture and Jaiswal 2009). As depicted in the crystal structures presented in Figure 1.7 both of these proteins interact with the positively charged loop regions located at the bottom vicinity of the Kelch domain (Lo, Li et al. 2006; Tong, Padmanabhan et al. 2007; Padmanabhan, Nakamura et al. 2008).

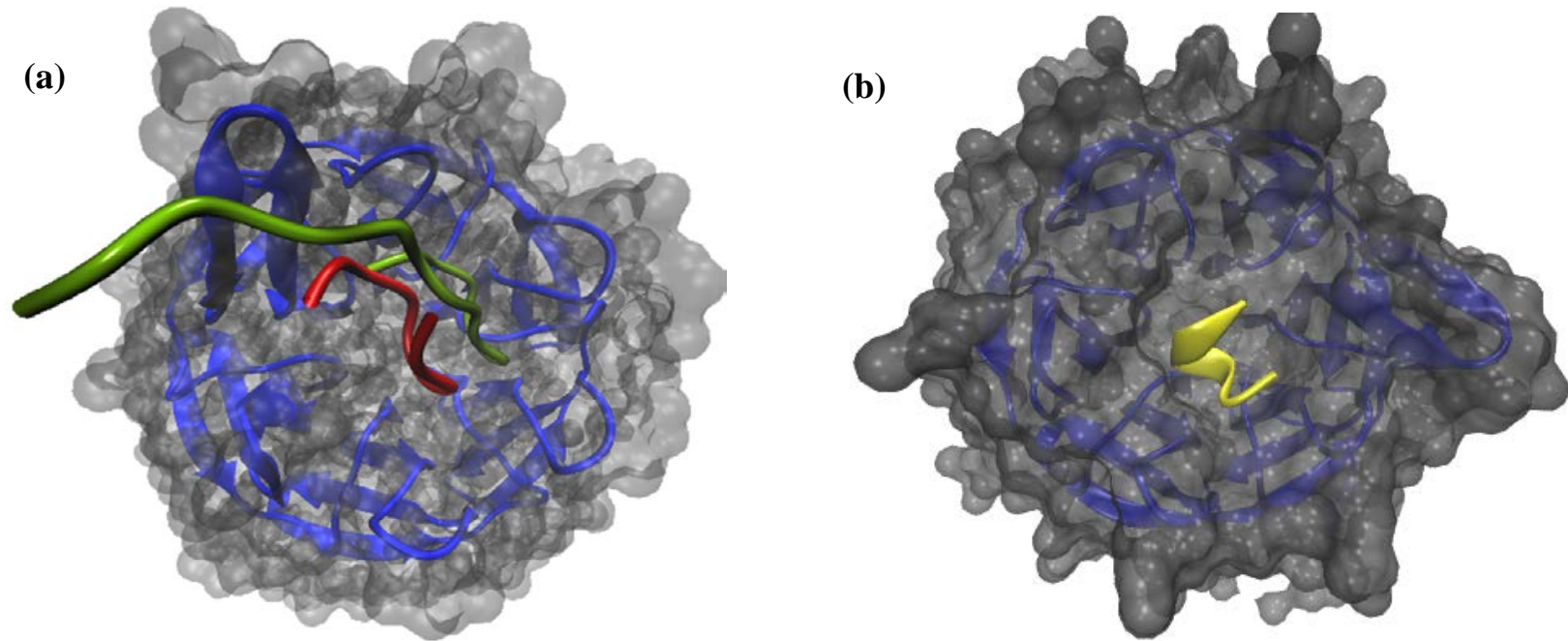
As illustrated by several examples, the biophysics of protein-protein interactions is quite unique for individual systems. Many IDPs undergo disorder to order transitions upon binding to targets, while others remain flexible even in the ligand bound states. Characterizing the binding interface of IDPs in protein-protein interaction can have important implication in drug discovery (Cheng, LeGall et al. 2006). Since a protein-protein interaction involving one structured and one disordered partner seems more

**Table 1.1. A list of somatic mutations identified within the Kelch domain of Keap1** (Forbes, Bindal et al. 2011). \* mutations studied in this work

<b>Residue</b>	<b>Amino Acid Mutation</b>	<b>Reoccurrence</b>	<b>Type</b>	<b>Tissue</b>
<b>Mutations in the binding interface</b>				
<b>334</b>	Y334H		Substitution - Missense	Lung
<b>336</b>	R336		Substitution - Nonsense	Lung
<b>364</b>	G364S		Substitution - Missense	Lung
<b>364</b>	<b>G364C*</b>		Substitution - Missense	Lung
<b>413</b>	<b>R413L*</b>		Substitution - Missense	Lung
<b>415</b>	<b>R415G*</b>		Substitution - Missense	Lung
<b>415</b>	R415C		Substitution - Missense	Lung
<b>460</b>	R460G		Substitution - Missense	Lung
<b>460</b>	R460S		Substitution - Missense	Lung
<b>461</b>	I461V		Substitution - Missense	Lung
<b>479</b>	D479G		Substitution - Missense	Lung
<b>480</b>	G480W		Substitution - Missense	Lung
<b>483</b>	R483C		Substitution - Missense	Lung
<b>509</b>	G509W		Substitution - Missense	Lung
<b>524</b>	G524C		Substitution - Missense	Lung
<b>527</b>	G527F		Substitution - Missense	Lung
<b>554</b>	R554Q		Substitution - Missense	Lung
<b>555</b>	S555C		Substitution - Missense	Lung
<b>556</b>	A556S		Substitution - Missense	Lung
<b>572</b>	Y572C		Substitution - Missense	Lung
<b>601</b>	R601W		Substitution - Missense	Lung
<b>603</b>	G603W		Substitution - Missense	Lung
<b>Mutations in the double GG repeats</b>				
<b>333</b>	G333S		Substitution - Missense	Lung
<b>333</b>	<b>G333C*</b>		Substitution - Missense	Lung
<b>369</b>	V369L		Substitution - Missense	Lung
<b>379</b>	<b>G379D*</b>		Substitution - Missense	Biliary tract
<b>428</b>	V428V		Substitution - coding silent	Pancreas
<b>430</b>	<b>G430C*</b>		Substitution-Missense	Lung
<b>475</b>	V475L		Substitution - Missense	Skin
<b>476</b>	<b>G476R*</b>		Substitution - Missense	Lung
<b>522</b>	A522V		Substitution - Missense	Breast
<b>524</b>	G524C		Substitution - Missense	Lung

<b>Residue</b>	<b>Amino Acid Mutation</b>	<b>Type</b>	<b>Tissue</b>
<b>Other mutations</b>			
<b>324</b>	V324M	Substitution – Missense	Oesophagus
<b>326</b>	R326H	Substitution – Missense	Endometrium
<b>350</b>	<b>G350S*</b>	Substitution-Missense	Gastric, AD
<b>362</b>	R362Q	Substitution – Missense	Lung
<b>389</b>	D389Y	Substitution – Missense	Lung
<b>397</b>	N397	Deletion - In frame	Lung
<b>407</b>	A407V	Substitution - Missense	Prostate
<b>409</b>	M409T	Substitution - Missense	Kidney
<b>417</b>	G417R	Substitution - Missense	Lung
<b>417</b>	G417E	Substitution - Missense	Lung
<b>418</b>	V418M	Substitution - Missense	Lung
<b>418</b>	V418L	Substitution - Missense	Lung
<b>419</b>	G419W	Substitution - Missense	Lung
<b>427</b>	<b>A427V*</b>	Substitution-Missense	Lung
<b>444</b>	E444	Substitution-Nonsense	Kidney
<b>449</b>	E449	Substitution - Nonsense	Breast
<b>452</b>	L452	Deletion - Frameshift	Breast
<b>453</b>	V453	Insertion - Frameshift	Breast
<b>459</b>	R459Q	Substitution - Missense	Skin
<b>470</b>	R470S	Substitution - Missense	Lung
<b>470</b>	R470C	Substitution - Missense	Lung
<b>470</b>	R470H	Substitution - Missense	Lung
<b>471</b>	L471L	Substitution - coding silent	Stomach
<b>488</b>	E488D	Substitution - Missense	Endometrium
<b>493</b>	E493D	Substitution - Missense	Lung
<b>497</b>	W497L	Substitution - Missense	Lung
<b>503</b>	M503K	Substitution - Missense	Lung
<b>506</b>	I506V	Substitution - Missense	Lung
<b>519</b>	I519	Insertion - Frame shift	Endometrium
<b>544</b>	W544R	Substitution - Missense	Kidney
<b>544</b>	W544C	Substitution - Missense	Lung
			Large
<b>558</b>	G558G	Substitution - coding silent	Intestine
<b>559</b>	I559I	Substitution - coding silent	Lung
<b>563</b>	Q563E	Substitution - Missense	Lung
<b>572</b>	Y572C	Substitution - Missense	Lung
<b>579</b>	D579Y	Substitution - Missense	Kidney
<b>584</b>	Y584	Deletion - Frameshift	Kidney
<b>588</b>	T588	Insertion - Frameshift	Lung
<b>593</b>	E593	Substitution - Nonsense	Liver





**Figure 1.7. Crystal structures of the ProT $\alpha$  and Neh2 peptides bound to the Kelch domain of Keap1.**

(a) peptides mimicking binding motifs of ProT $\alpha$  (red) (PDB id: 2Z32) and the high affinity ETGE motif of Neh2 domain (green) (PDB id: 2FLU) bound to the Kelch propeller (Lo, Li et al. 2006; Padmanabhan, Nakamura et al. 2008). (b) The DLG motif of Neh2 bound to the Kelch domain of Keap1 (PDB id: 2DYH) (Tong, Padmanabhan et al. 2007). All three peptides interact with the same surface of the Kelch domain.

promising in therapeutic intervention (Cheng, LeGall et al. 2006; Dunker, Oldfield et al. 2008), understanding the mechanism of binding, structural and dynamic characterization of IDPs in their free and target-bound states is necessary for successful drug design.

Nrf2 and ProT $\alpha$  are great examples to understand the part IDPs play in biological processes as these two proteins have vital roles in cellular maintenance. Their activity in oxidative stress response pathway, especially, the dual nature of Nrf2 in cancers, makes them interesting candidates for delineating the mechanisms by which IDPs function in the cell and thus great drug targets. In this work the nature of Nrf2 and ProT $\alpha$  interactions in full length have been investigated with Keap1. Additionally, target binding with Keap1 has been scrutinized in the presence of somatic mutations associated with cancers (listed in Table 1.1).

## **1.8 Scope of thesis**

The main objectives that are investigated in this work include:

- I. Determining the underlying mechanism of interaction of full length ProT $\alpha$  and Neh2 domain of Nrf2 with Keap1.**
- II. A new high yield purification protocol for the Neh2 domain of human Nrf2 is presented and allows studying full length target binding with the Kelch domain of human Keap1**
- III. Investigation of the role of selected disease associated mutations, reported in Keap1, on target binding to ProT $\alpha$  and Nrf2**

In chapter 1, the interaction occurring between the disordered ProT $\alpha$  and the Kelch domain of human Keap1 has been extensively characterized. ITC, NMR and peptide array techniques were used to delineate the amino acids in ProT $\alpha$  involved in affecting binding to the Kelch domain of human Keap1. Additionally, spin relaxation and amide exchange experiments helped elucidate that the nature of disordered ProT $\alpha$  interaction with the Keap1 as a fuzzy one, where ProT $\alpha$  retains its disordered nature even in target bound states.

A new high yield purification protocol, in chapter 2, provided pure Neh2 sample used for the structural characterization of the full length Neh2 domain of human Nrf2 and for binding studies with the Kelch domain of human Keap1. Higher content of residual structures including  $\alpha$ -helical and  $\beta$ -turn propensities were observed at the N and C-terminals, respectively, in the Neh2 domain. Additionally, NMR studies of the DLG and ETGE motifs in full-length context suggested possible contacts between the N and C terminal residues.

Nine somatic mutations occurring in the Kelch domain of the human Keap1 (highlighted in Table 1.1) were selected and their effects on the structural integrity of Kelch domain and binding to ProT $\alpha$  and Nrf2 were scrutinized. Mutations in the double glycine repeat regions destabilized the Kelch domain and resulted in protein aggregation. Amino acid substitutions in the interface and in other regions did not affect the secondary structural content of the Kelch domain. Nonetheless, mutations in the binding interface compromised complex formation with targets, especially to the DLG motif of Neh2. These results provide new insights into IDP protein-protein interaction and their roles in diseases.

## 1.9 References

- Alam, J., C. Wicks, et al. (2000). "Mechanism of heme oxygenase-1 gene activation by cadmium in MCF-7 mammary epithelial cells. Role of p38 kinase and Nrf2 transcription factor." J Biol Chem **275**(36): 27694-27702.
- Andrews, N. C., H. Erdjument-Bromage, et al. (1993). "Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein." Nature **362**(6422): 722-728.
- Aoki, Y., H. Sato, et al. (2001). "Accelerated DNA adduct formation in the lung of the Nrf2 knockout mouse exposed to diesel exhaust." Toxicol Appl Pharmacol **173**(3): 154-160.
- Baird, L., D. Lleres, et al. (2013). "Regulatory flexibility in the Nrf2-mediated stress response is conferred by conformational cycling of the Keap1-Nrf2 protein complex." Proc Natl Acad Sci U S A **110**(38): 15259-15264.
- Balogun, E., M. Hoque, et al. (2003). "Curcumin activates the haem oxygenase-1 gene via regulation of Nrf2 and the antioxidant-responsive element." Biochem J **371**(Pt 3): 887-895.
- Baxevanis, C. N., S. Frilingos, et al. (1990). "Enhancement of human T lymphocyte function by prothymosin alpha: increased production of interleukin-2 and expression of interleukin-2 receptors in normal human peripheral blood T lymphocytes." Immunopharmacol Immunotoxicol **12**(4): 595-617.
- Bell, S., C. Klein, et al. (2002). "p53 contains large unstructured regions in its native state." J Mol Biol **322**(5): 917-927.
- Bode, A. M. and Z. Dong (2004). "Post-translational modification of p53 in tumorigenesis." Nat Rev Cancer **4**(10): 793-805.
- Brzovic, P. S., C. C. Heikaus, et al. (2011). "The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex." Mol Cell **44**(6): 942-953.
- Calabrese, V., R. Lodi, et al. (2005). "Oxidative stress, mitochondrial dysfunction and cellular stress response in Friedreich's ataxia." J Neurol Sci **233**(1-2): 145-162.
- Chan, J. Y., X. L. Han, et al. (1993). "Cloning of Nrf1, an NF-E2-related transcription factor, by genetic selection in yeast." Proc Natl Acad Sci U S A **90**(23): 11371-11375.
- Chanas, S. A., Q. Jiang, et al. (2002). "Loss of the Nrf2 transcription factor causes a marked reduction in constitutive and inducible expression of the glutathione S-

- transferase Gsta1, Gsta2, Gstm1, Gstm2, Gstm3 and Gstm4 genes in the livers of male and female mice." Biochem J **365**(Pt 2): 405-416.
- Chen, W., Z. Sun, et al. (2009). "Direct interaction between Nrf2 and p21(Cip1/WAF1) upregulates the Nrf2-mediated antioxidant response." Mol Cell **34**(6): 663-673.
- Cheng, M., P. Olivier, et al. (1999). "The p21(Cip1) and p27(Kip1) CDK 'inhibitors' are essential activators of cyclin D-dependent kinases in murine fibroblasts." EMBO J **18**(6): 1571-1583.
- Cheng, Y., T. LeGall, et al. (2006). "Rational drug design via intrinsically disordered protein." Trends Biotechnol **24**(10): 435-442.
- Chevillard, G. and V. Blank (2011). "NFE2L3 (NRF3): the Cinderella of the Cap'n'Collar transcription factors." Cell Mol Life Sci **68**(20): 3337-3348.
- Choy, W. Y. and J. D. Forman-Kay (2001). "Calculation of ensembles of structures representing the unfolded state of an SH3 domain." J Mol Biol **308**(5): 1011-1032.
- Cino, E. A., R. C. Killoran, et al. (2013). "Binding of disordered proteins to a protein hub." Sci Rep **3**: 2305.
- Clements, C. M., R. S. McNally, et al. (2006). "DJ-1, a cancer- and Parkinson's disease-associated protein, stabilizes the antioxidant transcriptional master regulator Nrf2." Proc Natl Acad Sci U S A **103**(41): 15091-15096.
- Cullinan, S. B., D. Zhang, et al. (2003). "Nrf2 is a direct PERK substrate and effector of PERK-dependent cell survival." Mol Cell Biol **23**(20): 7198-7209.
- Dawson, R., L. Muller, et al. (2003). "The N-terminal domain of p53 is natively unfolded." J Mol Biol **332**(5): 1131-1141.
- Deshaies, R. J. (1999). "SCF and Cullin/Ring H2-based ubiquitin ligases." Annu Rev Cell Dev Biol **15**: 435-467.
- Dinkova-Kostova, A. T., W. D. Holtzclaw, et al. (2002). "Direct evidence that sulfhydryl groups of Keap1 are the sensors regulating induction of phase 2 enzymes that protect against carcinogens and oxidants." Proc Natl Acad Sci U S A **99**(18): 11908-11913.
- Dunker, A. K., C. J. Brown, et al. (2002b). "Intrinsic disorder and protein function." Biochemistry **41**(21): 6573-6582.
- Dunker, A. K., C. J. Brown, et al. (2002a). "Identification and functions of usefully disordered proteins." Adv Protein Chem **62**: 25-49.

- Dunker, A. K., M. S. Cortese, et al. (2005). "Flexible nets. The roles of intrinsic disorder in protein interaction networks." FEBS J **272**(20): 5129-5148.
- Dunker, A. K., J. D. Lawson, et al. (2001). "Intrinsically disordered protein." J Mol Graph Model **19**(1): 26-59.
- Dunker, A. K., Z. Obradovic, et al. (2000). "Intrinsic protein disorder in complete genomes." Genome Inform Ser Workshop Genome Inform **11**: 161-171.
- Dunker, A. K., C. J. Oldfield, et al. (2008). "The unfoldomics decade: an update on intrinsically disordered proteins." BMC Genomics **9** Suppl 2: S1.
- Dunker, A. K. and V. N. Uversky. (2010). "Drugs for 'protein clouds': targeting intrinsically disordered transcription factors." Curr Opin Pharmacol **10**(6): 782-788.
- Dyson, H. J. and P. E. Wright. (2002). "Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance." Adv Protein Chem **62**: 311-340.
- Dyson, H. J. and P. E. Wright. (2004). "Unfolded proteins and protein folding studied by NMR." Chem Rev **104**(8): 3607-3622.
- Dyson, H. J. and P. E. Wright. (2005). "Intrinsically unstructured proteins and their functions." Nat Rev Mol Cell Biol **6**(3): 197-208.
- Eliezer, D. (2007). "Characterizing residual structure in disordered protein States using nuclear magnetic resonance." Methods Mol Biol **350**: 49-67.
- Enomoto, A., K. Itoh, et al. (2001). "High sensitivity of Nrf2 knockout mice to acetaminophen hepatotoxicity associated with decreased expression of ARE-regulated drug metabolizing enzymes and antioxidant genes." Toxicol Sci **59**(1): 169-177.
- Evstafieva, A. G., G. A. Belov, et al. (2003). "Apoptosis-related fragmentation, translocation, and properties of human prothymosin alpha." Exp Cell Res **284**(2): 211-223.
- Favreau, L. V. and C. B. Pickett. (1991). "Transcriptional regulation of the rat NAD(P)H:quinone reductase gene. Identification of regulatory elements controlling basal level expression and inducible expression by planar aromatic compounds and phenolic antioxidants." J Biol Chem **266**(7): 4556-4561.
- Finkel, T. and N. J. Holbrook. (2000). "Oxidants, oxidative stress and the biology of ageing." Nature **408**(6809): 239-247.

- Forbes, S. A., N. Bindal, et al. (2011). "COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer." Nucleic Acids Res **39**(Database issue): D945-950.
- Frimpong, A. K., R. R. Abzalimov, et al. (2010). "Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: conformational heterogeneity of alpha-synuclein." Proteins **78**(3): 714-722.
- Fulda, S., A. M. Gorman, et al. (2010). "Cellular stress responses: cell survival and cell death." Int J Cell Biol **2010**: 214074.
- Furukawa, M., T. Ohta, et al. (2002). "Activation of UBC5 ubiquitin-conjugating enzyme by the RING finger of ROC1 and assembly of active ubiquitin ligases by all cullins." J Biol Chem **277**(18): 15758-15765.
- Furukawa, M. and Y. Xiong. (2005). "BTB protein Keap1 targets antioxidant transcription factor Nrf2 for ubiquitination by the Cullin 3-Roc1 ligase." Mol Cell Biol **25**(1): 162-171.
- Fuxreiter, M. and P. Tompa. (2012). "Fuzzy complexes: a more stochastic view of protein function." Adv Exp Med Biol **725**: 1-14.
- Garza, A. S., N. Ahmad, et al. (2009). "Role of intrinsically disordered protein regions/domains in transcriptional regulation." Life Sci **84**(7-8): 189-193.
- Gast, K., H. Damaschun, et al. (1995). "Prothymosin alpha: a biologically active protein with random coil conformation." Biochemistry **34**(40): 13211-13218.
- Gerothanassis, I. P., T. Anastassios, et al. (2002). "Nuclear Magnetic Resonance (NMR) Spectroscopy: Basic Principles and Phenomena, and their Application to Chemistry, Biology and Medicine." Chemistry Education: Research and Practice in Europe **3**(2): 229-252.
- Gomez-Marquez, J., F. Segade, et al. (1989). "The expression of prothymosin alpha gene in T lymphocytes and leukemic lymphoid cells is tied to lymphocyte proliferation." J Biol Chem **264**(15): 8451-8454.
- Gsponer, J., M. E. Futschik, et al. (2008). "Tight regulation of unstructured proteins: from transcript synthesis to protein degradation." Science **322**(5906): 1365-1368.
- Gulbis, J. M., Z. Kelman, et al. (1996). "Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA." Cell **87**(2): 297-306.
- Haarmann, C. S., D. Green, et al. (2003). "The random-coil 'C' fragment of the dihydropyridine receptor II-III loop can activate or inhibit native skeletal ryanodine receptors." Biochem J **372**(Pt 2): 305-316.

- Hammoudeh, D. I., A. V. Follis, et al. (2009). "Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc." J Am Chem Soc **131**(21): 7390-7401.
- Hannappel, E. and T. Huff. (2003). "The thymosins. Prothymosin alpha, parathymosin, and beta-thymosins: structure and function." Vitam Horm **66**: 257-296.
- Haritos, A. A., O. Tsolas, et al. (1984). "Distribution of prothymosin alpha in rat tissues." Proc Natl Acad Sci U S A **81**(5): 1391-1393.
- Hast, B. E., E. W. Cloer, et al. (2014). "Cancer-derived mutations in KEAP1 impair NRF2 degradation but not ubiquitination." Cancer Res **74**(3): 808-817.
- Hayes, J. D. and M. McMahon (2006). "The double-edged sword of Nrf2: subversion of redox homeostasis during the evolution of cancer." Mol Cell **21**(6): 732-734.
- Hayes, J. D. and M. McMahon (2009). "NRF2 and KEAP1 mutations: permanent activation of an adaptive response in cancer." Trends Biochem Sci **34**(4): 176-188.
- Hershey, P. E., S. M. McWhirter, et al. (1999). "The Cap-binding protein eIF4E promotes folding of a functional domain of yeast translation initiation factor eIF4G1." J Biol Chem **274**(30): 21297-21304.
- Hochstrasser, M. (1996). "Ubiquitin-dependent protein degradation." Annu Rev Genet **30**: 405-439.
- Homchaudhuri, L., M. De Avila, et al. (2010). "Secondary structure and solvent accessibility of a calmodulin-binding C-terminal segment of membrane-associated myelin basic protein." Biochemistry **49**(41): 8955-8966.
- Hoshino, H., A. Kobayashi, et al. (2000). "Oxidative stress abolishes leptomycin B-sensitive nuclear export of transcription repressor Bach2 that counteracts activation of Maf recognition element." J Biol Chem **275**(20): 15370-15376.
- Huang, S. and D. E. Ingber. (2002). "A discrete cell cycle checkpoint in late G(1) that is cytoskeleton-dependent and MAP kinase (Erk)-independent." Exp Cell Res **275**(2): 255-264.
- Iakoucheva, L. M., C. J. Brown, et al. (2002). "Intrinsic disorder in cell-signaling and cancer-associated proteins." J Mol Biol **323**(3): 573-584.
- Iakoucheva, L. M., P. Radivojac, et al. (2004). "The importance of intrinsic disorder for protein phosphorylation." Nucleic Acids Res **32**(3): 1037-1049.
- Igarashi, K., K. Kataoka, et al. (1994). "Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins." Nature **367**(6463): 568-572.



- Iizuka, T., Y. Ishii, et al. (2005). "Nrf2-deficient mice are highly susceptible to cigarette smoke-induced emphysema." Genes Cells **10**(12): 1113-1125.
- Inamdar, N. M., Y. I. Ahn, et al. (1996). "The heme-responsive element of the mouse heme oxygenase-1 gene is an extended AP-1 binding site that resembles the recognition sequences for MAF and NF-E2 transcription factors." Biochem Biophys Res Commun **221**(3): 570-576.
- Ioannou, K., P. Samara, et al. (2012). "Prothymosin alpha: a ubiquitous polypeptide with potential use in cancer diagnosis and therapy." Cancer Immunol Immunother **61**(5): 599-614.
- Itoh, K., T. Chiba, et al. (1997). "An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements." Biochem Biophys Res Commun **236**(2): 313-322.
- Itoh, K., N. Wakabayashi, et al. (1999). "Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain." Genes Dev **13**(1): 76-86.
- Jain, A. K. and A. K. Jaiswal. (2006). "Phosphorylation of tyrosine 568 controls nuclear export of Nrf2." J Biol Chem **281**(17): 12132-12142.
- Jain, A. K. and A. K. Jaiswal. (2007). "GSK-3beta acts upstream of Fyn kinase in regulation of nuclear export and degradation of NF-E2 related factor 2." J Biol Chem **282**(22): 16502-16510.
- Jaiswal, A. K. (1991). "Human NAD(P)H:quinone oxidoreductase (NQO1) gene structure and induction by dioxin." Biochemistry **30**(44): 10647-10653.
- Jeffery, C. J. (1999). "Moonlighting proteins." Trends Biochem Sci **24**(1): 8-11.
- Jeffery, C. J. (2004). "Molecular mechanisms for multitasking: recent crystal structures of moonlighting proteins." Curr Opin Struct Biol **14**(6): 663-668.
- Juneja, J. and J. B. Udgaonkar (2003). "NMR studies of Protein Folding." Current Science **84**: 157-172.
- Kang, H. J., Y. B. Hong, et al. (2010). "CR6-interacting factor 1 (CRIF1) regulates NF-E2-related factor 2 (NRF2) protein stability by proteasome-mediated degradation." J Biol Chem **285**(28): 21258-21268.
- Kang, M. I., A. Kobayashi, et al. (2004). "Scaffolding of Keap1 to the actin cytoskeleton controls the function of Nrf2 as key regulator of cytoprotective phase 2 genes." Proc Natl Acad Sci U S A **101**(7): 2046-2051.
- Kansanen, E., S. Kuosmanen, et al. (2013). "The Keap1-Nrf2 pathway: Mechanisms of activation and dysregulation in cancer." Redox Biology **1**: 45-49.

- Karapetian, R. N., A. G. Evstafieva, et al. (2005). "Nuclear oncoprotein prothymosin alpha is a partner of Keap1: implications for expression of oxidative stress-protecting genes." Mol Cell Biol **25**(3): 1089-1099.
- Karetsou, Z., A. Kretsovali, et al. (2002). "Prothymosin alpha interacts with the CREB-binding protein and potentiates transcription." EMBO Rep **3**(4): 361-366.
- Kay, L. E. (2005). "NMR studies of protein structure and dynamics." J Magn Reson **173**(2): 193-207.
- Keum, Y. S., S. Yu, et al. (2006). "Mechanism of action of sulforaphane: inhibition of p38 mitogen-activated protein kinase isoforms contributing to the induction of antioxidant response element-mediated heme oxygenase-1 in human hepatoma HepG2 cells." Cancer Res **66**(17): 8804-8813.
- Kikuchi, N., Y. Ishii, et al. (2010). "Nrf2 protects against pulmonary fibrosis by regulating the lung oxidant level and Th1/Th2 balance." Respir Res **11**: 31.
- Kim, Y. R., J. E. Oh, et al. (2010). "Oncogenic NRF2 mutations in squamous cell carcinomas of oesophagus and skin." J Pathol **220**(4): 446-451.
- Klein, C. and L. T. Vassilev (2004). "Targeting the p53-MDM2 interaction to treat cancer." Br J Cancer **91**(8): 1415-1419.
- Kobayashi, A., E. Ito, et al. (1999). "Molecular cloning and functional characterization of a new Cap'n' collar family transcription factor Nrf3." J Biol Chem **274**(10): 6443-6452.
- Kobayashi, A., M. I. Kang, et al. (2004). "Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2." Mol Cell Biol **24**(16): 7130-7139.
- Kobayashi, M. and M. Yamamoto (2006). "Nrf2-Keap1 regulation of cellular defense mechanisms against electrophiles and reactive oxygen species." Adv Enzyme Regul **46**: 113-140.
- Kover, K. E., M. Bruix, et al. (2008). "The solution structure and dynamics of human pancreatic ribonuclease determined by NMR spectroscopy provide insight into its remarkable biological activities and inhibition." J Mol Biol **379**(5): 953-965.
- Lee, J. M., K. Chan, et al. (2004). "Targeted disruption of Nrf2 causes regenerative immune-mediated hemolytic anemia." Proc Natl Acad Sci U S A **101**(26): 9751-9756.
- Li, X., D. Zhang, et al. (2004). "Crystal structure of the Kelch domain of human Keap1." J Biol Chem **279**(52): 54750-54758.

- Linding, R., R. B. Russell, et al. (2003). "GlobPlot: Exploring protein sequences for globularity and disorder." Nucleic Acids Res **31**(13): 3701-3708.
- Lo, S.-C., X. Li, et al. (2006). "Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling." EMBO J **25**(15): 3605-3617.
- Ma, J., H. Cai, et al. (2012). "PALB2 interacts with KEAP1 to promote NRF2 nuclear accumulation and function." Mol Cell Biol **32**(8): 1506-1517.
- Ma, Q., L. Battelli, et al. (2006). "Multiorgan autoimmune inflammation, enhanced lymphoproliferation, and impaired homeostasis of reactive oxygen species in mice lacking the antioxidant-activated transcription factor Nrf2." Am J Pathol **168**(6): 1960-1974.
- MacLeod, A. K., M. McMahon, et al. (2009). "Characterization of the cancer chemopreventive NRF2-dependent gene battery in human keratinocytes: demonstration that the KEAP1-NRF2 pathway, and not the BACH1-NRF2 pathway, controls cytoprotection against electrophiles as well as redox-cycling compounds." Carcinogenesis **30**(9): 1571-1580.
- Martindale, J. L. and N. J. Holbrook (2002). "Cellular response to oxidative stress: signaling for suicide and survival." J Cell Physiol **192**(1): 1-15.
- McMahon, M., N. Thomas, et al. (2004). "Redox-regulated turnover of Nrf2 is determined by at least two separate protein domains, the redox-sensitive Neh2 degron and the redox-insensitive Neh6 degron." J Biol Chem **279**(30): 31556-31567.
- Mihaela Ilie, D. M., Ed. (2012). Lipid Peroxidation: Trends in the Evaluation of Lipid Peroxidation. Biochemistry, Genetics and Molecular Biology. Chapter 5: 111-130. ISBN 9789535107163.
- Mileo, E., M. Lorenzi, et al. (2013). "Dynamics of the intrinsically disordered protein CP12 in its association with GAPDH in the green alga *Chlamydomonas reinhardtii*: a fuzzy complex." Mol Biosyst **9**(11): 2869-2876.
- Milles, S. and E. A. Lemke (2011). "Single molecule study of the intrinsically disordered FG-repeat nucleoporin 153." Biophys J **101**(7): 1710-1719.
- Minezaki, Y., K. Homma, et al. (2007). "Intrinsically disordered regions of human plasma membrane proteins preferentially occur in the cytoplasmic segment." J Mol Biol **368**(3): 902-913.
- Mittag, T., L. Kay, et al. (2009). "Protein Dynamics and Conformational disorder in molecular recognition." J.Mol. Recognit. **23**(2): 105-16.

- Mittag, T., S. Orlicky, et al. (2008). "Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor." Proc Natl Acad Sci U S A **105**(46): 17772-17777.
- Mohan, A., C. J. Oldfield, et al. (2006). "Analysis of molecular recognition features (MoRFs)." J Mol Biol **362**(5): 1043-1059.
- Moi, P., K. Chan, et al. (1994). "Isolation of NF-E2-related factor 2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the beta-globin locus control region." Proc Natl Acad Sci U S A **91**(21): 9926-9930.
- Moinova, H. R. and R. T. Mulcahy (1998). "An electrophile responsive element (EpRE) regulates beta-naphthoflavone induction of the human gamma-glutamylcysteine synthetase regulatory subunit gene. Constitutive expression is mediated by an adjacent AP-1 site." J Biol Chem **273**(24): 14683-14689.
- Mulcahy, R. T., M. A. Wartman, et al. (1997). "Constitutive and beta-naphthoflavone-induced expression of the human gamma-glutamylcysteine synthetase heavy subunit gene is regulated by a distal antioxidant response element/TRE sequence." J Biol Chem **272**(11): 7445-7454.
- Muto, A., H. Hoshino, et al. (1998). "Identification of Bach2 as a B-cell-specific partner for small maf proteins that negatively regulate the immunoglobulin heavy chain gene 3' enhancer." EMBO J **17**(19): 5734-5743.
- Network, C. G. A. R. (2012). "Comprehensive genomic characterization of squamous cell lung cancers." Nature **489**(7417): 519-525.
- Ng, K. P., G. Potikyan, et al. (2007). "Multiple aromatic side chains within a disordered structure are critical for transcription and transforming activity of EWS family oncoproteins." Proc Natl Acad Sci U S A **104**(2): 479-484.
- Nioi, P., T. Nguyen, et al. (2005). "The carboxy-terminal Neh3 domain of Nrf2 is required for transcriptional activation." Mol Cell Biol **25**(24): 10895-10906.
- Niture, S. K. and A. K. Jaiswal (2009). "Prothymosin- $\alpha$  Mediates Nuclear Import of the INrf2/Cul3{middle dot}Rbx1 Complex to Degrade Nuclear Nrf2." J Biol Chem **284**(20): 13856-13868.
- Niture, S. K. and A. K. Jaiswal (2011). "INrf2 (Keap1) targets Bcl-2 degradation and controls cellular apoptosis." Cell Death Differ **18**(3): 439-451.
- Ohta, T., K. Iijima, et al. (2008). "Loss of Keap1 function activates Nrf2 and provides advantages for lung cancer cell growth." Cancer Res **68**(5): 1303-1309.

- Ohta, T., J. J. Michel, et al. (1999). "ROC1, a homolog of APC11, represents a family of cullin partners with an associated ubiquitin ligase activity." *Mol Cell* **3**(4): 535-541.
- Olashaw, N., T. K. Bagui, et al. (2004). "Cell cycle control: a complex issue." *Cell Cycle* **3**(3): 263-264.
- Oldfield, C. J., Y. Cheng, et al. (2005). "Coupled folding and binding with alpha-helix-forming molecular recognition elements." *Biochemistry* **44**(37): 12454-12470.
- Oldfield, C. J., J. Meng, et al. (2008). "Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners." *BMC Genomics* **9** Suppl 1: S1.
- Oyake, T., K. Itoh, et al. (1996). "Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site." *Mol Cell Biol* **16**(11): 6083-6095.
- Paci, E., M. Vendruscolo, et al. (2002). "Determination of a transition state at atomic resolution from protein engineering data." *J Mol Biol* **324**(1): 151-163.
- Padmanabhan, B., Y. Nakamura, et al. (2008). "Structural analysis of the complex of Keap1 with a prothymosin alpha peptide." *Acta Cryst F* **64**(Pt 4): 233-238.
- Padmanabhan, B., K. I. Tong, et al. (2006). "Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer." *Mol Cell* **21**(5): 689-700.
- Pan, L. X., A. A. Haritos, et al. (1986). "Human prothymosin alpha: amino acid sequence and immunologic properties." *Arch Biochem Biophys* **250**(1): 197-201.
- Perdomo, J., E. L. Fock, et al. (2010). "A monopartite sequence is essential for p45 NF-E2 nuclear translocation, transcriptional activity and platelet production." *J Thromb Haemost* **8**(11): 2542-2553.
- Pierce, M. M., C. S. Raman, et al. (1999). "Isothermal titration calorimetry of protein-protein interactions." *Methods* **19**(2): 213-221.
- Pontius, B. W. and P. Berg (1990). "Renaturation of complementary DNA strands mediated by purified mammalian heterogeneous nuclear ribonucleoprotein A1 protein: implications for a mechanism for rapid molecular assembly." *Proc Natl Acad Sci U S A* **87**(21): 8403-8407.
- Radhakrishnan, I., G. C. Perez-Alvarado, et al. (1998). "Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB." *FEBS Lett* **430**(3): 317-322.
- Radivojac, P., L. M. Iakoucheva, et al. (2007). "Intrinsic disorder and functional proteomics." *Biophys J* **92**(5): 1439-1456.

- Rock, R. S., B. Ramamurthy, et al. (2005). "A flexible domain is essential for the large step size and processivity of myosin VI." Mol Cell **17**(4): 603-609.
- Romero, P., Z. Obradovic, et al. (1997). "Identifying disordered regions in proteins from amino acid sequences." Proc. IEEE Int. Conf. Neural Netw **1**: 90-95.
- Romero, P., Z. Obradovic, et al. (2001). "Sequence complexity of disordered protein." Proteins **42**(1): 38-48.
- Rushmore, T. H., M. R. Morton, et al. (1991). "The antioxidant responsive element. Activation by oxidative stress and identification of the DNA consensus sequence required for functional activity." J Biol Chem **266**(18): 11632-11639.
- Rushmore, T. H. and C. B. Pickett (1990). "Transcriptional regulation of the rat glutathione S-transferase Ya subunit gene. Characterization of a xenobiotic-responsive element controlling inducible expression by phenolic antioxidants." J Biol Chem **265**(24): 14648-14653.
- Sarikas, A., T. Hartmann, et al. (2011). "The cullin protein family." Genome Biol **12**(4): 220.
- Satoh, H., T. Moriguchi, et al. (2010). "Nrf2-deficiency creates a responsive microenvironment for metastasis to the lung." Carcinogenesis **31**(10): 1833-1843.
- Savvides, S. N., S. Raghunathan, et al. (2004). "The C-terminal domain of full-length E. coli SSB is disordered even when bound to DNA." Protein Sci **13**(7): 1942-1947.
- Seibenhener, M. L., T. Geetha, et al. (2007). "Sequestosome 1/p62--more than just a scaffold." FEBS Lett **581**(2): 175-179.
- Shibata, T., A. Kokubu, et al. (2008). "Genetic alteration of Keap1 confers constitutive Nrf2 activation and resistance to chemotherapy in gallbladder cancer." Gastroenterology **135**(4): 1358-1368.
- Singh, A., V. Misra, et al. (2006). "Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer." PLoS Med **3**(10): e420.
- Stellwagen, E., R. Rysavy, et al. (1972). "The conformation of horse heart apocytochrome c." J Biol Chem **247**(24): 8074-8077.
- Sugase, K., H. J. Dyson, et al. (2007). "Mechanism of coupled folding and binding of an intrinsically disordered protein." Nature **447**(7147): 1021-1025.
- Sun, J., M. Brand, et al. (2004). "Heme regulates the dynamic exchange of Bach1 and NF-E2-related factors in the Maf transcription factor network." Proc Natl Acad Sci U S A **101**(6): 1461-1466.

- Sun, Z., Y. E. Chin, et al. (2009). "Acetylation of Nrf2 by p300/CBP augments promoter-specific DNA binding of Nrf2 during the antioxidant response." Mol Cell Biol **29**(10): 2658-2672.
- Sykiotis, G. P. and D. Bohmann (2010). "Stress-activated cap'n'collar transcription factors in aging and human disease." Sci Signal **3**(112): re3.
- Szabo, P., D. Ehleiter, et al. (1992). "Prothymosin alpha expression occurs during G1 in proliferating B or T lymphocytes." Biochem Biophys Res Commun **185**(3): 953-959.
- Taguchi, K., N. Fujikawa, et al. (2012). "Keap1 degradation by autophagy for the maintenance of redox homeostasis." Proc Natl Acad Sci U S A **109**(34): 13561-13566.
- Taguchi, K., H. Motohashi, et al. (2011). "Molecular mechanisms of the Keap1-Nrf2 pathway in stress response and cancer evolution." Genes Cells **16**(2): 123-140.
- Theodore, M., Y. Kawai, et al. (2008). "Multiple nuclear localization signals function in the nuclear import of the transcription factor Nrf2." J Biol Chem **283**(14): 8984-8994.
- Tkachev, V. O., E. B. Menshchikova, et al. (2011). "Mechanism of the Nrf2/Keap1/ARE signaling system." Biochemistry (Mosc) **76**(4): 407-422.
- Todd, M. J., G. H. Lorimer, et al. (1996). "Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism." Proc Natl Acad Sci U S A **93**(9): 4030-4035.
- Tompa, P. (2009). "Structural disorder in amyloid fibrils: its implication in dynamic interactions of proteins." FEBS J **276**(19): 5406-5415.
- Tompa, P. and P. Csermely (2004). "The role of structural disorder in the function of RNA and protein chaperones." FASEB J **18**(11): 1169-1175.
- Tompa, P. and M. Fuxreiter (2008). "Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions." Trends Biochem Sci **33**(1): 2-8.
- Tompa, P., C. Szasz, et al. (2005). "Structural disorder throws new light on moonlighting." Trends Biochem Sci **30**(9): 484-489.
- Tong, K. I., Y. Katoh, et al. (2006). "Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model." Mol Cell Biol **26**(8): 2887-2900.
- Tong, K. I., A. Kobayashi, et al. (2006). "Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism." Biol Chem **387**(10-11): 1311-1320.

- Tong, K. I., B. Padmanabhan, et al. (2007). "Different electrostatic potentials define ETGE and DLG motifs as hinge and latch in oxidative stress response." Mol Cell Biol **27**(21): 7511-7521.
- Uversky, V. N. (2002). "Natively unfolded proteins: a point where biology waits for physics." Protein Sci **11**(4): 739-756.
- Uversky, V. N. (2011). "Intrinsically disordered proteins from A to Z." Int J Biochem Cell Biol **43**(8): 1090-1103.
- Uversky, V. N. (2012). "Intrinsically disordered proteins and novel strategies for drug discovery." Expert Opin Drug Discov **7**(6): 475-488.
- Uversky, V. N. (2013a). "A decade and a half of protein intrinsic disorder: biology still waits for physics." Protein Sci **22**(6): 693-724.
- Uversky, V. N. and A. K. Dunker (2013b). "The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure." F1000 Biol Rep **5**: 1.
- Uversky, V. N., J. R. Gillespie, et al. (2000a). "Why are "natively unfolded" proteins unstructured under physiologic conditions?" Proteins **41**(3): 415-427.
- Uversky, V. N., J. R. Gillespie, et al. (2000b). "Zn(2+)-mediated structure formation and compaction of the "natively unfolded" human prothymosin alpha." Biochem Biophys Res Commun **267**(2): 663-668.
- Uversky, V. N., J. R. Gillespie, et al. (1999). "Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH." Biochemistry **38**(45): 15009-15016.
- Uversky, V. N. and S. Longhi (2011). Instrumental Analysis of Intrinsically Disordered proteins.
- Uversky, V. N., C. J. Oldfield, et al. (2008). "Intrinsically disordered proteins in human diseases: introducing the D2 concept." Annu Rev Biophys **37**: 215-246.
- Vacic, V., C. J. Oldfield, et al. (2007). "Characterization of molecular recognition features, MoRFs, and their binding partners." J Proteome Res **6**(6): 2351-2366.
- Vassilev, L. T. (2004). "Small-molecule antagonists of p53-MDM2 binding: research tools and potential therapeutics." Cell Cycle **3**(4): 419-421.
- Vassilev, L. T., B. T. Vu, et al. (2004). "In vivo activation of the p53 pathway by small-molecule antagonists of MDM2." Science **303**(5659): 844-848.
- Venugopal, R. and A. K. Jaiswal (1996). "Nrf1 and Nrf2 positively and c-Fos and Fra1 negatively regulate the human antioxidant response element-mediated expression



- of NAD(P)H:quinone oxidoreductase1 gene." Proc Natl Acad Sci U S A **93**(25): 14960-14965.
- Venyaminov, S., A. T. Gudkov, et al. (1981). "Absorption and Circular Dichroism Spectra of Individual Proteins from Escherichia coli Ribosomes."
- Vucetic, S., C. J. Brown, et al. (2003). "Flavors of protein disorder." Proteins **52**(4): 573-584.
- Wakabayashi, N., A. T. Dinkova-Kostova, et al. (2004). "Protection against electrophile and oxidant stress by induction of the phase 2 response: fate of cysteines of the Keap1 sensor modified by inducers." Proc Natl Acad Sci U S A **101**(7): 2040-2045.
- Ward, J. J., J. S. Sodhi, et al. (2004). "Prediction and functional analysis of native disorder in proteins from the three kingdoms of life." J Mol Biol **337**(3): 635-645.
- Weikl, T., K. Abelmann, et al. (1999). "An unstructured C-terminal region of the Hsp90 co-chaperone p23 is important for its chaperone function." J Mol Biol **293**(3): 685-691.
- Weinreb, P. H., W. Zhen, et al. (1996). "NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded." Biochemistry **35**(43): 13709-13715.
- Wild, A. C., J. J. Gipp, et al. (1998). "Overlapping antioxidant response element and PMA response element sequences mediate basal and beta-naphthoflavone-induced expression of the human gamma-glutamylcysteine synthetase catalytic subunit gene." Biochem J **332** ( Pt 2): 373-381.
- Wright, P. E. and H. J. Dyson (1999). "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm." J Mol Biol **293**(2): 321-331.
- Wright, P. E. and H. J. Dyson (2009). "Linking folding and binding." Curr Opin Struct Biol **19**(1): 31-38.
- Wu, C. L., A. L. Shiau, et al. (1997). "Prothymosin alpha promotes cell proliferation in NIH3T3 cells." Life Sci **61**(21): 2091-2101.
- Xie, H., S. Vucetic, et al. (2007). "Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins." J Proteome Res **6**(5): 1917-1932.
- Xu, C., X. Yuan, et al. (2006). "Mechanism of action of isothiocyanates: the induction of ARE-regulated genes is associated with activation of ERK and JNK and the phosphorylation and nuclear translocation of Nrf2." Mol Cancer Ther **5**(8): 1918-1926.

