ELUCIDATING THE UNDERSTUDIED FRUCTOSAMINE-3-KINASE (FN3K) FAMILY: A COMBINED EXPERIMENTAL AND COMPUTATIONAL STUDY ON STRUCTURE, FUNCTION, EVOLUTION, AND REGULATION

by

SAFAL SHRESTHA

(Under the Direction of Natarajan Kannan)

ABSTRACT

Fructosamine-3-kinases (FN3Ks) are evolutionarily conserved protein kinase-like (PKL) fold enzymes that play an important role in cellular homeostasis by repairing proteins. Through phosphoryl transfer, the enzyme removes sugar adducts from glycated lysine residues in proteins. Despite being conserved across all domains of life, the family has received very little attention.

In this dissertation, I apply a combination of experimental and computational methods to illuminate the structure, function, evolution, and regulation of the enzyme family. By solving the crystal structure of the FN3K ortholog from Arabidopsis thaliana, for the first time, we provide a structural basis for redox mode of regulation. We also demonstrated its conservation across the family, including in humans. Next, by leveraging the FN3K knock out (KO) in HepG2 cells and using a multi-omics (transcriptomics, metabolomics, interactomics, and genomics) approach, we provide new links of FN3K to lipid and co-factor metabolism as well as oxidative stress response. We also find new links between FN3K and Nicotinamide adenine dinucleotide (NAD) and show that human FN3K can specifically bind to different NAD compounds. Moreover, using phylogenetics and Bayesian statistics based evolutionary sequence constraints method, I derived

evolutionary relationship between FN3Ks and related families in the PKL superfamily as well as define sequence constraints distinguishing FN3K family. I also identified several divergent bacterial sub-groups within the FN3K family that will require further characterization. Through high resolution crystal structures, I show two distinct modes of adenine binding in the FN3K family. This dissertation sheds light on the multifaceted roles and regulatory mechanisms of FN3K enzymes, expanding our understanding of their function and evolution, and highlighting potential areas for further research within the field.

INDEX WORDS: Fructosamine-3-kinases, protein kinase like, evolution, redox, sequence analysis, structure

ELUCIDATING THE UNDERSTUDIED FRUCTOSAMINE-3-KINASE (FN3K) FAMILY: A COMBINED EXPERIMENTAL AND COMPUTATIONAL STUDY ON STRUCTURE, FUNCTION, EVOLUTION, AND REGULATION

by

SAFAL SHRESTHA

B.S. in Bioinformatics, Ramapo College of New Jersey, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

ELUCIDATING THE UNDERSTUDIED FRUCTOSAMINE-3-KINASE (FN3K) FAMILY: A COMBINED EXPERIMENTAL AND COMPUTATIONAL STUDY ON STRUCTURE, FUNCTION, EVOLUTION, AND REGULATION

by

SAFAL SHRESTHA

Major Professor:	Natarajan Kannan
Committee:	Eileen J. Kennedy
	Eva-Maria Strauch
	Jorge C. Escalante
	Y. George Zheng

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2023

To my family and friends for continued love and support

****

To Dr. Ashley Stuart for teaching me about protein structure and modeling

****

To Dr. Sandra Suarez for spending hours of her time helping me draft my grad school applications

****

To Millie for the joy

****

ACKNOWLEDGEMENTS

I wish to extend my deepest thanks to my dissertation advisor, Natarajan Kannan, for his unwavering guidance and support throughout my Ph.D. journey. I also extend my appreciation to my committee members, Eileen J. Kennedy, Eva-Maria Strauch, Jorge C. Escalante, and Y. George Zheng, for their invaluable advice. I also want to thank the past and present lab members: Dr. Samiksha Katiyar, Dr. Rahil Taujale, Dr. Wayland Yeung, Aarya Venkat, Dr. George Bendzunas, Dr. Carlos Sanz, Dr. Annie Kwon, Dr. Liang-Chin Huang, Dr. Zheng Ruan, Brady O'Boyle, Nathan Gravel, Zhongliang Zhou, Niral Thaker, Ishan Aggarwal, Grace Watterson, and Jason Lu. Finally, I want to express my gratitude to Dr. Patrick A. Eyers and Dr. Dominic Byrne at the University of Liverpool for their highly productive collaborations and invaluable scientific insights.

TABLE OF CONTENTS

Page

APPENDICES

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Motivation

Cells rely on a fine-tuned balance of metabolic processes to maintain their proper function and overall homeostasis. At the heart of these processes are metabolic enzymes, which facilitate the countless reactions necessary for life. However, imbalances or malfunctions within these pathways, such as the accumulation of certain intermediates or spontaneous reactions due to high cellular concentrations, can lead to detrimental disease states [1].

Advanced Glycation End (AGE) products are a heterogeneous group of cross-linking compounds that form over time through spontaneous process called glycation and have been implicated in a variety of chronic illnesses such as diabetes, neurodegenerative conditions, cardiovascular issues, and kidney disease [2-5]. Distinct from glycosylation, which is enzyme driven, glycation is non-enzymatic attachment of reducing sugars such as glucose and ribose to free amines on proteins, lipids, and nucleic acids [6, 7]. To counter this, organisms have evolved a protein repair enzyme known as Fructosamine-3-kinase (FN3K). FN3Ks remove the sugar adduct from lysine residues by transferring phosphate from an ATP molecule [8-10].

While prior research has largely focused on elucidating the role of FN3Ks in deglycation and understanding substrate specificity mainly through works on the human paralogs [8, 9, 11-14], there is a notable gap in the literature looking into how the enzyme family is regulated. Moreover, despite being conserved across all domains of life, there is a clear lack of deeper evolutionary

analysis on the FN3K family utilizing both sequence and structural datasets to gain mechanistic insights. While there is a paucity of structural information available for the family, we propose to bridge this gap by applying evolutionary systems biology in conjunction with experimental techniques from biochemistry and structural biology to explore the regulation and functional evolution of FN3K family members. Our working hypothesis is that critical residues associated with these functions can be identified through comparative sequence analysis of FN3Ks from diverse species, as well as with related kinase families. Through this approach, we aim to unveil signature residues that are fundamental for the enzyme's regulatory and functional properties. Such findings would not only advance our understanding of FN3K but could also provide insights into the regulation of other related kinases such as aminoglycosides phosphotransferase (APH) involved in antibiotics resistance, choline kinases and so on. Moreover, the work will pave the way for the development of targeted therapeutic interventions.

1.2 Background

1.2.1 Glycation, Advanced Glycation End products, and Deglycation by FN3Ks

Spontaneous reactions of reducing sugars such as glucose and ribose with free amine groups in proteins found in basic amino acids such as lysine and arginine residues leads to the formation of unstable Schiff base through a process called glycation [15]. Rearrangement of the Schiff base can lead to the formation of stable Amadori products called ketosamines [6, 16, 17]. If not removed, over time these compounds are converted to a heterogeneous group of crosslinking compounds called Advanced Glycation End (AGE) products through further oxidation and/or elimination reactions. Importantly, AGEs have been implicated in many chronic disease states

such as diabetes, cardiovascular, and neurodegenerative including Parkinson's and Alzheimer's, and are even considered a marker for aging [3-5, 18-23]. However, nature has evolved an enzymatic repair mechanism to prevent the formation of such molecules. FN3Ks are a family of small molecule kinases belonging to the eukaryotic like kinases (ELKs) group that adopt the Protein Kinase Like (PKL) fold and are thus distantly related to Eukaryotic Protein Kinases (EPKs). FN3Ks remove the protein bound sugars by transferring ɣ-phosphate from ATP to 3'-hydroxyl (-OH) group on the ketosamine moiety (Fig. 1.1) [8, 10]. The proximity of the ketone group next to the phosphate destabilizes the intermediate resulting in recovery of the free amine and formation of an inorganic phosphate and 3-deoxyglucosone (3-DG) [8, 10, 12]. Interestingly 3-deoxyglucosone (3-DG) is a di-carbonyl compound which is a potent glycating agent and must undergo detoxification within cells [19, 20, 24-27].

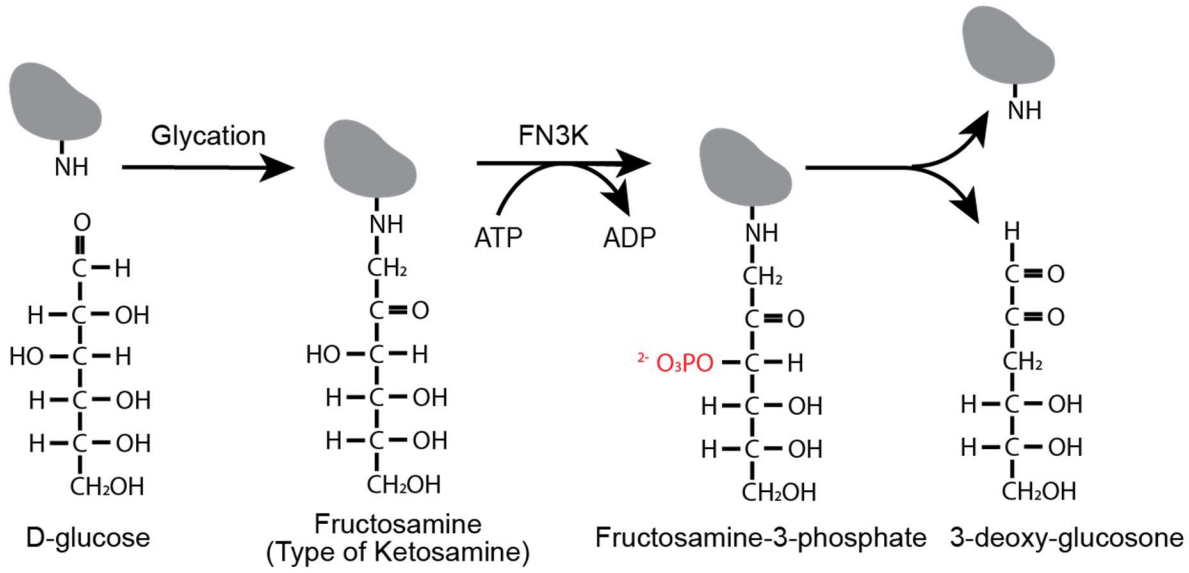


Figure 1.1: De-glycation by FN3K.

A schematic diagram showing deglycation reaction by FN3K enzyme.

1.2.2 Protein Kinase Fold (PKL) Superfamily

The PKL superfamily consists of three major groups of proteins: EPKs, ELKs, and the more divergent Atypical Kinases (APK) [28, 29]. EPKs are involved in various cellular signaling pathways, phosphorylating protein substrates on serine, threonine, or tyrosine residues [30-32] while ELKs phosphorylate diverse sets of small molecule substrates. Together with FN3K, they include several functionally important families such as Aminoglycoside phosphotransferases (APHs), Choline Kinases (ChoK), and Methylthioribose kinase (MTRK). APHs phosphorylate aminoglycosides and are involved in antibiotic resistance in bacteria [33-35]. ChoK phosphorylate the choline molecule that leads to formation of phosphatidylcholine which is a major phospholipid in eukaryotic membranes [36, 37] whereas MTRKs phosphorylate 5-Methylthioribose in bacteria and plants and are involved in methionine salvage pathway [38]. On the other hand, APKs consists of protein families such as Phosphatidylinositol phosphate kinases (PIPKs) and AarF domain-containing protein kinases (ADCKs) involved in synthesis of membrane localized phosphoinositide phospholipids [39] and Coenzyme Q biosynthesis pathways respectively [40].

Despite their high divergence and having very low sequence identity between each other, at the structural level the major groups within the PKL fold superfamily still conserve the PKL fold. The PKL fold is characterized by a bilobal structure with β strand rich N-lobe for ATP binding and α-helical rich C lobe important for substrate binding. Importantly, among the major groups, there is more similarity observed for the ATP binding N-lobe as compared to the C-lobe potentially due to the diversity in substrates that are phosphorylated.

1.2.3 Evolutionary systems biology approach to studying the FN3K family.

FN3Ks are one of the few members of the PKL superfamily that are conserved across all domains of life (archaea, bacteria, and eukaryotes). As a result, there is a plethora of sequence information available for the family but there is a paucity of structural information. Only two apo-forms of bacterial homologs have been crystallized through the Structural Genomics effort with no related publications. However, the emergence of highly accurate deep learning-based AlphaFold [41] modeling tool provides unprecedented structural information and modeling opportunity to gain mechanistic insight into the family.

While traditional phylogenetic approaches provide evolutionary relationships between protein sequences and help identify divergent subgroups, it fails to identify sequence constraints that define the subgroup. The optimal multiple-category Bayesian Partitioning with Pattern Selection (omcBPPS) procedure identifies hierarchically arranged subgroups that conserve signature residues or sequence constraints based on a pattern of both sequence conservation and divergence [42-44]. The procedure has previously been applied successfully to study various EPKs and ELKs families to identify functionally important residues [45-49]. We will apply the procedure to study the FN3K family and define family and sub-family level sequence constraints that can be used to generate testable hypothesis.

1.3 Key challenges and unresolved questions

The role of FN3Ks in protein repair through deglycation, combined with its conservation across all domains of life, provides a compelling rationale for deeper study of the family, both for functional characterization and potential therapeutic interventions. While some aspects of the enzyme function, such as substrate specificity have been elucidated previously through

biochemical approaches, there is still a clear lack of understanding on how the enzyme family is regulated. While deglycation has been reported as one of the functions for the enzyme family, the role of the enzyme in a larger pathway context is still missing. While the Protein Databank (PDB) database contains two apo crystal structures of the bacterial homologs, courtesy of structural genomics efforts, it lacks information on the structural binding of various ligands to the enzyme. There is a wealth of sequence information available for the FN3K family. However, a comprehensive evolutionary analysis utilizing available sequences, crystal structures, and predicted models to pinpoint FN3K-specific sequence features is still missing. Defining the sequence features in the context of structural data will provide mechanistic insights into the FN3K family and provide a framework to generate a new testable hypothesis.

1.4 Major research questions addressed.

To address the aforementioned questions, subsequent chapters employ a combination of bioinformatics, biochemical, and structural methodologies to shed light on the evolutionary development of various functional facets of the FN3K family. For the first time, we provide a structural basis for a novel mode of redox regulation in the FN3K family. Notably, we show that the human FN3K is also regulated in a similar manner. We extend the work further and leverage the human FN3K Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) knowck out (KO) in HepG2 cell lines to show new links of FN3Ks to pathways related lipid synthesis, oxidative stress response, and carbon metabolism. Moreover, we link FN3Ks to Nicotinamide adenine dinucleotide (NAD) molecules and show specific binding of NAD compounds to human FN3K. Next, we perform a comprehensive evolutionary analysis of the FN3K family and identify sequence and structural constraints unique to the FN3K family as well as identify divergent

subgroups within the family. Survey of the bacterial operon structure, gene fusion events, and modeling provides new insights into how FN3Ks interact with diverse domains to potentially perform diverse functions in different organisms. Moreover, through structural characterization, we gather new insights into the nucleotide binding modes within the FN3K family.

1.4.1 Identifying a novel mode of redox regulation through the P-loop cysteine residues.

Based on the crystal structures of apo forms of the bacterial homologs from *T. fusca* and *H. somnus*, it was known that FN3Ks adopt the protein kinase fold and are distantly related to highly regulated EPKs. Despite being involved in an important cellular process of protein repair and being conserved across the tree of life, no previous work had investigated how the activity of the enzyme was regulated.

In this work, by solving the crystal structure of the FN3K ortholog from *A. thaliana*, for the first time, we provide the structural basis for redox mode of regulation in the family through the P-loop cysteine. Moreover, based on the conservation of the P-loop cysteine residue, we show existence of similar mechanism in human FN3K both *in vitro* and *in vivo*.

1.4.2 Multi-omics approach comparing FN3K WT and KO HepG2 cell lines provide new links to nicotinamides adenine dinucleotide molecules.

While the role of human FN3K in deglycation has been established in previous work, the need for the enzyme to be redox regulated and its importance in a larger pathway context was less understood. To address this knowledge gap, we leveraged the FN3K CRISPR Knock-Out (KO) cell line and performed an integrative multi-omics study (transcriptomics, interactomics, metabolomics, and genomics). Based on the study, we found new links of FN3K function to lipid

biosynthesis (cholesterol and fatty acids), carbon and co-factor metabolism as well as oxidative stress response. Based on significant enrichment of nicotinamide adenine dinucleotide (NAD) binding proteins, we report specific binding of human FN3K to NAD compounds in a metal and concentration-dependent manner. Thus, the study identified potential links between FN3Ks and NAD-mediated energy metabolism and redox balance.

1.4.3 Evolutionary and structural analyses reveal new insights into nucleotide binding.

FN3Ks are conserved across all domains of life and are distantly related to EPKs. However, it is more closely related to a subset of eukaryotic like kinases (ELKs) that phosphorylate small molecule substrates. While previous work has compared members of the PKL enzymes based on core domain shared by the PKL members, a more focused analysis between FN3Ks and related ELKs is lacking. In this work, we identify evolutionary relationship between FN3K and other related ELKs and define FN3K specific signature residues. We further classify FN3K sequence into multiple divergent subgroups and identify sequence constraints related to subgroups. Detailed look into the bacterial operon structure as well as gene fusion events in conjunction with AlphaFold modeling provide new insights into how divergent subgroups of FN3Ks could interact with other domains to perform diverse functions. Finally, we provide structural evidence for variable mode of adenine binding in *T. fusca* as well as ADP ribose binding in *A. thaliana* orthologs. Together these data provide a framework to explore other functions of FN3Ks and its orthologs.

Bibliography

1. Bommer, G.T., E. Van Schaftingen, and M. Veiga-da-Cunha, *Metabolite Repair Enzymes Control Metabolic Damage in Glycolysis.* (0968-0004 (Print)).
2. Bank, R.A., et al., *Ageing and zonal variation in post-translational modification of collagen in normal human articular cartilage. The age-related increase in non-enzymatic glycation affects biomechanical properties of cartilage.* Biochem J, 1998. **330 (Pt 1)**: p. 345-51.
3. Böhm, P.G.a.M., *Advanced glycation end products. Key players in skin aging?* Dermato-Endocrinology, 2012. **4**(3): p. 259-270.
4. Hause, F., et al., *Accumulation of glycated proteins suggesting premature ageing in lamin B receptor deficient mice.* Biogerontology, 2018. **19**(1): p. 95-100.
5. Josephine M. Forbes, L.T.L.Y., Vocki Thallas, Markus Lassila, Riccardo Candido, Karin A. Jandeleit-Dahm, Merlin C. Thomas, Wendy C. Burns, Elizabeth K. Deemer, Susan R. Thorpe, Mark E. Cooper and Terri J. Allen., *Advanced Glycation End Product Interventions Reduce Diabetes-Accelerated Atherosclerosis.* Diabetes, 2004. **53**(7): p. 1813-1823.
6. Baynes, J.W., et al., *The Amadori product on protein: structure and reactions.* Prog Clin Biol Res, 1989. **304**: p. 43-67.
7. Teodorowicz, M., J. van Neerven, and H. Savelkoul, *Food Processing: The Influence of the Maillard Reaction on Immunogenicity and Allergenicity of Food Proteins.* Nutrients, 2017. **9**(8).

8. Benjamin S. Szwerglod, S.H., and Paul J. Beisswenger, *Human Fructosamine-3-Kinase. Purification, Sequencing, Substrate Specificity, and Evidence of Activity In Vivo.* Diabetes, 2001. **50**: p. 2139-2147.

9. François Collard, E.W., Niki Bergans, Juliette Fortpied, Didier Vertommen, Florent Vanstapel, Ghislain Delpierre, and Emil Van Shaftingen, *Fructosamine 3-kinase-related protein and deglycation in human erythrocytes.* Biochemical Journal, 2004. **382**: p. 137-143.

10. Ghislain Delpierre, F.C., Juliette Fortpied and Emile Van Schaftingen, *Fructosamine-3-kinase is involved in an intracellular deglycation pathway in human erythrocytes.* Biochemical Journal 2002. **365**: p. 801-808.

11. François Collard, G.D., Vincent Stroobant, Gert Matthijs, Emil Van Shaftingen, *A Mammalian Protein Homologous to Fructosamine-3-Kinase Acting on Psicosamines and Ribulosamines but not on Fructosamines.* Diabetes, 2003. **52**: p. 2888-2895.

12. Ghislain Delpierre, M.H.R., François Collard, Vincent Stroobant, Florent Vanstapel, Helena Santos, and Emile Van Schaftingen, *Identification, Cloning, and Heterologous Expression of a Mammalian Fructosamine-3-Kinase.* Diabetes, 2000. **49**: p. 1627-1634.

13. James R. Conner, P.J.B., Benjamin S. Szwergold, *The expression of the genes for fructosamine-3-kinase and fructosamine-3-kinase-related protein appears to be constitutive and unaffected by environmental signal.* Biochemical and Biophysical Research Communications, 2004. **323**: p. 932-936.

14. Jérôme Delplanque, G.D., Fred R, Opperdoes, and Emile Van Schaftingen, *Tissue Distribution and Evolution of Fructosamine 3-Kinase and Fructosamine 3-Kinase-related Protein.* The Journal of Biological Chemistry, 2004. **279**(45): p. 46606-46613.

15. Higgins, H.F.B.a.P.J., *Reaction of Monosaccharides with Proteins: Possible Evolutionary Significance.* Science, 1981. **213**: p. 222-224.

16. Emile Van Schaftingen, F.C., Elsa Wiame, Maria Veiga-da-Cunha, *Enzymatic repair of Amadori products.* Amino Acids, 2012. **42**: p. 1143-1150.

17. Hodge, J.E., *The Amadori rearrangement.* Adv Carbohydr Chem, 1955. **10**: p. 169-205.

18. Paul, R.G. and A.J. Bailey, *Glycation of collagen: the basis of its central role in the late complications of ageing and diabetes.* (1357-2725 (Print)).

19. Niwa, T., et al., *Presence of 3-Deoxyglucosone, a Potent Protein Crosslinking Intermediate of Maillard Reaction, in Diabetic Serum.* Biochemical and Biophysical Research Communications, 1993. **196**(2): p. 837-843.

20. Niwa, T. and S. Tsukushi, *3-deoxyglucosone and AGEs in uremic complications: inactivation of glutathione peroxidase by 3-deoxyglucosone.* Kidney Int Suppl, 2001. **78**(0098-6577 (Print)): p. S37-41.

21. Varun Parkash Singh, A.B., Nirmal Singh, and Amteshwar Singh Jaggi, *Advanced Glycation End Products and Diabetic Complications.* Korean Journal of Physiology and Pharmacology, 2014. **18**: p. 1-14.

22. Shun-Yao Ko, H.-A.K., Kuo-Hsiung Chu, Tzong-Ming Shieh, Tzong-Xheng Chi, Hong-I Chen, Weng-Cheng Chang, Shu-Shing Chang, *The Possible Mechanism of Advanced Glycation End Products (AGEs) for Alzheimer's Disease.* PLoS One, 2015. **10**(11): p. e0143345.

23. Sharma, C., et al., *Advanced glycation End-products (AGEs): an emerging concern for processed food industries.* J Food Sci Technol, 2015. **52**(12): p. 7561-76.

24. Kevin J. Knecht, M.S.F., and John W. Baynes, *Detection of 3-Deoxyfructose and 3-Deoxyglucosone in HUman Urine and Plasma: Evidence for Intermediate Stages of the Maillard Reaction in Vivo.* Archives of Biochemistry and Biophysics, 1992. **294**(1): p. 130-137.

25. Niwa, T., *3-Deoxyglucosone: metabolism, analysis, biological activity, and clinical implication.* Journal of Chromatography B, 1999. **731**: p. 23-36.

26. T. R. Brown, B.S., K.A. Brown, M.A. Schwartz, A.M. Tobia and F. Kappler, *Modulation of in vivo 3-deoxyglucosone levels.* Biochemical Society Transactions, 2003. **31**(6): p. 1433-1437.

27. Kusunoki, H., et al., *Relation Between Serum 3-Deoxyglucosone and Development of Diabetic Microangiopathy.* Diabetes Care, 2003. **26**(6): p. 1889.

28. Kannan, N., et al., *Structural and functional diversity of the microbial kinome.* PLoS Biol, 2007. **5**(3): p. e17.

29. Scheeff, E.D. and P.E. Bourne, *Structural evolution of the protein kinase-like superfamily.* PLoS Comput Biol, 2005. **1**(5): p. e49.

30. Manning, G., et al., *Evolution of protein kinase signaling from yeast to man.* Trends Biochem Sci, 2002. **27**(10): p. 514-20.

31. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.

32. Taylor, S.S., et al., *Evolution of the eukaryotic protein kinases as dynamic molecular switches.* Philos Trans R Soc Lond B Biol Sci. **367**(1602): p. 2517-28.

33. Hon, W.C., et al., *Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases.* Cell, 1997. **89**(6): p. 887-95.

34. Daigle, D.M., D.W. Hughes, and G.D. Wright, *Prodigious substrate specificity of AAC(6')-APH(2"), an aminoglycoside antibiotic resistance determinant in enterococci and staphylococci.* Chem Biol, 1999. **6**(2): p. 99-110.

35. Kaplan, E., et al., *Aminoglycoside binding and catalysis specificity of aminoglycoside 2''-phosphotransferase IVa: A thermodynamic, structural and kinetic study.* Biochim Biophys Acta, 2016. **1860**(4): p. 802-13.

36. Peisach, D., et al., *The crystal structure of choline kinase reveals a eukaryotic protein kinase fold.* Structure, 2003. **11**(6): p. 703-13.

37. Hudson, C.S., et al., *Kinetic and mechanistic characterisation of Choline Kinase-alpha.* Biochim Biophys Acta, 2013. **1834**(6): p. 1107-16.

38. Ku, S.Y., et al., *Structures of 5-methylthioribose kinase reveal substrate specificity and unusual mode of nucleotide binding.* (0021-9258 (Print)).

39. Llorente, A., et al., *PIP kinases: A versatile family that demands further therapeutic attention.* (2212-4934 (Electronic)).

40. Stefely, J.A., et al., *Mitochondrial ADCK3 employs an atypical protein kinase-like fold to enable coenzyme Q biosynthesis.* Mol Cell, 2015. **57**(1): p. 83-94.

41. Jumper, J.A.-O., et al., *Highly accurate protein structure prediction with AlphaFold.* (1476-4687 (Electronic)).

42. Neuwald, A.F., *Evolutionary clues to DNA polymerase III beta clamp structural mechanisms.* Nucleic Acids Res, 2003. **31**(15): p. 4503-16.

43. Neuwald, A.F., *The CHAIN program: forging evolutionary links to underlying mechanisms.* Trends in Biochemical Sciences, 2007. **32**(11): p. 487-493.

44. Neuwald, A.F., *A Bayesian sampler for optimization of protein domain hierarchies.* J Comput Biol, 2014. **21**(3): p. 269-86.

45. Kannan, N., et al., *The hallmark of AGC kinase functional divergence is its C-terminal tail, a cis-acting regulatory module.* Proc Natl Acad Sci U S A, 2007. **104**(4): p. 1272-7.

46. Nguyen, T., et al., *Co-conserved MAPK features couple D-domain docking groove to distal allosteric sites via the C-terminal flanking tail.* PLoS One, 2015. **10**(3): p. e0119636.

47. Oruganty, K., et al., *Identification and classification of small molecule kinases: insights into substrate recognition and specificity.* BMC Evol Biol, 2016. **16**: p. 7.

48. Mohanty, S., et al., *Hydrophobic Core Variations Provide a Structural Framework for Tyrosine Kinase Evolution and Functional Specialization.* PLoS Genet, 2016. **12**(2): p. e1005885.

49. Kwon, A., et al., *Tracing the origin and evolution of pseudokinases across the tree of life.* Science Signaling, 2019. **12**(578): p. eaav3810.

CHAPTER 2

A REDOX-ACTIVE SWITCH IN FRUCTOSAMINE-3-KINASES EXPANDS THE REGULATORY REPERTOIRE OF THE PROTEIN KINASE SUPERFAMILY

Shrestha S, Katiyar S, Sanz-Rodriguez CE, Kemppinen NR, Kim HW, Kadirvelraj R, Panagos C, Keyhaninejad N, Colonna M, Chopra P, Byrne DP, Boons GJ, van der Knaap E, Eyers PA, Edison AS, Wood ZA, Kannan N. A redox-active switch in fructosamine-3-kinases expands the regulatory repertoire of the protein kinase superfamily. Sci Signal. 2020 Jul 7;13(639):eaax6313. doi: 10.1126/scisignal.aax6313. PMID: 32636308; PMCID: PMC8455029.
Reprinted here with permission of the publisher.

2.1 Abstract

Aberrant regulation of metabolic kinases by altered redox homeostasis substantially contributes to aging and various diseases, such as diabetes. We found that the catalytic activity of a conserved family of fructosamine-3-kinases (FN3Ks), which are evolutionarily related to eukaryotic protein kinases, is regulated by redox-sensitive cysteine residues in the kinase domain. The crystal structure of the FN3K homolog from *Arabidopsis thaliana* revealed that it forms an unexpected strand-exchange dimer in which the ATP-binding P-loop and adjoining beta strands are swapped between two chains in the dimer. This dimeric configuration is characterized by strained interchain disulfide bonds that stabilize the P-loop in an extended conformation. Mutational analysis and solution studies confirmed that the strained disulfides function as redox "switches" to reversibly regulate the activity and dimerization of FN3K. Human FN3K, which contains an equivalent P-loop Cys, was also redox-sensitive, whereas ancestral bacterial FN3K homologs, which lack a P-loop Cys, were not. Furthermore, CRISPR-mediated knockout of FN3K in human liver cancer cells altered the abundance of redox metabolites, including an increase in glutathione. We propose that redox regulation evolved in FN3K homologs in response to changing cellular redox conditions. Our findings provide insights into the origin and evolution of redox regulation in the protein kinase superfamily and may open new avenues for targeting human FN3K in diabetic complications.

