THE N-END RULE PATHWAY: MOLECULAR PRINCIPLES OF STRUCTURAL RECOGNITION AND RATIONAL DESIGN OPPORTUNITIES

by

Shashi Kanth Murthy Sriram

Bachelors of Pharmacy, Jawaharlal Nehru Technological University, 2004

Masters of Science in Molecular Sciences and Nanotechnology, Louisiana Tech University, 2006

Submitted to the Graduate Faculty of

School of Pharmacy in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

SCHOOL OF PHARMACY

This dissertation was presented

by

Shashi Kanth Murthy Sriram

It was defended on

April 18, 2012

and approved by

Song Li, MD, PhD, Associate Professor, Pharmaceutical Sciences

Samuel M. Poloyac, PharmD, PhD, Associate Professor, Pharmaceutical Science

Yong Wan, PhD, Associate Professor, Department of Cell Biology & Physiology

Wen Xie, MD, PhD, Professor, Pharmaceutical Science

Dissertation Advisor: Yong Tae Kwon, PhD, Associate Professor, Pharmaceutical Science

Copyright © by Shashi Kanth Murthy Sriram

2012

THE N-END RULE PATHWAY: MOLECULAR PRINCIPLES OF STRUCTURAL RECOGNITION AND RATIONAL DESIGN OPPORTUNITIES

Shashi Kanth Murthy Sriram, PhD

University of Pittsburgh, 2012

The N-end rule relates the regulation of the *in vivo* half-life of a protein to the identity of its N-terminal residue. A set of N-terminal degradation signals is targeted by the recognition components (N-recognins) of the N-end rule pathway. Recent reports on the N-end rule substrates, components, functions and structural basis of substrate recognition have provided critical insights. The N-end rule pathway is now emerging as a major ubiquitin-dependent cellular proteolytic system. The scope of this dissertation is to understand the structural principles of substrate recognition and utilize this structural basis to provide insights on the functions and underlying mechanisms of the N-end rule pathway. We were also interested in exploring the N-end rule pathway as an intracellular target for heterodivalent interactions. We discuss and demonstrate the basis of thermodynamics and kinetics principles that govern these interactions between the N-recognins and the heterodivalent ligands. We further pursued to exploit these principles in the design and development of high affinity ligands that target the N-end rule pathway.

TABLE OF CONTENTS

PREFACE X

1.0 THE N-END RULE PATHWAY 1

1.1 MAMMALIAN N-END RULE 4

1.2 YEAST N-END RULE 6

1.3 BACTERIAL N-END RULE 11

1.4 PLANT N-END RULE 13

1.5 DROSOPHILA N-END RULE 16

2.0 STRUCTURAL BASIS OF THE N-END RULE RECOGNITION 18

2.1 STRUCTURAL BASIS OF TYPE 1 RECOGNITION 21

2.2 STRUCTURAL BASIS OF TYPE 2 RECOGNITION 25

2.3 MOLECULAR PRINCIPLES OF THE N-END RULE RECOGNITION 28

2.4 METHODS 32

3.0 RATIONAL DESIGN OF INHIBITORS TARGETING THE N-END RULE 33

3.1 MULTIVALENT INTERACTIONS 33

3.1.1 Principles 35

3.1.2 Thermodynamics And Kinetics 36

3.2 HETEROVALENT INHIBITORS OF THE N-END RULE 39

3.2.1 RF-C11 Prototype and L1L2-Cn Family 42

3.2.2 Optimal Linker Length 45

3.2.3 Monovalent Derivatives 54

3.3 METHODS 63

4.0 DISCUSSION 65

4.1 N-END RULE CODE 65

4.2 UBR BOX 67

4.3 FUNCTIONS AND IMPLICATIONS 69

4.4 FRAGMENT BASED DRUG DESIGN 73

4.5 VISION AND FUTURE DIRECTION 74

APPENDIX A 76

BIBLIOGRAPHY 84

LIST OF TABLES

Table 1. N-degrons and Modification 5

Table 2. Substrates of the N-end rule pathway 17

LIST OF FIGURES

Figure 1. Ubiquitin-dependent Proteolytic System (UPS). 3

Figure 2. The N-end Rule Pathway. 8

Figure 3. Mammalian UBR family and known and putative N-recognins from other species. 9

Figure 4. Eukaryotic N-recognins. 10

Figure 5. N-end rule pathway functions. 15

Figure 6. Examples of the N-end rule in physiology. 11

Figure 7. Sequence homology of UBR box and N-domain. 19

Figure 8. Zinc finger in UBR box. 23

Figure 9. Structural basis of type1 binding. 24

Figure 10. Structural basis of type2 recognition. 27

Figure 11. Molecular Principles of substrate recognition. 31

Figure 12. Types of multivalent interactions. 34

Figure 13. Influence of linker length. 39

Figure 14. L_1L_2-Cn family. 41

Figure 15. Effect of linker length on RF-Cn activity. 47

Figure 16. RF-Cn activity trend with linker length. 48

Figure 17. Width and linker length. 50

Figure 18. Multivalency-assisted enhancement. 51

Figure 19. L_1L_2-C5 controls. 53

Figure 20. Role of α-amino group. 55

Figure 21. Role of peptide bond. 57

Figure 22. Amphetamine as N-degron. 58

Figure 23. Trace amines as N-degrons. 60

Figure 24. Role of C-terminal side. 62

PREFACE

Dedicated to my parents, Smt. S. Laxmi and Sri. S. Guru Murthy, to my niece Ms. Adhya Sriram and nephew Mstr. Arya Sriram. I am very grateful to the support of my brother Mr. Vishnu Sriram, my sister-in-law Mrs. Shilpa Sivakumaran and my wife Mrs. Sumi Bose. Thank you to my uncles, Dr. Katakam Veerabadra Rao and Dr. Meenakshi Sivakumaran. Lastly, I wish to express my sincere gratitude towards my advisor Dr. Yong Tae Kwon, his lab and the graduate program.

1.0 THE N-END RULE PATHWAY

Proteins play a critical role in essentially all cellular processes and are considered as principal biochemical compounds required for cellular functions. The concentration of a protein determines its propensity to perform a given function and the individual protein concentration is regulated by a balance between the rate of synthesis (transcription and translation) and the rate of degradation (proteolysis). Traditionally proteolysis was considered a bulk, non-selective process by cytosolic proteases and organelles like lysosomes. Discovery of ubiquitin (Ub) field and regulated proteolysis by the ubiquitin–dependent proteolytic system (UPS) in 1980's changed the prevailing understanding about intracellular proteolysis (Bachmair et al., 1986). Highlighting the importance of regulated protein degradation over that of transcription and translation on physiological regulation.

Ub is a 76-amino acid protein whose conjugation to other proteins regulates a variety of biological processes. UPS involves the tagging of a target protein through covalent conjugation of Ub to an internal Lys residue mediated by the E1-E2-E3 enzymatic cascade (**Fig. 1**) (Varshavsky, 1997). The UPS degrades an enormous variety of proteins that contain specific degradation signals (degrons). E1 is the ATP-dependent Ub-activating enzyme, which forms a thioester bond between the C-terminal Gly of Ub and a specific Cys of E1. The activated Ub is transesterified to a Cys residue on E2 enzyme. E3 recognizes a substrate's degradation signal (degron) and through complexation with E2, conjugates Ub to the ε-amino group of a Lys

residue of a substrate protein. Subsequent polyubiquitylated substrate is recognized by the 26S proteasome for final degradation (**Fig. 1**).

The N-end rule relates the destabilizing activity of a given N-terminal amino acid and its post-translational modification (Bachmair et al., 1986; Varshavsky, 2011). The N-end rule pathway is a subset of the UPS, where recognition E3 components called N-recognins recognize the N-terminal degradation signals (N-degrons) on the substrate (**Fig. 1)**. The pathway involved in the N-end rule operates in all organisms examined so far, ranging from mammals and plants to bacteria (Gonda et al., 1989; Stary et al., 2003; Tobias et al., 1991). Although there are variations in hierarchical structures and components, N-end rule pathways in eukaryotes typically involve specific recognition components called N-recognins, which recognizes the N-terminal degrons (N-degrons) of substrates and tag them with ubiquitin for subsequent proteasomal degradation (Tasaki and Kwon, 2007). In prokaryotes, which lack ubiquitin and ubiquitin-like proteins, the recognition component binds and delivers target substrates directly to the ClpAP protease complex, a bacterial equivalent of the 26S proteasome (**Fig. 2)** (Mogk et al., 2007; Sriram et al., 2011; Tasaki, 2012).

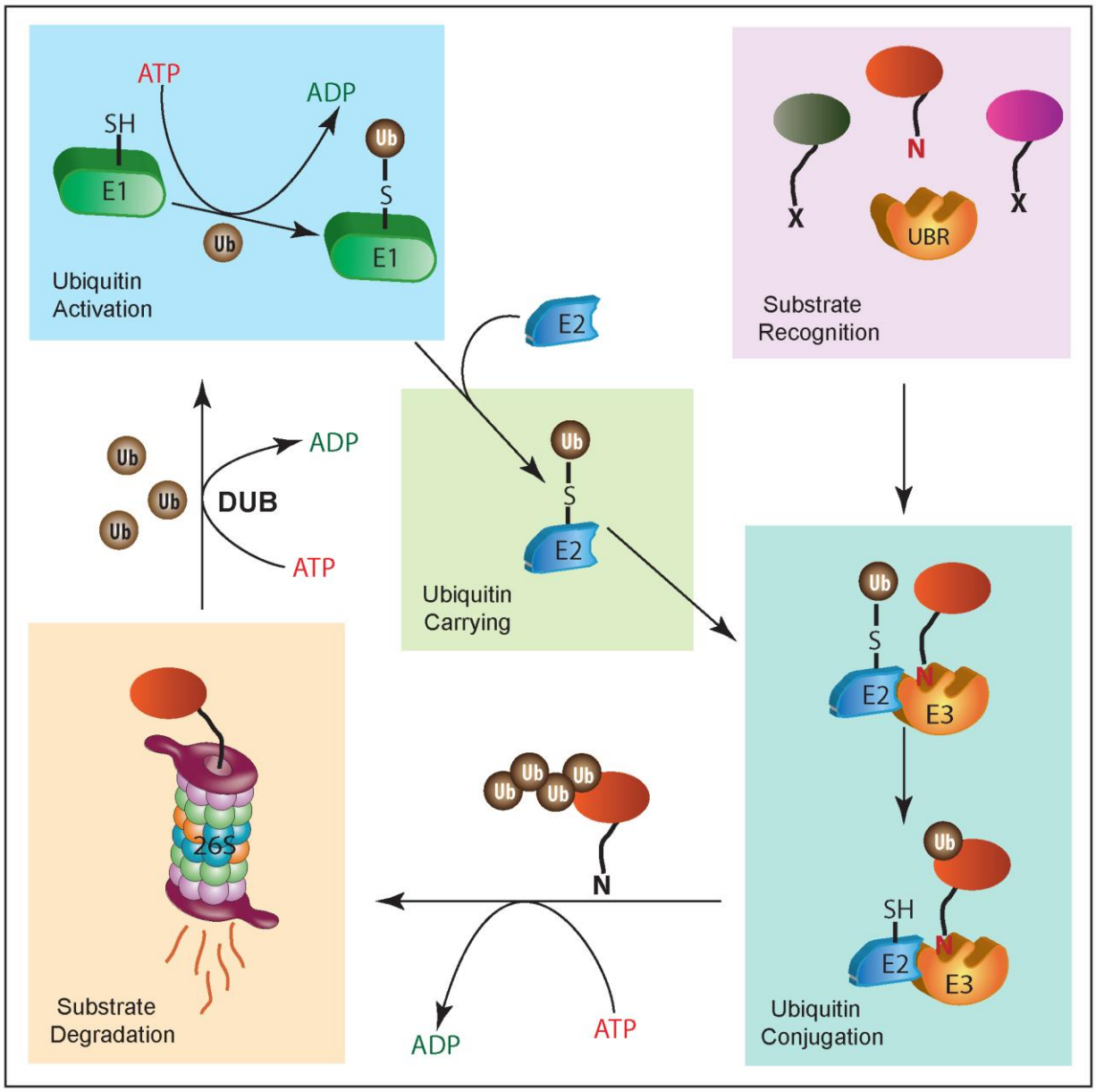

Figure 1. Ubiquitin-dependent Proteolytic System (UPS). The substrates are ubiquitylated through multiple rounds of a linear reaction catalyzed by E1, E2, and E3. Shown, as an example is the N-end rule pathway. E1, Ub activating enzyme; E2, Ub conjugating enzyme; E3, Ub protein ligase; N, N-degron; 26S, 26S proteasome, DUB, deubiquitinating enzyme.

1.1 MAMMALIAN N-END RULE

The complexity of the eukaryotic N-end rule allows a hierarchical structure, where pro-N-degrons (tertiary and secondary destabilizing residues) can be modified to N-degrons (primary destabilizing residues) (**Fig. 2)**. Primary destabilizing residues can be classified as type1 degrons (positively charged residues; Arg, Lys and His) and type2 degrons (bulky hydrophobic residues; Phe, Leu, Trp, Ile and Tyr) (**Table 1)** (Bachmair et al., 1986; Bachmair and Varshavsky, 1989). In mammalian proteins, the tertiary destabilizing residues Asn and Gln are deamidated into Asp and Glu by two distinct N-terminal amidases, NTAN1 and NTAQ1 (Grigoryev et al., 1996; Kwon et al., 2000). The N-terminal Asp and Glu are then conjugated with Arg by R-transferases encoded by *ATE1*(Kwon et al., 1999*)*. In addition to Asn and Gln, an N-terminal Cys functions as a tertiary destabilizing residue in mammalian cells through a two-step modification involving oxidation and arginylation (**Fig. 2)**(Hu et al., 2005; Kwon et al., 2002). Comprehensibly, the degradation of N-end rule substrates bearing the Cys degron, such as RGS4 (regulators of G-protein signaling protein-4), is inhibited by depletion of oxygen or nitric oxide(Lee et al., 2005).

The mammalian genome encodes at least seven UBR box-containing proteins, UBR1 through UBR7 (**Fig. 3**) (Tasaki et al., 2005). UBR1, UBR2, and UBR3 are referred as canonical, owing to their sequelogy, size (about 200 kDa), and conserved domains, including the UBR box, RING finger (ubiquitylation domain), and autoinhibitory domain (which sterically blocks certain regions via intramolecular interaction). UBR4, UBR5, UBR6 and UBR7, are referred as noncanonical UBR box proteins, since they are evolutionarily divergent and nonsequelogous to each other. UBR1, UBR2 and UBR4 can bind both type1 and type2 degrons, whereas UBR5 shows a preference for only type1 N-degrons (Tasaki et al., 2009). On the basis of binding and degradation assays, UBR1, UBR2, UBR4, and UBR5 are classified as N-recognins, and UBR3,

UBR6, and UBR7 as non-N-recognins (**Fig. 4**) (Sriram et al., 2011; Tasaki, 2012). The canonical N-recognins form an E2-E3 complex with the E2 enzymes HR6A and HR6B, which are highly homologous (95% identity) and are functional counterparts of *S. cerevisiae* Ubc2/Rad6 (Tasaki and Kwon, 2007). This complexation leads to polyubiquitylation of the substrates and their subsequent degradation. The known mammalian substrates targeted through N-degrons include RGS4, RGS5, and RGS16 (**Table 2**) (**Fig. 5 & 6**). In addition, N-recognins can target substrates through an internal degron (I-degron) (Byrd et al., 1998). The I-degron is a non-N-degron embedded in the body of the substrate. Although the binding site of I-degrons is not fully characterized, but it is known that this binding site can be blocked through an intramolecular interaction with the C-terminal autoinhibitory domain. And the accessibility through this autoinhibitory domain is regulated by the status of type1 and type2 binding sites. Substrates with I-degrons recognized by N-recognins include c-Fos (a component of the AP1 transcription factor) and histone H2A (An et al., 2010; Sasaki et al., 2006; Tasaki, 2012; Varshavsky, 2011).

Table 1. N-degrons and Modification **(Sriram and Kwon, 2010).**

	1°	2°	3° destabilizing residues
Eukaryotes	Type 1: Arg, Lys, His Type 2: Leu, Phe, Trp, Tyr, Ile		
Mammals		Asp, Glu, oxidized Cys (Arg-ylation)	Asn, Gln (deamidation); Cys (oxidation)
Yeast		Met, Ala, Val, Ser, Thr, Cys (acetylation)	
Escherichia coli	Type 2: Leu, Phe, Trp, Tyr	Arg, Lys (Leu or Phe-ylation)	

1.2 YEAST N-END RULE

In *Saccharomyces cerevisiae* proteins, both Asn and Gln (tertiary destabilizing residues) are deamidated into Asp and Glu (secondary destabilizing residues) by a single N-terminal amidase, Nta1 (Baker and Varshavsky, 1995). The N-terminal Asp and Glu are conjugated with Arg by *Ate1*-encoded Arg^{tRNA}-protein transferase (R-transferase), which creates the primary destabilizing residue Arg at the N-terminus of an otherwise stable protein (Balzi et al., 1990; Li and Pickart, 1995). Interestingly, in contrast to mammals, N-terminal Cys is stabilizing in yeast (**Fig. 2**). The N-terminal Arg and other primary destabilizing residues are recognized and directly bound by the N-recognin Ubr1, a 225-kDa RING finger E3 ligase (Bachmair et al., 1986). Following substrate recognition, Ubr1 mediates substrate polyubiquitination as a complex with the Rad6/Ubc2 E2 conjugating enzyme, leading to its proteasomal degradation (**Fig. 3&4**) (Dohmen et al., 1991; Pickart, 2001).

Known yeast substrates carrying N-degrons include Scc1, a subunit of the cohesin complex (**Table 2**) (Rao et al., 2001), and those harboring internal degrons include the homeodomain protein Cup9 (a transcriptional repressor of the Ptr2 peptide transporter), GPA1 (the Gα subunit that controls signal transduction during mating) and Mgt1 (the O6-alkylguanine-DNA alkyltransferases) (Turner et al., 2000). In addition, Ubr1 has been shown to mediate degradation of misfolded proteins in the cytosol in conjunction with chaperons, such as Hsp70 and Sse1 (**Fig. 5**) (Eisele and Wolf, 2008; Heck et al., 2010).

Interestingly in *S. cerevisiae*, an acetylated N-terminal residue, including the retained initiator Met, can act as an N-degron, thereby functioning as an alternative signal to initiate the N-end rule pathway (Hwang et al., 2010b). The Doa10 E3 ligase (**Fig. 3**), in concert with the Ubc6 or Ubc7 E2 enzymes, functions as a new type of N-recognin that mediates the

polyubiquitylation of acetylated N-degrons (AcN-degrons). Doa10-dependent degradation was verified in eight acetylation-permissive yeast proteins, although it has yet to be determined whether AcN-degron-dependent degradation is functionally relevant to most cellular proteins or just a few. The acetylation-based N-end rule pathway in mammals remains to be characterized (see **4.3**).

Also, a recent study (Hwang et al., 2010a) in yeast showed that Ubr1 and Ufd4, two distinct recognition components for the N-end rule pathway and the UFD (Ub fusion degradation) pathway, respectively, form a complex and synergistically mediate the ubiquitylation for both pathways. In the case of the N-end rule pathway, Ufd4 does not recognize an N-degron but instead increases the efficacy and processivity of Ubr1 only after Ubr1 recognizes an N-degron and initiates ubiquitylation. Thus, Ufd4 acts as an E4-like cofactor to enhance processivity. It has yet to be determined whether the mammalian E3 of the UFD pathway, thyroid receptor-interacting protein 12 (TRIP12), can also act as an E4 for the N-end rule pathway (see **4.3**).

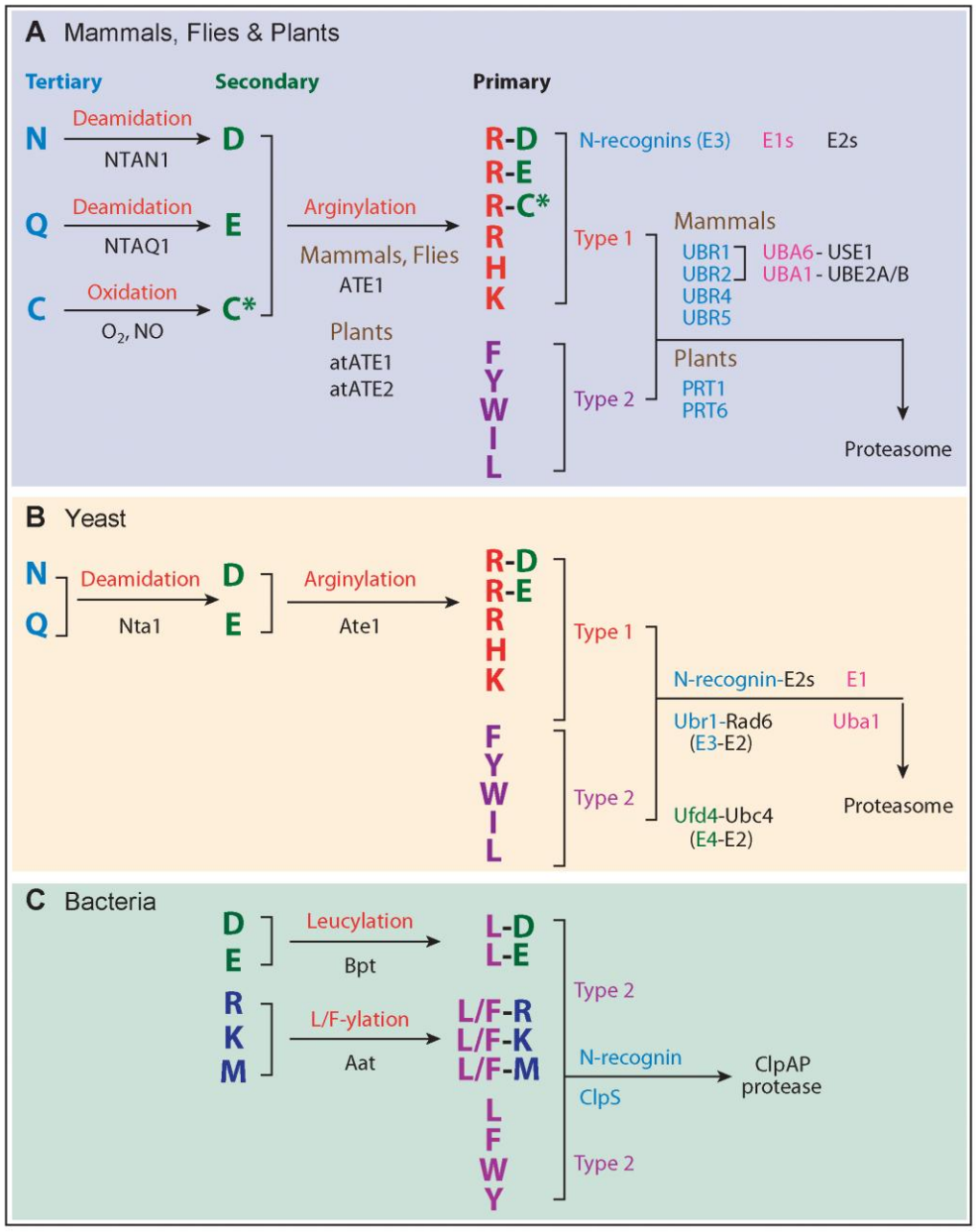

Figure 2. The N-end Rule Pathway. The classical N-end rule pathway in various eukaryotes and prokaryotes. (A) The N-end rule pathway in mammals, flies, and plants. (B) The S. cerevisiae N-end rule pathway. (C) The bacterial N-end rule pathway. Abbreviations: Aat, aminoacyl transferase; Bpt, bacterial protein transferase; C^*, oxidized Cys; E4, Ub conjugation factor; L/F-ylation, leucylation/phenylalanylation; N, asparagine; Q, glutamine; Uba1 and Uba6, Ub activating enzymes 1 and 6; Ubc4, Ub conjugating enzyme 4; Ube2A/B, Ub conjugating enzyme E2 A/B; UBR box, Ub ligase N-recognin box; Ubr1, -2, -4, -5, Ub ligase N-recognin1, -2, -4, -5; Ufd4, Ub fusion degradation 4; Use1, Uba6-specific E2 1 (**Tasaki, 2012**).