- Yi, S., B. L. Boys, et al. (2007). "Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: evidence for metalation as an entropic switch." Biochemistry **46**(45): 13120-13130.
- Yoh, K., K. Itoh, et al. (2001). "Nrf2-deficient female mice develop lupus-like autoimmune nephritis." Kidney Int **60**(4): 1343-1353.
- Yoo, N. J., H. R. Kim, et al. (2012). "Somatic mutations of the KEAP1 gene in common solid cancers." Histopathology **60**(6): 943-952.
- Yu, R., C. Chen, et al. (2000). "Activation of mitogen-activated protein kinase pathways induces antioxidant response element-mediated gene expression via a Nrf2-dependent mechanism." J Biol Chem **275**(51): 39907-39913.
- Yusupov, M. M., G. Z. Yusupova, et al. (2001). "Crystal structure of the ribosome at 5.5 Å resolution." Science **292**(5518): 883-896.
- Zhang, D. D., S. C. Lo, et al. (2004). "Keap1 is a redox-regulated substrate adaptor protein for a Cul3-dependent ubiquitin ligase complex." Mol Cell Biol **24**(24): 10941-10953.
- Zhang, J., T. Hosoya, et al. (2007). "Nrf2 Neh5 domain is differentially utilized in the transactivation of cytoprotective genes." Biochem J **404**(3): 459-466.
- Zhang, O. and J. D. Forman-Kay (1995). "Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer." Biochemistry **34**(20): 6784-6794.
- Zipper, L. M. and R. T. Mulcahy (2000). "Inhibition of ERK and p38 MAP kinases inhibits binding of Nrf2 and induction of GCS genes." Biochem Biophys Res Commun **278**(2): 484-492.
- Zor, T., B. M. Mayr, et al. (2002). "Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators." J Biol Chem **277**(44): 42241-42248.

## Chapter 2

### Fuzzy complex formation between Prothymosin $\alpha$ and the Kelch domain of Keap1

#### 2.1 Introduction

Prothymosin  $\alpha$  (ProT $\alpha$ ) is a 12-kDa acidic protein with multiple biological functions (Haritos, Goodall et al. 1984; Pan, Haritos et al. 1986; Frillingos, Frangou-Lazaridis et al. 1991). It is highly conserved in mammals, is widely distributed in various tissues, and has a distinct amino acid composition (Haritos, Goodall et al. 1984; Pan, Haritos et al. 1986; Frillingos, Frangou-Lazaridis et al. 1991). Its protein sequence lacks aromatic and Cys residues, and about 50% of the sequence is composed of either Glu or Asp. These acidic residues are mostly localized in the center of the protein sequence (Haritos, Goodall et al. 1984; Pan, Haritos et al. 1986; Frillingos, Frangou-Lazaridis et al. 1991; Diaz-Jullien, Perez-Estevez et al. 1996; Segade and Gomez-Marquez 1999). A nuclear localization signal, TKKQK, exists in the C-terminal region (Rubtsov, Zolotukhin et al. 1997; Trumbore, Wang et al. 1997). Early studies of ProT $\alpha$  had shown that the protein is unstructured under non-denaturing conditions but can adopt a partially folded conformation at acidic pH (Gast, Damaschun et al. 1995; Uversky, Gillespie et al. 1999). We have extensively characterized the structural and dynamic properties of this protein under physiological buffer conditions by using nuclear magnetic resonance (NMR) spectroscopy and electrospray ionization mass spectrometry (ESI-MS) (Yi, Boys et al. 2007). The results clearly demonstrated that ProT $\alpha$  is largely disordered and extremely dynamic, with only a slight propensity for  $\beta$ -strand structure. However, the

protein can adopt more compact conformations upon binding to zinc ions (Yi, Boys et al. 2007).

ProT $\alpha$  is an oncoprotein expressed at high levels in proliferating cells. It increases cell proliferation by shortening the G1 phase of the cell cycle (Rodriguez, Vinuela et al. 1998). Studies have shown that in the absence of ProT $\alpha$ , cells undergo apoptosis or are unable to divide (Sburlati, Manrow et al. 1991; Rodriguez, Vinuela et al. 1999; Evstafieva, Belov et al. 2000). ProT $\alpha$  also binds histones (Papamarcaki and Tsolas 1994; Diaz-Jullien, Perez-Estevez et al. 1996; Karetso, Sandaltzopoulos et al. 1998; Martic, Karetso et al. 2005), plays a role in chromatin remodeling (Karetso, Sandaltzopoulos et al. 1998; Karetso, Martic et al. 2004), acts as a transcription activator through CREB binding protein (Karetso, Kretsovali et al. 2002; Subramanian, Hasan et al. 2002), prevents formation of the apoptosome (Qi, Wang et al. 2010; Jiang, Kim et al. 2003; Piacentini, Evangelisti et al. 2003; Malicet, Giroux et al. 2006), and facilitates movement of charged particles within the nucleus and surrounding environment (Enkemann, Ward et al. 2000). Further, the protein suppresses HIV-1 by inducing production of type 1 interferon (Mosoian, Teixeira et al. 2010; Kubota, Adachi et al. 1995; Mosoian, Teixeira et al. 2007), plays a neuroprotective role after ischemic events (Ueda, Fujita et al. 2007; Fujita, Ueda et al. 2009), and functions as a tumor-associated protein in human colon cancer, lung cancer, urinary tract cancer, and breast cancer (Dominguez, Magdalena et al. 1993; Tsitsiloni, Stiakakis et al. 1993; Traub, Jost et al. 2006; Gou, Tong et al. 2009; Tsai, Jou et al. 2009). Recent studies have shown that ProT $\alpha$  also has an important role in the oxidative stress response (Karapetian, Evstafieva et al. 2005; Niture and Jaiswal 2009).

Cells are constantly attacked by reactive oxygen species, electrophiles, carcinogens, and xenobiotics (Itoh, Chiba et al. 1997; Primiano, Sutter et al. 1997). Pathological conditions, such as neurodegenerative diseases, arthritis, cancer, and cardiovascular diseases, can arise if cells fail to counteract oxidative stress by inducing cytoprotective gene expression (Itoh, Chiba et al. 1997; Primiano, Sutter et al. 1997; Pratico, Rokach et al. 2004; Kim and Guengerich 2005; Schulze and Lee 2005). Nuclear factor erythroid 2-related factor 2 (Nrf2) is one of the key regulators of the cellular responses to oxidative stress (Venugopal and Jaiswal 1996; Itoh, Chiba et al. 1997; Itoh, Tong et al. 2004). Nrf2 is a bZIP transcription activator of cytoprotective genes (Venugopal and Jaiswal 1996; Itoh, Chiba et al. 1997). The protein consists of six domains, Neh1 to Neh6 (Itoh, Wakabayashi et al. 1999). The N-terminal Neh2 domain, which is intrinsically disordered (Tong, Katoh et al. 2006), is involved in maintaining homeostatic levels of Nrf2 through binding to Keap1 (Kelch-like ECH-associated protein 1) (Itoh, Wakabayashi et al. 1999).

ProT $\alpha$  plays a crucial role in the Nrf2 signaling pathway by mediating the nuclear import of Keap1/Cul3-RBX1 ligase complex in order to lower Nrf2 levels in the nucleus, allowing the cell to return to normal conditions (Niture, Kaspar et al. 2010; Niture and Jaiswal 2009). Once in the nucleus, ProT $\alpha$  dissociates from the Keap1/Cul3-RBX1 ligase complex, allowing Nrf2 to bind to the complex for degradation (Niture, Kaspar et al. 2010; Niture and Jaiswal 2009).

The interaction between mouse ProT $\alpha$  and Keap1 has been studied previously. Using the yeast two hybrid system and different protein constructs, amino acids 32 to 52

in the mouse ProT $\alpha$  have been identified to be crucial for the interaction with the Kelch domain of Keap1 (Karapetian, Evstafieva et al. 2005). Padmanabhan *et al.* expanded on these findings by structurally analyzing the mouse ProT $\alpha$ -Kelch complex (Padmanabhan, Nakamura et al. 2008). A 16-mer peptide composed of <sup>39</sup>AQNEENGEQEADNEVD<sup>54</sup> of mouse ProT $\alpha$  was crystallized with the Kelch domain (Padmanabhan, Nakamura et al. 2008). The structure reveals that the negatively charged ProT $\alpha$  peptide binds to the basic bottom face of the  $\beta$ -propeller structure of Kelch, forming many electrostatic interactions, especially via the ENGE motif of ProT $\alpha$ . It is noteworthy that the electron density corresponding to residues 49-54 of the ProT $\alpha$  peptide was missing, suggesting this part of the peptide may not adopt a stable conformation upon binding to Kelch (Padmanabhan, Nakamura et al. 2008). Compared to the structures of Neh2-Kelch (Lo, Li et al. 2006; Padmanabhan, Tong et al. 2006; Padmanabhan, Nakamura et al. 2008), the EENGE motif of ProT $\alpha$  and the EETGE motif in the Neh2 domain bind to the same site on the Kelch domain (Padmanabhan, Nakamura et al. 2008).

Although ProT $\alpha$  has multiple biological functions and many targets, very limited structural information about this disordered protein is available, especially concerning its target interactions. The crystallographic study of ProT $\alpha$ -Kelch peptide-protein complex is the only detailed structural characterization of ProT $\alpha$  in a target bound-state form (Padmanabhan, Nakamura et al. 2008). However, the molecular mechanism by which full-length ProT $\alpha$  binds to Kelch remains unclear. Many disordered proteins undergo disorder-to-order transition upon binding to their targets (Wright and Dyson 1999; Dyson and Wright 2005; Wright and Dyson 2009). These protein-protein interactions are

frequently enthalpy driven because the loss of conformational freedom of an intrinsically disordered protein (IDP) upon target binding is usually significant, with a resulting unfavorable entropy change (Wright and Dyson 1999; Bordelon, Montegudo et al. 2004; Fink 2005; Meszaros, Tompa et al. 2007; Sugase, Dyson et al. 2007). IDPs can also interact with targets via preformed structural elements, perhaps lessening the entropy loss and modulating the affinity of binding (Kumar, Ma et al. 2000; Tsai, Ma et al. 2001; Fuxreiter, Simon et al. 2004). Intriguingly, some IDPs remain disordered even in complexes with targets (Fuxreiter 2012 ; Fuxreiter and Tompa 2008; Zhuo, Ilangovan et al. 2010). Due to the crucial roles ProT $\alpha$  and Keap1 have in the oxidative stress response, in this study, experimental approaches were employed to delineate the molecular basis of their binding. The result will not only lead to a better understanding how these two proteins function in the Nrf2 pathway, it will also provide insight into designing drugs that specifically target the interactions between Keap1 and its disordered partners (Cino, Fan et al. 2013; Uversky 2012 ; Wang, Cao et al. 2011 ; Cheng, LeGall et al. 2006; Hammoudeh, Follis et al. 2009).

## **2.2 Materials and Methods**

### ***2.2.1 Protein Expression and Purification***

The expression and purification of human ProT $\alpha$  was reported previously (Yi, Brickenden et al. 2008). The pET15b plasmid carrying the human Kelch cDNA, a kind gift from Dr. Mark Hannink at the University of Missouri-Columbia, was transformed into *E. coli* BL21 (DE3) cells (Novagen). The protein was overexpressed in M9 minimal media and induced by adding 0.5 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG;

Bioshop). After 18 h induction at 16 °C, the bacteria were harvested and stored at -20 °C. The his-tagged protein was purified by affinity chromatography using Ni-Sepharose 6 FF beads (Amersham Biosciences). The eluted protein was dialyzed against TE buffer (20 mM Tris-HCl, 1 mM EDTA, 1 mM DTT, 100 mM NaCl). The His-tag was cleaved with human alpha-thrombin (Haematologic Technologies), and protein was further purified with HiLoad Superdex 75 column (GE Healthcare). The final protein sample was dialyzed in 50 mM sodium phosphate buffer, 100 mM NaCl, 1 mM DTT, pH 7.

### ***2.2.2 Peptide synthesis***

The ProT $\alpha$  peptide (98% in purity),  $^{38}\text{NANEENGEQEA}^{48}$  was ordered from GenScript USA Inc. The lyophilized peptide was dissolved in 50 mM Sodium phosphate buffer, 100 mM NaCl, 1 mM DTT for ITC and NMR experiments.

### ***2.2.3 Site-directed mutagenesis of wild-type ProT $\alpha$***

Wild-type ProT $\alpha$  in pET15b plasmid was subjected to point mutagenesis using the Stratagene Quikchange II site-directed mutagenesis kit. The primers used for each point mutation are included in Supplemental data. Over expression and purification of mutant variants of ProT $\alpha$  were carried out with the similar procedure used for the wild-type ProT $\alpha$  (Yi, Brickenden et al. 2008).

### ***2.2.4 Primers used for point mutations.***

E41P: 5' CCCTGCTAACGGGAATGCTAATCCGGAAAATGGGGAGC 3'

E45G: 5' TAATGAGGAAAATGGGGGGCAGGAGGCTGACAATG 3'



E45A: 5' TAATGAGGAAAATGGGGCGCAGGAGGCTGACAATG 3'

E51A: 5' CAGGAGGCTGACAATGCGGTAGACGAAGAAGAG 3'

E51G: 5' CAGGAGGCTGACAATGGGGTAGACGAAGAAGAG 3'

E51Y: 5'GGGGAGCAGGAGGCTGACAATTACGTAGACGAAGAA 3'

E47A: 5' GAAAATGGGGAGCAGGCGGCTGACAATGAGGTA 3'

E47R: 5' GAGGAAAATGGGGAGCAGCGCGCTGACAATGAGGTAGAC 3'

D49A: 5' GGAGCAGGAGGCTGCCAATGAGGTAGACG 3'

N36A: 5' ATGGAAGAGACGCCCCTGCTGCCGGGAATGCTAATG 3'

G44A: 5' GGAATGCTAATGAGGAAAATGCGGAGCAGGAGGCTG 3'

### **2.2.5 NMR spectroscopy**

All NMR experiments (except the amide hydrogen exchange experiments) were performed on a Varian Inova 600MHz spectrometer equipped with xyz-gradient triple resonance probe at 25 °C in buffer containing 50 mM sodium phosphate, 100 mM NaCl, 1 mM DTT, and 10% (v/v) D<sub>2</sub>O (Cambridge Isotope Laboratories). 1 mM of sodium 2, 2-dimethyl-2-silapentane-5-sulfonate (DSS, Sigma) was added to each NMR sample as an internal standard for the chemical shift referencing (Wishart, Bigam et al. 1995). All NMR data were analyzed with NMRPipe (Delaglio, Grzesiek et al. 1995), and the spectra were analyzed with NMRView (Johnson 2004).

For the NMR titration experiments, unlabeled Kelch was titrated into 600  $\mu\text{L}$  of 150  $\mu\text{M}$   $^{15}\text{N}$ -labeled ProT $\alpha$  until a 1:2 molar ratio (ProT $\alpha$  : Kelch) point was reached.  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were collected for each titration point.

Backbone  $^{15}\text{N}$  longitudinal relaxation rates ( $R_1$ ), relaxation rates in rotating frame ( $R_{1\rho}$ ), and steady-state  $^1\text{H}$ - $^{15}\text{N}$  NOE relaxation parameters of 150  $\mu\text{M}$   $^{15}\text{N}$ -labeled ProT $\alpha$  were measured in both the free state and the Kelch bound state (unlabeled 300  $\mu\text{M}$  Kelch) using identical parameters as described previously (Yi, Boys et al. 2007).

Amide hydrogen exchange (HX) rate measurements were performed at 25  $^\circ\text{C}$  on a Bruker Avance 800 MHz spectrometer equipped with cryogenic probe (National University of Singapore) using the method recently proposed by Fan *et al* (Fan, Lim et al.). The HX rates of  $^{15}\text{N}$ -labeled ProT $\alpha$  (180  $\mu\text{M}$ ) were measured in both the free state and the Kelch-bound state (unlabeled 450  $\mu\text{M}$  Kelch) using the following mixing times: 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 190, 220, 260, and 300 ms. For each mixing time, data sets with and without the suppression of radiation damping were collected in interleaved manner and each FID was acquired with 6 scans. A reference spectrum was collected for each sample using a long interscan delay (10 s) and by removing the first  $^1\text{H}$  pulse and the mixing period in the pulse sequence as mentioned in the reference (Fan, Lim et al. 2011). The data were processed using NMRPipe (Delaglio, Grzesiek et al. 1995). For each isolated peak, signal intensities were fitted to Equation 1 (Equation 9 in the reference (Fan, Lim et al. 2011) in order to determine the exchange rates.

$$\frac{I_{EX}(t)}{I_{ref}} = \frac{f * k_{ex}}{R_{1H} + k_{ex} - R_{1w}} [\exp(-R_{1w}(t - t_0)) - \exp(-(R_{1H} + k_{ex})(t - t_0))] \quad (1)$$

where  $f$  is the fractional steady-state water magnetization;  $R_{1H}$  is the relaxation rate of an amide proton;  $I_{ref}$  is the equilibrium magnetization of the amide proton, which is equal to two times of the peak intensity of the amide in the reference spectrum when it is recorded with the same number of scans as the other spectra;  $I_{EX}(t)$  is the peak intensity of the amide correlation at mixing time  $t$ , which is obtained from the difference spectrum of two interleaved data sets;  $t_0$  is the latency interval; and  $R_{1w}$  is the longitudinal relaxation rate of water. In the data fitting,  $f$ ,  $t_0$ , and  $R_{1w}$  were set to 0.90, 17.4 ms, and  $0.3 \text{ s}^{-1}$ , respectively. The protection factor ( $P$ ) of the backbone amide proton of each amino acid was calculated as the ratio of the intrinsic HX rate ( $k_{int}$ ) of random coil to the experimentally determined HX rate. The  $k_{int}$  values were estimated with SPHERE web server (<http://www.fccc.edu/research/labs/roder/sphere/sphere.html>) (Bai, Milne et al. 1993; Connelly, Bai et al. 1993).

### 2.2.6 Analytical ultracentrifugation

Analytical ultracentrifugation technique was employed to determine the binding stoichiometry of ProT $\alpha$  to the Kelch domain of Keap1. Sedimentation equilibrium experiments were performed in a Beckman-Coulter Optima XL-A analytical ultracentrifuge (Beckman Coulter, Palo Alto, CA) with an AN60Ti rotor. Six-sector Epon-charcoal centerpieces with quartz windows were used. Protein samples were dialyzed in 50 mM phosphate, 100 mM NaCl, 1 mM DTT, at pH 7.0. Data were collected at 22,000 rpm following 24 h equilibration period at 4°C. Kelch absorbance at 280 nm was measured with 10 replicates with a step size of 0.002 cm. Data were fitted

to a single species model using Microsoft Origin 7.0 software. Partial specific volume for Kelch was calculated to be 0.7219 mL/g based on the amino acid composition. The solvent density was measured to be 1.008 g/mL. Protein concentrations were calculated by protein quantification methods, Lowry for ProT $\alpha$  and Bradford assay for Kelch, as well as by amino acid analysis (Advanced Protein Technology Centre, The Hospital for Sick Children, Toronto, ON).

### **2.2.7 Isothermal titration calorimetry (ITC)**

ITC experiments were carried out on a MicroCal<sup>TM</sup> VP-ITC. All purified protein samples were dialyzed into 50 mM sodium phosphate, 100 mM NaCl, 1 mM DTT at pH 7.0. For titrations, purified Kelch and ProT $\alpha$  were prepared to 60  $\mu$ M and 790  $\mu$ M, respectively, in dialysis buffer, and each was degassed prior to titration. Kelch was loaded into the 1.42 mL cell, and ProT $\alpha$  was loaded into the syringe (300  $\mu$ L). Titrations were performed at 25 °C starting with initial injection of 3  $\mu$ L, followed by 59 larger injections of 5  $\mu$ L, with spacing of 360 s (except for the ProT $\alpha$  peptide with spacing of 120 s). The sample cell was stirred at 300 rpm throughout the experiment. The buffer blank was performed under the same conditions and showed negligible heats of binding. The dissociation constant ( $K_d$ ), stoichiometry of binding ( $N$ ), binding enthalpy ( $\Delta H$ ), and entropy ( $\Delta S$ ) were obtained by non-linear least squares fitting of the data to a single-site binding model provided with the data analysis software (Origin 7). Baselines were subtracted from final data using Origin software. All protein concentrations were determined using Lowry (for ProT $\alpha$ ) and Bradford (for Kelch) assays. Protein

concentrations were confirmed by amino acid analysis (Advanced Protein Technology Centre, The Hospital for Sick Children, Toronto, ON).

### **2.2.8 Peptide array**

A twenty-residue,  $^{38}\text{NANEENGEQEADNEVDEEEE}^{57}$ , ProT $\alpha$  peptide spot array was ordered from the Peptide and Peptide Array Synthesis Facility at the University of Western Ontario. The peptide membrane was hydrated, blocked with 5% skim milk, and then probed with the Kelch domain of Keap1. His Probe-HRP was used to detect the ProT $\alpha$  peptide-Kelch interaction. The peptide array membrane was developed with the ECL Lightening Plus kit (Perkin Elmer, Mississauga, ON) before exposing to film.

## **2.3 Results**

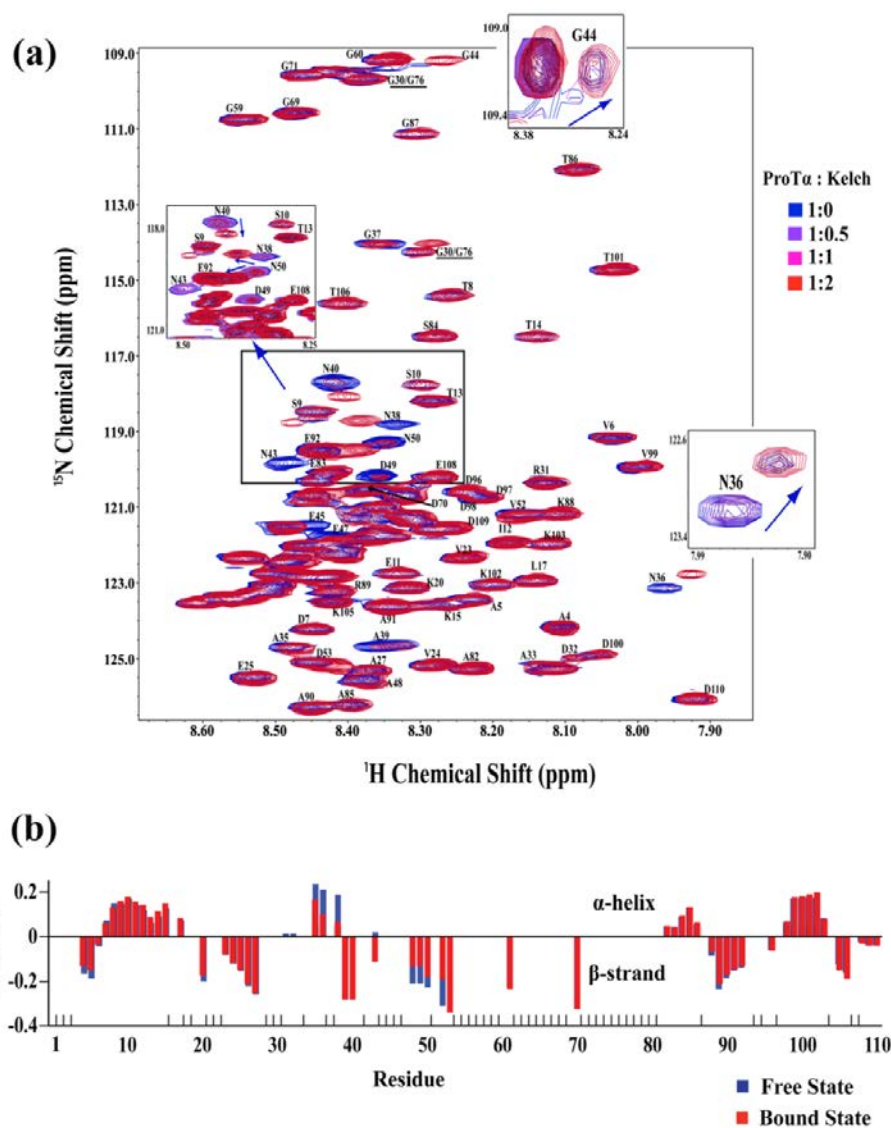
### **2.3.1 Mapping the Kelch-binding region on ProT $\alpha$ by NMR spectroscopy**

The Keap1-binding region on the full-length human ProT $\alpha$  was identified by using NMR spectroscopy. A two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) spectrum of  $^{15}\text{N}$ -labeled ProT $\alpha$  was collected (Figure. 2.1a). The lack of chemical shift dispersion in the  $^1\text{H}$  dimension of the spectrum indicates that the protein was intrinsically disordered in the absence of Kelch. This result is consistent with our previous structural characterization of ProT $\alpha$  carried out in a buffer with lower ionic strength (Yi, Boys et al. 2007).

Changes in the chemical environment of ProT $\alpha$  upon binding to the Kelch domain were then probed by a series of  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments acquired with

increasing concentrations of unlabeled Kelch (Figure 2.1a). Additions of Kelch led to the perturbations of several ProT $\alpha$  backbone  $^1\text{H}/^{15}\text{N}$  resonances. Significant chemical shift changes or attenuations in peak intensity were observed for N36, G37, N38, A39, N40, N43, G44, D49, and N50, as well as several unassigned residues (Yi, Boys et al. 2007). Notably, the dispersion of NMR signals in the  $^1\text{H}$  dimension remained very narrow, indicating that ProT $\alpha$  did not undergo global disorder-to-order transition upon binding to the Kelch domain of Keap1. Aside from the resonance changes mentioned above, most of the peaks in the  $^1\text{H}-^{15}\text{N}$  HSQC spectra remain unchanged in position. This result strongly suggests that the dynamic nature of the rest of ProT $\alpha$  was retained upon target binding.

To identify the conformational changes of ProT $\alpha$  induced by the binding of Kelch, backbone  $^{13}\text{C}\alpha/\beta$  chemical shifts were analyzed using the residue-specific secondary structure propensity (SSP) program developed by Forman-Kay and co-workers (Marsh, Singh et al. 2006). This method combines different chemical shifts and generates a single score for each residue between 1 and -1, indicative of fully formed  $\alpha$ -helical structure and  $\beta$ -strand structure, respectively. SSP scores of ProT $\alpha$  in the absence and presence of Kelch were compared. The result (Figure 2.1b) shows that moderate changes occurred in SSPs of ProT $\alpha$  in the segment spanning residues 36-52; however, the rest of the protein had similar SSP scores in the absence and presence of Kelch. Notably, in the bound state, amino acids in region 35-53 had SSP scores ranging only from -0.34 to 0.17, indicating a lack of stable secondary structure in this region.



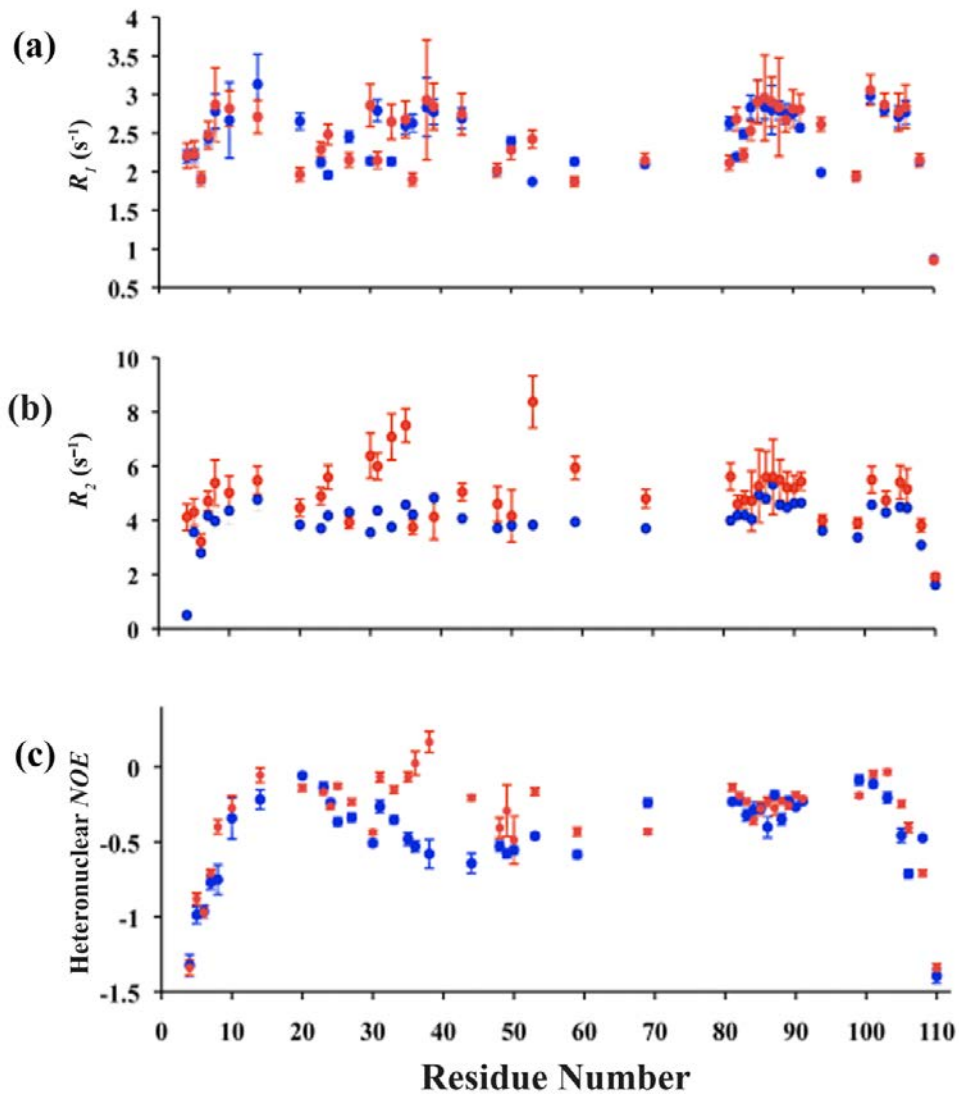
**Figure 2.1. NMR titration experiments of ProT $\alpha$  with Kelch.**  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of ProT $\alpha$  alone (blue) and with 2 equivalents of Kelch domain. Close-up views of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra, showing selected chemical shift changes during the titration, are shown. (b) SSP scores for ProT $\alpha$  in the free (blue) and bound state with Kelch, calculated using  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  chemical shifts. Residues for which SSP scores were not calculated are represented by a tick mark. The regions showing significant changes in SSP correspond to peaks on the HSQC spectrum that undergo large chemical shift changes.

### 2.3.2 *Disordered ProT $\alpha$ forms a fuzzy complex with the Kelch domain of Keap1*

To further characterize the dynamic properties of the ProT $\alpha$ -Kelch complex, backbone  $^{15}\text{N}$  spin relaxation and steady-state  $^1\text{H}$ - $^{15}\text{N}$  NOE experiments were performed on the free and Kelch-bound states of ProT $\alpha$ . These experiments are commonly used in studying protein dynamics and are particularly valuable for probing motions on the ps-ns timescale (Kay 1998). Figure 2.2 shows the  $^{15}\text{N}$   $R_1$ ,  $R_2$ , and steady-state  $^1\text{H}$ - $^{15}\text{N}$  NOEs measured for  $^{15}\text{N}$ -labeled ProT $\alpha$  (150  $\mu\text{M}$ ) in the absence and presence of 300  $\mu\text{M}$  of unlabeled Kelch. Forty-three assigned residues (out of the 110 ProT $\alpha$  residues) with well resolved peaks were included in the data analysis. In the absence of Kelch, all of these 43 residues have  $R_1$  values within a narrow range of 0.87 to 3.13  $\text{s}^{-1}$  (2.45  $\text{s}^{-1}$  in average) (Figure 2.2a). Interestingly, addition of Kelch led to insignificant changes in the  $R_1$  values of ProT $\alpha$  (0.85 – 3.06  $\text{s}^{-1}$ , average = 2.47  $\text{s}^{-1}$ ). In the absence of Kelch, the  $R_2$  values ranged from 0.5 to 5.3  $\text{s}^{-1}$ . Notably, for a globular protein with similar molecular weight of ProT $\alpha$ , residues in structured regions are expected to have  $R_2 \sim 10 \text{ s}^{-1}$  (Walma, Aelen et al. 2004). Upon addition of Kelch, small, yet significant, increases in  $R_2$  values were observed for most of the residues (Figure 2.2b). Larger increases were noted for residues in the region 30-53, which are involved in the interaction with Kelch, as identified by the chemical shift mapping result mentioned in the previous section.

The steady-state  $^1\text{H}$ - $^{15}\text{N}$  NOE is sensitive to fast internal motion, a good indicator of protein mobility on the ps-ns timescale (Peng and Wagner 1992; Kay 1998). Generally, large negative NOE values are observed in flexible regions of a protein (Peng and Wagner 1992). In the free state, NOE values of ProT $\alpha$  were mostly negative, which

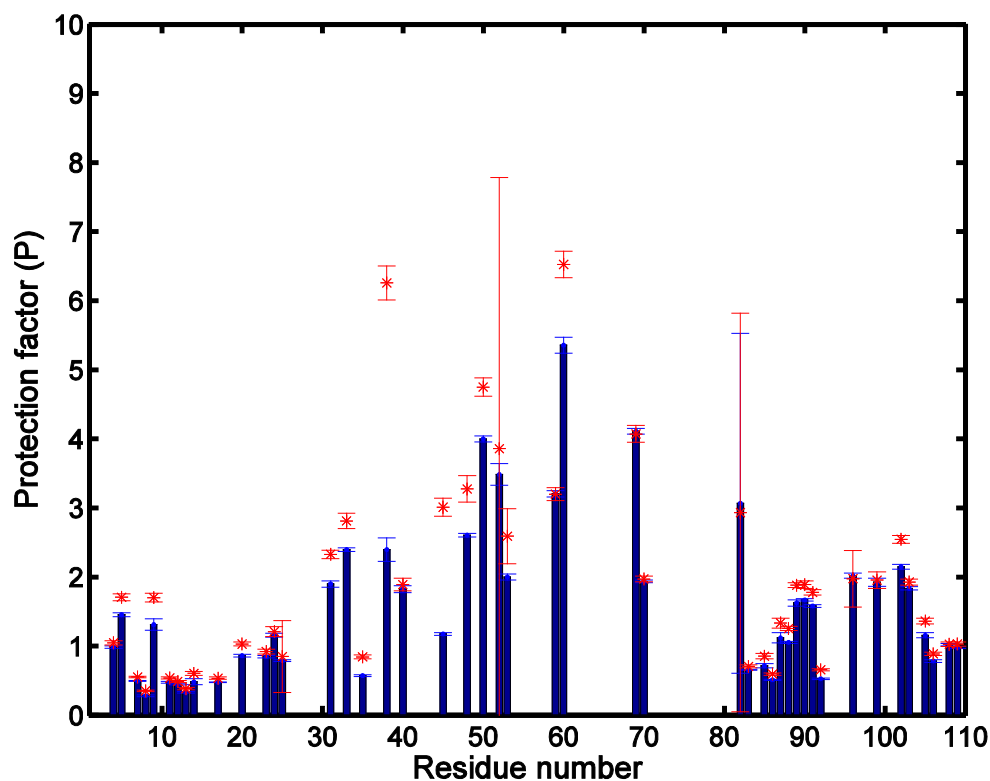




**Figure 2.2.** Backbone  $^{15}\text{N}$  relaxation experiments of ProT $\alpha$  in the free (blue) and bound state with 2 equivalents of Kelch. (a)  $R_1$  (longitudinal) relaxation rate. (b)  $R_2$  (transverse) relaxation rate. (c) steady-state  $^1\text{H}$ - $^{15}\text{N}$  NOE.

is consistent with the disordered nature of this protein (Figure 2.2c). In the Kelch-bound state, increases in the NOE values were observed for residues 30 to 59 (-0.49 to 0.17), indicating that this region became more restricted in motion (Figure 2.2c). Notably, for residues in a structured complex of size ~46 kDa, the NOE values are expected to be ~0.8. Note that the errors of the measurements are significant for those peaks with NOEs close to zero.

To further investigate the nature of the ProT $\alpha$ -Kelch interaction, we conducted NMR amide hydrogen exchange experiments to probe the changes in solvent accessibility of ProT $\alpha$  upon binding to Kelch (Fan, Lim et al. 2011). Protection factor, which is defined as the ratio of the intrinsic to the measured exchange rate, was determined for backbone amides of ProT $\alpha$  (with isolated signals) in the absence and presence of Kelch (Figure 2.3). The low protection factor values of ProT $\alpha$  (0.3 – 5.5; average of 1.6) in the absence of Kelch indicate that the backbone amides were largely exposed to solvent, as expected for a highly unstructured protein like ProT $\alpha$ . Interestingly, no dramatic increases in protection factor values were observed upon addition of Kelch. Even though moderate increases in protection factor values for N38, E45, N50, and G60, were observed, these values were significantly smaller than those observed in folded proteins. For instance, residues located in the secondary structure elements of lysozyme have amide protection factors of  $10^3$ - $10^8$  (Radford, Buck et al. 1992).



**Figure 2.3. Protection factor ( $P$ ) of backbone amides in ProT $\alpha$  in the absence (blue bars) and presence of Kelch (red \*).**

The  $P$  value of each amino acid was calculated as the ratio of the intrinsic HX rate ( $k_{int}$ ) of random coil to the experimentally determined HX rate. The  $k_{int}$  values were estimated using the SPHERE web server

(<http://www.fccc.edu/research/labs/roder/sphere/sphere.html>) (Bai, Milne et al. 1993; Connelly, Bai et al. 1993)

### ***2.3.3 Identifying crucial residues of ProT $\alpha$ for Kelch binding***

Permutation peptide arrays were used to examine the roles that individual residues in ProT $\alpha$  play in the interaction with the Kelch domain. A series of 20-mer peptides, spanning residues 38-57 of human ProT $\alpha$ , were synthesized on a functionalized cellulose membrane and then probed with the Kelch protein. Each of the wild-type residues 38-57 was substituted with any of the 19 amino acids (Figure 2.4). Cys residues were excluded in the permutations to avoid formation of covalent linkages to Kelch during probing, which would result in false positives. Based on the signal intensities observed in the peptide array experiments, mutations that significantly attenuated the binding of the ProT $\alpha$  peptides to Kelch were identified. Many mutations in the <sup>38</sup>NANEENGE<sup>45</sup> region were found to be detrimental to Kelch binding, whereas replacing other residues outside this region had negligible effects on protein-peptide interactions. In particular, mutations of residues E42, G44, and E45 resulted in loss of interaction with Kelch, indicating that these three amino acids are critical for the binding. On the other hand, mutation of N43 in the <sup>42</sup>ENGE<sup>45</sup> motif to a Thr or Ser enhanced the binding to Kelch.

In the peptide array experiment, the E41P mutation also resulted in a higher affinity of binding between Kelch and the ProT $\alpha$  peptide. The side chain of E41 forms hydrogen bonds across the binding interface with Kelch (Padmanabhan, Nakamura et al. 2008). Replacing the glutamate in this position with a non-polar proline was expected to lead to loss of favorable interactions. Therefore, based on the peptide array result, it was unclear why the E41 P mutant of ProT $\alpha$  displayed a higher affinity to Kelch.

AA Position (wt)	Amino acid substituted in ProT $\alpha$ and probed with Kelch  K R H D E N Q S T A G P I L V M F Y W +	Effect of substitution on Kelch binding			Essential for binding
		<i>Weaker</i>	<i>Equal</i>	<i>Stronger</i>	
N38		X	KRFYW		N
A39		HGPA	ILVMFYSW	KRDEQT	A
N40		X	HN		N
E41		NSTGVMYW	X	P	E
E42		X	E		E
N43		X	NH	ST	N
G44		X	G		G
E45		X	E		E
Q46		Y	X		-
E47		X	RDEW		-
A48		Y	X		-
D49		Y	X		-
N50		Y	X		-
E51		Y	X		-
V52		Y	X		-
D53		Y	X		-
E54		Y	X		-
E55		Y	X		-
E56		X	X		-
E57		X	X		-

**Figure 2.4. Permutation peptide array experiments.**

20-mer peptides of ProT $\alpha$  (residues 38-57) were immobilized on a cellulose membrane and then probed with the Kelch domain of Keap1. The left column shows the wild-type sequence and positions of the amino acids that were permuted (mutations specified at the top of second column). The '+' in the second column indicates wt 20-mer sequence and is used as a positive control. Array was repeated twice. Intensity of each permutation was qualitatively measured and compared to wild-type, and results reported are based on what was observed consistently in the two arrays. Columns, on the right side indicate amino acid residues that weaken, strengthen, or have no effect on the binding affinity between the 20-mers and Kelch. As shown in the array, mutations in the EENGE motif are poorly tolerated. AA, amino acid; X, every amino acid except Cys and those mentioned as weaker or stronger; wt, wild-type.

### 2.3.4 Quantitative analysis of ProT $\alpha$ and Kelch interaction

To obtain quantitative information for the ProT $\alpha$ -Kelch interaction, isothermal titration calorimetry (ITC) was used to measure the thermodynamic parameters of binding. Parameters, including the binding stoichiometry ( $n$ ), dissociation constant ( $K_d$ ), enthalpy change ( $\Delta H$ ), and entropy change ( $\Delta S$ ) of binding, were obtained for the wild-type and a series of mutational variants of full-length ProT $\alpha$ . The results are summarized in Table 2.1. To evaluate the contributions of ProT $\alpha$  residues distant from the Kelch binding motif to the affinity of binding, we first compared the affinity of Kelch to the full-length ProT $\alpha$  and an 11-mer peptide spanning residues 38-48 (Padmanabhan, Nakamura et al. 2008). Figure 2.5a & b show the integrated isotherms exhibiting a single-phase curve, which were fitted best to a single-site binding model. The binding stoichiometry obtained for both the peptide and full-length protein are close to 1, indicating 1:1 binding between ProT $\alpha$  and Kelch protein. The result is also in agreement with the stoichiometry derived from sedimentation equilibrium experiments (Figure 2.6 and Table 2.3).

The dissociation constant of the full-length ProT $\alpha$ -Kelch complex was determined to be  $2.6 \pm 0.4 \mu\text{M}$  by ITC. Interestingly, the 11-mer ProT $\alpha$  peptide bound slightly tighter ( $K_d = 1.79 \pm 0.09 \mu\text{M}$ ) to Kelch compared to full-length ProT $\alpha$ . Despite the similar  $K_d$  values, the  $\Delta H$  and  $\Delta S$  of these two binding processes were very different. The results (Table 2.1) indicate a larger entropy loss when the ProT $\alpha$  peptide bound to Kelch. The more negative  $\Delta S$  value was compensated for by the more favorable  $\Delta H$ , resulting in similar  $K_d$  values for both full-length ProT $\alpha$  and its peptide.

**Table 2.1. Thermodynamic parameters for binding of the full-length ProT $\alpha$ , ProT $\alpha$  peptide, and ProT $\alpha$  mutants to Kelch domain of Keap1 at 25 °C (first set)**

WT Kelch	$n$	$K_a$ ( $10^5 \text{ M}^{-1}$ )	$K_d$ ( $\mu\text{M}$ )	$\Delta H$ (kcal/mol)	$T\Delta S$ (kcal/mol)	$\Delta G$ (kcal/mol)
Full length	$1.06 \pm 0.02$	$3.8 \pm 0.6$	$2.6 \pm 0.4$	$-14.8 \pm 0.4$	-7.19	$-7.61 \pm 0.09$
ProT $\alpha$						
ProT $\alpha$ Peptide	$0.93 \pm 0.00(4)$	$5.6 \pm 0.3$	$1.79 \pm 0.09$	$-22.6 \pm 0.1$	-14.76	$-7.84 \pm 0.03$
EENGE						
N36A	$0.95 \pm 0.00(5)$	$1.53 \pm 0.05$	$6.5 \pm 0.2$	$-10.1 \pm 0.6$	-3.03	$-7.07 \pm 0.02$
E41P	$1.10 \pm 0.00(2)$	$19.2 \pm 0.8$	$0.52 \pm 0.02$	$-14.17 \pm 0.04$	-5.60	$-8.57 \pm 0.02$
E47A	$1.08 \pm 0.16$	$0.069 \pm 0.006$	$145 \pm 13$	$-15.2 \pm 0.3$	-9.96	$-5.24 \pm 0.05$
E47R	$0.93 \pm 0.09$	$0.33 \pm 0.03$	$30 \pm 3$	$-13.1 \pm 0.2$	-6.94	$-6.16 \pm 0.05$
D49A	$1.04 \pm 0.02$	$1.5 \pm 0.2$	$6.7 \pm 0.9$	$-12.6 \pm 0.3$	-5.53	$-7.07 \pm 0.08$
E51A	$1.06 \pm 0.01$	$1.14 \pm 0.05$	$8.8 \pm 0.4$	$-12.7 \pm 0.2$	-5.80	$-6.90 \pm 0.03$
E51G	$1.04 \pm 0.01$	$1.40 \pm 0.06$	$7.1 \pm 0.3$	$-15.23 \pm 0.02$	-8.21	$-7.02 \pm 0.02$

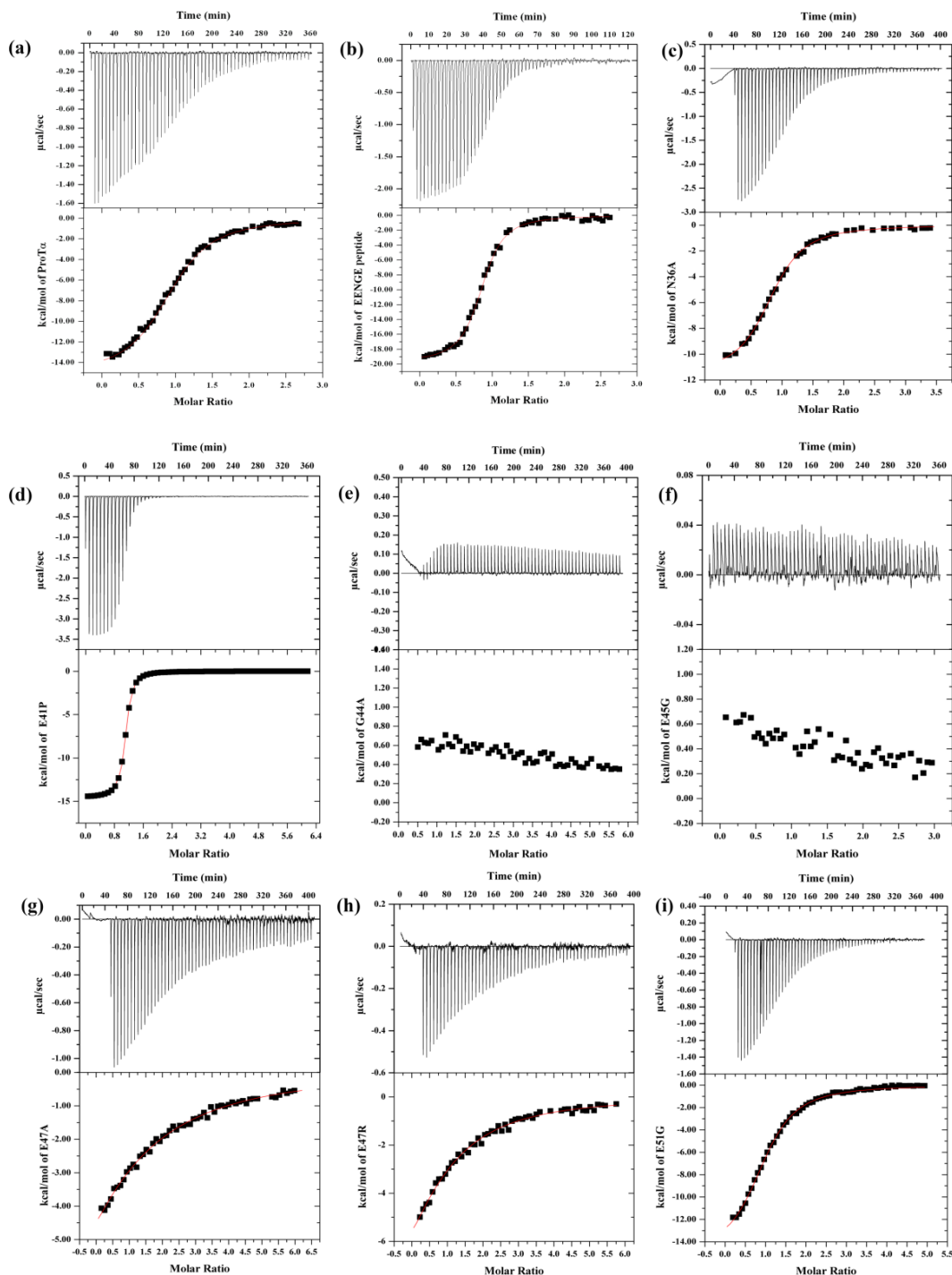
\*  $n$  is the stoichiometry,  $K_a$ , and  $K_d$  are the association and dissociation constants, respectively.  $\Delta G = -RT\ln K_a$ , where  $T$  is the temperature in Kelvin,  $R$  is the gas constant.  $\Delta H$  and  $\Delta S$  are the enthalpy and entropy changes, respectively. Each experiment was repeated twice; the values obtained between the two runs were similar. Values listed in the tables are based on one run, and the data for the duplicate run is provided in the supplementary data (Table 2.2).