2.2 **Introduction**

Glycation is a universal post-translational modification in which reducing sugars such as glucose and fructose are non-enzymatically added to free amine groups on peptides, proteins, and lipids. This non-enzymatic modification occurs endogenously in all living organisms, as well as exogenously in foods we consume [1, 2]. Sugars attached to amine groups can undergo Amadori rearrangements to form stable linkages with biomolecules [3-5]. Because such linkages can adversely affect biomolecular functions, organisms have evolved deglycation mechanisms [6-8] to repair the potentially toxic effects of reactive sugars. Fructosamine-3 kinases (FN3Ks) are a conserved family of deglycation enzymes that remove ribose and fructose sugars attached to surface-exposed lysine (Lys) residues (ketosamines) in proteins [9-12]. They do this by catalyzing the transfer of the gamma phosphate from ATP to the 3' hydroxyl group in the ketosamine substrate. Phosphorylation of ketosamines by FN3Ks results in an unstable ketosamine 3-phosphate intermediate, which spontaneously decomposes into inorganic phosphate [9]. Since the original unmodified Lys is regenerated as a consequence of FN3K deglycation activity, FN3Ks are believed to function as protein repair enzymes [6, 13].

FN3Ks are conserved across the tree of life [14]. While simple eukaryotes and prokaryotes both contain a single copy of the FN3K gene, complex eukaryotes, including mammals, encode two copies, FN3K and FN3K related protein (FN3K-RP), presumably due to a gene duplication event in amphibians. Although the functions of FN3K homologs in lower eukaryotes and bacteria are yet to be equivocally established, it is proposed that they repair proteins modified by ribose-5-phosphate, a potent glycating agent generated by the metabolic pentose phosphate pathway [6, 12, 14, 15]. While FN3K activity is essential for normal cellular functions, an uncontrolled activity can result in altered cellular homeostasis and disease [16, 17]. For example, accumulation of 3-

deoxyglucosone, a byproduct of human FN3K activity, is causatively associated with diabetic complications such as retinopathy and neuropathy [18, 19]. Increased 3-deoxyglucosone levels also contribute to oxidative stress by inhibiting glutathione peroxidase, a potent cellular antioxidant [20]. Thus, tight regulation of FN3K activity is crucial for maintaining cellular homeostasis and to prevent oxidative stress resulting from its repair functions. However, despite the remarkable conservation of FN3Ks across the tree of life and their fundamental roles in cellular functions and disease, a mechanistic understanding of FN3K function and mode(s) of regulation is currently lacking.

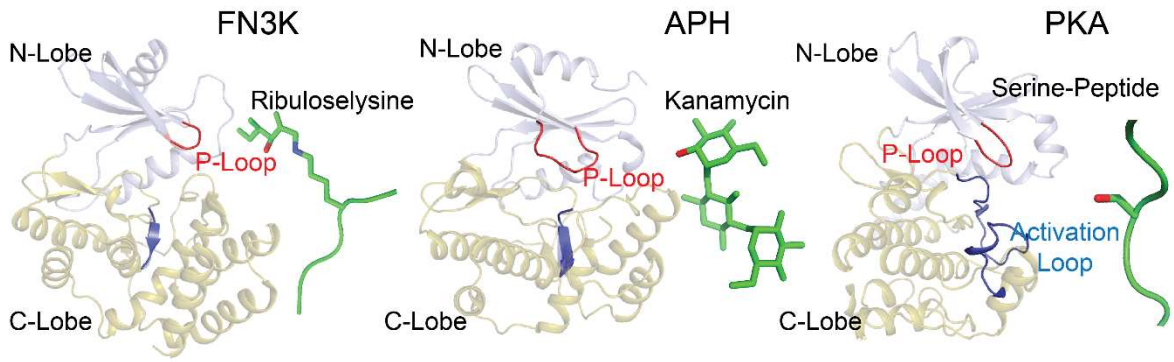


Figure 2.1: FN3K adopts a Protein Kinase fold.

Comparison of the overall fold of FN3K from *T. fusca* (TfFN3K) (PDB ID: 3F7W), Aminoglycoside Phosphotransferase (APH) (PDB ID: 1L8T) [21] and Protein Kinase A (PKA) (PDB ID: 1ATP) [22]. The structures are shown as cartoon where N-lobe is colored in light blue and the C-lobe in olive. The substrates are shown as either sticks (ribuloselysine, kanamycin) or cartoons (serine-peptide), and colored green. The oxygen atom on the hydroxyl group where the phosphate group is transferred is colored in red. The P-loop and activation loop are colored in red and blue respectively.

We previously showed that FN3Ks belong to a large superfamily of protein kinase-like (PKL) enzymes that include eukaryotic protein kinases (EPKs), small-molecule kinases, and atypical kinases [23, 24]. FN3Ks are more closely related to small-molecule kinases, such as aminoglycoside kinase (APH) and choline kinases, than EPKs and more distantly related to atypical pseudokinases such as Fam20C and SelO [25-27]. Through quantitative comparisons of the evolutionary constraints acting on diverse PKL-fold enzymes, we previously demonstrated that

EPKs share sequence and structural similarity with small-molecule kinases in the N-terminal ATP binding lobe, but diverge significantly in the C-terminal substrate-binding lobe [23, 24, 28]. In particular, the extended activation segment connecting the ATP and substrate binding lobes that classically controls catalytic activity through conformational changes driven by reversible phosphorylation of serine, threonine and tyrosine residues [29] is unique to EPKs and absent in small molecule kinases, including FN3Ks (Fig. 2.1). An accompanying paper submitted side-by-side with this manuscript demonstrates that in addition to reversible phosphorylation, oxidation and reduction of a conserved Cys residue in the activation segment is a much more common mode of Ser/Thr protein kinase regulation than had been previously appreciated [30].

Here, we identify a critical role for the ATP binding P-loop in the redox regulation of FN3Ks. By solving the first crystal structure of a eukaryotic FN3K homolog, *Arabidopsis thaliana* FN3K (AtFN3K), we find that the P-loop is stabilized in an extended conformation by a Cys-mediated disulfide bond connecting two chains to form a covalently linked dimer. We show that the reduction of disulfides results in AtFN3K activation. Consistently, the human FN3K (HsFN3K) conserving the P-loop cysteine is redox-active and displays altered oligomerization when proliferating cells are exposed to acute oxidative stress.

We propose that redox control mediated by the P-loop Cys is an ancient mechanism of FN3K regulation that emerged progressively during FN3K evolution from bacteria to humans. Because many protein kinases contain an equivalent redox-active cysteine in the P-loop, our studies also have broad implications for understanding the structure, function, and evolution of all PKL-fold enzymes, in particular, tyrosine kinases, which contain a cysteine residue at the equivalent position. Our detailed mechanistic characterization of FN3Ks also opens new avenues

for the design of FN3K-targeted small molecule inhibitors for diabetic complications associated with increased protein glycation.

2.3 Results

2.3.1 Crystal structure of AtFn3K reveals a novel strand-exchange dimer.

To investigate the structural basis for biological FN3K regulation, we solved the first crystal structure of a plant FN3K homolog (AtFN3K) in complex with the Adenosine triphosphate (ATP) mimic Adenylyl-imidodiphosphate (AMP-PNP) at a resolution of 2.37 Å using a multiple model molecular replacement strategy. The asymmetric unit of AtFN3K contains two molecules with a small degree of disorder at the N- and C- termini (residues 1-6 and 296, 297) (Fig. 2.2A). Each chain contains a well-ordered molecule of AMP-PNP in the active site, albeit with missing electron density for γ-phosphate (Fig. 2.2B). AMP-PNP is known to hydrolyze slowly over time, as shown previously for Protein Kinase A [31]. Because the nitrogen atom in the β-phosphate of AMP-PNP could not be identified, we have modeled the ligand as Adenosine diphosphate (ADP).

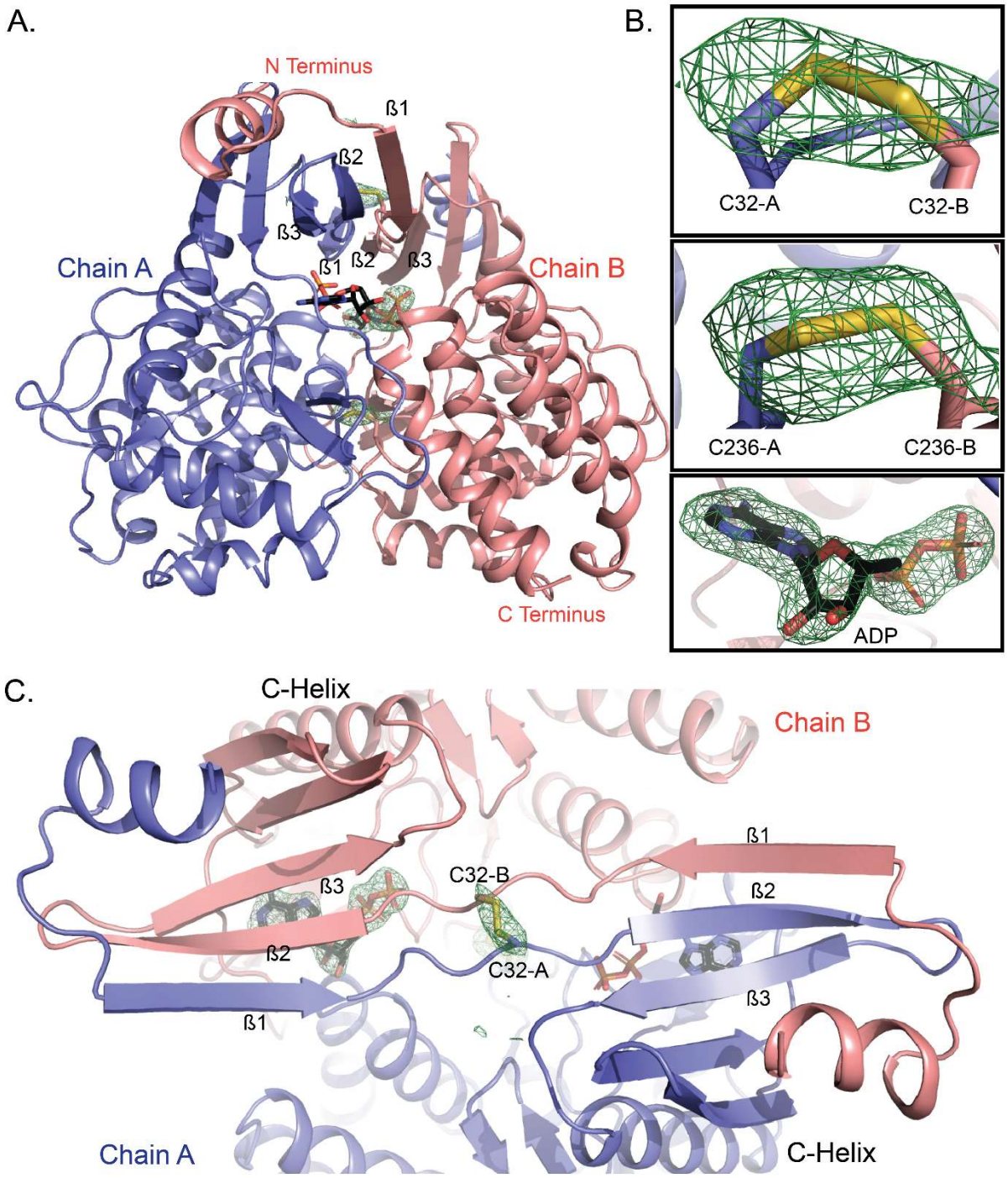


Figure 2.2: **AtFN3K WT is a novel beta-strand exchange disulfide mediated dimer.**

(A) Cartoon representation of the crystal structure of *A. thaliana* FN3K (AtFN3K) homodimer. The two disulfide bridges between two Cys32 and two Cys236 as well as the ADP molecules are shown as sticks. **(B)** Simulated annealing omit difference maps (Fo-Fc) calculated at 2.4 Å resolution and contoured at 4.5 rmsd. Maps were calculated after substituting both cysteines

with Ala (top two panels) and removing the ADP molecule (bottom panel). **(C)** Top view of AtFN3K showing the beta-strand exchange.

AtFN3K adopts a canonical protein kinase fold (PKL-fold) with an N-terminal ATP binding lobe and C-terminal substrate-binding lobe. A comparison to the bacterial search model (PDB code 5IGS) shows that the chains superimpose 218 Cα atoms with an rmsd of 3.3 Å [32]. Due to the strong similarity between the two structures and the fact that the closely related aminoglycoside phosphotransferase structure from *E. coli* (5IGS) has already been described in detail [33], we focus here on the unique structural features of AtFN3K. Unlike other kinases, the AtFN3K structure is a 'strand-exchange' dimer, in which the β1 strand in one chain forms an anti-parallel β-sheet with the β2 strand of the adjacent chain (Fig. 2.2C). In most PKL-fold crystal structures solved to date, the β1 strand forms an intra-chain anti-parallel beta-sheet with β2, and the P-loop connecting the two strands typically positions the ATP for catalysis, as observed in *Thermobifida fusca* (TfFN3K), aminoglycoside kinase and the prototypic eukaryotic Protein Kinase (EPK) A (Figs. 2.1, 2.3A). However, in the AtFN3K dimer, the P-loop is unfolded, and the extended P-loop conformation is stabilized by an intermolecular disulfide bond between Cys^{32} in chain A and Cys^{32} in chain B (Figs. 2.2B, 2.3A). The electron density of the Cys^{32} suggests that a small fraction of the disulfides may have cleaved by X-rays during data collection, but we could not model both conformations with confidence given the relatively low 2.4 Å resolution. A second intermolecular disulfide is formed between Cys^{236} in each chain (Figs. 2.2B, 2.3B). Cys^{236} is located in the F-G loop, which is typically involved in substrate binding in both small-molecule kinases and protein kinases (Fig. 2.3B). In the AtFN3K dimer, the substrate-binding lobes are covalently tethered to create a unique interface, presumably for phosphorylating ketosamine and related substrates (Fig. 2.3C). The AtFN3K dimer buries nearly 2500 $Å^2$ (16.4%) of solvent-

accessible area of each monomer. The total solvation free energy (ΔG) gain upon the formation of the interface is calculated to be -36.2 kcal/mol (p-value of 0.014) and a total of 36 hydrogen bonds and 2 disulfide bridges are formed at the dimer interface [34].

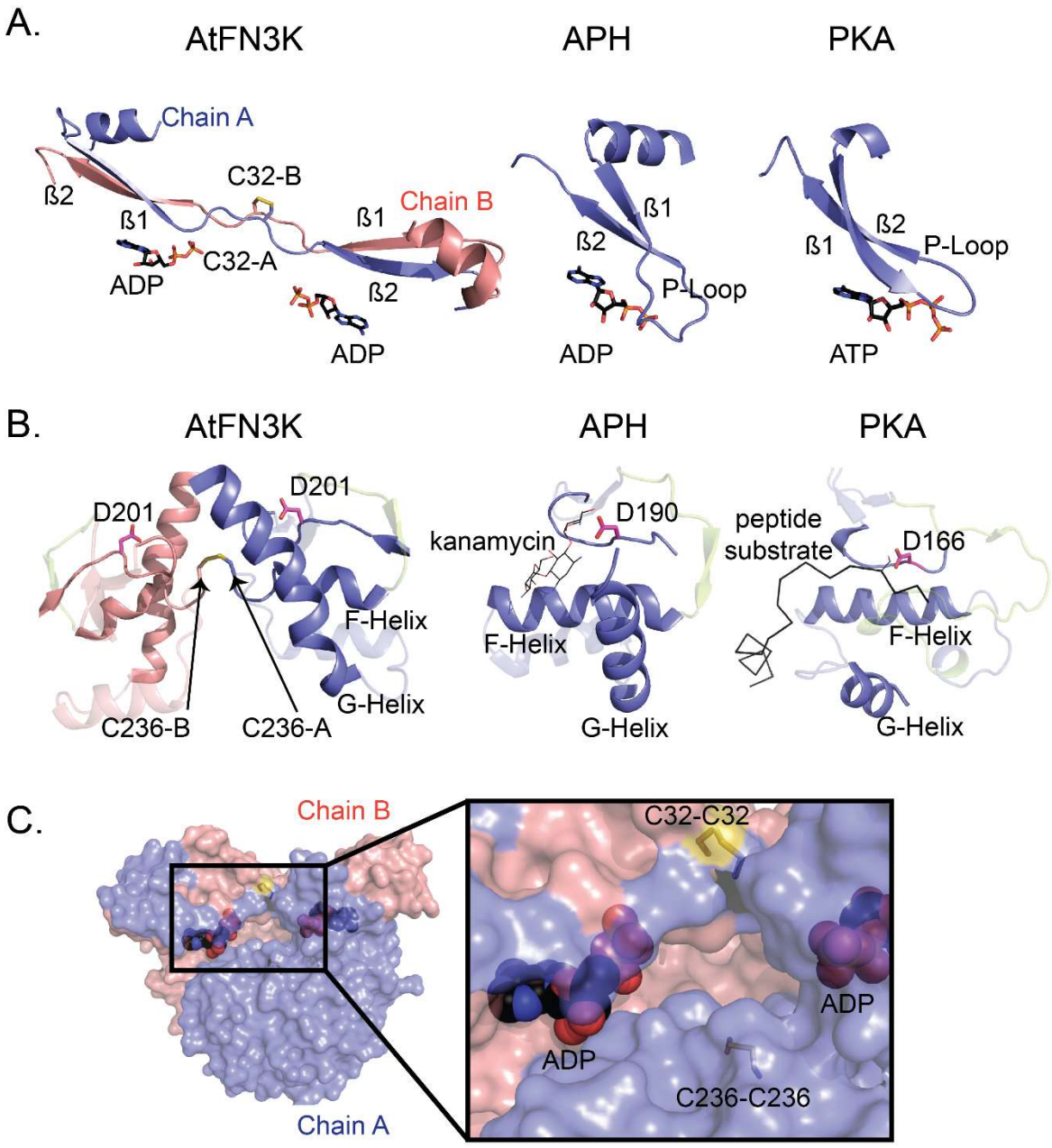


Figure 2.3: Comparison of AtFN3K ATP and substrate binding regions with APH and PKA.

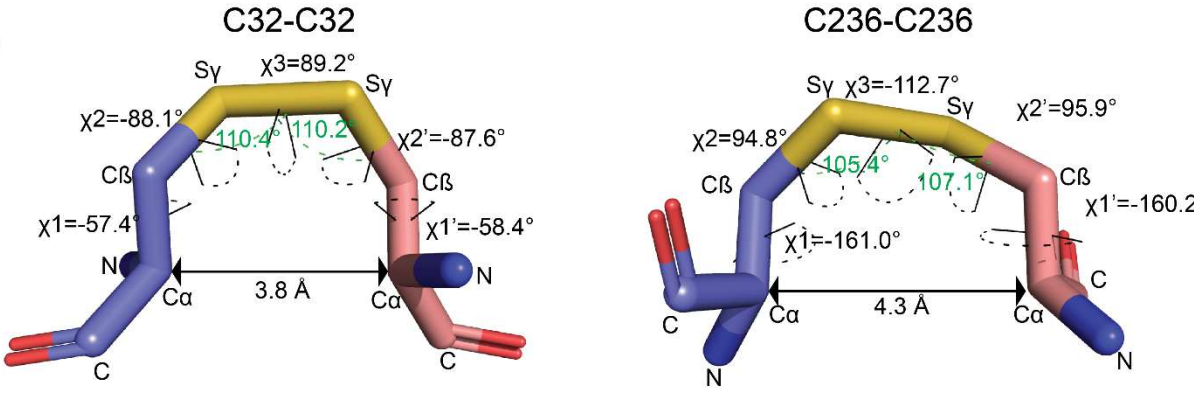
(A) Comparison of the P-loop of AtFN3K with that of APH and PKA. PDBs 1L8T and 1ATP were used for APH and PKA respectively. Carbon atoms of ADP and ATP molecules are colored in black, and the oxygen atoms are colored in red. Chain A and chain B of AtFN3K are colored in slate and salmon, respectively. **(B)** Comparison of the substrate binding lobe of AtFN3K with APH and PKA. Catalytic aspartate is shown as sticks with carbon atoms colored in magenta. The activation loop is colored

in limon. The APH substrate kanamycin is shown as lines with carbon atoms colored in magenta. The PKA peptide substrate is shown as ribbon and colored black. The serine residue in the peptide is modelled and shown as lines. PDBs used as in (A). **(C)** Surface representation of AtFN3K. Chains A and B and ADP-associated carbons are colored as described in (A). The two disulfide bridges, C32-C32 and C236-C236 are indicated with sticks with sulfur atoms colored in yellow.

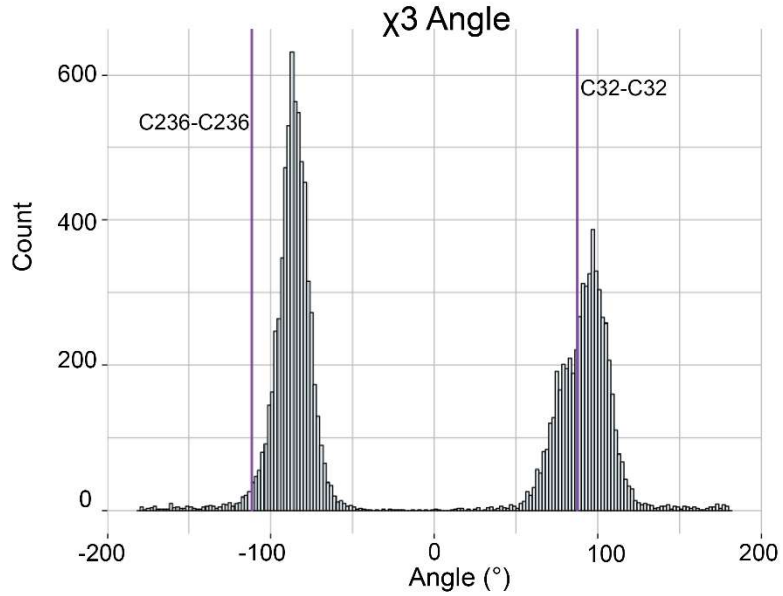
2.3.2 Disulfides in the AtFN3K dimer interface are conformationally strained.

Disulfide bonds in proteins can be classified into various categories based on their geometry and torsional strain energy [35-37]. Conformational analysis of inter-molecular disulfides in AtFN3K indicates non-ideal geometry (Fig. 2.4A). In particular, the χ3 dihedral of -112.7° (Cβ-Sγ- Sγ- Cβ) for Cys^{236}-Cys^{236} deviates from the peak distribution of ~90° observed for disulfides in the Protein Data Bank (PDB) (Fig. 2.4B). Likewise, the Cα-Cα distance of 3.8 Å for Cys^{32}-Cys^{32} is much shorter than 5.7 Å peak observed for Cα-Cα distances in the PDB [38, 39]. These non-ideal geometries suggest that the intermolecular disulfides, in particular, the Cys^{236}-Cys^{236} disulfide, are conformationally strained.

A.


C32-C32
Sγ
χ3=89.2°
Sγ
χ2=-88.1°
110.4°
110.2°
χ2'=-87.6°
Cß
Cß
χ1=-57.4°
χ1'=-58.4°
N
Cα
3.8 Å
Cα
N
C
C
C236-C236
Sγ
χ3=-112.7°
Sγ
χ2'=95.9°
χ2=94.8°
105.4°
107.1°
Cß
Cß
χ1'=-160.2
χ1=-161.0°
C
Cα
4.3 Å
Cα
C
N
N

B.


χ3 Angle
600
400
200
0
Count
C236-C236
C32-C32
-200
-100
0
100
200
Angle (°)

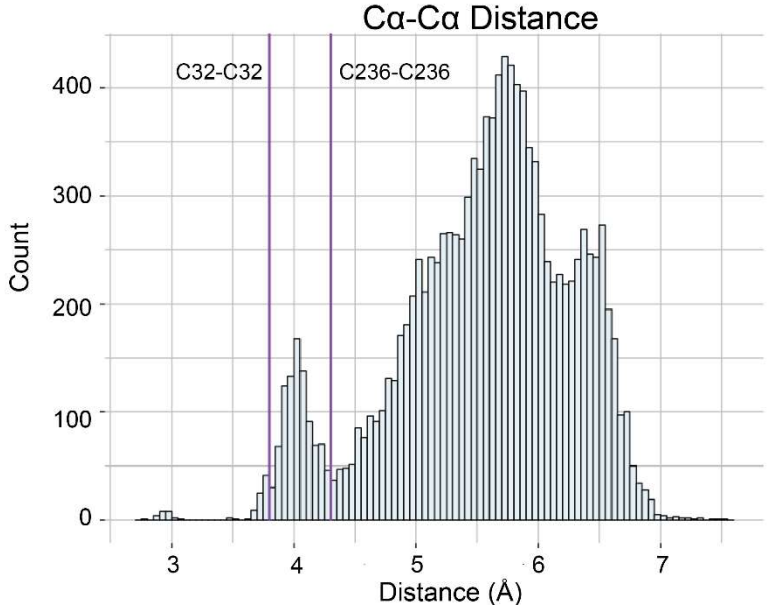

Cα-Cα Distance
400
300
200
100
0
Count
C32-C32
C236-C236
3
4
5
6
7
Distance (Å)

Figure 2.4: Geometric analysis of the disulfides in AtFN3K.

(A) Distances and dihedral values are labelled for the disulfide bridges between Cys^{32}-Cys^{32} and Cys^{236}-Cys^{236} between chain A and chain B. The angle (Cβ-Sγ-Sγ) is shown with a green dashed line. The dihedrals are shown with dashed line colored in black. **(B)** Distribution of the Cα- Cα distance and χ3 angle (Cβ-Sγ- Sγ- Cβ) for PDB structures with resolution less than 1.5 Å. Values for the Cys^{32}-Cys^{32} and Cys^{236}-Cys^{236} disulfides in AtFN3K is represented by a vertical line colored in magenta.

2.3.3 P-loop cysteine is critical for the formation of disulfide linked dimer species.

Next, we wanted to test if either or both cysteines are essential for the observed disulfide-linked dimer. We mutated the cysteine residues individually and together to alanine residues and subjected them to non-reducing SDS-PAGE analysis in the presence of redox agents, 1,4-Dithiothreitol (DTT) and hydrogen peroxide (H_2O_2). The disulfide-linked dimer is absent in the P-loop cysteine mutants (Cys32Ala and Cys32Ala/Cys236Ala) but not in the Cys236Ala mutants (Fig. 2.5A), suggesting that the P-loop cysteine is critical for the observed disulfide-linked dimer. In addition, we identified multiple monomeric species on the gel, including a reduced monomer (M_{Red}) and monomers with intramolecular disulfide ($M_{S\text{-}S}$) (Fig. 2.5A). The presence of $M_{S\text{-}S}$ band in the double cysteine mutant led us to hypothesize that the two remaining cysteines, Cys^{196} and Cys^{222} (Fig. 2.5B), in the kinase domain could form intramolecular disulfides. To test if an intramolecular disulfide is formed in solution or upon denaturation, we treated the Size Exclusion Chromatography (SEC) purified WT dimer with N-Ethylmaleimide (NEM) before denaturation. NEM blocks free thiols and prevents their oxidation and subsequent disulfide formation [40]. The absence of $M_{S\text{-}S}$ and the presence of M_{Red} in the NEM treated sample suggests that the $M_{S\text{-}S}$ band is a consequence of denaturation. As a control, we also ran a non-reducing SDS page on the triple cysteine mutant (Cys32Ala/Cys236Ala/Cys196Ala), which is incapable of forming any intramolecular disulfides. As expected, only M_{Red} band was observed in the triple mutant, further suggesting that the $M_{S\text{-}S}$ band is a consequence of denaturation during SDS page analysis. For this reason, we primarily focus on the inter-chain disulfides observed in the crystal structure.