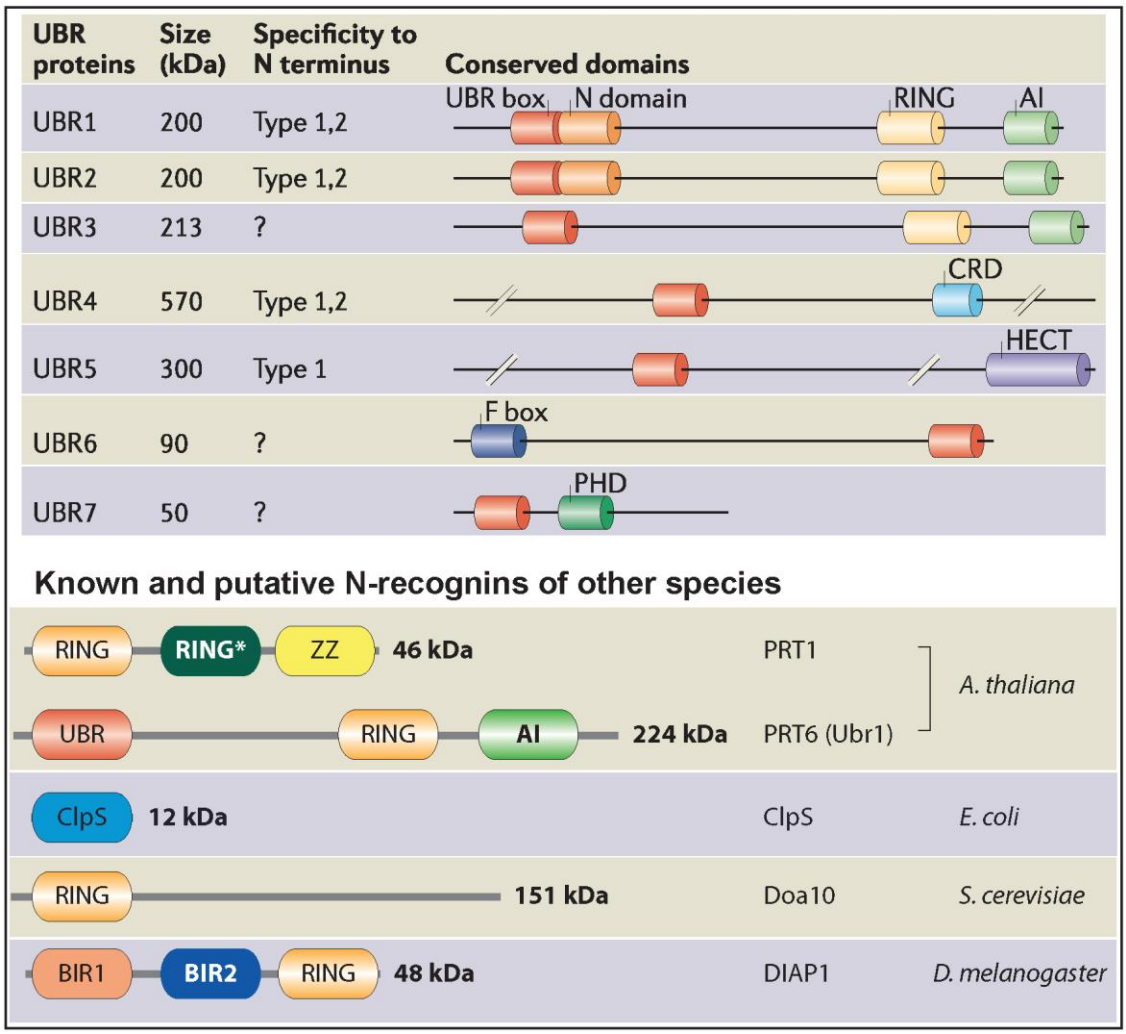

Figure 3. Mammalian UBR family and known and putative N-recognins from other species. A schematic diagram of UBR box proteins. UBR indicates UBR box; RING, RING finger; CRD, cysteine-rich domain; HECT, HECT domain;PHD, plant homeodomain finger; AI, UBR specific autoinhibitory domain. (b) Known and putative N-recognins of other species. Abbreviations: AI, autoinhibitory domain; BIR; baculoviral inhibition of apoptosis protein repeat; ClpS, ATP-dependent Clp protease adaptor protein; CRD, cystein-rich domain; Doa10, the ER-localized ubiquitin ligase Doa10; DIAP1, Drosophila inhibitor of apoptosis 1; PHD, plant homeodomain finger; HECT, homologous to the E6-AP C terminus; N, N domain; PABC, poly(A)-binding protein C-terminal domain; RING∗, composite domain containing RING and CCCH-type Zn fingers; RING, RING finger; UBA domain, Ub association domain; UBR, UBR box; ZZ, a specific zinc finger domain that binds to two zinc ions **(Tasaki, 2012; Sriram et al., 2011)**.

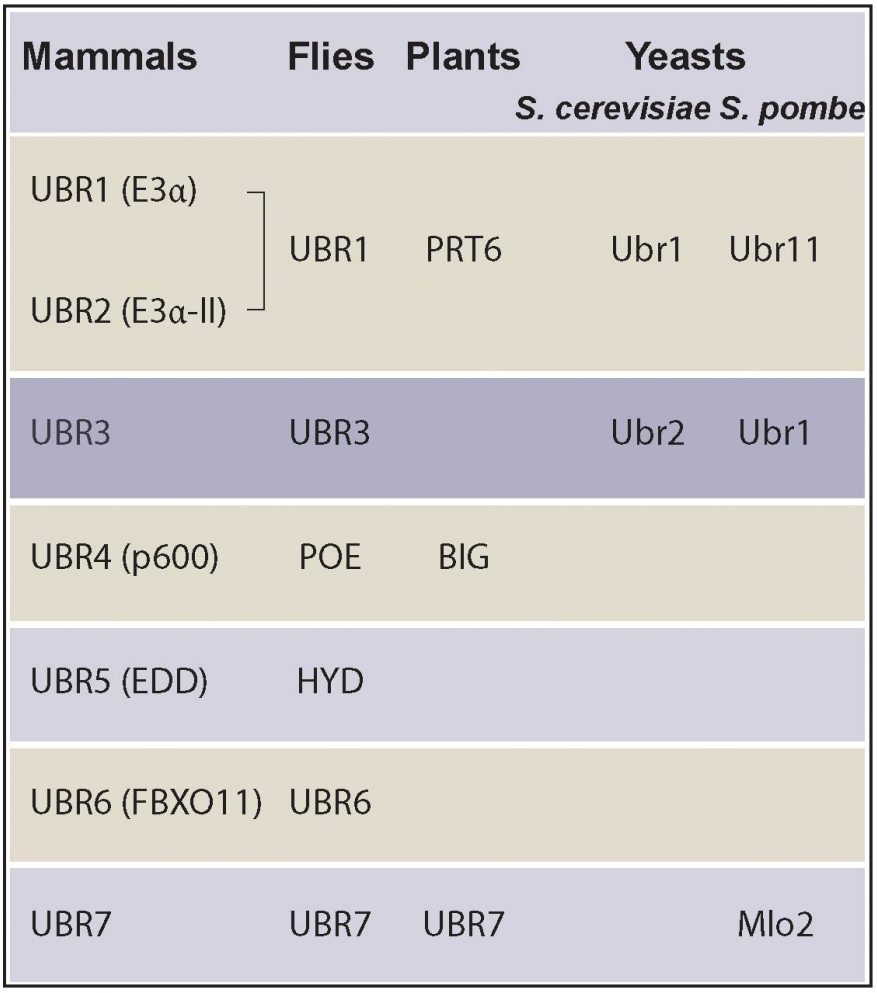

Figure 4. Eukaryotic N-recognins. UBR1 and UBR2 (200 kDa) are functionally overlapping canonical UBR box N-recognins. UBR3 (213 kDa) is a canonical UBR box protein but does not show affinity to N-end rule peptides. Knockout of UBR3 in mice resulted in neonatal death associated with female-specific anosmia, a finding consistent with its unique expression in neural tissues of the so-called five senses (olfaction, hearing, vision, touch, and taste). UBR3 mediates degradation of the DNA repair protein APE1 and is required for genomic stability. UBR3 is sequelogous and thought to be a functional homolog to S. cerevisiae Ubr2 and Schizosaccharomyces pombe Ubr1. The homologs in yeasts are involved in transcriptional regulation of the proteasome (through degradation of the transcriptional activator Rpn4), sexual differentiation, nuclear enrichment of the proteasome (through degradation of the nuclear envelope protein Cut8), and the oxidative stress response. UBR4 is a sequelog of the Arabidopsis BIG, which plays a role in auxin transport, root hair elongation, hormone and light responses, and the regulation of sulfur deficiency-responsive genes. The Drosophila homolog of UBR4, POE/PUSH/CALO, has been implicated as an interactor of calmodulin in the retina and the polar granule molecules Vasa and Tudor in germ plasm from early

embryos; synaptic transmission; perineurial glial growth; male sterility; and meiotic chromosome pairing, recombination, and segregation in females. UBR5 is a sequelog of the Drosophila HYD, whose mutations result in imaginal disc hyperplasia associated with uncontrolled cell proliferation through the independent activation of hedgehog and decapentaplegic. UBR6/FBXO11 (94 kDa), a component of a SCF E3 complex, has been implicated in the neddylation of p53 and the human diseases vitiligo (a skin disorder) and otitis media (a common childhood disease characterized by middle ear inflammation following infection). UBR7 (48 kDa) and its sequelogs in multicellular organisms have a PHD domain, which resembles the RING domain. S. pombe mlo2, a putative UBR7 homolog, is implicated in chromosome transmission fidelity in mitosis **(Tasaki, 2012)**.

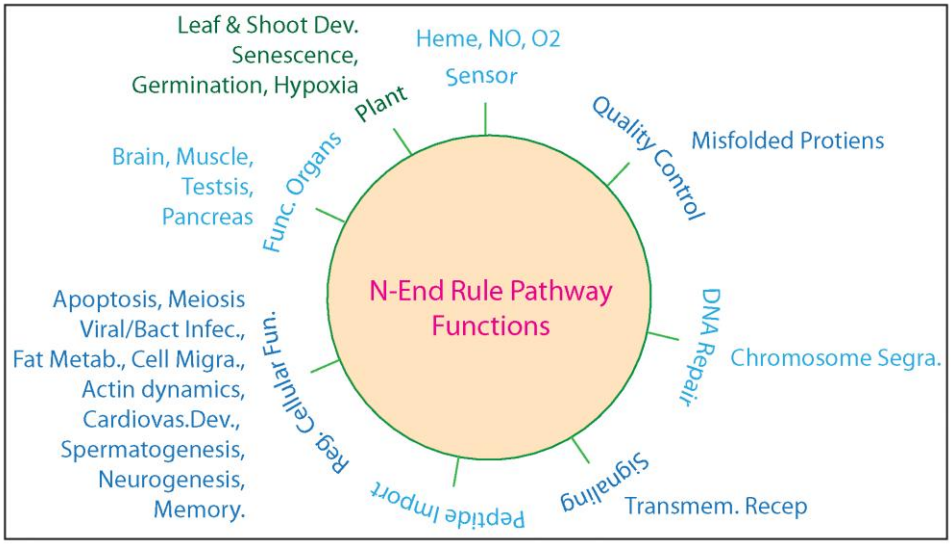

Figure 5. N-end rule pathway functions.

1.3 BACTERIAL N-END RULE

In prokaryotes, ClpS mediates the degradation of N-end rule substrates without ubiquitin-like molecules. ClpS-bound substrates are directly delivered to the ClpAP complex, a ring-shaped

proteolytic machinery which functions like the 26S proteasome (**Fig. 2**) (Erbse et al., 2006; Tasaki, 2012). The bacterial N-recognin ClpS recognizes N-terminal Leu, Phe, Trp and Tyr, among which Leu and Phe can be derived from aminoacyl-tRNA through leucylation or phenylalanylation (**Table 1**)(Ninnis et al., 2009; Tobias et al., 1991). Two classes of aminoacyl transferases are known to mediate leucylation and phenylalanylation in the N-end rule pathway, leucyl/pheylalanyltRNA-protein (L/F)-transferases and leucyltRNA-protein (L)-transferases (**Fig. 2**) (Dougan et al., 2010).

The E. coli L/F-transferase, encoded by aat, transfers Leu or Phe to the acceptors Arg and Lys, which are type1 primary residues in eukaryotes. However, recent identification of the first substrate of the bacterial N-end rule pathway, PATase (putrescine aminotransferase) (Ninnis et al., 2009), led to an unexpected finding that the L/F-transferase conjugates Leu or Phe to the N-terminal Met (not Arg or Lys) of PATase, suggesting that the specificity of L/F-transferase may be broader than previously reported. The second aminoacyl transferase of the bacterial N-end rule pathway is Bpt L-transferase (Graciet et al., 2006). In contrast to the Aat L/Ftransferase, the Bpt L-transferase transfers Leu to N-terminal Asp and Glu, which are arginylation acceptors in eukaryotes. Consistent with R-transferase-like acceptor specificity, the Bpt L-transferase is sequelogous to eukaryotic R-transferases but possesses the donor tRNA specificity of prokaryotic Aat L/F-transferase.

Its physiological substrates include E. coli Dps (DNA protection during starvation) and PATase (YgjG putrescine-aminotransferase) (Ninnis et al., 2009; Schmidt et al., 2009). Dps is an 18-kDa DNA-binding protein that condenses the nuclear material of starving cells to form a highly ordered, stable structure, thus protecting DNA during stress. After the initiator fMet, is cleaved by MetAP, an unknown protease removes the next four residues, exposing Leu6 as a

primary degron for the ClpS-ClpAP protease system (**Table 2**). Interestingly, Dps is also destroyed as a full-length protein by the ClpXP protease that targets the N-terminal segment (Dps1-5), which would be otherwise cleaved off in the process of N-end rule degradation(Dougan et al., 2010). Thus, the N-end rule pathway counteracts the degradation of Dps by ClpXP, functioning as a proteolytic switch that modulates the cellular function of Dps. E. coli PATase, an enzyme involved in the catabolism of putrescine (**Fig. 6**), is targeted by the N-end rule pathway through N^t-leucylation or N^t-phenylalanylation by Aat L/F-transferase. *E. coli* PATase, an enzyme involved in the catabolism of putrescine, is targeted by the N-end rule pathway through N^t-leucylation or N^t-phenylalanylation by Aat L/F-transferase (Ninnis et al., 2009; Schmidt et al., 2009). Interestingly, the attachment of Leu or Phe was not to the pro-N-degron Arg or Lys but to the initiator Met, suggesting that the specificity of L/F-transferase may be broader than previously reported (Ninnis et al., 2009).

1.4 PLANT N-END RULE

The *Arabidopsis thaliana* genome encodes sequelogs of mammalian amidases NTAN1 and NTAQ1, which generate arginylation-permissive pro-N-degrons (**Fig. 2**) (Baker and Varshavsky, 1995).. In contrast to mammalian N-end rule, arginylation in *A. thaliana* is mediated by two distinct R-transferases, ATE1 and ATE2, which are encoded by separate genes. To date, two plant N-recognins, PROTEOLYSIS 1 (PRT1) and PRT6, have been identified (**Table 2**) (Bachmair et al., 1993; Garzón et al., 2007).

PRT1 is a 45-kDa protein with two RING domains and one ZZ domain. The mutation of *PRT1* impaired the degradation of model N-end rule substrate bearing aromatic hydrophobic

residues (Phe, Trp and Tyr) at their N-termini, but not aliphatic hydrophobic residues, such as Leu and Ile (Potuschak et al., 1998). As PRT1 is not a sequelog of mammalian N-recognins and has neither of the UBR box nor the N-domain, it remains to be determined how this divergent N-recognin in plants recognizes aromatic hydrophobic residues. PRT6 is a 224-kDa canonical N-recognin that contains the UBR box but lacks the N-domain (**Fig. 3&4**) (Garzón et al., 2007).

Arabidopsis prt6 mutants are impaired in degradation of N-end rule substrates bearing type1 degrons, such as Arg, but not type2 degrons, indicating discrete substrate selectivity for PRT1 and PRT6. Although mutants of *prt1* and *prt6* are defective in targeting degrons with aromatic and basic side chains, respectively, these mutants retain virtually intact activities for Leu and Ile, indicating the presence of an addition N-recognin. Such candidates include two noncanonical UBR proteins (UBR4/BIG and UBR7) and ClpT, a sequelog of E. coli ClpS, which is predicted to localize in chloroplasts (**Fig. 4**) (Peltier et al., 2004).

Recently it was reported that the N-end rule pathway of the plant Arabidopsis functions as an oxygen sensor similar to mammalian N-end rule through regulated proteolysis of the hypoxia-sensitive transcription factor family carrying the pro-N-degron Cys2 (**Fig. 6**) (Gibbs DJ, 2011; Licausi F, 2011). In normoxia, the ethylene response factor group VII transcription factors, including hypoxia-responsive element 1 and 2 (HRE1 and HRE2) and related to AP2.12 (RAP2.12), are downregulated through proteasomal degradation in a manner depending on the pro-N-degron Cys2. In hypoxia, however, these hypoxia-sensitive transcription factors are accumulated, resulting in transcriptional induction of genes that promote anaerobic metabolism and survival of hypoxia (**Fig. 6**). As hypoxia-inducible factor-1 (HIF-1), a known oxygen sensor in animals, is absent in plants, the Cys branch of the N-end rule pathway may represent an oxygen-sensing mechanism in plants (Gibbs DJ, 2011; Licausi F, 2011). The Arabidopsis and

human genomes encode at least 206 and 502 proteins, respectively, with the Met-Cys motif (**Fig. 6**). Thus, these Met-Cys proteins may represent a unique proteome, whose functions include sensing oxygen and other cellular stresses through oxidation and arginylation of the pro-N-degron Cys.

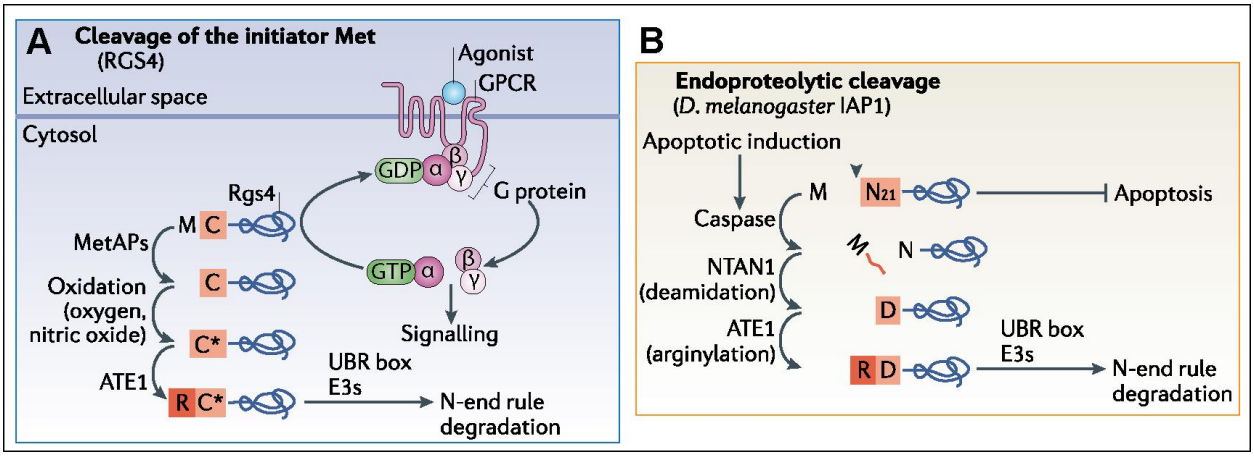

Figure 6. Examples of the N-end rule in physiology. (A) The activation of an amino-terminal degron (N-degron) of regulator of guanine-nucleotide-binding protein (G) protein signalling 4 (RGS4) through the cleavage of N-terminal Met. The N-terminally exposed Cys2 of RGS4 undergoes oxidation into CysO2 or CysO3, which can be inhibited by depletion of oxygen or nitric oxide37–39. The oxidized Cys2 (C*) is structurally similar to an arginylation-permissive residue, Asp, and arginylated by arginyl-tRNA–protein transferase 1 (ATE1; a member of the R-transferase family). The resulting type 1 degron, Arg, is targeted by UBR box-containing E3 ligases. (B) The activation of an N-degron of *Drosophila melanogaster* IAP1 (DIAP1) through an endoproteolytic cleavage. DIAP1 normally inhibits initiator and effector caspases. Upon apoptotic induction, the effector caspases, such as ICE and DCP1, cleave the N-terminal 20-residue fragment, exposing an N-degron, Asp21, which subsequently enters N-end rule modifications: deamidation by the N-terminal amidase NTAN1 and arginylation by ATE1 (**Sriram et al., 2011**).

1.5 DROSOPHILA N-END RULE

The *Drosophila melanogaster* has a similar hierarchical structure and components as the mammalian N-end rule (**Fig. 2**) (Baker and Varshavsky, 1995). In *Drosophila*, Ntan1 mediates the degradation of an antiapoptotic regulator, *Drosophila* inhibitor of apoptosis 1 (DIAP1) that prevents unwanted cell death by apoptosis (**Table 2**) (Ditzel et al., 2003). When apoptosis is activated, the effector caspase DrICE or DCP-1 cleaves DIAP1 after Asp at position 20, thus creating a C-terminal fragment harboring the tertiary destabilizing residue Asn (**Fig. 6**). This exposed Asn undergoes sequential N-end rule modifications, including deamidation by Ntan1, arginylation by Ate1, and ubiquitination by UBR box E3 ligases (**Fig. 3&4**) (Ditzel et al., 2008; Ditzel and Meier, 2005; Ditzel et al., 2003). The apoptosis inhibitors DIAP1 and DIAP2 and their mammalian homologs (XIAP, cIAPs, and ML-IAP) are RING E3s that suppress undesirable apoptotic activities by inhibiting the functions of initiator and effecter caspases (**Fig. 5**) (Orme and Meier, 2009).

Table 2. Substrates of the N-end rule pathway **(Sriram et al., 2011)**

Substrate	Species	Terminal sequence	Pro-N-degron	N-degron	Modifications
Scc1 cohesin subunit	Saccharomyces cerevisiae	RLGESIMSEE	–	Arg269 type 1	Endoproteolytic cleavage by separase
RAD21 (Scc1 homologue) cohesin subunit	Homo sapiens	EGSAFEDDD	Glu173	Arg* type 1	Endoproteolytic cleavage by separase, arginylation
IAP1	Drosophila melanogaster	NNTNATQLFK	Asn21	Arg* type 1	Endoproteolytic cleavage by caspase, deamidation, arginylation
GNG2	Bos taurus	NTASIAQARK	Asn5	Arg* type 1	MetAP, protease, deamidation, arginylation
RGS4 GAP	Mus musculus	CKGLAGLPAS	Cys2	Arg* type 1	N terminal Met cleavage by MetAPs, oxidation, arginylation
RGS5 GAP	Mus musculus	CKGLAALPHS	Cys2	Arg* type 1	N terminal Met cleavage by MetAPs, oxidation, arginylation
RGS16 GAP	Mus musculus	CRTLATFPNT	Cys2	Arg* type 1	N terminal Met cleavage by MetAPs, oxidation, arginylation
Sindvis virus nsP4 RNA polymerase	Alphavirus	YIFSTDTGPG	–	Tyr189 type 2	Endoproteolytic cleavage of proprotein
HIV-1 integrase DNA intergrase	Lentivirus	FLDGIDKAQE	–	Phe1 type 2	Endoproteolytic cleavage of proprotein
LLO pore-forming cytolysin	Listeria monocytogenes	KDASAFHKED	–	Lys25 type 1	Bacterial signal peptidase during secretion
PATase	Escherichia coli	MNRLPSSASA	Met1	Leu* or Phe*	Leucylation or phenylation by Leu- or Phe transferase
Dps	Escherichia coli	LVKSKATNLL	–	Leu6	Endoproteolytic cleavage of proprotein

Dps, DNA protection during starvation; GAP, GTPase-activating protein; GNG2, G protein subunit γ2; IAP1, inhibitor of apoptosis; LLO, listeriolysin O; MetAP, Met aminopeptidase; nsP4, non-structural protein 4; PATase, putrescine aminotransferase; RGS, regulator of G protein signalling; Scc1, sister chromatid cohesion 1. *Residues that have been transferred from aminoacyl-tRNA to pro-N-degrons by aminoacyl tRNA transferases.

2.0 STRUCTURAL BASIS OF THE N-END RULE RECOGNITION

Our lab had previously established a family of N-recognins identified by a ~70-residue zinc-finger domain, termed the UBR box (**Fig. 3 & 4**). In eukaryotes, N-recognins have two distinct substrate recognition domains: the UBR box for positively charged type1 N-degrons (Arg, Lys, and His) and the N-domain for bulky hydrophobic type2 N-degrons (Phe, Tyr, Trp, Leu, and Ile) (**Table 1**) (Tasaki et al., 2005; Tasaki et al., 2009). In bacteria, which lack the UPS, the N-end rule pathway mediates the degradation of substrates bearing Trp, Phe, Tyr, or Leu N-degrons through the recognition by ClpS (**Fig. 2**) (**Table 1**). UBR box and N-domain share a high level of homology between different organisms, highlighting the conserved mechanism of structural recognition (**Fig. 7**) (Tasaki and Kwon, 2007; Tasaki et al., 2009). Recently determined crystal structures of UBR boxes and ClpS provide new insights into molecular principles of N-end rule interactions for type1 and type2 degrons (Choi et al., 2010; Matta-Camacho et al., 2010; Roman-Hernandez et al., 2009). The structural comparison of the UBR box, together with the structures of ClpS, led to the hypothesis that eukaryotic type1 domains and prokaryotic type2 N-recognins adopt different folding strategies to accommodate a variety of amino acids that vary in size and shape. We have found that despite the fundamental structural differences and a billion years of evolutionary distance, the UBR box and ClpS still share a remarkable similarity in the molecular principles of substrate recognition (Sriram and Kwon, 2010).