**Table 2.2. Thermodynamic parameters for binding of the ProT $\alpha$  peptide, full-length ProT $\alpha$  and its mutants to Kelch domain of Keap1 at 25 °C (second set)**

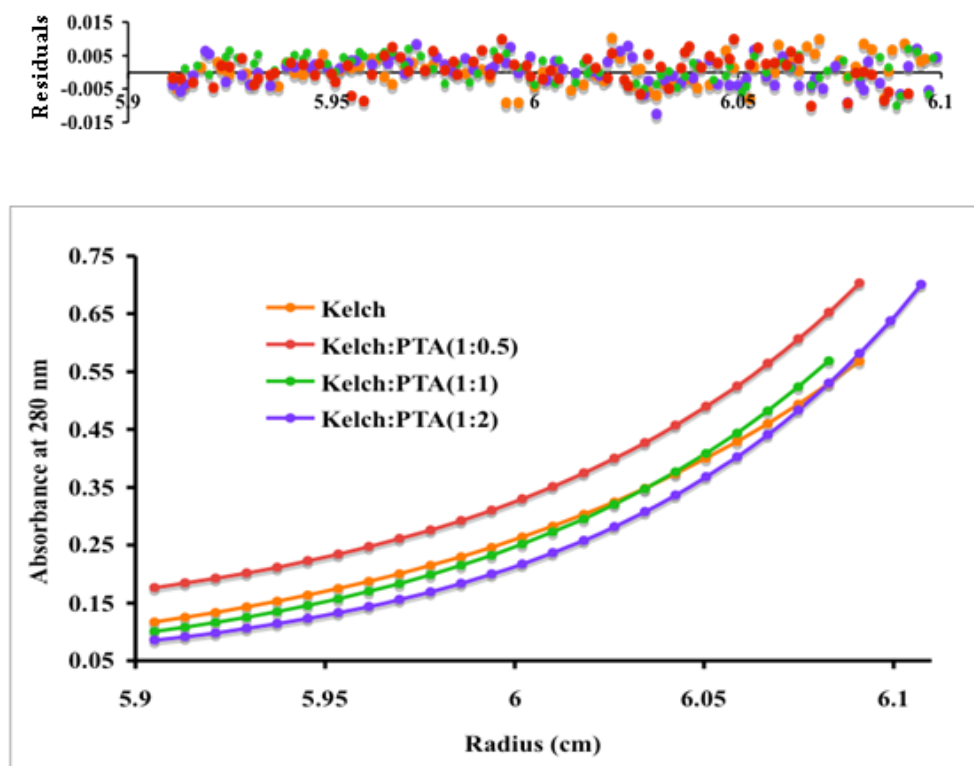
WT Kelch	$n$	$K_a$ ( $10^5 \text{ M}^{-1}$ )	$K_d$ ( $\mu\text{M}$ )	$\Delta H$ (kcal/mol)	$T\Delta S$ (kcal/mol)	$\Delta G$ (kcal/mol)
Full length ProT $\alpha$	$1.04 \pm 0.02$	$3.3 \pm 0.3$	$3.04 \pm 0.2$	$-15.5 \pm 0.4$	-7.97	$-7.53 \pm 0.05$
ProT $\alpha$ Peptide EENGE	$1.13 \pm 0.004$	$5.7 \pm 0.3$	$1.74 \pm 0.08$	$-23.1 \pm 0.1$	-15.24	$-7.86 \pm 0.03$
N36A	$0.95 \pm 0.04$	$1.6 \pm 0.5$	$6.1 \pm 1.8$	$-9.9 \pm -0.5$	-2.8	$7.1 \pm 0.2$
E41P	$1.16 \pm 0.002$	$20.3 \pm 0.8$	$0.49 \pm 0.02$	$-15.05 \pm 0.05$	-6.45	$-8.60 \pm 0.02$
E47A	$1.10 \pm 0.12$	$0.061 \pm 0.002$	$164 \pm 5$	$-15.9 \pm 0.2$	-10.74	$-5.16 \pm 0.02$
E47R	$0.91 \pm 0.04$	$0.303 \pm 0.02$	$33 \pm 2$	$-13.20 \pm 0.08$	-7.09	$-6.11 \pm 0.03$
D49A	$1.06 \pm 0.05$	$2.2 \pm 0.8$	$4.5 \pm 1.6$	$-12.0 \pm 0.7$	-4.7	$-7.3 \pm 0.2$
E51A	$0.95 \pm 0.01$	$1.18 \pm 0.06$	$8.5 \pm 0.4$	$-13.2 \pm 0.2$	-6.28	$-6.92 \pm 0.03$
E51G	$1.05 \pm 0.01$	$1.49 \pm 0.05$	$6.7 \pm 0.2$	$-15.63 \pm 0.02$	-8.57	$-7.06 \pm 0.02$

\*  $n$  is the stoichiometry,  $K_a$ , and  $K_d$  are the association and dissociation constants, respectively.  $\Delta G = -RT \ln K_a = \Delta H - T\Delta S$ , where  $T$  is the temperature and  $R$  is the gas constant.  $\Delta H$  and  $\Delta S$  are the enthalpy and entropy changes, respectively.





**Figure 2.5. ITC profiles of Keap1 with full length ProT $\alpha$  and its mutant variants.**  
 (a) Full length wild-type ProT $\alpha$  (b) EENGE ProT $\alpha$  11-mer peptide (c) N36A (d) E41P (e) G44A (f) E45G (g) E47A (h) E47R (i) E51G



**Figure 2.6. Sedimentation equilibrium analysis of Kelch and ProT $\alpha$  binding.** Pure Kelch protein at a concentration of 10  $\mu$ M was centrifuged with various concentrations of ProT $\alpha$  (to achieve a ratio of 1:0.5, 1:1, and 1:2 with the Kelch domain) at 4°C at 22 000 rpm. The upper panel represents the residual values for the fits.

**Table 2.3. Molecular weight values of Kelch calculated from the sedimentation equilibrium experiment.**

<b>Protein</b>	<b>Concentration (<math>\mu\text{M}</math>)</b>	<b>Calculated MW (Da)</b>	<b>Observed MW (Da)</b>
Kelch	10	33915.85	34565 $\pm$ 338
Kelch	10	39952.76	41551 $\pm$ 331
ProT $\alpha$	5		
Kelch	10	45989.67	44978 $\pm$ 396
ProT $\alpha$	10		
Kelch	10	45989.67	46793 $\pm$ 379
ProT $\alpha$	20		

The charge complementarity between ProT $\alpha$  and Kelch is crucial for their interaction. The highly acidic Kelch-binding motif of ProT $\alpha$  binds to the positively charged bottom face of the  $\beta$ -propeller structure of Kelch, forming multiple salt bridges and hydrogen bonds across the interface (Padmanabhan, Nakamura et al. 2008). As aforementioned, Karapetian et al. demonstrated that the E45G/E51G double mutant of ProT $\alpha$  is not capable of binding to Kelch (Karapetian, Evstafieva et al. 2005). To further delineate the roles that E45 and E51 play in the interface, binding affinities of the E45G, E45A, E51A, and E51G full-length ProT $\alpha$  mutants to Kelch were measured. The results show that E45 is critical for the interaction with Kelch, as the mutation of this residue to Ala or Gly completely abolished the binding to Kelch (Table 2.1 and Figure 2.5f). This observation is consistent with the peptide array result and is in agreement with the crystallographic data (Padmanabhan, Nakamura et al. 2008), which illustrate that the side-chain of E45 forms hydrogen bonds with multiple amino acids on Kelch. Mutations of E51 (Table 2.1 and Figure 2.5i), on the other hand, had little effect on the binding affinity to Kelch, indicating that this acidic residue plays only a minor role in mediating the interaction of these two proteins. The peptide array result clarifies that E45, but not E51, is indispensable for the protein-protein interaction.

The peptide array result also indicates that the mutation of E47, which is located close to the <sup>42</sup>ENGE<sup>45</sup> motif, to any other amino acid (except for R, D, and W) leads to a decrease in binding affinity to Kelch. Based on the charge complementarity at the ProT $\alpha$  and Kelch interaction sites, the result of the E47R mutation is unexpected. To validate this finding, binding affinities of E47R and E47A to Kelch were determined by ITC (Table 1 and Figure 2.5g & 2.5h). Results show that while the Ala mutation caused a 55-

fold decrease in the binding affinity to Kelch, the increase in  $K_d$  for E47R was less significant ( $K_d = 30 \mu\text{M}$  compared to  $2.6 \mu\text{M}$  for the wild-type).

E41 is another acidic residue that showed an interesting result in the peptide array experiment. The crystal structure of the mouse ProT $\alpha$ -Kelch peptide-protein complex reveals that the side-chain of E41 is involved in hydrogen bonding with Kelch (Padmanabhan, Nakamura et al. 2008). However, mutation of this acidic amino acid to a non-polar proline residue increased the binding affinity to Kelch (Table 2.1 & Figure 2.5d). ITC measurements confirmed that the E41P mutant has a more favorable  $\Delta G_{bind}$  compared to the wild-type ProT $\alpha$  (-8.57 kcal/mol vs. -7.61 kcal/mol; Table 2.1), mainly due to the less negative value of the  $\Delta S$  component.

Based upon the peptide array result, G44 is another residue that is essential for binding Kelch. To confirm this finding, the binding affinity of the G44A mutant to the Kelch domain was determined. As expected, no interaction of this mutant with Kelch was detected by ITC (Figure 2.5e). Nevertheless, this specific mutation was very helpful in validating the backbone assignments around this amino acid. The other two mutants that were constructed for the same purpose were N36A and D49A. Both mutational variants bound to Kelch with moderately lower binding affinities (with  $K_d$  of 6.5 and 6.7  $\mu\text{M}$ , respectively), compared to wild-type ProT $\alpha$  (Table 2.1 & Figure 2.5c). Although the N36 was excluded from the crystallographic study of the peptide-protein complex of ProT $\alpha$  and Kelch, and the electron density of D49 was absent in the crystal structure solved (Padmanabhan, Nakamura et al. 2008), both the NMR chemical shift perturbations

and ITC results indicate that these two flanking residues also contribute moderately to the binding of ProT $\alpha$  and Kelch.

## 2.4 Discussion

To better understand the molecular mechanism underlying the oxidative stress response, the intrinsically disordered ProT $\alpha$  interaction with the Kelch domain of Keap1 was extensively studied. Many examples of IDPs are provided in literatures that undergo disorder-to-order transition upon binding to their targets through the coupled binding and folding process (Sugase, Dyson et al. 2007; Wright and Dyson 2009). Remarkably, our studies of the ProT $\alpha$ -Kelch complex have revealed that ProT $\alpha$  retains its disordered nature, even in the bound state. This finding is supported by the observations that the resonance signals of ProT $\alpha$  in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments were clustered in a narrow range from 7.9 ppm to 8.6 ppm, both in the free and bound states (Figure 2.1a). Significant chemical shift changes were observed only for amino acids 35-50. Most of these residues have also been identified as crucial in mediating the interaction between the mouse homologs of ProT $\alpha$  and Kelch (Karapetian, Evstafieva et al. 2005; Padmanabhan, Nakamura et al. 2008). The crystal structure of mouse Kelch in complex with a ProT $\alpha$  peptide spanning amino acids 39-54 (corresponding to residues 38-53 in human) reveals that this particular segment of ProT $\alpha$  forms a  $\beta$ -turn conformation upon binding to the bottom side of the  $\beta$ -propeller structure of Kelch (Padmanabhan, Nakamura et al. 2008). The binding interface is found to be highly complementary in charge, with the acidic amino acids in ProT $\alpha$  playing crucial roles in interacting with multiple Arg residues located on the binding surface of Kelch

(Padmanabhan, Nakamura et al. 2008). In particular, the EENGE motif in ProT $\alpha$  (residues 41-45 in human ProT $\alpha$ ) was shown to be essential for the Kelch binding (Padmanabhan, Nakamura et al. 2008). Our NMR titration results are in agreement with these previous findings, and importantly, we show that several other residues flanking the EENGE motif also display attenuation in resonance signals, indicating they may also be involved in Kelch binding.

Compared to the changes in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC of  $^{15}\text{N}$ -labeled Neh2 upon binding to Kelch, the spectral changes observed for free and bound ProT $\alpha$  were less dramatic. Tong *et al.* have demonstrated that sharp resonance signals of disordered Neh2 were severely broadened out upon addition of Kelch (Tong, Katoh et al. 2006). In the presence of a two molar ratio of Kelch, most of the resonance signals of Neh2 disappeared, likely due to the significant increase in molecular tumbling time through the formation of a large molecular weight complex or as the result of intermediate conformational exchange between the free and bound states. In contrast, the sharpness of resonance signals prevailed throughout the NMR titration experiments for the majority of residues in ProT $\alpha$ . Moreover, spin relaxation experiments showed that addition of Kelch led to insignificant changes in the  $R_1$  values of ProT $\alpha$  ( $0.85 - 3.06 \text{ s}^{-1}$ , average =  $2.47 \text{ s}^{-1}$ ). These observed  $R_1$  values are systematically larger than expected for a folded protein with similar molecular weight of ProT $\alpha$  (van Ingen, Baltussen et al. 2006; Kim, Fuzery et al. 2009). For instance, the D39A mutant of the 128-residue apo-IscU has an averaged  $R_1$  value of  $1.64 \text{ s}^{-1}$  measured under the same magnetic field strength and temperature (i.e., 600 MHz; 298 K), while the averaged  $R_1$  value of the 105-residue PAH2B domain of mSin3B is  $1.45 \text{ s}^{-1}$  (at 293 K) (van Ingen, Baltussen et al. 2006; Kim, Fuzery et al. 2009).

The higher  $R_1$  values for ProT $\alpha$  are the result of a large amplitude of internal motion on the ps-ns timescale in the protein. Meanwhile, small, yet significant, increases in  $R_2$  values were observed for most of the residues (Figure 2.2b) upon addition of Kelch, particularly in the region 30-53. The increases in  $R_2$  suggest that this particular region became more restricted in motion upon binding to Kelch (Kay 1998). However, the observed  $R_2$  values are still significantly smaller than expected for a 46 kDa (the sum of molecular weights of ProT $\alpha$  and Kelch) complex (Walma, Aelen et al. 2004). Further, the NOE values of ProT $\alpha$  remained predominantly negative upon binding to Kelch, indicative of a flexible protein lacking secondary structure (Figure 2.2c). NOEs of residues in the region directly involved in forming the binding interface with Kelch became positive (increases up to 0.17), but these positive values were still significantly lower than expected for a 46-kDa complex with well-defined structure. Moreover, the small protection factors determined in the amide exchange experiments are in agreement with the low  $R_2$  and small NOE values observed, suggesting that ProT $\alpha$  forms a dynamic complex with Kelch.

Peptide array and ITC experiments were conducted to elucidate further the roles individual amino acids in the binding region of ProT $\alpha$  play in the ‘fuzzy’ interaction with the Kelch domain. The results are compared to the Neh2-Kelch binding. The binding of Nrf2 to Keap1 has been described as a “hinge and latch” mechanism (Tong, Kobayashi et al. 2006). The disordered Neh2 domain binds to two Kelch domains of a Keap1 dimer via the N-terminal DLG and the C-terminal ETGE motifs, respectively. It has been proposed that the complex formation facilitates the ubiquitination of a Lys residue located between the DLG and ETGE motifs in sequence, leading to the degradation of



Nrf2 (Tong, Katoh et al. 2006; Tong, Kobayashi et al. 2006). Previous studies have demonstrated that the ETGE motif of Nrf2 has a higher binding affinity ( $K_{a, \text{ETGE}} = 12.4 \times 10^7 \text{ M}^{-1}$ ) compared to the DLG ( $K_{a, \text{DLG}} = 2 \times 10^6 \text{ M}^{-1}$ ) motif for the Kelch domain (Tong, Katoh et al. 2006; Tong, Kobayashi et al. 2006). Our ITC results revealed that the binding affinity of ProT $\alpha$  ( $K_a = (3.8 \pm 0.6) \times 10^5 \text{ M}^{-1}$ ) was also significantly weaker than that of the Neh2 ETGE motif. Tong *et al.* reported that the  $\Delta H$  and  $T\Delta S$  of binding between Kelch and mouse Neh2 [ $\Delta 1-33$ ] with the low affinity DLG motif deleted were -21.2 and -10.2 kcal/mol, respectively (Tong, Katoh et al. 2006). Large and negative values of  $\Delta S$  are frequently observed for binding involving IDPs mainly due to the significant reduction in the conformational entropy (Meszaros, Tompa et al. 2007). Interestingly, compared to the thermodynamic parameters obtained in this study, the entropic penalty was lower for the dynamic ProT $\alpha$ -Kelch complex formation (Table 2.1). On the other hand, the interaction between ProT $\alpha$  and Kelch was weaker based on the values of the enthalpy changes.

Our peptide array result showed evidently that the <sup>42</sup>ENGE<sup>45</sup> motif is critical for binding, since mutations in this region were disruptive. However, exceptions were seen. Mutation of N43 to a Thr or Ser enhanced the binding to Kelch. Intriguingly, with the ETGE motif, the Neh2 domain of Nrf2 has been shown to have a greater affinity for the Kelch domain (Tong, Katoh et al. 2006). Therefore, conversion of the ENGE to an ETGE motif in ProT $\alpha$  resulting in a higher binding affinity is not unexpected. Based on the crystal structures of ProT $\alpha$ -Kelch and Neh2-Kelch peptide-protein complexes, both the Asn in ProT $\alpha$  and the Thr in Neh2 are not directly involved in the interaction with Kelch.

Instead, the side-chains of each of these two amino acids play important roles in stabilizing the  $\beta$ -turn conformation of the binding motif by forming hydrogen bonds with the peptide backbone (Tong, Katoh et al. 2006; Padmanabhan, Nakamura et al. 2008). This finding explains why the N43S mutant of ProT $\alpha$  also showed Kelch binding affinity that is similar to N43T, as the hydroxyl group of Ser may still participate in the intramolecular hydrogen bonding (Tong, Katoh et al. 2006; Padmanabhan, Nakamura et al. 2008). Notably, hydrogen bonds between the N/N and D/T pairs in the  $^{40}\text{NEEN}^{43}$  and  $^{77}\text{DEET}^{80}$  motifs of ProT $\alpha$  and Neh2, respectively, also contribute to stabilization of the  $\beta$ -turn conformation of these proteins in their free states (Cino, Choy et al. 2012; Cino, Wong-ekkabut et al. 2011).

Meanwhile, the binding affinity between ProT $\alpha$  and Kelch increased by 5-fold when E41, the residue N-terminal to the ENGE motif, of ProT $\alpha$  was mutated to a Pro residue ( $K_a = 19.2 \pm 0.8 \times 10^5 \text{ M}^{-1}$ ). One possible explanation for the higher affinity of the ProT $\alpha$  E41P mutant is that the conformational entropy of the free state is reduced, lowering the amount of entropic penalty upon binding. The  $^{40}\text{NEENGE}^{45}$  region of ProT $\alpha$  adopts  $\beta$ -turn conformations when bound to the Kelch domain (Komatsu, Kurokawa et al. 2010; Padmanabhan, Nakamura et al. 2008). Pro residues are commonly found in position  $i+1$  of  $\beta$ -turns in proteins (Hutchinson and Thornton 1994). The cyclic side chain of Pro leads to drastic restriction in the sampling of backbone dihedral angles. MD simulation of the E41P peptide (in its free state) showed that it has a high tendency to form the bound-state-like  $\beta$ -turn structure (Khan, 2013). Therefore, it seems plausible that the incorporation of a Pro at this position leads to a less severe conformational

entropy loss upon binding, resulting in a higher binding affinity. Indeed, lowered free state conformational entropy is thought to play an important role in the high affinity of the Neh2-Kelch domain interaction (Cino, Wong-ekkabut et al 2011. ; Tong, Kobayashi et al. 2006). Another possible explanation is that, since E41 is involved in forming salt bridges and hydrogen bonds across the binding interface with Kelch, mutating this residue to a non-polar Pro would result in loss of these interactions. The change would render this region of the protein more flexible, thus lowering the reduction in conformational entropy of the system upon binding.

Intriguingly, p62, a protein involved in regulation of autophagy, was recently found to bind to the Kelch domain of Keap1(Lau, Wang et al. 2010). Crystal structures show that ProT $\alpha$ , p62, and Neh2 all interact with the same site on the Kelch domain (Lau, Wang et al. 2010; Tong, Katoh et al. 2006; Padmanabhan, Nakamura et al. 2008). Even though the Kelch-binding motif in p62, <sup>350</sup>PSTGE<sup>354</sup>, differs from the <sup>41</sup>EENGE<sup>45</sup> and <sup>78</sup>EETGE<sup>82</sup> motifs of ProT $\alpha$  and Neh2, respectively, the Gly and Glu residues at the 4<sup>th</sup> and 5<sup>th</sup> positions are conserved. This observation is consistent with our results that these two residues are crucial for the interaction with Kelch (Lau, Wang et al.2010; Tong, Katoh et al. 2006; Padmanabhan, Nakamura et al. 2008). Interestingly, the corresponding E41 in ProT $\alpha$  is replaced with a Pro residue in p62 (Lau, Wang et al. 2010). ITC experiments performed by Komatsu *et al.* show that the binding affinity of p62 for Kelch is  $K_a = 5.4 \pm 0.3 \times 10^5 \text{ M}^{-1}$ , (Lau, Wang et al. 2010) which is slightly higher than what was determined for ProT $\alpha$  and Kelch ( $K_a = 3.8 \pm 0.6 \times 10^5 \text{ M}^{-1}$ ).

E47 is another residue in ProT $\alpha$  that may help to stabilize the bound-state  $\beta$ -hairpin structure. Our ITC results (Table 2.1) revealed that mutation of this residue to an Ala or Arg decreased the binding affinity to Kelch by 55 fold, and ~10 fold, respectively. Since the side-chain of the corresponding Glu was invisible in the crystal structure of mouse ProT $\alpha$ -Kelch peptide-protein complex (Padmanabhan, Nakamura et al. 2008), it is unclear whether this residue was involved in any intra- or intermolecular interactions. We speculate that the side-chain of E47 may be involved in intra-molecular hydrogen bonding with N38, and the interaction can still be retained in the E47R mutant. The transient interaction between these two residues may also help to stabilize the bound-state-like  $\beta$ -hairpin structure in the free state of ProT $\alpha$ .

The NMR and ITC experiments show that even though peptides can be used to mimic specific regions of a protein when studying binding, the target recognition mechanisms of peptides can differ from that of the full-length protein. As seen in Table 1, although the ProT $\alpha$  peptide and protein had similar binding affinities to Kelch, their  $\Delta H$  and  $\Delta S$  values differed. Based on the observed chemical shift changes of ProT $\alpha$  upon target binding and the spin relaxation parameters determined in both the free and bound-states, only residues around the binding motif were involved in the interaction with Kelch while the rest of the protein remained highly disordered. One possible explanation for the significant differences in the  $\Delta H$  and  $\Delta S$  values is that a greater loss in the conformational entropy of the peptide occurred upon binding to Kelch. This idea is in agreement with our earlier computational study of the free-state structure of full-length ProT $\alpha$  and peptide using MD simulations (Cino, Wong-ekkabut et al. 2011). In that work, we had demonstrated that the full-length ProT $\alpha$  has a higher propensity than the peptide in

forming the bound-state-like  $\beta$ -hairpin structure, which can result in a less entropic penalty upon coupled binding and folding for the full-length protein. On the other hand, the enthalpy change of the ProT $\alpha$  peptide is more negative compared to the full-length protein. Since the measured  $\Delta H$  is the sum of the enthalpy changes due to the binding and folding processes, the larger  $\Delta H$  value of the peptide is likely due to the folding from the less bound-state-like conformations.

Despite the highly dynamic nature of disordered proteins, it has been generally accepted that they will adopt a well-defined structure once bound to the target, at least in the regions that are directly involved in the interaction. Several recent examples, however, challenge this view and suggest that disordered proteins can still be flexible even in the bound states (Fuxreiter 2012; Fuxreiter and Tompa 2008 ; Zhuo, Ilangovan et al.2010). For instance, the disordered cyclin-dependent kinase inhibitor Sic1 does not undergo a disorder-to-order transition upon binding to Cdc4 (Mittag, Orlicky et al. 2008). Instead, NMR studies showed that the multiple and dispersed phosphorylated motifs on Sic1 interact with a single binding site on Cdc4 in a dynamic equilibrium manner. This unusual binding mechanism has been proposed to lead to the ultrasensitivity to phosphorylation, which is crucial for the function of Sic1 in cell cycle regulation (Mittag, Orlicky et al. 2008). Another example of a fuzzy complex is the dynamic interaction between clathrin assembly protein AP180 and clathrin (Zhuo, Ilangovan et al.2010). NMR studies demonstrated that a large fragment of the clathrin binding domain (CBD) of AP180 is disordered. Interestingly, this protein fragment remains unstructured even in the bound form with clathrin. This dynamic binding mode was speculated to be important for the assembly of the clathrin lattice (Zhuo, Ilangovan et al.2010).

In summary, the results presented in this chapter show that even though ProT $\alpha$  binds to the Kelch domain of Keap1 with low affinity, the interaction is highly specific. Only a small region of ProT $\alpha$  is involved in the interaction with the Kelch domain, and the protein does not undergo large conformational change upon binding. These properties may allow ProT $\alpha$  to also interact with other targets simultaneously. In addition, the fuzzy complex formed between the two proteins is probably what allows ProT $\alpha$  to quickly dissociate from Keap1 after the latter is translocated into the nucleus (Niture and Jaiswal 2009).

## 2.5 References

- Bai, Y., J. S. Milne, et al. (1993). "Primary structure effects on peptide group hydrogen exchange." *Proteins* **17**(1): 75-86.
- Bordelon, T., S. K. Montegudo, et al. (2004). "A disorder to order transition accompanies catalysis in retinaldehyde dehydrogenase type II." *J Biol Chem* **279**(41): 43085-43091.
- Cheng, Y., T. LeGall, et al. (2006). "Rational drug design via intrinsically disordered protein." *Trends Biotechnol* **24**(10): 435-442.
- Cino, E., J. Fan, et al. (2013). "(1)H, (15)N and (13)C backbone resonance assignments of the Kelch domain of mouse Keap1." *Biomol NMR Assign.* **7**(2): 149-153.
- Cino, E. A., W. Y. Choy, et al. (2012). "Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations." *J Chem Theory Comput* **8**(8): 2725-2740.
- Cino, E. A., J. Wong-ekkabut, et al. (2011). "Microsecond molecular dynamics simulations of intrinsically disordered proteins involved in the oxidative stress response." *PLoS One* **6**(11): e27371.
- Connelly, G. P., Y. Bai, et al. (1993). "Isotope effects in peptide group hydrogen exchange." *Proteins* **17**(1): 87-92.
- Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." *J Biomol NMR* **6**(3): 277-293.
- Diaz-Jullien, C., A. Perez-Estevez, et al. (1996). "Prothymosin alpha binds histones in vitro and shows activity in nucleosome assembly assay." *Biochim Biophys Acta* **1296**(2): 219-227.
- Dominguez, F., C. Magdalena, et al. (1993). "Tissue concentrations of prothymosin alpha: a novel proliferation index of primary breast cancer." *Eur J Cancer* **29A**(6): 893-897.
- Dyson, H. J. and P. E. Wright (2005). "Intrinsically unstructured proteins and their functions." *Nat Rev Mol Cell Biol* **6**(3): 197-208.
- Enkemann, S. A., R. D. Ward, et al. (2000). "Mobility within the nucleus and neighboring cytosol is a key feature of prothymosin-alpha." *J Histochem Cytochem* **48**(10): 1341-1355.
- Evstafieva, A. G., G. A. Belov, et al. (2000). "Prothymosin alpha fragmentation in apoptosis." *FEBS Lett* **467**(2-3): 150-154.

- Fan, J. S., J. Lim, et al. (2011). "Measurement of amide hydrogen exchange rates with the use of radiation damping." J Biomol NMR **51**(1-2): 151-162.
- Fink, A. L. (2005). "Natively unfolded proteins." Curr Opin Struct Biol **15**(1): 35-41.
- Frillingos, S., M. Frangou-Lazaridis, et al. (1991). "Isolation and partial sequence of goat spleen prothymosin alpha." Mol Cell Biochem **108**(1): 85-94.
- Fujita, R., M. Ueda, et al. (2009). "Prothymosin-alpha plays a defensive role in retinal ischemia through necrosis and apoptosis inhibition." Cell Death Differ **16**(2): 349-358.
- Fuxreiter, M. (2012). "Fuzziness: linking regulation to protein dynamics." Mol Biosyst **8**(1): 168-177.
- Fuxreiter, M., I. Simon, et al. (2004). "Preformed structural elements feature in partner recognition by intrinsically unstructured proteins." J Mol Biol **338**(5): 1015-1026.
- Fuxreiter, M. and P. Tompa. (2008). "Fuzzy complexes: a more stochastic view of protein function." Adv Exp Med Biol **725**: 1-14.
- Gast, K., H. Damaschun, et al. (1995). "Prothymosin alpha: a biologically active protein with random coil conformation." Biochemistry **34**(40): 13211-13218.
- Gou, L. T., A. P. Tong, et al. (2009). "Altered protein-expressing profile in hPNAS4-induced apoptosis in A549 human lung adenocarcinoma cells." J Cell Biochem **108**(5): 1211-1219.
- Hammoudeh, D. I., A. V. Follis, et al. (2009). "Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc." J Am Chem Soc **131**(21): 7390-7401.
- Haritos, A. A., G. J. Goodall, et al. (1984). "Prothymosin alpha: isolation and properties of the major immunoreactive form of thymosin alpha 1 in rat thymus." Proc Natl Acad Sci U S A **81**(4): 1008-1011.
- Hutchinson, E. G. and J. M. Thornton (1994). "A revised set of potentials for beta-turn formation in proteins." Protein Sci **3**(12): 2207-2216.
- Itoh, K., T. Chiba, et al. (1997). "An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements." Biochem Biophys Res Commun **236**(2): 313-322.
- Itoh, K., K. I. Tong, et al. (2004). "Molecular mechanism activating Nrf2-Keap1 pathway in regulation of adaptive response to electrophiles." Free Radic Biol Med **36**(10): 1208-1213.



- Itoh, K., N. Wakabayashi, et al. (1999). "Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain." Genes Dev **13**(1): 76-86.
- Jiang, X., H. E. Kim, et al. (2003). "Distinctive roles of PHAP proteins and prothymosin-alpha in a death regulatory pathway." Science **299**(5604): 223-226.
- Johnson, B. A. (2004). "Using NMRView to visualize and analyze the NMR spectra of macromolecules." Methods Mol Biol **278**: 313-352.
- Karapetian, R. N., A. G. Evstafieva, et al. (2005). "Nuclear oncoprotein prothymosin alpha is a partner of Keap1: implications for expression of oxidative stress-protecting genes." Mol Cell Biol **25**(3): 1089-1099.
- Karetsou, Z., A. Kretsovali, et al. (2002). "Prothymosin alpha interacts with the CREB-binding protein and potentiates transcription." EMBO Rep **3**(4): 361-366.
- Karetsou, Z., G. Martic, et al. (2004). "Prothymosin alpha associates with the oncoprotein SET and is involved in chromatin decondensation." FEBS Lett **577**(3): 496-500.
- Karetsou, Z., R. Sandaltzopoulos, et al. (1998). "Prothymosin alpha modulates the interaction of histone H1 with chromatin." Nucleic Acids Res **26**(13): 3111-3118.
- Kay, L. E. (1998). "Protein dynamics from NMR." Nat Struct Biol **5 Suppl**: 513-517.
- Kim, D. and F. P. Guengerich (2005). "Cytochrome P450 activation of arylamines and heterocyclic amines." Annu Rev Pharmacol Toxicol **45**: 27-49.
- Kim, J. H., A. K. Fuzery, et al. (2009). "Structure and dynamics of the iron-sulfur cluster assembly scaffold protein IscU and its interaction with the cochaperone HscB." Biochemistry **48**(26): 6062-6071.
- Kobayashi, A., M. I. Kang, et al. (2004). "Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2." Mol Cell Biol **24**(16): 7130-7139.
- Komatsu, M., H. Kurokawa, et al. (2010). "The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1." Nat Cell Biol **12**(3): 213-223.
- Kubota, S., Y. Adachi, et al. (1995). "Binding of human prothymosin alpha to the leucine-motif/activation domains of HTLV-I Rex and HIV-1 Rev." Eur J Biochem **233**(1): 48-54.
- Kumar, S., B. Ma, et al. (2000). "Folding and binding cascades: dynamic landscapes and population shifts." Protein Sci **9**(1): 10-19.

- Lau, A., X. J. Wang, et al. (2010). "A noncanonical mechanism of Nrf2 activation by autophagy deficiency: direct interaction between Keap1 and p62." Mol Cell Biol **30**(13): 3275-3285.
- Li, X., D. Zhang, et al. (2004). "Crystallization and initial crystallographic analysis of the Kelch domain from human Keap1." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 2): 2346-2348.
- Lo, S. C., X. Li, et al. (2006). "Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling." Embo J **25**(15): 3605-3617.
- Malicet, C., V. Giroux, et al. (2006). "Regulation of apoptosis by the p8/prothymosin alpha complex." Proc Natl Acad Sci U S A **103**(8): 2671-2676.
- Marsh, J. A., V. K. Singh, et al. (2006). "Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation." Protein Sci **15**(12): 2795-2804.
- Martic, G., Z. Karetsov, et al. (2005). "Parathymosin affects the binding of linker histone H1 to nucleosomes and remodels chromatin structure." J Biol Chem **280**(16): 16143-16150.
- Meszaros, B., P. Tompa, et al. (2007). "Molecular principles of the interactions of disordered proteins." J Mol Biol **372**(2): 549-561.
- Mittag, T., S. Orlicky, et al. (2008). "Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor." Proc Natl Acad Sci U S A **105**(46): 17772-17777.
- Mosoian, A., A. Teixeira, et al. (2007). "Influence of prothymosin-alpha on HIV-1 target cells." Ann N Y Acad Sci **1112**: 269-285.
- Mosoian, A., A. Teixeira, et al. (2010) "Prothymosin-alpha inhibits HIV-1 via Toll-like receptor 4-mediated type I interferon induction." Proc Natl Acad Sci U S A **107**(22): 10178-10183.
- Niture, S. K. and A. K. Jaiswal (2009). "Prothymosin-alpha mediates nuclear import of the INrf2/Cul3 Rbx1 complex to degrade nuclear Nrf2." J Biol Chem **284**(20): 13856-13868.
- Niture, S. K., J. W. Kaspar, et al. (2010). "Nrf2 signaling and cell survival." Toxicol Appl Pharmacol **244**(1): 37-42.
- Padmanabhan, B., Y. Nakamura, et al. (2008). "Structural analysis of the complex of Keap1 with a prothymosin alpha peptide." Acta Crystallogr Sect F Struct Biol Cryst Commun **64**(Pt 4): 233-238.

- Padmanabhan, B., K. I. Tong, et al. (2006). "Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer." Mol Cell **21**(5): 689-700.
- Pan, L. X., A. A. Haritos, et al. (1986). "Human prothymosin alpha: amino acid sequence and immunologic properties." Arch Biochem Biophys **250**(1): 197-201.
- Papamarcaki, T. and O. Tsolas (1994). "Prothymosin alpha binds to histone H1 in vitro." FEBS Lett **345**(1): 71-75.
- Peng, J. W. and G. Wagner (1992). "Mapping of the spectral densities of N-H bond motions in eglin c using heteronuclear relaxation experiments." Biochemistry **31**(36): 8571-8586.
- Piacentini, M., C. Evangelisti, et al. (2003). "Does prothymosin-alpha act as molecular switch between apoptosis and autophagy?" Cell Death Differ **10**(9): 937-939.
- Pratico, D., J. Rokach, et al. (2004). "F2-isoprostanes as indices of lipid peroxidation in inflammatory diseases." Chem Phys Lipids **128**(1-2): 165-171.
- Primiano, T., T. R. Sutter, et al. (1997). "Antioxidant-inducible genes." Adv Pharmacol **38**: 293-328.
- Qi, X., L. Wang, et al. (2010). "Novel small molecules relieve prothymosin alpha-mediated inhibition of apoptosome formation by blocking its interaction with Apaf-1." Biochemistry **49**(9): 1923-1930.
- Radford, S. E., M. Buck, et al. (1992). "Hydrogen exchange in native and denatured states of hen egg-white lysozyme." Proteins **14**(2): 237-248.
- Rodriguez, P., J. E. Vinuela, et al. (1998). "Overexpression of prothymosin alpha accelerates proliferation and retards differentiation in HL-60 cells." Biochem J **331** ( Pt 3): 753-761.
- Rodriguez, P., J. E. Vinuela, et al. (1999). "Prothymosin alpha antisense oligonucleotides induce apoptosis in HL-60 cells." Cell Death Differ **6**(1): 3-5.
- Rubtsov, Y. P., A. S. Zolotukhin, et al. (1997). "Mutational analysis of human prothymosin alpha reveals a bipartite nuclear localization signal." FEBS Lett **413**(1): 135-141.
- Sburlati, A. R., R. E. Manrow, et al. (1991). "Prothymosin alpha antisense oligomers inhibit myeloma cell division." Proc Natl Acad Sci U S A **88**(1): 253-257.
- Schulze, P. C. and R. T. Lee (2005). "Oxidative stress and atherosclerosis." Curr Atheroscler Rep **7**(3): 242-248.
- Segade, F. and J. Gomez-Marquez (1999). "Prothymosin alpha." Int J Biochem Cell Biol **31**(11): 1243-1248.

- Subramanian, C., S. Hasan, et al. (2002). "Epstein-Barr virus nuclear antigen 3C and prothymosin alpha interact with the p300 transcriptional coactivator at the CH1 and CH3/HAT domains and cooperate in regulation of transcription and histone acetylation." J Virol **76**(10): 4699-4708.
- Sugase, K., H. J. Dyson, et al. (2007). "Mechanism of coupled folding and binding of an intrinsically disordered protein." Nature **447**(7147): 1021-1025.
- Tong, K. I., Y. Katoh, et al. (2006). "Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model." Mol Cell Biol **26**(8): 2887-2900.
- Tong, K. I., A. Kobayashi, et al. (2006). "Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism." Biol Chem **387**(10-11): 1311-1320.
- Traub, F., M. Jost, et al. (2006). "Peptidomic analysis of breast cancer reveals a putative surrogate marker for estrogen receptor-negative carcinomas." Lab Invest **86**(3): 246-253.
- Trumbore, M. W., R. H. Wang, et al. (1997). "Prothymosin alpha in vivo contains phosphorylated glutamic acid residues." J Biol Chem **272**(42): 26394-26404.
- Tsai, C. J., B. Ma, et al. (2001). "Structured disorder and conformational selection." Proteins **44**(4): 418-427.
- Tsai, Y. S., Y. C. Jou, et al. (2009). "Aberrant prothymosin-alpha expression in human bladder cancer." Urology **73**(1): 188-192.
- Tsitsiloni, O. E., J. Stiakakis, et al. (1993). "Expression of alpha-thymosins in human tissues in normal and abnormal growth." Proc Natl Acad Sci U S A **90**(20): 9504-9507.
- Ueda, H., R. Fujita, et al. (2007). "Identification of prothymosin-alpha1, the necrosis-apoptosis switch molecule in cortical neuronal cultures." J Cell Biol **176**(6): 853-862.
- Uversky, V. N. (2012). "Intrinsically disordered proteins and novel strategies for drug discovery." Expert Opin Drug Discov **7**(6): 475-488.
- Uversky, V. N., J. R. Gillespie, et al. (1999). "Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH." Biochemistry **38**(45): 15009-15016.
- van Ingen, H., M. A. Baltussen, et al. (2006). "Role of structural and dynamical plasticity in Sin3: the free PAH2 domain is a folded module in mSin3B." J Mol Biol **358**(2): 485-497.

- Venugopal, R. and A. K. Jaiswal (1996). "Nrf1 and Nrf2 positively and c-Fos and Fra1 negatively regulate the human antioxidant response element-mediated expression of NAD(P)H:quinone oxidoreductase1 gene." Proc Natl Acad Sci U S A **93**(25): 14960-14965.
- Walma, T., J. Aelen, et al. (2004). "A closed binding pocket and global destabilization modify the binding properties of an alternatively spliced form of the second PDZ domain of PTP-BL." Structure **12**(1): 11-20.
- Wang, J., Z. Cao, et al. (2011). "Novel Strategies for Drug Discovery Based on Intrinsically Disordered Proteins (IDPs)." Int J Mol Sci **12**(5): 3205-3219.
- Wishart, D. S., C. G. Bigam, et al. (1995). "<sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shift referencing in biomolecular NMR." J Biomol NMR **6**(2): 135-140.
- Wright, P. E. and H. J. Dyson (1999). "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm." J Mol Biol **293**(2): 321-331.
- Wright, P. E. and H. J. Dyson (2009). "Linking folding and binding." Curr Opin Struct Biol **19**(1): 31-38.
- Yi, S., B. L. Boys, et al. (2007). "Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: evidence for metalation as an entropic switch." Biochemistry **46**(45): 13120-13130.
- Yi, S., A. Brickenden, et al. (2008). "A new protocol for high-yield purification of recombinant human prothymosin alpha expressed in Escherichia coli for NMR studies." Protein Expr Purif **57**(1): 1-8.
- Zhuo, Y., U. Ilangoan, et al. (2010). "Dynamic interactions between clathrin and locally structured elements in a disordered protein mediate clathrin lattice assembly." J Mol Biol **404**(2): 274-290.

## Chapter 3

# Structural characterization of the full-length Neh2 domain of human Nrf2 and its interaction with the Kelch domain of human Keap1

### 3.1 Introduction

The Nrf2 protein belongs to the Cap 'n' Collar (CNC) group of transcription factors that contain a conserved basic-leucine zipper region (Moi, Chan et al. 1994). The protein contains six domains, and the Neh2 domain at the N-terminal region of Nrf2 is responsible for the regulation of the antioxidant pathway (Itoh, Wakabayashi et al. 1999). Mouse and human Neh2 domains are typically comprised of 98 residues and the two homologs have about 94 % sequence homology (Figure 3.1). The structure of the human and mouse Neh2 domain was simulated using secondary structure predictor PrDOS software (Ishida and Kinoshita 2007). Figure 3.2 shows that there is no difference in the disordered nature of the human and mouse Neh2 domain. Because of the high similarity in sequence and the similar structure prediction, differences in structure or function between the human and mouse Neh2 homologs are highly unlikely.

Most Nrf2/Keap1 binding studies have been performed with peptides encoding the ETGE and DLG motifs (Lo, Li et al. 2006; Tong, Katoh et al. 2006; Tong, Padmanabhan et al. 2007). It is costly to use synthetic peptides mimicking regions of interest. Moreover, peptides do not always allow proper elucidation of the mechanism of interaction with targets as noted for ProTα/Keap1 binding in full-length and peptide

```

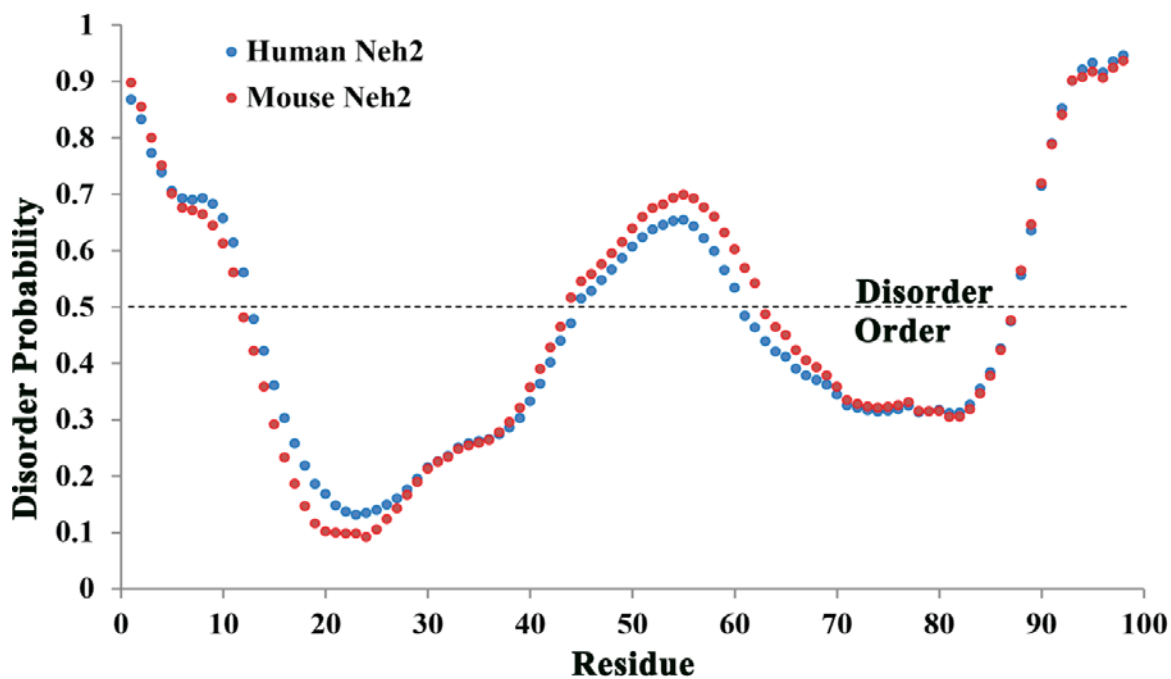
Human Neh2 MMDLELPPPGLPPSQQDMDLIDILWRQDIDLGVSREVFDFSQRRKEYELEK 50
Mouse Neh2 MMDLELPPPGLPQSQQDMDLIDILWRQDIDLGVSREVFDFSQRQKDELEK

Human Neh2 QKKLEKERQEQLQKEQEKAFFAQLQLDEETGEFLPIQPAQHIQSETSG 98
Mouse Neh2 QKKLEKERQEQLQKEQEKAFFAQFQLDEETGEFLPIQPAQHIQTDTSG

```

**Figure 3.1. Sequence alignment of the mouse and human Neh2 domains of Nrf2.**

The Neh2 domain of the human and mouse are composed of 98 residues. The amino acid sequence is well conserved between the human and mouse with differences in six amino acids (highlighted in red).



**Figure 3.2. Disorder prediction for the human and mouse Neh2.**

The overlay of the human Neh2 domain (blue) with the mouse Neh2 domain (red) shows very minor differences in secondary structure prediction. A score below 0.5 is indicative of structured regions, and above 0.5 is indicative of disordered regions. Disorder was predicted by PrDOS (Ishida and Kinoshita 2007)



(Khan, Cino et al. 2013). Tong et al. have reported a purification protocol for the Neh2 domain of mouse Nrf2 in which 4.0 mg of recombinant Neh2 was obtained. This method was used to extract the recombinant full-length Neh2 domain for binding studies. However, the low yield and purity of the sample using the method by Tong *et al.* were not suitable for the human Nrf2. In this work, an alternative high yield protocol, free from impurities, is established for the recombinant Neh2 domain, Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG of human Nrf2. The method described here is not only time and cost effective, but importantly, allowed investigation of the full-length Neh2 interaction with Kelch in presence of cancer causing mutations (discussed in Chapter 4).