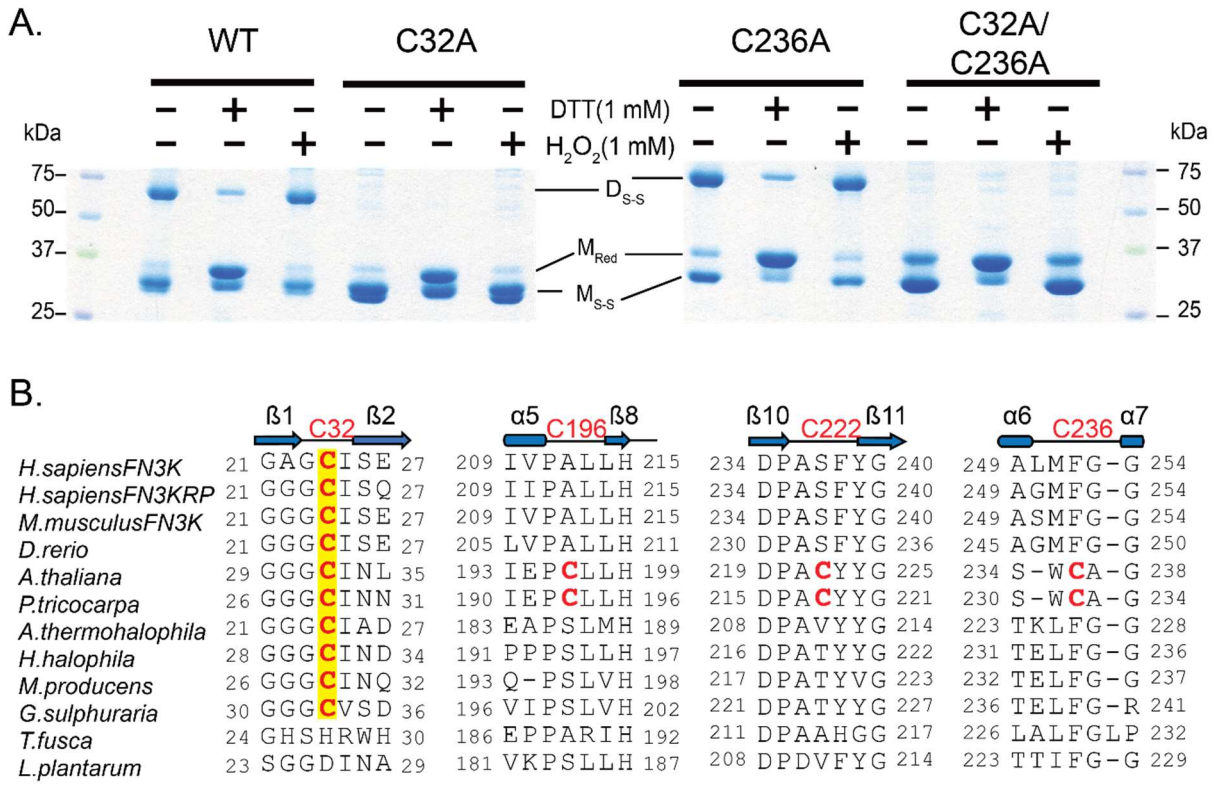


Figure 2.5: P-loop cysteine (Cys32) is critical for the formation of disulfide-linked dimer.

(A) Non-reducing SDS-PAGE of AtFN3K WT and Cys to Ala mutants. 15 µg of protein was incubated with 1 mM DTT or 1mM H_2O_2 for 20 minutes and was then subjected to SDS PAGE under non-reducing conditions. $D_{S\text{-}S}$: Disulfide linked dimer; M_{Red}: Monomer Reduced; $M_{S\text{-}S}$: Monomer with intramolecular disulfide. **(B)** Multiple sequence alignment of FN3K orthologs. Two additional cysteines (Cys196 and Cys222) specific to plant FN3Ks are shown. The alignment was generated using MUSCLE [41].

2.3.4 AtFN3K WT dimer is activated by redox agents.

In order to determine the functional state of the native enzyme (WT) in solution, we performed SEC on the purified protein. We identified two peaks, a dimer and a monomer (Fig. 2.6A) with the dimer peak being dominant in solution. Compared to a reducing SDS-PAGE, the purified dimer resolved on non-reducing SDS-PAGE gel as two distinct bands. To test sensitivity to DTT, we performed Pyruvate Kinase/Lactate Dehydrogenase (PK/LDH) assay on the dimer and

monomer fractions in the presence and absence of 2 mM DTT. The activity of the dimer species increased nearly 40fold in the presence of the reductant (Fig. 2.6B). In contrast, the monomer species was active but insensitive to DTT (Fig. 2.6B), suggesting that only the dimer species is redox-sensitive. The activity of the dimer species increased with increasing concentration of reduced glutathione (GSH), a physiological reductant. We also performed PK/LDH assays on single and double cysteine to alanine or serine mutants. P-loop cysteine (Cys32/Ala/Ser) mutants were less sensitive to DTT compared to C-lobe cysteine mutants (Cys236/Ala/Ser). Moreover, AtFN3K WT also showed sensitivity to DTT when activity was measured independently using Nuclear Magnetic Resonance (NMR)-based assay which measures ribuloselysine phosphorylation.

To test if AtFN3K could dimerize without the cysteines, we performed SEC on the triple cysteine mutant (Cys32Ala/Cys236Ala/Cys196Ala). Like the native enzyme, we identified both dimer and monomer peaks (Fig. 2.6C). We hypothesized that the two species (monomer and dimer) would be insensitive to DTT since they are incapable of forming disulfide bridge. As expected, both the dimer and monomer species were insensitive to DTT (Fig. 2.6D). Taken together, these data suggest that AtFN3K can still dimerize in the absence of two cysteine residues (Cys^{32} and Cys^{236}), which lie at the dimer interface.

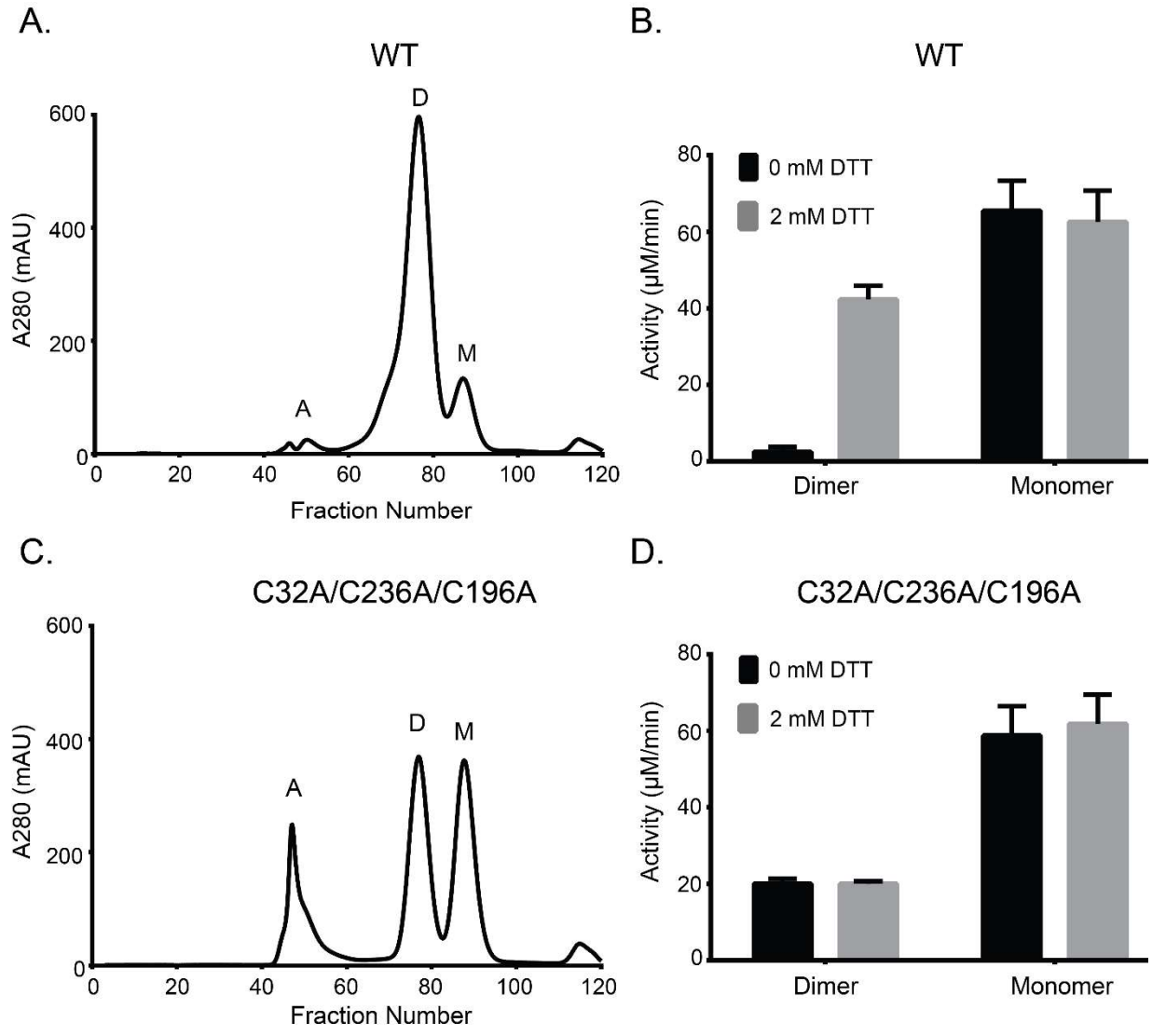


Figure 2.6: Both WT and triple cysteine mutant (Cys32Ala/Cys236Ala/Cys196Ala) exist as two distinct species in solution with the WT dimer being redox sensitive.

(A) Size Exclusion Chromatography (SEC) of AtFN3K WT protein. **(B)** Activity of WT protein in the presence and absence of 2 mM DTT. **(C)** Size Exclusion Chromatography (SEC) of the triple cysteine mutant protein. **(D)** Activity of WT protein in the presence and absence of 2 mM DTT. **(A, C)** Each fraction was 1 ml in volume. A: Aggregates, D: Dimer, M: Monomer. **(B, D)** 1 µg of protein was assayed using PK/LDH assay as mentioned in the methods. Ribulose-N-α-Ac-lysine was used as the substrate. Error bar represents the standard error of six independent experiments

To test whether the monomer and dimer species were in equilibrium, we first performed sedimentation velocity analysis on the purified Cys32Ala/Cys236Ala/Cys196Ala triple mutant. c(s) analysis revealed a distribution consisting of a monomer peak at 2.0 S and a dimer peak at 3.1 S which are predicted to be at 2.2 S and 3.6 S, respectively. Glycerol is required to maintain the recombinant AtFN3K stably in solution and tends to form a self-gradient during centrifugation which will disproportionately reduce the S-values of sedimenting species. The monomer-dimer distribution is consistent with the size exclusion profile suggesting that the protein exists in two predominant species (Fig 2.6C). Secondly, we performed sedimentation velocity analysis across a moderate change (5x) in protein concentration but did not detect any changes in the ratio of monomers and dimers present. To confirm that the species are not exchanging to any significant extent, we purified the dimer and monomer species using SEC and repeated the sedimentation velocity analysis. The c(s) distribution of the purified dimer shows no significant dissociation but does contain a 1.3 S species that may correspond to a small amount of misfolded monomer. The c(s) distribution of the purified monomer suggests only a small amount of aggregation, as evidenced by a slight tailing of the c(s) peak toward larger s-values.

2.3.5 Evolution of P-loop cysteine redox regulation

Multiple sequence alignment of FN3K orthologs from diverse organisms indicates that while the P-loop cysteine (Cys^{32}) is conserved across diverse organisms, the other three cysteines (Cys^{236}, Cys^{196}, and Cys^{222}) are unique to plant FN3Ks (Fig. 2.5B). Notably, human FN3K (HsFN3K) and human FN3K-Related Protein (HsFN3KRP) both contain an equivalent cysteine in the P-loop. HsFN3K contains three additional cysteines, two of which are specific to both FN3K and FN3K-RP. We expressed and purified HsFN3K and assayed its activity under reducing conditions. The activity of HsFN3K increased in the presence of DTT like that observed for

AtFN3K (Fig. 2.7A). Mutating the equivalent P-loop cysteine (Cys^{24}) to alanine results in loss of sensitivity to DTT. We also tested the activity in the presence of GSH. The activity increased for the WT whereas no change was observed for the Cys24Ala mutant. As a control, we expressed, purified, and assayed AtFN3K homologs from *Thermobifida fusca* (TfFN3K) and *Lactobacillus plantarum* (LpFN3K), which contain histidine and aspartate in place of the P-loop cysteine, respectively (Fig. 2.5B). Notably, TfFN3K and LpFN3K catalytic activity did not change in the presence of any of the redox agents tested (Fig. 2.7B). Also, non-reducing SDS PAGE for the WT and Cys24Ala HsFN3K confirmed that the P-loop cysteine is essential for dimerization.

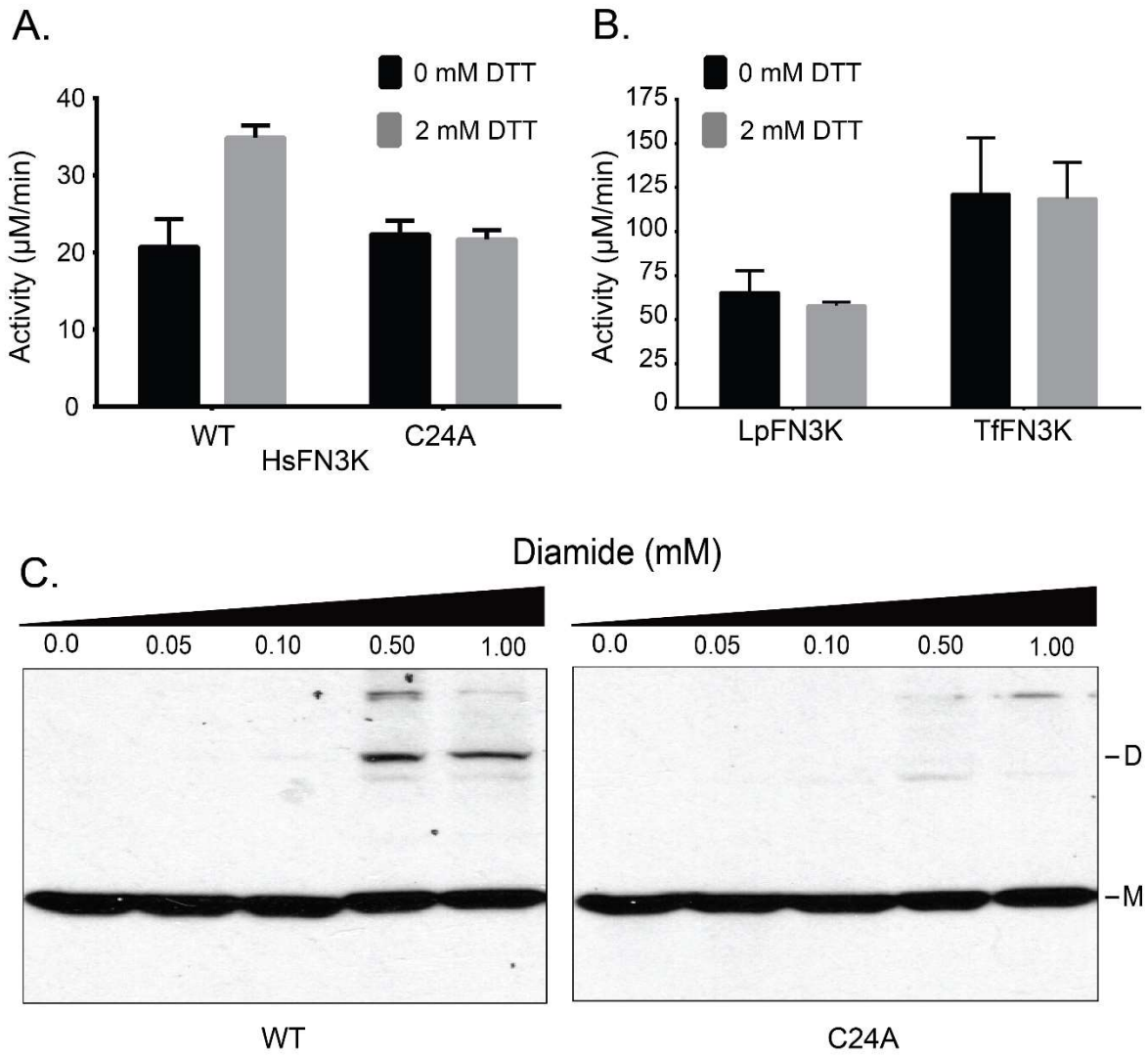


Figure 2.7: P-loop cysteine contributes to redox-sensitivity in HsFN3K.

(A and B) PK/LDH assays performed with 10.0 μg of HsFN3K (A) or 1.0 μg each of TfFN3K and *L. plantarum* FN3K (LpFN3K) **(B)** Proteins were incubated with buffer (0 mM DTT) or 2 mM DTT, and ribulose-N-α-Ac-lysine was used as the substrate. Data are mean ± standard error of three independent experiments. **(C)** Effect of different diamide concentrations on transfected Flag-tagged WT and C24A HsFN3K in HEK293 cells. Total cell lysates were immunoblotted for Flag. $D_{S\text{-}S}$: disulfide-linked dimer, M: monomer. Blot is representative of 3 experiments.

Finally, to extend these findings in mammalian cells, we overexpressed Flag-tagged HsFN3K in HEK293T cells using transient transfection. Cells were grown for 48 hours and treated with different concentrations of diamide. Diamide decreases the cellular concentration of GSH and promotes disulfide formation [42]. It can also react with free thiols in proteins and lead to disulfide

bond formation [43]. Diamide treatment induced higher-order oligomerization of WT HsFN3K, but not for Cys24Ala mutant (Fig. 7C).

2.3.6 FN3K CRISPR knockout alters redox sensitive cellular metabolites.

Cellular functions of FN3Ks are currently unknown. Our analysis of FN3K expression in various cancer cell lines identified significant overexpression in the liver and eye cancer cells. To assess the functional significance of this increased expression, we generated a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) knock out of HsFN3K (FN3K-KO) in HepG2 liver cancer cell line and compared the metabolome of WT and FN3K-KO cells using untargeted 1H nuclear magnetic resonance (^{1}H NMR) metabolomics. This revealed a significant difference in metabolite abundance in FN3K-KO compared to WT (Fig. 2.8A, B). Of special note, glutathione and lactate levels were increased in FN3K knockout cells relative to wild type HepG2 cells, while pantothenate, phosphocreatine/creatine ratio, aspartate, glycine, and serine levels were decreased. Glutathione is a major cellular redox regulator [44], while intracellular levels of pantothenate, glycine, serine, and aspartate, are known to be reactive to the redox status of the cell [45-47]. Additionally, phosphocreatine/creatine ratio and lactate control ATP production and glycolysis, respectively [48]. The enrichment of these metabolites suggests potential links between FN3Ks, redox levels, and ATP production.

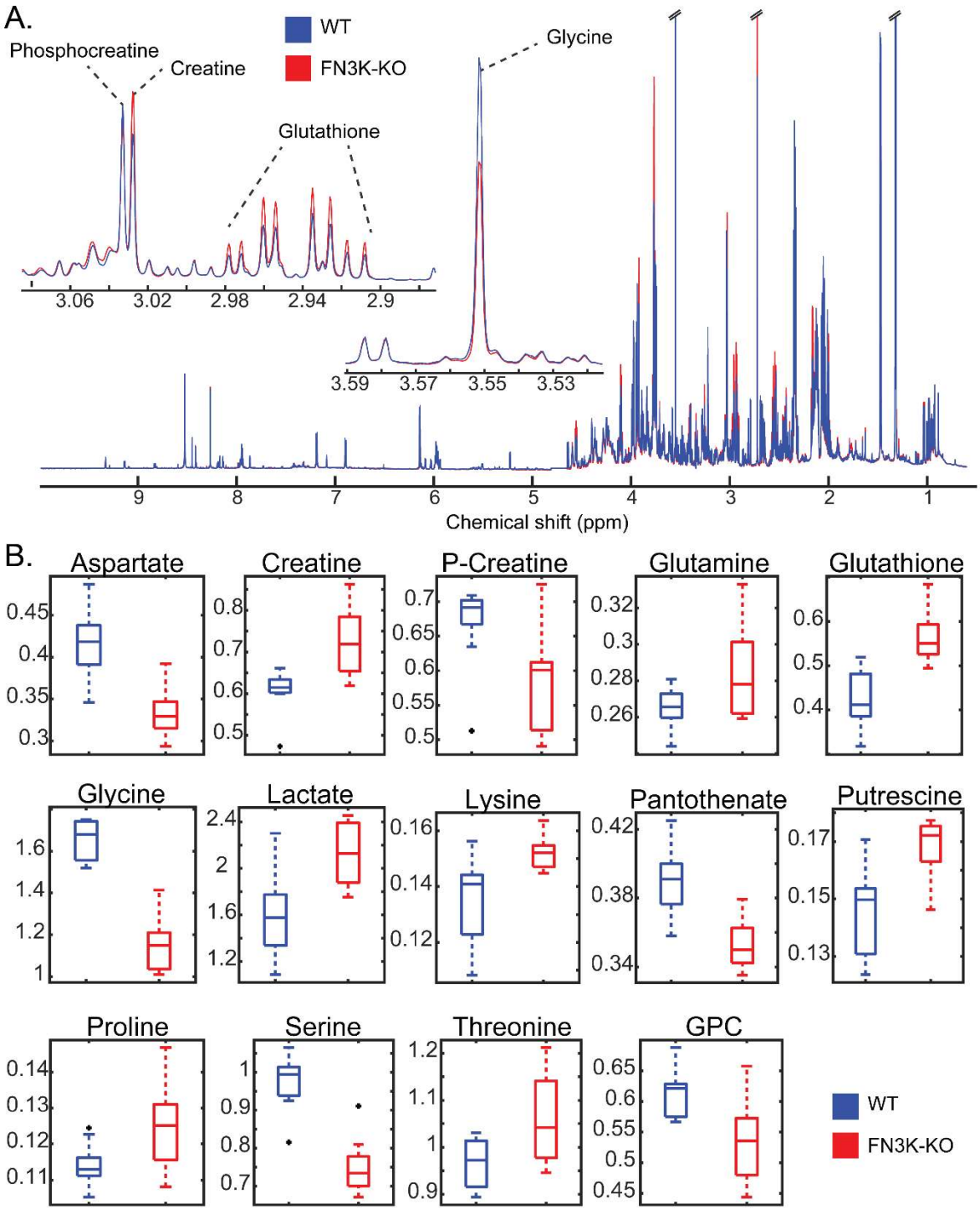


Figure 2.8: Redox sensitive metabolites are altered in HsFN3K CRISPR knockout.

(A) ^{1}H NMR spectra of HepG2 (WT) and FN3K-Knock Out (FN3K-KO) cells. Traces are the average for each group (HepG2 N=10, FN3K-KO N=9). Insets highlight examples of regions containing annotated metabolites observed to be significantly different between cell lines. **(B)** Box and whisker plots of significant (FDR p-value < 0.05) metabolites annotated with highest

confidence. Black points indicate outliers. Two-tailed T-test was performed, and the p-values were corrected for false discovery rate using the Benjamini-Hochberg method.

2.4 Discussion

Here we report the structural basis for redox regulation in an ancient family of phosphorylation-based enzymes associated with protein repair. The crystal structure of *Arabidopsis thaliana* FN3K reveals a novel strand exchange dimer in which two chains are covalently linked by disulfide bonds emanating from the P-loop in the ATP binding lobe and F-G loop in the substrate binding lobe. We propose that these disulfide bridges are conformationally strained and are essential for maintaining AtFN3K in an inhibited dimeric conformation. Our studies also show that inhibited native AtFN3K (WT dimer) is activated more than 40-fold under reducing conditions. We speculate that reduction of the disulfides releases the constraint on the P-loop and the substrate-binding F-G loop to allow ATP and substrate access, and catalysis in the dimeric state. Thus, AtFN3K toggles between an inhibited "oxidized" dimer state and active "reduced" dimer to phosphorylate ketosamine substrates. Using SEC, we also identified a monomeric species insensitive to DTT. Although crystallization attempts to isolate the monomeric species have been unsuccessful, we predict that a monomer would adopt a canonical kinase fold without any beta-strand exchange. As a result, both P-loop and substrate-binding loops cannot be constrained through disulfide bridges, preventing redox regulation.

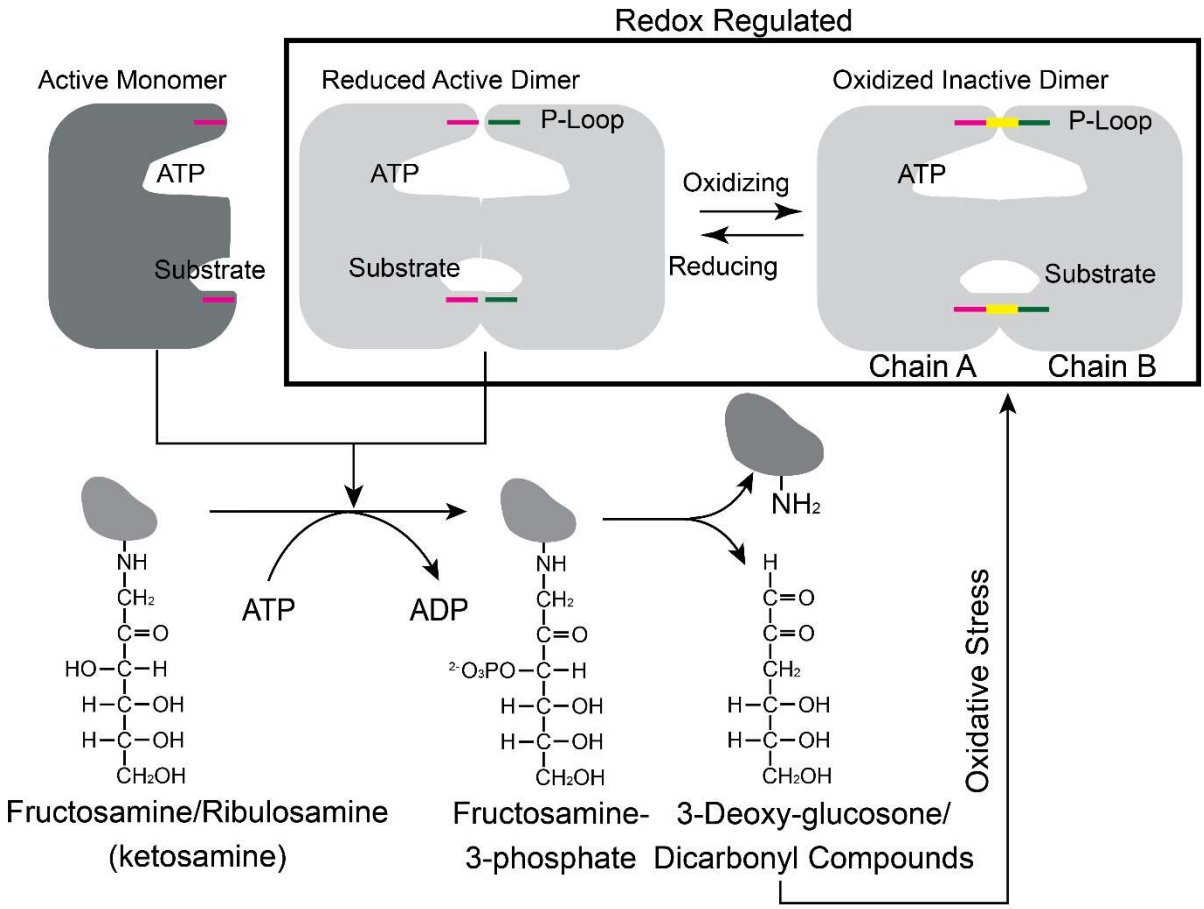


Figure 2.9: Proposed redox feedback regulation of plant and mammalian FN3Ks.

Cartoon showing the possible relationship between redox regulation of FN3K activity and its physiological function. The disulfide is colored in yellow.

Similarly, SEC employed on the Cys32Ala/Cys236Ala/Cys196Ala mutant revealed the presence of both dimeric and monomeric species, indicating that the enzyme can dimerize without these cysteines. PISA analysis of the crystal structure had shown that nearly 2500 $Å^2$ (16.4%) of each monomer is buried as part of the dimer interface. Sedimentation velocity analysis further revealed that the two species were distinct and did not exist in equilibrium.

Our cellular studies show that dimerization and higher-order oligomerization of human FN3K is altered after exposure to the thiol-oxidizing agent, diamide. Consistently, an unbiased

proteomic study identified a sulfenylated form of the P-loop cysteine in human FN3K-RP in HeLa cells [49], suggesting that a regulatory mechanism of partial Cys P-loop oxidation to sulfenic acid (which is a reversible modification) is likely to be operative in cells.

The identification of redox-active cysteines in FN3Ks also opens up the possibility of a feedback regulation mechanism in which FN3K activity is controlled by its catalytic byproduct, 3-deoxyglucosone, which is known to contribute to oxidative stress [20]. We speculate that the accumulation of advanced glycation end products such as 3-deoxyglucosone drives the conformational ensemble of FN3K species towards an inactive dimeric form through the oxidation of the P-loop cysteine, while reduced levels maintain FN3K in an active reduced form (Fig. 2.9). Such a feedback inhibition mechanism might explain how the essential deglycation functions of FN3K are carried out in cells without contributing to oxidative stress. Consistent with this view, our biochemical and mutational studies confirm that inhibited dimeric species predominate in the absence of reducing agents, while active dimeric or monomeric species may dominate in the presence of reducing agents. Moreover, our comparative metabolomics study of the HsFN3K WT and KO in HepG2 liver cells revealed alteration in redox-sensitive metabolites, namely glutathione. Increased cellular glutathione in the KO relative to WT suggests potential links between FN3K functions and oxidative stress response that needs to be tested in future studies.

Redox control of FN3K may also explain the altered glycated patterns observed in plant and human proteomes [50-53] and facilitate deglycation functions to be carried out in different cellular compartments. AtFN3K contains a chloroplast signal peptide N-terminus of the kinase domain, and removal of the signal peptide results in the localization of AtFN3K in different cellular compartments, including nucleus and mitochondria. Human FN3K is localized in the mitochondria [54] and has been detected in red blood cells and serum samples. Mitochondria generates reactive

oxygen species (ROS), which can mediate redox signaling through oxidation of SH groups in cysteine residues in proteins [55-57]. The presence of a redox-sensitive P-loop cysteine (Cys24) in HsFN3K further suggests that redox regulation of FN3K is physiologically relevant.