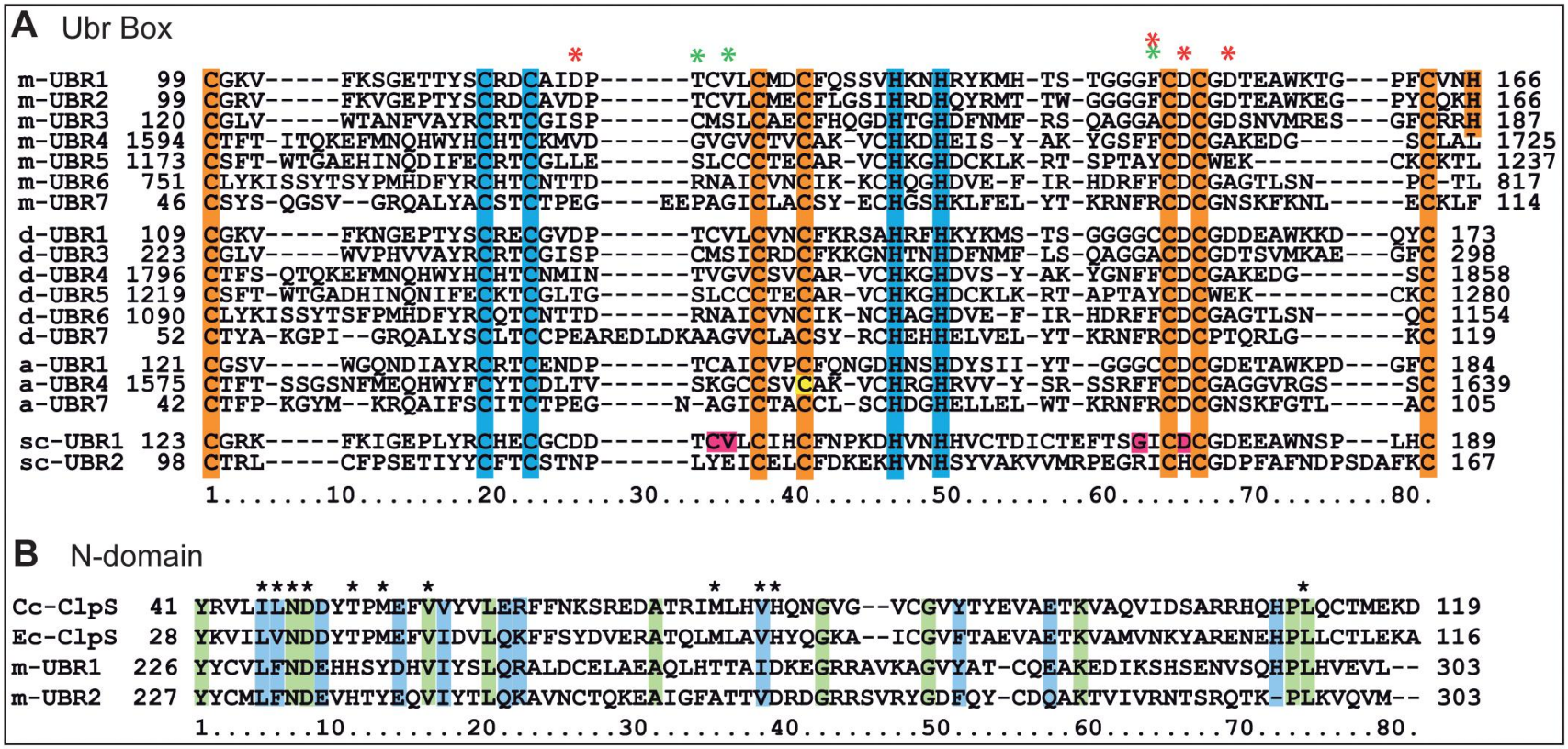

Figure 7. Sequence homology of UBR box and N-domain. (A) Sequence alignment of UBR boxes with the zinc coordinating histidine and cysteine residues. Highlighted blue are the residues (Cys112, Cys115, His133, and His 136) of human UBR2 UBR box, which forms the typical C2H2 zinc finger. Highlighted yellow are those (Cys99, Cys124, Cys127, Cys149, Cys151, His163, and His 166) involved in the atypical zinc finger. Cys127 is shared by two zinc ions. The N-terminal amino acid makes contacts with Asp118, Asp150, Asp153, and Phe148 (red asterisks). The N-terminal α-amino group interacts with Phe148 and Asp150, while the N-terminal side chain with Asp153 and Asp150 and, indirectly, with Asp118 through water bridge. The peptide bond interacts with Phe148 (-C=O) and Thr120 (-NH), and with Val122 through hydrophobic interaction (green asterisks). The second residue forms hydrogen bonding with Thr120, Val122,

and Phe148 (green asterisks). Indicated by yellow highlight is the Cys residue of *Arabidopsis* BIG whose missence mutation perturbs auxin transport. Indicated by pink highlight are the residues of *S. cerevisiae* Ubr1, essential for degradation of type 1 N-end rule substrates. Note that different organisms contain different set of UBR box proteins. m, *Mus musculus*, d, *Drosophila melanogaster*, a, *Arabidopsis thaliana*, sc, *Saccharomyces cerevisiae*. (B) Sequence alignment of N-domains of mouse UBR1 and UBR2 with *C. crescentus* and *E. coli* ClpS N-recognins. The type-2 recognition domains of UBRN-recognins and Clps proteins share significant homology in secondary structure and substrate specificity to type-2 N-degrons. Asterisks indicate the residues of *C. crescentus* ClpS in contact with N-degrons. Blue and green highlights indicate residues partially and completely conserved, respectively **(Sriram et al., 2011).**

2.1 STRUCTURAL BASIS OF TYPE 1 RECOGNITION

Yeast Ubr1 and human UBR1 and UBR2 unbound and bound crystal structures have been recently solved (Choi et al., 2010; Matta-Camacho et al., 2010). It was reported that these N-recognins bind type1 residues through a shallow binding groove, which is supported by two zinc fingers of the UBR box. The UBR box forms a unique folding characterized by two contiguous zinc fingers (**Fig. 8**). The first zinc finger is a typical Cys_2His_2 motif containing one zinc ion, and the second one is an atypical binuclear Cys_6His_1 motif in which two zinc ions are shared by a common Cys for tetrahedral coordination. These zinc coordination residues are well conserved across UBR family members, including non-N-recognins, suggesting that the UBR box may have yet other function(s) outside N-end rule recognition, possibly, as a site of redox modification (see **4.2**) (Sriram and Kwon, 2010).

Substrate binding residues of the UBR box localize near the shallow cleft, with negatively charged residues enriched on the surface to recognize the basic N-terminal side chain. The UBR box binds to a degron through specific interactions with the protonated α-amino group of the first residue, the side chains of the first and second residues and the backbone atoms of the first three residues (**Fig. 9**) (Choi et al., 2010; Matta-Camacho et al., 2010). These interactions are established through hydrogen bonding, with some contribution by non-covalent, electrostatic interactions between the positively charged N-terminal side chain and the negatively charged Asp residue (Asp179 in *S. cerevisiae* and Asp153 in humans) of the UBR box. This electrostatic interaction, together with hydrogen bonding, forms a salt bridge between the N-terminal side chain and Asp179 (or Asp153). In addition, the protonated α-amino group forms three conserved

hydrogen bonds, two with *S. cerevisiae* Ubr Asp176 (Asp150 in humans) and one with *S. cerevisiae* Ile174 (Phe148 in humans) (**Fig. 9**) (Choi et al., 2010; Matta-Camacho et al., 2010; Sriram and Kwon, 2010).

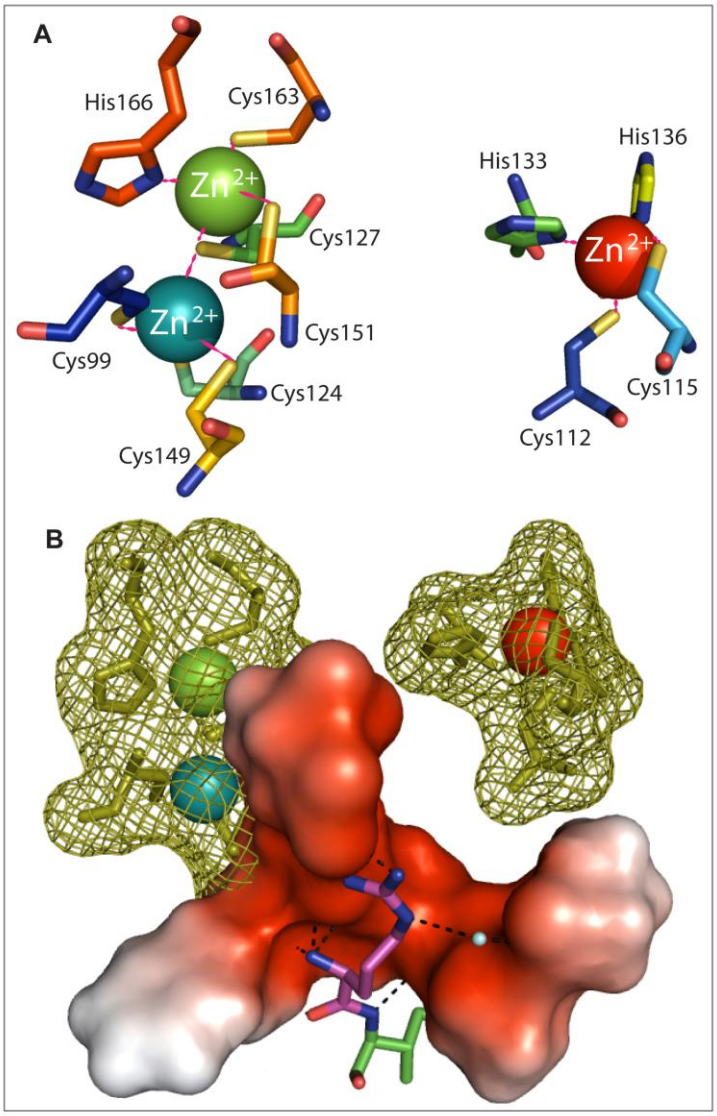

Figure 8. Zinc finger in UBR box. (A) The structure of the zinc-fingers of the human UBR1 (ubiquitin ligase N-recognin 1) UBR box in coordination with zinc ions. The conserved Cys and His residues are shown in a typical motif (CX2CX20HX2H) containing a single zinc ion (red sphere) and an atypical motif (CX24CX2CX21CXCX11CX2H) containing two zinc ions (green and teal spheres) in coordination with Cys127. B (B) A schematic of the zinc-fingers, which stabilize the substrate-binding pocket. The electrostatic potential surface of the residues (Asp118, Thr120, Val122, Phe148, Asp150 and Asp153) that interact with the Arg-Ile peptide through hydrogen bonds (black dotted lines) is shown, water molecule is shown as blue sphere. Behind the substrate-binding pocket, the surfaces of the two zinc-fingers are shown as meshes. Note that the atypical zinc-finger is in a direct contact with the binding pocket, whereas the typical zinc-finger is away from it **(Sriram et al., 2011)**.

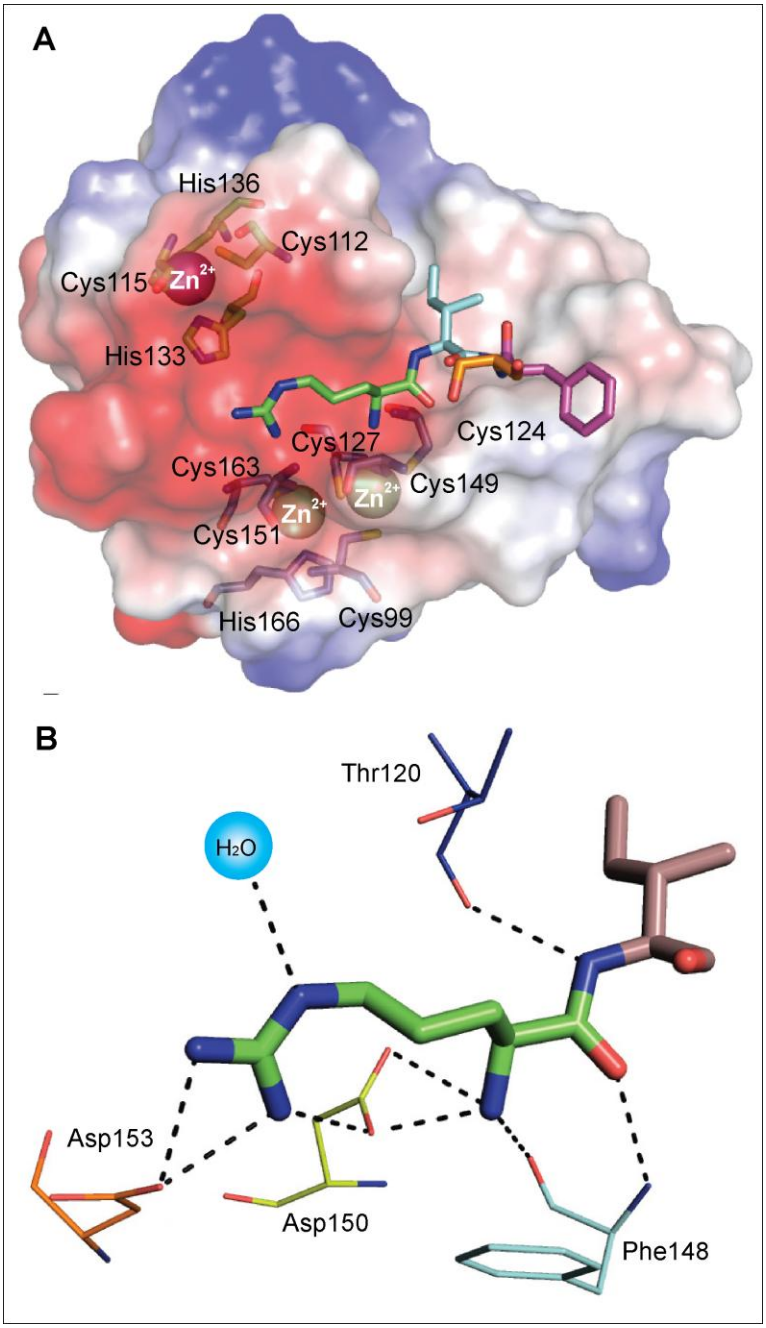

Figure 9. Structural basis of type1 binding. (A) The electrostatic potential of the UBR2 UBR box bound with Arg-Ile (Protein Databank (PDB) code: 3NY3). The amino-terminal residue binds to a negatively charged, shallow binding groove, and the second residue binds to a hydrophobic pocket. Surface colours: red, negative; white, hydrophobic; blue, positive. (B) Key hydrogen bonds (dotted lines) of UBR box residues (line representation) with Arg-Ile (stick representation). Atoms are coded by colour: Red, oxygen; blue, nitrogen **(Sriram et al., 2011)**.

2.2 STRUCTURAL BASIS OF TYPE 2 RECOGNITION

Structures of *Caulobacter crescentus* and *E. coli* ClpS unbound or in a complex with either a peptide or ClpA have been previously solved (Dougan et al., 2010; Erbse et al., 2006; Wang et al., 2008b). This monomeric, cone-shaped protein is composed of a short N-terminal region and two C-terminal regions, one for interacting with ClpA of the ClpA/P protease complex and the other to form a hydrophobic binding pocket for destabilizing N-terminal residues (**Fig. 10**). *E.coli* and *C. crescentus* ClpS have similar structures (root-mean-square deviation (RMSD) = 0.9 Å) and bind their substrates through structurally conserved hydrophobic pockets that are preformed, since there was no significant change in this pocket induced by binding to a ligand (Roman-Hernandez et al., 2009). The pockets are relatively small with a volume of ~200 $(Å)^3$, but remarkably it shows relatively high binding affinity of low μM dissociation constant to degrons (Schmidt et al., 2009).

The ClpS binds to different substrates in a similar manner through a network of hydrogen bonding with the backbone atoms of the first two residues of the substrates. In these interactions, the α-amino group and the N-terminal side chain serve as the principal recognition determinant. *C. crescentus* ClpS forms three hydrogen bonds with the α-amino group of the N-degron of a bacterial model substrate, two with His79 and Asn47 directly and the third through a water molecule that also forms a hydrogen bond to Asp49, which are conserved in UBR boxes (**Fig. 10**) (Dougan et al., 2010; Roman-Hernandez et al., 2009).

In ClpS, the substrate specificity is further stabilized by its substrate binding pocket, whose fine adjustment in size and shape govern the accessibility to the substrate (Sriram and Kwon, 2010). ClpS excludes β-branched Ile, Thr and Val hydrophobic, non-N-degrons from

other hydrophobic Leu, Phe, Tyr and Trp N-degrons. This difference is contributed by the gatekeeper Met at the entrance of the substrate binding site, whose side chain will clash with the β-branched side chain of Ile, Thr and Val but not with those of Leu, Trp and Phe (Guo et al., 2002; Roman-Hernandez et al., 2009; Schmidt et al., 2009). This was supported by mutating this Met gatekeeper to Ala with has a smaller side chain. Through his mutation *E. coli* and *C. crescentus* ClpS can be engineered to acquire a new specificity to Ile while retaining the ability to recognize specific features of N-terminal amino acids (Roman-Hernandez et al., 2009; Wang et al., 2008b). In eukaryotic N-recognins, this Met gatekeeper is replaced by a highly conserved Tyr, which excludes Val but not the Ile, Leu, Phe, Tyr and Trp degrons, explaining why mammalian N-recognins can recognize Ile as a type2 substrate (**Table 1**) (Sriram and Kwon, 2010).

In the genetic code, the default N-terminal residue is Met in eukaryotes and fMet in prokaryotes. One remaining question is how N-recognins exclude the straight chain Met but not its structural cousin, Leu, with a branched side chain. Recent structural modeling has show that the Cε methyl group of Met, which corresponds to the Cδ1 methyl group of Leu, occupies an unfavorable position, creating an adverse change in its van der Waals packing (Dougan et al., 2010; Roman-Hernandez et al., 2009). Modeling studies have also predicted the existence of a Met rotamer that can bind to the pocket without significant steric clash, but this N-recognin-acceptable rotamer represents only about 2% of the whole Met residues present in structural database as opposed to 59% for Leu rotamer bound to ClpS. Moreover, the affinity of this rotamer is about 30-fold lower compared to Leu due to the side-chain rotamer entropic penalty (Leu = 0.3 kcal/mol and Met = 2.4 kcal/mol). Together, the unfavorable van der Waals packing

combined with side-chain entropy penalty results in 1,000-fold lowered binding affinity for Met compared to Leu (Wang et al., 2008a; Wang et al., 2008b).

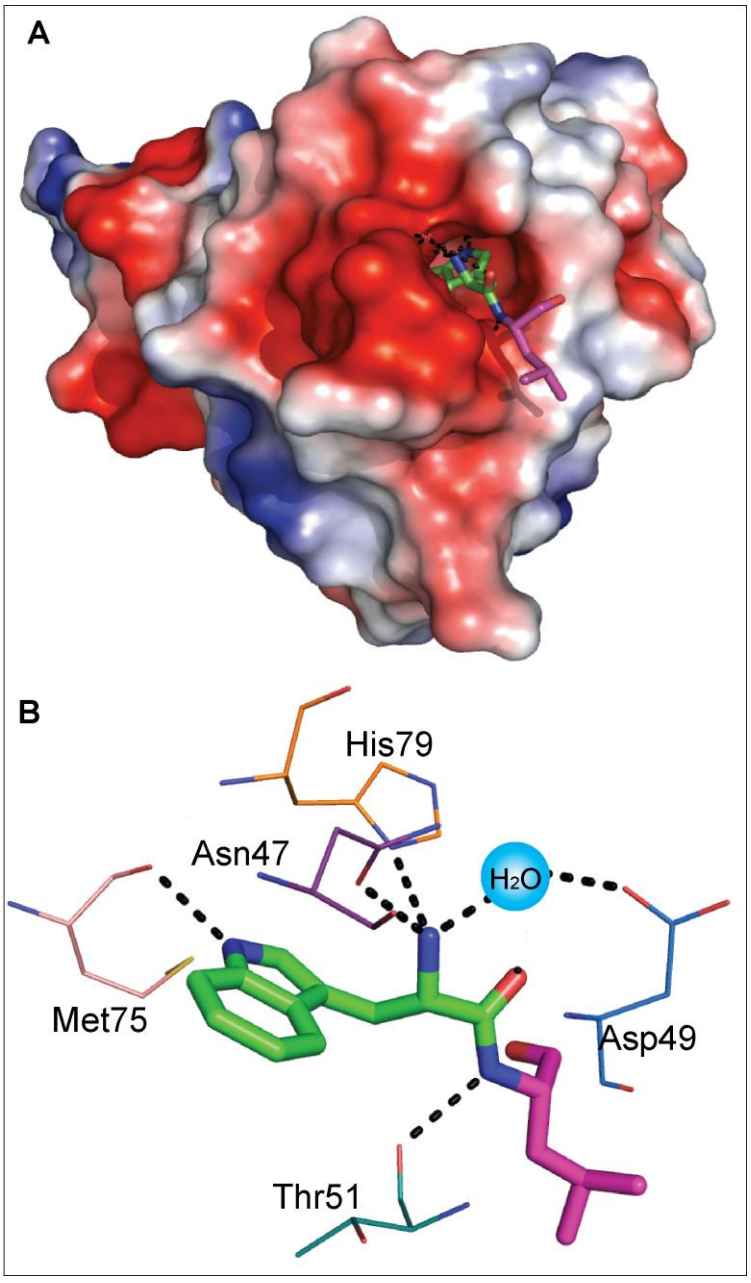

Figure 10. Structural basis of type2 recognition. (A) The electrostatic potential of *Caulobacter crescentus* ClpS bound with Trp-Leu (PDB code: 3GQ1). The hydrophobic N-terminal residue binds to a deep hydrophobic pocket,

and the interaction of the second residue occurs outside of the pocket. (B) Key hydrogen bonds (dotted lines) of ClpS residues (line representation) with Trp-Leu (stick representation) **(Sriram et al., 2011)**.

2.3 MOLECULAR PRINCIPLES OF THE N-END RULE RECOGNITION

The recent structures of the UBR box, together with the previous structures of ClpS (and, thus, the N-domain in eukaryotes), suggest that eukaryotic type1 domains and prokaryotic type2 N-recognins adopt different folding strategies to accommodate a variety of amino acids differing in size and shape (Sriram and Kwon, 2010). Specifically, the UBR boxes of mammalian and yeast N-recognins interact with type1 peptides on the surface of a relatively shallow, acidic groove (**Fig. 9**). This substrate interaction is mechanistically distinct from the ClpS, in which the hydrophobic N-terminal side chain of a type2 peptide is deeply buried in a hydrophobic pocket through substrate-selective interactions (**Fig. 10**). Despite the differences in structure and after being separated by a billion years of evolutionary distance, the UBR box and ClpS, however share a remarkable similarity in their molecular principle of substrate recognition (**Fig. 11**) (Sriram and Kwon, 2010). For instance, both the UBR box and ClpS recognize the free α-amino group of the N-terminal amino acid via three highly conserved, rigidly positioned hydrogen bonds.

Given that the N-terminal α-amino group is present in all native proteins, this weak and transient interaction is not substrate-selective but an essential entry step in N-end rule selection, which might enable the N-recognin to rapidly scan a large pool of the protein N-termini. During this scanning, once the N-recognin is engaged with a genuine substrate, the interaction with the

α-amino group is stabilized by substrate selective interactions with the side chain of a basic residue (the UBR box) or a bulky hydrophobic residue (ClpS), which is further stabilized by hydrogen bonds with the first peptide bond and the side chain of the second residue. In all cases, the side chain of the third residue does not participate considerably in N-end rule interactions, i.e., it either turns away from the surface of the UBR box or is excluded from the ClpS hydrophobic pocket (**Fig. 11**). We believe that by restricting the major interactions to the first two amino acids, N-recognin can minimize non-N-end rule interactions with the rest part of the protein.

N-recognins achieve high processivity and selectivity to the destabilizing N-terminal residues of a substrate through a multi-step strategy, which is similar in eukaryotes and bacteria. One critical step in both type1 and type2 interactions is the formation of hydrogen bonds with the free α-amino group of the N-terminal residue (Sriram and Kwon, 2010). This weak, transient and reversible interaction enables the N-recognin to rapidly scan the N-termini of proteins, including those that are being translated as in the case of cotranslational degradation. The second, substrate-selective step is the formation of hydrogen bonds with the N-terminal side chain. This involves stabilization with the positively charged side chains in type1 degrons and the bulk hydrophobic side chains in type2 degrons. In a type1 interaction, these hydrogen bonds are further supported by the electrostatic interactions between the negatively charged surface of the UBR box and a positively charged N-terminal side chain, forming a salt bridge (Sriram and Kwon, 2010). Likewise, in a type2 interaction, the hydrophobicity of the N-domain pocket provides the hydrophobic interaction with the bulky hydrophobic N-terminal side chain.

Once the N-recognin binds the N-terminal residue, the binding is further stabilized by additional hydrogen bonds with the first peptide bond and the side chain of the second residue.

While the first two amino acids are engaged with a substrate-binding site, the side chain at position 3 stays away from the surface of the type1 binding groove or is excluded from the type2 hydrophobic pocket. Therefore, by restricting the major interactions to the first two residues, N-recognins can select and bind to its substrate based on the identity of a substrate's N-terminus (**Fig. 11**) (Sriram and Kwon, 2010).