Structural characterization of the full length Neh2 demonstrated that it is a disordered protein with high secondary structure content, thereby, limiting its flexibility unlike ProT $\alpha$  (Tong, Katoh et al. 2006; Yi, Boys et al. 2007; Khan, Cino et al. 2013). Additionally, hydrodynamic radius of the Neh2 domain indicates a more compact structure than the disordered ProT $\alpha$  (Yi, Boys et al. 2007). The Neh2 $\Delta$ DLG and Neh2 $\Delta$ ETGE NMR HSQC in the free and Kelch bound state implicated presence of long range interactions between the N and C terminal regions of the Neh2 domain.

## **3.2 Materials and Methods**

### ***3.2.1 Construction of Neh2, Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG in E. coli Expression***

#### ***Vectors***

The pENTR™221 vector containing the gene for the full length human Nrf2 (NM\_006164) was purchased from Invitrogen. Using Gateway cloning, the Neh2 domain of the human Nrf2 was transferred to a pDEST17 vector containing a his-tag and

a TEV cleavage sequence. The Quickchange Multi site directed mutagenesis kit (Stratagene) was used to delete the <sup>79</sup>ETGE<sup>82</sup> region in full length Neh2 domain thus creating the Neh2ΔETGE construct using the Neh2 in pDEST17 vector as a template. The Neh2ΔDLG, in which the first thirty-three residues of the Neh2 domain are deleted, was constructed using the full length Nrf2 in pENTR™221 as a template. Using Gateway cloning methods, the Neh2ΔDLG construct was transferred to a pDEST17 vector containing a his-tag and a TEV cleavage sequence. The Neh2ΔETGE and Neh2ΔDLG constructs were first prepared by Tong *et al* to study the DLG and ETGE interaction with the Kelch domain of mouse Keap1 (Tong, Katoh et al. 2006).

### **3.2.2 Expression of the Neh2, Neh2ΔETGE and Neh2ΔDLG protein**

Neh2, Neh2ΔETGE and Neh2ΔDLG were over-expressed in the Rosetta 2(DE3) pLysS *E. coli* strain in M9 minimal medium. The cultures were grown at 37 °C in the presence of Carbenicillin (50 µg/mL) to a density of (A 600) 0.8 AU and then induced with 0.5 mM IPTG. The cultures were allowed to grow over night at 22 °C in a shaking incubator at 250 rpm. Bacterial cells were harvested by centrifugation at 6,000 x g for 15 minutes. The bacterial pellets were resuspended in 40 mL of 1 x Phosphate buffered saline (PBS) and centrifuged at 6,000 x g for 15 minutes at 4 °C in 50 mL centrifuge tubes. Bacterial wet pellets weighing between 2.0 to 2.5 g were routinely obtained. For NMR experiments, the media was prepared with 1 g of <sup>15</sup>NH<sub>4</sub>Cl as the nitrogen source for uniformly <sup>15</sup>N-labeled protein samples. For the Neh2 protein backbone assignment, the media was prepared with 1 g of <sup>15</sup>NH<sub>4</sub>Cl and 3 g of <sup>13</sup>C<sub>6</sub>-glucose as the sole nitrogen and carbon sources, respectively. Analysis of bacterial lysate on SDS-PAGE showed that

Neh2, Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG are present in both the soluble and insoluble fractions.

### ***3.2.3 Purification of the Soluble Neh2, Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG protein***

The bacterial pellet of Neh2 was resuspended in lysis buffer (20 mM Tris-HCl, 1 mM EDTA, pH 8.0). The resuspension was incubated at 37 °C for 15 minutes and then sonicated 3 x 10 second bursts on ice. Cell debris was removed by centrifugation at 40,000 x g for 30 min at 4 °C. The supernatant was applied to Ni-Sepharose 6FF column equilibrated with binding buffer (20 mM Tris-HCl, 1 mM EDTA, 10 mM imidazole, 500 mM NaCl, 10% glycerol, pH 8.0) and was incubated for 2 hours at room temperature. The Neh2 protein was eluted in 5 x 5 mL fractions with elution buffer (20 mM Tris-HCl, 300 mM NaCl, 750 mM imidazole, 10% glycerol, pH 8.0). Protein fractions were dialyzed over night at 4 °C against 2 L of dialysis buffer (20 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, 10% glycerol, pH 8.0 and 1 mM DTT). After a second dialysis step, the protein sample was centrifuged at 40,000 x g for 15 min to remove aggregates. Protein concentration was determined with the Bio-Rad assay (Bradford 1976). The his-tag of Neh2 was removed with TEV protease (1 mg per 15 mg of protein). The TEV cleaved protein sample was applied to equilibrated Ni-sepharose column and incubated for 30 min at room temperature. The unbound protein fraction containing the Neh2 domain was collected and its concentration was determined by Bio-Rad protein assay (Bradford 1976). Protein samples were analyzed on SDS-PAGE for purity. For experimental work, the Neh2 was dialyzed in 50 mM sodium phosphate, 100 mM NaCl, 1 mM DTT, 5% glycerol, pH 7.0. The same purification procedure was applied to the Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG pellets to obtain soluble protein.

### ***3.2.4 Purification of Neh2 and Neh2 $\Delta$ ETGE protein from Inclusion bodies***

An insoluble bacterial pellet for the full-length recombinant Neh2 domain was resuspended in lysis buffer (50 mM Tris-HCl, 1.0% Triton X-100, 100 mM NaCl, 1 mM EDTA, 5 mM DTT, pH 8.0). An equal volume of wash buffer (50 mM Tris-HCl, 0.5% Triton X-100, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 8.0) was added to homogenize the inclusion body suspension and then centrifuged at 25 000 x g for 30 min at 4 °C. The pellet was resuspended in solubilization buffer (25 mM Tris-HCl, 300 mM NaCl, 1mM EDTA, 8 M Urea, pH 7.8) and then centrifuged at 16 000 x g for 30 min at 20 °C. The supernatant was applied to Ni-Sepharose beads that were equilibrated with solubilization buffer (no EDTA added) and incubated at room temperature for 2 hours. The Neh2 protein was eluted in 5 x 5 mL volume fractions with elution buffer (25 mM Tris-HCl, 8 M Urea, 300 mM NaCl, 1 M imidazole, pH 7.8). The eluted fractions were dialyzed in buffer I (25 mM Tris-HCl, 8 M Urea, pH 7.8) over night at room temperature. The next day, protein was dialyzed out of urea into buffer II (25 mM Tris-HCl pH 7.8, 10% glycerol) at 4 °C. The his-tag of the Neh2 was removed with TEV protease. The TEV cleaved protein sample was applied to equilibrated Ni-Sepharose beads and incubated for 30 min at room temperature. Unbound protein was collected, protein concentration was determined by Bio-Rad protein assay, and protein samples were analyzed on SDS-PAGE (Bradford 1976). The same purification procedure was applied to the Neh2 $\Delta$ ETGE insoluble pellet.

### ***3.2.5 Purification of Neh2, Neh2 $\Delta$ ETGE, and Neh2 $\Delta$ DLG protein with 8 M Urea***

A 2.6 g wet pellet from 1 L of bacterial growth was resuspended in solubilization buffer (25 mM Tris-HCl, 8 M Urea, 5 mM imidazole, 300 mM NaCl, and 1 mM EDTA, pH 7.5). The cell lysate was then sonicated for 3 x 10 seconds and centrifuged at 40 000 x g for 30 min at 20 °C. The supernatant was incubated with Ni-Sepharose beads equilibrated with solubilization buffer (without EDTA) at room temperature for 2 hours. Protein was eluted in 6 x 5 mL volume fractions with Elution buffer (25 mM Tris-HCl, 8 M Urea, 750 mM imidazole, 300 mM NaCl, pH 7.5). The eluted protein fractions were dialyzed in buffer I (25 mM Tris-HCl, 8 M Urea, 1 mM EDTA, 100 mM NaCl) overnight at room temperature. The following day, the protein was refolded by exchanging the protein into dialysis buffer II (25 mM Tris-HCl, 100 mM NaCl, pH 7.5). Next, his-tag was removed from Neh2 by adding 1 mg of his-tagged TEV protease per 15 mg of protein and incubated over night at room temperature. Sample was centrifuged at 40 000 x g for 30 min at 4 °C to remove aggregates and precipitate. Protein was unfolded again by adjusting the final urea concentration of the protein sample to 8 M. The TEV cleaved protein sample was then incubated with Ni-Sepharose beads. Unbound protein was collected and dialyzed against buffer I. The protein was refolded using dialysis buffer II at 4 °C. For experimental work, the Neh2 sample was dialyzed in 50 mM sodium phosphate buffer, 100 mM NaCl, 1 mM DTT, pH 7.0. The protein concentrations were confirmed by Bio-Rad assay (Bradford 1976). Protein samples were examined on SDS-PAGE for purity, and samples were sent for Mass spectrometry analysis to ensure the protein band corresponded to the correct molecular weight.

### **3.2.6 Analytical Ultracentrifugation Experiments**

Sedimentation velocity and equilibrium experiments were conducted on a Beckman Optima XL-A Analytical Ultracentrifuge (Biomolecular Interactions and Conformations Facility, University of Western Ontario). For sedimentation equilibrium, the Neh2 sample was run in 6-channel cells with epon-charcoal centerpieces, at a rotor speed of 35,000 rpm at 4 °C in an An60Ti rotor. Absorbance was measured at 280 nm, using 0.002-cm radial steps and averaging over 10 readings. Samples were allowed to equilibrate for 16 hours at the desired speed. Apparent molecular weights of the Neh2 domain of Nrf2 was calculated using ideal single species model in GraphPad Prism program (Briere and Dunn 2006). For sedimentation velocity, two-channel cell with Epon-charcoal centerpieces were used for Neh2 in an An60Ti rotor. Absorbance was monitored at 280 nm at rotor speed of 60,000 rpm (4 °C). Scans were collected at 10-min intervals in 0.003-cm radial steps, averaged over three replicates. Data was analyzed with the SEDFIT software (Brown and Schuck 2006). The partial specific volume of the protein was calculated from the amino acid composition, using the freeware program SEDNTERP (Tom Laue, University of New Hampshire).

### **3.2.7 Circular Dichroism**

Secondary structure contents of Neh2, Neh2 $\Delta$ ETGE, Neh2 $\Delta$ DLG were examined by Circular Dichroism (CD) spectropolarimetry using a Jasco J-810 instrument (Biomolecular Interactions and Conformations Facility, University of Western Ontario). For each protein, ten scans from 260 – 190 nm (100 nm/min at 0.5 increments) were recorded using a 1 mm path-length cell at 20 °C, averaged and the buffer background was

subtracted. Protein concentrations were quantified by Bio-Rad protein quantification assay (Bradford 1976). The CDSSTR method along with reference set SP175 available on the Dichroweb was used to deconvolute the CD data (Lees, Miles et al. 2006). Protein denaturation was followed in the CD at 222 nm. The temperature was increased at a rate of 1 °C/min.

### 3.2.8 *NMR Spectroscopy*

All NMR experiments were acquired on a Varian INOVA 600 MHz spectrometer equipped with a pulsed field gradient triple resonance probe at 25 °C. All samples contained 10% (v/v) D<sub>2</sub>O (Cambridge Isotope Laboratories), and 1 mM DSS (Sigma) as an internal standard for chemical shift referencing. Protein samples were in buffer containing 50 mM sodium phosphate, 100 mM NaCl, 1 mM DTT, pH 7.0. All data were processed with NMR Pipe and NMRDraw (Delaglio, Grzesiek et al. 1995). Spectra were analyzed with NMRView (Johnson 2004).

The <sup>1</sup>H-<sup>15</sup>N HSQC spectra of the Neh2, Neh2ΔETGE and Neh2ΔDLG were collected with protein concentrations ranging from 100 μM to 200 μM. A 0.4 mM uniformly <sup>15</sup>N, <sup>13</sup>C-labeled sample was used to acquire CBCA(CO)NH and HNCACB for the backbone resonance assignment of the full-length Neh2 domain of human Nrf2 (Sattler 1999). The Cara software was used for the backbone assignment of the human Neh2 domain (Keller 2004). Secondary structure propensity (SSP) scores were calculated using the <sup>13</sup>Ca/β chemical shifts extracted from the HNCACB experiment (Marsh, Singh et al. 2006). For the NMR titration experiments, a 100 μM <sup>15</sup>N-labeled samples with and without 200 μM unlabeled Kelch were prepared for Neh2, Neh2ΔETGE and Neh2ΔDLG.

$^1\text{H}$ - $^{15}\text{N}$  HSQC NMR experiments were recorded to monitor spectral changes induced upon the addition of Kelch protein.

Backbone  $^{15}\text{N}$  longitudinal relaxation rates ( $R_1$ ), relaxation rates in rotating frame ( $R_{1\rho}$ ), and steady-state  $^1\text{H}$ - $^{15}\text{N}$  NOE relaxation parameters of 400  $\mu\text{M}$   $^{15}\text{N}$ -labeled Neh2 and Neh2 $\Delta\text{DLG}$  were measured. Spectra were collected as 128 x 1072 complex points in the  $t_1$  and  $t_2$  dimensions with spectral widths of 8384.9 and 1458.4 Hz for  $^1\text{H}$  and  $^{15}\text{N}$  dimensions, respectively. The  $R_1$ ,  $R_{1\rho}$  relaxation rates and the steady-state  $^1\text{H}$ - $^{15}\text{N}$  NOE were calculated using the method described previously (Yi, Boys et al. 2007).

### 3.3 Results

#### 3.3.1 Purification of the Neh2 domain

The initial method used for the purification of the Neh2 domain involved isolating the protein from the soluble fraction of the cell lysate (Tong, Yamamoto et al. 2008). In this method, a large amount of the protein is lost in the unbound and wash fractions (Figure 3.3a, lane 4 & 5). The loss of protein in the unbound fraction and wash step suggested that the His-tag is not accessible to interact with the Ni-beads. In this purification procedure, more than 50% of the protein is lost in the dialysis, and the His-tag cleavage step combined. Additionally, when concentrating the protein to the desired volume, protein precipitation was observed leading to further loss in the overall protein yield.

A 10 g wet bacterial pellet obtained from four liters of M9 minimal medium prepared with  $^{15}\text{N}$ - $\text{NH}_4\text{Cl}$  as the sole ammonium source was used to obtain a 100  $\mu\text{M}$



$^{15}\text{N}$ -labeled Neh2 sample to acquire an NMR HSQC of the Neh2 domain. This purification yielded ~3 mg of Neh2 and included contaminants (Lanes 7 and 8 in Figure 3.3a). A  $^1\text{H}$ - $^{15}\text{N}$  HSQC of the Neh2 domain obtained from this impure sample is shown in Figure 3.4. The narrow dispersion of the amide proton chemical shifts confirms that the human Neh2 domain is also disordered like the mouse Neh2 domain of Nrf2 (Tong, Katoh et al. 2006). Although an NMR HSQC of the Neh2 domain was acquired, nonetheless, the Neh2 stability was compromised with this purification method, and due to presence of impurities, target binding studies could not be performed. Alternatively, the recombinant Neh2 domain was purified from inclusion bodies in an attempt to increase its purity and the final yield. The purification from the inclusion bodies resulted in an even lower yield of 0.6 mg, however, the sample appeared free of contaminants (Figure 3.3b, lane 8).

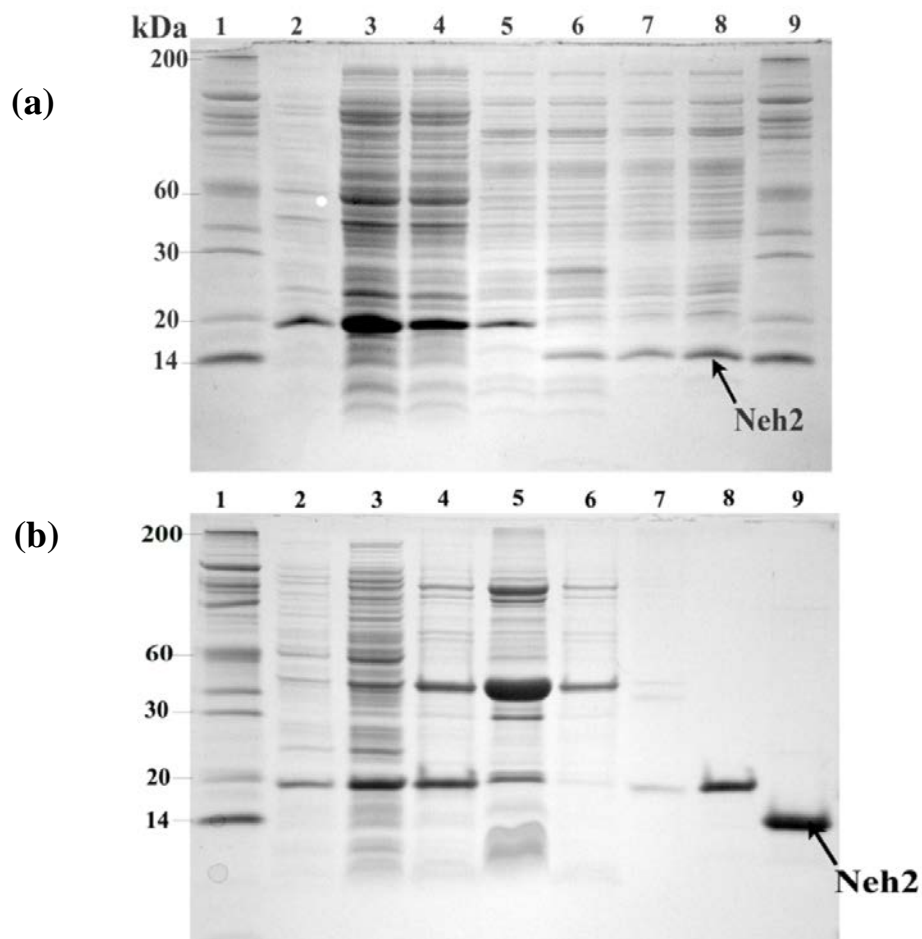
An alternative strategy was adopted to obtain a pure sample of the recombinant Neh2 domain of human Nrf2 for target binding studies. A 2.5 g pellet obtained from 1 L of M9 minimal media growth was suspended in 8M urea and purified under denaturing conditions. The sample was refolded and then subjected to TEV protease to cleave off the his-tag. The Neh2 and TEV mixture was denatured in 8M urea a second time and incubated with Ni-sepharose beads to remove the TEV protease as well as any uncleaved Neh2. This method not only resulted in pure Neh2 sample (as depicted in lane 3 of Figure 3.5b), but also the overall yield was improved from 0.8 mg to 35 mg, an over 40 fold increase.

Overlay of the NMR HSQC spectra of the refolded Neh2 and the soluble Neh2 (Figure 3.6) confirms that the protein sample obtained by the urea denaturation technique

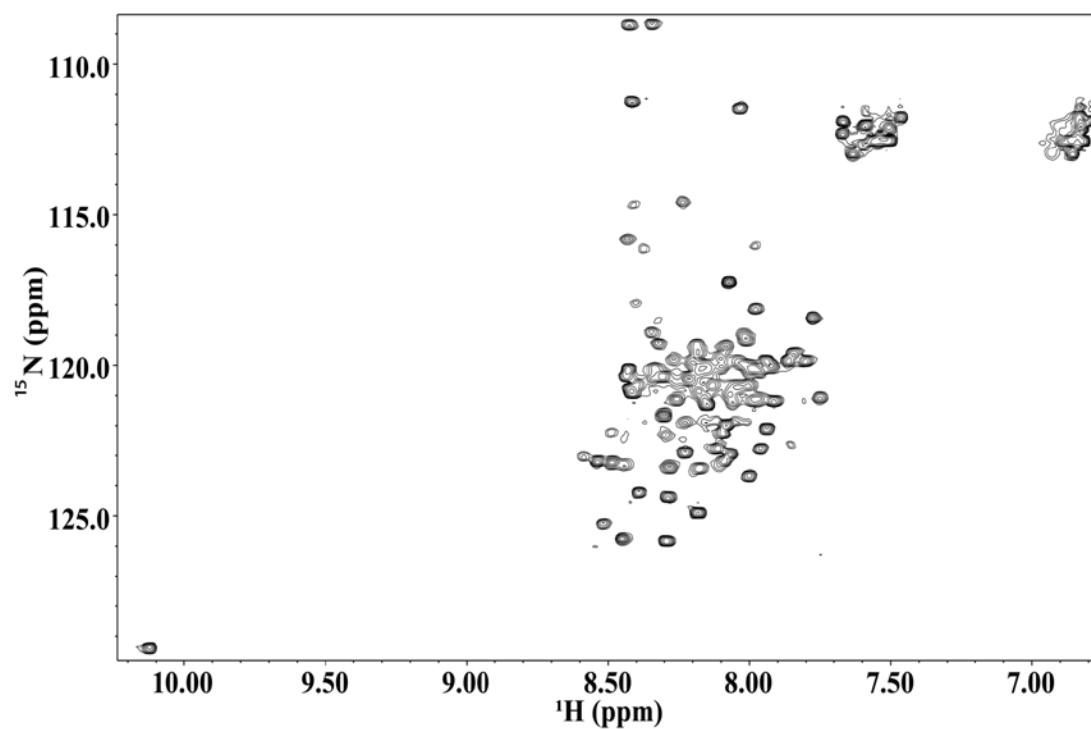
does not result in misfolding of this domain. The Neh2 sample of  $^{15}\text{N}$ -labeled Neh2 acquired from the urea denaturation process was sent for Mass spectrometry to assess its molecular weight. Figure 3.7a shows a mass spectrum of one specie with a molecular weight of 11 810 Da. The peak at 11 810 Da is in good agreement with the calculated molecular weight for the full-length Neh2 domain ( $\text{MW}_{\text{cal}} = 11\ 812\text{Da}$ ).

Another useful analytical method for determining protein molecular weight as well as protein oligomeric state in solution is sedimentation equilibrium. This method measures the equilibrium concentration distribution in the presence of a centrifugal field (Lewis 1994). Equilibrium is reached when the effect of centrifugal force on the samples is balanced by diffusion (Lewis 1994). The distribution is dependent on the molecular weight of the macromolecule and not its shape (Lewis 1994). The molecular weight and oligomer state of the Neh2 domain were assessed by carrying out sedimentation equilibrium experiment at 35 000 rpm. The results revealed that the Neh2 domain exists in the monomeric form in solution corresponding to an average molecular weight of 11 700 Da (Figure 3.7b). This observed molecular weight is on par with the expected molecular weight of 11 812 Da.

The Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG constructs of the Neh2 domain were purified by denaturing the cell lysate from 1 L of bacterial culture in 8 M urea. Table 3.1 presents the Neh2, Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG yields obtained by the different purification procedures. The solubilization of the cell lysate in 8 M urea significantly increased the final yield for all three constructs and the resulting protein was of high purity.

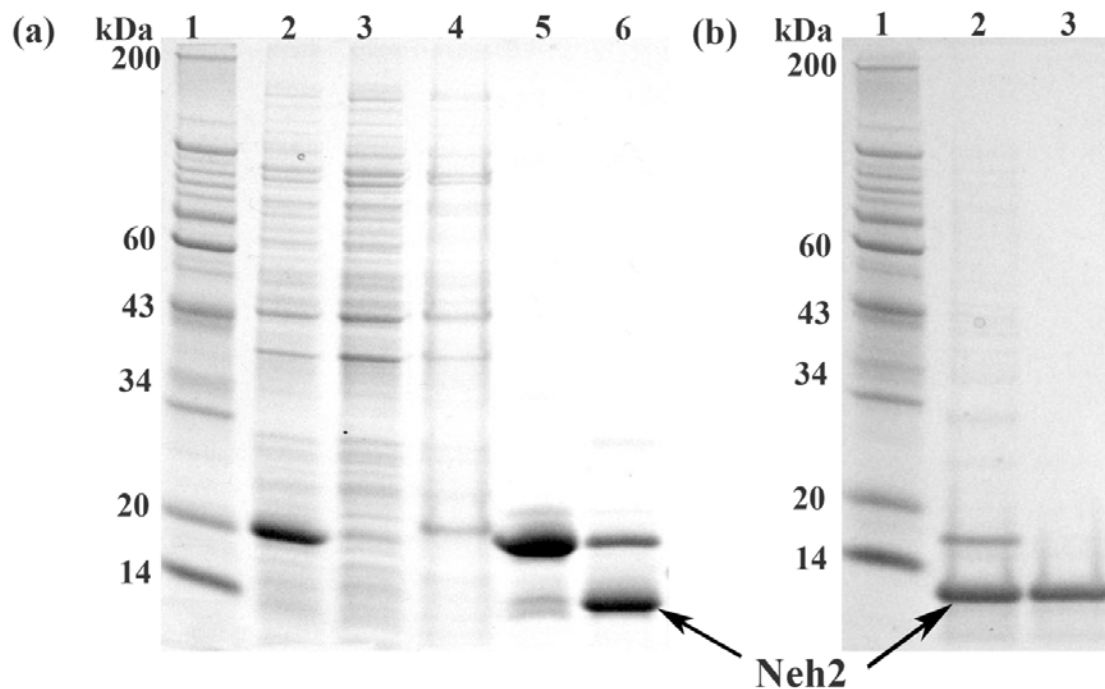


**Figure 3.3. SDS-PAGE (12%) analysis of the overexpression and purification of the human Neh2 domain.** Purification of Neh2 (a) from soluble fraction. Lanes 1 and 9 contain the protein molecular weight markers, with molecular weights labeled on the left of the gel. Lane 2 is the post-induction sample. Lanes 3 and 4 contain soluble and unbound fractions, respectively. Lane 5 is the wash step. Lane 6 is the dialysis precipitate fraction. Lanes 7 and 8 represent 5  $\mu$ g and 10  $\mu$ g of the final Neh2 protein, respectively. The Neh2 protein runs lower in lanes 6, 7 and 8 as the Neh2 was subjected to TEV protease cleavage to remove the his-tag present on the N-terminus of the Neh2 domain. (b) Purification from inclusion bodies. Lane 1, molecular weight marker ; Lane 2, post-induction sample ; Lane 3, cell lysate ; Lane 4, urea soluble ; Lane 5, urea-insoluble ; Lane 6, unbound ; Lane 7, bound to beads ; Lane 8, his-tagged Neh2 ; Lane 9, his-tag free Neh2.



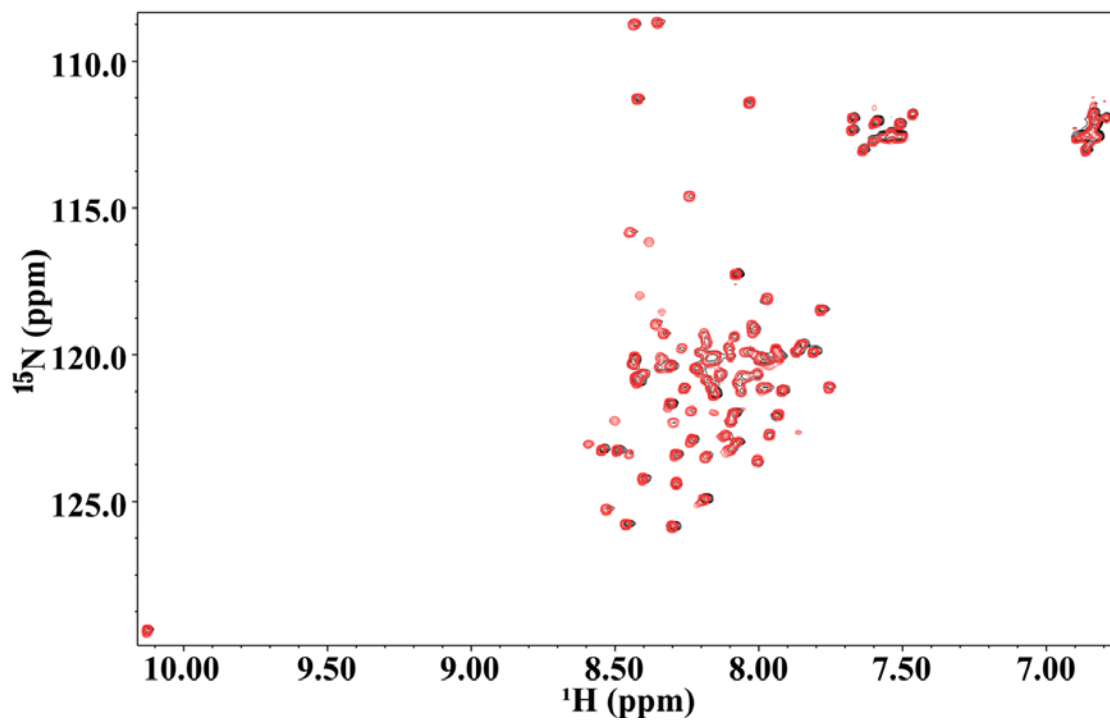
**Figure 3.4.**  $^1\text{H}$ - $^{15}\text{N}$  HSQC of the human Neh2 domain of Nrf2.

$^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectrum of the 0.1 mM human Neh2 domain of Nrf2. The spectrum was collected in 50 mM sodium phosphate, 100 mM NaCl, 1 mM DTT, pH 7.0 at 25 °C.



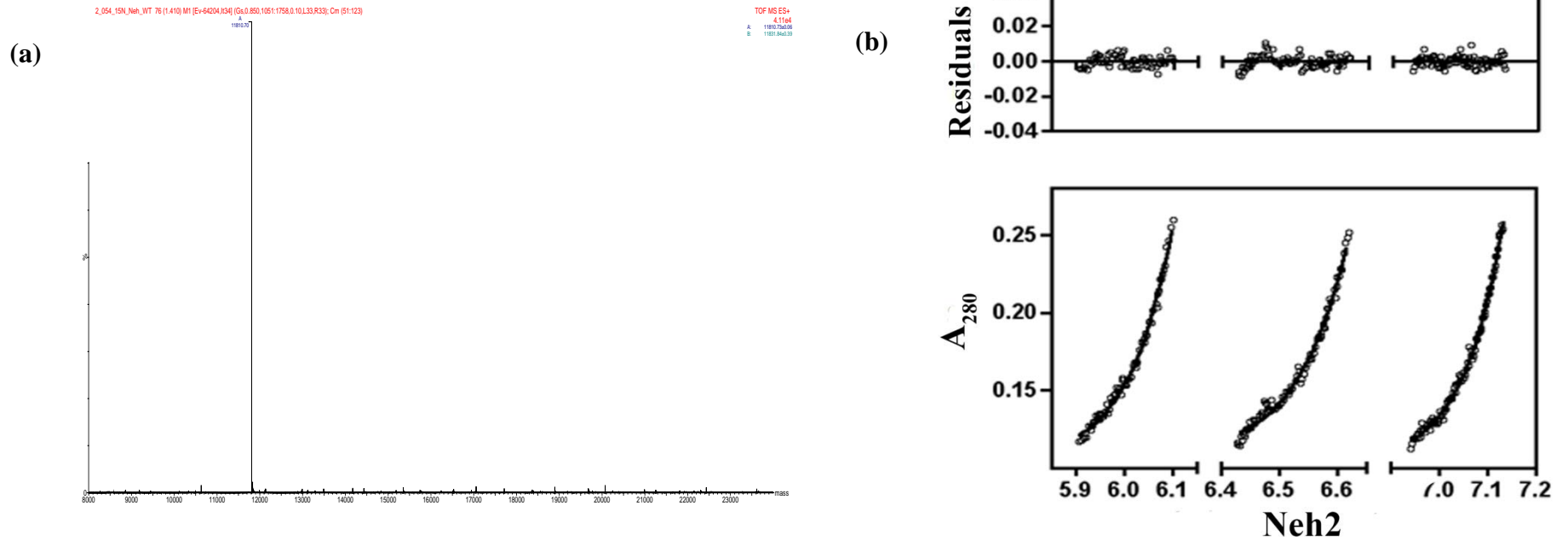
**Figure 3.5. Purification of Neh2 by denaturation in 8M Urea.**

SDS-PAGE (12% gel) (a) analysis of the protein sample during the first step of purification. Lane 1, molecular weight markers; lane 2, cell lysate; lane 3, unbound fraction; lane 4, dialysis precipitate; lanes 5, his-tagged Neh2; lanes 6, his-tagged cleaved Neh2 protein. (b) analysis of Neh2 samples after second denaturation step. Lane 1, molecular weight markers; lane 2, Neh2 sample pre Ni-bead binding; Lane 3, the unbound fraction of Neh2 post Ni-bead binding.



**Figure 3.6.**  $^1\text{H}$ - $^{15}\text{N}$  HSQC of the Neh2 domain of human Nrf2.

100  $\mu\text{M}$  of  $^{15}\text{N}$ -labeled Neh2 obtained from the soluble fraction of the cell lysate (black) is overlaid with 100  $\mu\text{M}$  of  $^{15}\text{N}$ -labeled Neh2 obtained from urea denaturation method (red).



**Figure 3.7. Neh2 Mass spectrum and Stoichiometry.** (a) Mass spectrum of  $^{15}\text{N}$ -labeled Neh2 domain of human Nrf2. The resulting mass spectrum revealed a peak at 11 810, which is in close agreement with the 11 812 calculated molecular weight of the Neh2 domain. (b) Sedimentation equilibrium analysis of the Neh2 domain revealed that Neh2 exists as a monomer in solution and corresponds to MW of 11 700 Da, in close agreement with 11 812 calculated molecular weight for the domain.

**Table 3.1. Protein yield in mg per liter of culture media from different purification methods**

	<b>Neh2<sup>wt</sup></b>	<b>Neh2<math>\Delta</math>ETGE</b>	<b>Neh2<math>\Delta</math>DLG</b>
<b>Soluble</b>	0.8 mg	0.95 mg	4.0 mg
<b>Inclusion bodies</b>	0.6 mg	2.9 mg	ND *
<b>8 M Urea</b>	35 mg	12 mg	30 mg

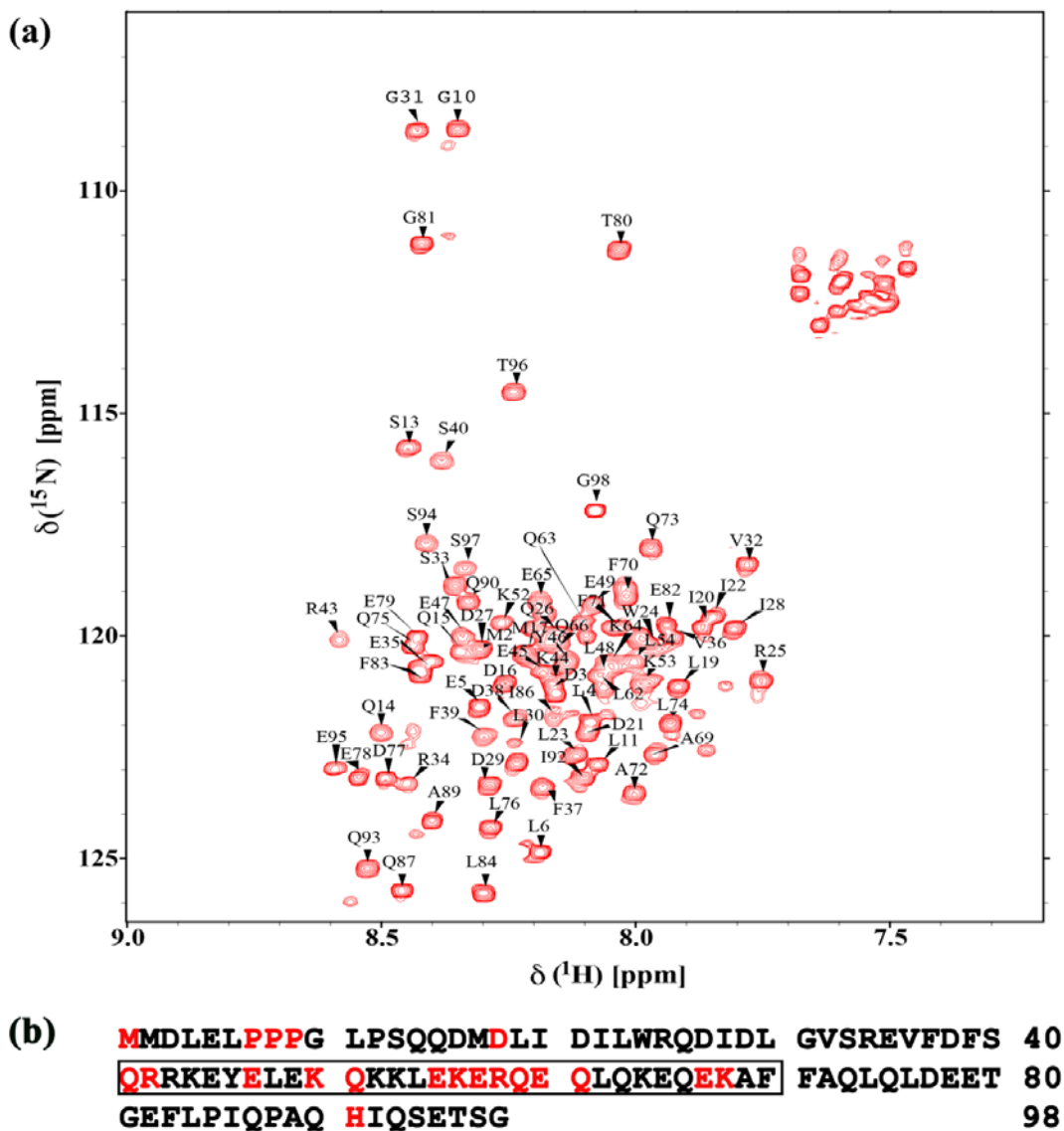
\*ND (Not determined). The Neh2 $\Delta$ DLG was not purified from the inclusion bodies as enough protein was obtained from the soluble fraction and from urea denaturation protocol. The urea denaturation protocol was preferred as the final sample from this procedure was free of impurities.



### 3.3.2 Backbone Assignment of the human Neh2 domain of Nrf2

The backbone assignment of the human Neh2 domain was completed in order to identify amino acids that are involved in forming the binding interface with targets and to characterize the structural changes the protein undergoes in the presence of binding targets. Using the HNCACB and CBCA(CO)NH three-dimensional NMR experiments, the backbone resonance assignment of the Neh2 domain was completed. Figure 3.8 shows a representative  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of the disordered Neh2 domain that is narrow in the proton dimension. Because the Neh2 domain appears to be a disordered protein, extensive overlap of resonances is observed making it challenging to complete the backbone assignment of this domain. The backbone assignment was further complicated by the presence of glutamine, glutamic acid and lysine residues (amino acids with similar carbon chemical shifts) in the central region of the Neh2 domain. Regardless of the high spectral degeneracy  $\sim 84\%$  of the  $^1\text{H}^{\text{N}}$  and  $^{15}\text{N}$  resonances of non-proline residues,  $83\%$  of all  $\text{C}_\alpha$  and  $\sim 79\%$  of all  $\text{C}_\beta$  resonances of the Neh2 domain of human Nrf2 were successfully completed (Table 3.2). The assignment of the Neh2 domain in this work is comparable to the backbone resonance assignment of the Neh2 domain of the mouse Nrf2 (Tong, Yamamoto et al. 2008).

Figure 3.9 shows the residue specific secondary structure propensity scores (SSP) calculated using the assigned  $^{13}\text{C}\alpha/\beta$  chemical shifts with the SSP software developed by Marsh *et al* (Marsh, Singh et al. 2006). Positive values indicate an  $\alpha$ -helical secondary structural propensity, and negative values indicate a  $\beta$ -strand structural propensity. The SSP scores calculated for the Neh2 domain show that most residues from 17 to 40 in the N-terminal region have a tendency to form an  $\alpha$ -helical structure. The center, residues 45

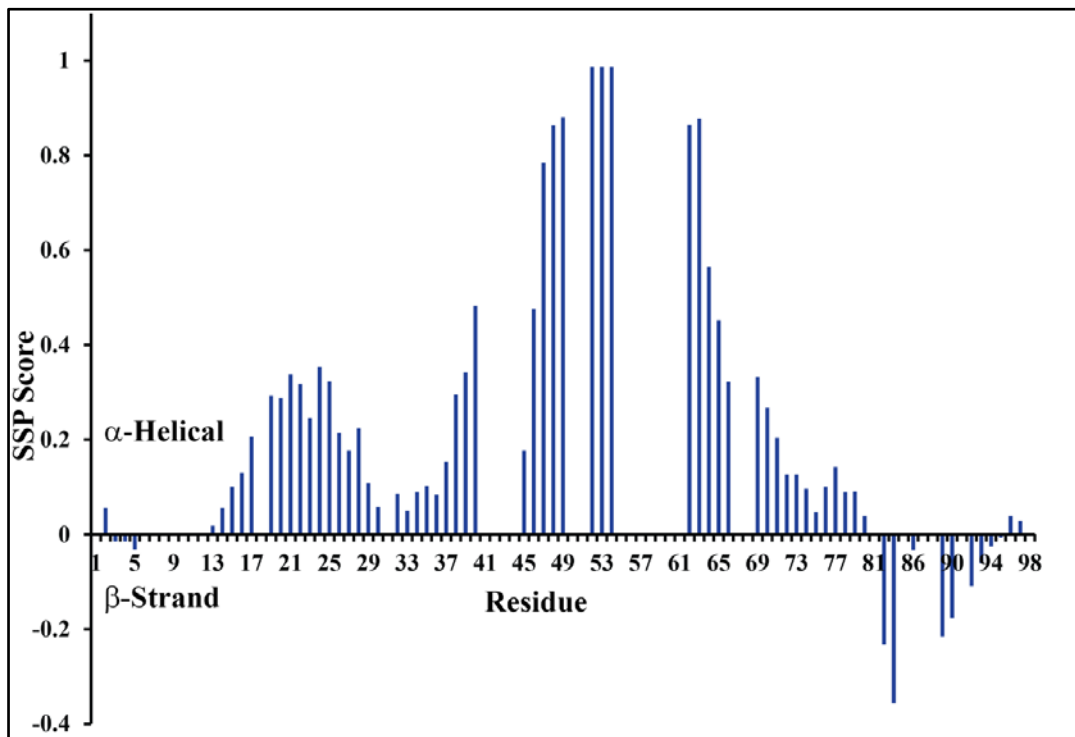


**Figure 3.8. Backbone resonance assignment of the Neh2 domain of the human Nrf2 protein.** (a)  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum and backbone resonance assignment of the  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled Neh2 domain of Nrf2. (b) Neh2 protein sequence with unassigned residues colored in red and the center region of high degeneracy is boxed.

**Table 3.2.  $^{15}\text{N}$ ,  $^{13}\text{C}$  and  $^1\text{H}$  resonance assignments for the human Neh2 domain**

<b>Residue</b>	<b><math>\text{H}^{\text{N}}</math></b>	<b>N</b>	<b><math>\text{C}^{\alpha}</math></b>	<b><math>\text{C}^{\beta}</math></b>
2	8.3	120.36	54.74	29.47
3	8.15	121.23	54.71	41.05
4	8.08	121.89	55.03	42.44
5	8.3	121.56	56.12	30.34
6	8.19	124.84	52.84	41.74
10	8.35	108.55	44.99	
11	8.07	122.87	53.06	41.91
13	8.45	115.76	58.15	64.06
14	8.5	122.11	56.41	29.28
15	8.34	120.3	56.4	29.41
16	8.25	121.01	54.97	40.96
17	8.21	120.36	56.17	32.75
19	7.91	121.12	55.72	42.06
20	7.86	119.7	62.3	38.26
21	8.09	122.11	55.34	41.43
22	7.84	119.48	62.44	38.57
23	8.11	122.65	56.12	41.82
24	7.96	120.16	57.81	29.3
25	7.75	120.9	56.56	30.82
26	8.15	119.92	58.58	29.5
27	8.3	120.25	54.65	40.82
28	7.8	119.81	61.4	38.86
29	8.28	123.31	54.43	41.13
30	8.23	122.34	55.62	42.22
31	8.42	108.55	45.71	
32	7.78	118.39	62.09	32.99
33	8.35	118.83	58.32	63.71
34	8.45	123.31	56.62	30.6
35	8.4	120.47	57.28	30.34
36	7.92	119.98	62.93	32.48
37	8.18	123.31	58.28	39.78
38	8.23	121.78	54.16	40.88
39	8.3	122.22	59.59	39.07
40	8.38	115.98	60.51	63.01
43	8.58	120.03	55.66	32.53
44	8.16	121.03	54.81	38.77
45	8.18	120.8	58.11	29.9
46	8.13	120.58	60.13	38.21
47	8.34	119.94	58.82	29.33
48	8.06	120.8	57.49	41.81
49	8.08	119.26	58.11	28.37
52	8.26	119.7	58.96	29.48
53	7.98	121.01	59.05	32.51
54	8	120.53	57.72	41.88
62	8.06	121.12	57.65	41.83
63	8.09	119.92	58.79	29.07
64	8.05	120.58	58.78	32.36

65	8.19	119.15	58.55	28.36
66	8.15	120.3	54.71	41.09
69	7.96	122.55	53.51	18.74
70	8.02	119.05	59.03	39.24
71	8.03	119.81	58.79	39.23
72	8	123.53	53.12	19
73	7.97	117.95	56.13	29.17
74	7.93	121.89	55.27	42.99
75	8.43	120.22	57.66	30.09
76	8.28	124.3	55.31	42.59
77	8.49	123.2	54.4	41.72
78	8.54	123.09	57.68	30.17
79	8.42	120.03	57.53	30.19
80	8.03	111.28	61.94	70.31
81	8.42	111.17	45.5	
82	7.93	119.7	55.96	31.02
83	8.42	120.8	57.48	39.59
84	8.3	125.72	52.52	42.39
86	8.16	121.79	61.01	38.81
87	8.46	125.72	53.43	28.95
89	8.4	124.08	52.48	19.2
90	8.32	119.15	55.73	29.69
92	8.1	123.09	60.94	38.69
93	8.53	125.17	55.79	29.45
94	8.41	117.84	58.36	63.82
95	8.59	122.98	56.72	30.35
96	8.23	114.45	61.5	69.74
97	8.33	118.39	58.32	63.86
98	8.08	117.08	46.32	



**Figure 3.9. Secondary structure propensity of the Neh2 domain of Nrf2.**

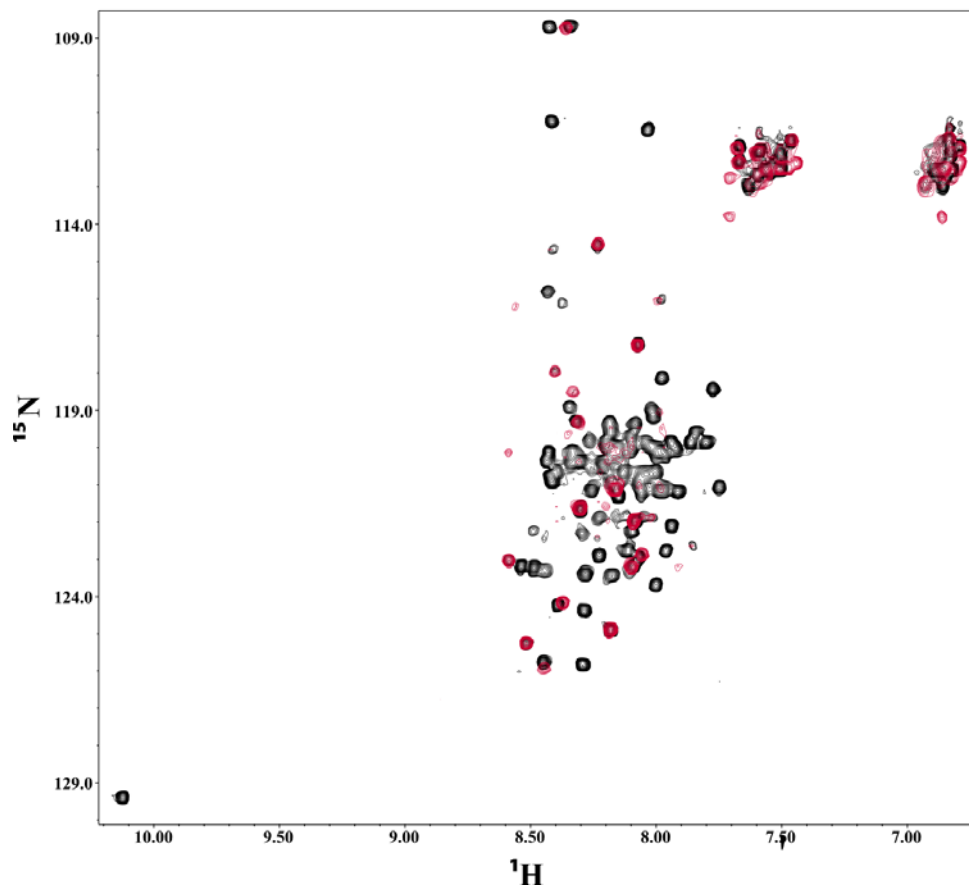
A score of 1 represents a fully helical structure and -1 represents a  $\beta$ -sheet structure. The score was calculated based on the  $C\alpha$  and  $C\beta$  chemical shifts obtained from the HNCACB and CBCA(CO)NH experiments.

to 65, of the Neh2 domain appear to be an extended helix, whereas, the C-terminus of the domain has a  $\beta$ -structure like propensity based on the assigned residues.