The P-loop Cys is conserved amongst FN3K orthologs from diverse prokaryotic and eukaryotic organisms, and our studies indicate that the FN3K homolog in humans (HsFN3K) is also redox-sensitive. In contrast to the activation loop, which appears to have evolved as a flexible motif for eukaryotic protein kinase regulation by phosphorylation, the P-loop is fundamentally conserved in diverse ATP fold enzymes, where it plays a critical role in clamping ATP and positioning it for efficient catalysis. Here we show that conformational control of the P-loop by reactive cysteines is an ancient mode of regulation. In this context, it is interesting to note that several tyrosine protein kinases also possess conserved Cys residues embedded within the Gly-rich motifs in the P-loop. These include a Cys at the position equivalent to Cys^{32} of AtFN3K found in eukaryotic SRC, FGFR, YES1 and FYN tyrosine kinases. Notably, both SRC and FGFR have previously been shown to be redox-regulated through Cys oxidation mechanisms in the P-loop [58, 59]. However, whether these kinases also form a strand exchange dimer like AtFN3K is not known, and it will be interesting to attempt to solve the structure of Tyr kinases crystallized under different oxidizing conditions.

Finally, human FN3K and FN3K-RP are implicated in diseases such as diabetic neuropathy and retinopathy, which are associated with increased protein glycation. Both kinases are also highly expressed in human tumors. However, developing therapeutic strategies for FN3Ks is potentially a double-edged sword because inhibition of FN3K activity can result in the accumulation of glycated proteins, while activation of FN3Ks might result in the accumulation of 3-deoxyglucosone, a potent generator of oxidative stress. Based on our new findings, we propose

a combinatorial drug development approach, in which the active monomeric species and the reduced dimeric species are targeted through the development of thiol-interacting agents. These might be developed to help normalize cellular protein glycan homeostasis, depending upon the indication and redox balance in the appropriate tissue and the availability of redox-active Cys residues in FN3K targets.

2.5 Materials and Methods

2.5.1 Expression and purification of AtFN3K WT and Cys-Ala variants

AtFN3K inserted into pET-15b expression vector was provided by Prof. Emile Van Shaftingen (Université Catholique of Louvain, Brussels, Belgium). Site-directed mutagenesis was performed using Q5 Hot Start High Fidelity Kit from New England Biolabs (NEB). Vectors were transformed into E. coli BL21(DE3) pLysS (Roche) cells and co-expression with GroEL was performed at 16°C for 18 hours in 500 ml of Luria–Bertani medium containing 100 μg/l ampicillin and 25 μg/l chloramphenicol. The bacterial extract was resuspended in 50 mM sodium phosphate buffer, pH 7.8, 300 mM NaCl, 10 % v/v glycerol (Buffer A) with 1mM Phenylmethylsulfonyl Fluoride (PMSF) and 3 mM 2-Mercaptoethanol (βME). The solution was sonicated on ice and the lysate was centrifugated at 15,000 x g for 45 minutes at 4°C. The supernatant was then passed through 0.5 ml of Talon® resin previously equilibrated with Buffer A without βME. The column was washed with 20 ml of Buffer B (50 mM Sodium phosphate buffer, pH 7.8, 1 M NaCl,10 % v/v glycerol) followed by a wash with 100 ml of Buffer B containing 10 mM imidazole. The protein was then eluted with Buffer A without βME containing 250 mM Imidazole. The purity of the protein was checked by SDS-PAGE, and the fractions with the protein were dialyzed against

buffer A without glycerol and later against Buffer C (25 mM HEPES, pH 7.8, 300 mM NaCl, 10% v/v Glycerol). The protein was snap-frozen with liquid nitrogen and stored at -80°C. Recombinant His-HsFN3K was co-expressed with GroEL in *E.coli* as well.

2.5.2 Crystallization, phasing, and refinement

Recombinant AtFN3K WT was dialyzed into a buffer containing 50 mM NaCl and 15 mM HEPES at pH 8.0, concentrated to ~30 mg/mL and quantified using the extinction coefficient (Abs 0.1%: 1.576) calculated by Protparam [60]. The protein was crystallized at 20°C using hanging drop vapor diffusion with 2 µl drops (1:1 protein to reservoir ratio). Crystals grew from a reservoir of 2 M ammonium sulfate, 0.1 M MES pH 5.5, 3.28 mM $MgCl_2$ and 1.9 mM AMP-PNP. Crystals were cryoprotected with the reservoir solution supplemented with 10 % cryoprotectant (1:1:1 ratio of ethylene glycol, DMSO, and glycerol) and flash cooled in liquid nitrogen. The diffraction data were collected at the SER-CAT 22-ID beamline at the Argonne National Laboratory using a Rayonix 300HS detector and processed using XDS [61]. Five percent of the data was set aside for cross validation. The structure was solved by molecular replacement in Phenix [62] using a multiple model template (structures from *T. fusca* (PDB code 3F7W), *H. somnus* (PDB code 3JR1) and *E. coli* (PDB code 5IGS) were used) with trimmed side chains and weighted according to structure and homology. Sequence matching and model building was performed using Phenix Autobuild [62] . Automated refinement in Phenix [62] and iterative manual fitting using Coot [63] produced the final model. The B-factors were refined using TLS [64].

2.5.3 Distribution of χ3 angle and Cα-Cα distance.

Table S1 titled "Excel spreadsheet of unique disulfides in a culled set of X-ray structures described by G.Wang and R. Dunbrack, Jr." from [65] was downloaded and PDBs with a resolution less than or equal to 1.5Å was selected to plot the histogram showing the distribution of the χ3 angle and Cα-Cα distance.

2.5.4 SDS-PAGE

15 µg of the enzyme was incubated with distilled H_2O or redox agents (DTT/H_2O_2) in a total volume of 15 µL. After 20 minutes, 5 µL of 4X Sample Buffer without 2-Mercaptoethanol (βME) was added and the protein was denatured at 100˚C for 5 minutes followed by loading 10 µL of the sample onto a 12% SDS-PAGE gel. The gels were stained with Coomassie for an hour and de-stained with water for several hours. Similarly, 3-5 µg of the enzyme (15 µL volume) was incubated with 20 mM NEM or buffer (15mM HEPES, pH 7.8, 300 mM NaCl, 10 % v/v Glycerol) for 5 mins at room temperature before adding 4X Sample Buffer with (Reducing) or without (Non-reducing) 2.8 M 2-Mercaptoethanol (βME).

2.5.5 Sequence Alignment

Representative FN3K orthologs were downloaded from Uniprot [66], NCBI [67] and aligned using MUSCLE [41]. Eukaryotic protein kinases (EPKs) with P-loop cysteine conserved were identified using previously curated alignment profiles that included EPKs and small molecule kinase sequences including FN3Ks [23, 24, 68, 69].

2.5.6 Size Exclusion Chromatography (SEC)

Approximately 4.5 mg of AtFN3K WT and triple cysteine mutants purified using co-expression with GroEL/ES chaperone (9 mg/ml in concentration) was passed through HiLoad 16/1600 Superdex 200 pg column at a flow rate of 1 ml/min with fractionation volume of 1 ml. 25 mM HEPES, pH 7.8, 300 mM NaCl, 10 % v/v glycerol was used as running buffer. SEC was performed at 4 °C. For BSA-Lysozyme standard, 250 µl of 4 mg/ml of BSA and Lysozyme each in running buffer were mixed and loaded on the column. For AtFN3K WT and triple cysteine mutants purified without GroEL chaperone, SEC was performed at a flowrate of 0.15 ml/min.

2.5.7 Enzyme Assays

PK/LDH: 1.0 µg (5 µL) of FN3K homologs (At, Tf, Lp)/mutants/dimer and 10 µg (5µL) of HsFN3K, 28 mM of Ribulose-N-α-Ac-lysine (10 µL) in the presence or absence of 20 mM DTT (5 µL) (Final concentration: 2 mM) were mixed with 20 µl of solution prepared by mixing the following: 150 µl of 5X Kinase Buffer (160 mM HEPES, pH 7.4, 80 mM MgCl2, 1.2 M NaCl, 40 % v/v Glycerol), 15 µl of 250 mM Phosphoenolpyruvic acid (PEP), 45 µl of PK/LDH mix [600-100 units/ml pyruvate kinase (PK), 900-1400 units/ml lactic dehydrogenase (LDH)], 90 µl of 37.5 mM Nicotinamide adenine dinucleotide (NADH). The reaction was started by adding 10 µL of 5 mM ATP (final 1mM). The final reaction volume was 50 µL per well. The 96well plate was immediately placed in a plate reader (Biotek Synergy H4) and the absorbance measured at 340 nm at 35°C continuously for two and half hours. Proteins were stored in Buffer D (25 mM HEPES, pH 7.4, 300 mM NaCl, 10% v/v Glycerol). Buffer D was also used as mock buffer as needed.

NMR real time assay: 100 μg of protein was incubated with 2 mM of ribuloselysine and 600 μM of ATP in 25 mM HEPES, pH 8.0, 5 mM of $MgCl_2$ prepared in D_2O. The reaction was monitored for 50 minutes by acquiring subsequent 1D 1H PURGE (Presaturation Utilizing Relaxation Gradients and Echoes) spectra. All spectra were acquired at 22 °C on an 800 MHz Avance Neo (Bruker) NMR spectrometer equipped with a z-gradient triple resonance TCI cryoprobe and normalized and references using 1H signals of 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). 32 scans were acquired during each experiment with an acquisition time of 1.25 sec. The spectra were processed using NMRpipe and MATLAB.

2.5.8 Sedimentation velocity

Triple cysteine mutant (Cys32Ala/C236Ala/Cys196Ala) was dialyzed into 25 mM HEPES (pH 7.8), 300 mM NaCl, 2 mM $MgCl_2$, 100 μM ATP, and 1 M glycerol, and the protein was quantified using an Agilent 8453 UV/vis with an ε280 of 55810 M-1cm-1 determined by ProtParam [60] . The sample was diluted to a final protein concentration of 10 μM then loaded into a 12 mm double-sector Epon centerpieces equipped with quartz windows. The cell was loaded into An60 Ti rotor and equilibrated for 1 hour at 20°C. Sedimentation velocity data were collected using an Optima XLA analytical ultracentrifuge (Beckman Coulter) at a rotor speed of 50000 RPM at 20°C. Data were recorded at 280 nm in radial step sizes of 0.003 cm. SEDNTERP [70] was used to model the partial specific volume of AtFN3K (0.713293 mL/g), as well the density (1.0353 g/mL) and the viscosity (0.013277 P) of the buffer. SEDFIT [71] was used to analyze the raw sedimentation data. Data were modeled as continuous c(s) distribution and were fit using baseline, meniscus, frictional coefficient, and systematic time-invariant and radial-invariant noise. Fit data for the experiment had an RMSD in the range of 0.005–0.007AU. Predicted sedimentation

coefficient (s) values for AtFN3K monomer (2.2 S) and dimer (3.6 S) were calculated from the atomic coordinate of AtFN3K using HYDROPRO [72]. Data fit and c(s) distribution plots were generated using GUSSI [73].

2.5.9 Diamide treatment of HEK293T cells

HEK293T cells were cultured in Dulbecco's Modified Eagle Media (DMEM) containing 10% Fetal Bovine Serum (FBS) on a 6 cm plate (total 6 plates) and allowed to grow overnight. Cells were transfected with 10 μg of Flag-tagged HsFN3K (EX-W1392-M46, GeneCopoeia) using Calcium Phosphate Transfection protocol (43). Cells were allowed to grow for 48h. After 48h cells were treated with indicated concentrations of Diamide (Sigma) for 2 h. Cells were lysed in buffer containing 50 mM Tris.HCl, pH 7.5, 150 mM NaCl, 10% glycerol, 1% TX-100 and 1x protease inhibitor cocktail (EMD-Millipore). Total Cell Lysate (TCL) was spun at 15,000 rpm for 10 min in a refrigerated centrifuge. Proteins were resolved on a 12% SDS-PAGE gel, transferred on PVDF membrane and detected by Western Blotting using anti-Flag antibody (Cell Signaling Technology).

2.5.10 CRISPR Knock Out (KO) cell line and Metabolomics

HepG2 cells were transfected with FN3K Double Nickase Plasmid (sc-412985-NIC) (Santa Cruz Biotechnology) and selected with puromycin according to manufacturer's protocol. FN3K deletion was confirmed by Western Blotting using an anti-FN3K antibody (Invitrogen). Stable cell line stocks were made from single cell colony. Adherent cells were harvested in ice-cold 80% methanol extraction solvent and flash-frozen in liquid nitrogen. Aqueous metabolites were extracted by vortexing/lysing cell pellets in the extraction solvent and collecting the

supernatant. 10% of the supernatant was taken from each sample to form an internal pooled sample. The solvent was then evaporated to produce dried extracts using a CentriVap Benchtop Vacuum Concentrator (Labconco, Kanas City, MO, USA). Extracts were reconstituted in a deuterium oxide phosphate buffer (pH 7.4) before data acquisition. 1H NMR) spectra were acquired on all samples using noesypr1d pulse sequence on an 800 MHz Bruker Avance III HD spectrometer. 1H-13C heteronuclear single quantum correlation (1H-13C HSQC) and 1H-1H total correlation spectroscopy (1H-1H TOCSY) spectra were acquired on the internal pooled sample and used for metabolite annotation with COLMARm [74]. 1H NMR spectra were annotated and analyzed using an in-house MATLAB toolbox to identify significant differences in metabolites between cell lines.

2.5.11 Subcellular Localization

Construction of plasmids for N. benthamiana transformations: For constructs used in the transient assays, full-length wild type or mutant coding sequences (CDS) of AtFN3K were cloned into pENTR/D-TOPO Gateway entry vector (Invitrogen) following the manufacturer's protocol. The coding regions were recombined into binary destination expression vector: pSITE-2CA N-terminal GFP (green fluorescent protein) fusion.

Transient expression of proteins in Nicotiana benthamiana: Agrobacterium tumefaciens strain C58C1 was used for the transient transformation. Colonies carrying the binary plasmids were grown at 28 °C on LB medium plates that contained 50 µg/ml gentamycin and 25 µg/ml rifampicin for selection of the strain, and 100 µg/ml spectinomycin for selection of the binary vector. For agroinfiltration, single colonies were grown overnight in 3 ml LB (28°C, 220 rpm). 50 µl of the agro suspension was added to 5 ml LB and the culture was grown overnight. The agro

was pelleted by centrifuging at 4,000 rpm for 20 minutes or 5,0000 rpm for 15 minutes and the suspension was adjusted to an OD_{600} of 0.2-0.3 in infiltration buffer containing 10 mM $MgCl_2$, 10 mM MES, pH 5.7 and 150 mM acetosyringone, pH 5.6, and incubated at room temperature for 3 hours before infiltration. To enhance transient expression of the fusion proteins, the viral suppressor of gene silencing p19 protein was co-expressed. For co-infiltration, equal volumes of cultures were mixed and infiltrated into *N. benthamiana* leaves through the abaxial surface using a 1 ml needleless syringe (Becton, Dickinson, and Company). Plants were then kept in a growth room at 24/22 °C with a 16/8 hours light/dark photoperiod for 48-72 hrs.

Epifluorescence and confocal microscopy: *N. benthamiana* leaf samples 70 to 90 hours post-infiltration (approximately 0.25 cm2 from the infiltrated area) were mounted in water and viewed directly with a Zeiss LSM 880 confocal scanning microscope using an oil immersion objective 40× Plan-Apochromat 1.4NA (numerical aperture of 1.4). Fluorescence was excited using 488 nm light for GFP. GFP emission fluorescence was selectively detected at 490-540 nm using the Zen 2.3 SP1 software. Each experiment was repeated three times.

2.5.12 Synthesis of Ribulose-N- α -Ac-lysine

A solution of *N*-α-acetyl-L-lysine (0.5 g, 2.66 mmol) and D-ribose (1.6 g, 10.64 mmol) in methanol (100 mL) was stirred at 50°C for 4 h under argon gas. The solvent was evaporated under vacuum and the residue obtained was purified over silica gel column by flash chromatography (R_f = 0.14, EtOAc/MeOH/H_2O, 5/3/2, v/v/v). The product obtained was further purified over a column of Dowex® 50 x 8 H^+ resin (0.6 x 5 cm) cation exchange resin using 150 mM NaCl as an eluent and desalted over a P-2 column using H_2O as an eluent. The desired fractions were combined, concentrated and lyophilized to afford the product as white solid (0.55 g, 64.7%); LRMS (ESI):

calculated for C13H25N2O7 $[M+H]^+$ 321.16, found 321.09. Synthesized Ribulose-*N*-α-Ac-lysine was characterized using NMR and ESI.

Bibliography

1. Maillard, L.C., *The action of amino acids on sugar; the formation of melanoidin by a methodic route.* Comptes Ren- dus Hebdomadaires Des Seances De L Academie Des Sciences 1912. **154**: p. 66-68.
2. Teodorowicz, M., J. van Neerven, and H. Savelkoul, *Food Processing: The Influence of the Maillard Reaction on Immunogenicity and Allergenicity of Food Proteins.* Nutrients, 2017. **9**(8).
3. Hodge, J.E., *The Amadori rearrangement.* Adv Carbohydr Chem, 1955. **10**: p. 169-205.
4. Higgins, H.F.B.a.P.J., *Reaction of Monosaccharides with Proteins: Possible Evolutionary Significance.* Science, 1981. **213**: p. 222-224.
5. Baynes, J.W., et al., *The Amadori product on protein: structure and reactions.* Prog Clin Biol Res, 1989. **304**: p. 43-67.
6. Emile Van Schaftingen, F.C., Elsa Wiame, Maria Veiga-da-Cunha, *Enzymatic repair of Amadori products.* Amino Acids, 2012. **42**: p. 1143-1150.
7. Veronika Maria Deppe, J.B., Timithy O'Connell, Karl-Heinz Maurer, Friedhelm Meinhardt, *Enzymatic deglycation of Amadori products in bacteria:mechanisms, occurrence and physiology functions.* Applied Microbiology Biotechnology, 2011. **90**: p. 399-406.
8. Böhm, P.G.a.M., *Advanced glycation end products. Key players in skin aging?* Dermato-Endocrinology, 2012. **4**(3): p. 259-270.
9. François Collard, G.D., Vincent Stroobant, Gert Matthijs, Emil Van Shaftingen, *A Mammalian Protein Homologous to Fructosamine-3-Kinase Acting on Psicosamines and Ribulosamines but not on Fructosamines.* Diabetes, 2003. **52**: p. 2888-2895.

10. Ghislain Delpierre, M.H.R., François Collard, Vincent Stroobant, Florent Vanstapel, Helena Santos, and Emile Van Schaftingen, *Identification, Cloning, and Heterologous Expression of a Mammalian Fructosamine-3-Kinase.* Diabetes, 2000. **49**: p. 1627-1634.

11. Benjamin S. Szwerglod, S.H., and Paul J. Beisswenger, *Human Fructosamine-3-Kinase. Purification, Sequencing, Substrate Specificity, and Evidence of Activity In Vivo.* Diabetes, 2001. **50**: p. 2139-2147.

12. Juliette Fortpied, R.G., Vincent Stroobant and Emile van Schaftingen, *Plant ribulosamine/erythrulosamine 3-kinase, a putative protein-repair enzyme.* Biochemical Journal, 2005. **388**: p. 795-802.

13. James R. Conner, P.J.B., Benjamin S. Szwergold, *The expression of the genes for fructosamine-3-kinase and fructosamine-3-kinase-related protein appears to be constitutive and unaffected by environmental signal.* Biochemical and Biophysical Research Communications, 2004. **323**: p. 932-936.

14. Jérôme Delplanque, G.D., Fred R, Opperdoes, and Emile Van Schaftingen, *Tissue Distribution and Evolution of Fructosamine 3-Kinase and Fructosamine 3-Kinase-related Protein.* The Journal of Biological Chemistry, 2004. **279**(45): p. 46606-46613.

15. Rita Gemayel, J.F., Rim Rzem, Didier Vertommen, Maria Veiga-da-Cunha and Emile Van Schaftingen, *Many fructosamine 3kinase homologues in bacteria are ribulosamine/erythrulosamine 3-kinase potentially involved in protein deglycation.* The FEBS Journal, 2007. **274**: p. 4360-4374.

16. T. R. Brown, B.S., K.A. Brown, M.A. Schwartz, A.M. Tobia and F. Kappler, *Modulation of in vivo 3-deoxyglucosone levels.* Biochemical Society Transactions, 2003. **31**(6): p. 1433-1437.

17. S. M. A. Pascal, M.V.-d.-C., P. Gilon, E. Van Schaftingen, and J. C. Jonas, *Effects of fructosamine-3-kinase deficiency on function and survival of mouse pancreatic islets prolonged culture in high glucose or ribose concentrations.* American Journal of Physiology - Endocrinology and Metabolism, 2010. **298**: p. E586-E596.

18. Kusunoki, H., et al., *Relation Between Serum 3-Deoxyglucosone and Development of Diabetic Microangiopathy.* Diabetes Care, 2003. **26**(6): p. 1889.

19. Niwa, T., et al., *Presence of 3-Deoxyglucosone, a Potent Protein Crosslinking Intermediate of Maillard Reaction, in Diabetic Serum.* Biochemical and Biophysical Research Communications, 1993. **196**(2): p. 837-843.

20. Niwa, T. and S. Tsukushi, *3-deoxyglucosone and AGEs in uremic complications: inactivation of glutathione peroxidase by 3-deoxyglucosone.* Kidney Int Suppl, 2001. **78**(0098-6577 (Print)): p. S37-41.

21. Fong, D.H. and A.M. Berghuis, *Substrate promiscuity of an aminoglycoside antibiotic resistance enzyme via target mimicry.* The EMBO Journal, 2002. **21**(10): p. 2323.

22. Zheng, J., et al., *2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor.* Acta Crystallographica Section D, 1993. **49**(3): p. 362-365.

23. Kannan, N., et al., *Structural and functional diversity of the microbial kinome.* PLoS Biol, 2007. **5**(3): p. e17.

24. Oruganty, K., et al., *Identification and classification of small molecule kinases: insights into substrate recognition and specificity.* BMC Evol Biol, 2016. **16**: p. 7.

25. Tagliabracci, V.S., et al., *A Single Kinase Generates the Majority of the Secreted Phosphoproteome.* Cell, 2015. **161**(7): p. 1619-1632.

26. Sreelatha, A., et al., *Protein AMPylation by an Evolutionarily Conserved Pseudokinase.* Cell, 2018. **175**(3): p. 809-821.e19.

27. Kwon, A., et al., *Tracing the origin and evolution of pseudokinases across the tree of life.* Science Signaling, 2019. **12**(578): p. eaav3810.

28. Kannan, N. and A.F. Neuwald, *Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component?* J Mol Biol, 2005. **351**(5): p. 956-72.

29. Nolen, B., S. Taylor, and G. Ghosh, *Regulation of Protein Kinases: Controlling Activity through Activation Segment Conformation.* Molecular Cell, 2004. **15**(5): p. 661-675.

30. Bryne, D.P., et al., *Aurora A regulation by reversible cysteine oxidation reveals evolutionary-conserved redox control of Ser/Thr protein kinase activity.* Science Signaling, 2020.

31. Bastidas, A.C., et al., *Phosphoryl transfer by protein kinase A is captured in a crystal lattice.* (1520-5126 (Electronic)).

32. Holm, L. and P. Rosenström, *Dali server: conservation mapping in 3D.* Nucleic acids research, 2010. **38**(Web Server issue): p. W545-W549.

33. Fong, D.H., et al., *Structural Basis for Kinase-Mediated Macrolide Antibiotic Resistance.* Structure, 2017. **25**(5): p. 750-761 e5.

34. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.

35. Schmidt, B., L. Ho, and P.J. Hogg, *Allosteric Disulfide Bonds.* Biochemistry, 2006. **45**(24): p. 7429-7433.

36. Haworth, N., L. L Feng, and M. Wouters, *High torsional energy disulfides: Relationship between cross-strand disulfides and right-handed staples*. Vol. 4. 2006. 155-68.

37. Wouters, M.A., S.W. Fan, and N.L. Haworth, *Disulfides as Redox Switches: From Molecular Mechanisms to Functional Significance.* Antioxidants & Redox Signaling, 2009. **12**(1): p. 53-91.

38. Srinivasan, N., et al., *Conformations of disulfide bridges in proteins.* International Journal of Peptide and Protein Research, 1990. **36**(2): p. 147-155.

39. Hanson, G.T., et al., *Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators.* J Biol Chem, 2004. **279**(13): p. 13044-53.

40. Gregory, J.D., *The Stability of N-Ethylmaleimide and its Reaction with Sulfhydryl Groups.* Journal of the American Chemical Society, 1955. **77**(14): p. 3922-3923.

41. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Research, 2004. **32**(5): p. 1792-1797.

42. Kosower, N.S. and E.M. Kosower, *[11] Diamide: An oxidant probe for thiols*, in *Biothiols Part A Monothiols and Dithiols, Protein Thiols, and Thiyl Radicals*. 1995, Academic Press. p. 123-133.

43. Kosower, N.S. and E.M. Kosower, *Formation of disulfides with diamide*, in *Methods in Enzymology*. 1987, Academic Press. p. 264-270.

44. Ferrer-Sueta, G., et al., *Factors affecting protein thiol reactivity and specificity in peroxide reduction.* Chem Res Toxicol, 2011. **24**(4): p. 434-50.

45. Gout, I., *Coenzyme A, protein CoAlation and redox regulation in mammalian cells.* Biochemical Society transactions, 2018. **46**(3): p. 721-728.

46. Amelio, I., et al., *Serine and glycine metabolism in cancer.* Trends Biochem Sci, 2014. **39**(4): p. 191-8.

47. Garcia-Bermudez, J., et al., *Aspartate is a limiting metabolite for cancer cell proliferation under hypoxia and in tumours.* (1476-4679 (Electronic)).

48. Fiske, C.H. and Y. Subbarow, *The Nature of the "Inorganic Phosphate" in Voluntary Muscle.* Science, 1927. **65**(1686): p. 401-3.

49. Akter, S., et al., *Chemical proteomics reveals new targets of cysteine sulfinic acid reductase.* Nature Chemical Biology, 2018. **14**(11): p. 995-1004.

50. Bilova, T., et al., *A Snapshot of the Plant Glycated Proteome: STRUCTURAL, FUNCTIONAL, AND MECHANISTIC ASPECTS.* J Biol Chem, 2016. **291**(14): p. 7621-36.

51. Greifenhagen, U., et al., *Sensitive and site-specific identification of carboxymethylated and carboxyethylated peptides in tryptic digests of proteins and human plasma.* J Proteome Res, 2015. **14**(2): p. 768-77.

52. Schmidt, R., et al., *Specific tandem mass spectrometric detection of AGE-modified arginine residues in peptides.* J Mass Spectrom, 2015. **50**(3): p. 613-24.

53. Zhang, Q., et al., *Proteomic profiling of nonenzymatically glycated proteins in human plasma and erythrocyte membranes.* J Proteome Res, 2008. **7**(5): p. 2025-32.

54. Thul, P.J., et al., *A subcellular map of the human proteome.* Science, 2017. **356**(6340): p. eaal3321.

55. Handy, D.E. and J. Loscalzo, *Redox regulation of mitochondrial function.* Antioxidants & redox signaling, 2012. **16**(11): p. 1323-1367.

56. Mailloux, R.J., X. Jin, and W.G. Willmore, *Redox regulation of mitochondrial function with emphasis on cysteine oxidation reactions.* Redox Biology, 2014. **2**: p. 123-139.

57. Go, Y.-M. and D.P. Jones, *Redox compartmentalization in eukaryotic cells.* Biochimica et biophysica acta, 2008. **1780**(11): p. 1273-1290.

58. Kemble, D.J. and G. Sun, *Direct and specific inactivation of protein tyrosine kinases in the Src and FGFR families by reversible cysteine oxidation.* Proceedings of the National Academy of Sciences, 2009. **106**(13): p. 5070.

59. Heppner, D.E., et al., *Direct cysteine sulfenylation drives activation of the Src kinase.* Nature Communications, 2018. **9**(1): p. 4522.

60. Gasteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server*, in *The Proteomics Protocols Handbook*, J.M. Walker, Editor. 2005, Humana Press: Totowa, NJ. p. 571-607.

61. Kabsch, W., *XDS.* Acta Crystallographica Section D, 2010. **66**(2): p. 125-132.

62. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallographica Section D, 2010. **66**(2): p. 213-221.

63. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

64. Urzhumtsev, A., P.V. Afonine, and P.D. Adams, *TLS from fundamentals to practice.* Crystallography reviews, 2013. **19**(4): p. 230-270.

65. Pijning Aster, E., et al., *Identification of allosteric disulfides from labile bonds in X-ray structures.* Royal Society Open Science, 2018. **5**(2): p. 171058.

66. UniProt Consortium, T., *UniProt: the universal protein knowledgebase.* Nucleic Acids Research, 2018. **46**(5): p. 2699-2699.

67. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Research, 2016. **44**(D1): p. D733-D745.

68. McSkimming, D.I., et al., *KinView: a visual comparative sequence analysis tool for integrated kinome research.* Mol Biosyst, 2016. **12**(12): p. 3651-3665.

69. Finn, R.D., et al., *The Pfam protein families database.* Nucleic Acids Res, 2010. **38**(Database issue): p. D211-22.

70. Laue, T.M., *Computer-aided interpretation of analytical sedimentation data for proteins.* Analytical Ultracentrifugation in Biochemistry and Polymer Science, 1992: p. 90-125.

71. Schuck, P., *On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation.* Anal Biochem, 2003. **320**(1): p. 104-24.

72. Ortega, A., D. Amoros, and J. Garcia de la Torre, *Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models.* Biophys J, 2011. **101**(4): p. 892-8.

73. Brautigam, C.A., *Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data.* Methods Enzymol, 2015. **562**(1557-7988 (Electronic)): p. 109-33.