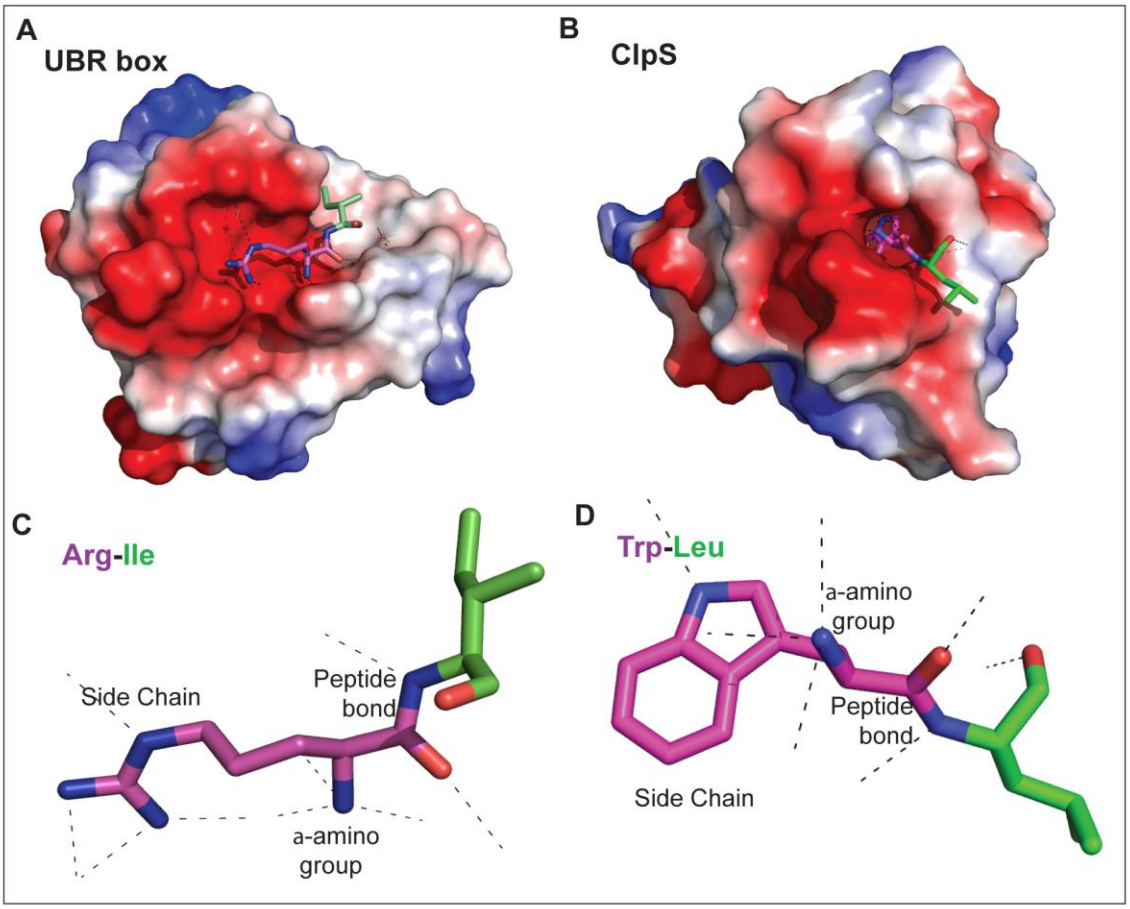

Figure 11. Molecular Principles of substrate recognition. (A,B) Electrostatic potential of UBR2 UBR box bound with Arg-Ile (PDB: 3NY3) and Caulobacter crescentus ClpS bound with Trp-Leu (PDB: 3GQ1). Surface color: red, negative; white, hydrophobic; blue, positive. (C,D) Key hydrogen bonds of N-degron dipeptides, Arg-Ile bound to the Ubr box (C) and Trp-Leu bound to ClpS (D). Blue, nitrogen; red, oxygen. Note the characteristic interactions from, 1) N-terminal side-chain, 2) α-amino group, and 3) the peptide link **(Sriram and Kwon, 2010).**

2.4 METHODS

Molecular Principle Graphics: X-ray crystallographic structures were obtained from the Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do) and the graphics were generated by using MacPyMol three-dimensional molecular visualization program (http://www.pymol.org/).

Structure prediction: The open access Robetta server (http://robetta.bakerlab.org/) was employed for obtaining the predicted structures. Robetta provides both *ab initio* and comparative models of protein domains. Domains without a detectable PDB homolog are modeled with the Rosetta de novo. Comparative models are built from Parent PDBs detected by UW-PDB-BLAST or HHSEARCH and aligned by various methods, which include HHSEARCH, Compass, and Promals. Loop regions are assembled from fragments and optimized to fit the aligned template structure. The procedure is fully automated.

3.0 RATIONAL DESIGN OF INHIBITORS TARGETING THE N-END RULE

3.1 MULTIVALENT INTERACTIONS

Nature employs multivalent interactions to exploit the inherent multivalency-assisted enhancement thereby increasing the selectivity and avidity of protein/protein or protein/ligand interactions in various processes, such as antigen-antibody, virus-cell and bacterial toxin-cell interactions (each with a K_d ~10^{-10} M) (**Fig. 12**)(Sriram et al., 2009); (Basha et al., 2006; Choi, 2004; Huskens, 2006; Kiessling et al., 2000, 2006). Examples of natural multivalent molecules include the trimeric hemagglutinin complex of the influenza virus that recognizes host cells through multivalent binding to N-acetyl neuraminic acid. The enhancement, often dramatic, in selectivity and avidity of multivalent interaction is manifested by synthetic multivalent sialic acid molecules capable of binding to the hemagglutinin receptor on the viral surface with a multivalent enhancement factor of greater than 10^7 (Choi et al., 1996). As such, natural and synthetic multivalent interactions have been extensively investigated to explain the basis of multivalency and in an attempt to inhibit undesired ligand-receptor interactions or to induce desired biological responses (Sriram et al., 2009). Various synthetic multivalent compounds were proven to be able to efficiently control physiological processes in different contexts, including receptor clustering, receptor selectivity, bacterial toxins, pathogen-cell adhesion, and protein-protein interaction (Choi, 2004; Sriram et al., 2009).

Most of multivalent molecules synthesized to date are interhomovalent in that two identical ligands target the same binding site of two identical proteins on the surface of viruses, bacteria, or cells (**Fig. 12**) (Choi, 2004). In contrast, rapamycin, an immunosuppressant drug produced from the bacterium *Streptomyces hygroscopicus*, is an interheterodivalent compound that can simultaneously bind two cytoplasmic proteins, FKBP12 (FK506 binding protein) and FRB (FKBP-rapamycin binding domain), to form the FKBP-rapamycin-FRB ternary complex (**Fig. 12**) (Sabatini et al., 1994). Some synthetic rapamycin derivatives were demonstrated to alter various intracellular pathways, including protein relocalization, conditional induction of apoptosis, protein degradation, and conditional protein splicing (Sriram et al., 2009). We hypothesized that similar heterodivalent interactions can be exploited in the N-end rule pathway targeting the UBR proteins.

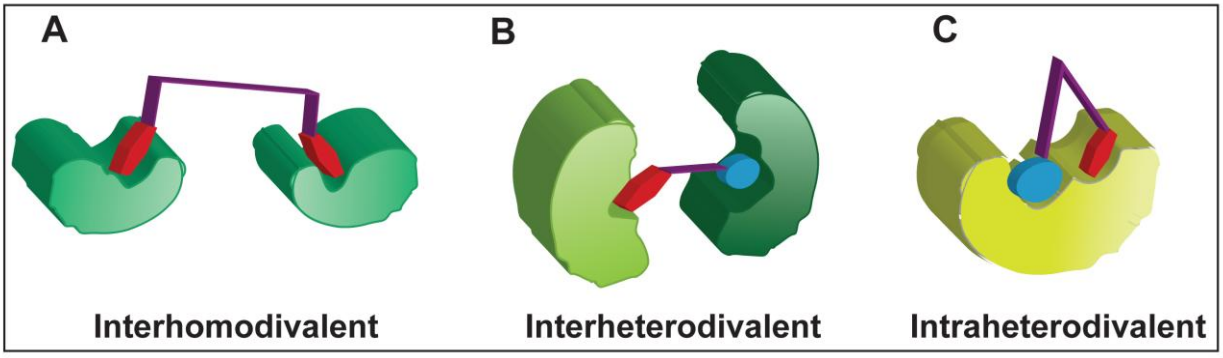

Figure 12. Types of multivalent interactions. Shown are interhomovalent (A), interheterovalent (B), and intraheterovalent (C) molecules. Interhomovalent ligands bind to multiple identical receptors while interheterovalent inhibitors to multiple non-identical proteins. Intraheterovalent inhibitors interact with multiple binding sites in a single protein. Interheterovalent ligand targets cognate binding sites of two different intracellular proteins **(Sriram et al., 2009).**

3.1.1 Principles

Intracellular signaling is often mediated by a family of functionally overlapping signal mediators that contain one or more structurally conserved domain(s) interacting with other ligands or proteins (Sriram et al., 2009). Protein-protein and ligand-protein interactions are the combined effect of multiple microscopic interactions, such as electrostatic interactions between amino acids and van der Waals interactions between atoms.

The communication between many signaling molecules is governed by weak, transient interactions (K_d > μM), as opposed to high-affinity drug-receptor interactions estimated to have mean K_d of $10^{-7.3}$ (Houk et al., 2003). Therefore the paradigm in drug discovery has been focused on screening or synthesizing the highest-affinity ligand (K_d, sub-μM or nM) on the hopes that the resulting ligand will lead to a compound with maximal therapeutic and minimal side effects (Sriram et al., 2009).

Under this paradigm, weak-affinity molecules are neglected based on a general notion that a weak-affinity molecule binds to a target with low selectivity and, thus, is pharmacologically useless. It is increasingly clear, however, that many weak-affinity biological interactions can become a useful target when multiple low-affinity ligands are combined into a multivalent molecule (Sriram et al., 2009). Various compounds with tethered ligands were designed to enhance affinity for intracellular targets, such as carbonic anhydrase-I, glutathione S-transferase, and thrombin (Banerjee et al., 2005; Erlanson et al., 2004; Maeda et al., 2006; Tolkatchev et al., 2005).

Bearing in mind the demonstrated effectiveness of multivalency in various interactions, one would speculate that a multivalent molecule targeting multiple sites within a single domain

or of multiple domains conserved in signaling molecules would enable the control of the entire protein family within a specific intracellular signaling pathway (Sriram et al., 2009).

3.1.2 Thermodynamics And Kinetics

Multivalent and monovalent molecules differ, thermodynamically and kinetically, in the nature of their interaction with target molecules (Sriram et al., 2009). Whereas the binding of a monovalent molecule is mainly determined by the ligand's binding affinity, the overall avidity of a multivalent molecule to the target is affected not only by the affinity of individual ligands but also by other parameters such as the linker connecting the individual ligands (Kiessling et al., 2000; Krishnamurthy, 2006; Mammen, 1998).

As noted by Kitov and Bundle (Kitov and Bundle, 2003), the free energy of binding for a multivalent interaction (ΔG^0_{multi}) can be described by the equation:

$$\Delta G^0_{multi} = n\Delta G^0_{mono} + \Delta G^0_{interaction} \quad (1)$$

where, ΔG^0_{mono} is the free energy of binding for the corresponding monovalent interaction, n represents the number of ligands that are bound to receptors, and $\Delta G^0_{interaction}$ contains contributions from the favorable and unfavorable effects of tethering. The various factors that contribute to $\Delta G^0_{interaction}$ are illustrated in the expression for ΔG^0_{multi} proposed by Krishnamurthy et al. (Krishnamurthy, 2006):

$$\Delta G^0_{multi} = n\Delta G^0_{mono} + (n-1)(T\Delta S^0_{mono,trans+rot} + \Delta H^0_{linker} - T\Delta S^0_{conf} + \Delta G^0_{coop}) - RT\ln(\Omega_n/\Omega_0) \quad (2).$$

The term $[(n-1)\ T\Delta S^0_{mono,trans+rot}]$ is based on the assumption that the unfavorable translational and rotational entropy of binding is approximately the same for a multivalent interaction as for a monovalent one. The term $[(n-1)\ \Delta H^0_{linker}]$ represents the change in enthalpy

due to interactions between the linker and the target. The term [-(n-1) $T\Delta S^0_{conf}$] represents the loss of conformational entropy of the linkers following binding of the multivalent ligand to the target. The term [(n-1) ΔG^0_{coop}] represents contributions from cooperativity, i.e. the influence of one binding event on subsequent binding events. The final term is a statistical factor based on the degeneracy (Ω_n) for the multivalent ligand-receptor complex (Kitov and Bundle, 2003).

The above discussion can be used to guide the design of high avidity multivalent or divalent ligands, by focusing on the various contributions to $\Delta G^0_{interaction}$. For instance, as noted by Krishnamurthy et al. (Krishnamurthy, 2006), the magnitude of the contribution due to "entropic enhancement", [(n-1) $T\Delta S^0_{mono,trans+rot}$], may be reduced by enthalpy/entropy compensation (EEC), since binding events with more favorable enthalpies of binding are associated with more unfavorable translational and rotational entropies of binding. They related $T\Delta S^0_{mono,trans+rot}$ to ΔH^0_{mono} by the expression:

$$T\Delta S^0_{mono,trans+rot} = c\ \Delta H^0_{mono} \qquad (3)$$

where c is a constant ($0<c<1$). Collectively, equations 2 and 3 suggest that for a constant ΔG^0_{mono}, the highest avidity multivalent ligands will be generated from monovalent ligands that bind with the most favorable enthalpy, ΔH^0_{mono}.

The avidity of a multivalent ligand is influenced not only by the choice of monovalent ligand, but also by the choice of linker. Equation 2 suggests that the use of rigid linkers might be optimal, as it would lower the conformational entropy penalty [-(n-1) $T\Delta S^0_{conf}$]; however, a rigid linker might also result in unfavorable interactions between the linkers or ligands and the receptor. On the other hand, a flexible linker would facilitate multivalent binding without steric obstruction, but might result in a significant loss in conformational entropy on binding. Flexible linkers have, however, been used successfully to design potent multivalent ligands (Kane, 2010),

and models based on effective concentration (C_{eff}) predict a much smaller loss in conformational entropy on binding for long and flexible linkers than models based on bonds that become completely restricted following multivalent binding (see **APPENDIX A**) (Krishnamurthy et al., 2007). Therefore, an effective strategy for the design of multivalent ligand might be to connect the individual ligands by a flexible linker with an optimal length that is close to or slightly longer than the spacing between the binding sites on the target protein (**Fig. 13**) (Kane, 2010; Sriram et al., 2009).

While the above discussion focused primarily on thermodynamics, the kinetics of interaction of multivalent ligands with their targets is also of interest. Studies on the kinetics of multivalent interaction suggest that enhancements in avidity are primarily due to a decrease in the rate of dissociation (k_{off}) of the multivalent entities than due to an increase in the rate of association (k_{on}) (Mammen, 1998). There are also fundamental differences between the dissociation of high avidity multivalent complexes and the dissociation of high affinity monovalent complexes. In multivalent binding, dissociation occurs in stages, enabling the rate of dissociation to be enhanced by the addition of sufficiently high concentrations of competing monovalent ligand (Rao et al., 1998). Therefore the design principles based on the thermodynamics and kinetics for multivalent ligands guides towards utilizing high affinity monovalent ligands that are linked with a flexible linker, whose length is longer than the distance between the binding sites (Kane, 2010). We note that the principles described above should be applicable not only for the design of multivalent ligands, but also for the design of homodivalent and heterodivalent ligands, including heterodivalent molecules that target simultaneously the type1 and type2 binding sites of N-recognins (Sriram et al., 2009).

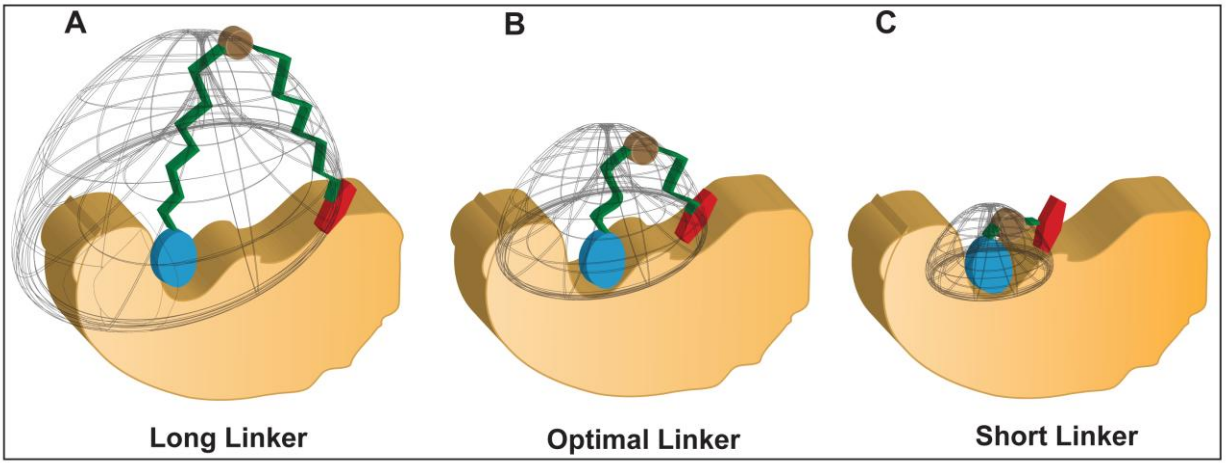

Figure 13. Influence of linker length. Model Showing the Influence of the Linker on Effective Concentrations of Divalent Molecules. The bound ligand in a divalent molecule confines the other ligand to the hemispherical proximity, influencing the effective concentration (C_{eff}) as a function of its linker length. Shown are RF-Cn-type molecules (see below), in which the linker is longer (A), optimal (B), or shorter (C) compared with the distance between two binding sites of the target **(Sriram et al., 2009)**.

3.2 HETEROVALENT INHIBITORS OF THE N-END RULE

Known mammalian N-recognins, termed UBR1, UBR2, UBR4 and UBR5, are characterized by the UBR box, a ~70-residue zinc finger-like domain that functions as a general substrate binding domain (**Fig. 3**) (Tasaki et al., 2005; Tasaki et al., 2009). The UBR box provides a structural element for binding to N-termini, in which specific residues in the UBR box (for type1) or the N-domain (for type2) provide substrate selectivity through interaction with the side group of an N-terminal residue. UBR box-containing fragments of UBR1 exhibit moderate affinity and high selectivity to destabilizing N-terminal residues with K_d of 1.6-3.4 μM (Tasaki et al., 2009; Xia et

al., 2008b). This moderate affinity allows an appropriate balance between substrate selectivity and enzymatic processivity, ensuring both selective binding to a substrate and rapid dissociation from the N-terminus for an optimal rate of polyubiquitylation. To explore the model of heterovalent interaction targeting an intracellular pathway, our lab had previously designed the heterodivalent molecule RF-C11 whose type1 and type2 ligands bind to multiple N-recognins (**Fig. 14**) (Lee et al., 2008). Heterovalent interaction to N-recognins was demonstrated to be an efficient way to control the function of this posttranslational modification pathway *in vitro* and in mammalian cells, such as cardiomyocytes. RF-C11 is a prototype compound, in which each of four replaceable components can be further optimized in affinity, stability and cell permeability. The techniques described here are likely to be useful for finding and developing multivalent compounds that modulate the function of other intracellular pathways *in vitro* and *in vivo* (Sriram et al., 2009).

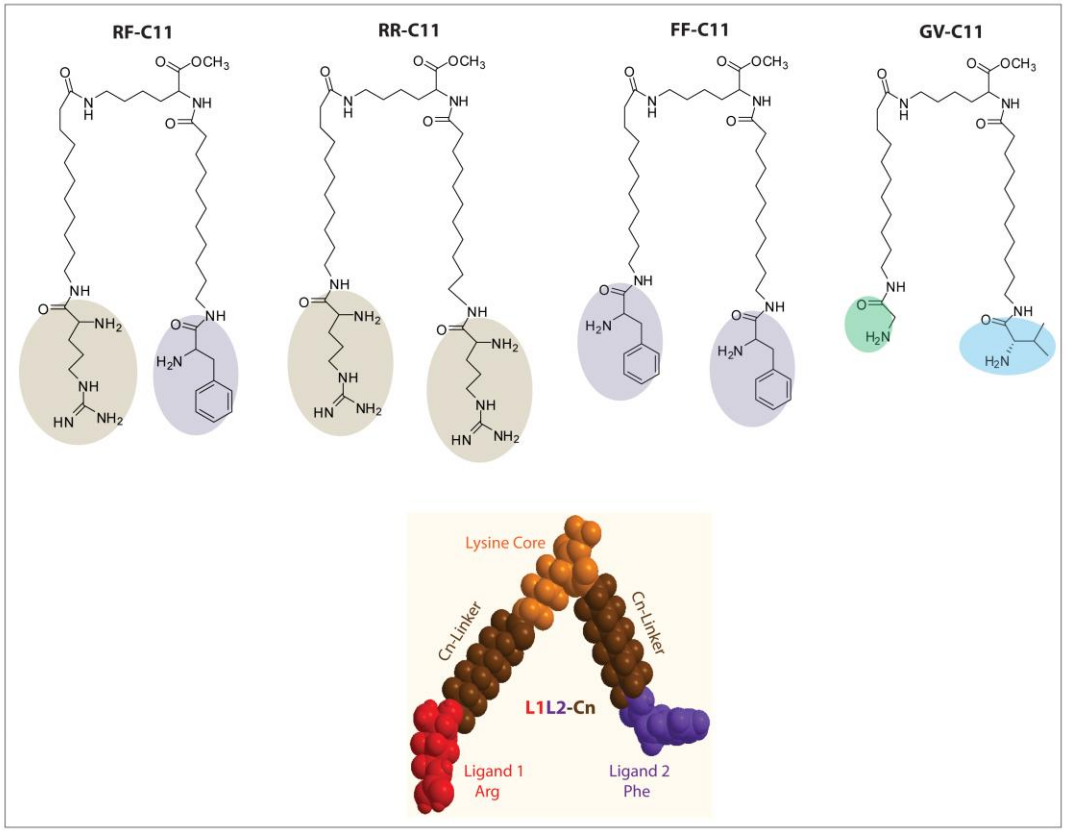

Figure 14. L_1L_2-Cn family. The Heterodivalent Inhibitor RF-C11 and its Control Compounds (A) Chemical structures of RF-C11 and its control compounds. Terminal moieties are indicated by colored background. N-terminal Arg and Phe are indicated by gold and blue backgrounds, respectively. Core Lys moiety and biotin modification are indicated by blue and red background, respectively. C11 hydrocarbon chain was used as linker/adapter. The carboxylic group of core Lys was esterified to curb further reaction. All the building blocks were connected through amide bonds. (B) A space-filling model of L_1L_2-Cn. It has mainly four building blocks; a core with multiple functional groups, two different N-degrons and two appropriate linker/spacer (Sriram et al., 2009).

3.2.1 RF-C11 Prototype and L1L2-Cn Family

Lee et al. (Lee et al., 2008) designed and characterized the synthetic heterovalent inhibitor, RF-C11, whose two different ligands cooperatively bind to two binding sites of the N-recognin family. RF-C11 was synthesized as one of the model compounds in L_1L_2-Cn family (**Fig. 14**). L_1L_2-Cn is composed of four replaceable components: ligand (L_1L_2), linker (Cn), core (lysine) and tag (e.g., biotin) (Sriram et al., 2009).. The amino acid lysine was chosen as the core component as it has trifunctional groups, among which ε-amine and α-amine are conjugated to two identical hydrocarbon chain linkers. In RF-C11, two C11 hydrocarbon chains were conjugated to the type1 substrate Arg and the type2 substrate Phe. Two homodivalent compounds, RR-C11 (bearing Arg at its termini) and FF-C11 (bearing Phe at its termini), were synthesized to compare heterodivalent vs. homodivalent interactions. The structural control GV-C11, with the stabilizing residues Gly and Val at its termini, was synthesized to evaluate the potential interaction of the linkers (**Fig. 14**) (Sriram et al., 2009)..

Polyubiquitylation involves an enzymatic cascade comprising E1, E2, E3, and the proteasome, in which crosstalk between E3-substrate interaction spatiotemporally modulates the metabolic stability of a short-lived protein (**Fig. 1**) (Sriram et al., 2009).. Accordingly, various assays are needed to verify biochemical and functional interaction of a small molecule to the N-end rule pathway. One efficient assay is to monitor the inhibitory efficacy of a small molecule on the degradation of an N-end rule substrate that is expressed in transcription-translation coupled reticulocyte lysates; this provides parameters concerning an empirical binding event rather than the actual affinity (Lee et al., 2008). Model N-end rule substrates (X-nsP4^f) can be created by

cotranslational cleavage of a Ub-protein fusion (^{f}DHFR^{h}-Ub^{R48}- X-nsP4^{f}) by deubiquitylating enzymes (DUBs), which yields a set of proteins bearing either type1, type2, or stabilizing residues (Sriram et al., 2009)..

Using Arg-nsP4 (type1) and Tyr-nsP4 (type2) as model substrates, Lee et al. (Lee et al., 2008) observed that the type1 dipeptide Arg-Ala inhibited degradation of the type1 substrate Arg-nsP4 with IC_{50} of 283 μM but showed no efficacy for the type2 substrate. Reciprocally, the type2 dipeptide Trp-Ala inhibited degradation of the type2 substrate Tyr-nsP4 (IC_{50}, 21 μM) but not type1 substrates. In contrast to monovalent compounds, RF-C11 inhibited *both* type1 and type2 substrates and, moreover, with significantly higher efficacy (IC_{50}, 16 μM for Arg-nsP4; 2.7 μM for Tyr-nsP4). RF-C11 also showed significantly higher efficacy compared to type1 homodivalent RR-C11 (67 μM for Arg-nsP4) and type2 homodivalent FF-C11 (151 μM for Tyr-nsP4). The activity of these L_1L_2-C11 compounds should be specific to the L_1L_2 ligands, as the structural control GV-C11 did not affect the degradation. The possibility that the enhanced efficacy of RF-C11 is due to allosteric conformational change of binding sites was ruled out because mixtures of monovalent or homodivalent compounds did not give significantly additive effects (Sriram et al., 2009)..