### 3.3.3 *Neh2 binding to the Kelch domain of the human Keap1*

To ensure that the purified recombinant Neh2 domain was folded properly, the binding of Neh2 to the Kelch domain of human Keap1 was examined. The overlay of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of the Neh2 domain in the free and Kelch bound states is shown in Figure 3.10. A significant amount of signal broadening is observed in the Kelch bound state spectrum suggesting that the Neh2 protein is able to interact with the Kelch domain of Keap1. Since it is known that the mouse Neh2 domain interacts with two independent Kelch domains, the complex formed is almost 80 kDa in size (Tong, Katoh et al. 2006). Assuming that the human Neh2 interacts with the Kelch domain in a 2:1 ratio similar to the mouse Neh2-Kelch binding, the formation of such a large complex can result in slowing the tumbling time of the Neh2 protein, thus providing an explanation for the observed disappearance of the resonance signals in the Neh2-Kelch bound state HSQC spectrum (Figure 3.10).

The Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG, low affinity binding motif and high affinity binding motif, respectively, were created to study their binding properties with the Kelch domain of the human Keap1 (Tong, Katoh et al. 2006) . An overlay of the two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of the wild-type Neh2 with the Neh2 $\Delta$ DLG and the Neh2 $\Delta$ ETGE is shown in Figure 3.11a&b. In the  $^1\text{H}$ - $^{15}\text{N}$  HSQC of the Neh2 $\Delta$ DLG, we observe that the first 33 residues deleted in the construct disappear in the HSQC spectrum. Additionally, peak broadening is observed for residues in the surrounding



**Figure 3.10. An overlay of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of Neh2 domain (black) and in complex with the Kelch domain of Keap1 (red).**

0.1 mM of  $^{15}\text{N}$ -labeled Neh2 full length protein was used to collect the Neh2 spectrum in the absence of Kelch. In the Kelch bound state, 0.2 mM of unlabeled Kelch domain of Keap1 was added to achieve a 1:2 ratio of Neh2 to Kelch



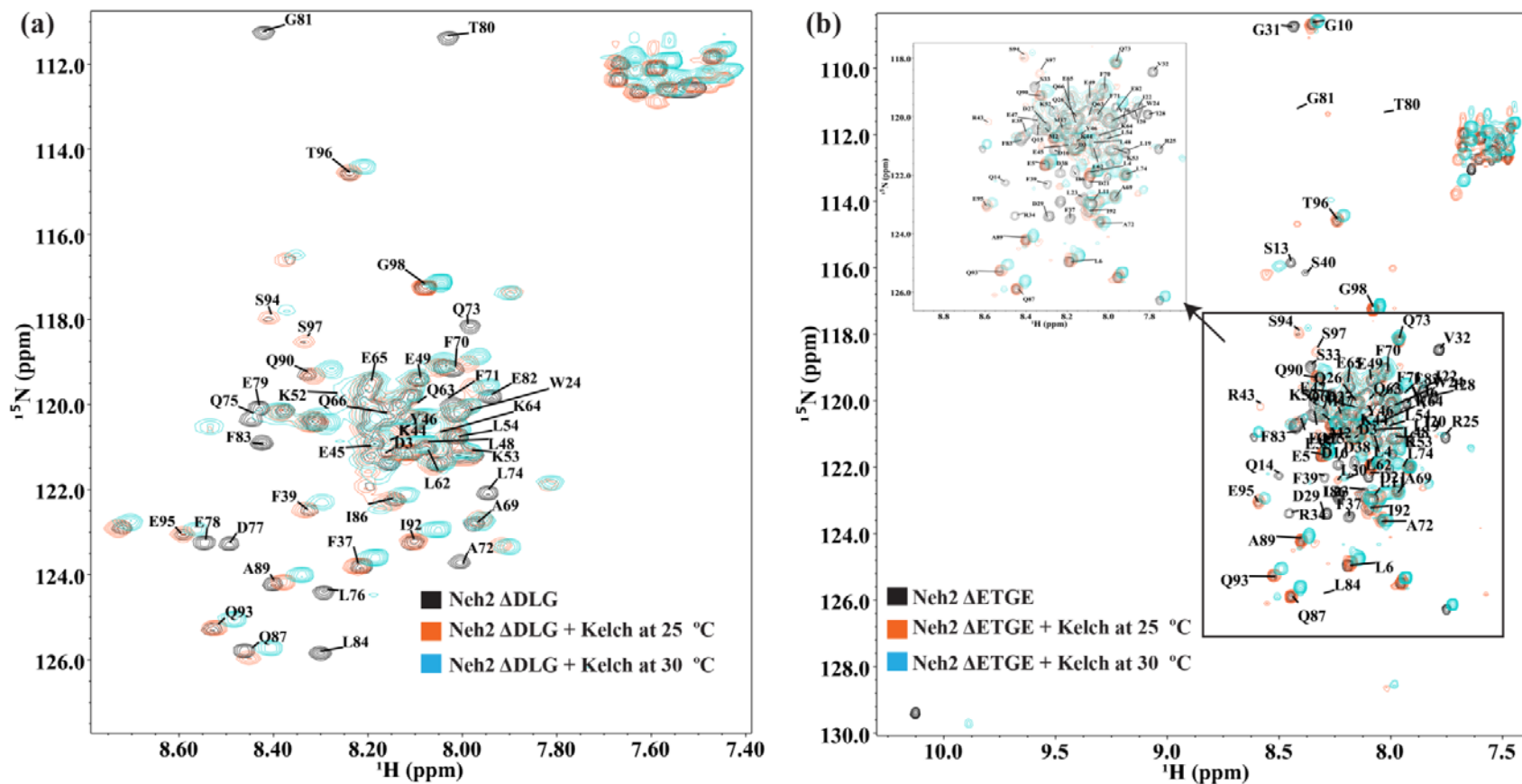


environment including R34, F37, D38 and R43. Interestingly, chemical shift attenuations for the C-terminal residues A69, F70, A72, Q73, L74, Q75, and L76 are also seen. The remaining residues that appear in the wild-type and the Neh2 $\Delta$ DLG spectra align well. The proton dimension of the Neh2 $\Delta$ DLG is also narrow like the wild-type Neh2 domain, representative of a disordered protein. This indicates that the deletion of the first thirty-three residues in the Neh2 domain does not alter the disorder nature of the Neh2 domain.

The  $^1\text{H}$ - $^{15}\text{N}$  HSQC overlay of the full length wild type Neh2 domain with the Neh2 $\Delta$ ETGE in Figure 3.11b reveals that peaks corresponding to the deleted residues E79, T80, G81, E82 disappear confirming that the backbone resonance assignment is accurate for these residues. Additional resonance signal loss is detected for residues Q75, L76, D77 and E78. Chemical shift changes are also observed in the C-terminal region for residues A69, A72, Q73, Q87, S97 and S94. Minimal signal perturbations are observed for N-terminal residues L11, D28, I28, I22, R25, S33 and S40. Like the Neh2 $\Delta$ DLG deletion, the ETGE deletion also has no impact on the disordered nature of the Neh2 domain as most of the protein peaks align well with the wild type spectrum in a narrow proton dimension. Therefore, the Neh2 $\Delta$ ETGE also retains the disordered structure of the wild-type Neh2 domain.

### ***3.3.4 Two site substrate recognition of the human Nrf2-Keap1 system***

The Neh2 $\Delta$ DLG and Neh2 $\Delta$ ETGE binding to the Kelch domain of the Keap1 protein were studied by NMR to ensure that the human Neh2 domain indeed interacts via the ETGE and DLG motifs with the Keap1 protein as observed for the mouse Nrf2-



**Figure 3.12.**  $^1\text{H}$ - $^{15}\text{N}$  HSQC of Neh2 variants in free and Kelch bound states at 25 °C and 30°C.

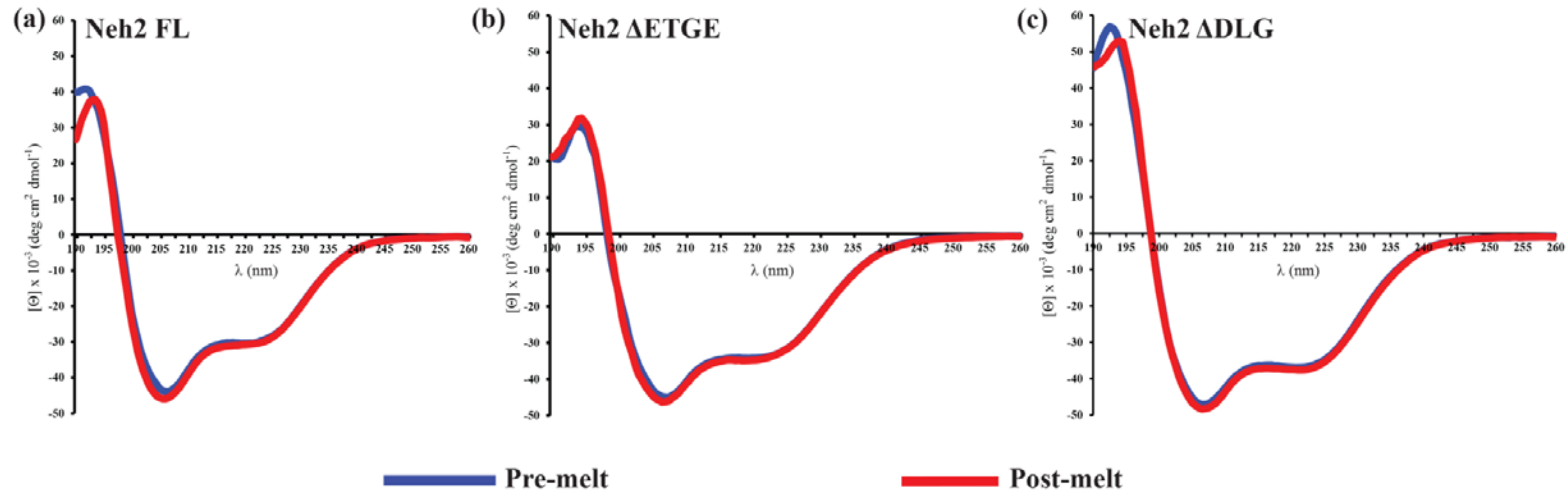
(a) Neh2 $\Delta$ DLG (black), Neh2  $\Delta$ DLG and Kelch at 25 °C (orange), and Neh2 $\Delta$ DLG and Kelch at 30 °C (cyan). (b) Neh2 $\Delta$ ETGE (black), Neh2 $\Delta$ ETGE and Kelch at 25 °C (orange), and Neh2 $\Delta$ ETGE and Kelch at 30 °C (cyan). The bound state of Neh2 variants with Kelch was in a 1:2 molar ratio, respectively.

Keap1 system (Tong, Katoh et al. 2006).  $^1\text{H}$ - $^{15}\text{N}$  HSQC overlay of the free and Kelch bound states of Neh2 $\Delta$ DLG and Neh2 $\Delta$ ETGE are depicted in Figure 3.11a&b. The Neh2 $\Delta$ DLG bound to Kelch at 25 °C shows that the resonance signals for L74, L76, D77, E78, E79, T80, G81, E82, L84 residues in the C-terminal region of the Neh2 domain undergo line broadening. This suggests that the ETGE motif of the Neh2 domain interacts with the human Kelch domain. In an attempt to minimize the line broadening effect, the temperature was increased by 5 °C.  $^1\text{H}$ - $^{15}\text{N}$  HSQC of the Neh2 $\Delta$ DLG bound to Kelch obtained at 30 °C (Figure 3.12a in cyan) indicates that the resonance signals are still not recovered for peaks that undergo line broadening at 25 °C. The temperature was not increased significantly as the Kelch domain is unstable and aggregates at higher temperatures.

Similarly, the  $^1\text{H}$ - $^{15}\text{N}$  HSQC of the Neh2 $\Delta$ ETGE-Kelch complex displays resonance signal loss for residues Q14, D16, I20, W24, I28, D29, L30, G31, V32, S33, R34, F37, F39 and S40. Chemical shift changes are also observed for L6, L11, S13, Q14, L23, and R25. Increasing the temperature to 30 °C reduced peak broadening. The indole ring of the tryptophan as well as a weak peak for Q14 is detected at 30 °C (Figure 3.12b in cyan). Remarkably, both the Neh2 $\Delta$ DLG and Neh2 $\Delta$ ETGE retain their signature spectra of a disordered protein, a narrow proton dimension, in the bound state with the Kelch domain. This indicates that both Neh2 $\Delta$ DLG and Neh2 $\Delta$ ETGE do not undergo a large structural transition, from a disordered to ordered state, in the target bound state.

### 3.3.5 *Secondary structure analysis of the Neh2 domain, Neh2 $\Delta$ DLG and Neh2 $\Delta$ ETGE*

Upon the deletion of the first thirty-three residues of the N-terminal, chemical shift changes were observed in the C-terminal region of the Neh2 domain. Similarly, deletion of the ETGE region led to minor changes in the N-terminal region as observed in the NMR HSQCs in Figure 3.11b. The data suggested that the N-terminal region maybe interacting with the C-terminal region of the protein. To study if there is any secondary structural elements in the Neh2 domain that maybe leading to this weak interaction between the two termini of the domain, the secondary structure of the Neh2 and its variants were studied by CD spectroscopy. The far-UV CD spectra of the Neh2, Neh2 $\Delta$ DLG and Neh2 $\Delta$ ETGE in Figure 3.13a,b&c indicate that the Neh2 domain is helical in nature. The secondary structure content of the Neh2 and its variants were quantitatively determined by CDSSTR analysis program available on the Dichroweb website (Whitmore and Wallace 2008). The results indicate that the Neh2 domain has 76% helical, 7%  $\beta$ -strand, 17% disordered/turn structure in 50 mM sodium phosphate buffer at pH 7.0. Based on the CD spectra the Neh2 $\Delta$ DLG contains 78% helical, 4%  $\beta$ -strand and 17% disordered/turn structure. The Neh2 $\Delta$ ETGE CD data analysis revealed 65% helical, 6%  $\beta$ -strand and 17% disordered/turn structure in 50 mM sodium phosphate buffer at pH 7.0. When a protein contains a region of  $\alpha$ -helical content, the spectral contribution from the helix may swamp the signal of other secondary structure content (Whitmore. and Wallace. 2007) . This is because, firstly the  $\phi/\psi$  angles of helices are well defined, and secondly, helical components produce very intense CD signals. Thus



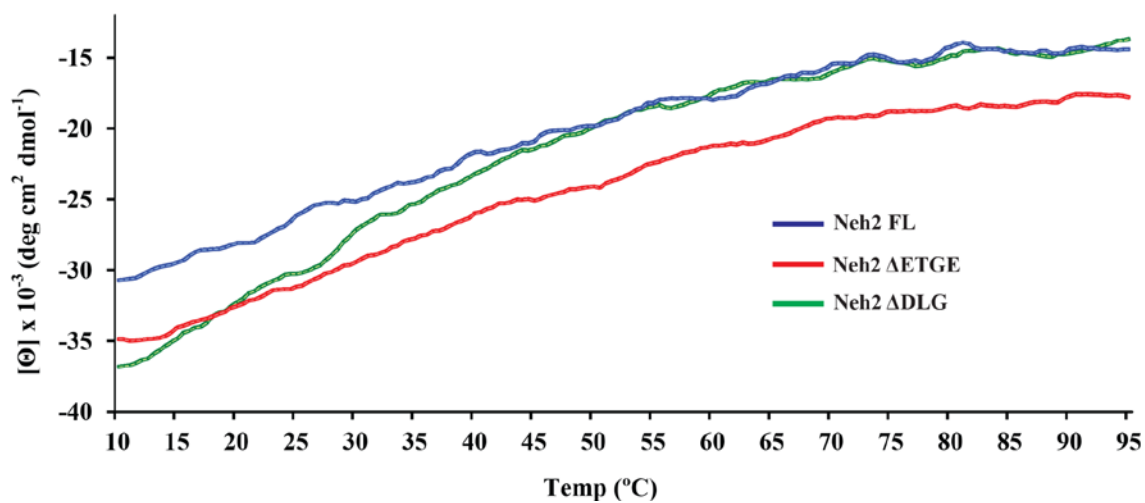
**Figure 3.13. Secondary structure of the Neh2 protein and its variants measured by far-UV CD spectropolarimetry.** The figure displays the spectra of (a) Neh2 FL (full length), (b) Neh2 $\Delta$ ETGE, (c) Neh2 $\Delta$ DLG. The far-UV CD spectra of Neh2 and its variants show a high degree of  $\alpha$ -helical structure.

the accuracy of the disordered content by CD is compromised (Whitmore and Wallace 2008). Hence, the observed high helical content in Neh2 and its variants could be resulting from the intense signal of the central helix. Interestingly, the  $\theta_{222/208}$  for the full length Neh2 is 0.72, 0.78 for Neh $\Delta$ ETGE, and 0.80 for Neh2 $\Delta$ DLG indicative of isolated non-interacting helices (Lavigne, Kondejewski et al. 1995).

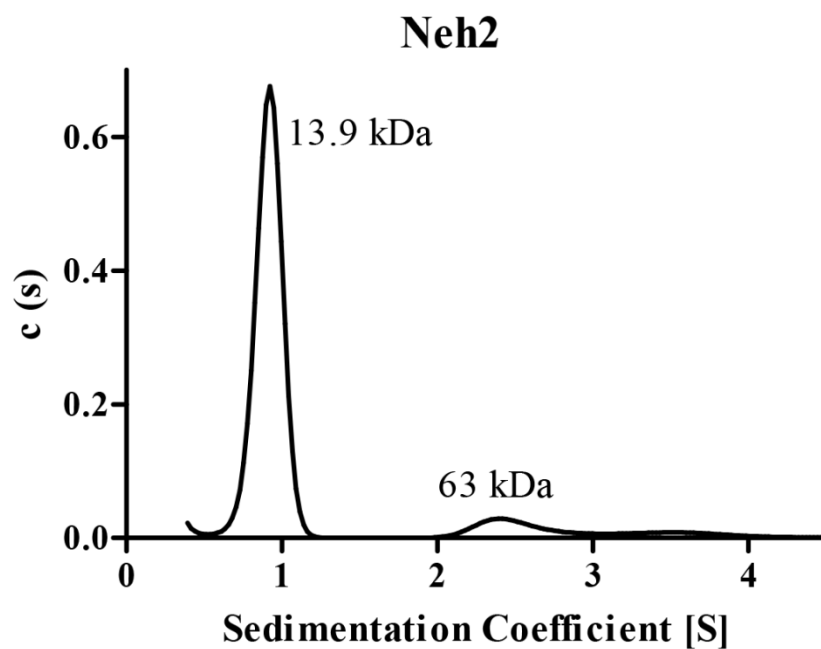
The stability of the helical structure for Neh2, Neh2 $\Delta$ DLG and Neh2 $\Delta$ ETGE was tested by thermal unfolding experiments. The protein was properly refolded after the thermal denaturation as shown in the overlay of the post melt CD spectra with the pre melt CD spectra in Figure 3.13a,b&c. Remarkably, the Neh2 domain and its two variants did not show a clear cooperative sigmoidal melt with the increase in solution temperature (Figure 3.14). The lack of a cooperative unfolding indicates the absence of a stable structure and thereby confirms the disordered nature of Neh2, Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG.

### **3.3.6 Hydrodynamic radius of Neh2 protein**

Sedimentation velocity, an analytical ultra-centrifugation (AUC) technique, is an alternative method to determine the molecular weight and hydrodynamic (Stokes) radius of proteins in solution. For the purpose of this study, sedimentation velocity technique was used to further enhance our understanding of the intrinsic nature of the Neh2 domain. Although the NMR data and the CD data agree that the Neh2 is a disordered protein with some helical structure, understanding of the hydrodynamic radius would provide further information about the compactness of the Neh2 domain. In the sedimentation velocity run, a 0.2 mg/mL concentration of the Neh2 domain was used. The data was fitted using



**Figure 3.14. Thermal melting of Neh2 FL and its variants Neh2 $\Delta$ ETGE and Neh2 $\Delta$ DLG at 222 nm.** Unfolding is monitored by the increase of  $[\Theta]$  as a function of temperature.



**Figure 3.15. Analytical ultracentrifugation experiment of**

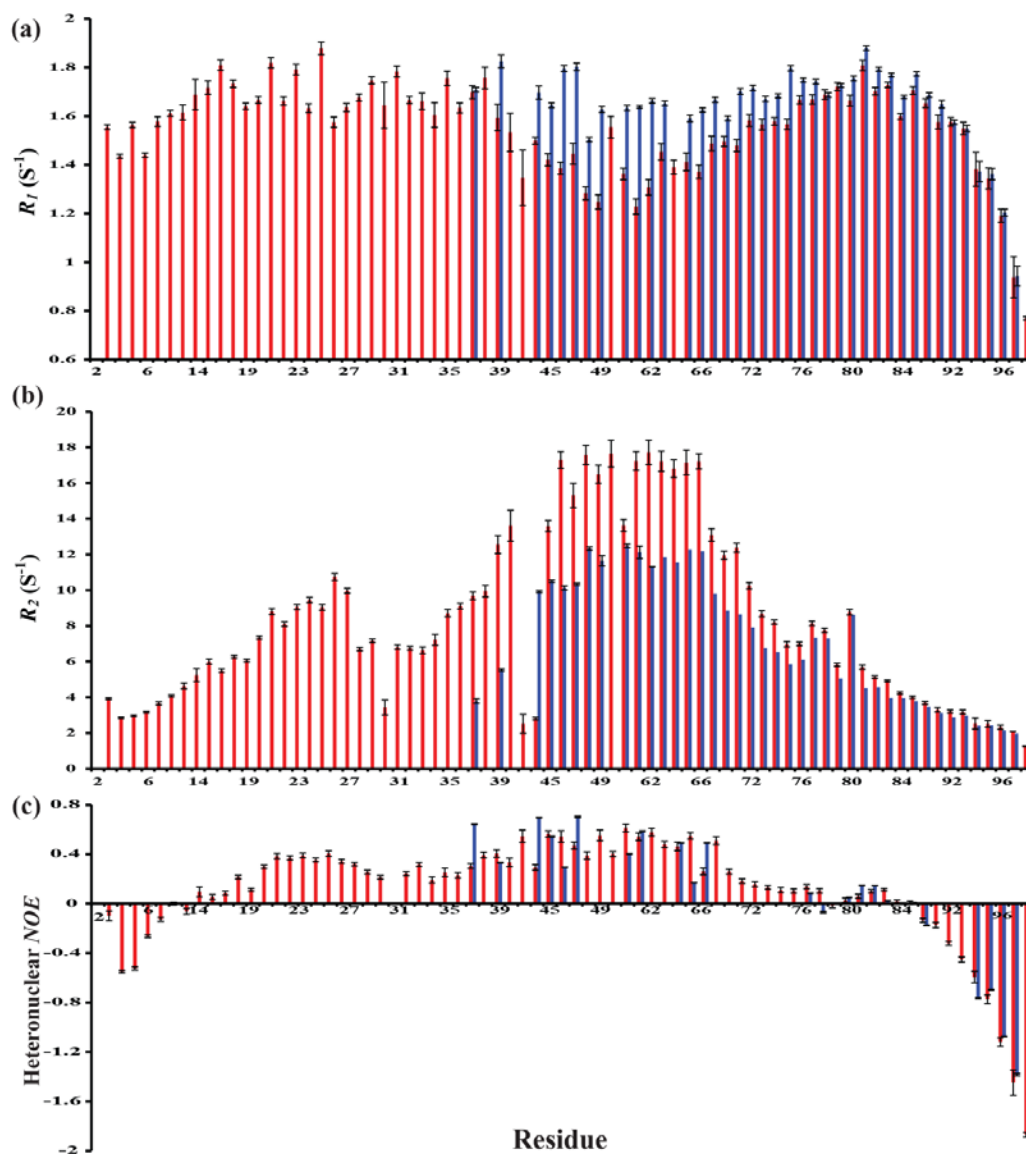
**Neh2 domain.** The sedimentation of Neh2 (40  $\mu$ M) was analyzed with a Beckman Coulter XL-I analytical ultracentrifuge at 4°C and at 60,000 rpm.



a size distribution analysis in a program called SEDFIT (Brown and Schuck 2006). The presence of more rapidly sedimenting peaks in Figure 3.15, corresponding to a molecular mass of 63 kDa, indicates the existence of aggregates in the sample. A peak with molecular mass of 13.9 kDa is also observed and this peak corresponds to the Neh2 domain. The SEDFIT program calculated the  $sw$  to be 0.7795,  $sw_{(20, w)}$  to be 0.8005 and a stokes radius of 2.55 nm (25.5 Å).

### 3.3.7 *Backbone Dynamics of the Neh2 protein*

The lack of chemical shift dispersion in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of the human Neh2 domain is typical of an unstructured protein. However, the SSP scores and the CD data indicate the presence of a pronounced helical as well as a  $\beta$ -sheet structure in the Neh2 domain. This suggests that Neh2 domain is amongst the biologically active disordered proteins that retain some secondary structure (Eliezer, Kutluay et al. 2001; Bienkiewicz, Adkins et al. 2002; Mark, Liao et al. 2005). To relate the average secondary structure with Neh2 dynamics, the amide backbone dynamics of the Neh2 domain and the Neh2 $\Delta$ DLG construct were studied by NMR relaxation experiments. The  $R_1$ ,  $R_2$ , and the steady-state heteronuclear NOE values provide information of motion on the picosecond-to-microsecond timescale thus helping identify regions restricted in motion and regions with high flexibility (Kay, Torchia et al. 1989). For the Neh2 domain, of the 84 assigned resonances, 74 were used for the measurement of the  $R_1$ ,  $R_2$  and the heteronuclear NOE values (Figures 3.16). The  $R_1$  values range from  $0.77 \pm 0.01$  s $^{-1}$  to  $1.88 \pm 0.03$  s $^{-1}$  with an average of  $1.56 \pm 0.03$  s $^{-1}$ . The  $R_2$  values range from  $1.25 \pm 0.03$  s $^{-1}$  to  $17.71 \pm 0.68$  s $^{-1}$  with an average of  $8.27 \pm 0.23$  s $^{-1}$ . The steady-state NOE values range from  $-1.87 \pm 0.02$  s $^{-1}$  to  $0.61 \pm 0.03$  s $^{-1}$  with an average of  $0.10 \pm 0.02$  s $^{-1}$ .



**Figure 3.16.** Backbone  $^{15}\text{N}$  NMR relaxation data for Neh2 full-length (*red*) and Neh2  $\Delta$ DLG (*blue*). (a) Longitudinal relaxation rate ( $R_1$ ); (b) transverse relaxation rate ( $R_2$ );  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear NOE.

The  $R_1$  values, of the 24 resonances of the Neh2 $\Delta$ DLG, range from  $0.78 \pm 0.003 \text{ s}^{-1}$  to  $1.88 \pm 0.01 \text{ s}^{-1}$  with an average of  $1.62 \pm 0.01 \text{ s}^{-1}$ . The  $R_2$  values range from  $1.12 \pm 0.024 \text{ s}^{-1}$  to  $12.47 \pm 0.17 \text{ s}^{-1}$  with an average of  $7.07 \pm 0.17 \text{ s}^{-1}$ . The steady-state NOE values range from  $-1.54 \pm 0.001 \text{ s}^{-1}$  to  $0.7 \pm 0.01 \text{ s}^{-1}$  with an average of  $0.004 \pm 0.003 \text{ s}^{-1}$ .

### 3.4 Discussion

The human Nrf2 protein is a major regulator of the oxidative stress response pathway. Limited number of studies have examined the Kelch-Nrf2 interaction in full-length context (Lo, Li et al. 2006; Tong, Katoh et al. 2006; Tong, Padmanabhan et al. 2007). In this work, a new high yield purification method for Neh2 domain of human Nrf2 allowed investigation of the Keap1-Nrf2 in full-length context.

Purification of the Neh2 domain from the soluble fraction yielded about ~0.8 mg of impure protein from 1 L of M9 bacterial culture (Figure 3.3a), therefore, not suitable for structural and binding experimental studies. One of the main objectives achieved in this part of the project was the increase in the final yield of Neh2 by more than 40 fold. The purity of the Neh2 in the final sample was also enhanced significantly (Figure 3.5b line 3). Identical NMR HSQC spectra were obtained through the different protein purification methods, suggesting that the final structure of the Neh2 is not compromised with the high yield purification procedure. Additionally, the Neh2 obtained with the revised purification method is functional as it is capable of binding to the Kelch domain of Keap1 (Figure 3.10).

Structural properties of the Neh2 domain were also assessed to ensure that it behaves similarly to the mouse Neh2 domain of Nrf2 (Tong, Katoh et al. 2006). NMR

data revealed that the human Neh2 domain is intrinsically disordered. However, analyses of the secondary structure propensity (SSP) scores show the N-terminal region with an  $\alpha$ -helical tendency, and central region to be helical in structure. It has been recently reported that the N-terminal region forms a helical structure upon binding to Keap1 (Fukutomi, Takagi et al. 2014). The C-terminal region shows a tendency to form  $\beta$ -sheet like structure. This data is consistent with what has been observed for the mouse Neh2 domain (Tong, Katoh et al. 2006) and also with the PrDOS prediction (Figure 3.2). Residues 39 to 71 of the mouse Neh2 domain have been shown to form a central helix (Tong, Katoh et al. 2006). Residues between 10 to 30 also have CSI values close to 1 for the mouse Neh2 domain indicating a tendency towards forming an  $\alpha$ -helical structure (Tong, Katoh et al. 2006). It has also been known from the crystal structures of the Neh2 domain bound to the Kelch domain of Keap1 that both the mouse and human ETGE regions form a  $\beta$ -hairpin loop (Lo, Li et al. 2006; Padmanabhan, Tong et al. 2006). The ETGE motif of the Neh2 domain of human Nrf2 has also been shown to have a tendency to form a  $\beta$ -turn structure in its free form (Cino, Wong-ekkabut et al. 2011). Thus the SSP scores obtained for the full-length Neh2 domain of human Nrf2 are in a good agreement with previously published data and confirm that the Neh2 domain has a partial secondary structure.

Two constructs corresponding to the low binding affinity motif, Neh2 $\Delta$ ETGE, and the high binding affinity motif, Neh2 $\Delta$ DLG, were constructed to study the two site binding of the Neh2 domain with the Kelch domain of Keap1 (Tong, Katoh et al. 2006). Binding of the full-length Neh2 domain to Kelch undergoes substantial line broadening which could be attributed to the increase in molecular tumbling time due to the formation

of a large complex (~80 kDa) and to the intermediate conformational exchange between the free and bound states (Tong, Katoh et al. 2006). However, the two constructs, form a complex of approximately 44 kDa each when bound to the Kelch domain individually. ProT $\alpha$  in the bound state to Kelch forms a complex of 46 kDa. The bound state  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of the ProT $\alpha$  shows no resonance broadening (Khan, Cino et al. 2013). When compared to ProT $\alpha$ , the bound state  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of the Neh2 $\Delta$ DLG and Neh $\Delta$ ETGE (Figure 3.12 a&b) show dramatic signal broadening, even at higher temperatures. The observed line broadening effect can mostly be ascribed to the intermediate chemical exchange effects of the bound and the unbound forms of the Neh2 variants.

The Neh2 $\Delta$ ETGE-Kelch bound HSQC spectrum shows peak attenuations in flanking residues of the previously identified DLG motif (Tong, Katoh et al. 2006; Tong, Padmanabhan et al. 2007). Recently, it was reported that residues M17 to Y46 are part of the DLG motif and crucial for interaction with the Keap1 protein (Fukutomi, Takagi et al. 2014). The observed chemical shift changes for the Neh2 $\Delta$ ETGE-Kelch are in par with this recent finding and emphasize the study of ligand binding in full-length proteins.

Interestingly, NMR HSQC of the Neh2 $\Delta$ DLG and Neh2 $\Delta$ ETGE reveal that upon introducing these deletions, the Neh2 domain retains its intrinsic disordered structure (Figure 3.11). A narrow spectra in the proton dimension is also observed in the Kelch bound state for both Neh2 constructs (Figure 3.12) suggesting that Neh2 does not undergo a substantial structural transition in the target bound state similar to ProT $\alpha$  (Khan, Cino et al. 2013). Moreover, the HSQC spectra of the two Neh2 variants in the free and the Kelch bound state revealed the possibility that the N and C termini of the

Neh2 domain maybe involved in interactions. The CD data of the Neh2 domain indicates the presence of a pronounced helical structure in the Neh2 domain and its variants. However, the  $\theta_{222/208}$  ratios below 1 indicate that the observed helical spectra are of a protein containing independent non-interacting helices (Lavigne, Kondejewski et al. 1995).

The compactness and the flexibility of the Neh2 domain were examined to explain the relation between the N and C termini. Sedimentation velocity experiments revealed that the Neh2 has a hydrodynamic radius of around 25.5 Å. For a structured protein with similar number of residues as that of the Neh2 domain, the hydrodynamic radius of the folded state is calculated to be 19 Å and for the fully denatured state the hydrodynamic radius of the same protein is around 32 Å (Wilkins, Grimshaw et al. 1999). ProTα is also similar in size as the Neh2 domain and has a hydrodynamic radius of around 34 Å (Yi, Boys et al. 2007). The fact that the hydrodynamic radius of the Neh2 domain is less than that of the ProTα and other denatured proteins of similar size strongly suggests that it is more compact in shape than the average unstructured protein of the same size. Based on this it is possible that the N and C terminal residues maybe involved in long range interactions.

The relaxation rates and the heteronuclear *NOEs* shed further light on the flexibility and compactness of the human Neh2 domain. The average  $R_1$  rate observed for the Neh2 domain is  $1.56 \pm 0.03 \text{ s}^{-1}$ .  $R_1$  values for Neh2 domain fall within the expected range of a folded protein of the same size. For instance, the average  $R_1$  rate for the globular 89 amino acid Barstar is close to  $2 \text{ s}^{-1}$ , obtained at 300 K, under 600 MHz

magnetic field (Sahu, Bhuyan et al. 2000). When compared to ProT $\alpha$  (average  $R_1$  value of  $2.45 \text{ s}^{-1}$ ), Neh2 domain is more restricted in motion (Khan, Cino et al. 2013).

For a globular protein with the same size as the Neh2 domain, the average  $R_2$  values are approximately  $10 \text{ s}^{-1}$  (Walma, Aelen et al. 2004). The high  $R_2$  rates of the Neh2 (average  $\sim 8 \text{ s}^{-1}$ ) suggest that this domain is structured. However, the N and C termini of the Neh2 domain have  $R_2$  values ranging from  $1.25 \text{ s}^{-1}$  to  $6.0 \text{ s}^{-1}$  and are comparable to ProT $\alpha$  (Khan, Cino et al. 2013). Whereas, residues 19 to 80 (with the exception of residue 30, 43 and 44) have  $R_2$  values that range from  $7 \text{ s}^{-1}$  to  $17.7 \text{ s}^{-1}$ . The high  $R_2$  values for this helical region imply that it is dynamically restricted or experiences conformational exchange, or both.

The average  $R_2$  values for the 99 residue  $\beta_2$ -microglobulin, obtained under same magnetic field strength and temperature, are also very high, ranging from  $7 \text{ s}^{-1}$  to  $18 \text{ s}^{-1}$  (Platt, McParland et al. 2005). Experimental evidence suggests that these high  $R_2$  values for  $\beta_2$ -microglobulin are a result of long range and non-native interactions between the aromatic residues found in the central region of the protein (Platt, McParland et al. 2005). In the Neh2 domain, we do not observe large cluster of aromatic residues; however, the phenylalanine residues, F37, F39, F70 and F71, and one Tyr 46 could possibly be involved in long range interactions. These long range interactions, if any, would provide an explanation for the observed motional restriction of the central helical region as well as the lower hydrodynamic radius of the Neh2 domain.

The steady-state  $^1\text{H}$ - $^{15}\text{N}$  NOE values are also in agreement with the  $R_2$  rates. The negative NOE values of the N and C termini indicate these regions to be lacking

structure, therefore, are highly flexible. Whereas the positive NOE values of the central region confirm the motional restriction of the helix.

The relaxation rates reveal no significant difference between the  $R_1$  values of the full-length Neh2 domain and the Neh2 $\Delta$ DLG. The  $R_2$  values for residues 38 to 80 of the Neh2 $\Delta$ DLG are higher than what is expected of a globular protein with the same size, confirming the lack of flexibility in this region or conformation exchange (Sahu, Bhuyan et al. 2000). The NOE values of the central region increase slightly, making this region more restricted in motion, whereas, the C-terminal region remains disordered as observed for the wild type Neh2. The relaxation data also confirms that deleting the first thirty-three residues does not lead to folding of the Neh2 domain.

The work completed in this chapter allowed us to directly investigate the interaction of the full-length Neh2 domain of human Nrf2 with the Kelch domain in chapter 4. Recently, there has been great interest in drug therapies using the Neh2 domain as a target (Suzuki, Motohashi et al. 2013). Binding studies with the full-length Neh2 domain can be used to study the effect of potential drug targets on the Nrf2-Keap1 system. Additionally, in Chapter 1, positive and negative regulators such as p62 and p21<sup>Cip1/WAF1</sup> of the Nrf2 been mentioned. This work lays the foundation for direct binding studies with p21<sup>Cip1/WAF1</sup>, which competes for Keap1 binding with Nrf2 (Taguchi, Motohashi et al. 2011).



### 3.5 References

- Bienkiewicz, E. A., J. N. Adkins, et al. (2002). "Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1)." Biochemistry **41**(3): 752-759.
- Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Anal Biochem **72**: 248-254.
- Briere, L. K. and S. D. Dunn (2006). "The periplasmic domains of Escherichia coli HflKC oligomerize through right-handed coiled-coil interactions." Biochemistry **45**(28): 8607-8616.
- Brown, P. H. and P. Schuck (2006). "Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation." Biophys J **90**(12): 4651-4661.
- Cino, E. A., J. Wong-ekkabut, et al. (2011). "Microsecond molecular dynamics simulations of intrinsically disordered proteins involved in the oxidative stress response." PLoS One **6**(11): e27371.
- Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-293.
- Eliezer, D., E. Kutluay, et al. (2001). "Conformational properties of alpha-synuclein in its free and lipid-associated states." J Mol Biol **307**(4): 1061-1073.
- Fukutomi, T., K. Takagi, et al. (2014). "Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degron and Keap1." Mol Cell Biol **34**(5): 832-846.
- Ishida, T. and K. Kinoshita (2007). "PrDOS: prediction of disordered protein regions from amino acid sequence." Nucleic Acids Res **35**(Web Server issue): W460-464.
- Itoh, K., N. Wakabayashi, et al. (1999). "Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain." Genes Dev **13**(1): 76-86.
- Johnson, B. A. (2004). "Using NMRView to visualize and analyze the NMR spectra of macromolecules." Methods Mol Biol **278**: 313-352.
- Kay, L. E., D. A. Torchia, et al. (1989). "Backbone dynamics of proteins as studied by <sup>15</sup>N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease." Biochemistry **28**(23): 8972-8979.
- Keller, R. (2004). "Computer Aided Resonance Assignment Tutorial." CANTINA Verlag.
- Khan, H., E. A. Cino, et al. (2013). "Fuzzy complex formation between the intrinsically disordered prothymosin alpha and the Kelch domain of Keap1 involved in the oxidative stress response." J Mol Biol **425**(6): 1011-1027.

- Lavigne, P., L. H. Kondejewski, et al. (1995). "Preferential heterodimeric parallel coiled-coil formation by synthetic Max and c-Myc leucine zippers: a description of putative electrostatic interactions responsible for the specificity of heterodimerization." J Mol Biol **254**(3): 505-520.
- Lees, J. G., A. J. Miles, et al. (2006). "A reference database for circular dichroism spectroscopy covering fold and secondary structure space." Bioinformatics **22**(16): 1955-1962.
- Lewis, M. S., Shrager, R.I., and Kim, S.J. (1994). Modern Analytical Ultracentrifugation: Acquisition and Interpretation of Data for Biological and Synthetic Polymer Systems. **1** 94-115.
- Lo, S. C., X. Li, et al. (2006). "Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling." Embo J **25**(15): 3605-3617.
- Mark, W. Y., J. C. Liao, et al. (2005). "Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions?" J Mol Biol **345**(2): 275-287.
- Marsh, J. A., V. K. Singh, et al. (2006). "Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation." Protein Sci **15**(12): 2795-2804.
- Moi, P., K. Chan, et al. (1994). "Isolation of NF-E2-related factor 2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the beta-globin locus control region." Proc Natl Acad Sci U S A **91**(21): 9926-9930.
- Padmanabhan, B., K. I. Tong, et al. (2006). "Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer." Mol Cell **21**(5): 689-700.
- Platt, G. W., V. J. McParland, et al. (2005). "Dynamics in the unfolded state of beta2-microglobulin studied by NMR." J Mol Biol **346**(1): 279-294.
- Sahu, S. C., A. K. Bhuyan, et al. (2000). "Backbone dynamics of barstar: a (15)N NMR relaxation study." Proteins **41**(4): 460-474.
- Sattler, M., Schleucher, J., and Griesinger, C. (1999). "Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients." Progress in NMR spec **34**: 93-158.
- Suzuki, T., H. Motohashi, et al. (2013). "Toward clinical application of the Keap1-Nrf2 pathway." Trends Pharmacol Sci **34**(6): 340-346.
- Taguchi, K., H. Motohashi, et al. (2011). "Molecular mechanisms of the Keap1-Nrf2 pathway in stress response and cancer evolution." Genes Cells **16**(2): 123-140.
- Tong, K. I., Y. Katoh, et al. (2006). "Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model." Mol Cell Biol **26**(8): 2887-2900.

- Tong, K. I., B. Padmanabhan, et al. (2007). "Different electrostatic potentials define ETGE and DLG motifs as hinge and latch in oxidative stress response." Mol Cell Biol **27**(21): 7511-7521.
- Tong, K. I., M. Yamamoto, et al. (2008). "A simple method for amino acid selective isotope labeling of recombinant proteins in *E. coli*." J Biomol NMR **42**(1): 59-67.
- Walma, T., J. Aelen, et al. (2004). "A closed binding pocket and global destabilization modify the binding properties of an alternatively spliced form of the second PDZ domain of PTP-BL." Structure **12**(1): 11-20.
- Whitmore, L. and B. A. Wallace (2008). "Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases." Biopolymers **89**(5): 392-400.
- Whitmore., L. and B. A. Wallace. (2007). "Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases." Wiley InterScience ([www.interscience.wiley.com](http://www.interscience.wiley.com)) **89**(5): 392-400.
- Wilkins, D. K., S. B. Grimshaw, et al. (1999). "Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques." Biochemistry **38**(50): 16424-16431.
- Yi, S., B. L. Boys, et al. (2007). "Effects of zinc binding on the structure and dynamics of the intrinsically disordered protein prothymosin alpha: evidence for metalation as an entropic switch." Biochemistry **46**(45): 13120-13130.

## Chapter 4

### **Molecular effects of cancer-associated somatic mutations on the structural and target recognition properties of Keap1**

#### **4.1 Introduction**

The Nrf2/Keap1 antioxidant pathway plays critical roles both in chemoprevention and cancer pathogenesis (Kwak and Kensler 2010; Müller and Hengstermann 2012; Jaramillo and Zhang 2013; Suzuki, Motohashi et al. 2013). The transcription factor Nuclear factor erythroid 2-related factor-2 (Nrf2) induces the cellular defenses against reactive oxygen species and toxic substances by initiating the transcription of an array of cytoprotective genes, which are essential for the elimination of electrophilic and oxidative stresses before damage is caused to the cellular macromolecules (Itoh, Chiba et al. 1997; Itoh, Wakabayashi et al. 1999). Under normal conditions, its activity is suppressed by Keap1 (Kelch-like ECH-associated protein 1) in conjunction with the Cullin 3 (Cul3) E3 ubiquitin ligase complex via ubiquitin-mediated proteasomal degradation (Itoh, Wakabayashi et al. 1999; Kobayashi, Kang et al. 2004; Zhang, Lo et al. 2004; McMahon, Thomas et al. 2006). In the presence of oxidative stress, however, specific cysteine residues in Keap1 are subjected to modifications by the oxidants (Zhang and Hannink 2003). The modified Keap1 is incapable of targeting Nrf2 for ubiquitination, allowing Nrf2 to translocate into the nucleus to initiate gene transcription (Itoh, Chiba et al. 1997).

Accumulation of Nrf2 in the nucleus however, can lead to aberrant cytoprotective gene expression if not regulated (McMahon, Itoh et al. 2003; Singh, Misra et al. 2006; Hayes and McMahon 2009). Keap1 is a cytosolic protein that lacks a nuclear localization signal and cannot enter the nucleus on its own to regulate Nrf2 levels (Niture and Jaiswal 2009). Recent studies have revealed that the Keap1 protein can be shuttled to the nucleus in complex with ProT $\alpha$ , an intrinsically disordered protein that contains a nuclear localization signal (Niture and Jaiswal 2009). Once in the nucleus, Keap1 mediates ubiquitination and proteasomal degradation of Nrf2, causing Nrf2 to return to its basal level. This regulatory mechanism is essential as the continuous expression of cytoprotective proteins can promote cell proliferation leading to tumorigenesis (Ohta, Iijima et al. 2008; Shibata, Kokubu et al. 2008; Jaramillo and Zhang 2013).

As mentioned in chapter 1, the three main functional domains of Keap1, called the Broad complex, Tramtrack and Bric-a-Brac (BTB) domain, the intervening region (IVR), and the Kelch domain, all contribute to Nrf2 repression. The N-terminal BTB domain is responsible for Keap1 dimerization. This domain, together with the IVR domain, mediates the Keap1 interaction with the Cul3-E3 ligase complex required for the ubiquitination of Nrf2 (Cullinan, Gordan et al. 2004; Kobayashi, Kang et al. 2004; Zhang, Lo et al. 2004). The C-terminal Kelch domain is responsible for target recognition. Besides Nrf2 and ProT $\alpha$ , this domain has also been shown to interact with WTX (Camp, James et al. 2012), p62 (Komatsu, Kurokawa et al. 2010), PGAM5 (Lo and Hannink 2006), PALB2 (Ma, Cai et al. 2012), FAC1 (Strachan, Morgan et al. 2004), and IKK $\beta$  (Kim, You et al. 2010) to name a few (Hast, Goldfarb et al. 2013). In the case of Nrf2, the N-terminal Neh2 domain of the protein harbors two Kelch-binding motifs via

the ETGE and DLG motifs (Tong, Katoh et al. 2006; Tong, Kobayashi et al. 2006). This binding mechanism allows the lysine residues located in between these two motifs in the Neh2 domain to be properly positioned for ubiquitination (McMahon, Thomas et al. 2006; Tong, Katoh et al. 2006; Tong, Kobayashi et al. 2006). On the other hand, other targets of Keap1 like ProT $\alpha$  form a complex with the Kelch domain in a one to one stoichiometry (Khan, Cino et al. 2013).