74. Bingol, K., et al., *Comprehensive Metabolite Identification Strategy Using Multiple Two-Dimensional NMR Spectra of a Complex Mixture Implemented in the COLMARm Web Server.* Analytical Chemistry, 2016. **88**(24): p. 12411-12418.

CHAPTER 3

ILLUMINATING THE FUNCTIONS OF UNDERSTUDIED FRUCTOSAMINE-3- KINASE (FN3K) USING A MULTI-OMICS APPROACH REVEALS NEW LINKS TO NICOTINAMIDE ADENINE DINUCLEOTIDES (NAD)

Safal Shrestha, Rahil Taujale, Samiksha Katiyar, and Natarajan Kannan.
Submitted to Nature Scientific Report, 2023

3.1 Abstract

Fructosamine-3-kinases (FN3Ks) are a conserved family of repair enzymes that phosphorylate reactive sugars attached to lysine residues in peptides and proteins. Although FN3Ks are present across the tree of life and share detectable sequence similarity to eukaryotic protein kinases, the biological processes regulated by these kinases are largely unknown. To address this knowledge gap, we leveraged the FN3K CRISPR Knock-Out (KO) cell line alongside an integrative multi-omics study combining transcriptomics, interactomics, metabolomics, and genomics to place these enzymes in a pathway context. The integrative analyses revealed the enrichment of pathways related to oxidative stress response, lipid biosynthesis (cholesterol and fatty acids), carbon and co-factor metabolism, and protein translational machinery. Moreover, statistically significant enrichment of metallothioneins and nicotinamide adenine dinucleotide (NAD) binding proteins suggest potential links between FN3Ks and NAD-mediated energy metabolism and redox balance. We report specific binding of human FN3K to NAD compounds in a metal and concentration-dependent manner and provide insight into their binding mode using modeling and experimental site-directed mutagenesis. By identifying a potential link between FN3Ks, redox regulation, and NAD-dependent metabolic processes, our studies provide a framework for targeting these understudied kinases in diabetic complications and other metabolic disorders.

3.2 Introduction

Glycation is a universal post-translational modification in which reducing sugars such as glucose and fructose are non-enzymatically added to free amine groups on peptides, proteins, and lipids [1]. This non-enzymatic modification occurs endogenously in all living organisms and exogenously in the foods we eat [2]. Sugars attached to amine groups can undergo Amadori rearrangements to form stable linkages with biomolecules [3, 4]. Because such linkages can adversely affect biomolecular functions, organisms have evolved detoxification/deglycation mechanisms to repair the potential toxic effects of sugars, an essential nutrient for life. Fructosamine-3 kinases (FN3Ks) are a family of deglycation enzymes that remove ribulose, psicose, and glucose sugars attached to surface exposed lysine residues (ketosamines) in proteins [5, 6]. They do so by catalyzing the transfer of the gamma phosphate from ATP to the 3' hydroxyl group in the ketosamine substrate. Phosphorylation of ketosamines by FN3Ks results in an unstable intermediate, ketosamine 3-phosphate, which spontaneously decomposes into inorganic phosphate and dicarbonyl sugar, 3-deoxyglucosone (3-DG), thereby regenerating the unmodified lysine [6]. The breakdown of ketosamine can result in the creation of Advanced Glycation End (AGEs) products [7], which are associated with oxidative stress due to the heightened production of reactive oxygen species (ROS) [8].

Deglycation by FN3Ks is an ancient mechanism for protein repair as FN3K homologs are present in diverse organisms, from archaea and bacteria to humans [9, 10] (Fig. 3.1A). While lower eukaryotes and prokaryotes contain a single copy of the FN3K gene, most tetrapod genomes encode two copies: FN3K and FN3K-related protein (FN3KRP). Interestingly, two independent gene duplication events have led to the tetrapods having two copies: one in reptiles/birds and the other in placental mammals [11]. Although the functions of FN3K homologs in lower eukaryotes

and bacteria are yet to be established, it has been proposed that they repair proteins modified by ribose-5-phosphate, a potent glycating agent formed from the conserved pentose phosphate pathway [9, 10]. Despite the remarkable conservation of FN3Ks across diverse organisms and their emerging role in protein repair, the biological and cellular processes linked to this important class of proteins remain unknown.

FN3K activity is highly substrate-specific with human FN3K (HsFN3K) phosphorylating ketosamines resulting from glycation of both L and D orientation sugars, whereas FN3KRP orthologs are limited to only D-orientation sugars [6, 9, 10, 12]. At the transcriptional level, in human tissues, HsFN3K is highly expressed in the brain but has low expressions in the thymus and tonsils (Fig. 3.1B). On the other hand, FN3KRP expression levels are somewhat uniform throughout the different tissues (Fig. 3.1B). Moreover, subcellular localization using immunohistochemistry (IHC) studies found FN3K localized to the mitochondria in HepG2 cells whereas FN3KRP is localized to the nucleoplasm in several different cell lines [13].

Previously, we reported the crystal structure of FN3K ortholog from *Arabidopsis thaliana* (AtFN3K), revealing an evolutionarily conserved redox regulation mechanism involving conserved cysteines in the ATP binding P-loop [14]. Specifically, we demonstrated that AtFN3K adopts a disulfide-linked dimer that can be reversed under reducing conditions and that the equivalent cysteine in human FN3K ($C24^{HsFN3K}$) performs an analogous role conferring redox sensitivity and disulfide-mediated oligomerization. Furthermore, a recent study has proposed a potential correlation between the deglycation function of HsFN3K and the development of hepatocellular carcinoma (liver cancer) through the involvement of the nuclear transcription factor NRF2 [15] which controls the expression of multiple antioxidant enzymes [16, 17]. Acting as an adaptor protein, Kelch-like ECH-associated protein (Keap1) exerts negative control over NRF2 by

facilitating its ubiquitination through Cullin3 (Cul3), thus marking it for proteasomal degradation [18]. Keap1 encompasses redox-sensitive cysteines that undergo modifications during oxidative stress, impeding the interaction between Keap1 and NRF2. The progressive buildup of NRF2 triggers the induction of antioxidant enzymes, further underscoring the role of FN3K in maintaining cellular redox balance and altered expression in liver cancer (Fig. 3.1C). It is worth noting that higher FN3K expression levels are also observed in ocular and adrenal cancers (Fig. 3.1C). However, despite the altered expression of HsFN3K in multiple cancer types and their established roles in metabolic diseases such as diabetes and associated complications [19, 20], our understanding of FN3K's functions within human cells and across different species, particularly in the context of redox regulation, remains incomplete.

Here we leverage the CRISPR Knock Out (KO) of HsFN3K in the HepG2 liver cancer cell line (Fig. 3.1D) and employ a multi-omics approach of transcriptomics, interactomics, metabolomics, and genomics to place these enzymes in the oxidative stress response pathway involving metallothioneins, lipid biosynthesis (fatty acid and cholesterol), gluconeogenesis, and protein translational machinery pathways. Moreover, enrichment analysis using the transcriptomics and the metabolomics datasets revealed enrichment of glutathione and co-factor metabolisms, including nicotinic acid metabolism. Based on the enrichment of Nicotinamide dinucleotide (NAD) binding superfamily in the transcriptomics and interactomics network as well as the use of the NAD compounds by the proteins in the same operon as FN3K/RP orthologs, we validated HsFN3K binding to NAD compounds using Differential Scanning Fluorimetry (DSF). We further probe the binding specificity through mutagenesis and identify key NADH binding residues in the adenosine binding pocket. In sum, our study uncovers a novel connection between FN3K and redox regulation through the NAD molecules.

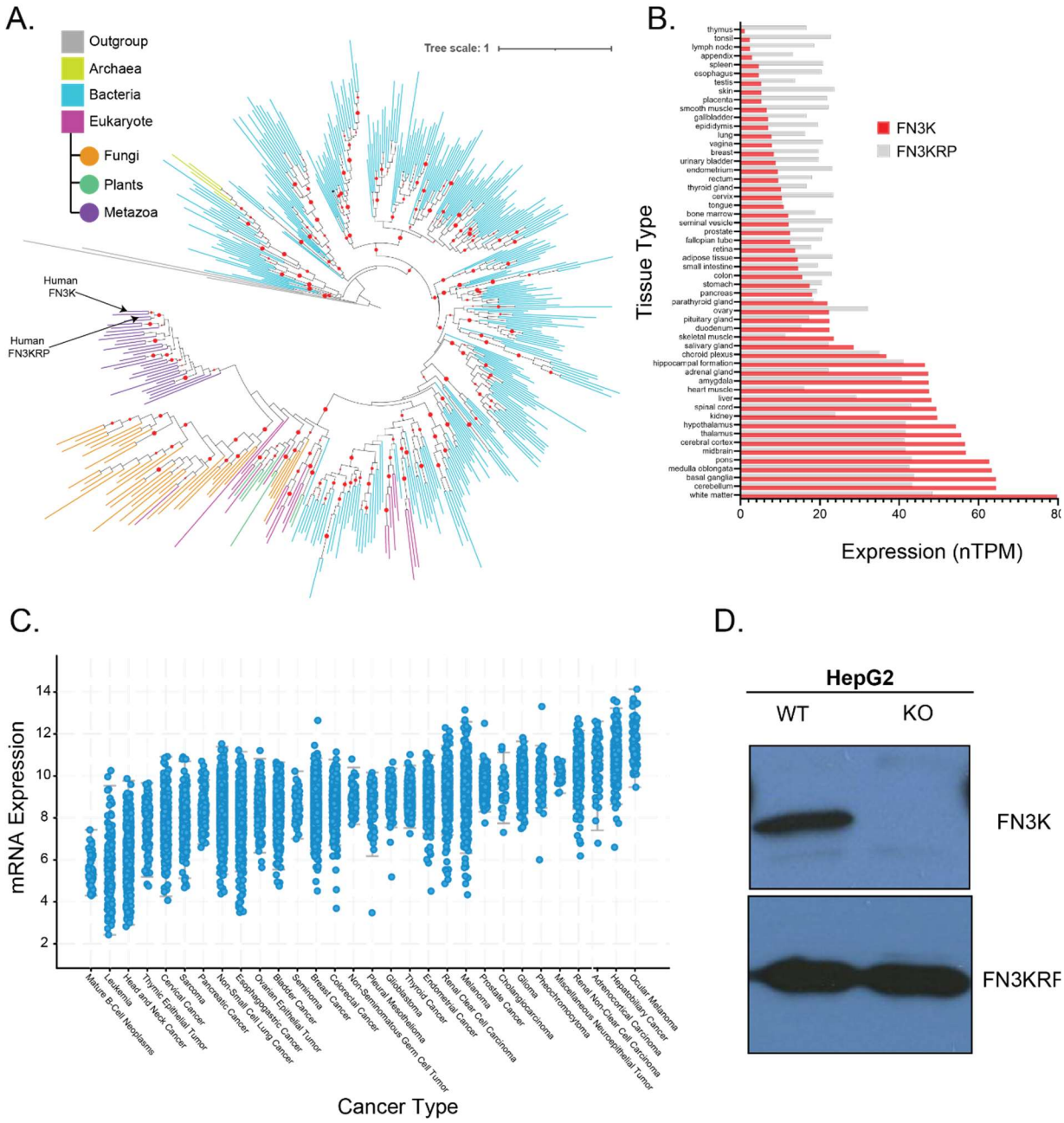


Figure 3.1: Study of FN3K conservation and expression.

(A) An evolutionary tree depicting the phylogenetic relationship among FN3K orthologs and paralogs. Maximum likelihood tree was constructed using FastTree 2.1.11[21]. Clades with bootstrap values greater than 0.8 are labeled with red dots. **(B)** Tissue-specific expression levels of human FN3K and human FN3KRP, based on data from the Protein Atlas Database. **(C)** The levels of human FN3K RNA expression across various types of cancer, arranged by median expression values. The figure was produced using cBioPortal. Unit for mRNA expression levels: RSEM (Batch normalized from Illumina HiSeq_RNASeqV2) (log2(value + 1)) **(D)** A Western blot analysis displaying the total quantities of human FN3K and FN3KRP in both wild-type and FN3K knockout (KO) HepG2 cells. The blots were from separate gels.

3.3 Results

3.3.1 Enrichment of Metallothioneins, Lipid Biosynthesis, and Signaling Pathways revealed through integrative transcriptomics and interactomics analyses.

To investigate the molecular pathways regulated by HsFN3K, we compared the transcriptomes (RNA sequencing) and interactomes (Immunoprecipitation, IP) of WT and FN3K Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Knock-Out (KO) HepG2 cells[14]. HsFN3K knockout did not affect protein levels of the paralog FN3KRP (Fig. 3.1D). We identified a total of 408 differentially expressed genes (DEGs) (greater than 2 log-fold change and p-value < 0.05), out of which 305 genes were upregulated in the KO (Fig. 3.2A, B). Moreover, the DEGs included a total of 23 RNA or pseudogenes. From the interactomics study, we identified 209 potential interacting partners, 18 of which were present in two or more replicates with SLC25A13, an aspartate/glutamate carrier protein, present in all three.

We explored the relationship between the two complementary datasets by combining the transcriptomics with the interactomes dataset using the String App [22] in the Cytoscape framework (Fig. 3.2A). Next, we performed pathway enrichment analyses on the entire network. The significantly enriched pathways were then ordered based on fold enrichment. This allows us to quantify pathway over-representation by comparing the total genes identified for a given pathway to the background frequency of total genes annotated for that same pathway [23].

Among the most enriched pathways, "*metallothioneins bind metals*" ranked the highest based on fold enrichment (Fig. 3.2C). Metallothioneins are low molecular weight (6-7 kDA) cysteine-rich proteins involved in metal homeostasis and oxidative stress response to shield against

DNA damage and cytotoxicity[24]. Notably, 7 of the 11 metallothioneins in the pathway were significantly upregulated in the KO cell line (Fig. 3.2D).

In addition to Metallothioneins, we identified pathways related to lipid biosynthesis (cholesterol and fatty acid) and signaling enriched in the FN3K KO cells relative to WT. Some of the enriched pathways related to cholesterol synthesis were "*Mevalonate arm of the cholesterol biosynthesis pathway", "Steroid biosynthesis", "Cholesterol metabolism with Bloch and Kandutsch-Russell pathways", "Statin inhibition of cholesterol production"* and "*Sterol regulatory element-binding protein (SREBP) signaling"* (Fig. 3.2C). Notably, critical rate-limiting enzymes in cholesterol biosynthesis such as Mevalonate diphosphate decarboxylase *(*MVD), 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (HMGCR), and Squalene monooxygenase (SQLE) are all up-regulated in the KO cells (Fig. 3.2D). Moreover, 7-Dehydrocholesterol reductase (DHCR7) that catalyzes cholesterol production from 7-Dehydrocholesterol by utilizing reduced nicotinamide adenine dinucleotide phosphate (NADPH) [25] was an up-regulated DEG and an interacting partner of HsFN3K in our IP experiment (Fig. 3.2D). Similarly, the Fatty acid synthase (FASN) gene that synthesizes long-chain fatty acids from acetyl coenzyme A (acetyl-CoA), malonyl-CoA, and NADPH[26] is upregulated DEG and immunoprecipitated with HsFN3K. Other DEGs in the lipid synthesis pathways are indicated in Fig. 3.2D.

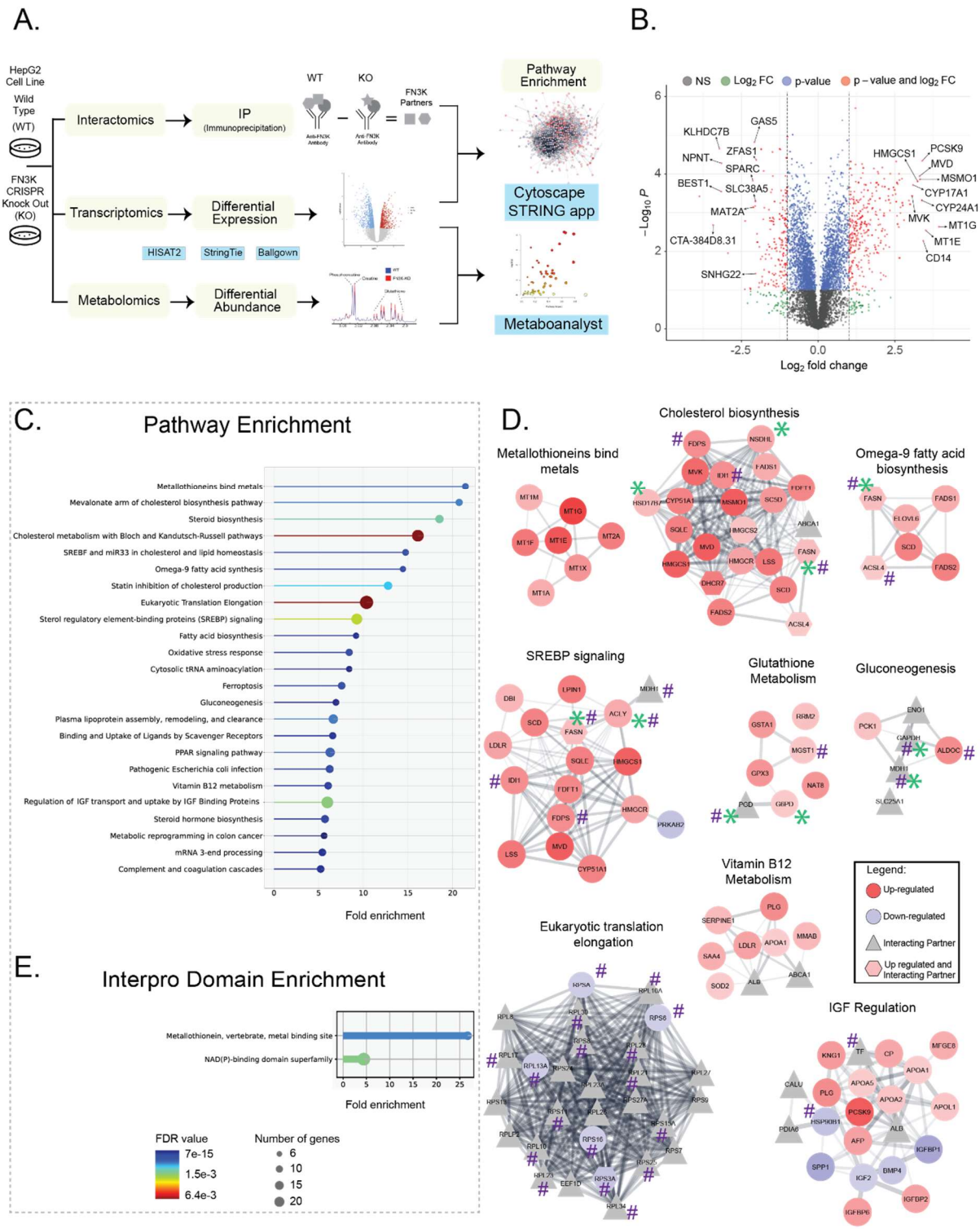


Figure 3.2: Pathway and domain enrichment of Cytoscape network combining transcriptomics and interactomics datasets.

(A) Schematic diagram describing the method to perform integrative analyses of Differentially expressed genes (DEGs), FN3K interaction partners, and Differentially abundant metabolites (DAMs). **(B)** A volcano plot exhibiting significantly upregulated and downregulated genes in the FN3K knockout HepG2 cell line compared to the wild type. **(C)** A lollipop graph showcasing enriched biological pathways ordered according to fold enrichment. **(D)** Pathways were enriched from the integrated network

of transcriptomic and interactomic data sets. Differentially Expressed Genes (DEGs) identified from comparative transcriptomics are represented by circles, with up-regulated and down-regulated genes colored in red and blue, respectively. Interaction partners discovered through the Immunoprecipitation (IP) experiments are represented by triangles and colored gray. Proteins identified in both transcriptomic and interactomic data sets are represented as hexagons. Members of the NAD(P) binding protein superfamily are marked with a green asterisk (*). Proteins with sulfenylated cysteines are marked with a purple hash (#). The edges between the nodes are based on StringApp, with a medium confidence cut-off value of 0.4. **(E)** A lollipop graph demonstrating enriched protein domains ranked by fold enrichment.

3.3.2 Oxidative stress response and glutathione metabolism genes are upregulated and protein translational machinery is downregulated in FN3K KO cells.

Besides the lipid biosynthesis and signaling pathways, "*Oxidative stress response through glutathione metabolism"* was also enriched (Fig. 3.2C, D). Glutathione transferases such as Glutathione S-transferase A1 (GSTA1) and Microsomal glutathione S-transferase 1 (MGST1), as well as Glutathione peroxidase 3 (GPX3), a glutathione peroxidase, were all upregulated in the KO (Fig. 3.2D). GPX3 reduces hydrogen and organic peroxides using glutathione and is involved in oxidative stress response [27]. Moreover, the upregulated glucose-6-phosphate dehydrogenase (G6PD) and the interacting partner Phosphogluconate dehydrogenase (PGD) are involved in carbohydrate metabolism through the Pentose phosphate pathway (PPP). Importantly, G6PD is a major source of cellular NADPH critical in maintaining reduced glutathione levels [28]. Similarly, the pathway within carbohydrate metabolism, "*gluconeogenesis"* which is essentially a reverse glycolytic pathway, was also enriched. It included both upregulated DEGs as well as interacting partners (Fig. 3.2D). The upregulated Phosphoenolpyruvate Carboxykinase 1 (PCK1) is the major indication of gluconeogenesis [29].

While multiple enriched pathways consisted solely of upregulated DEGs, the enriched "*Eukaryotic translation initiation"* pathway comprised only downregulated DEGs. It includes several small ribosomal proteins (RPS) identified as interacting partners, such as the eukaryotic translation elongation factor 1 delta (EEF1D) (Fig. 3.2D).

Additional enrichment pathways were associated with "*Insulin-Like Growth Factor (IGF) regulation"* and "*vitamin B12 metabolism"*. Of interest, the most upregulated gene in the FN3K KO, Proprotein convertase subtilisin/kexin type 9 (PCSK9) [30], is central to IGF regulation and crucial in managing plasma cholesterol homeostasis. Enrichment was also observed in Vitamin B12 metabolism, with the upregulated Metabolism of cobalamin associated B (MMAB) gene converting vitamin B12 into adenosylcobalamin (AdoCbl) [31] (Fig. 3.2D).

Given the enrichment of the oxidative stress response pathway and the redox regulation of FN3K through the P-loop cysteine, we explored whether any proteins in the network exhibited post-translationally modified (PTM) cysteine residues. To detect such modifications, we used the iCysMod [32] database, which contains curated, experimentally identified PTMs on cysteine residues. We searched for proteins with S-sulfenylation (-SOH) modifications within our network. Interestingly, several proteins—including FASN, malate dehydrogenase 1 (MDH1), G6PD, and a host of large and small Ribosomal proteins—exhibited sulfenylated cysteines (indicated by #, Fig. 3.2D). Many of these proteins contained more than one sulfenylated cysteine.

3.3.3 Nicotinamide dinucleotide Phosphate (NADP) binding domains are enriched in FN3K KO cells.

Next, using the same integrated network, we identified enriched protein domains. Only two different protein domains were significantly enriched. These include the metallothioneins and nicotinamide dinucleotide (NAD) binding domain superfamily (Fig. 3.2E). Nicotinamide dinucleotide (NAD) or its phosphorylated form (NADP) are co-factors that exist as redox couples (NAD+/NADH and NADP+/NADPH) in cells to serve distinct functions [33]. While NAD+/NADH is involved in cellular energy metabolism, NADP+/NADPH is involved in redox

balance and fatty acid and nucleic acid synthesis [33]. By mapping the enriched NADP binding domain superfamily proteins (indicated by *) on the enriched pathways (Fig. 3.2D), we find these are involved in several enriched pathways, including cholesterol synthesis, fatty acid synthesis, gluconeogenesis, and glutathione metabolism. Notably, enrichment analysis on subnetworks containing only the first and second neighbors of FN3K, recapitulate the enriched pathways observed in Fig. 3.2. Importantly, domain enrichment on this subnetwork also showed enrichment of NAD(P) binding domain superfamily (FDR value: 1.9 E-9) (Fig. 3.3A). Using Uniprot's reaction annotations, we probed the network further and identified several additional proteins that used NAD(P) compounds in their enzymatic reaction (Fig. 3.3B). These included several key enzymes in lipid metabolism, including SQLE and HMGCR. In addition to existing redox couples in cells, NAD can also be used to transfer ADP-ribose moiety. Indeed, we find ADP-ribosyl transferases such as Poly (ADP-ribose) polymerase 1 (PARP1) and ADP-Ribosyltransferase 4 (ART4) that use NAD+ to transfer ADP-ribose moiety to arginine residues in proteins as a reversible post-translational modification. Moreover, we find Sirtuin 3 (SIRT3) in our network, a NAD-dependent mitochondrial deacetylase [34]. Interestingly, decapping DXO, de-capping exoribonuclease protein that specifically de-caps NAD-capped RNAs [35], was also present in our network.

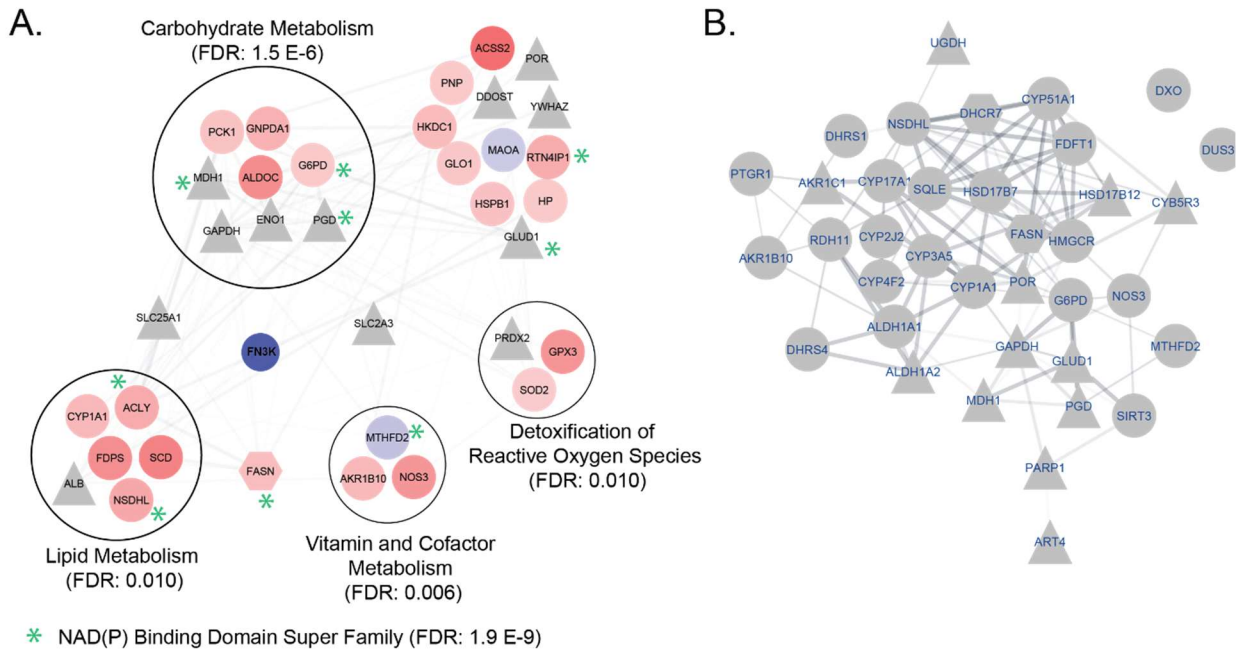


Figure 3.3: Enrichment of NAD(P) binding proteins in the integrated (transcriptomics and metabolomics) network.

(A) First and second order neighbors of HsFN3K in the integrated (transcriptomics and metabolomics) network. Differentially Expressed Genes (DEGs) identified from comparative transcriptomics are represented by circles, with up-regulated and down-regulated genes colored in red and blue, respectively. Interaction partners discovered through the Immunoprecipitation (IP) experiments are represented by triangles and colored gray. Proteins identified in both transcriptomic and interactomics data sets are represented as hexagons. Members of the NAD(P) binding protein superfamily are marked with a green asterisk (*). FDR stands for False Discovery Rate. **(B)** Proteins in the integrative network in Fig. 2 that has NAD related compounds (NADH, NAD+, NADPH, NADP+) in its reaction annotation on Uniprot.

3.3.4 Joint pathway analysis using transcriptomics and metabolomics datasets reveal enrichment of glutathione, carbon, and co-factor metabolisms including nicotinate/nicotinamide metabolism.

We had previously performed a comparative metabolomics study between the KO and WT HepG2 cells [14] and had identified several significantly abundant metabolites. In light of the new transcriptomics datasets, we performed joint pathway analysis using the MetaboAnalyst [36] and identified several enriched pathways (Fig. 3.4A). Matched features for enriched pathways such as "*terpenoid backbone biosynthesis*", "*retinol metabolism*", and "*steroid biosynthesis*" were exclusive to DEGs. In contrast, "*aminoacyl-tRNA biosynthesis*" and "*Valine, leucine, and isoleucine biosynthesis*" were exclusive to differentially abundant metabolites (DAMs).