To further verify the effect of L_1L_2-C11 on the E3 activity of N-recognins, Lee et al. (Lee et al., 2008) showed that RF-C11 inhibits the *in vitro* ubiquitylation of N-end rule substrates with much higher efficacy than homodivalent compounds, that RF-C11 directly binds to a 50-kDa UBR box-containing fragment of UBR1, and that RF-C11 is capable of pulldown multiple endogenous N-recognins from rat testes extracts. These results provide experimental evidence that heterodivalent interaction to multiple N-recognins, in the midst of the mammalian proteome, leads to inhibition of both type1 and type2 N-end rule activities with higher efficacy compared to

homodivalent or monovalent interaction. Maly et al. (Maly et al., 2000) showed that a heterodivalent inhibitor, composed of carbazole and catechol units linked by a flexible alkane chain, bound to the c-Src kinase with the heterodivalent IC_{50} of 0.064μM, compared to the monovalent IC_{50} of ~40 μM. Rao and Whitesides (Rao, 1997) reported an enhancement factor of 10^3 for homodivalent vancomycin and D-Ala-D-Ala interaction.

A relatively moderate enhancement factor of RF-C11 heterovalent interaction can be attributed to the linker length and the ligand affinity to targets, if the off-target interaction of the linker and ligands with themselves or other cellular macromolecules is ignored (Sriram et al., 2009). Based on earlier discussion (see **3.1.2**) the parameters effecting the heterovalent interaction are binding enthalpy (ΔH^{mono}), conformational entropic penalty ($T\Delta S^{di}_{conf}$) and effective concentration (C_{eff}). C_{eff}, can be better explained as the enhanced local concentration of the ligands near the binding sites (see **APPENDIX A**). Specifically, during RF-C11 interaction, the bound Phe ligand to the type2 site will partially constrain the unbound Arg ligand of the same molecule within the hemisphere of radius equivalent to the linker length, and thereby increases the local Arg concentration in the proximity of the type1 site (**Fig. 13**). This will increase the probability of Arg binding to the type1 site. Reciprocally, the bound Arg, whose binding has been facilitated by the bound Phe, in turn increases the local Phe concentration in the proximity of the type2 site, further facilitating the Phe interaction to N-recognin. This mutual enhancement of local ligand concentrations is inversely correlated to the linker length, until the linker matches the distance between two targets (**Fig. 13**). Future strategy includes developing amino acid derivatives with high affinity to achieve high monovalent enthalpy and optimizing the linker to achieve the maximal effective concentration without contributing significantly to the conformational entropic penalty(Sriram et al., 2009).

3.2.2 Optimal Linker Length

Optimal linker length and high affinity monovalent ligands are the two key components in a heterodivalent ligand that are attributed to obtaining maximum multivalency-assisted enhancement (see **3.1.2)** (**Fig. 13**). We characterized compounds with varying linker lengths to understand the effect of linker length on the binding affinity of a heterodivalent ligand targeting the N-end rule pathway. L_1L_2-Cn family of heterovalent compounds were tested by using *in vitro* model substrates, which were generated through the cleavage of a tripartite fusion protein, $^{f}DHFR^{h}$-Ub^{R48}-X-$nsP4^{f}$, using an *in vitro,* coupled transcription and translation (TnT) system in rabbit reticulocytes. For a reference protein, DHFR-Ub fragment was stably detected. Type1 and type2 N-degrons were recognized and ubiquitylated by the N-recognins for degradation. Therefore, the model substrates bearing type1 and type2 N-degron, such as Arg-nsP4 and Tyr-nsP4, respectively, are short-lived in normal condition, while proteins bearing non-N-degron, such as Met-nsP4 and Ala-nsP4, were robustly expressed as reference (**Fig. 15**). Since the half-lives of X-nsP4s were closely related with the identity of X, the N-degron, the affinity of L_1L_2-Cn compounds to N-recognins can be indirectly observed by detecting the inhibitory potential of L_1L_2-Cn compounds to inhibit the X-nsP4 degradation (see **3.2.1**).

A subset of L_1L_2-Cn family members with differing linker lengths were studied, in these compounds L_1L_2 (RF) was kept constant and only linker length (Cn) was varied to understand the influence of linker length on the binding affinity of these compounds. We tested these RF-Cn (n = 2, 3, 5, 8, 9, 11 & 15) compounds, using our *in vitro* rabbit reticulocyte TnT reaction (**Fig. 15**). They were characterized by time-course immunoblotting, using both type1 and type2 model substrates at 10, 100 and 250 μM concentrations. A common trend (with few exceptions)

that emerged from these studies was that upto a certain increase in the linker length, the inhibitory potential (hence the binding affinity) increased and then it started decreasing. Based on the linker lengths (n=2, 3, 5, 8, 9, 11 and 15) tested, the maximum activity was observed for n=9 (RF-C9), followed by RF-C5 and RF-C11, as clearly evident in the lowest (10 μM) concentration profile at 100 mins. Therefore the optimal linker length for RF-Cn family to target N-recognins was around n=9 (**Fig. 15&16**).

Additionally, by looking at $\Delta G_{\Delta}^{multi-mono}$, i.e the difference between the change in free energies of a multivalent and a single monovalent interaction (see **APPENDIX A**), which is the enhancement due to multivalency, we have:

$$\Delta G_{\Delta}^{multi-mono} = (n-1)\Delta H^{mono} - T\Delta S_{nconf}^{multi} \quad (4)$$

Suggesting that when the contribution of monovalent enthalpic component in multivalent interactions is kept constant (i.e., L_1L_2=RF), then the conformational entropic component (i.e., Cn) becomes the significant factor in determining the overall avidity. Therefore the linker length, hence the linker conformation after binding becomes critical in determining the enhancement in a heterodivalent interaction. Similarly C_{eff}, which is a function of the linker length and its flexibility (see **APPENDIX A**) and which determines the probability of the ligands being bound to the binding sites, can also be influenced by an optimal linker length (**Fig. 13**). Since we observed RF-C9 to have the highest binding affinity among all the linker lengths tested, this suggests that the linker with n ~ 9 offers least conformation penalty and perhaps the most C_{eff}, therefore this becomes the most favorable length to attain the maximum binding affinity (**Fig. 15&16**).

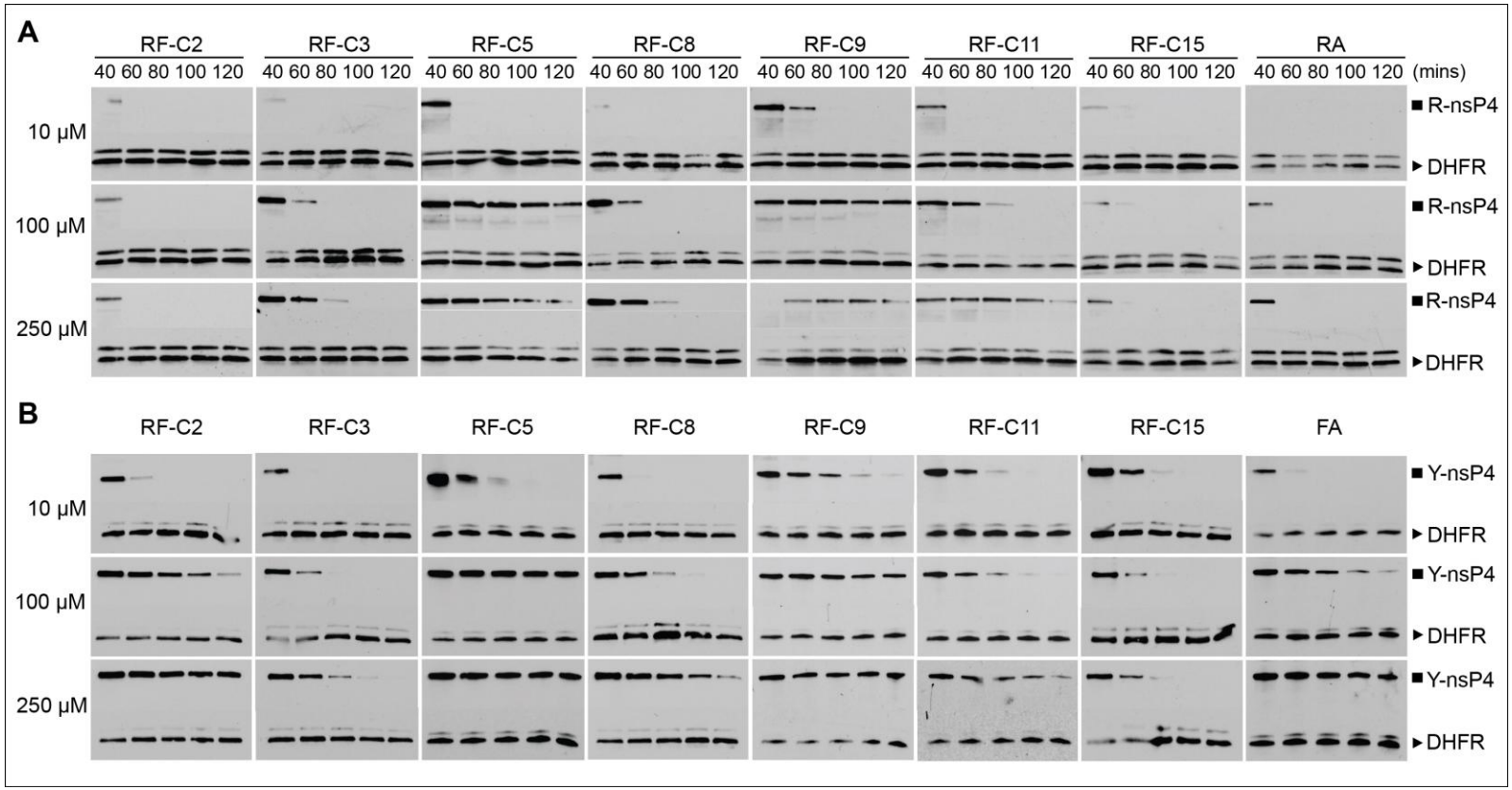

Figure 15. Effect of linker length on RF-Cn activity. *In vitro* inhibition of (A) Arg-nsP4 and (B) Tyr-nsP4 degradation by the divalent ligands. The model substrates were expressed in reticulocytes in the presence of different inhibitors or controls, and their degradations were analyzed in time- and concentration-dependent manner using traditional Western blotting. 150μM bestatin was included for RA and FA dipeptides.

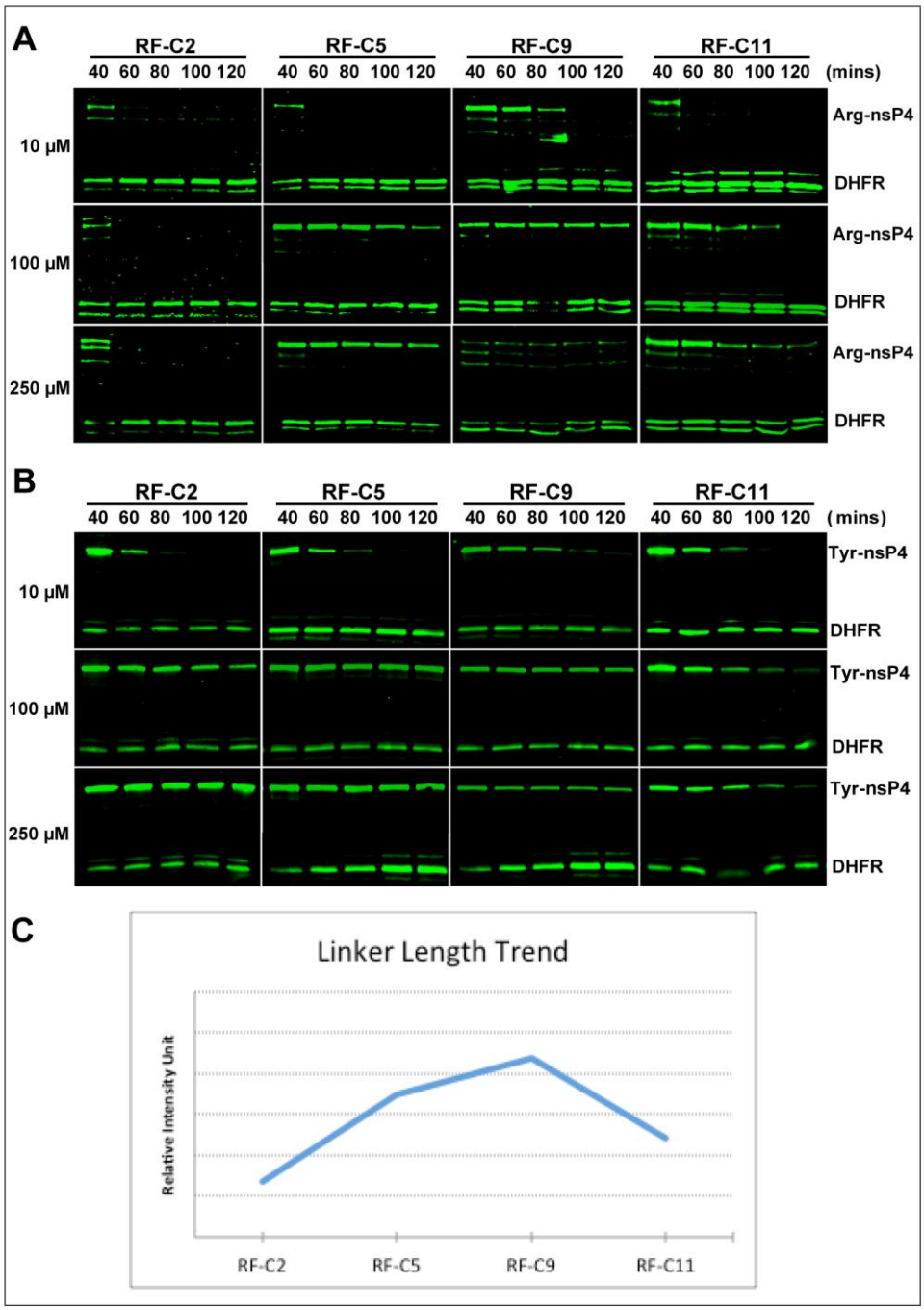

Figure 16. RF-Cn activity trend with linker length. *In vitro* inhibition of (A) Arg-nsP4 and (B) Tyr-nsP4 degradation by the divalent ligands. Same as previous condition expect IR dye-streptavidin based Li-COR system was used for detection. (C) The curve depicts the combined trend of activity on type1 and type2 with respect to the linker length.

Interestingly linker lengths n=3 (RF-C3) and n=8 (RF-C8) exhibited a different behavior outside the trend observed for other lengths (**Fig. 15**). After excluding the experimental errors and chemical integrity to this behavior, we modeled the predicted width of these compounds. The expected order was RF-C2 < RF-C3 < RF-C5 < RF-C8 < RF-C9 < RF-C11 < RF-C15, based on the assumption that width and the linker length have a linear relationship. But the predicted width followed the order RF-C3 < RF-C2 < RF-C5 = RF-C9 < RF-C11 < RF-C15 = RF-C8 (**Fig. 17**).

Suggesting that the width of RF-C3 was narrower and that of RF-C8 was broader than expected. And surprisingly this anomaly in widths, overlapped with the lowered activity of RF-C3 and RF-C8 compounds. Also the widths of RF-C5 and RF-C9, was almost similar and they exhibited the highest affinity, suggesting that this width was the most suitable for binding to the N-recognins (**Fig. 17**). This insight perhaps points at the fact that RF-C3 and RF-C8 are burdened with a huge conformational penalty, which leads to their suboptimal binding affinities (See **3.1.2**). However this analysis is based on unsubstantiated modeling results, so the interpretation based on this outcome needs to be cautiously taken into consideration, till this hypothesis is either experimentally validated or otherwise.

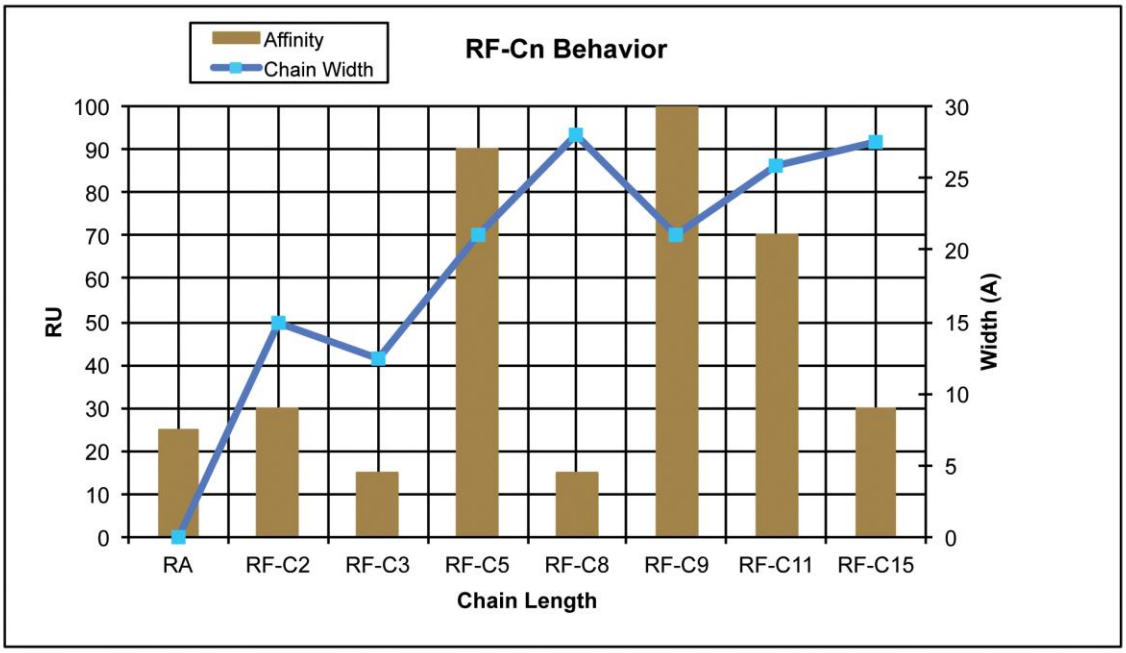

Figure 17. Width and linker length. This a plot with the relative activity of heterodivalent ligands normalized to RF-C9 (100%), and overlapped with the calculated widths of different linker lengths.

Our lab had previously reported true multivalency-assisted enhancement with RF-C11 by comparing with a combination of monovalent RA+FA and homodivalent RR-C11+FF-C11 (chapter 3.2.1). We wanted to corroborate this observation by utilizing a combination of monovalent ligands carrying a single linker (R-C5+F-C5) and by comparing them with our heterodivalent ligand (RF-C5), as it was not done earlier (**Fig. 18**). Here we employed a similar time-course immunoblotting using streptavidin (with IR dye)-biotin conjugation system on an Odyssey detection platform. Ligands were tested at a concentration of 100 μM and the samples were collected at 40, 60, 80 & 100 mins from the start of a TnT reaction. Y-nsP4 and R-nsP4, were used a type2 and type1 model substrates respectively. As expected, the combination of two

monovalent-linker ligands (R-C5+F-C5) still has a lower inhibitory potential than a single heterovalent ligand (RF-C5). In line with our earlier discussion (See **3.1.1 and 3.1.2**) this can be attributed to the inherent thermodynamic and kinetic enhancement of a multivalent interaction that is not obtained by combining two separate individual interactions. Also note that the activity of R-C5 (type1 ligand) alone towards R-nsP4 (type1 substrate) is weak compared to that of F-C5 (type2 ligand) alone towards Y-nsP4 (type2 substrate) (**Fig. 18**). This is an expected trend for type1 ligands based on previous observations including FA, WA and RA dipeptide results with Y-nsP4 and R-nsP4, respectively, so the possibility of an inactive R-C5 can be excluded here.

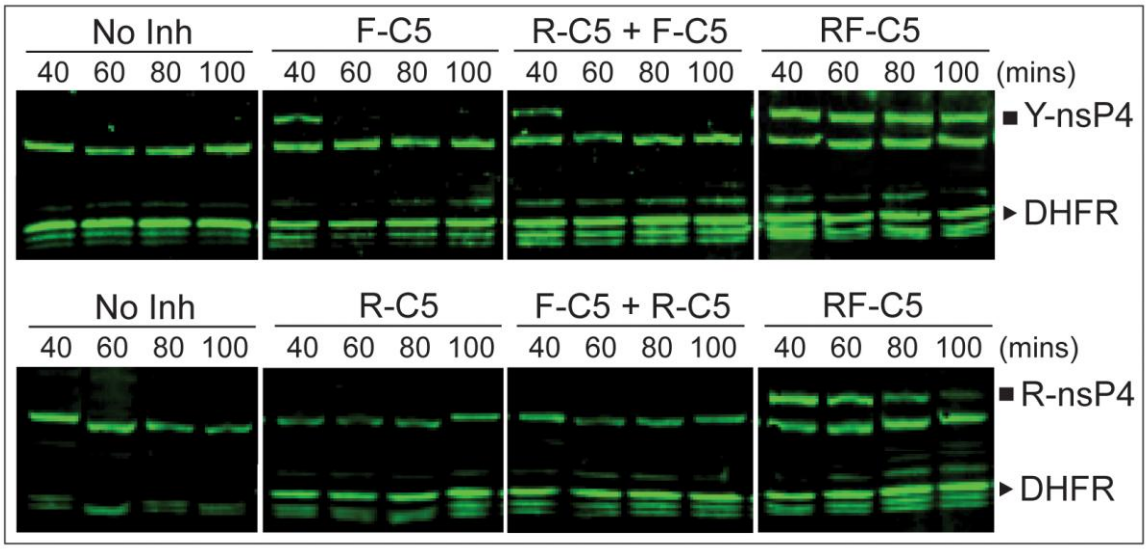

Figure 18. Multivalency-assisted enhancement. Y-and R-nsP4 was expressed in reticulocyte lysates, where its cotranslational cleavage yields the DHFR-Ub reference and the X-nsP4 substrate. The effects of homovalency vs. heterovalency at 100μM concentration on type1 and type2 degradation was monitored by using time course anti-biotin Western blotting and Li-COR detection system.

Further validation of the heterovalent nature of RF-C5 was studied by analyzing with RR-C5 (type1 homovalent control) and FF-C5 (type2 homovalent control), to compare heterodivalent vs. homodivalent interactions (**Fig. 19**). The negative control GV-C5, with the stabilizing residues Gly and Val at its termini, was compared to evaluate any non-specific interactions from the lysine core or the linker (**Fig. 14**). As seen previously with the prototype RF-C11 controls, RF-C5 controls exhibited a similar behavior. RR-C5 had no detectable activity against type2 substrate and vice versa with FF-C5, demonstrating the heterovalent nature of the binding sites on the N-recognins. RF-C5 was still the strongest ligand compared to RR-C5 and FF-C5 in stabilizing R-nsP4 and Y-nsP4, respectively. Interestingly, FF-C5 exhibited stronger affinity than FA compared to FF-C11, which demonstrated a lower affinity than FA (See **3.2.1**).

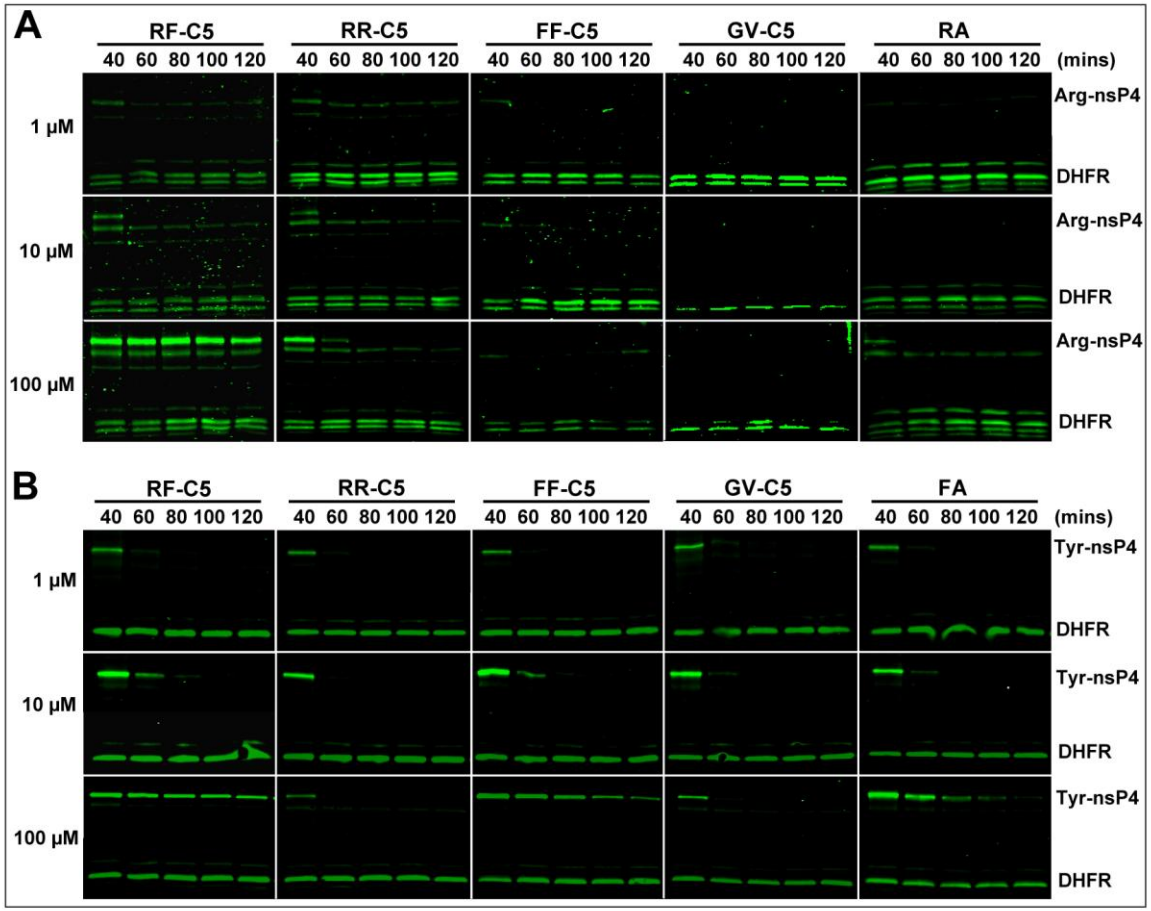

Figure 19. L_1L_2-C5 controls. The model substrates were expressed in reticulocytes in the presence of RF-C5 or its controls, and their degradations were analyzed in time- and concentration-dependent manner using anti-biotin Western blotting with Li-COR detection sytem. GV-C5 is a structural control of RF-C5, and RR-C5 and FF-C5 are heterovalent controls of the bivalent RF-C5. 150μM bestatin was included for RA and FA dipeptides.