Over 200 somatic mutations within Keap1 have been reported in various types of cancer tissues and cancer-derived cell lines (Singh, Misra et al. 2006; Ohta, Iijima et al. 2008; Shibata, Kokubu et al. 2008; Network 2012; Yoo, Kim et al. 2012; Hast, Cloer et al. 2014). Several of them have been shown to be defective in inhibiting Nrf2 activity (Ohta, Iijima et al. 2008; Shibata, Kokubu et al. 2008; Suzuki, Maher et al. 2011), resulting in constitutive expression of cytoprotective enzymes in cancer cells thus promoting cell survival (Hayes and McMahon 2009). It has been speculated that these mutations compromise Keap1's structural integrity, rendering it incapable of promoting the ubiquitination of Nrf2 (Lo, Li et al. 2006; Hayes and McMahon 2009). However, only a few studies of the structural and target-binding properties of these mutants have been carried out to date (Lo, Li et al. 2006; Hast, Cloer et al. 2014). Therefore, the molecular links between these mutations and the diseases remain poorly understood.

To address this problem, we set out to elucidate the effects of disease-associating missense mutations on the structural integrity and target recognition of Keap1 by using nuclear magnetic resonance (NMR) spectroscopy, CD spectropolarimetry and ITC. We have chosen to focus on the Kelch domain of Keap1 in this study. Importantly, close to 50% of the missense somatic mutations identified in Keap1 so far are found to reside in

this domain (based on the data in the database of the Catalogue of Somatic Mutations in Cancer-COSMIC (Forbes, Bindal et al. 2011). Crystallographic studies show that the Kelch domain adopts a six-bladed  $\beta$ -propeller conformation with several conserved amino acids that are important for maintaining the hydrogen bond network, linking the propeller blades and forming the hydrophobic core of the protein (Li, Zhang et al. 2004; Lo, Li et al. 2006; Padmanabhan, Tong et al. 2006). In particular, there is a conserved glycine doublet found in each blade of the Kelch domain. These two glycine residues are involved in an intrablade hydrogen bond network that may be instrumental in the folding of individual blade structures. Both the ETGE and DLG motifs of Nrf2, as well as ProT $\alpha$  are found to bind to the same positively charged pocket located at the mouth of the central tunnel of the Kelch  $\beta$ -propeller (Lo, Li et al. 2006; Padmanabhan, Nakamura et al. 2008).

We have selected nine cancer-associated somatic mutations, G333C, G350S, G364C, G379D, R413L, R415G, A427V, G430C, G476R, to investigate their effects on the target recognition and structural integrity of the Kelch domain of Keap1 (Table 4.1) (Padmanabhan, Tong et al. 2006; Singh, Misra et al. 2006; Ohta, Iijima et al. 2008; Shibata, Kokubu et al. 2008; Hayes and McMahon 2009; Suzuki, Maher et al. 2011; Yoo, Kim et al. 2012). These mutants can be divided into three categories based on their location in the Kelch domain. The R415 residue is one of the conserved Arg residues located in the positively charged vicinity in the bottom face of the propeller (Figure 4.1a)

**Table 4.1. Somatic mutations included in this study.**

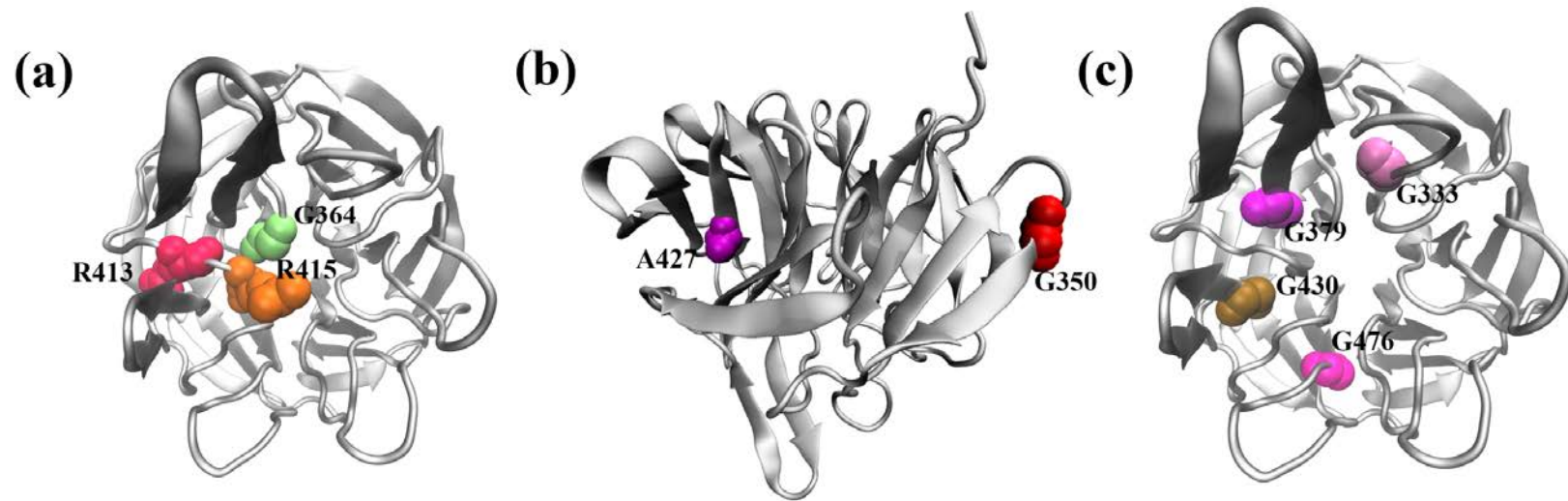
<b>Type</b>	<b>Location in Kelch</b>	<b>Specimen</b>	<b>Cancer Type</b>
<b>G364C</b>	Blade II	Cell line H1648	Lung (ADC)
<b>R415G</b>	Blade III	Patient	Lung (ADC)
<b>R413L</b>	Blade III	Patient	Lung
<b>G350S</b>	Blade I	Cell lines and Patient	Lung
<b>A427V</b>	Blade III	Patient	Lung (SCC)
<b>G333C</b>	Blade I	Cell line A549	Lung
<b>G379D</b>	Blade II	Patient	Gall Bladder (ADC)
<b>G430C</b>	Blade III	Patient	Lung (ADC)
<b>G476R</b>	Blade IV	Patient	Lung (LCC)

Abbreviations: ADC: Adenocarcinoma, SCC: Small cell carcinoma, LCC: Large cell carcinoma. References: Lo et al (2006) *EMBO J*, 25, 3605-3617; Hayes & McMahon (2009) *Trends Biochem Sci*, 34, 176-188.



and it is directly involved in the interface formation with ProT $\alpha$  and Nrf2 (Lo, Li et al. 2006; Padmanabhan, Tong et al. 2006; Padmanabhan, Nakamura et al. 2008). The G364C and R413L mutants are also included in the same category as the R415G. Even though these two residues do not directly participate in target binding, they are both next to residues that are involved in the ProT $\alpha$  and Nrf2 interactions (Figure 4.1a) (Lo, Li et al. 2006; Padmanabhan, Tong et al. 2006; Padmanabhan, Nakamura et al. 2008). Conversely, the second set of mutations, G350S and A427V, are distant from the binding interface. The G350S as shown in Figure 4.1b is at the opposite side of the basic pocket, and the A427V is found within a  $\beta$ -strand of blade III of the propeller. The last group, containing G333C, G379D, G430C, and G476R, are mutations that occur in the conserved double glycine repeats located in different blades (Figure 4.1c) (Li, Zhang et al. 2004). Three of these four mutations, G333C, G379D and G430C, occur in the second position of the GG repeat while the G476R occurs in the first position of the doublet. Since these glycine residues are expected to play an important structural role in the Kelch domain, it is important to assess the impacts of these mutations on the domain's structural integrity and function.

NMR and ITC were used to examine the interactions between Nrf2/ProT $\alpha$  and the Kelch mutants. Nrf2 and ProT $\alpha$  have been shown to have the strongest and the weakest binding affinity, respectively, to the wild-type Kelch domain among all the known partners (Cino, Killoran et al. 2013). Additionally, we have completed the backbone NMR chemical shift assignment of the 34 kDa Kelch domain of human Keap1 (Cino, Killoran et al. 2013). This affords an effective way to probe the conformational changes



**Figure 4.1. Mutation sites located in the Kelch domain of Keap1 (PDB: 1U6D (Li, Zhang et al. 2004)).**

- (a) Somatic mutations occurring in the binding interface. (b) Mutations found outside the binding interface.  
(c) Mutations in the conserved double glycine repeats

induced upon mutating specific residues in the Kelch domain and provides a manner to observe regions important for target binding. Our results show that these somatic mutations exert differential effects on the structural integrity and target recognition properties of Keap1, implying that their molecular links to the cancer pathogenesis can be very distinct.

## 4.2 Materials and Methods

### 4.2.1 Cloning, expression and purification of Keap1-Kelch variants

Mutant constructs of the Kelch domain were generated using the QuikChange II site-directed mutagenesis kit (Stratagene) with the wild-type human Kelch domain in the prokaryote expression vector pDEST17 (Life Technologies). The following are the primers used for the mutations:

G333C-5'-ctacaccgctggctgctacttccgaca-3';

G350S-5'-caaccccagtacagcacctggctccg-3';

G364C-5'-aggtgccgaggagctgcctggccggctgcg-3';

G379D-5'-gtacgccgtgggacaggaacaactcgcc-3';

R413L-5'-atgagcgtgccccttaaccgcatcggg-3';

R415G-5'-agcgtgccccgtaacggcatcggggtgg-3';

A427V-5'-gatggccacatctatgtcgtcggcggt-3';

G430C-5'-atatgccgtcggctgctcccacggctg-3';

G476R-5'-tgctttatgccgtgaggggctttgacggg-3'.

Wild-type and mutational variants were expressed and purified following the protocols described in (Khan, Cino et al. 2013).

#### **4.2.2 *Neh2 constructs, expression and purification***

The gene fragment of human Neh2 domain of Nrf2 (amino acid residues M1-G98) and Neh2 $\Delta$ DLG (deletion of residues 1-33 from Neh2) were cloned out from the full length Nrf2 (purchased from Invitrogen; NM\_006164) and then inserted into the Gateway pDEST17 vector. The Neh2 $\Delta$ ETGE mutant (deletion of <sup>79</sup>ETGE<sup>82</sup> from Neh2) was constructed from the full-length Neh2 fragment using the QuikChange II site-directed mutagenesis kit (Stratagene). For each construct, a Tobacco Etch virus (Diaz-Jullien, Perez-Estevez et al.) cleavage recognition site, ENLYFQG, was inserted between the histidine tag and the Neh2 fragment. All three constructs were transformed into the Rosetta 2 (DE3)pLysS (Novagen) strain of *Escherichia coli* (*E. coli*) for protein expression in M9 media. Cells were grown at 37°C and protein over-expression was induced at an OD<sub>600</sub> of 0.6 with 0.5 mM of isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG, BioShop). The cells were allowed to grow overnight at 25°C before harvesting. The Neh2 protein and its variants were purified batch wise by Nickel-Sepharose (GE Healthcare) affinity chromatography, followed by his-tag removal using TEV protease at 25°C for 21 hours. Uncleaved Neh2 and the his-tagged TEV protease were then removed by Nickel-Sepharose affinity chromatography. The final protein sample was analyzed on

SDS-PAGE for purity and dialyzed into 50 mM sodium phosphate buffer, 100 mM NaCl, 1 mM DTT, pH 7.0 for NMR and ITC experiments.

### **4.2.3 Peptides**

The ProT $\alpha$ -ENGE peptide (> 98% purity), APANGNANEENGEQEADNEV, and the Neh2-ETGE peptide (> 98% purity), AFFAQLQLDEETGEFL, were ordered from GenScript USA Inc. The Neh2-DLG peptide, IDILWRQDIDLGVSREVFDF, was ordered from the Tufts University Core Facility with a purity level above 95%. The lyophilized peptides were dissolved and dialyzed into 50 mM sodium phosphate buffer, 100 mM NaCl, and 1 mM DTT for ITC and NMR experiments.

### **4.2.4 Nuclear magnetic resonance (NMR) experiments**

For the backbone resonance assignment of the Kelch domain of human Keap1, NMR experiments were performed at 25 °C on a Bruker Avance 800 MHz (National University of Singapore) spectrometer equipped with a cryogenic probe. A 600  $\mu$ M  $^2\text{H}$ - $^{15}\text{N}$ - $^{13}\text{C}$  labeled sample of the Kelch domain was used to acquire  $^1\text{H}$ - $^{15}\text{N}$ -HSQC, HN(CO)CACB, HNCACB, and  $^{15}\text{N}$ -NOESY-HSQC spectra for sequential assignment of the protein. The data was processed with NMRPipe (Delaglio, Grzesiek et al. 1995) and analyzed using CARA (<http://cara.nmr.ch>). The chemical shifts were deposited in the Biological Magnetic Resonance Bank (accession number 19992).

$^{15}\text{N}$ -labeled samples of the Kelch domain were prepared at a 200  $\mu$ M concentration for the ligand-binding NMR experiments.  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments were performed on a Varian INOVA 600 MHz spectrometer with a cryogenic probe (UWO

Biomolecular NMR facility) at 25 °C. Data was analyzed using NMRPipe (Delaglio, Grzesiek et al. 1995) and NMRView (Johnson 2004).

#### **4.2.5 Circular Dichroism (CD) Spectropolarimetry**

Wild-type and Kelch mutant protein solutions (~2.2 mg/mL) were each prepared in 50 mM sodium phosphate buffer, 100 mM NaCl, 1 mM DTT, pH 7.0. Far-UV CD spectra were obtained in a 0.1 mm cuvette using a Jasco J-810 spectropolarimeter (Easton, MD), at 25 °C for wavelengths ranging from 190-260 nm. The final curves for each sample represent the mean of twenty separate scans. The CD data ( $\theta$ , in millidegrees) were converted to mean residue ellipticity ( $[\theta]$ ) using standard formula and the software supplied by the manufacturer. The programs SELCON3, CDSSTR, and CONTIN provided by the DichroWeb were used to deconvolute the CD spectra (Whitmore and Wallace 2008).

#### **4.2.6 Isothermal Titration Calorimetry**

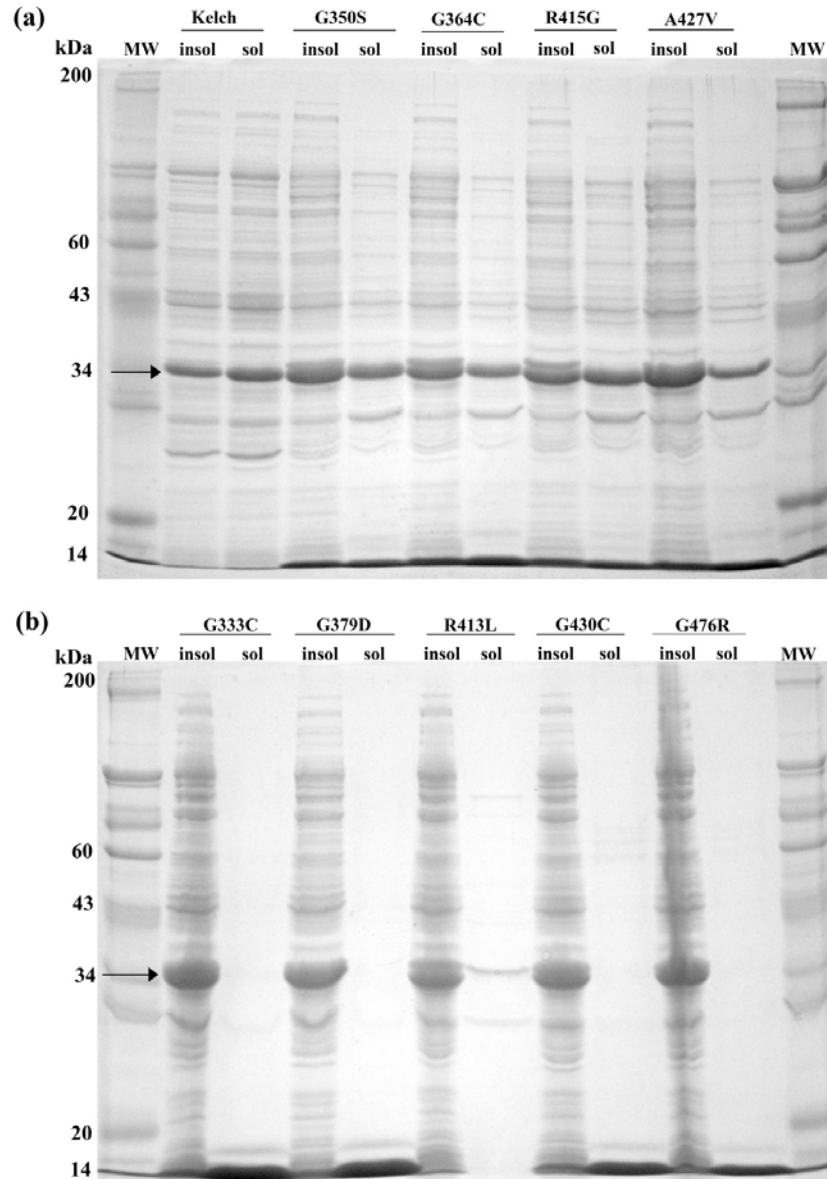
ITC experiments were carried out on a MicroCal™ VP-ITC. All purified protein samples were dialyzed into 50 mM phosphate buffer with 100 mM NaCl and 1 mM DTT at pH 7.0. In typical experiments, wild-type Kelch and mutational variants were prepared to a concentration of ~15  $\mu$ M for the Neh2-ETGE peptide (~150  $\mu$ M) titrations and ~60  $\mu$ M for the ProT $\alpha$  peptide (~800  $\mu$ M) and Neh2-DLG peptide (~1-1.5 mM) titrations. Kelch or its mutational variants were loaded into the 1.42-mL cell, and ligands were loaded into the syringe (300  $\mu$ L). For the full-length Neh2 and Neh2 $\Delta$ ETGE experiments with the wild-type and mutants of Kelch, the Kelch was loaded into the syringe and the specified otherwise) starting with initial injection of 3  $\mu$ L, followed by 59 injections of

5  $\mu$ L, with spacing of 240 seconds to 300 seconds. For the Neh2 $\Delta$ ETGE construct, titrations were performed at 25 °C, 20 °C, 15 °C, 10 °C and 5 °C. The buffer blank was performed under the same conditions and showed negligible heats of binding. The dissociation constant ( $K_d$ ), stoichiometry of binding ( $n$ ), binding enthalpy ( $\Delta H$ ), and entropy ( $\Delta S$ ) were obtained by fitting the data to a single-site binding model with the data analysis software (Origin 7) provided by the manufacturer. Baselines were subtracted from final data using the Origin software. For the full-length Neh2 experiments, the data were fitted either to a two-site binding model (wild-type Kelch, G350S, and A427V) or a one-site-binding model (G364C and R415G). Protein concentrations were confirmed by amino acid analysis (Advanced Protein Technology Centre, The Hospital for Sick Children, Toronto, ON).

## 4.3 Results

### 4.3.1 *Mutations exert differential effects on the solubility of the Kelch domain of Keap1*

The wild-type Kelch domain of human Keap1 was over-expressed in *E. coli* and purified by the procedure described previously (Khan, Cino et al. 2013). Four of the Kelch mutants, G350S, G364C, R415G and A427V, were found to have similar solubility as the wild-type and were purified with comparable yields (Figure 4.2a). The R413L, G333C, G379D, G430C and G476R mutants, on the other hand, were found almost exclusively in the insoluble fractions under various over-expression conditions tested. We selected to purify the G333C, G379D, and the R413L proteins from inclusion bodies using 8 M Urea and subjected to refolding. Severe protein aggregation occurred upon



**Figure 4.2. Coomassie-stained SDS-PAGE depicting the purification of Kelch and its mutational variants.** (a) Insoluble (insol) and soluble (sol) fractions of wild type Kelch, G350S, G364C, R415G and A427V mutants. (b) Insoluble (insol) and soluble (sol) fractions of G333C, G379D, R413L, G430C and G476R mutants of the Kelch domain. The MW lanes are the molecular weight markers with molecular weights labeled on the left. Arrow is drawn identifying the band corresponding to Kelch and its mutational variants.



refolding and only a small amount of soluble protein was obtained for the three mutants. The signals observed in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra (Figures 4.3a &b) of G333C and G379D proteins were broad and poorly dispersed in the proton dimension, suggesting that both mutants were aggregated. While the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of R413L (Figure 4.3c) resembles that of a folded protein; however, we were unable to obtain a stable (even at 4 °C) high purity protein sample.

#### ***4.3.2 Effects of mutations on the target recognition of the Kelch domain***

To assess the effects of mutations on the target recognition of the Kelch domain of Keap1, we employed ITC to determine the binding affinities of the wild-type Kelch domain and the soluble mutants (G350S, G364C, R415G and A427V) to ProT $\alpha$  and the Neh2 domain of Nrf2. Peptides Neh2 proteins were loaded into the cell. Titrations were performed at 25 °C (unless encoding the Neh2-ETGE, Neh2-DLG, and ProT $\alpha$ -ENGE Kelch-binding motifs were used in the studies. Table 4.2 lists the binding parameters of the wild-type and mutants of the Kelch domain to the ProT $\alpha$ -ENGE and Neh2-ETGE peptides. Results of duplicate runs are shown in Table 4.3.

The dissociation constant ( $K_d$ ) of the wild-type Kelch/ProT $\alpha$ -ENGE and Kelch/Neh2-ETGE complexes were determined to be  $2.7 \pm 0.1 \mu\text{M}$  and  $31 \pm 2 \text{ nM}$ , respectively. These values are on par with the values of  $\sim 2.6 \mu\text{M}$  and  $\sim 20 \text{ nM}$  for Kelch complex formation with full-length ProT $\alpha$  and with a 16-mer Neh2-ETGE peptide as reported in the literature (Lo, Li et al. 2006; Khan, Cino et al. 2013). The binding affinities of the mutants to ProT $\alpha$  and Neh2-ETGE peptides varied. While the G350S mutation has only minor effects on the target binding ( $3.0 \pm 0.1 \mu\text{M}$  for the ProT $\alpha$  and to

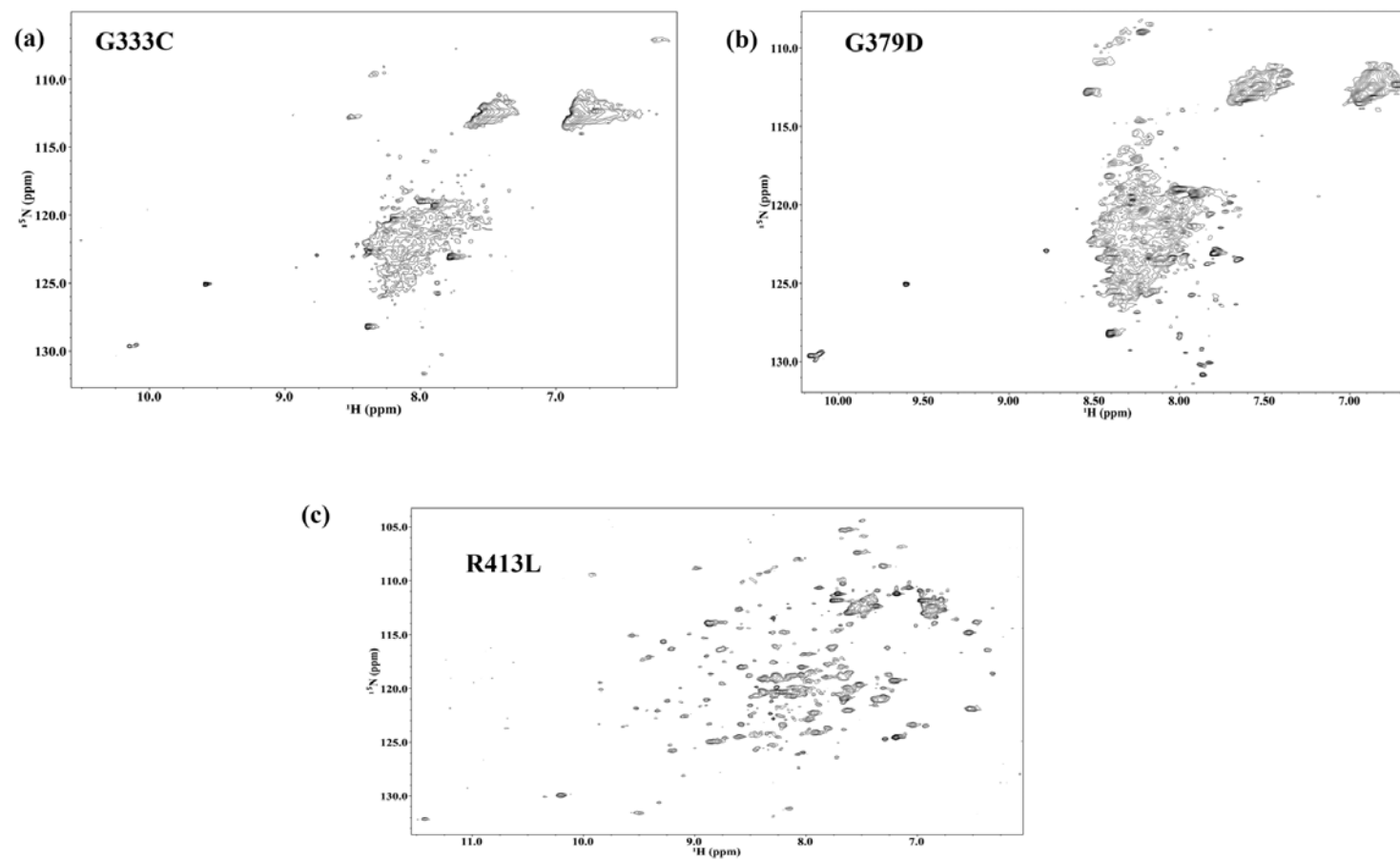


Figure 4.3.  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectra of G333C (a), G379D (b), and R413L (c) mutants of the Kelch domain.

**Table 4.2. Thermodynamic parameters for the binding of the Neh2 and ProT $\alpha$  peptides to the Kelch domain of human Keap1 at 25 °C.**

ENGE peptide of ProT $\alpha$	$n^*$	$K_a^*$ ( $10^5 M^{-1}$ )	$K_d^*$ ( $\mu M$ )	$\Delta H^*$ (kcal/mol)	$T\Delta S^*$ (kcal/mol)	$\Delta G^*$ (kcal/mol)
Wild-type Kelch	$1.07 \pm 0.004$	$3.78 \pm 0.15$	$2.65 \pm 0.11$	$-17.1 \pm 0.2$	<b>-9.50</b>	$-7.60 \pm 0.02$
G350S	$0.92 \pm 0.02$	$3.33 \pm 0.07$	$3.00 \pm 0.06$	$-18.1 \pm 0.6$	<b>-10.58</b>	$-7.52 \pm 0.01$
G364C	$0.98 \pm 0.01$	$0.32 \pm 0.02$	$31.3 \pm 2.0$	$-14.17 \pm 0.2$	<b>-8.03</b>	$-6.14 \pm 0.04$
R415G**	-	-	-	-	-	-
A427V	$1.06 \pm 0.003$	$5.25 \pm 0.1$	$1.90 \pm 0.04$	$-22.4 \pm 0.1$	<b>-14.60</b>	$-7.80 \pm 0.02$
ETGE peptide of Neh2	$n^*$	$K_a^*$ ( $10^7 M^{-1}$ )	$K_d^*$ (nM)	$\Delta H^*$ (kcal/mol)	$T\Delta S^*$ (kcal/mol)	$\Delta G^*$ (kcal/mol)
Wild-type Kelch	$1.03 \pm 0.007$	$3.18 \pm 0.23$	$31.4 \pm 2.3$	$-20.4 \pm 0.09$	<b>-10.17</b>	$-10.23 \pm 0.04$
G350S	$1.01 \pm 0.005$	$2.65 \pm 0.3$	$37.7 \pm 4.3$	$-18.9 \pm 0.17$	<b>-8.78</b>	$-10.12 \pm 0.06$
G364C	$1.07 \pm 0.03$	$0.29 \pm 0.01$	$345 \pm 12$	$-11.53 \pm 0.4$	<b>-2.72</b>	$-8.81 \pm 0.02$
R415G**	-	-	-	-	-	-
A427V	$1.11 \pm 0.002$	$6.44 \pm 0.67$	$15.5 \pm 1.6$	$-19.6 \pm 0.13$	<b>-8.95</b>	$-10.65 \pm 0.06$

\*  $n$  is the stoichiometry,  $K_a$ , and  $K_d$  are the association and dissociation constants, respectively.  $\Delta G = \Delta H - T\Delta S = -RT \ln K_a$ , where  $T$  is the temperature in Kelvin,  $R$  is the gas constant.  $\Delta H$  and  $\Delta S$  are the enthalpy and entropy changes, respectively. The corresponding ITC isotherms are shown in Figure 4.4. Values listed in the table are based on one run, and the data for the duplicate run is provided in Table 4.3. \*\* The binding observed was very weak; therefore, values for  $n$ ,  $K_a$ ,  $K_d$ ,  $\Delta H$ ,  $\Delta S$ , and  $\Delta G$  could not be determined accurately.

**Table 4.3. Thermodynamic parameters for the binding of the Neh2 and ProT  $\alpha$  peptides to the human Kelch domain of Keap1 at 25 °C (duplicate run).**

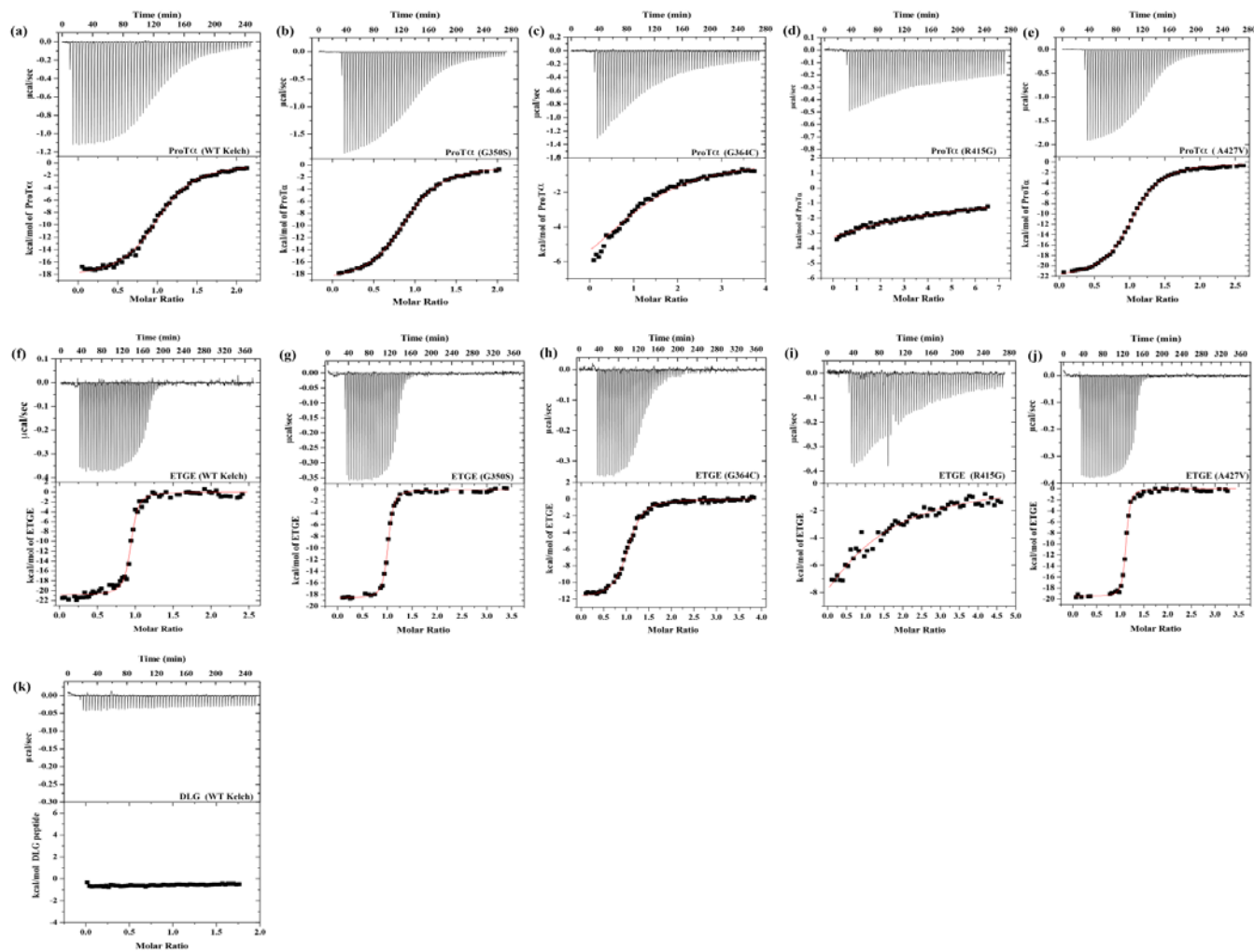
ENGE peptide of ProT $\alpha$	$n^*$	$K_a^*$ ( $10^5 \text{ M}^{-1}$ )	$K_d^*$ ( $\mu\text{M}$ )	$\Delta H^*$ (kcal/mol)	$T\Delta S^*$ (kcal/mol)	$\Delta G^*$ (kcal/mol)
Wild-type Kelch	$1.03 \pm 0.005$	$4.01 \pm 0.2$	$2.49 \pm 0.12$	$-17.7 \pm 0.12$	<b>-10.07</b>	$-7.64 \pm 0.03$
G350S	$0.96 \pm 0.003$	$2.73 \pm 0.05$	$3.66 \pm 0.07$	$-19.2 \pm 0.11$	<b>-11.79</b>	$-7.41 \pm 0.01$
G364C	$1.06 \pm 0.2$	$0.29 \pm 0.01$	$34.5 \pm 1.2$	$-15.9 \pm 0.03$	<b>-9.82</b>	$-6.08 \pm 0.02$
R415G**	-	-	-	-	-	-
A427V	$1.04 \pm 0.004$	$5.24 \pm 0.2$	$1.91 \pm 0.07$	$-22.3 \pm 0.15$	<b>-14.50</b>	$-7.80 \pm 0.02$
ETGE peptide of Neh2	$n^*$	$K_a^*$ ( $10^7 \text{ M}^{-1}$ )	$K_d^*$ (nM)	$\Delta H^*$ (kcal/mol)	$T\Delta S^*$ (kcal/mol)	$\Delta G^*$ (kcal/mol)
Wild-type Kelch	$1.01 \pm 0.002$	$3.43 \pm 0.5$	$29.2 \pm 4.4$	$-21.0 \pm 0.18$	<b>-10.73</b>	$-10.27 \pm 0.09$
G350S	$1.0 \pm 0.003$	$2.62 \pm 0.3$	$38.2 \pm 4.3$	$-19.2 \pm 0.11$	<b>-9.09</b>	$-10.11 \pm 0.06$
G364C	$1.08 \pm 0.03$	$0.4 \pm 0.02$	$250 \pm 13$	$-10.63 \pm 0.04$	<b>-1.63</b>	$-9.0 \pm 0.03$
R415G**	-	-	-	-	-	-
A427V	$1.05 \pm 0.003$	$4.33 \pm 0.5$	$23.1 \pm 2.7$	$-19.7 \pm 0.19$	<b>-9.29</b>	$-10.41 \pm 0.07$

\* $n$  is the stoichiometry,  $K_a$ , and  $K_d$  are the association and dissociation constants, respectively.  $\Delta G = \Delta H - T\Delta S = -RT\ln K_a$ , where  $T$  is the temperature in Kelvin,  $R$  is the gas constant.  $\Delta H$  and  $\Delta S$  are the enthalpy and entropy changes, respectively. Values listed in the table are based on one run, and the data for the duplicate run is provided in Table 4.2. \*\* The binding observed was very weak, therefore, accurate values for  $n$ ,  $K_a$ ,  $K_d$ ,  $\Delta H$ ,  $\Delta S$ , and  $\Delta G$  could not be calculated.

$38 \pm 4$  nM for the Neh2-ETGE peptides), the G364C mutation lowered the binding affinity for both Neh2-ETGE and ProT $\alpha$  by >10 fold compared to the wild-type complex ( $31 \pm 2$   $\mu$ M for the ProT $\alpha$  and to  $345 \pm 12$  nM for the Neh2-ETGE peptides). Our ITC data show that the interactions of G364C with targets are enthalpically less favorable compared to the wild-type interactions. Replacement of the Gly residue with a Cys may restrict the conformational freedom of the neighboring region and/or possibly disrupt the wild-type hydrogen bond network.

The binding affinities of Kelch to both the ProT $\alpha$  and Neh2-ETGE peptides were severely impaired by the R415G mutation to an extent that the affinities could not be accurately determined by ITC (Figures 4.4d & i). Intriguingly, in contrast to G350S, G364C and R415G, the interactions with ProT $\alpha$  and Neh2-ETGE peptides were modestly strengthened with the A427V mutation. The  $K_d$  values are  $1.90 \pm 0.04$   $\mu$ M and  $16 \pm 2$  nM, respectively. Interestingly, the A427V-ProT $\alpha$  complex formation was enthalpically more favorable, while the entropy change is more negative in comparison with the wild-type.

We have also attempted to measure the binding affinities of a 20-residue Neh2-DLG peptide (encoding the sequence of I20-F39 of Nrf2) to the wild-type Kelch (Figure 4.4 k) and the four mutants; however, no binding was detected by ITC at 25°C in any case. Our results suggest that either the interaction between the Kelch domain and the Neh2-DLG peptide is very weak or the  $\Delta H$  of the binding is close to zero. This observation is in good agreement with results obtained very recently by Fukutomi *et al.* (Fukutomi, Takagi et al. 2014) showing that the binding of the Kelch domain with a 16-



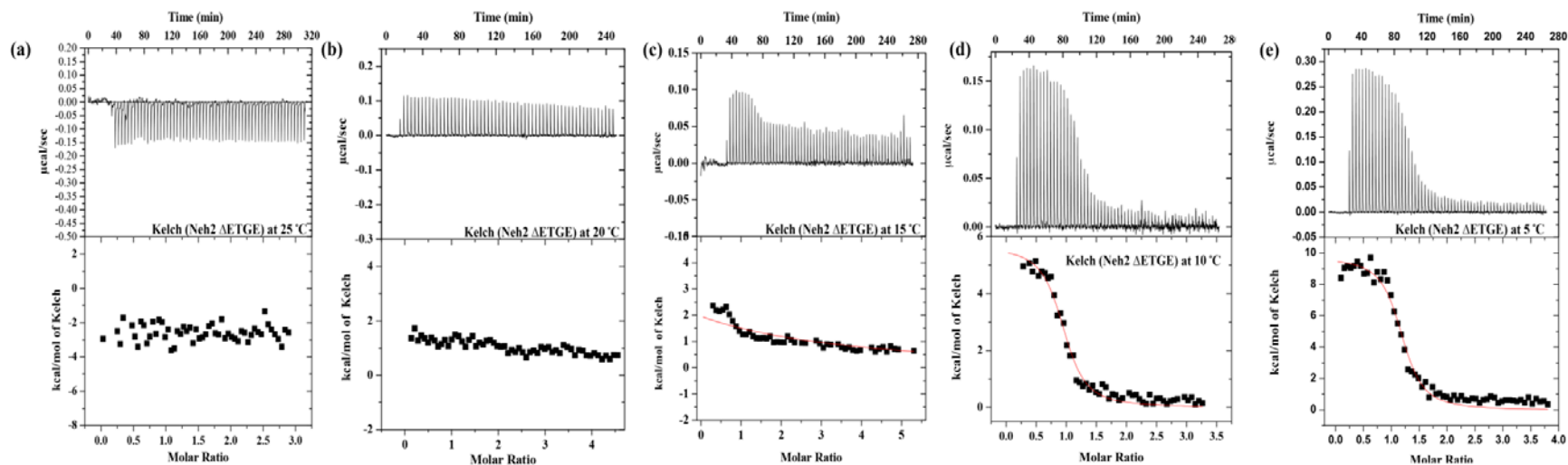
**Figure 4.4.** ITC profiles of titrating ProT $\alpha$ -ENGE peptide to wild-type Kelch. (a), G350S (b), G364C (c), R415G (d), and A427V (e), respectively; (f-j) show the ITC profiles of titrating Neh2-ETGE peptide to wild-type Kelch (f), G350S (g), G364C (h), R415G (i), and A427V (j), respectively; (k) ITC profile of titrating Neh2-DLG peptide to wild-type Kelch.

mer peptide containing the previously defined DLG motif of Nrf2 (I22-V36) (Tong, Padmanabhan et al. 2007) was too weak to be detected by the differential scanning calorimetry.

The Neh2 $\Delta$ ETGE (deletion of <sup>79</sup>ETGE<sup>82</sup> from full-length Neh2) construct was then used to investigate the Kelch-binding of the DLG motif in the context of full-length Neh2. Figures 4.5a-e show the isotherms for the binding of the Neh2 $\Delta$ ETGE protein to the Kelch domain at 25 °C, 20 °C, 15 °C, 10 °C and 5 °C, respectively. From the ITC isotherms in Figure 4.5 it is apparent that the complex formation is endothermic at low temperatures and as the temperature is increased, the enthalpy change becomes smaller. Table 4.4 lists the thermodynamic parameters for the Neh2 $\Delta$ ETGE-Kelch complex at 5 °C and 10 °C. With the buffer conditions used in this study, the  $K_a$  of the human Neh2 $\Delta$ ETGE/Kelch complex was determined to be  $\sim 2.8 \times 10^6 \text{ M}^{-1}$  at 10 °C. The value is similar to what was determined for the mouse Neh2 $\Delta$ ETGE/Kelch complex formation at 25 °C ( $2.0 \times 10^6 \text{ M}^{-1}$ ) (Tong, Katoh et al. 2006). We then estimated the  $\Delta C_p$  value based on the  $\Delta H$  values at the 5 °C and 10 °C. Based on the calculated  $\Delta C_p$  value of  $-0.47 \text{ kcal mol}^{-1} \text{ K}^{-1}$ , the  $\Delta H$  at 25 °C was estimated to be  $\sim -1.2 \text{ kcal/mol}$ . This small value of enthalpy change explains why binding between the Neh2 $\Delta$ ETGE and the Kelch domain could not be observed at 25 °C by ITC.

#### ***4.3.3 Somatic mutations in the Kelch domain affect its interaction with the full-length Neh2 domain***

To examine how the somatic mutations in the Kelch domain affect its interactions



**Figure 4.5.** ITC profiles of titrating wild-type Kelch to Neh2 $\Delta$ ETGE at 25 °C (a), 20 °C (b), 15 °C (c), 10 °C (d), and 5 °C, respectively.



**Table 4.4. Thermodynamic parameters for the binding of the Kelch domain of Keap1 to Neh2ΔETGE at different temperatures.**

Temp (° C)	$n^*$	$K_a^*$ ( $10^6 \text{ M}^{-1}$ )	$K_d^*$ (nM)	$\Delta H^*$ (kcal/mol)	$T\Delta S^*$ (kcal/mol)	$\Delta G^*$ (kcal/mol)
<b>Titration set 1</b>						
5	$0.90 \pm 0.03$	$3.03 \pm 0.14$	$330 \pm 15$	$8.13 \pm 0.11$	16.37	$-8.24 \pm 0.03$
10	$1.02 \pm 0.17$	$2.85 \pm 0.13$	$351 \pm 16$	$5.80 \pm 0.06$	14.16	$-8.36 \pm 0.03$
<b>Titration set 2</b>						
5	$1.10 \pm 0.02$	$3.13 \pm 0.18$	$319 \pm 18$	$8.70 \pm 0.15$	16.96	$-8.26 \pm 0.03$
10	$0.97 \pm 0.02$	$2.66 \pm 0.11$	$376 \pm 16$	$5.64 \pm 0.04$	13.96	$-8.32 \pm 0.02$

\* $n$  is the stoichiometry,  $K_a$ , and  $K_d$  are the association and dissociation constants, respectively.  $\Delta G = \Delta H - T\Delta S = -RT\ln K_a$ , where  $T$  is the temperature in Kelvin,  $R$  is the gas constant.  $\Delta H$  and  $\Delta S$  are the enthalpy and entropy changes, respectively.

with the DLG and ETGE motifs in the context of full-length Neh2, ITC experiments were carried out at 5 °C to determine the binding parameters of the wild-type Kelch and four mutational variants with the Neh2 domain (residues 1-98) of Nrf2. Kelch proteins were titrated to the full-length Neh2 and the isotherms were fit to either a two-site or one-site binding model. The results are listed in Table 4.5 (Figures 4.6 a-e) shows the ITC isotherms and the results of the duplicate runs are listed in Table 4.6). Similar to what was observed for the binding of the Kelch domain to the Neh2-ETGE peptide, both R415G and G364C mutations dramatically weakened Kelch's interaction with the ETGE motif of the full-length Neh2. Compared to the wild-type Kelch, which binds to the ETGE motif with a  $K_d$  of  $1.26 \pm 0.03$  nM, G364C and R415G have binding affinities that are 37 and 870 fold lower, respectively. On the other hand, the G350S and A427V mutations show no significant impact on the ETGE binding. It is noteworthy that the binding affinities of the Neh2-ETGE motif to the wild-type and mutational variants of Kelch are tighter when present in the full-length Neh2 protein as opposed to a peptide. For instance, the Kelch R145G provided a measurable affinity of  $\sim 1.1$   $\mu$ M when titrated to full-length Neh2 but could not be accurately determined for the ETGE peptide (Figure 4.6d).