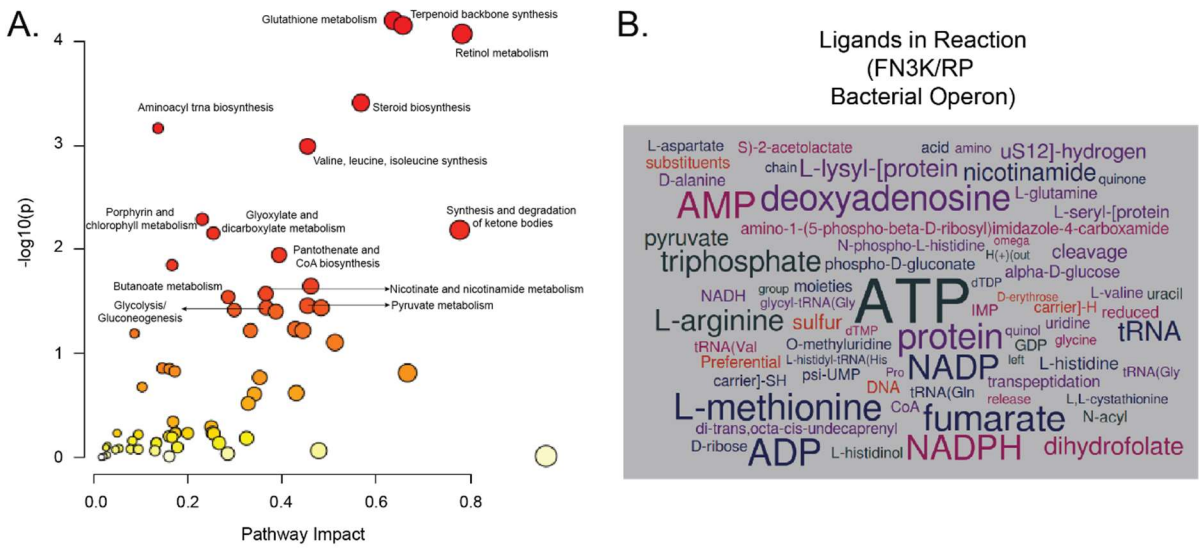


Figure 3.4: Enrichment of co-factors metabolism.

(A) Joint pathway enrichment of transcriptomics and metabolomics datasets using Metaboanalyst [36]. A multitude of pathways associated with glutathione, carbon, and co-factor metabolisms were found to be enriched. **(B)** Word plot showing ligands/metabolites used by proteins in the same operon as FN3K/RP in bacteria. The operons were identified using ProOpDB [37]. Uniprot reaction annotation was used to identified ligands used by the proteins in the operon.

Several of the enriched pathways matched both DAMs and DEGs. The highly enriched "*glutathione metabolism"* pathway included metabolites such as glutathione, glycine, and glutamate and DEGs such as GPX3, G6PD, N-acetyltransferase 8 (NAT8), GSTA1, MGST1, and Ribonucleotide reductase regulatory subunit M2 (RRM2). Other enriched pathways were related to carbon and co-factor metabolisms. Pathways within carbon metabolism included "*Glyoxylate and dicarboxylate metabolism*" (citrate, glycine, glutamate, formate, glutamine, Acetyl-CoA acetyltransferase 2 (ACAT2), Acyl-CoA synthetase short chain family member 2 (ACSS2)), *"Pyruvate metabolism"* (lactate, Phosphoenolpyruvate carboxykinase 1 (PCK1), Glyoxalase I (GLO1), ACSS2, ACAT2), *"Butanoate metabolism"* (glutamate, 3-hydroxy-3-methylglutaryl-CoA synthase 1 and 2 (HMGCS1/2), ACAT2), and *"Glycolysis/Gluconeogenesis"* (citrate, glucose, Aldolase C, fructose-bisphosphate (ALDOC), Hexokinase domain containing 1 (HKDC1), PCK1, ACSS2). Metabolism of co-factors included pathways such as *"Porphyrin and*

chlorophyll metabolism" (glycine, glutamate, MMAB, Ceruloplasmin (CP), and UDP glucuronosyltransferases (UGTs) 2B11,2B10, and 2A3), *"Pantothenate and CoA biosynthesis"* (pantothenate, valine, aspartate, Pantothenate kinase 1 (PANK1), Vanin 1 (VNN1)), and *"Nicotinate and nicotinamide metabolism"* (aspartate, Quinolinate phosphoribosyltransferase (QPRT), Purine nucleoside phosphorylase (PNP), SIRT3, 5'-nucleotidase ecto (NT5E)).

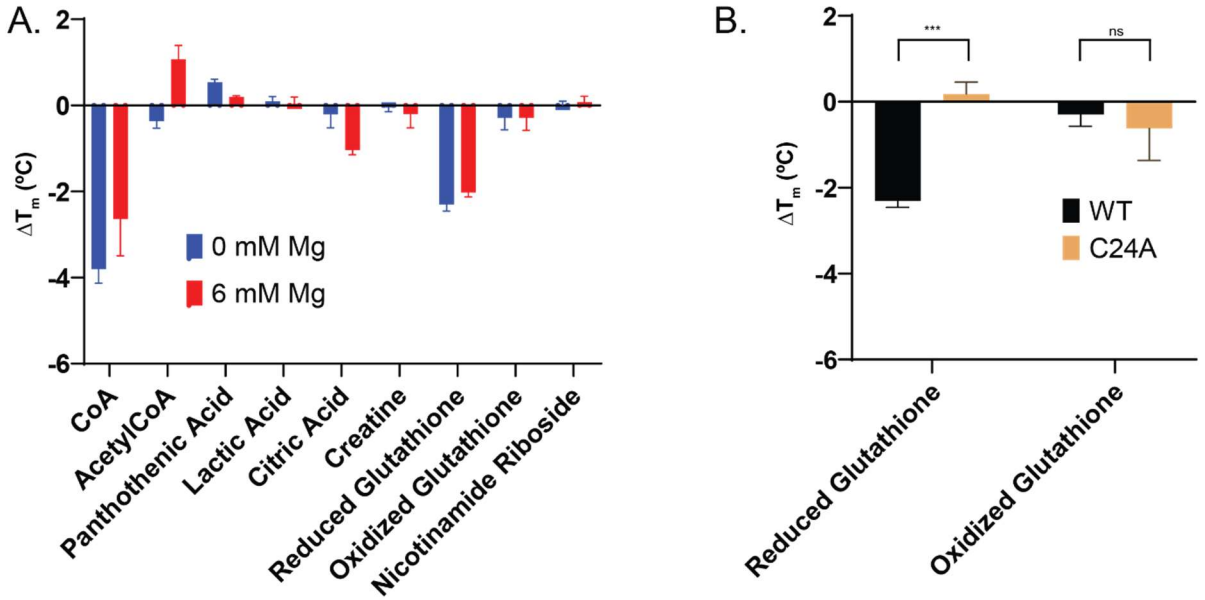


Figure 3.5: Binding of Human FN3K to different metabolites.

(A) Word plot showing protein family/domain that is found in the same operon as FN3K/RP orthologs in bacteria. **(B)** A bar chart showing shift in melting temperatures (ΔTm) for HsFN3K WT with 0 mM (blue) or 6 mM (red) magnesium (Mg). **(C)** A bar chart showing shift in melting temperatures (ΔTm) for HsFN3K WT (black) and C24A (yellow) with reduced and oxidized glutathione. Asterisks (*) indicate values that were significantly different from the C24A mutant as compared to the WT ($p<0.05$, Student's t-test). **(D)** A bar chart showing shift in melting temperatures (ΔTm) for HsFN3K WT and the mutants for different nucleotides in the presence of 6 mM Mg. **(B, C, D)** All tested compounds had a concentration of 5 mM.

Since FN3K/RPs are conserved across the tree of life, including several groups of bacteria (Fig. 3.1A), there is a plethora of evolutionary information available at the bacterial operon level. Using the operon structures on ProOpDB [37], we identified proteins present in the same operon as FN3K/RP orthologs. Protein families related to low molecular weight protein tyrosine phosphatase (LMWPTP), Short-chain dehydrogenases/reductases (SDRs), and Sirtuins were

common (Fig. 3.5A). Next, we utilized Uniprot's reaction annotations for the proteins and visualized the ligand and metabolites used by these proteins (Fig. 3.4B). Interestingly, these included nucleotides such as AMP, ADP, and ATP as well as NAD related compounds (NADH, NADPH, NAD+, NADP+) (Fig. 3.4B).

3.3.5 Human FN3K binds to NAD related compounds specifically and in a metal dependent manner.

Based on the enriched pathways and metabolites (Figs. 3.2, 3.3, and 3.4), we conducted a screening to assess the binding of purified Wild Type (WT) HsFN3K to a select group of compounds, utilizing the Differential Scanning Fluorimetry (DSF) assay. While compounds such as lactic acid, creatine, nicotinamide riboside, and oxidized glutathione exhibited no discernible changes in thermal stability, both coenzyme A (CoA) and reduced glutathione displayed destabilizing effects regardless of the presence or absence of magnesium (Fig. 3.5A). Notably, reduced glutathione induced significant destabilization of the WT enzyme when compared to the C24A P-loop cysteine mutant (P value: 1.7E-4) (Fig. 3.5B). Moreover, various NAD compounds exhibited distinct effects on stabilizing the WT HsFN3K, primarily when magnesium was present. Although the stabilities were relatively lower compared to ADP or ATP, they surpassed those of AMP (Fig. 3.6A). Interestingly, the reduced forms (NADH/NADPH) demonstrated greater thermal stability than their oxidized counterparts (Fig. 3.6A). Notably, among the NAD compounds, NADH exhibited the highest thermal stability. Subsequently, we evaluated the specificity of this interaction by subjecting the enzyme to DSF at varying concentrations of the NAD compounds. Indeed, all NAD compounds displayed saturation kinetics, suggesting a specific binding affinity to the enzyme (Fig. 3.6B). Notably, at sub-millimolar concentrations, NADH

exerted the most pronounced stabilizing effect on human FN3K when compared to other NAD compounds.

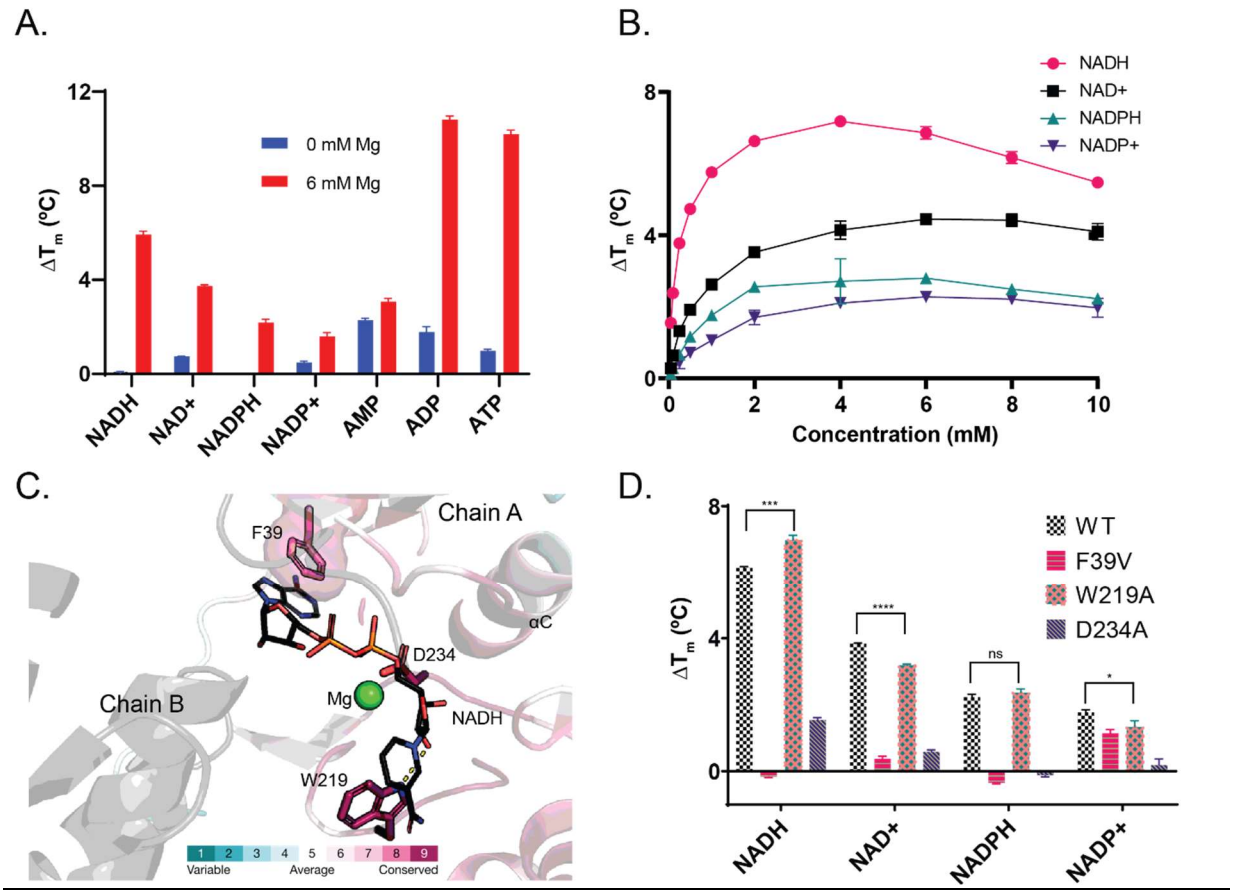


Figure 3.6: Interaction of Human FN3K (HsFN3K) and mutants with NAD compounds.

(A) A bar chart showing shift in melting temperatures (ΔTm) for HsFN3K WT with 0 mM (blue) or 6 mM (red) magnesium (Mg). All tested compounds had a concentration of 5 mM. **(B)** Changes in ΔTm of HsFN3K WT with 10 mM Mg at varied NAD compound concentrations ranging from 0.050 mM to 10 mM. **(C)** Homology model depicting the dimer of HsFN3K with NADH and Mg, utilizing Acedock and Metal Ion Binding (MIB) server, respectively. A. thaliana FN3K crystal structure (PDB ID: 6OID) was used as a template. The degree of conservation of the residues was evaluated and color-coded using the ConSurf server. **(D)** Bar chart showing ΔTm for HsFN3K WT and the mutants in the presence of various NAD compounds. Asterisks (*) indicate values that were significantly different from the W219A mutant as compared to the WT ($p<0.05$, Student's t-test). **(A, B, D)** Mean ΔTm values ± SD were calculated from 3 independent experiments (N= 3) and were calculated by subtracting the control Tm value (Apo).

3.3.6 The ATP binding pocket is a likely binding site for NADH.

To better understand how NAD compounds bind to HsFN3K, we conducted docking experiments using Acedock [38]. Since NADH exhibited the highest stabilization effect in our DSF assays, we proceeded to dock the NADH molecule onto the homology model of HsFN3K.

Leveraging the crystal structure of the plant homolog from *A. thaliana* FN3K (AtFN3K) (PDB ID: 6OID) as a template, which we had previously demonstrated forms a disulfide-linked dimer[14], we aimed to shed light on the binding mode of HsFN3K. Additionally, we introduced a magnesium ion into the homology model using the Metal Ion-Binding webserver[39], given that our DSF assays indicated magnesium dependency for NADH binding. Since NADH encompasses the adenine dinucleotide phosphate (ADP) moiety, we executed scaffold docking based on the ADP molecule's placement within the AtFN3K crystal structure.

Among the diverse predicted binding modes for NADH, the top-scoring mode revealed the involvement of highly conserved aromatic residues, F39 and W219, within the ADP binding pocket (Fig. 3.6C). The nicotinamide moiety of NADH was anticipated to establish hydrogen bonds along with CH-π interactions with W219 (Fig. 3.6C), while the adenine moiety engaged in π-π interaction with F39. Notably, F39 in HsFN3K corresponds to F47 in AtFN3K, which partakes in a π-π interaction with the adenine ring of the ADP molecule within the crystal structure (PDB ID: 6OID) [14].

To delve deeper into the specificity of the interaction between NAD compounds and HsFN3K, we performed site-directed mutagenesis guided by our docking findings. Subsequently, we conducted DSF assays on the mutant proteins. Specifically, we substituted the aromatic F39 and W219 with aliphatic valine (F39V) and alanine (W219A) residues, respectively. Additionally, we introduced an alanine mutation (D234A) for the metal chelating D234 as a control.

Among the three mutants, F39V led to the complete elimination of binding to NADH and NADPH, while retaining detectable thermal stabilities for NAD+ and NADP+ (Fig. 3.6D). Interestingly, the W219A mutant displayed subtle alterations in thermal stability between the reduced and oxidized states of NAD compounds. The W219A mutant exhibited increased stability

in the presence of reduced NAD compounds (NADH/NADPH) but diminished thermal stability when exposed to the corresponding oxidized forms (NAD+/NADP+), compared to the WT enzyme. Although nuanced, these stability changes proved statistically significant for NADH (P value: 7.0E-4) and NAD+ (P value: 1.4E-5). Intriguingly, the W219A mutation also reduced thermal stability for all tested nucleotides. As for the D234A mutant, detectable thermal stability was observed for NADH and NAD+, while the phosphorylated forms NADPH and NADP+ did not exhibit notable stability (Fig. 3.6D).

3.4 Discussion

We leveraged the FN3K CRISPR KO in HepG2 cells and performed the first comprehensive study to investigate the function of this understudied family of eukaryotic-like kinases. Based on the multi-omics analyses and biochemical assays, we uncovered a potential role for FN3K in oxidative stress response, lipid metabolism (cholesterol and fatty acid), and carbon metabolism pathways through specific interaction with NAD compounds in a metal-dependent manner (Figs. 3.2-3.6). While there have been no previous studies investigating the binding of FN3Ks to NAD compounds, an unbiased NAD interactome study using clickable, photoaffinity labeling identified HsFN3KRP in one of its replicates [40]. Notably, the study was done with HEK293T cell lysates where HsFN3K expression is very low (2.2 nTPM as compared to 51 nTPM for HsFN3KRP) [41]. To our knowledge, AMP-activated protein kinase (AMPK) is the only kinase previously reported to bind NAD compounds with NADH inhibiting AMPK activity and NAD+ activating it [42]. Interestingly, we saw downregulation of PRKAB2, the non-catalytic unit of AMPKs, in the FN3K KO cells. Moreover, the abundances of metabolites tied to energy production such as lactate, creatine, and phosphocreatine, significantly differed between the WT

and KO cell lines [14], suggesting links between FN3K and energy production through the NAD compounds.

In cells, two major redox couples are related to NAD compounds: NAD+/NADH and NADP+ and NADPH. The remaining two are glutathione (GSSG/GSH) and thioredoxin (Trx(SH)2/TrxSS) [43]. The redox state in the subcellular compartments is fine-tuned by controlling the ratio of the oxidized and reduced forms of these compounds. For instance, mitochondria have high NADH concentrations to provide reducing power for oxidative phosphorylation, whereas the cytosol has more NAD+ [43]. While there are neither mitochondrial nor nuclear localization signals on HsFN3K, HsFN3K is localized in mitochondria in HepG2 cells [13]. Although NADH stabilized HsFN3K the most among the NAD compounds (Fig. 3.6A), it is possible that in cells, HsFN3K is regulated by the changes in the ratios of the NAD redox couples.

NADP+/NADPH redox couple is critical to the antioxidant system as it donates electrons to glutathione and thioredoxin systems [43] (Fig. 3.7). Both these systems play a pivotal role in ROS response. ROS can covalently modify redox-active cysteines in proteins to form sulfenylated forms and disulfide bonds to trigger oxidative stress response signaling [44]. Notably, several of the proteins in the enriched pathways have evidence of sulfenylated cysteines, including many of the NAD(P) binding superfamily members (Fig. 3.2B). Indeed, we have previously shown that under oxidative stress conditions HsFN3K forms disulfide-linked dimer via the P-loop cysteine (C24) [14]. Moreover, other studies have found sulfenylation of the cysteine in both HsFN3K and HsFN3KRP [45]. The increased presence of reduced glutathione combined with the elevated expression of oxidative stress response genes (including GPX3, MGST1, GSTA1, and metallothioneins) in FN3K-KO provides additional support for direct links between HsFN3K and oxidative stress response. Since NRF2, the master regulator of oxidative stress response, is

deglycated by HsFN3K [15], it is possible that HsFN3K knockout alters oxidative stress response and redox balance through the NRF2 pathway. Alternatively, other proteins within these pathways could be direct substrates of HsFN3K, and mapping these substrates and sites of deglycation would be a major goal for future studies.

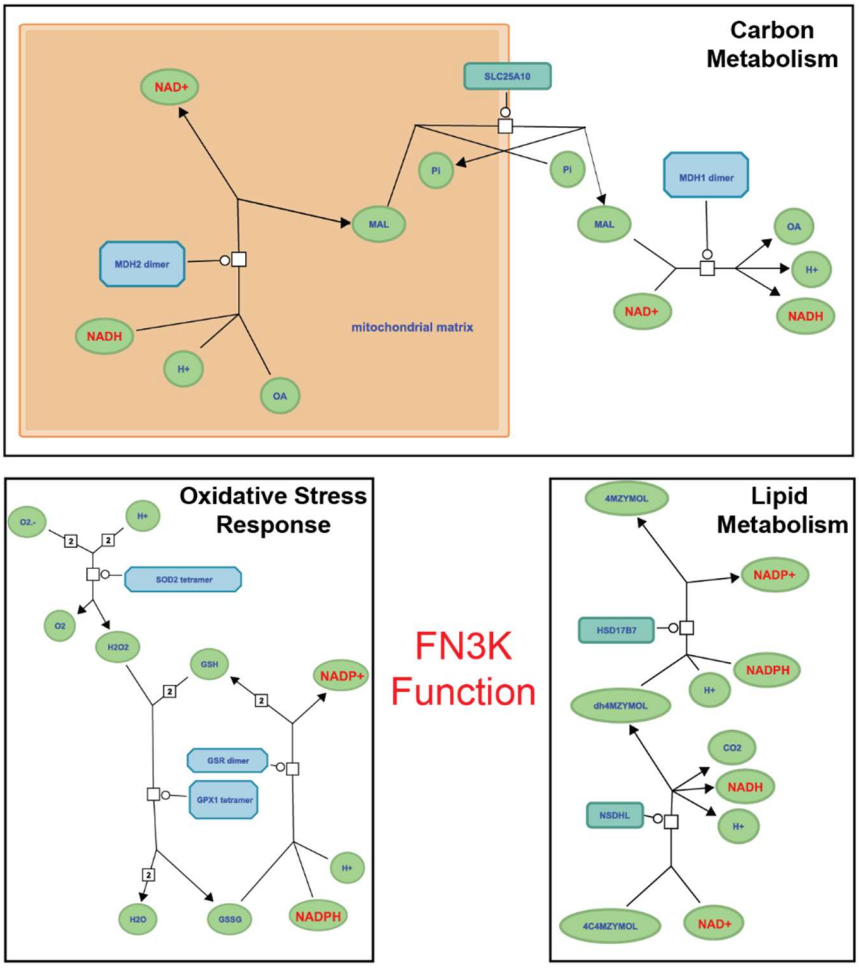


Figure 3.7: Model of HsFN3K cellular regulation and functions.

The enriched pathways based on integrative omics studies and the NAD compounds are shown in the Reactome [46] pathways.

3.5 Materials and Methods

3.5.1 RNA-seq sample preparation

FN3K KO was generated in HepG2 cells (bought from American Type Culture Collection (ATCC) HB-8065) as described previously [14]. HepG2 (WT and KO) cells were cultured on 6 am plates in triplicate to 70-80% confluency in Eagle's Minimum Essential Media (EMEM, ATCC) containing 10% Fetal Bovine Serum. Cells were washed with PBS and harvested. The cell pellet obtained was shipped on dry ice to Novogene Corporation Inc (1007 Slater Road, Suite 140, Durham, NC 27703) for RNA extraction and RNA seq analysis.

3.5.2 Immunoprecipitation and Mass Spectrometry sample preparation

HepG2 (WT and KO) cells were cultured on 10 cm plate to 80-90% confluency. Cells were washed with PBS and lysed in 1ml of lysis buffer (50 mM Tris.HCl, pH 7.4, 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 10% glycerol and protease inhibitor cocktail). Cell lysate was spun down at 12000 rpm, and supernatant was subjected to immunoprecipitation with rabbit FN3K Polyclonal Antibody (1:200 folds dilution) (Invitrogen Catalog # PA5-28603) and 40ul slurry of protein A/G plus agarose beads (Pierce 20423) at 4C for 18h. Beads were washed 3 x lysis buffer, and bound proteins were eluted in 1x Laemmli SDS sample buffer. Proteins were resolved on 10% SDS-PAGE gel by running gel for 10 min (avoid running gel to full to minimize gel amount). The whole lane was cut and sent for trypsinization and MS analysis. The second gel was run to analyze the enriched FN3K protein band in WT and the lack of it in KO.

In-gel trypsin digestion protocol was as follows. First, the gel bands were sliced into small pieces, and then rinsed with 50% acetonitrile/20 mM ammonium bicarbonate (~pH7.5-8) twice.

The gel pieces were dehydrated by adding 100% acetonitrile and dried out by a SpeedVac. A various amount of Trypsin solution (0.01μg/μL in 20 mM ammonium bicarbonate) was added until the gel pieces totally absorbed the Trypsin solution. The tubes were placed in an incubator at 37°C overnight. The tryptic peptides were extracted from gel pieces by incubating with 50% acetonitrile/0.1% formic acid twice. The extracts were dried down by a SpeedVac.

Mass spectrometry analyses were performed on a Thermo-Fisher LTQ Orbitrap Elite Mass Spectrometer coupled with a Proxeon Easy NanoLC system (Waltham, MA) located at Proteomics and Mass Spectrometry Facility at the University of Georgia.

The enzymatic peptides were loaded into a reversed-phase column (self-packed column/emitter with 200 Å 5 μM Bruker MagicAQ C18 resin), then directly eluted into the mass spectrometer. Briefly, the two-buffer gradient elution at the flow rate 450 nL/min (0.1% formic acid as A and 99.9% acetonitrile with 0.1% formic acid as B) starts with 0% B, holds at 0% B for 2 minutes, then increases to 30% B in 50 minutes, to 50% B in 10 minutes, and to 95% B in 10 minutes.

The data-dependent acquisition (DDA) method was used to acquire MS data. A survey MS scan was acquired first, and then the top 10 ions in the MS scan were selected for following CID MS/MS analysis. Both MS and MS/MS scans were acquired by Orbitrap at the resolutions of 120,000 and 15,000, respectively.

Data were analyzed using Xcalibur software (version 2.2, Thermo Fisher Scientific). Protein identification and modification characterization were performed using Thermo Proteome Discoverer (version 1.4) with Mascot (Matrix Science 2.7) and Uniprot database. The spectra of possible modified peptides were inspected further to verify the accuracy of the assignments.

3.5.3 RNA-Seq data analysis

We conducted RNA sequencing (RNA-Seq) on total RNA from the FN3K KO and the wild-type (WT) cell lines. RNA extraction, preparation of RNA library and transcriptome sequencing were conducted by Novogene Co., LTD. The obtained reads were then analyzed using the "new Tuxedo" package suite of tools [47]. HISAT2 [48] was first used to map the reads to the human reference genome separately for the KO and WT samples. This was performed using the default parameters. Then, the alignment was passed to StringTie [49] for assembly and quantification of the transcripts. The assembled transcripts were then passed to Ballgown [50] to perform statistical tests to identify differentially expressed transcripts between the WT and KO samples. We used Ballgown's stattest function to perform a statistical comparison of the FPKM values for all assembled transcripts at the gene level and report the p-value and the fold change for genes that were significantly differentially expressed. Only genes with p-value < 0.05 and a log fold change > 0.2 were considered as significantly differentially expressed and used in subsequent analyses.

3.5.4 Pathway enrichment analysis

We collected the UniProt IDs for i) all the genes that were found to be differentially expressed in the RNA-Seq analysis and ii) the genes that were identified as interacting partners using Co-IP. This list, comprising 513 unique genes, was uploaded into Cytoscape [51] and the STRING Cytoscape plugin [22] was used to build a STRING-based network of all these genes. We then used the STRING plugin to perform an enrichment analysis on this network of genes to find enriched pathways. Since we were interested primarily in identifying pathways, we only retained the enrichment results obtained for pathways from KEGG [52], Reactome [53], and Wiki

pathways [54]. We used a redundancy cutoff of 0.4 within STRING to remove pathways that had a high overlap of genes. This resulted in 76 enriched pathways. We further filtered this list to remove small pathways that had 5 or less genes resulting in 63 enriched pathways with an FDR value below 0.05.

We used fold enrichment, a quantitative measure of the over-representation of a pathway, to rank order the obtained pathways. Fold enrichment is obtained by comparing the background frequency of total genes annotated for that pathway on humans to the actual genes that mapped to that same pathway from the list that was provided. Specifically, fold enrichment for each pathway was calculated using the following proportion:

Fold enrichment = (n/M)/(Nb/Nt)

Where,

n is the total number of genes from our list that match the pathway.

M is the total number of genes in our list (513).

Nb is the total number of background genes (total genes in that pathway) in humans.

Nt is the total number of genes in humans (19566, for human genome used by the STRING database version 11.5)

3.5.5 Differential Scanning Fluorimetry (DSF)

Approximately 7 uM of Human FN3K WT or mutants (F39V, D234A, W219) were mixed with 1:500 SYPRO ORANGE (Sigma) in buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 5% glycerol). The nucleotides or nicotinamides were added in the presence or absence of 6 mM $MgCl_2$ to make the final concentration of 5 mM. The solution was then heated gradually from 20°C to 95

°C at the rate of 0.3°C, and the fluorescence was monitored using the StepOne Plus Real-time PCR instrument. Data was analyzed using Prism.