3.2.3 Monovalent Derivatives

The second key component for achieving maximum multivalency-assisted enhancement is to have monovalent ligands with strong binding affinities (See **3.1.2**). We started looking into this by tackling basic questions like, how can dipeptides be ligands but a single amino acid cannot be an N-degron? What is the role for the second amino acid? What are the minimal structural recognition features? We then started with type2 'Phe' derivatives, as most of the derivatives were available commercially and later the crystal structure of homologous type2 site (ClpS) was also solved. Our first goal was to define the minimal recognition requirements, since these minimal features are required to achieve the threshold binding energy for recognition with N-recognins. Then any additional binding feature will only add to the magnitude of this recognition, thus providing a derivative with higher affinity.

Based on the molecular principles of substrate recognition by N-recognins we know that there are three main attributes that form the minimal requirements for substrate recognition (See **2.3**). First two come from the N-terminal residue and third from the linkage between the N-terminal and second residue. They are 1) α-amino group of the N-terminal amino acid, 2) appropriate N-terminal side-chain and 3) proper functional groups beyond the N-degron, like a peptide bond between the N-terminal and second amino acid (**Fig. 11**). Phe derivatives were used to analyze these requirements and to characterize the recognition features so they can be used as a template for designing the ideal monovalent compound (See **3.1.1**).

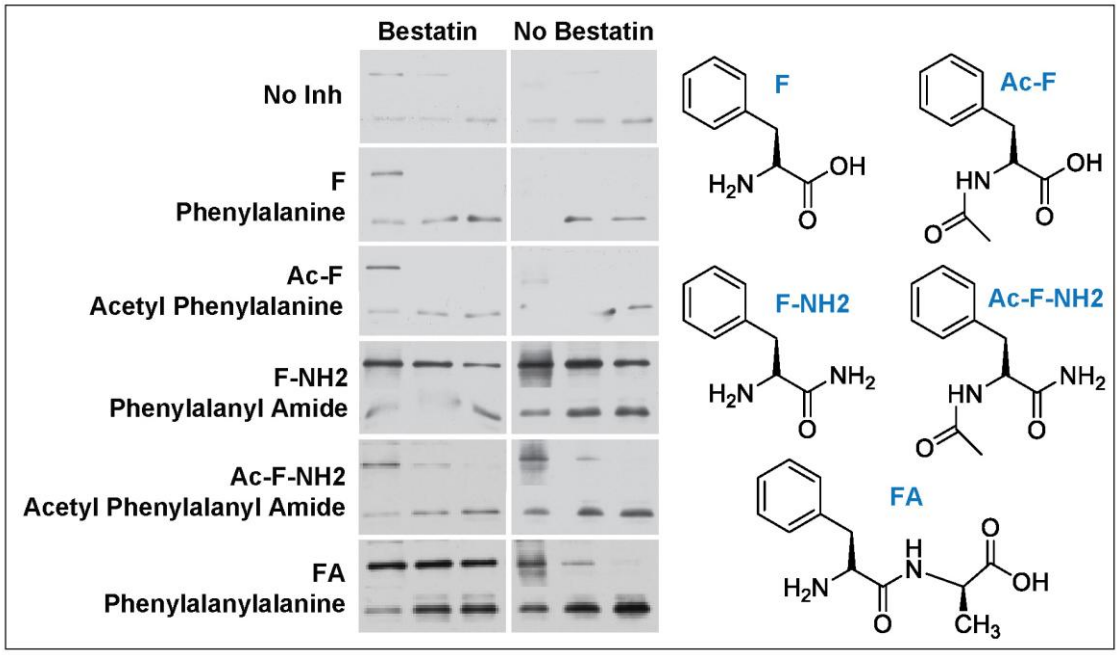

Figure 20. Role of α-amino group. Type2 model substrate was expressed in reticulocytes in the presence of different derivatives, and their degradations were analyzed in the presence and absence of bestatin in a time-course (40, 60, & 80 mins) dependent manner using anti-biotin Western blotting.

We started with the simplest compound, Phenylalanine (Phe or F), Phe has an appropriate N-terminal side-chain and a free α-amino group, so it fulfills two of the three criteria, but we failed to detect any activity (**Fig. 20**). Then we compared Ac-Phe, whose α-amino group is blocked with an acetate group. Even this provided no detectable binding. Then we compared Phe-amide (and Ac-Phe-amide), which has the side-chain, α-amino group and a primary amine group on the C-terminus. Interestingly, Phe-amide exhibited activity similar to FA (Phe-Ala) dipeptide and was even resistant to proteases as the activity was retained without bestatin. 150 uM bestatin was added to the reaction to protect dipeptides from proteases in the reticulocyte lysates. When the α-amino group on Phe-amide is blocked as in Ac-Phe-amide, there is only

negligible binding affinity. Based on these observation it is clear that the α-amino group is a key determinant and also the primary amine in Phe-amide substituted the role of a NH in a CO-NH peptide link. We know that in type2 interactions the N-terminal side-chain makes favorable interactions with the hydrophobic surface, the α-amino group forms three H-bonds and the NH in the CO-NH peptide link forms a H-bond (**Fig. 11**). Phe-amide had the ability to utilize all the key recognition features leading to its binding affinity. Therefore single amino acids cannot be N-degrons since they do not have sufficient features for recognition, and is therefore not able to provide the threshold binding affinity needed for recognition. A comparison of Arg and Arg-NH2 binding with UBR box has showed a difference of 30x in their binding affinity (Choi et al., 2010). Interestingly this NH (in both type1 and type2 N-degrons) interacts with a Thr residue of the N-recognin, which is highly conserved in both UBR box and ClpS interactions. And it seems that the need of a second amino acid is not a prerequisite for a small ligand (i.e., dipeptides) as shown with Phe-amide (See **2.3**), however the second residue in a protein/peptide, indeed contributes to the binding affinity (Choi et al., 2010; Matta-Camacho et al., 2010; Wang et al., 2008a).

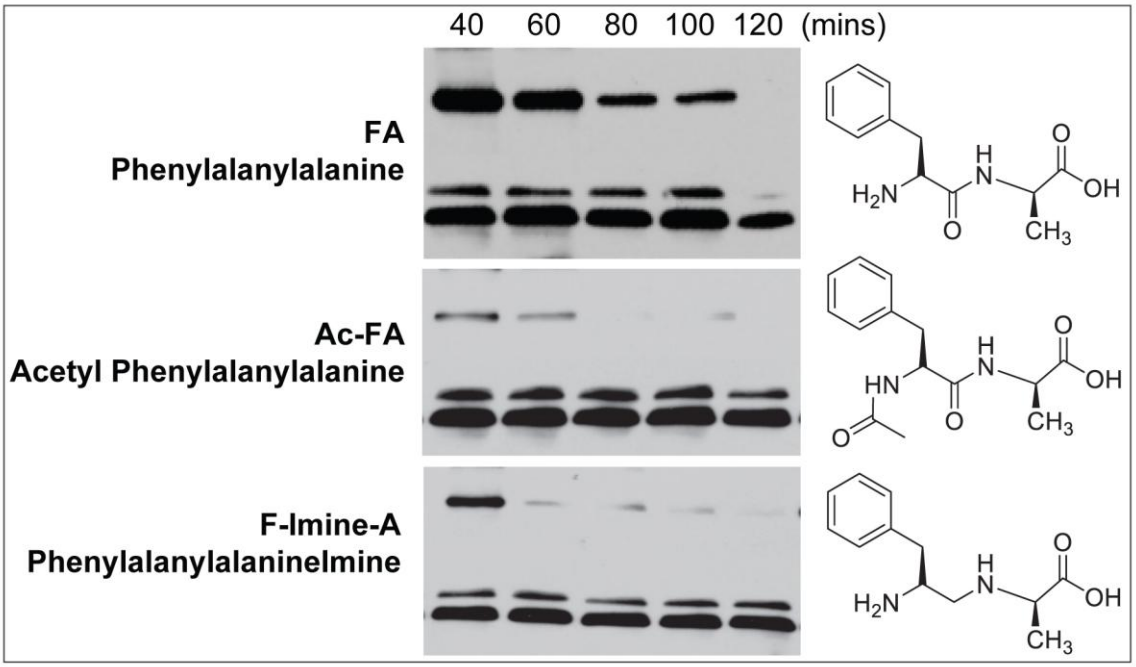

Figure 21. Role of peptide bond. Y-nsP4 model substrate was expressed in reticulocytes with different derivatives, and their degradations were analyzed in the presence of 150□M bestatin at 40, 60, 80, 100 and 120 mins using anti-biotin Western blotting**.**

We then examined the role of a peptide bond by comparing FA dipeptide and a Phe-imine-Ala compound that lacks a peptide link (**Fig. 21**). The activity of Ac-Phe-Ala and Phe-imine-Ala is very negligible compared to FA. Phe-imine-Ala lacks a peptide link but still has a NH group that does not seem to function, it perhaps lacks the ability to participate in H-bonding. Thus a proper peptide link is necessary to have its full functionality when there is a second residue (See **2.3**).

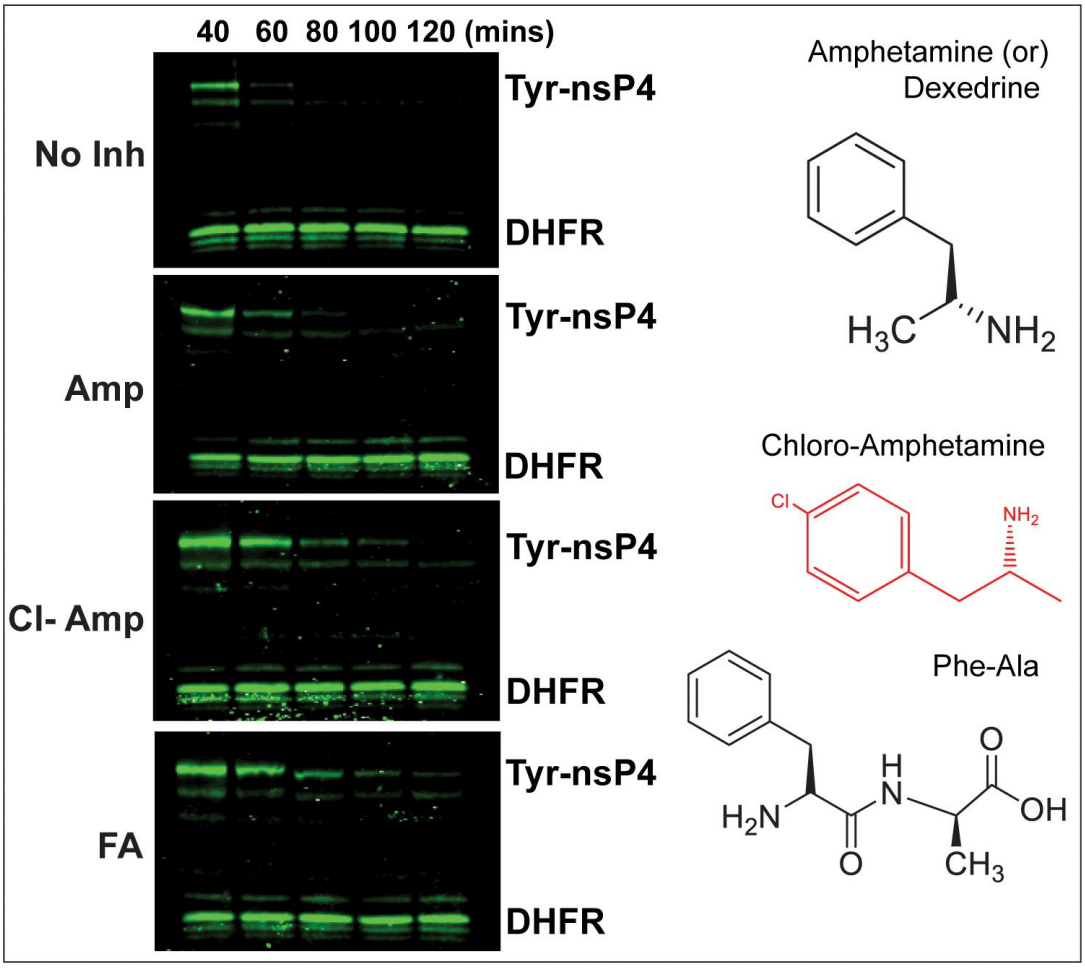

Figure 22. Amphetamine as N-degron. Type2, Y-nsP4 model substrate was expressed in reticulocytes with different derivatives, and their degradations were analyzed at 40, 60, 80, 100 and 120 mins using anti-biotin Western blotting. Only FA condition had 150μM bestatin

While we were screening for Phe-amide like analogs, we tested d-amphetamine (or Dexedrine) and its analogs (chloramphetamine and methamphetamine) (**Fig. 22**). This family had all the minimal recognition features of a N-degron (See **2.3**) (**Fig. 11**). Amphetamine and methamphetamine (data not shown) exhibited detectable binding to the type2 site. Moreover the addition of additional hydrophobic group on the N-terminal side-chain (chloramphetamine)

increased the binding affinity, since the N-terminal side-chain is lodged in a deep hydrophobic pocket. In contrast addition of charged groups (-OH) on the N-terminal side-chain led to decreased affinity as seen in the case of dopamine (data not shown) and tyramine (**Fig. 23**). Amphetamine is a central nervous system stimulant (psychostimulant) commonly used to treat several disorders, including attention deficit (ADHD), narcolepsy, and obesity, but has a propensity for substance abuse (Fleckenstein, 2007). This is the first report of a known therapeutic molecule identified as a N-degron in the N-end rule pathway, moreover they were stable even with out bestatin, suggesting that amphetamine-family might perhaps play a modulatory role by binding to N-recognins in the nervous system.

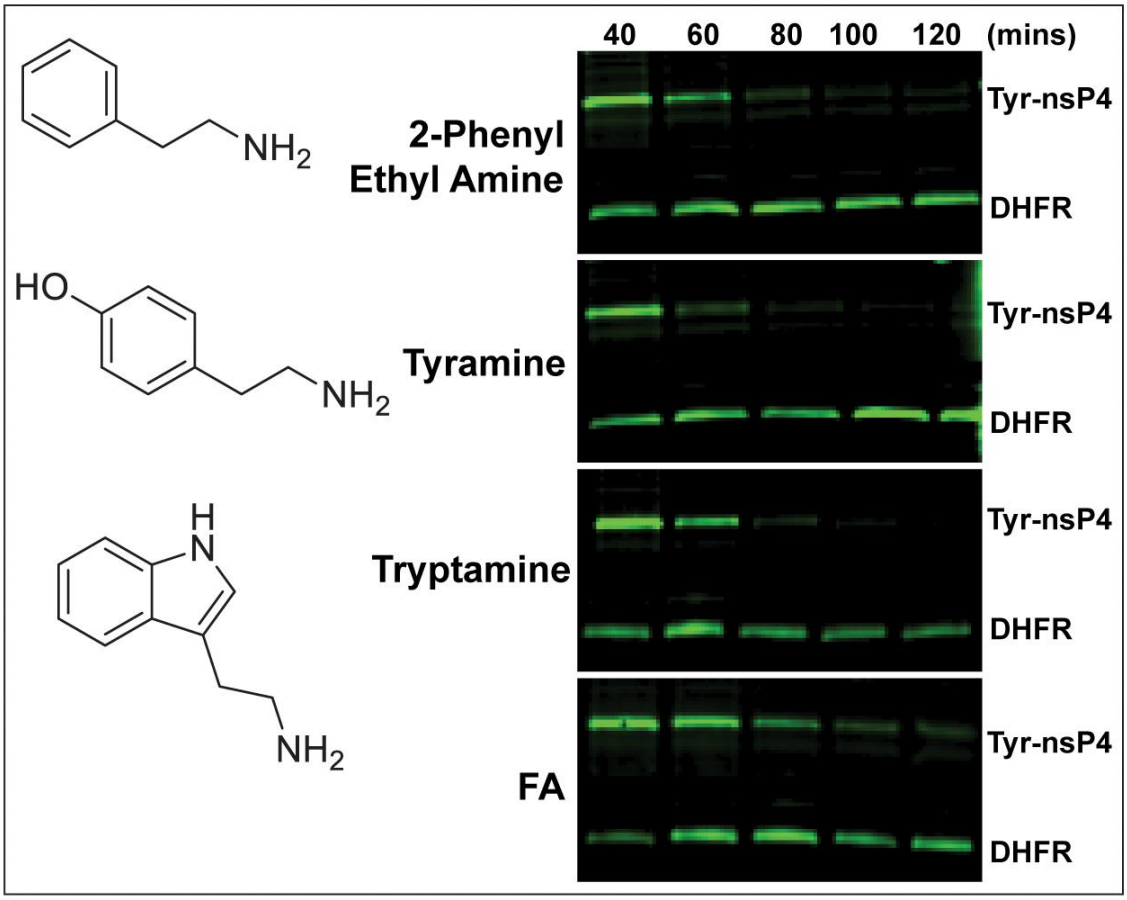

Figure 23. Trace amines as N-degrons. Y-nsP4 model substrate was expressed in reticulocytes with different derivatives, and their degradations were analyzed at 40, 60, 80, 100 and 120 mins using anti-biotin Western blotting. Only FA condition had 150μM bestatin

Additionally, our screening identified 2-phenylethylamine (2PEA), tyramine and tryptamine as N-degrons (**Fig. 23**). This group also includes the key structural recognition features discussed earlier; infact 2PEA seems stronger than amphetamine, which can be attributed to the subtle differences in their structures (See **2.3**). Also addition of a charged group

on the N-terminal side-chain (tyramine) disrupts the binding. And having a more hydrophobic group (indole group in tyramine) adds to the binding affinity (**Fig. 23**).

Interestingly, this group is categorized as 'trace amines'. Trace amines are physiological molecules synthesized in the brain and in the peripheral nervous system (Burchett, 2006). They have a high rate of metabolism and therefore are present in 'trace' amounts. At low concentrations they potentiate the activity of neurotransmitters (neuromodulators) and at high concentrations they exhibit neurotransmitter-like functions. Trace amines are associated with disorders similar to amphetamine abuse. Amphetamine and trace amines are identified as the first therapeutic and physiological molecule acting as N-degrons. This leads to the speculation that these small molecules binds to the N-recognins in specialized environments like neurons, to modulate the local N-end rule pathway (see **4.3**).

Based on the observations from these derivatives (**Fig. 20-23)**, we demonstrated the minimal structural features for a type2 N-degron, namely 1) α-amino group 2) hydrophobic side-chain and 3) peptide link (See **2.3**). In addition we have identified enhancement modifications, such as increasing the hydrophobicity of the N-terminal side-chain (chloramphetamine and tryptamine), avoiding charged functional groups near the side-chain (tyramine and dopamine) and having a primary amine (NH_2) group, like in a peptide link at the end of N-terminal amino acid (Phe-amide, 2PEA, and amphetamine). In addition to the N-terminal amino acid features, we looked at the down stream enhancement features towards the C-terminus (**Fig. 24**).

To understand these additional C-terminal interactions, we tested Phe-derivatice with C3 and C5 linkages that vary in their C-terminal functional groups. The lengths C3 and C5 correspond to position 2 and 2.5, respectively in a three amino acid trimer (**Fig. 24**). It was revealed that by having a functional group that could participates in H-bonding (F-C2-acid, F-C3-acid, F-C3-phos), increases the binding affinity of that derivative, as opposed to a derivative not having such a functional group (F-C5, F-C5-ester) (See **2.3**) (**Fig. 11**). Our mapping of these additional enhancement interactions can be combined with the minimal recognition features to yield a hybrid ligand with significantly high binding affinity (See **3.1.2**).

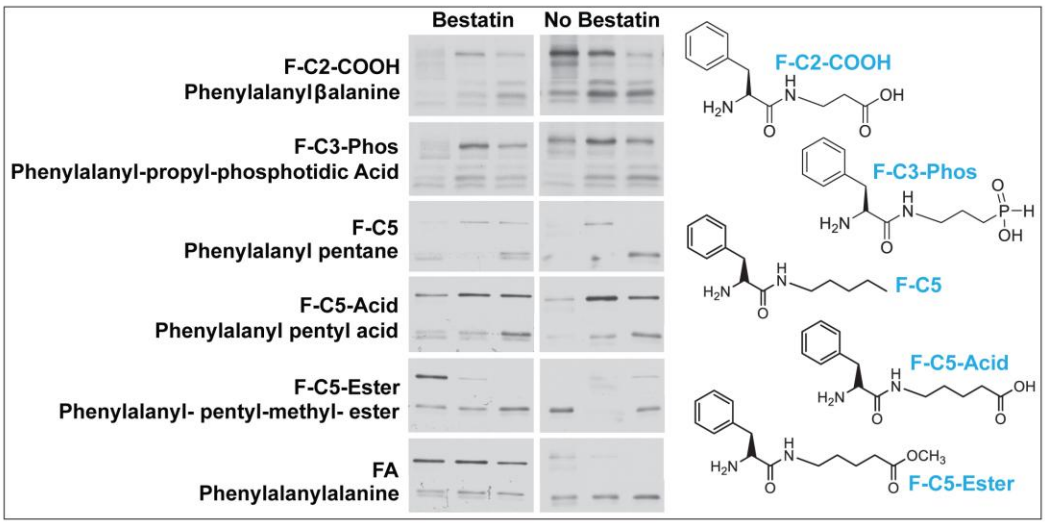

Figure 24. Role of C-terminal side. Y-nsP4 model substrate was expressed in reticulocytes with different derivatives, and their degradations were analyzed at 40, 60, 80, 100 and 120 mins using anti-biotin Western blotting in the presence and absence of 150μM bestatin.

3.3 METHODS

Synthesis of Ligands: The L_1L_2-Cn family was synthesized by our collaboarators at IICT, India. It was following the same strategy used for RF-C11 and its controls. Briefly, the intermediate compounds of L_1L_2-Cn and controls are connected by conventional protocols of amide bond formation using EDC with high yield. For example, to synthesize RF-Cn, the carboxyl end of Arg [Z-Arg(Z)2-OH] is conjugated to a Cn-carbon chain through an amide bond, resulting in an Arg-Cn chain. In parallel, the carboxyl end of a second Cn-carbon chain is conjugated to the ε-amino group of a Lys residue, yielding a Cn-Lys chain. Phe (BOC-Phe-OH) is conjugated to the Cn-Lys to make a Phe-Cn-Lys chain. The free amine of Lys of the Phe-Cn-Lys chain is conjugated to the carboxyl end of the Arg-Cn chain to yield RF-Cn. The reactive free carboxylic end of Lys is esterified to curb any further reaction through it. The intermediates and the final products were characterized by TLC, analytical HPLC, ^{1}H NMR, and mass spectroscopy. Phe derivatives, amphetamine family and trace amines were commercially available, other derivatives were a kind gift from Dr. Song Li's lab.

In Vitro Degradation Assay: Time- and concentration-dependent in vitro degradation and assays were done as described earlier (Chapter 3). Briefly, biotinylated lysine-tRNA complex (Transcend tRNA, Promega) was added into a reaction mixture for random labeling of expressed proteins and biotinylated proteins were detected against horseradish peroxidase (HRP)-conjugated streptavidin (Pierce) or IR dye-conjugated streptavidin (Li-COR). To evaluate L_1L_2-Cn efficacy on *in vitro* degradation, we expressed Y-nsP4 (type2), R-nsP4 (type1) model substrates from ^{f}DHFR^{h}-Ub^{R48}-X-nsP4^{f} fusion protein in transcription-translation-coupled reticulocyte lysate (TnT) in the presence or absence of inhibitors and 150μM bestatin, and measured the remaining protein levels using time-course immunoblotting.

Computational Modeling: Molecular modeling was conducted with ChemBio3D Ultra 12.0 (CambridgeSoft). Ligand structures were generated using ChemDraw. Ligand structures were energy minimized using the MM2 force field. Distances were calculated using the atomic distance tool in ChemBio3D Ultra 12.0.

4.0 DISCUSSION

4.1 N-END RULE CODE

Second Residue. In addition to the N-terminal residue, the second residue of a substrate also contributes to N-end rule selection, i.e., the second position influences the magnitude of binding affinity and not so much to the specificity (**see 2.3**). Yet there is no known principle or 'code' for the importance of the second residue. It was reported that yeast UBR box prefers hydrophobic residues (Leu & Ile) at the second position, followed by basic residues, whereas acidic residues are rather disfavored (Choi et al., 2010). In contrast, human UBR box prefers acidic residues (Asp and Glu), such as those found in the products of arginylation of pro-N-degrons (**Table 1**) (**Fig. 9**), these residues had 8x more binding affinity than Ala (Choi et al., 2010; Matta-Camacho et al., 2010). Interestingly, bacterial ClpS favors a secondary destabilizing residue, like basic Arg or Lys residues, followed by hydrophobic residues at the second position (**Fig. 10**) (Erbse et al., 2006; Wang et al., 2008a). It appears that the second residue significantly influences the affinity to the UBR box in a manner that is customized to specific N-recognins. This is achieved based on the complementarity with the charge distribution on the N-recognin sites that are in contact with the second residue, as seen in the structures of the yeast and human UBR boxes and ClpS (**Fig. 11**) (Sriram and Kwon, 2010). This knowledge has implications in developing ligands that can exhibit inter-N-recognin specificity.