As observed for the mouse Kelch-DLG complex (Tong, Katoh et al. 2006), the DLG motif in the human Neh2 also binds significantly weaker to the Kelch domain compared to the ETGE motif ( $K_d$  of 179 nM vs 1.26 nM). For the mutants, the binding

**Table 4.5. Thermodynamic parameters for the binding of the wild-type and mutational variants of Kelch to the full length Neh2 at 5° C.**

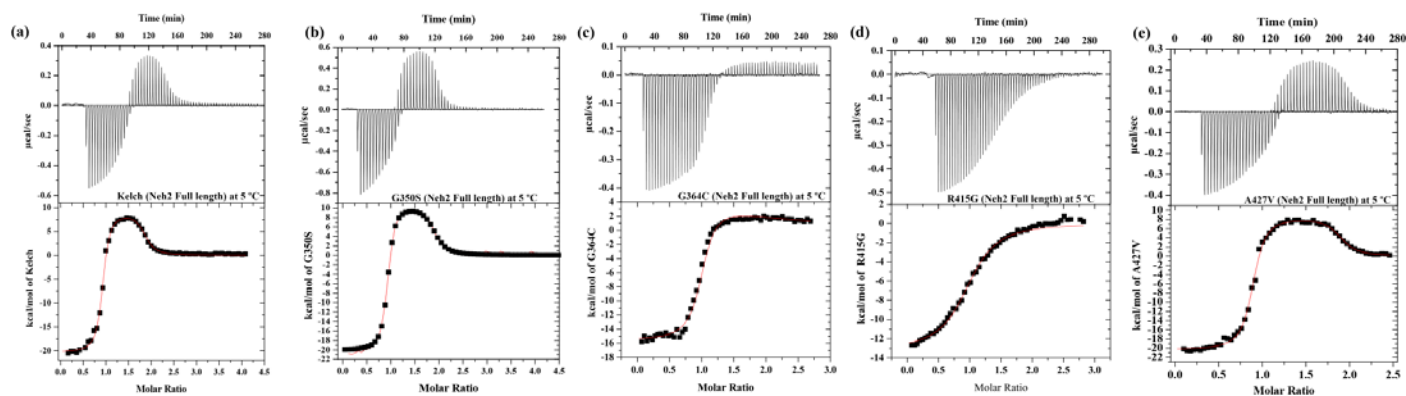
Protein		<i>n</i>	$K_a$ ( $10^7 \text{ M}^{-1}$ )	$K_d$ (nM)	$\Delta H$ (kcal/mol)	$T\Delta S$ (kcal/mol)	$\Delta G$ (kcal/mol)
Kelch	ETGE	$0.90 \pm 0.00(3)$	$79.2 \pm 2.07$	$1.26 \pm 0.03$	$-20.34 \pm 0.12$	-9.02	$-11.32 \pm 0.01$
	DLG	$0.93 \pm 0.02$	$0.56 \pm 0.012$	$179 \pm 4$	$9.04 \pm 0.03$	17.62	$-8.58 \pm 0.01$
G350S	ETGE	$0.91 \pm 0.01$	$60.9 \pm 2.99$	$1.64 \pm 0.08$	$-20.13 \pm 0.07$	-8.96	$-11.17 \pm 0.03$
	DLG	$0.93 \pm 0.01$	$0.35 \pm 0.014$	$286 \pm 11$	$10.61 \pm 0.12$	18.93	$-8.32 \pm 0.02$
G364C	ETGE	$0.92 \pm 0.01$	$2.15 \pm 0.18$	$46.5 \pm 3.9$	$-15.09 \pm 0.06$	-5.77	$-9.32 \pm 0.04$
	DLG	-	-	-	-	-	-
R415G	ETGE	$1.03 \pm 0.01$	$0.091 \pm 0.008$	$1099 \pm 97$	$-13.35 \pm 0.12$	-5.77	$-7.58 \pm 0.05$
	DLG	-	-	-	-	-	-
A427V	ETGE	$0.91 \pm 0.01$	$79.1 \pm 4.9$	$1.26 \pm 0.08$	$-20.17 \pm 0.17$	-8.85	$-11.32 \pm 0.03$
	DLG	$1.00 \pm 0.03$	$0.56 \pm 0.021$	$179 \pm 7$	$9.44 \pm 0.10$	18.02	$-8.58 \pm 0.02$

The isotherms of wild-type Kelch, G350S and A427V mutants were fit to a two-site-binding model, assuming that the binding affinities of the ETGE motif are higher than that of the DLG motif. The isotherms of G364C and R415G mutants were fit to a one-site-binding model. *n* is the stoichiometry,  $K_a$ , and  $K_d$  are the association and dissociation constants, respectively.  $\Delta G = \Delta H - T\Delta S = -RT\ln K_a$ , where T is the temperature in Kelvin, R is the gas constant,  $\Delta H$  and  $\Delta S$  are the enthalpy and entropy changes, respectively. Values listed in the table are based on one run, the data for the duplicate run is provided in Table 4.6

**Table 4.6. Thermodynamic parameters for the binding of the Kelch and Kelch mutational variants to full length Neh2 at 5° C (duplicate runs).**

Protein		$n^*$	$K_a^*$ ( $10^7 \text{ M}^{-1}$ )	$K_d^*$ (nM)	$\Delta H^*$ (kcal/mol)	$T\Delta S^*$ (kcal/mol)	$\Delta G^*$ (kcal/mol)
Kelch	ETGE	$1.06 \pm 0.04$	$73.3 \pm 1.84$	$1.36 \pm 0.03$	$-19.02 \pm 0.08$	-7.74	$-11.28 \pm 0.01$
	DLG	$1.12 \pm 0.02$	$0.46 \pm 0.010$	$217 \pm 5$	$8.46 \pm 0.10$	16.93	$-8.47 \pm 0.01$
G350S	ETGE	$1.07 \pm 0.01$	$51.8 \pm 1.79$	$1.93 \pm 0.07$	$-19.83 \pm 0.13$	-8.74	$-11.08 \pm 0.02$
	DLG	$1.1 \pm 0.03$	$0.36 \pm 0.011$	$278 \pm 8$	$9.52 \pm 0.09$	17.86	$-8.34 \pm 0.02$
G364C	ETGE	$0.92 \pm 0.01$	$2.34 \pm 0.12$	$42.7 \pm 2.2$	$-15.0 \pm 0.12$	-5.63	$-9.37 \pm 0.03$
	DLG	-	-	-	-	-	-
R415G	ETGE	$1.0 \pm 0.01$	$0.089 \pm 0.008$	$1124 \pm 101$	$-13.40 \pm 0.05$	-5.83	$-7.57 \pm 0.05$
	DLG	-	-	-	-	-	-
A427V	ETGE	$0.91 \pm 0.00(3)$	$68.4 \pm 1.95$	$1.46 \pm 0.04$	$-20.69 \pm 0.14$	-9.45	$-11.24 \pm 0.02$
	DLG	$1.00 \pm 0.01$	$0.53 \pm 0.014$	$189 \pm 5$	$9.52 \pm 0.02$	18.07	$-8.55 \pm 0.01$

The isotherms of wild-type Kelch, G350S and A427V mutants were fit to a two-site-binding model, assuming that the binding affinities of the ETGE motif are higher than that of the DLG motif. The isotherms of G364C and R415G mutants were fit to a one-site-binding model.  $n$  is the stoichiometry,  $K_a$ , and  $K_d$  are the association and dissociation constants, respectively.  $\Delta G = \Delta H - T\Delta S = -RT \ln K_a$ , where  $T$  is the temperature in Kelvin,  $R$  is the gas constant,  $\Delta H$  and  $\Delta S$  are the enthalpy and entropy changes, respectively.



**Figure 4.6. ITC profiles of titrating wild-type Kelch and mutational variants to full length Neh2 5 °C**

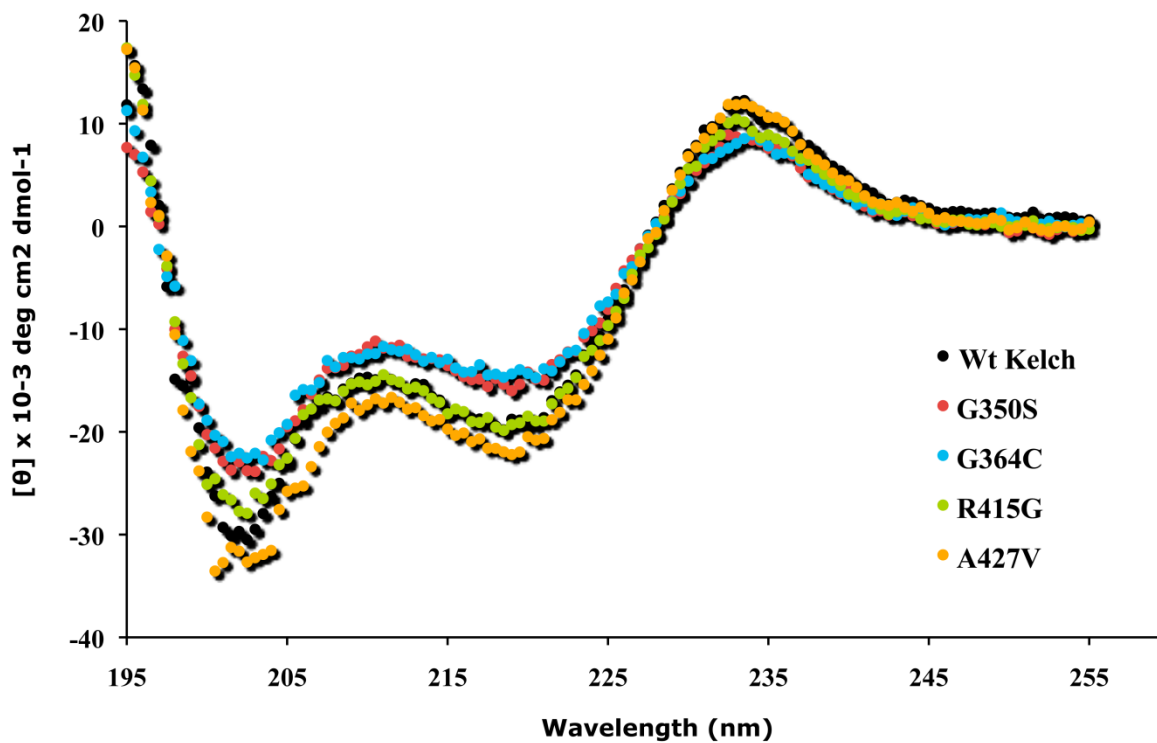
(a) wild-type Kelch (b), G350S (c), G464C (d), R415G (e), and A427V respectively.

affinity of the G350S to the DLG motif decreased by less than two-fold while the A427V mutant has the same affinity ( $K_d$  of  $179 \pm 4$  nM) as compared to the wild-type. The binding parameters indicate that in the presence of either of these two somatic mutations, the Kelch domain should still be able to interact effectively with both the ETGE and DLG motifs of Nrf2. On the other hand, the interactions of the DLG motif with G364C and R415G could not be detected by ITC (Figures 4.6 c & d).

#### ***4.3.4 Probing structural changes of Kelch and its mutational variants by CD and NMR***

CD and NMR were used to identify the effects of missense mutations on the overall folding and local conformation of the Kelch domain. The CD results revealed that the  $\beta$ -sheet secondary structure content, observed in the wild-type Kelch domain, is preserved in G364C, G350S, R415G and A427V (Figure 4.7). The percentages of  $\beta$ -sheet and disordered regions estimated from the deconvolution CD spectra are almost identical to the wild-type protein (Table 4.7), strongly suggesting that these four mutations do not have a significant effect on the overall fold of the Kelch domain.

We then used NMR spectroscopy to investigate the effects of somatic mutations on the local structural changes and ligand-binding of the Kelch domain. To be able to monitor the site-specific conformational changes upon mutation and target binding, backbone chemical shift assignment of the 34-kDa human Kelch domain was performed.



**Figure 4.7. CD spectra of Kelch and mutational variants.** The spectra of the mutants appears similar to the wild type Kelch domain suggesting no significant changes occur in the secondary structure content of the mutational variants of the Kelch domain listed.

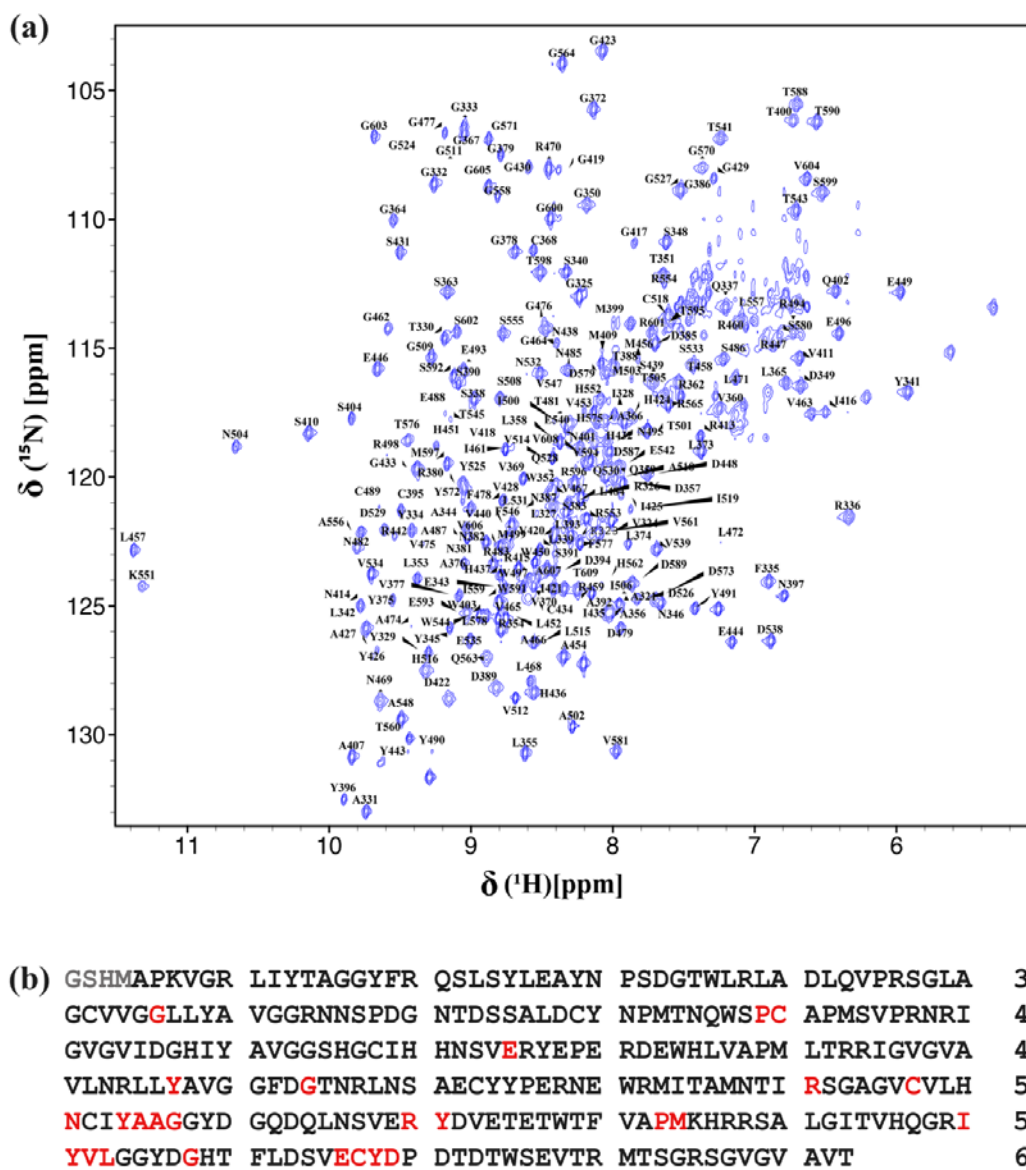
**Table 4.7. Deconvoluted CD data for the WT Kelch and mutational variants**

<b>Protein</b>	<b>Helix</b>	<b><math>\beta</math>-sheet</b>	<b>Turn</b>	<b>Unordered</b>
<b>WT-Kelch</b>	<b>0.063</b>	<b>0.432</b>	<b>0.120</b>	<b>0.385</b>
<b>G350S</b>	<b>0.063</b>	<b>0.433</b>	<b>0.120</b>	<b>0.384</b>
<b>G364C</b>	<b>0.063</b>	<b>0.432</b>	<b>0.118</b>	<b>0.386</b>
<b>R415G</b>	<b>0.062</b>	<b>0.415</b>	<b>0.120</b>	<b>0.394</b>
<b>A427V</b>	<b>0.061</b>	<b>0.436</b>	<b>0.119</b>	<b>0.384</b>

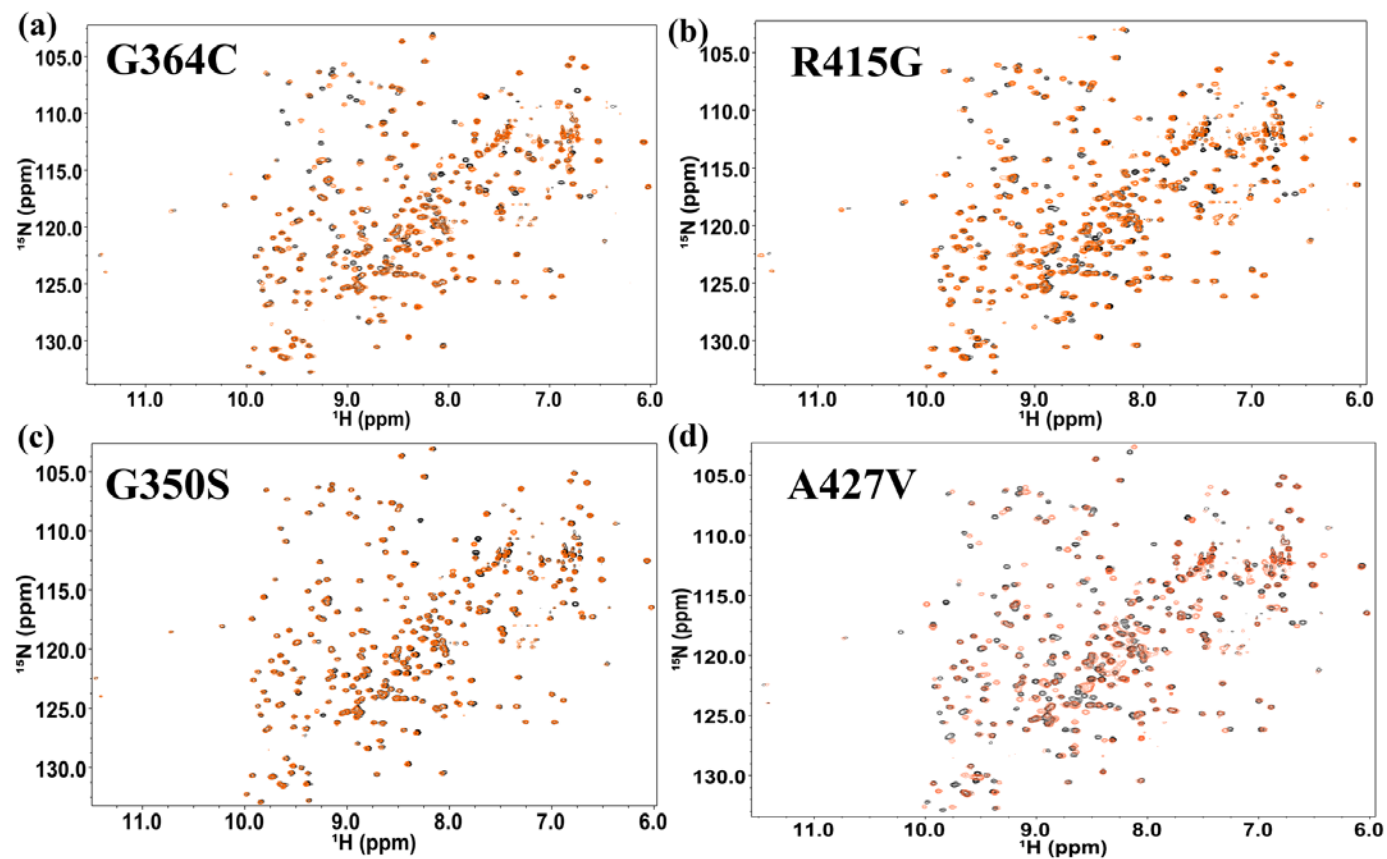


$^1\text{HN}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}\alpha$ ,  $^{13}\text{C}\beta$  resonance signals were assigned using the HNCACB, HN(CO)CACB, and  $^{15}\text{N}$  NOESY- HSQC experiments. The assigned  $^1\text{H}$ - $^{15}\text{N}$  TROSY- HSQC spectrum is showed in Figure 4.8. Although the Kelch amino acid sequence contains regions of degeneracy (Figure 4.8), we were able to assign 92 % of the  $^1\text{HN}$  and  $^{15}\text{N}$  resonances of the non-proline residues, 91.6 % of  $^{13}\text{C}\alpha$  and 90.7 % of  $^{13}\text{C}\beta$  of all residues. The chemical shift assignments of the human Kelch domain are comparable to the assignments of the Kelch domain of mouse Keap1 that we have completed previously (Cino, Killoran et al. 2013).

To probe the conformational changes of Kelch upon mutations,  $^1\text{H}$ - $^{15}\text{N}$  TROSY- HSQCs of G364C, R415G, G350S, and A427V were collected and compared to the wild-type Kelch spectrum (Figures 4.9). The dispersion of the amide peaks for all four mutants is similar to that of the wild-type Kelch. Further analysis of the chemical shift perturbations (Figure 4.10) revealed that A427V displayed the largest amplitude of peak shifts followed by G364C, R415G and G350S. It is noteworthy that large chemical shift changes are observed for some residues that are located far from the mutation sites in the protein sequence for A427V, G364C and R415G mutants. For instance, A427 is located in the middle of the second  $\beta$ -strand of Blade III. Mutating of this residue to a valine caused peak shifts of a significant number of residues clustered in Blade II (residues V360-M409) and Blade III (S410-L457), in particular, S363, G379, E446 and A466 residues (Figure 4.10c). The S363 and G379 residues also undergo a substantial change in the G364C spectrum (Figure 4.10b).

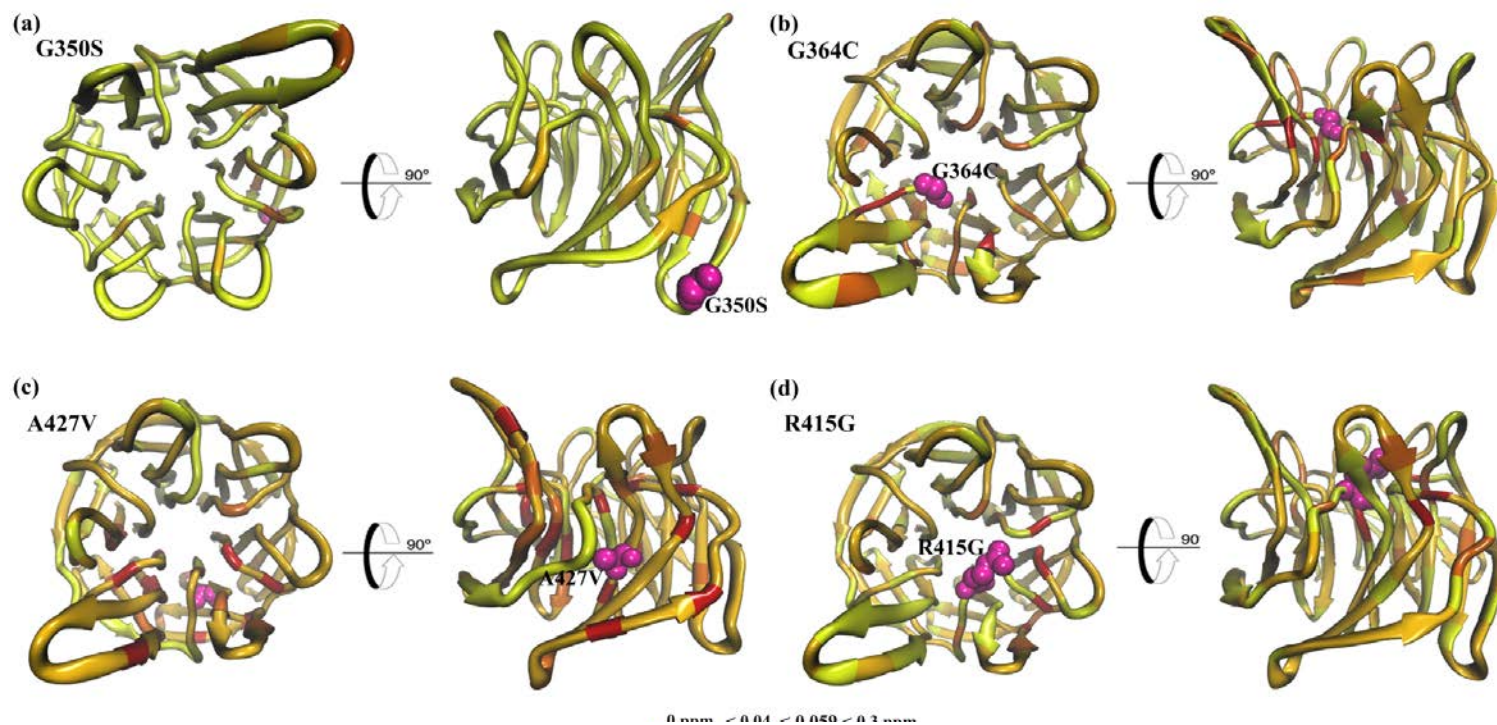


**Figure 4.8. Backbone resonance assignments of the Kelch domain of human Keap1.** (a)  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of  $^2\text{H}/^{15}\text{N}/^{13}\text{C}$  labeled Kelch domain of Keap1. (b) Amino acid sequence of the Kelch domain, unassigned residues are colored in red. Amino acids appearing in grey are residual residues from the affinity tag



**Figure 4.9.** Overlay of the  $^1\text{H}$ - $^{15}\text{N}$  TROSY-HSQC NMR spectra of wild-type Kelch (black) and mutants

(a) G364C, (b) R415G (c) G350S, and (d) A427V.



**Figure 4.10.**  $^1\text{H}_\text{N}$  and  $^{15}\text{N}$  chemical shift changes of the Kelch mutants. Residues with traceable assigned resonances are colored based on composite  $^1\text{H}_\text{N}$  and  $^{15}\text{N}$  chemical shift changes ( $\Delta\delta = [(\Delta\delta_{\text{HN}})^2 + (\Delta\delta_{\text{N}}/5)^2]^{1/2}$  in ppm). (a) G350S, (b) G364C, (c) A427V and (d) R415G.

#### 4.3.5 NMR analysis of the interactions of Kelch with Neh2 and ProT $\alpha$ peptides

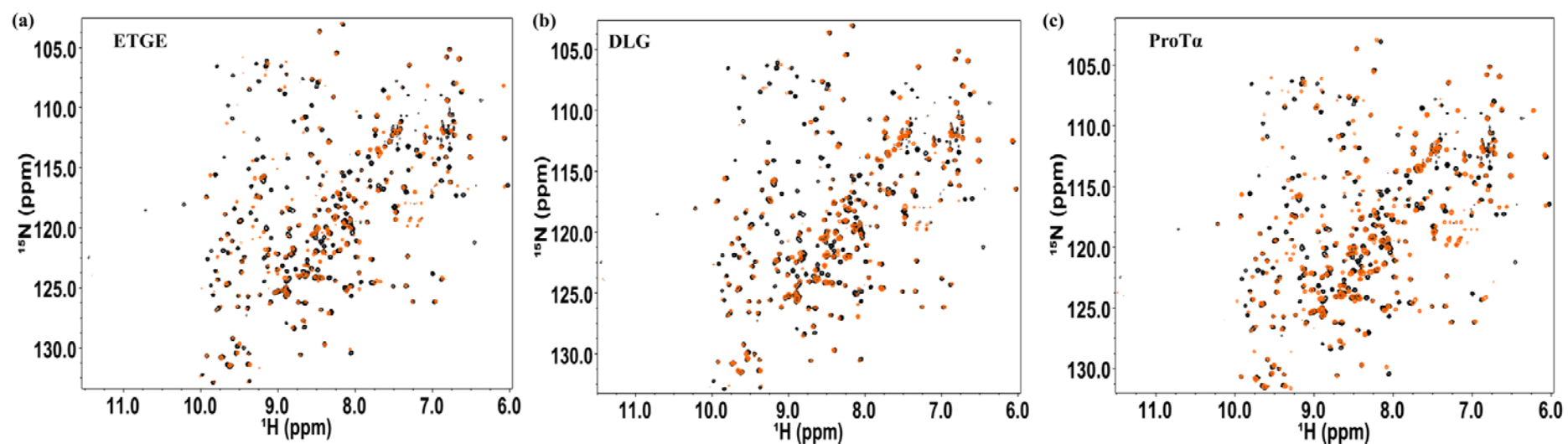
Figures 4.11a-c show the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of the wild-type Kelch domain in the presence of 5 molar equivalents of unlabeled Neh2-ETGE, Neh2-DLG and ProT $\alpha$ -ENGE peptides, respectively. Quantitative analysis of the Kelch chemical shift perturbations revealed large changes in S363, G417, G462, S508, G527, and T576 in the ETGE and ProT $\alpha$  bound state (Figures 4.11 a & c). Some of these residues are directly involved in complex formation with targets, whereas the peak shift of others may be due to the changes in their local chemical environment (Lo, Li et al. 2006; Padmanabhan, Nakamura et al. 2008). For example, S363 was found to form a hydrogen bond with the E46 of ProT $\alpha$  in the complex (Padmanabhan, Nakamura et al. 2008). Meanwhile, the G462 residue does not interact with ProT $\alpha$  directly, rather it forms an intramolecular hydrogen bond with R415 which is involved in a salt bridge with E43 of ProT $\alpha$  (Padmanabhan, Nakamura et al. 2008).

The amplitudes and pattern of chemical shift changes of G350S upon binding to the Neh2-ETGE and ProT $\alpha$ -ENGE peptides were comparable to those of the wild-type (Figure 4.12c, Figure 4.13c, Figures 4.15a&c). G364C and R415G displayed smaller amplitudes of chemical shift changes upon binding to the Neh2-ETGE and ProT $\alpha$ -ENGE peptides (Figures 4.12 a&b, Figures 4.13 a&b, Figures 4.15b&d). Despite the difference in the amplitudes of changes, the overall chemical shift perturbation pattern is similar to the wild-type, suggesting that in the presence of either one of these two mutations, Kelch can still bind to the Neh2-ETGE and ProT $\alpha$ -ENGE peptides in similar modes. This is in

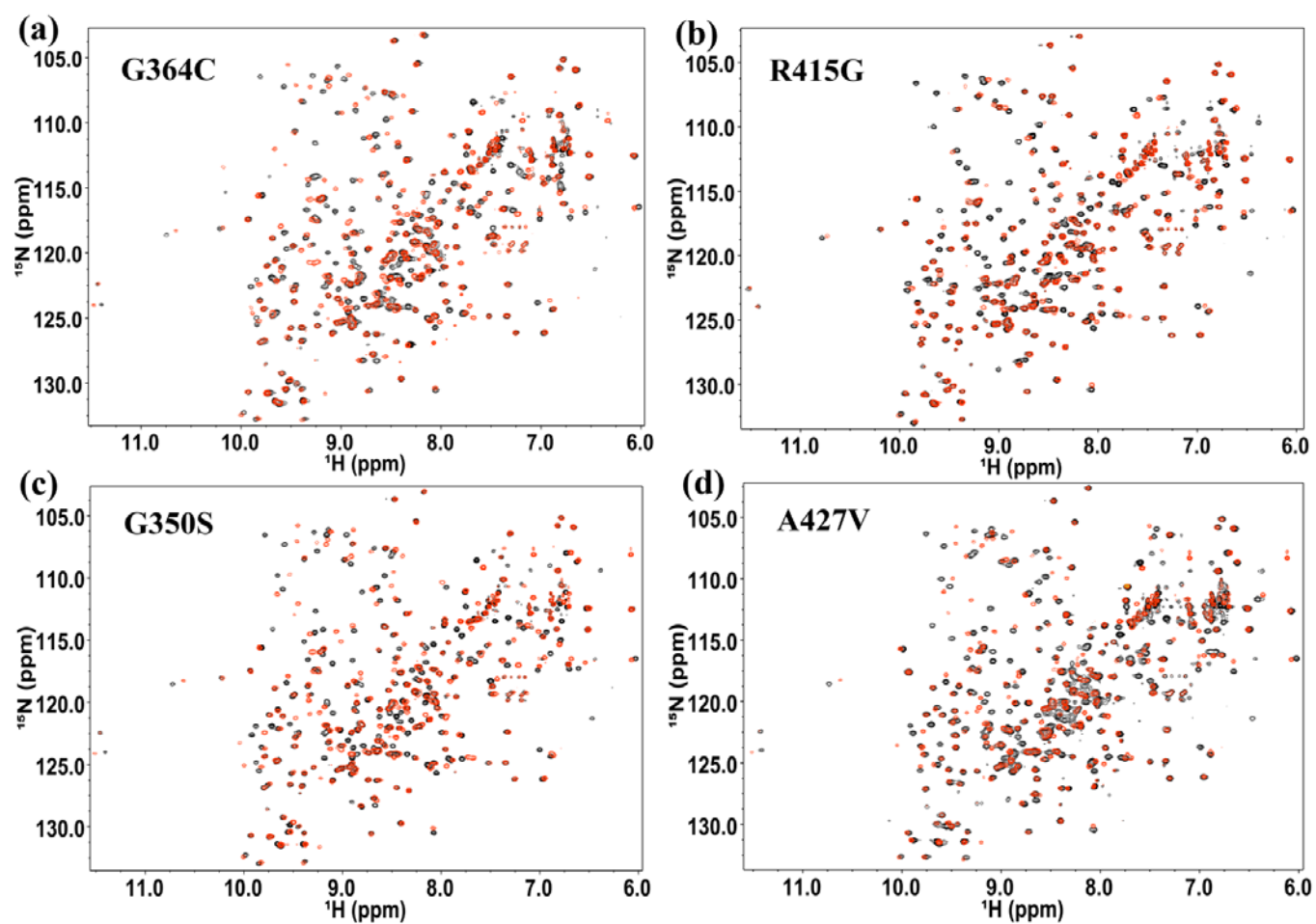
agreement with our ITC data showing that both R415G and G364C were able to interact with the ETGE motif in the full-length Neh2 but with lower affinities.

Titration of the Neh2-DLG peptide to the wild-type Kelch domain (Figures 4.11b & 4.15e) results in large resonance shifts of R354, D385, H436 and A607. The H436 residue was found to interact with the conserved G430 residue. The chemical environment of the G430 is influenced by the R415 residue, which takes part in hydrogen bonding with the targets (Tong, Padmanabhan et al. 2007). Similarly, the chemical environment of other Kelch residues undergoing large peak shifts are affected by amino acids involved in hydrogen bonding with targets. Figures 4.14a-d show the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of G364C, R415G, G350S, and A427V in the absence and presence of five molar equivalents of the Neh2-DLG peptides. For the G350S and A427V mutants, while many resonance signals broaden out in the ligand-bound state, both mutants display similar chemical shift attenuation patterns as the wild-type Kelch. Our NMR data is in accordance with the ITC data, indicating that both A427V and G350S mutants can bind to the Neh2-DLG motif.

On the other hand, the relatively minor chemical shift changes observed for the G364C and R415G upon addition of DLG peptides indicate that the complex formation is disrupted by the mutations (Figure 4.15 and Figures 4.14a & b). In fact, for the G364C, we observe no significant peak broadening and only minor resonance perturbations across the whole spectrum. The minor shifts observed in the bound state NMR spectra of G364C-DLG complex and R415G-DLG complex together with our ITC data suggest that the presence of either one of these two somatic mutations in the Kelch domain can depress the ability of Keap1 to interact with the DLG motif of Neh2.

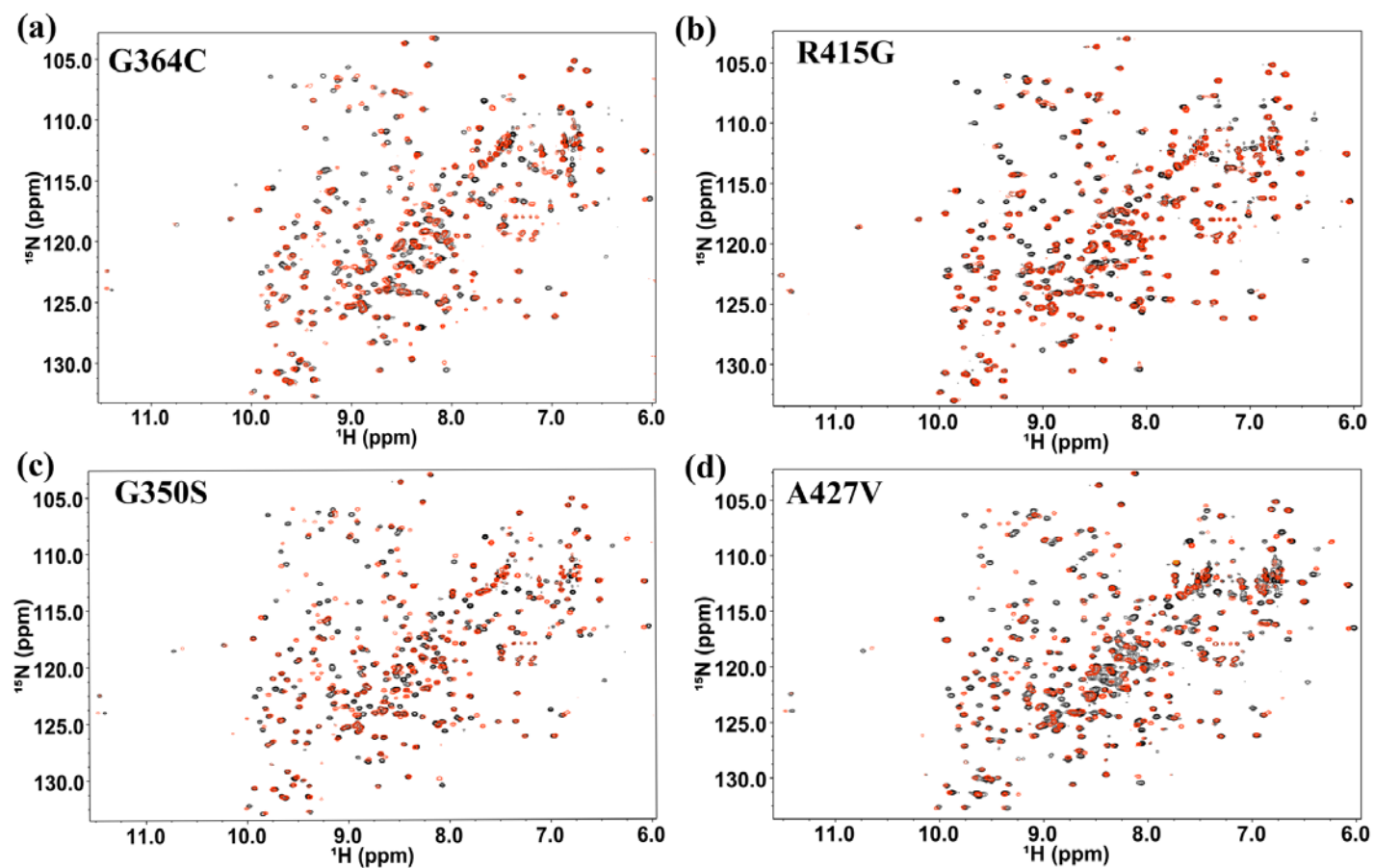


**Figure 4.11.**  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectra of the wild-type Kelch domain in the absence (black) and presence of five molar equivalents of (a) Neh2-ETGE peptide, (b) Neh2-DLG peptide, and (c) ProTu peptide.

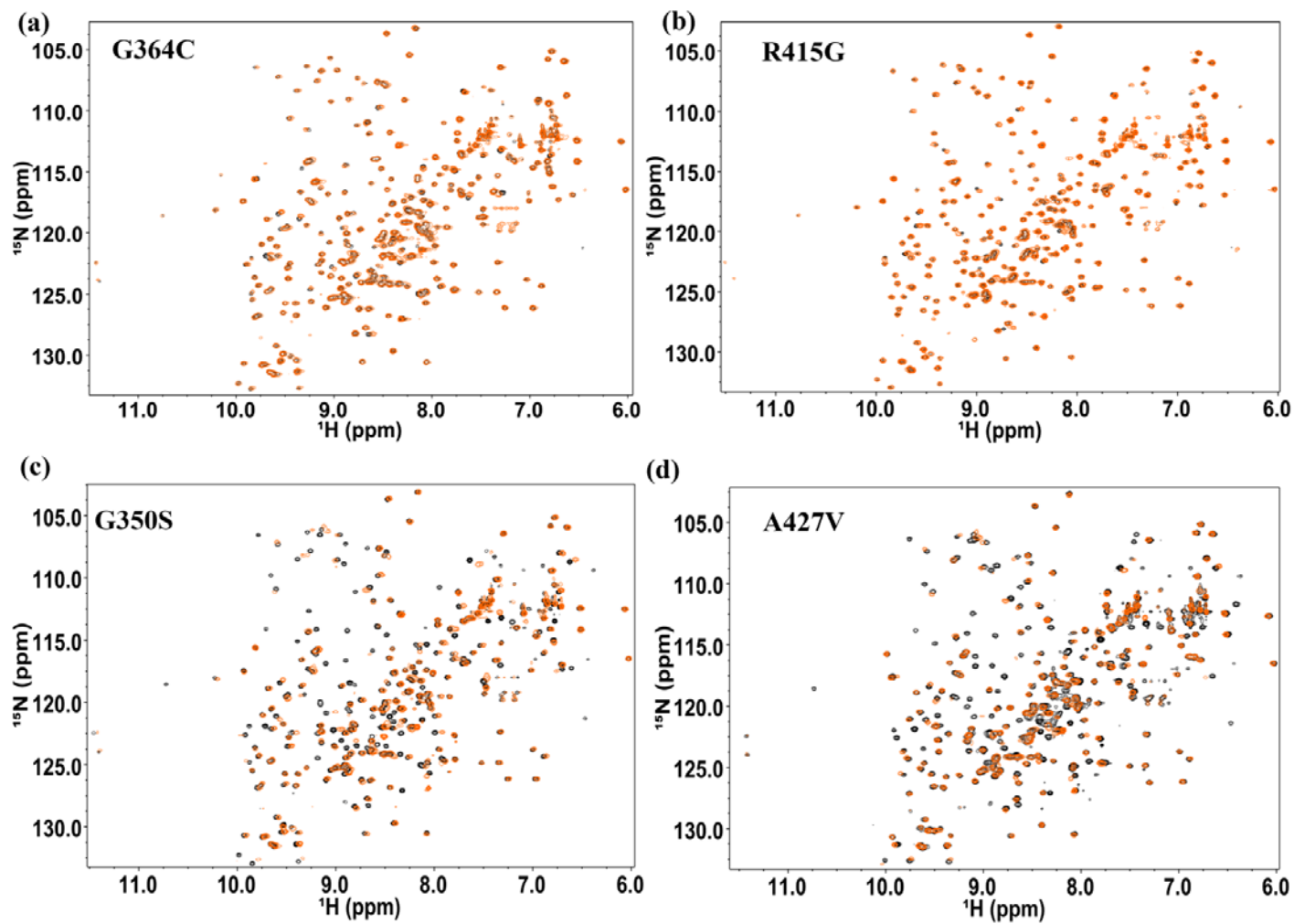


**Figure 4.12.** Overlay of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectra of the (a) G364C, (b) R415G, (c) G350S, and (d) A427V Kelch mutations in the absence (black) and presence (orange) of a two molar equivalent of the Neh2-ETGE peptide.

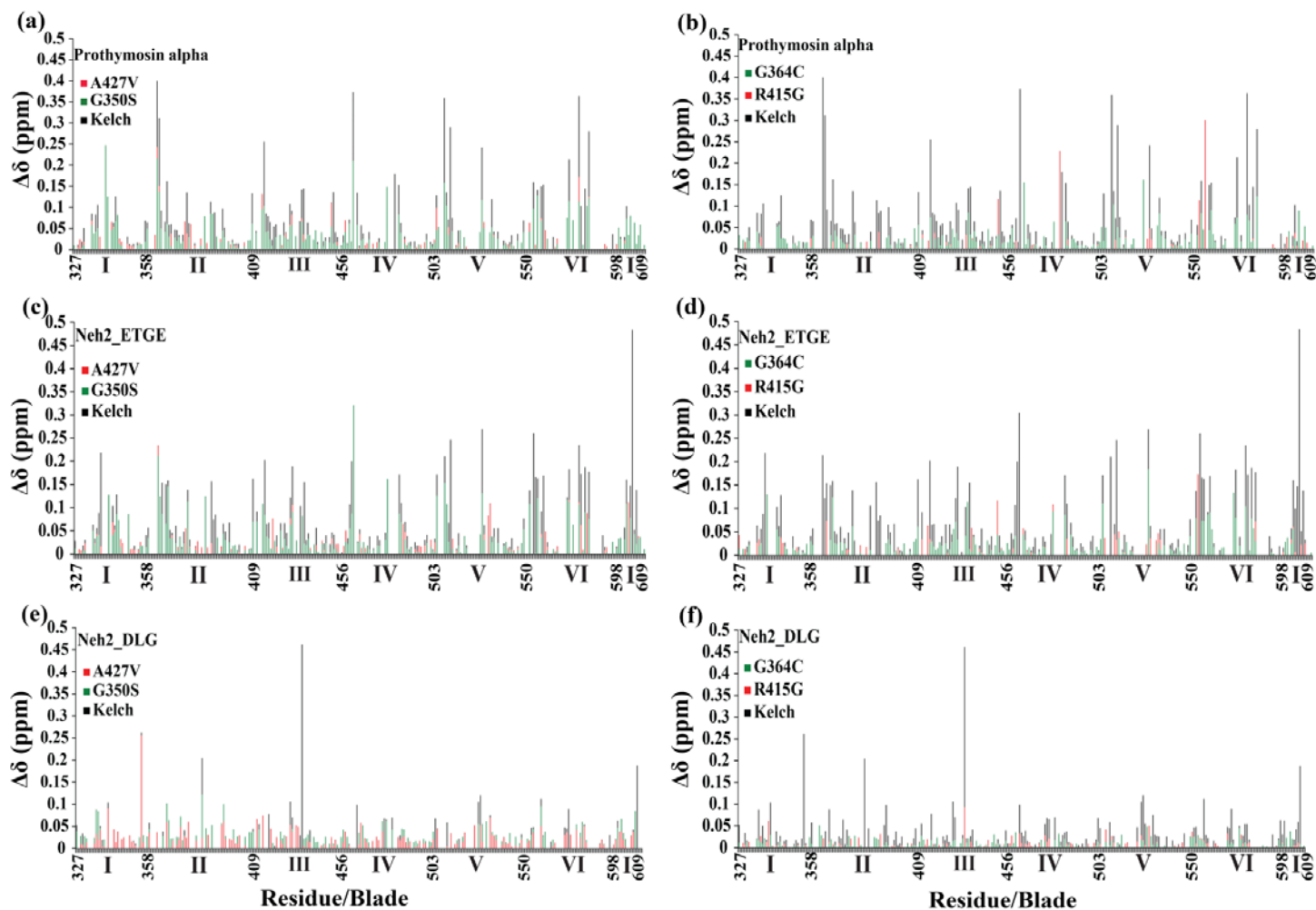




**Figure 4.13.** Overlay of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectra of the (a) G364C, (b) R415G, (c) G350S, and (d) A427V Kelch mutations in the absence (black) and presence (orange) of a two molar equivalent of the ProT $\alpha$ -ENGE peptide.



**Figure 4.13.** Overlay of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectra of the Kelch (a) G364C, (b) R415G, (c) G350S, and (d) A427V mutations in the absence (black) and presence (orange) of a five molar equivalent of the Neh2-DLG peptide.



**Figure 4.14. Composite  $^1\text{H}_\text{N}$  and  $^{15}\text{N}$  chemical shift perturbation ( $\Delta\delta = [(\Delta\delta_{\text{HN}})^2 + (\Delta\delta_{\text{N}}/5)^2]^{1/2}$ ) analysis of the wild type and the four mutants in the presence of two molar equivalents of ProT $\alpha$  peptide. (a, b), two molar equivalents of Neh2-ETGE peptide (c, d), and five molar equivalents of Neh2-DLG peptide (e, f). The boundaries of the propeller blades are defined based on (Li, Zhang et al. 2004).**

## 4.4 DISCUSSION

In this work, we have characterized the effects of cancer-associated somatic mutations on structure and target recognition of Keap1. Nine mutations identified in lung cancer patients or cancer cell lines, were selected based on their structural importance in the Kelch domain (Table 4.1). ITC, CD and NMR techniques were used to quantify the binding parameters of these mutational variants with two of the binding targets of Keap1 and to assess the conformational changes these mutations induce on the Kelch structure. Our results indicated that the disease-linked somatic mutations have differential effects on the structural integrity and the target recognition of Keap1.