3.5.6 Phylogenetic analysis

FN3K sequences were identified and aligned using previously curated FN3K profile [55, 56] by querying the Uniprot reference proteome [57]. Fasttree 2.1 [21] created a maximum likelihood tree based on the aligned sequences. The tree was annotated and visualized using the Interactive Tree of Life (iTOL) [58] webserver.

3.5.7 Homology modeling, docking and ConSurf Analysis

The crystal structure of the plant homolog from *A. thaliana* (PDB ID: 6OID) was used as a template for the homology model of HsFN3K. ADP was modeled in the pocket by aligning the crystal structure with the homology model. Magnesium ion was docked on the homology model using the Metal-Ion Binding(MIB) [39] webserver. NADH was docked onto the homology model using scaffold docking with ADP as a template on the Acedock [38] webserver. Chain A of the homology model was used as a query for the ConSurf [59] server.

Bibliography

1. Böhm, P.G.a.M., *Advanced glycation end products. Key players in skin aging?* Dermato-Endocrinology, 2012. **4**(3): p. 259-270.
2. Uribarri, J., et al., *Dietary advanced glycation end products and their role in health and disease.* (2156-5376 (Electronic)).
3. Baynes, J.W., et al., *The Amadori product on protein: structure and reactions.* Prog Clin Biol Res, 1989. **304**: p. 43-67.
4. Hodge, J.E., *The Amadori rearrangement.* Adv Carbohydr Chem, 1955. **10**: p. 169-205.
5. François Collard, G.D., Vincent Stroobant, Gert Matthijs, Emil Van Shaftingen, *A Mammalian Protein Homologous to Fructosamine-3-Kinase Acting on Psicosamines and Ribulosamines but not on Fructosamines.* Diabetes, 2003. **52**: p. 2888-2895.
6. Szwergold, B.S., P.J. Howell S Fau - Beisswenger, and P.J. Beisswenger, *Human fructosamine-3-kinase: purification, sequencing, substrate specificity, and evidence of activity in vivo.* (0012-1797 (Print)).
7. Chen, J.H., et al., *Role of advanced glycation end products in mobility and considerations in possible dietary and nutritional intervention strategies.* (1743-7075 (Print)).
8. Cepas, V., et al., *Redox Signaling and Advanced Glycation Endproducts (AGEs) in Diet-Related Diseases. LID - 10.3390/antiox9020142 [doi] LID - 142.* (2076-3921 (Print)).
9. Juliette Fortpied, R.G., Vincent Stroobant and Emile van Schaftingen, *Plant ribulosamine/erythrulosamine 3-kinase, a putative protein-repair enzyme.* Biochemical Journal, 2005. **388**: p. 795-802.
10. Rita Gemayel, J.F., Rim Rzem, Didier Vertommen, Maria Veiga-da-Cunha and Emile Van Schaftingen, *Many fructosamine 3kinase homologues in bacteria are*

ribulosamine/erythrulosamine 3-kinase potentially involved in protein deglycation. The FEBS Journal, 2007. **274**: p. 4360-4374.

11. Cunningham, F.A.-O., et al., *Ensembl 2022.* (1362-4962 (Electronic)).
12. François Collard, E.W., Niki Bergans, Juliette Fortpied, Didier Vertommen, Florent Vanstapel, Ghislain Delpierre, and Emil Van Shaftingen, *Fructosamine 3-kinase-related protein and deglycation in human erythrocytes.* Biochemical Journal, 2004. **382**: p. 137-143.
13. Thul, P.J., et al., *A subcellular map of the human proteome.* Science, 2017. **356**(6340): p. eaal3321.
14. Shrestha, S.A.-O., et al., *A redox-active switch in fructosamine-3-kinases expands the regulatory repertoire of the protein kinase superfamily. LID - 10.1126/scisignal.aax6313 [doi] LID - eaax6313.* (1937-9145 (Electronic)).
15. Sanghvi, V.R., et al., *The Oncogenic Action of NRF2 Depends on De-glycation by Fructosamine-3-Kinase.* Cell, 2019. **178**(4): p. 807-819.e21.
16. Ngo, V.A.-O. and M.L. Duennwald, *Nrf2 and Oxidative Stress: A General Overview of Mechanisms and Implications in Human Disease. LID - 10.3390/antiox11122345 [doi] LID - 2345.* (2076-3921 (Print)).
17. Rushmore, T.H., C.B. Morton Mr Fau - Pickett, and C.B. Pickett, *The antioxidant responsive element. Activation by oxidative stress and identification of the DNA consensus sequence required for functional activity.* (0021-9258 (Print)).
18. McMahon, M., et al., *Keap1-dependent proteasomal degradation of transcription factor Nrf2 contributes to the negative regulation of antioxidant response element-driven gene expression.* (0021-9258 (Print)).

19. Francesca Ávemaria, P.C., Annunziata Lapolla, Giovanni Sartore, Nino Cristiano Chilelli, Renata Paleari, Alessandro Ambrosi, Maurizio Ferrari and Andrea Mosca, *Possible role of fructosamine 3-kinase genotyping fot the management of diabetic patients.* Clinical Chemistry and Laboratory Medicine, 2015. **53**(9): p. 1315-1320.

20. Lorena Mosca, S.P., Maria C. Patrosso, Alessandro Marocchi, Annunziata Lapolla, Giovanni Sartore, Nino C. Chilelli, Renata Paleari and Andrea Mosca, *Genetic variability of the fructosamine-3-kinase gene in diabetic patients.* Clinical Chemistry and Laboratory Medicine, 2011. **49**(5): p. 803-808.

21. Price, M.N., A.P. Dehal Ps Fau - Arkin, and A.P. Arkin, *FastTree 2--approximately maximum-likelihood trees for large alignments.* (1932-6203 (Electronic)).

22. Doncheva, N.A.-O., et al., *Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data.* (1535-3907 (Electronic)).

23. Dalmer, T.R.A. and R.D. Clugston, *Gene ontology enrichment analysis of congenital diaphragmatic hernia-associated genes.* (1530-0447 (Electronic)).

24. Si, M. and J. Lang, *The roles of metallothioneins in carcinogenesis.* (1756-8722 (Electronic)).

25. Moebius, F.F., et al., *Molecular cloning and expression of the human delta7-sterol reductase.* (0027-8424 (Print)).

26. Jayakumar, A., et al., *Isolation and chromosomal mapping of genomic clones encoding the human fatty acid synthase gene.* (0888-7543 (Print)).

27. Takebe, G., et al., *A comparative study on the hydroperoxide and thiol specificity of the glutathione peroxidase family and selenoprotein P.* (0021-9258 (Print)).

28. Stanton, R.C., *Glucose-6-phosphate dehydrogenase, NADPH, and cell survival.* (1521-6551 (Electronic)).

29. Tang, Y., et al., *Overexpression of PCK1 Gene Antagonizes Hepatocellular Carcinoma Through the Activation of Gluconeogenesis and Suppression of Glycolysis Pathways.* (1421-9778 (Electronic)).

30. Zhang, L., et al., *Proprotein convertase subtilisin/kexin type 9 (PCSK9) in lipid metabolism, atherosclerosis and ischemic stroke.* (1563-5279 (Electronic)).

31. Leal, N.A., et al., *Human ATP:Cob(I)alamin adenosyltransferase and its interaction with methionine synthase reductase.* (0021-9258 (Print)).

32. Wang, P., et al., *iCysMod: an integrative database for protein cysteine modifications in eukaryotes. LID - bbaa400 [pii] LID - 10.1093/bib/bbaa400 [doi].* (1477-4054 (Electronic)).

33. Xiao, W., et al., *NAD(H) and NADP(H) Redox Couples and Cellular Energy Metabolism.* (1557-7716 (Electronic)).

34. Zhang, J., et al., *Mitochondrial Sirtuin 3: New emerging biological function and therapeutic target.* (1838-7640 (Electronic)).

35. Jiao, X., et al., *A mammalian pre-mRNA 5′ end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing.* (1097-4164 (Electronic)).

36. Chong, J., D.S. Wishart, and J. Xia, *Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis.* (1934-340X (Electronic)).

37. Taboada, B., et al., *ProOpDB: Prokaryotic Operon DataBase.* (1362-4962 (Electronic)).

38. Ruiz-Carmona, S., et al., *rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids.* (1553-7358 (Electronic)).

39. Lin, Y.F., et al., *MIB: Metal Ion-Binding Site Prediction and Docking Server.* (1549-960X (Electronic)).

40. Šileikytė, J.A.-O., et al., *Chemical Proteomics Approach for Profiling the NAD Interactome.* (1520-5126 (Electronic)).

41. Uhlén, M., et al., *Proteomics. Tissue-based map of the human proteome.* (1095-9203 (Electronic)).

42. Rafaeloff-Phail, R., et al., *Biochemical regulation of mammalian AMP-activated protein kinase activity by NAD and NADH.* (0021-9258 (Print)).

43. Berthiaume, J.M., et al., *Mitochondrial NAD(+)/NADH Redox State and Diabetic Cardiomyopathy.* (1557-7716 (Electronic)).

44. Chung, H.S., et al., *Cysteine oxidative posttranslational modifications: emerging regulation in the cardiovascular system.* (1524-4571 (Electronic)).

45. Xiao, H., et al., *A Quantitative Tissue-Specific Landscape of Protein Redox Regulation during Aging.* Cell, 2020. **180**(5): p. 968-983.e24.

46. Croft, D., et al., *Reactome: a database of reactions, pathways and biological processes.* Nucleic acids research, 2011. **39**(suppl 1): p. D691-D697.

47. Pertea, M., et al., *Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown.* (1750-2799 (Electronic)).

48. Kim, D.A.-O.X., et al., *Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype.* (1546-1696 (Electronic)).

49. Pertea, M., et al., *StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.* (1546-1696 (Electronic)).

50. Frazee, A.C., et al., *Ballgown bridges the gap between transcriptome assembly and expression analysis.* (1546-1696 (Electronic)).

51. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* (1088-9051 (Print)).

52. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* (0305-1048 (Print)).

53. Jassal, B., et al., *The reactome pathway knowledgebase.* (1362-4962 (Electronic)).

54. Martens, M., et al., *WikiPathways: connecting communities.* (1362-4962 (Electronic)).

55. Oruganty, K., et al., *Identification and classification of small molecule kinases: insights into substrate recognition and specificity.* BMC Evol Biol, 2016. **16**: p. 7.

56. Kannan, N., et al., *Structural and functional diversity of the microbial kinome.* PLoS Biol, 2007. **5**(3): p. e17.

57. UniProt Consortium, T., *UniProt: the universal protein knowledgebase.* Nucleic Acids Research, 2018. **46**(5): p. 2699-2699.

58. Letunic, I.A.-O. and P. Bork, *Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees.* (1362-4962 (Electronic)).

59. Ashkenazy, H., et al., *ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules.* (1362-4962 (Electronic)).

CHAPTER 4

EVOLUTIONARY AND STRUCTURAL ANALYSES OF THE FN3K FAMILY REVEALS NEW INSIGHT INTO NUCLEOTIDE BINDING

Safal Shrestha and Natarajan Kannan.
To be submitted to PNAS

4.1 Abstract

In humans, Fructosamine-3-kinase (FN3K) plays a crucial role in cellular homeostasis by performing protein repair through deglycation. Importantly, it is one of the few members of the protein kinase-like (PKL) fold family that is conserved across the tree of life. Despite its conservation across all domains, a focused, deep evolutionary analysis of the family is lacking. Previous work has identified the core domain shared by all PKL members and has derived evolutionary relationships between members in the superfamily. This work showed that FN3Ks are closely related to other small molecule kinases in the eukaryotic-like kinase (ELKs) group of the superfamily. These include families such as aminoglycoside phosphotransferases (APHs) and Choline Kinase (ChoK). In this study, we define the core domain shared by FN3Ks and related ELKs and identify sequence constraints distinctive to the FN3K family. We further derive evolutionary relationships between FN3K family members and identify several bacterial sub-groups that are divergent in the nucleotide and substrate-binding pockets. We also explored the bacterial operon structure containing FN3K and modeled the protein complexes formed by the proteins in the operons. Notably, we find an interface on FN3Ks that is consistently predicted to form interactions with diverse domains in the operon. We also structurally characterize FN3K orthologs from AtFN3K WT and TfFN3K WT bound to adenine-containing nucleotides. We find two distinct modes of adenine binding in TfFN3K and, through a structural search, show a similar mode of binding in the CcrZ protein.

4.2 Introduction

Glycation is a spontaneous reaction in which reducing sugars, such as glucose and ribose, bind to free amines on proteins, lipids, and DNA molecule [1]. Fructosamine-3-kinases (FN3K) play a central role in cellular homeostasis by repairing protein via the deglycation process [2-6]. Specifically, FN3Ks remove glycated adducts from lysine residues by transferring a phosphate from ATP molecules [7, 8]. Notably, the deglycation function of FN3K has been demonstrated to be vital for the proper functioning of the master regulator of oxidative stress response transcription factor gene, NRF2, in the development of hepatocarcinoma [9, 10]. Additionally, we previously highlighted a novel mode of redox regulation using the crystal structure of the FN3K ortholog from *A. thaliana* (AtFN3K) [11] and more recently provided links to lipid metabolic pathways, co-factor metabolisms, and nicotinamide adenine dinucleotides (NAD) molecules (Chapter 3).

FN3Ks are members of the protein kinase-like (PKL) fold superfamily and share a distant relationship with Eukaryotic Protein Kinases (EPKs), which phosphorylate protein substrates on serine, threonine, and tyrosine residues [12-16]. The PKL superfamily comprises three major groups: EPKs, eukaryotic-like kinases (ELKs), and atypical kinases (APKs). While previous studies [17, 18] on the PKL superfamily have elucidated evolutionary relationships among its members by defining a core set of secondary structure elements and identifying sequence motifs crucial for catalysis and divalent metal binding, a comprehensive evolutionary analysis of FN3K in context with related ELKs remains absent. FN3Ks and related ELKs share secondary elements beyond the core domain. Importantly, the subset of ELKs includes families such as aminoglycosides phosphotransferases (APHs) that phosphorylate antibiotics such as kanamycin at 2’ and 3’ hydroxyls [19-23], Choline Kinase (ChoK) that phosphorylate choline [24, 25], 5-Methylthioribose Kinase (MTRKs) that phosphorylate 5-Methylthioribose [26], among others.

In this study, we employed the DeepAlign [27], a multiple protein structure alignment tool to produce an accurate seed alignment, expanding the core domain to encompass secondary structure elements shared by FN3Ks and related ELKs. We derive a phylogenetic relationship between the ELK members and show that FN3Ks are more closely related to Glucosamine kinase (GluK) and stress response YiHE family members than to ChoK and APH families. We also define sequence constraints that distinguish FN3Ks from related ELKs using optimal multiple category Bayesian Partitioning with Pattern Selection (omcBPPS) [28] program. We further identified multiple divergent bacterial sub-groups of FN3Ks as well as sequence constraints associated with them. We also surveyed the diversity of bacterial operon with FN3K, and modelled complexes formed by proteins in the operon using AlphaFold multimer [29]. The models consistently predict the FN3K interface consisting of ATP and substrate binding pocket for protein-protein interaction. We also structurally characterized FN3K orthologs *Thermobifida fusca* (TfFN3K) and AtFN3K bound to different nucleotides and showed two distinct binding modes for the adenine moiety. Structural motif search using Fit3D [30] provided insights into how spatial arrangement of aromatic and aliphatic residues in the ATP binding pocket could facilitate distinct mode of adenine binding in the ELK family.

4.3 Results

4.3.1 Structural similarities and differences between FN3Ks and related ELKs to EPKs.

FN3Ks belong to the Protein Kinase Like (PKL) fold superfamily which consists of eukaryotic protein kinases (EPKs), atypical kinases (APKs), and eukaryotic like kinases (ELKs) [17, 31]. All PKL fold enzymes adopt a bilobal structure with β strand rich N lobe and alpha helices

rich C lobe. Through structural and sequence comparisons of the PKL members, previous works have identified a set of secondary structural elements conserved in all PKL members collectively known as the core domain (Fig. 4.1A, colored dark gray). The core domain includes the five β strands (β1- β5) and αC helix in the N lobe (α2 in FN3K) whereas in the C lobe, the core domain consists of the αE (α3 in FN3K), catalytic loop with the catalytic base (HRD-Asp in EPKs, HGD-Asp in FN3K), the magnesium binding loop containing the (DFG-Asp in EPK, DPA-Asp in FN3K), and the αF helix (α6 in FN3K) (Fig. 4.1A). Despite these conserved regions, there are major differences between EPKs and ELKs especially in the C-lobe due to differences in where insertions are present between the core domain elements. In EPKs, the insertion is located between β9 and the αF and forms the regulatory loop commonly known as the activation loop (Fig. 4.1A, B). The loop consists of phosphorylatable residues that dictate the switch between the active or inactivated states in EPKs [12]. On the other hand,small molecule kinases have an insertion between αE and the catalytic loop. The insertion varies in length between families but usually forms at least two helices (α4 and α5 in FN3K). Notably, in FN3Ks, the insertion consists of two additional antiparallel β strands (β6 and β7) that protrudes into the putative substrate binding pocket (Fig. 4.1A, B). Further differences are observed in the C terminus of the αF helix. In ELKs, out of the three terminal helices, two of the helices (α8 and α9 in FN3K) form a four helical bundle with the helices (α4 and α5) formed by the ELK specific insert region (Fig. 4.1B).

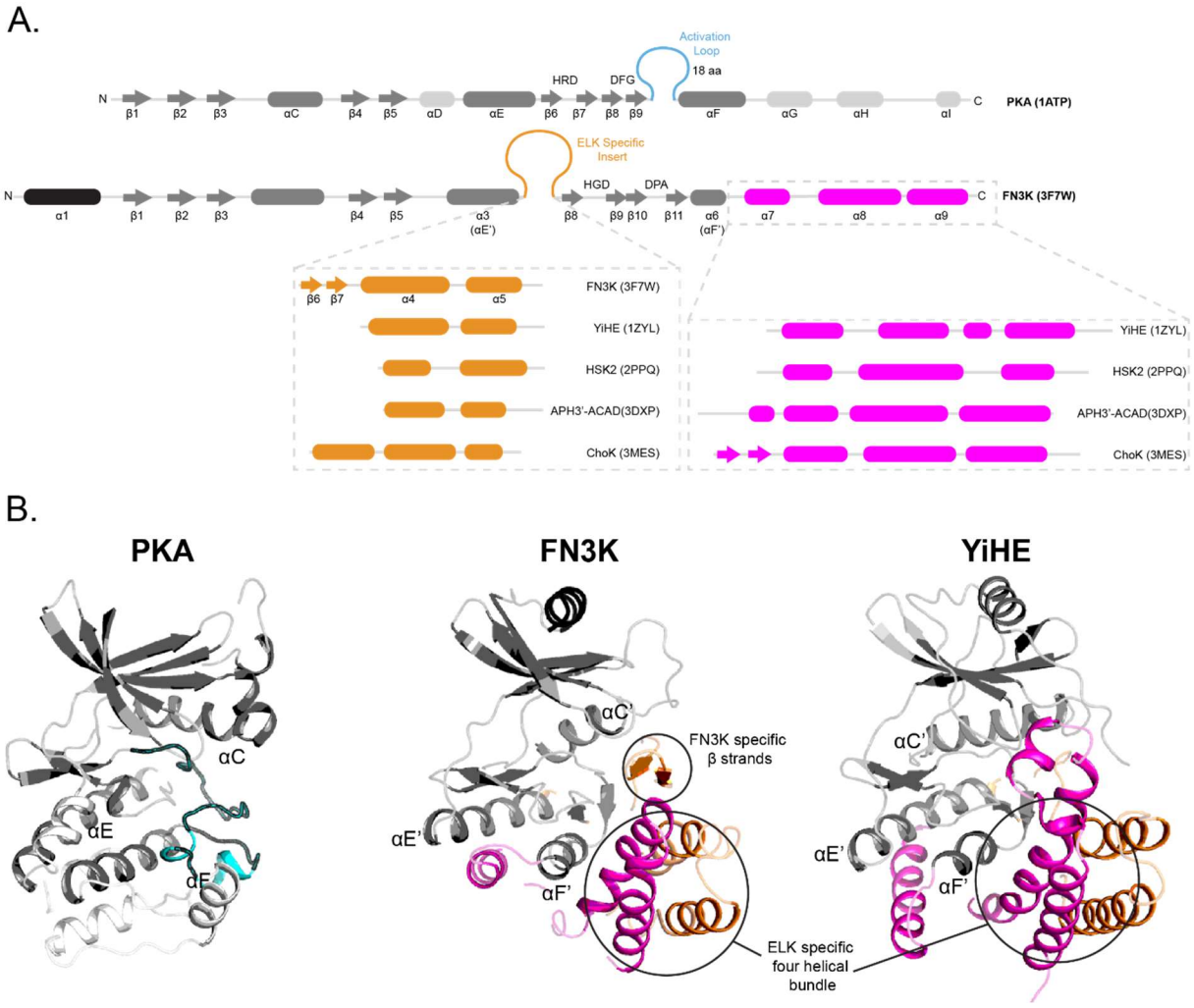


Figure 4.1. Differences in secondary structure elements between FN3K and EPKs.

(A) Secondary structure element for prototypic EPK, Protein Kinase A (PKA), FN3K and related ELKs. The core domain conserved across PKL is colored dark gray. **(B)** Cartoon representation of the crystal structure of the eukaryotic protein kinase PKA (PDB: 1ATP), and ELKs: FN3K (PDB: 3F7W), and YiHE (PDB: 1ZYL). The activation loop and ELK specific inserts are colored in cyan and orange respectively. The C terminal helices to αF are colored light gray and magenta for EPK and ELK respectively.

4.3.2 Evolutionary relationship between FN3Ks and related ELKs reveal sequence constraints in both ATP and substrate binding pockets.

While previous works compared the diverse PKL members and identified sequence constraints within the core domain shared by the PKL members, there is a lack of a deeper evolutionary comparison focusing solely on ELK members closely related to FN3K. Here, we

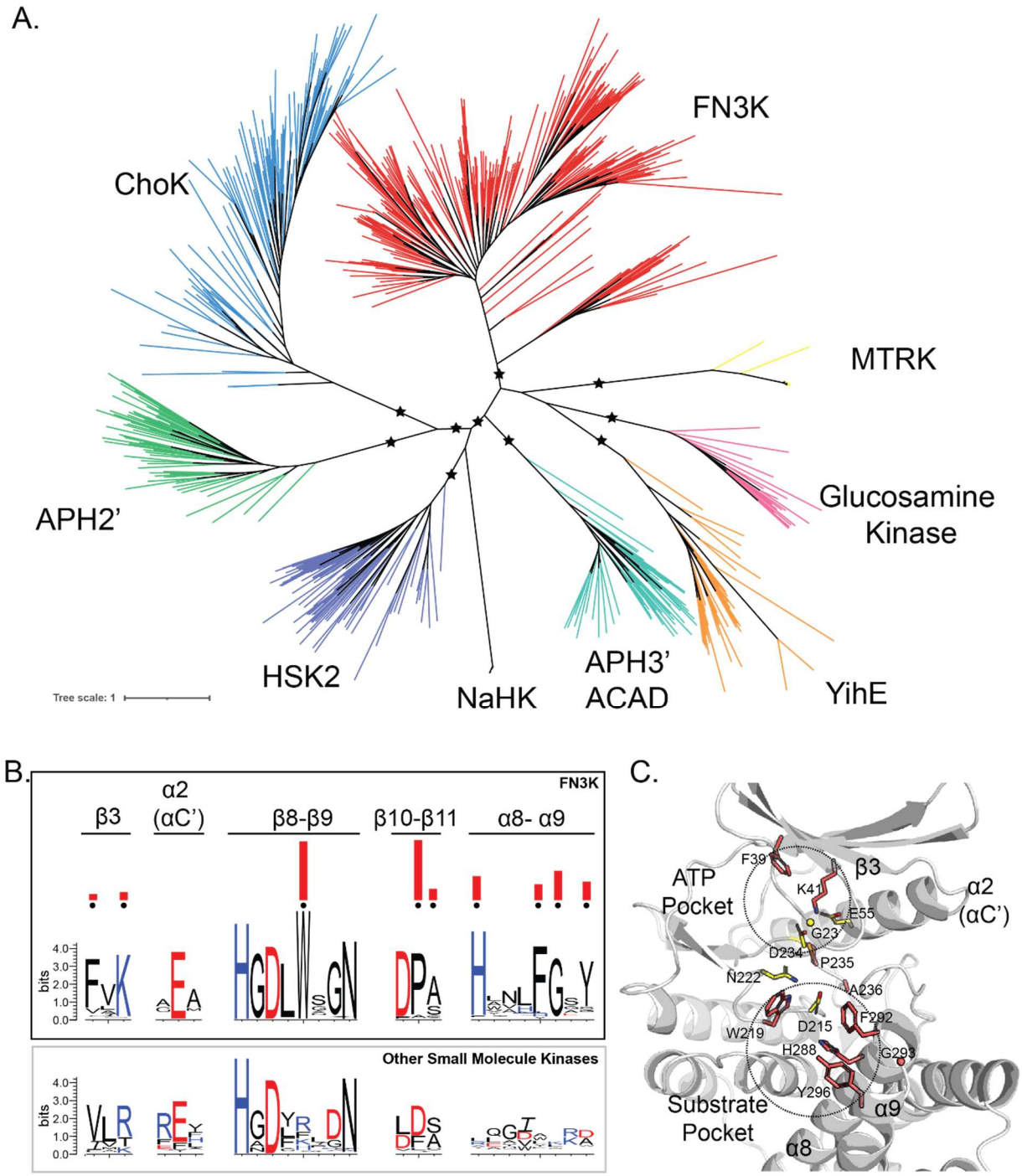


Figure 4.2. Evolutionary relationships between FN3Ks and other related ELKs based on the extended core alignment.

(A) Maximum likelihood phylogenetic tree showing relationship between FN3K and other related ELK family members. The tree was constructed using Fasttree. Notable nodes with bootstrap values greater than 0.8 are indicated with black stars. **(B)** Weblogo of the FN3K sequences (foreground; top) and other related ELKs (background; bottom). FN3K specific sequence constraints distinguishing FN3K from other ELK sequences identified through the omcBPPS procedure. The histogram (red) above the Weblogo indicates the extent to which distinguishing residues in the foreground alignment diverge from the corresponding position in the background alignment. Black dots mark the alignment positions used by the omcBPPS procedure. **(C)** FN3K specific sequence constraints mapped to human FN3K AlphaFold model (Uniprot ID: Q9H479). FN3K specific

sequence constraints are shown as sticks and with carbon atoms colored in magenta. FN3K-ELK shared sequence constraints are shown as sticks with carbon atoms colored in yellow.

define the extended core domain shared by FN3Ks and related ELKs using available sequences and crystal structures (See Methods). The extended core domain now includes the two helices (α4 and α5 in FN3K; Fig. 4.1) in the ELK specific insertion as well as the terminal three helices (α7-α9 in FN3K; Fig. 4.1). The profile, based on seed alignment with the extended core, was then used to align and identify diverse ELK members within the Uniprot reference proteome.

To understand the evolutionary relationship between the ELK members related to FN3Ks, we generated a maximum likelihood tree from the purged sequence sets representing each of the ELK families using Fasttree [32] (Fig. 4.2A). Notably, several ELK families formed monophyletic clades with high confidence (bootstrap > 0.8, indicated by black star). We find that FN3K is more closely related to ELK families such as MTRK, Glucosamine Kinase, and YihE. Interestingly, the branch consisting of ELK families such as ChoK, APH2', Homo serine kinase 2 (HSK2), N-acetyl hexosamine kinase (NaHK), and APH3' had strong bootstrap support (0.84) suggesting closer evolutionary relationship amongst themselves.

Next, we used the omcBPPS procedure to find sequence patterns distinctive of the FN3K family as compared to the other ELK groups. The analysis revealed sequence constraints in key regions associated with both nucleotide such as ATP and substrate binding. Notably, FN3Ks usually conserve an aromatic Phe (Phe39 in Human FN3K) residue in β3 (Fig. 4.2B, C). This residue is equivalent to the aliphatic VAIK-Ala in EPKs (A70, PKA numbering) that forms Van der Waals interactions with the adenine moiety of the ATP molecule [12, 13]. In other ELK families, like in EPKs, this residue is usually an aliphatic valine. The invariant β3 VAIK-Lys (K72, PKA numbering) found in EPKs is variable in ELKs and is substituted by residues such as arginine and threonine. However, like EPKs, in FN3Ks this is usually conserved as lysine (K41, HsFN3K

numbering) (Fig. 4.2B, C). Importantly, the most distinctive sequence patterns were localized to the catalytic loop, the magnesium binding loop, and the two terminal helices. In the catalytic loop, FN3Ks have an invariant tryptophan residue (W219), whereas in the magnesium binding loop, FN3K family members conserve a proline (P235) following the magnesium binding aspartic acid (Fig. 4.2B, C). In the two terminal helices, FN3Ks conserve the HxxFGxxY (H288, F292, G293, Y296) motif forming the other side of the putative substrate binding pocket (Fig. 4.2B, C).

4.3.3 Identification of several monophyletic clades in the FN3K family.

Using the phylogenetic approach, we identified the evolutionary relationships among FN3K orthologs. The maximum likelihood tree showed several monophyletic clades especially in bacterial FN3Ks (Fig. 4.3A). Interestingly, several species even within the same groups showed a closer relationship to other groups rather than their own; for instance, α-proteobacteria showed a closer relationship to FN3Ks belonging to the Actinomycetota group instead of ɣ-proteobacteria FN3Ks (Fig. 4.3A). Instead, ɣ-Proteobacteria FN3Ks showed a closer relationship to Cyanobacterial FN3Ks. Moreover, two distinct clades consisting of both Fibrobacterota, Chlorobiota, and Bacteroidota (FCB)/Bacillota groups were identified (Fig. 4.3A). Moreover, plants and fungi FN3Ks formed monophyletic clades and were more closely related to the metazoan FN3Ks where FN3K and FN3K related protein (FN3KRP) forming two distinct clades (Fig. 4.3A).