Physiological N-degrons. Another mystery in the N-end rule pathway has been that, despite the intense efforts for the last two decades, only a limited number of type1& type2 substrates were identified (**Table 2**), which cannot be fully explained by conditionality and complexity of N-degron. Recent structural data might hold the key to this puzzle. It was found that the UBR box of yeast Ubr1 structurally prefers Arg over other type1 substrates like Lys and His, because the side chains of Lys and His are not close enough for the full range of hydrogen bonding with Ubr1 residues (Choi et al., 2010). Also Lys and His show ~5 and ~15 times lower binding affinity to Ubr1 than Arg. This structural analyses, together with previous binding assays, suggest that the UBR box has been structurally and functionally optimized to the arginine and arginylation branch of the pathway, whereas Lys and His may not, and if any, may rarely occur in physiological N-degrons (**Table 1**) (Sriram and Kwon, 2010). Consistent with this possibility, all of classical (non-acetylated) N-end rule substrates of eukaryotes identified thus far bear the primary degron Arg or arginylation-permissive pre-N-degrons. Such substrates include *S. cerevisiae* Scc1 with N-terminal Arg, *Drosophila* DIAP1 with N-terminal Asn, and mammalian RSG4, RGS5, and RGS16 with N-terminal Cys (**Table 2**). The enrichment of Arg in N-degrons can be explained by its superiority as degron in that pre-N-degrons can be conditionally destabilized through arginylation as a checkpoint prior to irreversible proteolysis (see **1.1**). Because of this conditionality, short-lived proteins that entered various entry points of the pathway may have been merged to its arginylation branch in the course of evolution (Sriram and Kwon, 2010).

Evolution of the N-end rule. The established functional and the newly found structural similarities between the eukaryotes and prokaryotes N-end rule pathways, eludes at the N-degron code development through evolution (**Table 1, Fig. 2**). In this model, ancient Ubr1 of eukaryotes

had only the UBR box, which recognized Arg and other type1 residues as part of the UPS, whereas the ancient prokaryotes developed a distinct folding of ClpS to recognize type2 N-degrons. Later the substrate specificities of present eukaryotic N-recognins were established by the incorporation of *ClpS* into *Ubr1*, probably when prokaryotic endosymbionts in eukaryotic cells were converted to organelles, such as the mitochondrion and the ER (Dyall et al., 2004). Since the zinc finger motif of the UBR box offers a superior structure and regulation (e.g., redox modification and allosteric modulation by N-end rule ligands), the *ClpS* encoded type2 site became structurally dependent on the UBR box, to gain a second level of selectively. In parallel, ancient Ubr1 gained the ability to mediate pro-N-degron-dependent proteolysis when R-transferase was recruited from the bacterial genome. After the divergence between yeasts and mammals, the N-end rule pathway added yet another regulatory mechanism by recruiting $Nt^{N,Q}$-amidase with a eukaryotic origin, which mediates deamidation of the pro-N-degrons Asn and Gln. An analogous process might have occurred in mammals with Nt^{N}-amidase and Nt^{Q}-amidase through convergence evolution (**Fig. 2&3**) (Tasaki, 2012).

4.2 UBR BOX

Redox Sensor. While the UBR box is a characteristic domain present in all eukaryotic N-recognins, the UBR box itself is not sufficient for substrate recognition as some UBR proteins do not seem to bind the N-end rule peptides (e.g. UBR3, UBR6 & UBR7), indicating that the UBR box may have an additional function outside the N-end rule interaction (**Fig. 4**). Given that the Cys_2His_2 zinc finger is found in many proteins, such as DNA binding proteins, where the Cys_2His_2 zinc finger acts as a redox-sensitive molecular switch (Pabo et al., 2001), perhaps the

UBR box might also act as a redox sensor. As in typical Cys_2His_2 zinc fingers, cysteine thiols and imidazole nitrogen atoms of histidine residues act as Zinc coordinating ligands (**Fig 8**). Under physiological conditions, the cytoplasmic milieu of the cell is reducing, and then most of the Zn^{+2} ions are bound by proteins. When the internal environment becomes oxidative, then the zinc finger thiols are oxidized releasing Zn^{+2} from zinc fingers (Kroncke and Klotz, 2009). This disruption of zinc finger leads to conformational changes in a protein. Similar redox sensitivity might be relevant to UBR family members localized in endoplasmic reticulum or mitochondria, where the reactive oxygen species (ROS) are active. Interestingly two or more zinc fingers can exhibit cooperativity and lead to consequential conformational changes in the native protein (Lee, 2010), in UBR family members this might translate into cooperativity of binding sites and/or allosteric activation of binding site (Sriram et al., 2011).

Confirmation. It has been known that the binding of type1 and type2 dipeptides to the UBR box and the N-domain synergistically induces conformational changes in yeast Ubr1, leading to the accelerated degradation of a substrate bearing an I-degron (Du et al., 2002; Turner et al., 2000) Through this feedback mechanism, the binding of ligands to UBR box and the N-domain exposes the Cup9 binding site, leading to a chain reaction involving Cup9 degradation, PTR2 accumulation, and accelerated peptide import (Du et al., 2002; Turner et al., 2000). Thus, it has been speculated that this allosteric activation is triggered by a conformational change in the ligand bound UBR box (**see 1.1**). Intriguingly, both Gehring and colleagues in human UBR box (personal communication) and Song and colleagues in yeast UBR box observed that a conformational change is indeed induced at the substrate binding site of UBR box (Choi et al., 2010). This conformation change is speculated to trigger a cascade of conformational events leading to the allosteric activation of Cup9 binding site. As such, one important issue that

remains unresolved is how small molecule ligands regulate the function of N-recognins through the UBR box and the N-domain (**see 4.3**).

Putative N-recognin. The crystal structures of DIAP1 BIR2 (baculovirus IAP repeat 2) (Wu et al., 2001) domain to the N-terminally exposed IBM (IAP-binding motif) of caspases demonstrates that the BIR2-IBM interaction is similar to that of the UBR box with type1 residues. BIR2 binds the IBM on the surface of a hydrophobic groove formed by a ~70-residue zinc-finger motif. The binding with N-terminal Ala involves two hydrogen bonds between the N-terminal α-amino group and negatively charged BIR2 residues (Asp277 and Gln282), and two hydrogen bonds between the first peptide bond and surrounding BIR2 residues (Wu et al., 2001). The N-terminal side chain fits tightly in the binding pocket of BIR domain. As N-terminal Ala binds to BIR domain, the following six residues interact with the surface of the shallow groove. Therefore allowing the N-terminal Ala as an anchoring component in the BIR-IBM interaction, and functioning as an N-degron-like determinant (**Fig. 9**) (Wu et al., 2001). This action mode in ubiquitylation of caspases and IAP antagonists qualifies DIAP1 and probably other IAPs as N-recognin-like recognition components in regulated proteolysis (**Fig. 3**). However, it remains to be determined whether N-terminal Ala of the IBM is a strong degron (Tasaki, 2012).

4.3 FUNCTIONS AND IMPLICATIONS

Small molecule regulation. The *S. cerevisiae* homeodomain protein Cup9 inhibits the import of extracellular dipeptides and tripeptides through transcriptional repression of the peptide transporter *PTR2* gene (see **4.2**). The function of Cup9 in Ptr2-dependent peptide import is

regulated by its Ubr1-dependent proteolysis through the recognition of its I-degron (See **1.2**) (Hwang and Varshavsky, 2008; Varshavsky, 2011). Ubr1-dependent Cup9 degradation is allosterically activated by small peptides bearing destabilizing N-terminal residues, providing a positive feedback loop in which imported peptides bind to Ubr1 and accelerate Ubr1-dependent Cup9 proteolysis, leading to accelerated peptide import. Thus, peptide import in S. cerevisiae is tightly controlled through the allosteric activation of Ubr1 by small peptides. In addition, *S. cerevisiae* cells can also sense free amino acids, particularly Trp and Leu, in a nutrient-rich environment. It has been found that the availability of these extracellular free amino acids induces transcriptional induction of *PTR2* through Ubr1-dependent degradation of Cup9 linked with the amino acid sensing SPS (Ssy1-Ptr3-Ssy5) pathway (Varshavsky, 2011; Xia et al., 2008a). It is speculated that a similar UBR-regulated, amino acid sensing pathway might be present in multicellular eukaryotes (Varshavsky, 2011). This might be the case with trace amines targeting the N-recognins (See **3.2.2**) (**Fig. 23**). Similar to the feed back loop of amino acids/peptides which bind to yeast Ubr1 to induce the peptide transporter expression, trace amines when bound to N-domain of mammalian UBR proteins might induce neurotransmitter transporters on plasma membranes or other transporters on ER membrane in neurons (Tasaki, 2012).

Acetylation. In eukaryotes, majority (50-80%) of the proteins are acetylated at the N-terminal α-amino group by N-terminal acetyltransferases that are associated with the ribosomes (Arnesen, 2011). N-terminal acetylation (Nt-acetylation) is irreversible and typically occurs cotranslationally at the retained N-terminal Met or a newly exposed N-terminal residue (Ala, Val, Ser, Thr or Cys) after the N-terminal Met is constitutively removed by Met aminopeptidases (MetAPs) (Polevoda and Sherman, 2003). Since its discovery half a century ago, the function of

this abundant modification remained elusive (Narita, 1958). It was generally assumed that the acetyl group caps the exposed N-terminal residue and thus protects a protein from an undesirable modification or degradation or, in a few proteins, is involved in protein-protein interactions and translocation to the membrane, but rarely in a ubiquitin-dependent manner (Mayer et al., 1989). Interestingly, a recent study showed that at least in *S. cerevisiae*, an acetylated N-terminal residue, including the retained initiator Met, could act as an N-degron, thereby functioning as an alternative signal to initiate the N-end rule pathway (See **1.2**) (Hwang et al., 2010b). This discovery that most cellular proteins are produced with potential N-degrons raises a question on the biological function of acetylation-mediated proteolysis. Doa10 is a transmembrane protein in the endoplasmic reticulum (ER) and an ERAD (ER associated degradation) component that mediates ubiquitylation of misfolded proteins on the cytosolic surface of the ER membrane (**Fig. 3**) (Deng and Hochstrasser, 2006; Ravid et al., 2006; Swanson et al., 2001). Cellular stresses impair folding of nascent proteins in the ER lumen, activating the unfolded protein response (UPR). To alleviate ER stress, the UPR removes misfolded proteins from the ER lumen in concert with accelerated proteolysis via the ERAD and reduces the protein load into the ER lumen by phosphorylating eukaryotic translation initiation factor 2α (eIF2α), which slows down global protein synthesis by ribosomes. Thus, Doa10-mediated degradation of acetylated proteins may contribute to protein quality control by reducing the protein input into the stressed ER lumen. Indeed, a recent study showed that N-terminal acetylation inhibits protein targeting to the ER (Forte et al., 2011). The consequence of this acetylation-mediated inhibition would be the retention of proteins with signal peptides in the cytosol, which may subsequently be eliminated through the recognition of their acetylated N-degrons (Sriram et al., 2011).

UFD collaboration. The UFD and N-end rule pathways have been characterized independently for more than two decades, with no obvious functional interactions with each other(Johnson et al., 1992; Johnson et al., 1995). Interestingly, now it has been demonstrated that these two distinct recognition components for the N-end rule pathway and the UFD pathway, form a complex and synergistically mediate ubiquitylation for both pathways (See **1.2**) (Hwang et al., 2010a). What is the purpose of this co-operation between the two pathways? Numerous processes occur on the spatially limited cytosolic surface of the ER, including protein synthesis, transportation and degradation of misfolded proteins of the cytosol and those retrotranslocated from the ER lumen. One unique feature of the N-end rule pathway is its ability to mediate cotranslational modification and ubiquitylation, which is thought to be required for quick decision-making regarding the metabolic stability of a substrate. As a nascent N-end rule substrate emerges from the ribosome at the rate of 10-20 amino acids per second, an N-degron should be targeted within ~2-3 secs (equivalent to 20-60 residues) to ensure cotranslational degradation before the polypeptide gets folded. The N-end rule degradation involves a multi-step interaction with an N-degron, a search for the Lys residue as an ubiquitylation site, ubiquitin chain growth, and proteasomal degradation, possibly through an adaptor linking polyubiquitylation to proteolysis. In this process, the N-end rule interaction and possibly addition of one or two ubiquitin molecules are crucial for substrate specificity, leaving the remaining steps of polyubiquitylation and degradation nonselective. As such, it is not surprising that the two ligases cooperate with each other in nonselective ubiquitin chain growth and, possibly, to prioritize proteasomal recognition (Sriram et al., 2011).

4.4 FRAGMENT BASED DRUG DESIGN

Multivalency can be looked at as an approach to convert low affinity monovalent bindings to high affinity multivalent interaction (See **3.1**). Obviously this would translate into highly desired situation of higher therapeutic efficacy at a lower dose, however it should be noted that in some cases (like high molecular weight ligands) multivalency can lead to complications in absorption, distribution, metabolism and elimination (ADME) profile, due to their pharmacokinetic and pharmacodynamic characteristics. But still multivalency can be a potential tool for rational drug design like fragment based drug design, where a new monovalent ligand can be optimized to yield a medium affinity ligand or use existing monovalent ligand with low affinity to create a multivalent ligand that could yield higher affinity through stronger interactions (Sriram et al., 2009).

The concept of multivalency is successfully used in fragment-based drug discovery discovery (FBDD) (Congreve et al., 2008), where a functional drug with high affinity and selectivity is synthesized or screened in smaller pieces that have low affinity and selectivity. In this approach, initial high throughput screening identifies simple molecular fragments, which usually are small (120-250 Da) and of weak affinity (K_d, 10 μM-mM). However, some of the resulting fragment hits may have high unit affinity per atom, and the combination of these monovalent molecules may yield a drug-like compound with high selectivity and affinity to the target, thermodynamically (enhanced binding affinity) and kinetically (reduced dissociation rate). The concept of multivalency has been recently adopted in FBDD, and we are now witnessing a number of compounds entering into phase II clinical trials (Congreve et al., 2008). The concept of heterodivalency also may be exploited in drug repositioning, an approach to develop new use for an existing drug, in which two appropriate drugs are linked to yield higher

efficacy or lower adverse effects, provided that tethering of the drugs does not adversely affect the pharmacokinetic properties (Sriram et al., 2009). Future strategy includes identifying appropriate target molecules, which will require advances in structural and functional understanding on biological interactions. The linkers and ligands will need to be optimized in cell penetration, solubility and *in vivo* stability. New thermodynamic models may be needed to better explain the interactions of the linkers and ligands with themselves and other molecules within the cell.

4.5 VISION AND FUTURE DIRECTION

While a number of physiological substrates of the N-end rule pathway have been discovered, many more are likely to remain unknown. According to the 'rule book', all but 18 of the 20 amino acids (except Pro and Gly) listed in the genetic code can have destabilizing activities (**Fig 2**), yet the number of identified substrates are by far underrepresented (**Table 2**). Recent studies have suggested that the UBR box of mammalian has been structurally optimized to interact with the N-terminal Arg compared with other type 1 residues (see **4.1**). This raises an intriguing possibility that selective degradation by the N-end rule pathway may be more relevant or confined to the arginylation branch. It is urgent to establish a reproducible assay system to screen substrates of the N-end rule pathway (Sriram et al., 2011).

Studies in the recent years have revealed that the pathway has a broader than expected functions of selective proteolysis of a few substrates to the extended functions of quality control of misfolded proteins, acetylation-based pathway, cooperation with other pathways (like UFD), flooding survival in plants (see **4.0**). As the pathway's scope is rapidly expanding, it is pressing

to link cellular functions to specific substrates whose activities are controlled through the destabilizing activities of N-degrons. And to explore the possibility of N-recognins collaborating with other systems like autophagy in cellular stress, transport mechanisms in neurons, etc (Tasaki, 2012).

The structure of the N-domain is unknown. It is anticipated that the binding of small molecule ligands to the UBR box and the N-domain of canonical N-recognins synergistically induce a conformational change which exposes a site that recognizes internal degrons of substrates (see **4.2** and **4.3**). In mammals, multiple N-recognins share enzymatic specificities to N-terminal residues but differ in cellular functions, perhaps through a mechanism that should be understood at the level of organs and tissue-specific cell types. The structures of UBR proteins containing both the UBR box and the N-domain will be required to better understand how the N-end rule ligands regulate the function of N-recognins (Sriram et al., 2011).

It is not surprising that misregulation of the N-end rule pathway results in pathophysiological conditions in animals, such as defects in cardiac development, angiogenesis, meiosis, DNA repair, and secretion of pancreatic enzymes (**Fig 5**). Some of these mimic the symptoms of human genetic diseases, such as Johanson-Blizzard Syndrome that is causally linked to mutations in UBR1. It is therefore increasingly urgent to develop potent and specific inhibitors of the pathway. One promising way to design such inhibitors is to employ FBDD in the N-end rule pathway (see **4.4**) (Sriram et al., 2009). Future strategy includes identifying appropriate target molecules, which will require advances in structural and functional understanding on biological interactions. The linkers and ligands will need to be optimized in cell penetration, ADME and *in vivo* stability.

APPENDIX A

BIOENERGETICS EQUATIONS

The inherent characteristics of multivalent interactions can be mainly attributed to the thermodynamic and the kinetic aspects of ligand-receptor binding. These two aspects are mutually related and therefore have a considerable overlap. Bioenergetics can be principally used to understand the difference between mono and multivalent interactions and kinetic aspects can be used as a quantitative tool and by combining both will enable us to design better multivalent ligands.

Applying Gibbs free energy equation to binding we can study the thermodynamics of mono and multivalent interactions.

$$\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ} \qquad (1)$$

where ΔG° is the standard change in free energy, which is a measure of available energy in a form that can be used to do work, by convention negative value for ΔG indicates a spontaneous reaction or a binding event and the magnitude of the negative value indicated the extent of spontaneity. ΔH°, standard change in enthalpy reflects the change in heat content of the binding interaction and similarly by convention negative value indicates a energetically favorable, spontaneous binding event. T is the absolute temperature and is considered a constant. ΔS° is a

quantitative expression for the standard change in disorder or randomness involved in a binding event and by convention a positive value or more randomness indicates a stable binding event.

Mammen, M., et al (Mammen, 1998) in their review have described in great detail the thermodynamic nature of multivalent interactions, borrowing from their concept based on free energy of binding we can explain the mono and multivalent interactions as:

$$\begin{aligned} \Delta G^{mono} &= \Delta H^{mono} - T\Delta S^{mono} \\ \Delta G_n^{multi} &= \Delta H_n^{multi} - T\Delta S_n^{multi} \end{aligned} \tag{2,3}$$

where n is n^{th} ordered multivalent interaction involving n ligands and n receptors, mono represents single and multi represents n-multivalent interactions.

As free energy is equated using enthalpy and entropy components, we can look at these components to understand the difference between mono and multivalent change in free energies. If we consider the change in enthalpy, since it is the change in heat content during unstrained, independent binding event we can approximate;

$$n.\Delta H^{mono} = \Delta H_n^{multi} \tag{4}$$

Entropy is mainly contributed by changes in translational (ΔS_{trans}), rotational (ΔS_{rot}) and conformational (ΔS_{conf}) components.

$$\Delta S \approx \Delta S_{trans} + \Delta S_{rot} + \Delta S_{conf} \tag{5}$$

ΔS_{trans} of a molecule is the result of its freedom to translate independently in space, i.e, it is a measure of number of possible arrangements of a molecule in a given space and this depends on its mass and concentration as $\Delta S_{trans} \alpha \ln(M) \,\&\, \ln(conc)^{-1}$, ΔS_{rot} is the freedom to rotate around its three axes, i.e., the possible number of different rotational positions available to the molecule

and is proportional to its moment of inertia $\Delta S_{rot} \alpha \ln(I_x.I_y.I_z)$ and ΔS_{conf} is the freedom to have different conformations and depends on the flexibility of the linker.

If we consider the contribution of translational and rotational entropies of particles at a given concentration, then it is approximated to be same irrespective of the particles being mono or multivalent, as these two entropies are weakly (logarithmically) dependent on their mass and size and then also by considering that there is a net loss in free translational and rotational entropies of these multivalent particles when multiple particles associate and become one particle such that $\Delta S_{trans}^{multi} = \Delta S_{rot}^{multi} = 0$ and by considering the effect of linker after binding on conformational entropy, $\Delta S_{n\,conf}^{multi}$, we have;

$$\begin{aligned} \Delta S^{mono} &\approx \Delta S_{trans}^{mono} + \Delta S_{rot}^{mono} \\ \Delta S_n^{multi} &\approx \Delta S_{trans}^{mono} + \Delta S_{rot}^{mono} + \Delta S_{n\,conf}^{multi} \\ \Delta S_n^{multi} &\approx \Delta S^{mono} + \Delta S_{n\,conf}^{multi} \end{aligned} \qquad (6,7,8)$$

Now to understand the thermodynamic advantage of multivalent interaction over monovalent interaction we can look at the difference in change of free energies of these two interactions, i.e., one with multivalent interactions with n ligands and n receptors to form a single complex and second with n monovalent, independent interaction with n copies of ligand and receptor, therefore:

$$\begin{aligned} n\Delta G^{mono} &= n\Delta H^{mono} - nT\Delta S^{mono} \\ \Delta G_n^{multi} &= \Delta H_n^{multi} - T\Delta S_n^{multi} \\ \Delta G_n^{multi} &= n.\Delta H^{mono} - T\Delta S^{mono} - T\Delta S_{n\,conf}^{multi} \\ \Delta G_{\Delta}^{multi-n.mono} &= (n-1)T\Delta S^{mono} - T\Delta S_{n\,conf}^{multi} \end{aligned} \qquad (9.10,11,12)$$

Therefore by looking at $\Delta G_{\Delta}^{multi-n.mono}$, which is a term for the free energy advantage offered by multivalent over the same number of monovalent interactions, it can be suggested that

overall enhancement in multivalent association is mostly due to the entropic contribution of monovalent binding and that multivalent binding is thermodynamically favorable as it has lower cost of entropic penalty than the entropic cost of same number of multiple individual monovalent associations.

Similarly by looking at $\Delta G_{\Delta}^{multi-mono}$, i.e the difference between change in free energies of multivalent and a given single monovalent interaction, we have;

$$\Delta G_{\Delta}^{multi-mono} = (n-1)\Delta H^{mono} - T\Delta S_{n\,conf}^{multi} \qquad (13)$$

Suggesting, the contribution of monovalent enthalpic component towards the free energy change of multivalent interactions and the significance of conformational entropic component that can lead to negative effect on multivalent interaction. Overall the thermodynamic aspect of multivalent interaction suggests that a higher valency number (n), stronger monovalent association (ΔH^{mono}) and conformation after binding are critical in multivalent interactions and an overall total between penalties and bonuses between these parameters will yield a favorable multivalent effect.

An example to illustrate this multivalent effect is the thermodynamic study of trivalent vancomycin molecule to trivalent D-Ala-D-Ala molecule, where the monovalent K_d is 1.6 µM and the trivalent yielded a K_d of 0.04 fM, i.e., 10^{11} times stronger than its monovalent or 25 times stronger than one of the strongest known biotin-streptavidin interaction (Kd~10^{-15} M) (Choi, 2004).It was found that for trivalent system ΔG^{tri} = -94 kJ/mol, ΔH^{tri} = -167 kJ/mol and the calculated $T\Delta S^{tri}$ = -73 kJ/mol. Similarly for the corresponding monovalent system ΔG^{mono} = -33 kJ/mol, ΔH^{mono} = -50.2 kJ/mol and $T\Delta S^{mono}$ = -17.2 kJ/mol. So we can see that

$3.\Delta H^{mono}(-150.6kJ/mol) \cong \Delta H^{tri}(-167kJ/mol)$, also from

$T\Delta S^{tri} \approx T\Delta S^{mono} + T\Delta S^{tri}_{conf}$, we can say that $T\Delta S^{tri}_{conf}$ = -56 kJ/mol.

Therefore we can calculate $\Delta G^{tri-mono}_{\Delta} = (3-1)\Delta H^{mono} - T\Delta S^{tri}_{conf}$ = -44.4 kJ/mol, which is close to experimental $\Delta G^{tri-mono}_{\Delta}$= -66 kJ/mol. Also the difference of -56 kJ/mol which is the entropic penalty on trivalent system that is attributed to $T\Delta S^{tri}_{conf}$ is explained as a result of restrained conformational state after formation of trivalent complex and due to the linker.