The G333, G379, G430, G476 and R413 residues are conserved in the Kelch domain. Our protein expression, purification and NMR data (Figures 4.3) suggest that mutating these residues can result in the unfolding/misfolding of the Kelch domain. Further *in vivo* experiments are required to verify these findings. As mentioned earlier, the G333, G379, G430 and G476 all play key roles in the folding of the Kelch domain (Li, Zhang et al. 2004). The crystal structure of the human Kelch domain (PDB: 1U6D) shows that the backbone torsion angles of these four glycines are generally located in disallowed regions. Therefore, mutating these residues to any other amino acid with a  $\beta$ -carbon side chain may induce severe conformational strain in the structure, causing the protein to unfold/misfold. Previous studies using immunoprecipitation and luciferase assays have demonstrated that the G333C, G379D, and G430C mutants are defective in repressing Nrf2-mediated transcription of cytoprotective genes (Singh, Misra et al. 2006; Ohta, Iijima et al. 2008; Shibata, Kokubu et al. 2008). In agreement with these findings,

our results suggest that reduction in protein stability is the cause of the loss-of-function of these mutants. We speculate that any other mutations occurring in the double glycine repeats (e.g. G333S and G524C listed in the COSMIC database) will be detrimental to Kelch's structural integrity (Singh, Misra et al. 2006; Ohta, Iijima et al. 2008; Shibata, Kokubu et al. 2008).

Although the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of Kelch R413L resembled that of a folded protein (Figure 4.3), the mutant is not stable compared to the wild-type Kelch. R413, located in Blade III, is one of the conserved Arg residues found in all six blades of the  $\beta$ -propeller structure (Li, Zhang et al. 2004). The side-chain of this residue is involved in inter-blade hydrogen bonding with G379, the first residue of the glycine doublet in Blade II (Li, Zhang et al. 2004). It has been suggested that the inter-blade hydrogen bonding between these conserved Arg and Gly residues are important for maintaining the Kelch structure (Li, Zhang et al. 2004). Our results demonstrated that a single point mutation in the R413 position is sufficient to destabilize the protein, and probably leads to the loss of function. The impacts of mutations of other conserved Arg residues (e.g. R362Q, R460G, R460S, R554Q, and R601W listed in COSMIC) on the protein stability remain to be elucidated.

Unlike the G333C, G379D, R413L, G430C, and G476R mutants, G350S, G364C, R415G, and A427V have comparable solubility to the wild-type protein. CD analysis of these four mutational variants illustrated that they all possess similar secondary structure contents as the wild-type (Figure 4.7). On the other hand, our ITC results clearly demonstrated that, depending on their locations in the protein, these four mutations exert differential impacts on the target binding of the Kelch domain. The G350 residue is

located in a loop region at the opposite face of the ligand-binding pocket of the  $\beta$ -propeller structure, forming a hydrogen bond with the neighboring N346 residue (Li, Zhang et al. 2004). Crystal structures of the Kelch domain in complex with the Neh2-ETGE, Neh2-DLG, and ProT $\alpha$  peptides revealed that this residue was not directly involved in the interactions with targets (Lo, Li et al. 2006; Tong, Padmanabhan et al. 2007; Padmanabhan, Nakamura et al. 2008). Even though the G350S mutation has been identified in non-small-cell lung cancer cell lines (Singh, Misra et al. 2006) and gastric cancer tissue (Yoo, Kim et al. 2012), the mutation appears to cause no substantial changes to the binding of ProT $\alpha$  or the Neh2 domain of Nrf2: only a 1-2 fold decrease in the binding affinity to ProT $\alpha$ , Neh2-ETGE, and Neh2-DLG motifs compared to the wild-type Kelch were observed. Our NMR data demonstrate that the G350S mutation does not lead to any significant change in the protein structure. Together, our results suggest that this particular mutation does not hamper the complex formation of Keap1 with Nrf2 and ProT $\alpha$ . Effects of this mutation on Keap1's interactions with other targets, however, remain to be investigated (Cino, Killoran et al. 2013; Hast, Cloer et al. 2014).

The Kelch A427V is another cancer-linked mutation (identified in cancer tissue of a patient with small cell carcinoma (Ohta, Iijima et al. 2008)) that was observed to interact with Nrf2 and ProT $\alpha$  with similar affinities to wild-type Kelch. In fact, this particular Ala to Val substitution resulted in modest increases (~ 2 fold) in the binding affinity of Kelch to ProT $\alpha$ -ENGE and Neh2-ETGE peptides (Table 4.2). Our NMR results further confirm that the binding of this mutant to ProT $\alpha$ , Neh2-ETGE and Neh2-DLG motifs are similar to the wild-type interactions with targets. Intriguingly, while the G350S mutation resulted in subtle chemical shift changes to residues around the mutation

site (Figure 4.10a), the A427V mutation, found in the  $\beta 2$  strand of blade III, leads to local structural changes in blades II and III, which are reflected by the substantial chemical shift changes observed (Figure 4.10c). However, these local structural changes do not impair the binding of Kelch to either Neh2 or ProT $\alpha$ . Since the Kelch domain of Keap1 also mediates the interaction with WTX (Camp, James et al. 2012), p62 (Komatsu, Kurokawa et al. 2010), PGAM5 (Lo and Hannink 2006), PALB2 (Ma, Cai et al. 2012), FAC1 (Strachan, Morgan et al. 2004), IKK $\beta$  (Kim, You et al. 2010) and Bcl2 (Niture and Jaiswal 2011), characterizing the effect of the A427V mutation on the binding of these targets will provide further insight into the links of this particular somatic mutation to the associated disease. It is noteworthy that Hast *et al.* (Hast, Cloer et al. 2014) recently reported that the R470C, D422N, and G423V mutations also increase the binding to Nrf2 compared to the wild-type Keap1 based on their immunoprecipitation and Western blot analysis. However, the molecular effects of these cancer-associated mutations on the structure of Kelch remain unknown.

Meanwhile, our ITC and NMR results show that both the G364C and R415G mutations impair Kelch's ability to interact with Nrf2 and ProT $\alpha$ , albeit to different extents. The G364C substitution resulted in significantly lower affinity for both Neh2-ETGE and ProT $\alpha$  peptides ( $K_d(\text{Neh2-ETGE}) = 345 \text{ nM}$  and  $K_d(\text{ProT}\alpha) = 31 \text{ }\mu\text{M}$  vs  $K_d(\text{Neh2-ETGE}) = 31 \text{ nM}$  and  $K_d(\text{ProT}\alpha) = 2.7 \text{ }\mu\text{M}$  for the wild-type Kelch). The substantial decrease in affinity for the ETGE motif was also observed in the context of full-length Neh2 domain ( $K_d = 47 \text{ nM}$  vs  $K_d = 1.26 \text{ nM}$  for wild-type Kelch). The crystal structure of the human Kelch domain in complex with the ETGE peptide illustrates that G364 has  $\psi/\phi$  torsion angles of left-handed helix and is not involved in hydrogen bonding

with the target. However, this Gly residue is located next to S363, which forms hydrogen bonds with E82 of the ETGE peptides (Lo, Li et al. 2006; Padmanabhan, Nakamura et al. 2008). Replacing the Gly with a Cys residue may restrict the orientation of the S363 side chain from forming hydrogen bonds with E82. Additionally, the G364C mutation may also affect the intra-molecular hydrogen bonding with the G603 residue (Li, Zhang et al. 2004). It is noteworthy that even though the G364C and R415G mutations significantly weaken the interactions of Kelch with ProT $\alpha$  and the ETGE motif of Neh2, low binding affinities with these two targets could still be measured. On the other hand, no interaction between the DLG motif and these two mutants could be detected by ITC.

The R415 residue of Kelch is directly involved in the binding with targets (Lo, Li et al. 2006; Padmanabhan, Nakamura et al. 2008; Komatsu, Kurokawa et al. 2010). In the mouse Keap1-ProT $\alpha$  peptide complex, R415 is involved in electrostatic interactions with the E43 residue of ProT $\alpha$  (Padmanabhan, Nakamura et al. 2008). This Arg residue also forms salt bridges with the E79 residue in the binding interface with the Neh2-ETGE peptide (Lo, Li et al. 2006). As such, it is not a surprise that replacing this critical Arg residue in the positively charged pocket with a non-polar Gly can significantly weaken the interactions of the Kelch domain with its targets, and this was observed in our ITC and NMR binding studies with ProT $\alpha$  and Neh2. In agreement with this, a study by Tong *et al* (Tong, Katoh et al. 2006), which replaced R415 with a Lys also attenuated the binding of Kelch to the ETGE motif of Neh2 domain. Similarly, Lo *et al* (Lo, Li et al. 2006) detected no Nrf2 repressive activity for the R415A mutant. We speculate that mutations of other residues that are in direct contact with the ETGE and DLG motifs of Neh2 (e.g. Y334H, R483C, S555C, and Y572C) will ultimately impair the interaction of



Keap1 with Nrf2. Further *in vivo* and *in vitro* experiments will allow correlating the molecular effects of different categories of mutations with their impacts on the Nrf2-repression function systematically.

## 4.5 References

- Camp, N. D., R. G. James, et al. (2012). "Wilms tumor gene on X chromosome (WTX) inhibits degradation of NRF2 protein through competitive binding to KEAP1 protein." J Biol Chem 287(9): 6539-6550.
- Cino, E. A., R. C. Killoran, et al. (2013). "Binding of disordered proteins to a protein hub." Sci Rep 3: 2305.
- Cullinan, S. B., J. D. Gordan, et al. (2004). "The Keap1-BTB protein is an adaptor that bridges Nrf2 to a Cul3-based E3 ligase: oxidative stress sensing by a Cul3-Keap1 ligase." Mol Cell Biol 24(19): 8477-8486.
- Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J. Biomol. NMR 6(3): 277-293.
- Diaz-Jullien, C., A. Perez-Estevez, et al. (1996). "Prothymosin alpha binds histones in vitro and shows activity in nucleosome assembly assay." Biochim Biophys Acta 1296(2): 219-227.
- Dominguez, F., C. Magdalena, et al. (1993). "Tissue concentrations of prothymosin alpha: a novel proliferation index of primary breast cancer." Eur J Cancer 29A(6): 893-897.
- Forbes, S. A., N. Bindal, et al. (2011). "COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer." Nucleic Acids Res 39(Database issue): D945-950.
- Fukutomi, T., K. Takagi, et al. (2014). "Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degron and Keap1." Mol Cell Biol 34(5): 832-846.
- Hast, B. E., E. W. Cloer, et al. (2014). "Cancer-derived mutations in KEAP1 impair NRF2 degradation but not ubiquitination." Cancer Res 74(3): 808-817.
- Hast, B. E., D. Goldfarb, et al. (2013). "Proteomic analysis of ubiquitin ligase KEAP1 reveals associated proteins that inhibit NRF2 ubiquitination." Cancer Res 73(7): 2199-2210.
- Hayes, J. D. and M. McMahon (2009). "NRF2 and KEAP1 mutations: permanent activation of an adaptive response in cancer." Trends Biochem Sci 34(4): 176-188.

- Itoh, K., T. Chiba, et al. (1997). "An Nrf2/Small Maf Heterodimer Mediates the Induction of Phase II Detoxifying Enzyme Genes through Antioxidant Response Elements." Biochem Biophys Res Commun 236: 313-322.
- Itoh, K., N. Wakabayashi, et al. (1999). "Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain." Genes Dev 13(1): 76-86.
- Jaramillo, M. C. and D. D. Zhang (2013). "The emerging role of the Nrf2-Keap1 signaling pathway in cancer." Genes Dev 27(20): 2179-2191.
- Johnson, B. A. (2004). "Using NMRView to visualize and analyze the NMR spectra of macromolecules." Methods Mol Biol 278: 313-352.
- Khan, H., E. A. Cino, et al. (2013). "Fuzzy complex formation between the intrinsically disordered prothymosin alpha and the Kelch domain of Keap1 involved in the oxidative stress response." J Mol Biol 425(6): 1011-1027.
- Kim, J.-E., D.-J. You, et al. (2010). "Suppression of NF-kappaB signaling by KEAP1 regulation of IKKbeta activity through autophagic degradation and inhibition of phosphorylation." Cell Signal 22(11): 1645-1654.
- Kobayashi, A., M. I. Kang, et al. (2004). "Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2." Mol Cell Biol 24(16): 7130-7139.
- Komatsu, M., H. Kurokawa, et al. (2010). "The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1." Nat Cell Biol 12(3): 213-223.
- Kwak, M.-K. and T. W. Kensler (2010). "Targeting NRF2 signaling for cancer chemoprevention." Toxicol Appl Pharmacol 244(1): 66-76.
- Li, X., D. Zhang, et al. (2004). "Crystal structure of the Kelch domain of human Keap1." J Biol Chem 279(52): 54750-54758.
- Lo, S.-C. and M. Hannink (2006). "PGAM5, a Bcl-XL-interacting protein, is a novel substrate for the redox-regulated Keap1-dependent ubiquitin ligase complex." J Biol Chem 281(49): 37893-37903.
- Lo, S.-C., X. Li, et al. (2006). "Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling." EMBO J 25(15): 3605-3617.

- Ma, J., H. Cai, et al. (2012). "PALB2 Interacts with KEAP1 To Promote NRF2 Nuclear Accumulation and Function." Mol Cell Biol 32(8): 1506-1517.
- McMahon, M., K. Itoh, et al. (2003). "Keap1-dependent proteasomal degradation of transcription factor Nrf2 contributes to the negative regulation of antioxidant response element-driven gene expression." J Biol Chem 278(24): 21592-21600.
- McMahon, M., N. Thomas, et al. (2006). "Dimerization of substrate adaptors can facilitate cullin-mediated ubiquitylation of proteins by a "tethering" mechanism: a two-site interaction model for the Nrf2-Keap1 complex." J Biol Chem 281(34): 24756-24768.
- Müller, T. and A. Hengstermann (2012). "Nrf2: Friend and Foe in Preventing Cigarette Smoking-Dependent Lung Disease." Chem Res Toxicol 25: 1805-1824.
- Network, C. G. A. R. (2012). "Comprehensive genomic characterization of squamous cell lung cancers." Nature 489(7417): 519-525.
- Niture, S. K. and A. K. Jaiswal (2009). "Prothymosin- $\alpha$  Mediates Nuclear Import of the INrf2/Cul3-Rbx1 Complex to Degrade Nuclear Nrf2." J Biol Chem 284(20): 13856-13868.
- Niture, S. K. and A. K. Jaiswal (2011). "INrf2 (Keap1) targets Bcl-2 degradation and controls cellular apoptosis." Cell Death Differ 18(3): 439-451.
- Ohta, T., K. Iijima, et al. (2008). "Loss of Keap1 function activates Nrf2 and provides advantages for lung cancer cell growth." Cancer Res 68(5): 1303-1309.
- Padmanabhan, B., Y. Nakamura, et al. (2008). "Structural analysis of the complex of Keap1 with a prothymosin  $\alpha$  peptide." Acta Cryst F 64(Pt 4): 233-238.
- Padmanabhan, B., K. I. Tong, et al. (2006). "Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer." Mol Cell 21(5): 689-700.
- Shibata, T., A. Kokubu, et al. (2008). "Genetic alteration of Keap1 confers constitutive Nrf2 activation and resistance to chemotherapy in gallbladder cancer." Gastroenterology 135(4): 1358-1368.
- Singh, A., V. Misra, et al. (2006). "Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer." PLoS Med 3(10): e420.

Strachan, G. D., K. L. Morgan, et al. (2004). "Fetal Alz-50 clone 1 interacts with the human orthologue of the Kelch-like E2f-associated protein." Biochemistry 43(38): 12113-12122.

Suzuki, T., J. Maher, et al. (2011). "Select heterozygous Keap1 mutations have a dominant-negative effect on wild-type Keap1 in vivo." Cancer Res 71(5): 1700-1709.

## Chapter 5

### Summary

#### 5.1 Introduction

The last decade of structural studies led to the discovery of proteins that lack a stable tertiary structure yet are part of vital biological functions. These proteins, now widely referred to as, intrinsically disordered proteins (IDPs), have altered the long believed protein-function paradigm, which hypothesized that cellular processes were only performed by proteins with definite static structure. Nevertheless, it has been established that over one third of eukaryotic proteins have intrinsically disordered regions that are thirty residues or longer in length (Ward, Sodhi et al. 2004). Furthermore, with the advances in experimental technology, conformational flexibility of folded proteins has also been brought into light (Uversky 2013). Structural studies of IDPs have also enhanced our understanding of their conformational plasticity. It has been noted that IDPs do not exist as true random coils as the name suggests (Uversky 2013). Rather these proteins contain residual structure that is significant to executing biological activity. The new model defining the structure of functional proteins can be viewed as a continuous spectrum of differently disordered conformations extending from fully ordered to completely structure-less proteins (Uversky 2013).

The lack of structure provides several functional advantages to IDPs compared to ordered proteins and domains. Some of these advantages include the existence of a large interaction surface area, the presence of short linear motifs as well as structural motifs provide IDPs with the advantage to scaffold and interact with numerous targets, and the

diverse post-translational modifications of IDPs facilitate regulation of their function and stability within a cell (Gunasekaran, Tsai et al. 2003; Diella, Haslam et al. 2008; Fuxreiter, Tompa et al. 2008; Galea, Wang et al. 2008). IDPs interact with other proteins with high specificity and low affinity. Such binding mechanism allows fast association with targets and rapid dissociation after the signaling process is completed (Wright and Dyson 2009). Intriguingly, proteins that participate in binding, regulatory, and signaling functions such as transcription factors are enriched in intrinsically disordered regions, for the reason that the structural and biochemical properties of IDPs are well suited to carry out these tasks (Fuxreiter, Tompa et al. 2008; Galea, Wang et al. 2008; Wright and Dyson 2009).

Because IDPs play key roles in cellular processes, many diseases are a result of abnormal functions of IDPs, reviewed in (Uversky, Oldfield et al. 2008). Therefore, elucidating the means by which these proteins function is central to the design of therapeutic agents. In this thesis, the role of two intrinsically disordered proteins, Nrf2, a bZIP transcription factor, and Prothymosin  $\alpha$  (ProT $\alpha$ ), a protein with wide range of cellular functions, has been investigated in the oxidative stress response. The oxidative stress response regulated by Nrf2 is one of the cells' approach by which it eliminates electrophilic and oxidative toxins.

## **5.2 Previous Work**

The Nrf2 (nuclear factor erythroid 2-related factor 2) mediated signaling pathway is an adaptive response to environmental and endogenous stress stimuli whose goal is to retain a normal cellular environment by eliminating the stress factors (Itoh, Chiba et al.

1997). The Kelch-like ECH-associated protein 1 (Keap1) plays a central role in the regulation of Nrf2. The Keap1 protein was discovered as an inhibitor of the Nrf2 protein by the yeast two hybrid screening (Itoh, Wakabayashi et al. 1999). Keap1 anchors the Nrf2 in the cytoplasmic environment of the cell by binding to the N-terminal Neh2 domain of Nrf2. During normal cellular conditions, a single Nrf2 protein is bound to a Keap1 homo dimer. The Keap1 is a Nrf2 linker protein for interaction with the Cul3-based E2-ubiquitin ligase complex, leading to continuous ubiquitination of Nrf2 followed by proteosomal degradation (Kobayashi, Kang et al. 2004). High levels of Nrf2 and Nrf2 regulated genes have been observed in Keap1 disrupted mice, thus, demonstrating the important role Keap1 plays in the negative regulation of Nrf2 (Wakabayashi, Itoh et al. 2003).

The structure of the mouse and human Kelch domain has been crystalized in the free and bound state with the ETGE motif of the Neh2 domain (Li, Zhang et al. 2004; Lo, Li et al. 2006; Padmanabhan, Tong et al. 2006). Crystal structures of the mouse Kelch domain with the mouse ProT $\alpha$  peptide and DLG-Neh2 peptide are also reported in literature (Tong, Padmanabhan et al. 2007; Padmanabhan, Nakamura et al. 2008). It has been determined that the immobilization of the Nrf2 by Keap1 occurs in a sequential manner where the high affinity site, ETGE motif, binds first, followed by the docking of the DLG motif onto a second Kelch domain of the homo dimer Keap1 protein (Tong, Kobayashi et al. 2006). The Keap1-ProT $\alpha$  complex has been shown to shipped into the nucleus, where, ProT $\alpha$  dissociates and Nrf2 degradation is initiated again to reestablish homeostatic conditions (Niture and Jaiswal 2009).



Structural analysis of the ProTα and Nrf2 to the Keap1 protein have also been performed using molecular dynamic studies (Cino, Wong-ekkabut et al. 2011). It has been proposed that the binding of ProTα and the Neh2 domain to the Keap1 protein occurs via preformed bound-state-like β-turns, and that the residues outside of the motifs may also contribute to the stability of the structural elements (Cino, Wong-ekkabut et al. 2011). Preformed structural elements of other Kelch binding targets, PGAM5 (Lo and Hannink 2006), p62 (Komatsu, Kurokawa et al. 2010), WTX (Camp, James et al. 2012), FAC1 (Strachan, Morgan et al. 2004), PALB2 (Ma, Cai et al. 2012), and IKKβ (Kim, You et al. 2010) have also been examined and it has been suggested that their binding affinity to the Keap1 protein is dependent on the extent of preformed bound-state like conformation in the free structure (Cino, Killoran et al. 2013).

Mutations in Nrf2 and Keap1 in human cancers have been also reported in literature (Shibata, Kokubu et al. 2008; Hast, Cloer et al. 2014). Such mutations have been linked with constitutive expression of pro-survival enzymes in tumors, thus displaying resistance to chemotherapy and providing advantageous growth to tumor cells (Padmanabhan, Tong et al. 2006). Cancer therapy of patients with mutations in Keap1 and Nrf2 have shown poorer prognosis than cancer patients lacking mutations in the two proteins (Shibata, Kokubu et al. 2008). Effects of many of these somatic mutations on Keap1's ability to repress Nrf2 activity have been established. For example, Padmanabhan, *et al.* have demonstrated in cell studies that the G430C and G364C mutations are unable to suppress Nrf2 mediated transcription. Hast *et al* recently examined the R320 mutation in the Kelch domain of Keap1 and showed that it has a stronger affinity for Keap1 targets, Nrf2 and Cul3 (Hast, Cloer et al. 2014) and has

proposed the presence of “super binding” mutations. Moreover, a Keap1 protein with the R320 mutation was able to ubiquitinate Nrf2 but could not promote its degradation (Hast, Cloer et al. 2014).

Studies of the Keap1 and Nrf2 mutations have made it evident that inhibition of Nrf2 can repress tumor cell proliferation and enhance apoptosis (Shibata, Kokubu et al. 2008). It has been also demonstrated experimentally that introduction of Nrf2-specific small interfering RNA into cancer cells showed decreased in the growth rate of the cells, making it a great drug target (Shibata, Kokubu et al. 2008).

### **5.3 Fuzzy complex between the flexible ProT $\alpha$ and Kelch domain of human Keap1**

The protein structure-function paradigm was believed to be still applicable to disordered proteins in the target bound state, as it was observed that IDPs adopt a structure upon complex formation. This view was challenged by the discovery of IDPs that retained their flexible nature in the ligand bound states. Investigation of the ProT $\alpha$  complex formation with the Kelch domain of the human Keap1 in chapter 2 revealed that the ProT $\alpha$ -Kelch complex falls into this “fuzzy” category of disordered proteins. A narrow proton dimension in the HSQC spectra is hallmark of disordered proteins. The  $^1\text{H}$ - $^{15}\text{N}$  HSQC of ProT $\alpha$  in the free and bound states exhibited a narrow proton dimension signifying that the protein remained disordered upon complex formation. The  $R_1$ ,  $R_2$ , and steady state NOE parameters obtained through the spin relaxation experiments further strengthened the presence of a flexible state of ProT $\alpha$  in the ligand bound state. The  $R_1$  values averaged to be  $2.47 \text{ s}^{-1}$  in the bound state, larger than what is observed for a

protein with the same size as ProT $\alpha$ . Such high  $R_1$  values result from internal motions on a picosecond-to-nanosecond timescale. Significant increases in the  $R_2$  values for the binding region of the ProT $\alpha$  were observed in the bound state, however, these high values in the bound state are still smaller than what is expected for a 46-kDa folded protein. Additionally, the small NOE and amide exchange values reinforced that ProT $\alpha$  falls into the fuzzy complex category upon complex formation with the Kelch domain of the human Keap1.

#### **5.4 Binding interface of human ProT $\alpha$ with the Kelch domain of human Keap1**

Karapetian *et al.* were the first to identify the binding interface of the mouse ProT $\alpha$  to the Kelch domain (Karapetian, Evstafieva *et al.* 2005). A static image of the bound ProT $\alpha$  to the Kelch domain of the mouse Keap1 protein is also available (PDB id: 2Z32). In this crystal structure, a 16 residue ProT $\alpha$  peptide containing the ENGE motif is bound to the Kelch propeller structure (Padmanabhan, Nakamura *et al.* 2008). In chapter 2 of this work, binding region of the human ProT $\alpha$  to the Kelch domain of Keap1 is further investigated. NMR, peptide array, and ITC experimental work added to our understanding of the Keap1 interaction with the ProT $\alpha$  protein. The results indicated that the region in human ProT $\alpha$ , which binds to the bottom face of the Kelch domain involves more critical residues than previously probed in the mouse ProT $\alpha$  (Karapetian, Evstafieva *et al.* 2005). The peptide array clearly demonstrated that mutations in the  $^{38}\text{NANEENGE}^{45}$  region were damaging to the binding of the Kelch. Mutations of the E51, which had been previously shown to be important for the interaction with the Kelch domain, appeared to have no effect on the binding to target (Karapetian, Evstafieva *et al.* 2005). Thermodynamic parameters of mutational variants of ProT $\alpha$  and chemical shift

perturbations in the ProT $\alpha$  HSQC spectrum upon Kelch binding also confirmed that flanking residues surrounding the ENGE motif are essential for ProT $\alpha$ -Kelch complex formation.

Another important point experimental work presented in chapter 2 demonstrated is that although peptides are able to mimic binding motifs of proteins, their mechanism of association with targets can vary from the full length protein. The binding enthalpy and entropy of the ProT $\alpha$  peptide differed from that of the full length protein upon forming a complex with Kelch. MD simulation studies have revealed that the ProT $\alpha$  peptide has a greater loss of enthalpy due to the folding from a structure-less state to the bound-state  $\beta$ -hairpin like structure (Cino, Wong-ekkabut et al. 2011). Additionally, because the peptide has a lower propensity than the full-length ProT $\alpha$  protein in forming the bound-state  $\beta$ -hairpin like structure, the entropy loss associated with forming the bound state structure will be higher for the peptide as observed in this work (Cino, Wong-ekkabut et al. 2011). Thus it is important where possible to study ligand binding in full-length proteins as understanding the mechanism of interaction in full-length protein may aid in the proper design of drug agents.

## **5.5 The Neh2 domain of human Nrf2**

The purification of the Neh2 domain of the Nrf2 has been discovered to be challenging as demonstrated for the mouse system (Tong, Yamamoto et al. 2008). In this work, a new protocol for the purification of the human Neh2 domain of Nrf2 is established, which yields high quantity of pure protein allowing studying the role of the full-length Neh2 domain in the oxidative stress response. The amide backbone resonance

assignment of the human Neh2 domain has laid the foundation for future ligand binding work with the Kelch domain or other binding targets that interact with the Neh2 domain, such as the p21 protein (Chen, Sun et al. 2009).

Structural analysis of the Neh2 domain of human Nrf2 revealed that Neh2 has greater residual structure, compare to the flexible ProT $\alpha$ . The central region of the human Neh2 domain is helical in nature and the C-terminal end has a  $\beta$ -turn propensity as observed for the mouse Neh2 (Tong, Katoh et al. 2006). Based on the relaxation data, the N-terminal region of the Neh2 domain, residues 14 to 31, appears to possess residual structure, *NOE* values increase up to 0.4 indicating that the region is less flexible. Resonances of residues involved in complex formation undergo line broadening and cannot be observed even when temperature is increased by 5 °C. Nevertheless, the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of the Neh2 $\Delta$ DLG and Neh2 $\Delta$ ETGE in the Kelch bound state reveal a narrow proton dimension of the Neh2 protein. Additionally, relaxation data of the Neh2-Kelch complex is not available, yet, based on the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra one can conclude that the Neh2 regions not involved in forming binding interface predominantly remain unchanged in structure.

## **5.6 Interaction between the Neh2 and Kelch domain**

NMR and ITC studies of the complex formation between the Neh2 domain of mouse Nrf2 and the Kelch domain of mouse Keap1 demonstrated that the process is exothermic for both the DLG and ETGE motifs with a  $\Delta H$  of  $\sim -7$  for the DLG-Kelch complex formation (Tong, Katoh et al. 2006). Upon titration of the Neh2 to the Kelch domain of human Keap1 at room temperature and vice versa, a two site binding profile,

as seen in the mouse system was not observed for the binding of the Neh2 and Kelch domains (Tong, Katoh et al. 2006). Lowering the temperature revealed that the  $\Delta H$  associated with the DLG-Kelch binding is close to -1.2 kcal/mol providing an explanation why a two site model could not be detected at room temperature. ITC experiments performed at 5 °C exhibit a two site binding for the Kelch-Neh2 complex using human derived constructs. These results suggest that the mechanism of binding of the DLG motif of Neh2 to the Kelch domain of the human Keap1 maybe different than what is observed in the mouse system.

## **5.7 Order in disorder and its implication in target binding**

The Neh2 domain of Nrf2 and ProT $\alpha$  are great examples for studying target binding of IDPs. In this work, the binding of ProT $\alpha$  and Nrf2 with the Kelch domain has delineated the underlying mechanism of their differential binding to the Kelch domain. As stated earlier, IDPs are not true random coils (Uversky 2013). Some IDPS have secondary structures and preformed structural elements that have important implications in target binding. Previous studies of the Neh2 domain and ProT $\alpha$  combined with this work emphasize that ProT $\alpha$  and the Neh2 domain behave differently structurally in the free and target bound states with Kelch (Padmanabhan, Tong et al. 2006; Tong, Padmanabhan et al. 2007; Padmanabhan, Nakamura et al. 2008; Cino, Wong-ekkabut et al. 2011). It has been demonstrated that the full length ProT $\alpha$  is highly flexible and less compact compare to the Neh2 domain of Nrf2 in the free and target bound. A high entropic penalty is associated with attaining the  $\beta$ -turn structure required for

ProT $\alpha$  interaction with the Kelch domain resulting in a higher dissociation constant compare to the Neh2-ETGE –Kelch complex (Khan 2013).

The data presented in chapter 3 shows presence of secondary structure content and a compact Neh2 domain of the Nrf2 protein compare to ProT $\alpha$ . The Neh2 -ETGE motif has preformed  $\beta$ -turn like structure in the free state which allows less energy spent on adopting a bound state like structure and thus binds tighter to the Kelch domain (Cino, Wong-ekkabut et al. 2011). Additionally, the SSP scores of the N-terminal region of the Neh2 domain in chapter 3 displays an  $\alpha$ -helical propensity (Figure 5.1a). Recently, Fukutomi et al, showed that the Neh2-DLG motif under goes a disorder to order transition by adopting a stable helix in the target bound state (Fukutomi, Takagi et al. 2014).

These differences in the presence of preformed structural elements and the overall compactness between ProT $\alpha$  and the Neh2 domain of Nrf2 have important implication in their differential binding affinity to the Kelch domain of human Keap1. A detail list of differences that contribute to the binding affinity of these two disordered proteins to the Kelch domain is presented in table 5.1.

## **5.8 Somatic mutations in the Kelch domain and its impact on target recognition**

The G364C and G430C were amongst the first somatic mutations, isolated from lung cancer tissue of a patient and lung cancer-derived cell lines, respectively, exposed in the Kelch domain of Keap1 (Padmanabhan, Tong et al. 2006). Luciferase assays of the G430C and G364C had revealed that in the presence of these mutations, Keap1 could not

**Table 5.1. A table listing structural differences between Neh2 and ProTα.** The presence of preformed structural elements allows Neh2 to form a tighter complex with the Kelch domain of Keap1 compare to ProTα.

### Neh2

- Presence of a central helix
- Residues 13 to 29 have a helical propensity
- Residues 82 to 94 have a propensity for a  $\beta$ -turn
- Hydrodynamic radius of  $\sim 25\text{\AA}$  (more compact)
- DLG motif forms a helix upon complex formation with Kelch
- ETGE forms a stable  $\beta$ -turn in the bound state

### ProTα

- One binding motif for Kelch
- Mostly disordered and extended
- Hydrodynamic radius of  $34\text{\AA}$  (Extended structure)
- Dynamic even in the target bound state even though ENGE forms a  $\beta$ -turn upon binding Kelch



perform its Nrf2 repression role, leading to constitutive Nrf2 mediated enzyme expression. This continuous presence of cytoprotective enzymes aided tumor cell proliferation (Padmanabhan, Tong et al. 2006). Since the initial discovery of these mutations, numerous other somatic alterations within the Keap1 protein have been revealed. In this thesis, nine somatic mutations, in the Kelch domain of human Keap1, discovered either in patient tissues or cancer cell lines, were created and their impact on Kelch structural integrity and Kelch target recognition was elucidated.

The purification procedure applied to obtain pure soluble samples for G333C, G379D, G430C, and G476R mutants resulted in aggregation of the protein samples and could not be used for ligand binding studies. These glycine residues, occur as doublets in the  $\beta$ -propeller and are found to be critical for the structural integrity of the Kelch domain (Li, Zhang et al. 2004), thus, introduction of other amino acids with  $\beta$ -carbon side chain will result in misfolding or unfolding of the Kelch domain. Since ProT $\alpha$  and Nrf2 are the weakest and strongest binding partners of the Kelch domain, respectively, (Cino, Killoran et al. 2013), we speculate that the Keap1 protein containing mutations within this double glycine region of the Kelch domain will also be incapable of interacting with other known targets such as p62, PALB2, (Komatsu, Kurokawa et al. 2010).

Another mutation that rendered the overall structural stability of the Kelch propeller is the R413L mutation. It is a conserved arginine residue detected in all six blades of the Kelch domain and its side chain hydrogen bonds with the conserved G379 residue. The  $^1\text{H}$ - $^{15}\text{N}$  HSQC obtained from the soluble fraction of this mutant resulted in a spectrum that resembled a folded protein but could not be compared to the wild-type

spectrum of the Kelch domain. Additionally, the protein sample and purity were compromised even at low temperatures. Based on the experimental evidence presented in chapter 4, it can be concluded that a mutation in this conserved Arg residue has an impact on the  $\beta$ -propeller structure and is not able to maintain wild-type interactions with targets.

The G350S, G364C, R415G, and A427V mutations were also studied. It has been shown through CD and NMR that the secondary structure content of the wild-type is not disrupted in these mutants and these mutations do not undergo aggregation. Binding studies with ProT $\alpha$  and Neh2 revealed that the G350S and A427V are able to maintain wild-type interactions. In fact the comparatively stronger binding of the A427V mutant to ProT $\alpha$  and Neh2 suggests that it may fall into the “super binder” category as observed with some other mutants (Hast, Cloer et al. 2014). The G364C and R415G mutations that occur on the binding interface are unable to interact with the DLG motif of the Neh2 domain.

## **5.9 Conclusion**

The work in this thesis has added valuable information to the understanding of the underlying mechanism by which the Kelch domain of the human Keap1 interacts with its targets. A new purification protocol of the Neh2 domain of the human Nrf2 protein has opened doors for future studies of the oxidative stress response in the human system. Details on the binding affinity of Kelch mutational variants affirm that the double glycines and the conserved Arg residues are critical for the structural integrity of the

Kelch domain. Other mutations studied in this work did not affect the overall stability of the  $\beta$ -propeller but did result in differential binding to targets.

### 5.10 Future Work

The data presented in this thesis demonstrate that some mutations occurring within the Kelch domain of the Keap1 protein in cancer patients or derived from cancer cells lines are able to maintain wild-type target binding. Since poorer prognosis of cancer patients with mutations in Keap1 has been noted (Shibata, Kokubu et al. 2008), it will be of great interest to analyze the binding of these mutant proteins in in vivo functional assays such as luciferase assays and determine whether they are still able to carry out Nrf2 degradation in normal and tumor cells. Additionally, this work provides the NMR backbone resonance assignment of the Kelch domain allowing studying the impact of these mutations on binding to other targets such as p62 (Komatsu, Kurokawa et al. 2010), PGAM5(Lo and Hannink 2006), PALB2(Ma, Cai et al. 2012), WTX(Camp, James et al. 2012), IKK $\beta$  (Kim, You et al. 2010) or FAC1(Strachan, Morgan et al. 2004). Such studies can shed further light on how these mutations are pro cell survival in tumor cells. The binding studies can also aid in understanding the underlying mechanism of interaction between Keap1 and other targets.

Many proteins have been reported in literature that are in competition with Nrf2 for Keap1 binding and promote the activation of the Nrf2 signaling pathway. Recently, p21<sup>Cip1/WAF1</sup> has been shown to compete Keap1 for Nrf2 binding (Chen, Sun et al. 2009). Immunoprecipitation assays indicated that the C-terminal region of the p21 binds to the DLG motif of the Neh2 domain and thus prevents Nrf2 ubiquitination (Chen, Sun et al.

2009). The C-terminal of the p21 is disordered in conformation and adopts a helical structure in CaM bound state, and remains extended when interacting with the PCNA (Gulbis, Kelman et al. 1996; Yoon, Venkatachalam et al. 2009). Binding studies of the C-terminal region of the p21 to the Neh2 domain can add further to our understanding of how disordered proteins interact with other unstructured partners. Whether the Neh2 domain or the p21 adopt a structured state upon binding to a structure-less protein is also exciting to examine. Such studies show the structural flexibility of disordered proteins upon ligand binding, that is, the disordered protein can adopt diverse conformations upon binding different targets via the same binding site. This information is crucial to understanding the binding mechanism of Nrf2 with its partners and the information is crucial to designing drug agents that are used to repress Nrf2 activity in chemotherapy.

## 5.11 References

- Camp, N. D., R. G. James, et al. (2012). "Wilms tumor gene on X chromosome (WTX) inhibits degradation of NRF2 protein through competitive binding to KEAP1 protein." *J Biol Chem* **287**(9): 6539-6550.
- Chen, W., Z. Sun, et al. (2009). "Direct interaction between Nrf2 and p21(Cip1/WAF1) upregulates the Nrf2-mediated antioxidant response." *Mol Cell* **34**(6): 663-673.
- Cino, E. A., R. C. Killoran, et al. (2013). "Binding of disordered proteins to a protein hub." *Sci Rep* **3**: 2305.
- Cino, E. A., J. Wong-ekkabut, et al. (2011). "Microsecond molecular dynamics simulations of intrinsically disordered proteins involved in the oxidative stress response." *PLoS One* **6**(11): e27371.
- Diella, F., N. Haslam, et al. (2008). "Understanding eukaryotic linear motifs and their role in cell signaling and regulation." *Front Biosci* **13**: 6580-6603.
- Fukutomi, T., K. Takagi, et al. (2014). "Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degron and Keap1." *Mol Cell Biol* **34**(5): 832-846.
- Fuxreiter, M., P. Tompa, et al. (2008). "Malleable machines take shape in eukaryotic transcriptional regulation." *Nat Chem Biol* **4**(12): 728-737.
- Galea, C. A., Y. Wang, et al. (2008). "Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits." *Biochemistry* **47**(29): 7598-7609.
- Gulbis, J. M., Z. Kelman, et al. (1996). "Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA." *Cell* **87**(2): 297-306.
- Gunasekaran, K., C. J. Tsai, et al. (2003). "Extended disordered proteins: targeting function with less scaffold." *Trends Biochem Sci* **28**(2): 81-85.
- Hast, B. E., E. W. Cloer, et al. (2014). "Cancer-derived mutations in KEAP1 impair NRF2 degradation but not ubiquitination." *Cancer Res* **74**(3): 808-817.
- Itoh, K., T. Chiba, et al. (1997). "An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements." *Biochem Biophys Res Commun* **236**(2): 313-322.
- Itoh, K., N. Wakabayashi, et al. (1999). "Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain." *Genes Dev* **13**(1): 76-86.

- Khan, H., Cino, EA., Brickenden, A., Fan, J., Yang, D., Choy, WY. (2013). Fuzzy complex formation between the intrinsically disordered prothymosin  $\alpha$  and the Kelch domain of Keap1 involved in the oxidative stress response. *J.Mol.Biol.* 425: 1011-27.
- Karapetian, R. N., A. G. Evstafieva, et al. (2005). "Nuclear oncoprotein prothymosin alpha is a partner of Keap1: implications for expression of oxidative stress-protecting genes." *Mol Cell Biol* **25**(3): 1089-1099.
- Kim, J. E., D. J. You, et al. (2010). "Suppression of NF-kappaB signaling by KEAP1 regulation of IKKbeta activity through autophagic degradation and inhibition of phosphorylation." *Cell Signal* **22**(11): 1645-1654.
- Kobayashi, A., M. I. Kang, et al. (2004). "Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2." *Mol Cell Biol* **24**(16): 7130-7139.
- Komatsu, M., H. Kurokawa, et al. (2010). "The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1." *Nat Cell Biol* **12**(3): 213-223.
- Li, X., D. Zhang, et al. (2004). "Crystal structure of the Kelch domain of human Keap1." *J Biol Chem* **279**(52): 54750-54758.
- Lo, S. C. and M. Hannink (2006). "PGAM5, a Bcl-XL-interacting protein, is a novel substrate for the redox-regulated Keap1-dependent ubiquitin ligase complex." *J Biol Chem* **281**(49): 37893-37903.
- Lo, S. C., X. Li, et al. (2006). "Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling." *EMBO J* **25**(15): 3605-3617.
- Ma, J., H. Cai, et al. (2012). "PALB2 interacts with KEAP1 to promote NRF2 nuclear accumulation and function." *Mol Cell Biol* **32**(8): 1506-1517.
- Niture, S. K. and A. K. Jaiswal (2009). "Prothymosin-alpha mediates nuclear import of the INrf2/Cul3 Rbx1 complex to degrade nuclear Nrf2." *J Biol Chem* **284**(20): 13856-13868.
- Niture, S. K. and A. K. Jaiswal (2011). "INrf2 (Keap1) targets Bcl-2 degradation and controls cellular apoptosis." *Cell Death Differ* **18**(3): 439-451.
- Padmanabhan, B., Y. Nakamura, et al. (2008). "Structural analysis of the complex of Keap1 with a prothymosin alpha peptide." *Acta Crystallogr Sect F Struct Biol Cryst Commun* **64**(Pt 4): 233-238.
- Padmanabhan, B., K. I. Tong, et al. (2006). "Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer." *Mol Cell* **21**(5): 689-700.

- Shibata, T., A. Kokubu, et al. (2008). "Genetic alteration of Keap1 confers constitutive Nrf2 activation and resistance to chemotherapy in gallbladder cancer." Gastroenterology **135**(4): 1358-1368, 1368 e1351-1354.
- Strachan, G. D., K. L. Morgan, et al. (2004). "Fetal Alz-50 clone 1 interacts with the human orthologue of the Kelch-like ECH-associated protein." Biochemistry **43**(38): 12113-12122.
- Tong, K. I., Y. Katoh, et al. (2006). "Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model." Mol Cell Biol **26**(8): 2887-2900.
- Tong, K. I., A. Kobayashi, et al. (2006). "Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism." Biol Chem **387**(10-11): 1311-1320.
- Tong, K. I., B. Padmanabhan, et al. (2007). "Different electrostatic potentials define ETGE and DLG motifs as hinge and latch in oxidative stress response." Mol Cell Biol **27**(21): 7511-7521.
- Tong, K. I., M. Yamamoto, et al. (2008). "A simple method for amino acid selective isotope labeling of recombinant proteins in E. coli." J Biomol NMR **42**(1): 59-67.
- Uversky, V. N. (2013). "A decade and a half of protein intrinsic disorder: biology still waits for physics." Protein Sci **22**(6): 693-724.
- Uversky, V. N., C. J. Oldfield, et al. (2008). "Intrinsically disordered proteins in human diseases: introducing the D2 concept." Annu Rev Biophys **37**: 215-246.
- Wakabayashi, N., K. Itoh, et al. (2003). "Keap1-null mutation leads to postnatal lethality due to constitutive Nrf2 activation." Nat Genet **35**(3): 238-245.
- Ward, J. J., J. S. Sodhi, et al. (2004). "Prediction and functional analysis of native disorder in proteins from the three kingdoms of life." J Mol Biol **337**(3): 635-645.
- Wright, P. E. and H. J. Dyson (2009). "Linking folding and binding." Curr Opin Struct Biol **19**(1): 31-38.
- Yoon, M. K., V. Venkatachalam, et al. (2009). "Residual structure within the disordered C-terminal segment of p21(Waf1/Cip1/Sdi1) and its implications for molecular recognition." Protein Sci **18**(2): 337-347.





## Curriculum Vitae

**Halema Khan**

**Department of Biochemistry**

**The University of Western Ontario, London, ON N6A 5C1**

### **Education**

---

2008-2014	University of Western Ontario	London, ON
	Ph.D. (Biochemistry)	
2003-2008	University of Western Ontario	London, ON
	BMSc. Honours (BioMedical Sciences and Biochemistry)	

### **Awards Received**

---

1. Department of Biochemistry, University of Western Ontario,  
**Western Graduate Research Scholarship, 2008-2013**
2. Department of Science, University of Western Ontario  
**Deans Honor list, 2008**
3. Department of Science, University of Western Ontario  
**Western University Entrance Scholarship, 2003**
4. University of Western Ontario  
**WISE Scholarship, 2002**
5. Ministry of Education, London, Canada  
**Ontario Scholar Award, 2003**
6. Ministry of Education, London, Canada  
**Outstanding student award, 2003**

## **Academic and Research Work Experience**

Sept 2008-Present	University of Western Ontario
<i>Graduate student</i>	
Sept 2009-April 2011	University of Western Ontario
<i>Teaching Assistant</i>	
Jan 2007 to Dec. 2010	Agriculture Agri-Food Canada
<i>Research Assistant</i>	
May 2006 to Dec. 2006	Patheon Pharmaceuticals
<i>Research Assistant</i>	
May 2005 to May 2006	University of Western Ontario
<i>Summer Undergraduate Research Assistant</i>	

## **Publications**

Khan, H., Cino, EA., Brickenden, A., Fan, J., Yang, D., Choy, WY. (2013). Fuzzy complex formation between the intrinsically disordered prothymosin  $\alpha$  and the Kelch domain of Keap1 involved in the oxidative stress response. *J.Mol.Biol.* **425**: 1011-27.

### *Article Submitted*

Khan, H., Killoran, R., Brickenden, A., Fan, J., Yang, D., Choy, WY. (2014). Molecular effects of cancer-associated somatic mutations on the structural and target recognition properties of Keap1. *Biochem.J.*

## **Conference Presentations**

1. Khan, H., Choy, W. (2012). Effects of cancer-associated somatic mutations on the structural and target recognition properties of Keap1. Department of Biochemistry Seminar Series. Oral presentation, University of Western Ontario.
2. Khan, H., Choy, W. (2012). Effects of cancer-associated somatic mutations on the structural and target recognition properties of Keap1.

Department of Biochemistry Seminar Series, Poster. University of Western Ontario.

3. Khan, H., Choy, W. (2010). Structure-Function Relationship of Prothymosin $\alpha$  in the oxidative stress response. Poster. Gordon Research Conference: Intrinsically disordered proteins. Davidson College, Davidson, North Carolina.
4. Khan, H., Choy, W. (2010). Structure-Function Relationship of Prothymosin $\alpha$  in the oxidative stress response. Biochemistry seminar series. Oral Presentation University of Western Ontario.