Similarly for a divalent vancomycin and D-Ala-D-Ala peptide ligand system, the $\Delta G^{bi-mono}_{\Delta}$= -3.9 kJ/mol, which is the additional contribution of second ligand to bivalent system, but with divalent K_d being ~10^3 times stronger than monovalent K_d, which points at an additional contributing factor other than $\Delta G^{bi-mono}_{\Delta}$ that contributes to such a high K_d difference, which was found to be high effective concentration of the second ligand ~ 20mM in this bivalent system favoring binding. Applying n=2 in $\Delta G^{multi-mono}_{\Delta} = (n-1)\Delta H^{mono} - T\Delta S^{multi}_{nconf}$,

We have $\Delta G^{bi-mono}_{\Delta} = \Delta H^{mono} - T\Delta S^{bi}_{conf}$,

i.e., the additional contribution to from the second ligand towards favorable interaction from a bivalent ligand is a net effect of monovalent enthalpy and entropic conformational penalty. Therefore to increase this $\Delta G^{bi-mono}_{\Delta}$ term, i.e., to enhance the favorability of bivalent interaction we can increase the affinity of our monovalent ligands, thus increasing the ΔH^{mono} component and reduce the cost of conformational entropic penalty by exploiting the linker, which moreover also influences the effective concentration. We also found similar bivalent effects with our RF-C11. Our model of hetero-bivalent inhibitor corresponds with the above concepts, we think the enhancement in inhibitory potential of our heterobivalent molecule over monovalent is an affect of affinity, effective concentration and cooperation.

Kinetic perspective of multivalency gain over monovalency has been explained as a result of decreased dissociation rate k_{off} of a multivalent interaction, instead of an increase in the association rate k_{on}. As is evident from trivalent vancomycin with monovalent vancomycin to D-Ala-D-Ala peptide, where the ratio of k_{off} for tri to mono is about 7 and the ratio of k_{off} is about 10^9. Also kinetics of anti-DNP (dinitrophenyl), hapten antibody binding to DNP-lysine vs. anti-DNP antibody to DNP covered surface of ΦX174 bacteriophage, shows that the k_{on} varies by a meager factor of 2, where as the k_{off} varies by a factor of 10^4.

Kramer and Karpen (Kramer and Karpen, 1998) have related the dissociation rate at the second site of a bivalent interaction through the dissociation rate of a bound monovalent ligand. Elucidating heterobivalency is more challenging than homobivalency however both hetero and homo valent compounds share the same governing principle. So if we approximate that the rates for both sites for heterobivalency are quantitatively similar, then the dissociation of a bound heterobivalent ligand occurs in two stages and the dissociation rate at the second site can be expressed in terms of dissociation rate of monovalent ligand multiplied by a statistical factor of 2 as dissociation can happen at either of the two equivalent sites and the probability that the second site is not occupied:

$$k_{off}^{II} = 2k_{off}^{I}.{p_0}^{II}$$

Probability of a binding site not to be occupied is given by:

$$p_0^{II} = \frac{K_d}{K_d + C_{total}}$$

where K_d is the dissociation constant of the monovalent ligand and C_{total} is the total concentration of a ligand at that site. Total concentration C_{total} has been calculated as effective concentration C_{eff}, in a hemispherical volume defined by the linker length (r), using Avogadro's number (N_A).

$$C_{eff} = \frac{1}{N_A.vol} = \frac{1}{N_A.\frac{2}{3}\pi.r^3}$$

Concept of effective concentration has been borrowed from polymer chemistry, where it was utilized for incorporating intramolecular cyclization reactions in synthesis as a theoretical parameter for estimating the rate of a ring closure reaction, considering a linker acting as a random coil polymer between the reactive groups. Lees and co workers have extended this concept to binding, C_{eff} in multivalent interactions has be employed as a probability term for two ends of a ligand to be within a given distance, i.e, the chance that one entity would experience the presence of another counterpart in its proximity. Therefore C_{eff} can be utilized for predicting an optimal length for a linker that could provide maximum binding enhancement.

Here C_{eff} is defined using a volume term involving root mean square distance $\left\langle r^2 \right\rangle^{1/2}$, which is dependent on the flexibility and length of the linker (a), this term is a characteristic of a given linker and on the number of units making up that linker (n).

$$\left\langle r^2 \right\rangle^{1/2} = a\sqrt{n}$$

$$C_{eff} = \frac{1}{N_A.\left(\left\langle r^2 \right\rangle^{1/2}\right)^3}\left(\frac{3}{2\pi}\right)^{3/2}$$

As it has been suggested in the preceding sections, the multivalent effect seen through thermodynamic aspect is attributed to the entropic barrier and through kinetic aspect is due to the effective concentration. Contributing to these two parameters is the role of the linker, thus designing an ideal linker is a key attribute to multivalent design. Two critical characteristics of a linker are its length and stiffness. Although there is no black or white form, but it a balanced

optimal form that seems to offer the best results. General intuition regarding stiffness of a linker leans towards a rigid linker, as it will contribute less to the conformational entropic penalty, however since it is rigid it might not reach out to the binding site efficiently hence causing unfavorable interactions. Where as a flexible linker can facilitate binding as the linker allows sampling of the conformational space, so decreasing steric obstruction, but can also increase the conformational entropic penalty and depending on the hydrophobic nature of the linker it can lead to unfavorable interaction with receptor. An ideal linker will have an optimal flexibility offering a favorable effect. Similarly with linker length, an optimal length based on the inter-receptor distance should be designed, it was reported that for a rigid linker this length should match the inter-receptor distance which is practically difficult, however a flexible linker would allow for some leniency in this matter

BIBLIOGRAPHY

An, J.Y., Kim, E.A., Jiang, Y., Zakrzewska, A., Kim, D.E., Lee, M.J., Mook-Jung, I., Zhang, Y., and Kwon, Y.T. (2010). UBR2 mediates transcriptional silencing during spermatogenesis via histone ubiquitination. Proceedings of the National Academy of Sciences of the United States of America *107*, 1912-1917.

Arnesen, T. (2011). Towards a functional understanding of protein N-terminal acetylation. PLoS biology *9*, e1001074.

Bachmair, A., Becker, F., and Schell, J. (1993). Use of a reporter transgene to generate arabidopsis mutants in ubiquitin-dependent protein degradation. Proceedings of the National Academy of Sciences of the United States of America *90*, 418-421.

Bachmair, A., Finley, D., and Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. Science *234*, 179-186.

Bachmair, A., and Varshavsky, A. (1989). The degradation signal in a short-lived protein. Cell *56*, 1019-1032.

Baker, R.T., and Varshavsky, A. (1995). Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. The Journal of biological chemistry *270*, 12065-12074.

Balzi, E., Choder, M., Chen, W.N., Varshavsky, A., and Goffeau, A. (1990). Cloning and functional analysis of the arginyl-tRNA-protein transferase gene ATE1 of *Saccharomyces cerevisiae*. The Journal of biological chemistry *265*, 7464-7471.

Banerjee, A.L., Eiler, D., Roy, B.C., Jia, X., Haldar, M.K., Mallik, S., and Srivastava, D.K. (2005). Spacer-based selectivity in the binding of "two-prong" ligands to recombinant human carbonic anhydrase I. Biochemistry *44*, 3211-3224.

Basha, S., Rai, P., Poon, V., Saraph, A., Gujraty, K., Go, M.Y., Sadacharan, S., Frost, M., Mogridge, J., and Kane, R.S. (2006). Polyvalent inhibitors of anthrax toxin that target host receptors. Proceedings of the National Academy of Sciences of the United States of America *103*, 13509-13513.

Burchett, S.A., and Hicks, T.P. (2006). The mysterious trace amines: Protean neuromodulators of synaptic transmission in mammalian brain. Progress in Neurobiology *79*, 23.

Byrd, C., Turner, G.C., and Varshavsky, A. (1998). The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. The EMBO journal *17*, 269-277.

Choi, S.K. (2004). Synthetic Multivalent Molecules: Concepts and Biomedical Applications (Hoboken, John Wiley & Sons).

Choi, S.K., Mammen, M., and Whitesides, G.M. (1996). Monomeric inhibitors of influenza neuraminidase enhance the hemagglutination inhibition activities of polyacrylamides presenting multiple C-sialoside groups. Chemistry & biology *3*, 97-104.

Choi, W.S., Jeong, B.C., Joo, Y.J., Lee, M.R., Kim, J., Eck, M.J., and Song, H.K. (2010). Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. Nature structural & molecular biology *17*, 1175-1181.

Congreve, M., Chessari, G., Tisi, D., and Woodhead, A.J. (2008). Recent developments in fragment-based drug discovery. J Med Chem *51*, 3661-3680.

Deng, M., and Hochstrasser, M. (2006). Spatially regulated ubiquitin ligation by an ER/nuclear membrane ligase. Nature *443*, 827-831.

Ditzel, M., Broemer, M., Tenev, T., Bolduc, C., Lee, T.V., Rigbolt, K.T., Elliott, R., Zvelebil, M., Blagoev, B., Bergmann, A., *et al.* (2008). Inactivation of effector caspases through nondegradative polyubiquitylation. Molecular cell *32*, 540-553.

Ditzel, M., and Meier, P. (2005). Ubiquitylation in apoptosis: DIAP1's (N-)en(d)igma. Cell death and differentiation *12*, 1208-1212.

Ditzel, M., Wilson, R., Tenev, T., Zachariou, A., Paul, A., Deas, E., and Meier, P. (2003). Degradation of DIAP1 by the N-end rule pathway is essential for regulating apoptosis. Nature cell biology *5*, 467-473.

Dohmen, R.J., Madura, K., Bartel, B., and Varshavsky, A. (1991). The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme. Proceedings of the National Academy of Sciences of the United States of America *88*, 7351-7355.

Dougan, D.A., Truscott, K.N., and Zeth, K. (2010). The bacterial N-end rule pathway: expect the unexpected. Molecular microbiology *76*, 545-558.

Du, F., Navarro-Garcia, F., Xia, Z., Tasaki, T., and Varshavsky, A. (2002). Pairs of dipeptides synergistically activate the binding of substrate by ubiquitin ligase through dissociation of its autoinhibitory domain. Proceedings of the National Academy of Sciences of the United States of America *99*, 14110-14115.

Dyall, S.D., Brown, M.T., and Johnson, P.J. (2004). Ancient invasions: from endosymbionts to organelles. Science *304*, 253-257.

Eisele, F., and Wolf, D.H. (2008). Degradation of misfolded protein in the cytoplasm is mediated by the ubiquitin ligase Ubr1. FEBS letters *582*, 4143-4146.

Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D.A., and Bukau, B. (2006). ClpS is an essential component of the N-end rule pathway in *Escherichia coli*. Nature *439*, 753-756.

Erlanson, D.A., Wells, J.A., and Braisted, A.C. (2004). Tethering: fragment-based drug discovery. Annu Rev Biophys Biomol Struct *33*, 199-223.

Fleckenstein, A.F., Volz, T.J., Riddle, E.L., Gibb, J.W., Hanson, G.R. (2007). New insights into the mechanism of action of amphetamines. Annu Rev Pharmacol Toxicol *47*, 17.

Forte, G.M., Pool, M.R., and Stirling, C.J. (2011). N-terminal acetylation inhibits protein targeting to the endoplasmic reticulum. PLoS biology *9*, e1001073.

Garzón, M., Eifler, K., Faust, A., Scheel, H., Hofmann, K., Koncz, C., Yephremov, A., and Bachmair, A. (2007). *PRT6*/At5g02310 encodes an *Arabidopsis* ubiquitin ligase of the N-end rule pathway with arginine specificity and is not the *CER3* locus. FEBS letters *581*, 3189-3196.

Gibbs DJ, L.S., Isa NM, Gramuglia S, Fukao T, et al. (2011). Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. Nature *479*, 3.

Gonda, D.K., Bachmair, A., Wunning, I., Tobias, J.W., Lane, W.S., and Varshavsky, A. (1989). Universality and structure of the N-end rule. The Journal of biological chemistry *264*, 16700-16712.

Graciet, E., Hu, R.G., Piatkov, K., Rhee, J.H., Schwarz, E.M., and Varshavsky, A. (2006). Aminoacyl-transferases and the N-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen. Proceedings of the National Academy of Sciences of the United States of America *103*, 3078-3083.

Grigoryev, S., Stewart, A.E., Kwon, Y.T., Arfin, S.M., Bradshaw, R.A., Jenkins, N.A., Copeland, N.G., and Varshavsky, A. (1996). A mouse amidase specific for N-terminal asparagine. The gene, the enzyme, and their function in the N-end rule pathway. The Journal of biological chemistry *271*, 28521-28532.

Guo, F., Esser, L., Singh, S.K., Maurizi, M.R., and Xia, D. (2002). Crystal structure of the heterodimeric complex of the adaptor, ClpS, with the N-domain of the AAA+ chaperone, ClpA. The Journal of biological chemistry *277*, 46753-46762.

Heck, J.W., Cheung, S.K., and Hampton, R.Y. (2010). Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. Proceedings of the National Academy of Sciences of the United States of America *107*, 1106-1111.

Houk, K.N., Leach, A.G., Kim, S.P., and Zhang, X. (2003). Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes. Angew Chem Int Ed Engl *42*, 4872-4897.

Hu, R.G., Sheng, J., Qi, X., Xu, Z., Takahashi, T.T., and Varshavsky, A. (2005). The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. Nature *437*, 981-986.

Huskens, J. (2006). Multivalent interactions at interfaces. Curr Opin Chem Biol *10*, 537-543.

Hwang, C.S., Shemorry, A., Auerbach, D., and Varshavsky, A. (2010a). The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. Nature cell biology *12*, 1177-1185.

Hwang, C.S., Shemorry, A., and Varshavsky, A. (2010b). N-terminal acetylation of cellular proteins creates specific degradation signals. Science *327*, 973-977.

Hwang, C.S., and Varshavsky, A. (2008). Regulation of peptide import through phosphorylation of Ubr1, the ubiquitin ligase of the N-end rule pathway. Proceedings of the National Academy of Sciences of the United States of America *105*, 19188-19193.

Johnson, E.S., Bartel, B., Seufert, W., and Varshavsky, A. (1992). Ubiquitin as a degradation signal. The EMBO journal *11*, 497-505.

Johnson, E.S., Ma, P.C., Ota, I.M., and Varshavsky, A. (1995). A proteolytic pathway that recognizes ubiquitin as a degradation signal. The Journal of biological chemistry *270*, 17442-17456.

Kane, R.S. (2010). Thermodynamics of Multivalent Interactions: Influence of the Linker. Langmuir *26*, 4.

Kiessling, L.L., Gestwicki, J.E., and Strong, L.E. (2000). Synthetic multivalent ligands in the exploration of cell-surface interactions. Curr Opin Chem Biol *4*, 696-703.

Kiessling, L.L., Gestwicki, J.E., and Strong, L.E. (2006). Synthetic multivalent ligands as probes of signal transduction. Angew Chem Int Ed Engl *45*, 2348-2368.

Kitov, P.I., and Bundle, D.R. (2003). On the nature of the multivalency effect: a thermodynamic model. J Am Chem Soc *125*, 16271-16284.

Kramer, R.H., and Karpen, J.W. (1998). Spanning binding sites on allosteric proteins with polymer-linked ligand dimers. Nature *395*, 710-713.

Krishnamurthy, V.M., Estroff, L.A., and Whitesides, G.M. (2006). Multivalency in ligand design- Fragement Based Approaches in Drug Design (Weinheim, Germany, Wiley-VCH Verlag Gmbh & Co).

Krishnamurthy, V.M., Semetey, V., Bracher, P.J., Shen, N., and Whitesides, G.M. (2007). Dependence of effective molarity on linker length for an intramolecular protein-ligand system. J Am Chem Soc *129*, 1312-1320.

Kroncke, K.D., and Klotz, L.O. (2009). Zinc fingers as biologic redox switches? Antioxidants & redox signaling *11*, 1015-1027.

Kwon, Y.T., Balogh, S.A., Davydov, I.V., Kashina, A.S., Yoon, J.K., Xie, Y., Gaur, A., Hyde, L., Denenberg, V.H., and Varshavsky, A. (2000). Altered activity, social behavior, and spatial memory in mice lacking the NTAN1p amidase and the asparagine branch of the N-end rule pathway. Molecular and cellular biology *20*, 4135-4148.

Kwon, Y.T., Kashina, A.S., Davydov, I.V., Hu, R.G., An, J.Y., Seo, J.W., Du, F., and Varshavsky, A. (2002). An essential role of N-terminal arginylation in cardiovascular development. Science *297*, 96-99.

Kwon, Y.T., Kashina, A.S., and Varshavsky, A. (1999). Alternative splicing results in differential expression, activity, and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. Molecular and cellular biology *19*, 182-193.

Lee, J., Kim, JS., Seok, C. (2010). Cooperativity and Specificity of Cys2His2 Zinc Finger Protein. The Journal of Physical Chemistry B *114*, 9.

Lee, M.J., Pal, K., Tasaki, T., Roy, S., Jiang, Y., An, J.Y., Banerjee, R., and Kwon, Y.T. (2008). Synthetic heterovalent inhibitors targeting recognition E3 components of the N-end rule pathway. Proceedings of the National Academy of Sciences of the United States of America *105*, 100-105.

Lee, M.J., Tasaki, T., Moroi, K., An, J.Y., Kimura, S., Davydov, I.V., and Kwon, Y.T. (2005). RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. Proceedings of the National Academy of Sciences of the United States of America *102*, 15030-15035.

Li, J., and Pickart, C.M. (1995). Binding of phenylarsenoxide to Arg-tRNA protein transferase is independent of vicinal thiols. Biochemistry *34*, 15829-15837.

Licausi F, K.M., Weits DA, Giuntoli B, Giorgi FM, et al. (2011). Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. Nature *479*, 4.

Maeda, D.Y., Mahajan, S.S., Atkins, W.M., and Zebala, J.A. (2006). Bivalent inhibitors of glutathione S-transferase: the effect of spacer length on isozyme selectivity. Bioorg Med Chem Lett *16*, 3780-3783.

Maly, D.J., Choong, I.C., and Ellman, J.A. (2000). Combinatorial target-guided ligand assembly: identification of potent subtype-selective c-Src inhibitors. Proceedings of the National Academy of Sciences of the United States of America *97*, 2419-2424.

Mammen, M., Choi, S.K., and Whitesides, G.M. (1998). Polyvalent interactions in biological systems:implications for design and use of multivalent ligands and inhibitors. . Angew Chem Int Ed Engl *37*, 40.

Matta-Camacho, E., Kozlov, G., Li, F.F., and Gehring, K. (2010). Structural basis of substrate recognition and specificity in the N-end rule pathway. Nature structural & molecular biology *17*, 1182-1187.

Mayer, A., Siegel, N.R., Schwartz, A.L., and Ciechanover, A. (1989). Degradation of proteins with acetylated amino termini by the ubiquitin system. Science *244*, 1480-1483.

Mogk, A., Schmidt, R., and Bukau, B. (2007). The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. Trends in cell biology *17*, 165-172.

Narita, K. (1958). Isolation of acetylpeptide from enzymic digests of TMV-protein. Biochimica et biophysica acta *28*, 184-191.

Ninnis, R.L., Spall, S.K., Talbo, G.H., Truscott, K.N., and Dougan, D.A. (2009). Modification of PATase by L/F-transferase generates a ClpS-dependent N-end rule substrate in *Escherichia coli*. The EMBO journal *28*, 1732-1744.

Orme, M., and Meier, P. (2009). Inhibitor of apoptosis proteins in *Drosophila*: gatekeepers of death. Apoptosis : an international journal on programmed cell death *14*, 950-960.

Pabo, C.O., Peisach, E., and Grant, R.A. (2001). Design and selection of novel Cys2His2 zinc finger proteins. Annual review of biochemistry *70*, 313-340.

Peltier, J.B., Ripoll, D.R., Friso, G., Rudella, A., Cai, Y., Ytterberg, J., Giacomelli, L., Pillardy, J., and van Wijk, K.J. (2004). Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications. The Journal of biological chemistry *279*, 4768-4781.

Pickart, C.M. (2001). Mechanisms underlying ubiquitination. Annual review of biochemistry *70*, 503-533.

Polevoda, B., and Sherman, F. (2003). N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. Journal of molecular biology *325*, 595-622.

Potuschak, T., Stary, S., Schlogelhofer, P., Becker, F., Nejinskaia, V., and Bachmair, A. (1998). PRT1 of Arabidopsis thaliana encodes a component of the plant N-end rule pathway. Proceedings of the National Academy of Sciences of the United States of America *95*, 7904-7908.

Rao, H., Uhlmann, F., Nasmyth, K., and Varshavsky, A. (2001). Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. Nature *410*, 955-959.

Rao, J., and Whitesides, G.M. (1997). Tight binding of a dimeric derivative of vancomycin. Journal of American Chemical Society *119*, 4.

Rao, J., Lahiri, J., Isaacs, L., Weis, R.M., and Whitesides, G.M. (1998). A trivalent system from vancomycin.D-ala-D-Ala with higher affinity than avidin.biotin. Science *280*, 708-711.

Ravid, T., Kreft, S.G., and Hochstrasser, M. (2006). Membrane and soluble substrates of the Doa10 ubiquitin ligase are degraded by distinct pathways. The EMBO journal *25*, 533-543.

Roman-Hernandez, G., Grant, R.A., Sauer, R.T., and Baker, T.A. (2009). Molecular basis of substrate selection by the N-end rule adaptor protein ClpS. Proceedings of the National Academy of Sciences of the United States of America *106*, 8888-8893.

Sabatini, D.M., Erdjument-Bromage, H., Lui, M., Tempst, P., and Snyder, S.H. (1994). RAFT1: a mammalian protein that binds to FKBP12 in a rapamycin-dependent fashion and is homologous to yeast TORs. Cell *78*, 35-43.

Sasaki, T., Kojima, H., Kishimoto, R., Ikeda, A., Kunimoto, H., and Nakajima, K. (2006). Spatiotemporal regulation of c-Fos by ERK5 and the E3 ubiquitin ligase UBR1, and its biological role. Molecular cell *24*, 63-75.

Schmidt, R., Zahn, R., Bukau, B., and Mogk, A. (2009). ClpS is the recognition component for *Escherichia coli* substrates of the N-end rule degradation pathway. Molecular microbiology *72*, 506-517.

Sriram, S.M., Banerjee, R., Kane, R.S., and Kwon, Y.T. (2009). Multivalency-assisted control of intracellular signaling pathways: application for ubiquitin- dependent N-end rule pathway. Chemistry & biology *16*, 121-131.

Sriram, S.M., Kim, B.Y., and Kwon, Y.T. (2011). The N-end rule pathway: emerging functions and molecular principles of substrate recognition. Nature reviews Molecular cell biology *12*, 735-747.

Sriram, S.M., and Kwon, Y.T. (2010). The molecular principles of N-end rule recognition. Nature structural & molecular biology *17*, 1164-1165.

Stary, S., Yin, X.J., Potuschak, T., Schlogelhofer, P., Nizhynska, V., and Bachmair, A. (2003). PRT1 of Arabidopsis is a ubiquitin protein ligase of the plant N-end rule pathway with specificity for aromatic amino-terminal residues. Plant physiology *133*, 1360-1366.

Swanson, R., Locher, M., and Hochstrasser, M. (2001). A conserved ubiquitin ligase of the nuclear envelope/endoplasmic reticulum that functions in both ER-associated and Matalpha2 repressor degradation. Genes & development *15*, 2660-2674.

Tasaki, T., and Kwon, Y.T. (2007). The mammalian N-end rule pathway: new insights into its components and physiological roles. Trends in biochemical sciences *32*, 520-528.

Tasaki, T., Mulder, L.C., Iwamatsu, A., Lee, M.J., Davydov, I.V., Varshavsky, A., Muesing, M., and Kwon, Y.T. (2005). A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. Molecular and cellular biology *25*, 7120-7136.

Tasaki, T., Sriram, S.M., Park, K.S., Kwon, Y.T. (2012). The N-end Rule Pathway. Annual review of biochemistry *82*.

Tasaki, T., Zakrzewska, A., Dudgeon, D.D., Jiang, Y., Lazo, J.S., and Kwon, Y.T. (2009). The substrate recognition domains of the N-end rule pathway. The Journal of biological chemistry *284*, 1884-1895.

Tobias, J.W., Shrader, T.E., Rocap, G., and Varshavsky, A. (1991). The N-end rule in bacteria. Science *254*, 1374-1377.

Tolkatchev, D., Vinogradova, A., and Ni, F. (2005). Transforming bivalent ligands into retractable enzyme inhibitors through polypeptide-protein interactions. Bioorg Med Chem Lett *15*, 5120-5123.

Turner, G.C., Du, F., and Varshavsky, A. (2000). Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway. Nature *405*, 579-583.

Varshavsky, A. (1997). The ubiquitin system. Trends in biochemical sciences *22*, 383-387.

Varshavsky, A. (2011). The N-end rule pathway and regulation by proteolysis. Protein science : a publication of the Protein Society, [Epub ahead of print].

Wang, K.H., Oakes, E.S., Sauer, R.T., and Baker, T.A. (2008a). Tuning the strength of a bacterial N-end rule degradation signal. The Journal of biological chemistry *283*, 24600-24607.

Wang, K.H., Roman-Hernandez, G., Grant, R.A., Sauer, R.T., and Baker, T.A. (2008b). The molecular basis of N-end rule recognition. Molecular cell *32*, 406-414.

Wu, J.W., Cocina, A.E., Chai, J., Hay, B.A., and Shi, Y. (2001). Structural analysis of a functional DIAP1 fragment bound to grim and hid peptides. Molecular cell *8*, 95-104.

Xia, Z., Turner, G.C., Hwang, C.S., Byrd, C., and Varshavsky, A. (2008a). Amino acids induce peptide uptake via accelerated degradation of CUP9, the transcriptional repressor of the PTR2 peptide transporter. The Journal of biological chemistry *283*, 28958-28968.

Xia, Z., Webster, A., Du, F., Piatkov, K., Ghislain, M., and Varshavsky, A. (2008b). Substrate-binding sites of UBR1, the ubiquitin ligase of the N-end rule pathway. The Journal of biological chemistry *283*, 24011-24028.