Next, we applied omcBPPS to identify functionally important residues defining the divergent FN3K sub-groups. For the distinct subgroups, the sequence constraints defining the subgroups were localized to various parts of the kinase domain. Importantly, for some subgroups, the constraints were found in both the ATP and substrate binding pockets (Fig. 4.3B). Within the

ATP binding pocket, certain species from Bacillota which belong to Firmicutes like *L. brevis*, conserve an aliphatic Leucine (L44) residue instead of the positively charged lysine residue found in most FN3K sequences (K41; Human FN3K numbering) (Fig. 4.3B). Similarly, the FN3K specific aromatic phenylalanine in β3 (F39; Human FN3K numbering) is replaced by an aliphatic valine residue in α-proteobacteria including *C. papalotli* (Fig. 4.3A, B). In species belonging to ɣ-proteobacteria including *E. coli* and *S. enterica*, sequence constraints are in the substrate binding pocket. In the α8 helix, an arginine (R264) replaces the usually conserved histidine (H288, Human FN3K numbering) residue that is usually conserved in FN3Ks. In a group of Actinomycetota, the phenylalanine (F292, Human FN3K numbering) is replaced by a histidine residue (H234) (Fig. 4.3B, right). Interestingly, we also find major structural divergence within this subgroup. Unlike other FN3Ks that have five β-strands in the lobe *including E. coli* FN3K, the FN3K orthologs within the subgroup only have three β-strands (Fig. 4.3C, right). Notably, the subgroup still conserves the β3 lysine (K7).

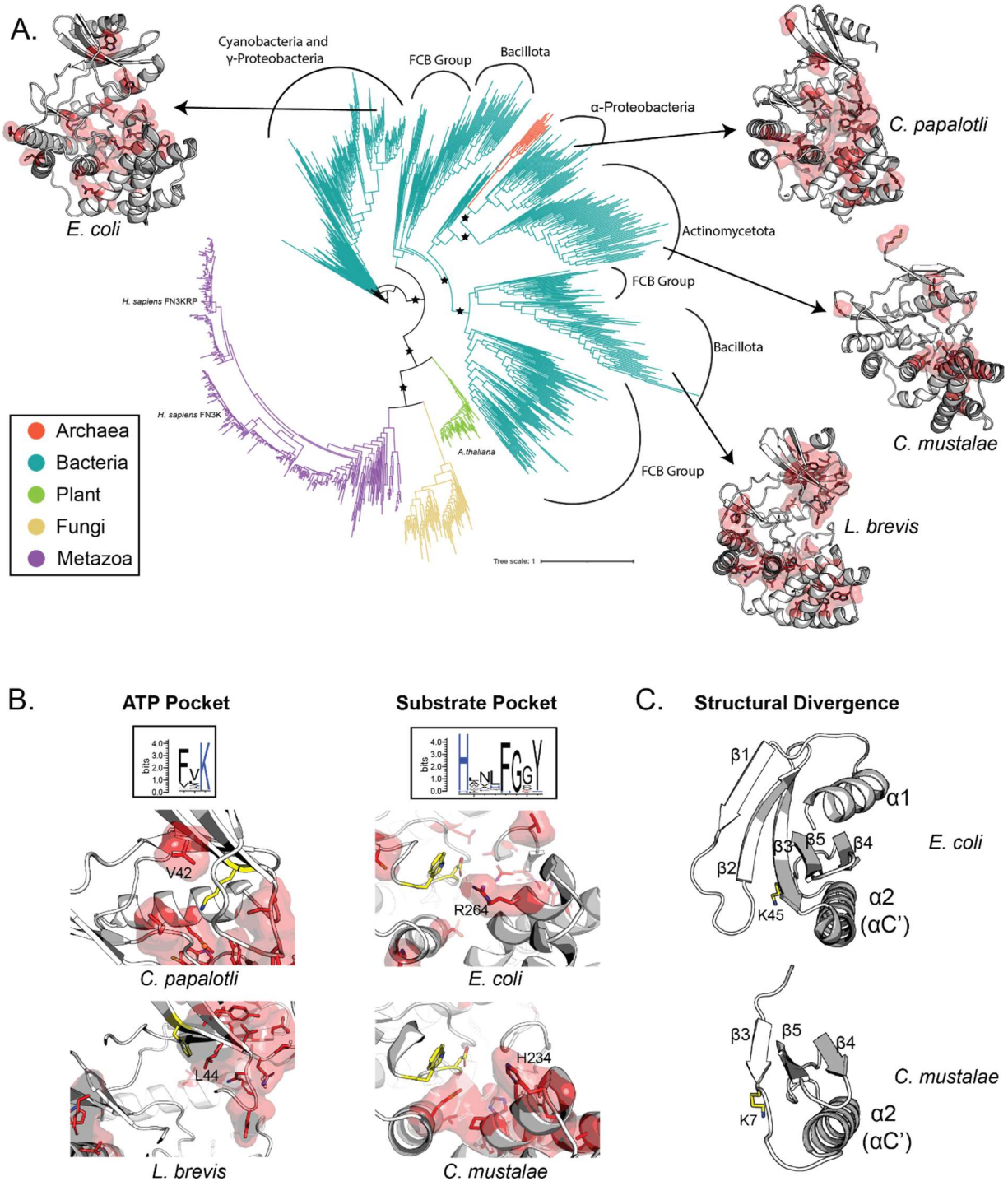


Figure 4.3: Evolutionary relationship between FN3K orthologs reveal several distinct clades.

(A) Maximum likelihood phylogenetic tree showing relationship between FN3K orthologs. Notable nodes with bootstrap values greater than 0.8 are indicated with black stars. Cartoon representation of AlphaFold models of bacterial FN3Ks with sequence constraints as identified by omcBPPS procedure specific to the sub-group mapped as sticks with carbon atoms colored in red. **(B)** Divergence within the FN3K orthologs at key positions in the ATP and substrate binding pockets. (Top) Weblogo of FN3K sequences at those positions showing canonical residues. FN3K specific constraint residues are shown as sticks and with carbon atoms colored yellow. The sub-group specific constraint residues are shown as sticks with carbon atoms colored in magenta.

(C) Cartoon representation of the N-lobe of E. coli FN3K as compared to C. mustalae FN3K orthologs. β3 lysine is shown as sticks and with carbon atoms colored in yellow.

4.3.4 Diversity in bacterial operon structure containing FN3K orthologs.

Given the diversity in bacterial FN3Ks, we next explored bacterial operon structures to investigate potential functions of FN3K in these bacterial species. An operon, which comprises a set of genes transcribed into a single mRNA, provides insight into gene regulation and function [33, 34]. We quantified the different Pfam domains or clans that are present in the same operon as FN3K orthologs. Notably, the Phosphatase clan appeared the greatest number of times in the same operon as FN3K (Fig. 4.4A) and included specific domains such as low molecular weight protein tyrosine phosphatase (LMWPTP) and Y_phosphatase. Both of these phosphatase domains act to dephosphorylate tyrosine residues. The LMWPTP domain was commonly observed in the same operon as FN3K across diverse bacterial groups (Terrabacteria, Pseudomonadota, PVC, and FCB group) alongside other domains (Fig. 4.4B). Similarly, the NADP_Rossman clan, which binds to nicotinamide adenine dinucleotide (NAD+) molecules, was the second most common clan observed (Fig. 4.4A). This clan encompassed oxidoreductase domains, including NAD/NADP-dependent adh_short and NADH-binding Pyr_redox. Additionally, it contained methyltransferase domains like Methyltransf 3/21/25. The clan was spread across several genera belonging to Pseudomonadota and Terrabacteria.

Other domains of interest included flavin binding domains such as FAD_DHS, FMN-binding, and FAD_Lum_binding. The FAD_DHS clan includes specific domains like SIR2, which is an NAD-dependent deacetylase (Fig. 4.4B). Additionally, the F420H(2)_quin_red domain, which belongs to FMN-binding, often co-occurs with adh_short and FN3K (Fig. 4.4B). Notably, auxiliary domains, such as the HTH domain (Helix-Turn-Helix) involved in DNA binding and the

2TM (2 Transmembrane) domain associated with membranes, were frequently observed (Fig. 4.4A). Although the Mce4_CUP1 domain, involved in cholesterol transport, is frequently observed, it is limited to the Mycolicibacterium genus of the Actinomycetota group and often appears in the same operon with multiple repeats. Lyase domains such as Lyase_1 and ASL_C2 were also found with the latter being specific to β-proteobacteria. The P-loop NTPase clan, frequently observed, comprises distinct domains like UvrD-helicase, AAA domains 19/30/33, and ABC_tran. Another notable clan included DHFred which consisted of DHFR_1 domain that catalyzes NADPH-dependent reduction of dihyrofolate to tetrahydrofolate. This domain is specific to the Flavobacteria group within the Bacteroidota of the FCB group. Similar to other commonly observed domains, DHFR_1 co-occurs with FN3K alongside other domains like 2TM, Acyl_transf_1, and spore_ytfj (Fig. 4.4B).

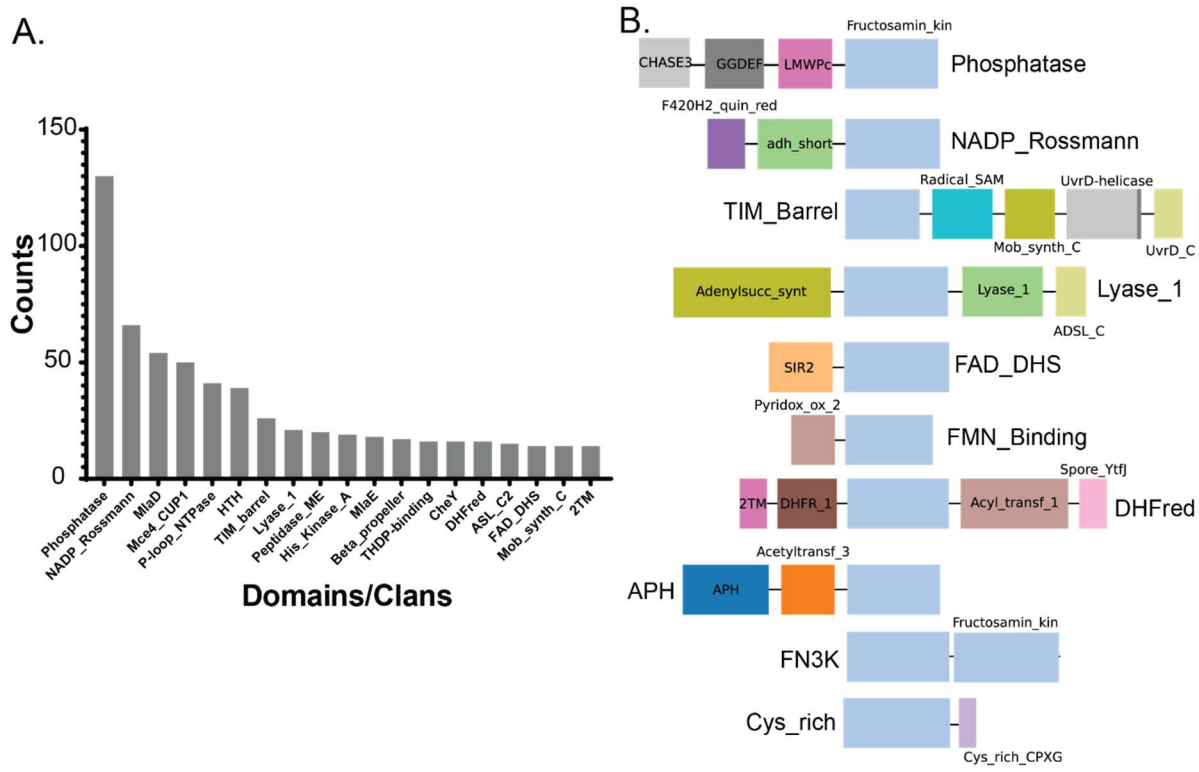


Figure 4.4: Domains found in the same operon as bacterial FN3K orthologs.

(A) Bar chart showing counts of individual Domain/Clan found in the same operon as FN3K. **(B)** Bacterial operon structures with FN3K orthologs and additional domains.

While infrequent, there were also other interesting operon structures. For example, FN3K appears in the same operon as the closely related APH domain (Fig. 4.4B), accompanied by the acetyl_transf_3 domain. In another instance, due to duplication event two FN3K domains were seen in the same operon (Fig. 4.4B).

4.3.5 AlphaFold multimer prediction of operon complexes and gene fusion.

Next, we leveraged the AlphaFold multimer prediction tool to gain insight into complexes formed by the operon structures in Fig. 4.4B. For several of these predicted complexes, the interface, which consists of the ATP and substrate binding pockets, was predicted to form interactions (Fig. 4.5A). When the operon contains multiple domains, the interface might either include just one domain, such as adh_short with FN3K, or encompass both domains, as seen with APH and acetyltransf_3 alongside FN3K (Fig. 4.5A). Using template independent AlphaFold multimer tool, we also predicted the complex for the operon structure with two FN3K domains. While both belong to FN3K domains, they only share 49.6% sequence identity (Fig. 4.5A). Interestingly, the predicted heterodimer closely resembled the crystal structure of the *A. thaliana* FN3K homodimer (Fig. 4.5B). Although it uses the same interface as in *A. thaliana* FN3K, the beta strand exchange is not predicted by the tool (Fig. 4.5B).

While the LMWPTP domain was commonly found in the same operon as FN3K, in certain instances, such as in *A. cellulolyticus*, it is gene-fused to FN3K. The AlphaFold model predicts docking of LMWPTP to the interface forming the ATP and substrate binding pockets (Fig. 4.5C). This interface is consistently predicted to form interactions with other domains in the operon as well (Fig. 4.5A). Similarly, in *C. glyciniphilum*, we observed the NUDIX domain fused to the N-

terminus of FN3K. Unlike other predicted complexes, in this case, the NUDIX domain interacts predominantly with the N-lobe beta-strands, away from the ATP and substrate binding pockets. The NUDIX domain is a pyrophosphohydrolase that hydrolyzes NUcleoside DIphosphate linked to another moiety, X (NDP-X), to form NMP and P-X (Fig. 4.5C).

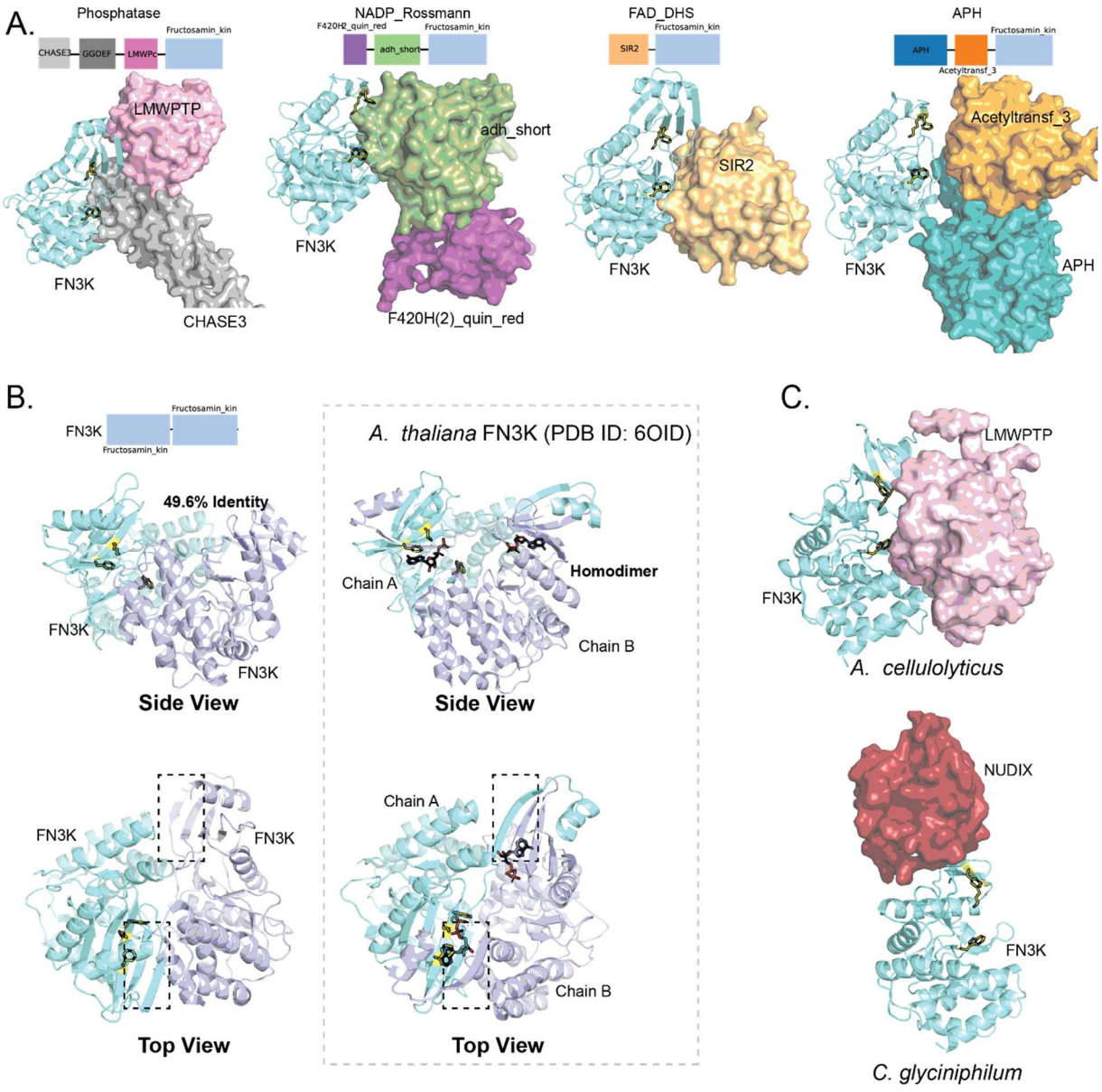


Figure 4.5: AlphaFold multimer models of FN3K complexes with other domains based on bacterial operon structures and gene fusion events.

(A) AlphaFold multimer model of FN3K in complex with other domains in the bacterial operon. **(B)** AlphaFold multimer model of FN3K heterodimer based on bacterial operon and the crystal structure of AtFN3K homodimer (PDB ID: 6OID). ADP molecule is shown as sticks with carbon atoms colored in black. **(C)** AlphaFold model of FN3K orthologs gene fused with

LMWPTP and NUDIX domain. **(A, C)** FN3K domain is shown as cartoon and colored in cyan. Additional domains are shown as surface and colored appropriately. **(A-C)** Residues in the β3 important for ATP binding (F39, K41; HsFN3K numbering) and substrate binding (W219; HsFN3K numbering) are shown as sticks with carbon atoms colored in yellow.

4.3.6 Structural characterization of FN3K orthologs provides new insight into nucleotide binding.

While our attempts to co-crystallize FN3K orthologs with ketosamine substrates did not succeed, we were able to co-crystallize FN3K orthologs from *T. fusca* (TfFN3K) and *A. thaliana* (AtFN3K) with different adenine-dinucleotide molecules. We determined the crystal structure of *T. fusca* FN3K (TfFN3K) WT in complex with AMP and a single magnesium ion at the resolution of 1.8 Å (Fig. 4.6A) and with ADP at 2.0 Å (Fig. 4.6B) using a multiple model molecular replacement strategy (See methods). Despite belonging to the same space group (P 4_3 2_1 2), the manner in which the nucleotide binds exhibit substantial differences in the adenine moeity. In the AMP-bound structure, the adenine moiety forms hydrogen bonds with the hinge region residues through the Watson-Crick edge (N1, N6 atoms), whereas in the ADP-bound structure, the adenine moiety engages in hydrogen bonding via the Hoogsteen edge (N6, N7 atoms) [35].

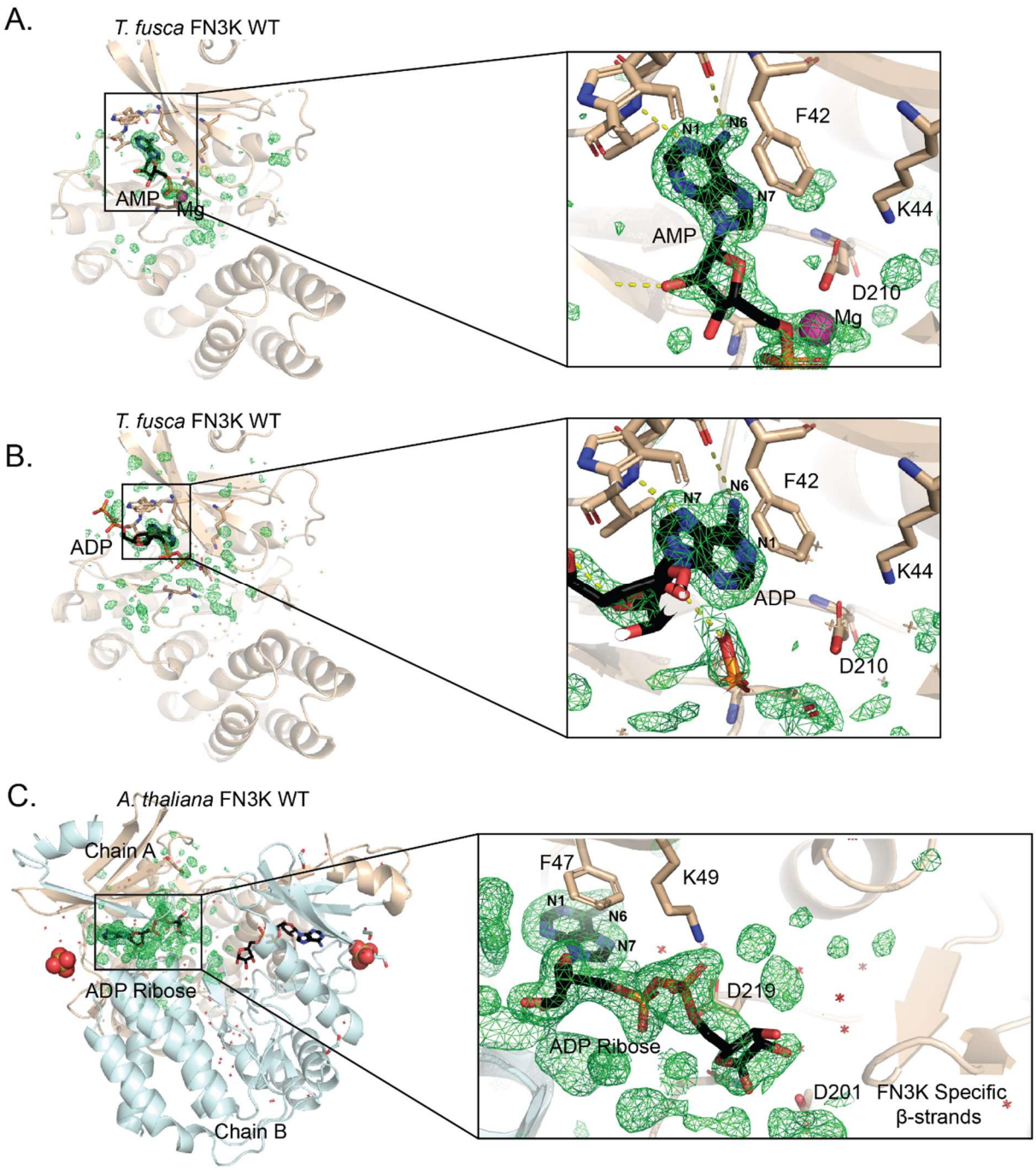


Figure 4.6: Structural characterization of FN3K orthologs reveal new insights into nucleotide binding.

(A) Cartoon representation of the crystal structure of TfFN3K bound to AMP and magnesium ion with zoomed in view. AMP is shown as sticks with carbon atoms colored in black. Magnesium ion is shown as a sphere and colored magenta. Simulated annealing map is shown as a green mesh and contoured at 3.0 σ. **(B)** Cartoon representation of the crystal structure of TfFN3K bound to ADP with zoomed in view shown. ADP is shown as sticks with carbon atoms colored in black. **(A, B)** Simulated annealing map is shown as green mesh and contoured at 3.0 σ. **(C)** Cartoon representation of the crystal structure of AtFN3K homodimer bound to ADP-ribose with zoomed in shown. ADP-ribose is shown as sticks with carbon atoms colored in black. Polder map is shown as green mesh and contoured at 3.0 σ.

In a previous study, we co-crystallized AtFN3K with ADP at a resolution of 2.4 Å [11]. In the present work, we determined the structure of AtFN3K complexed with ADP-ribose at a resolution of 1.7 Å (Fig. 4.6C). It is worth noting that the electron density for the ribose part of ADP-ribose is weak as compared to the ADP segment of ADP-ribose (Fig. 4.6C, Right). Interestingly, the ribose group is located near the FN3K-specific β-strand (Fig. 4.6C).

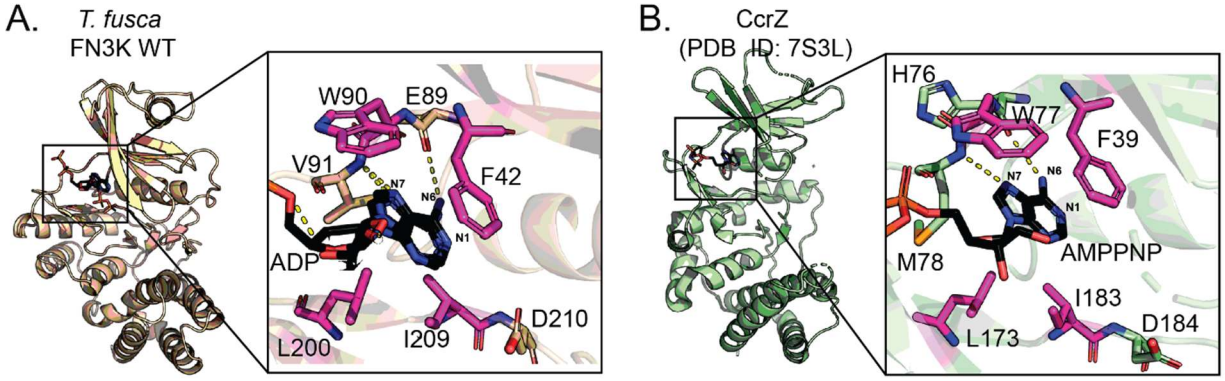


Figure 4.7: Structural motif search reveals potential shared mechanisms.

(A) Cartoon representation of the crystal structure of TfFN3K bound to ADP. (Inset) Zoomed in view of the ADP binding pocket. The residues surrounding the adenine moiety are shown as sticks with residues used for structural motif search using Fit3D are colored in magenta. **(B)** Cartoon representation of the crystal structure of CcrZ from *B. subtilis* bound to AMPPNP. (Inset) Zoomed in view of the AMPPNP binding pocket. The residues surrounding the adenine moiety are shown as sticks with residues identified by Fit3D are colored in magenta.

4.3.7 Structural motif search identifies CcrZ protein bound to AMPPNP in a similar manner to ADP bound to TfFN3K.

Given the differences in the adenine group's binding mode to TfFN3K in the two different crystal structures (Fig. 4.6A, B), we sought to identify if there were any structural motifs that were facilitating those conformations. We performed the spatial pattern search using Fit3D with query structural motif consisting of four hydrophobic residues surrounding the adenine moiety in TfFN3K bound to ADP (Fig. 4.7A, residues colored in magenta). We searched the PDB70 database using residues F42, W90, L200, and I209. Interestingly, only one significant hit was identified (p-value of 0.01), which corresponded to the crystal structure (PDB ID: 7S3L) of the

cell cycle regulator, CcrZ protein, from *B. subtilis* [36]. Notably, the CcrZ protein is complexed with the nucleotide, AMPPNP. Moreover, the matching residues from CcrZ (F39, W90, L173, and I183) formed the interaction surface for the adenine moiety (Fig. 4.7B). Like ADP bound to TfFN3K, the adenine moiety of AMPPNP in complex with CcrZ is using the Hoogsten edge (N6, N7 atoms) for hydrogen bonding with the protein.

4.4 Discussion

In this study, we performed a deep evolutionary analysis of the FN3K family and, for the first time, identified sequence constraints distinguishing the FN3K family from the related ELK families (Fig. 4.2B). Using a combination of phylogenetics and a sequence constraints-based method, we identified divergent bacterial FN3K subgroups and surveyed bacterial operon structures to gain insights into FN3K functions in these species. While the presence of the LMWPTP domain in the same operon as FN3K orthologs has previously been noted and described to function not only as a phospho-tyrosine phosphatase but also as ribulosamine-5-phosphatase [37, 38], the roles of other functional domains haven't been discussed in the context of FN3K function.

The second most prevalent domain/clan consisted of NADP_Rossman that bind NAD related molecules. In Chapter 3, we noted links between Human FN3K and NAD molecules and showed specific binding to various levels. It is possible that the interaction is evolutionary conserved and important in different species for FN3K function. Interestingly, we also find flavin adenine dinucleotide (FAD) binding domains in the same operon as FN3K orthologs. Like NAD molecules, FAD molecules also exist as redox couples in cells and act as co-factors for different proteins to be drive different reactions such as dehydrogenation, electron transfer, hydroxylation,

demethylation, and disulfide reduction [39-41]. It is conceivable that FN3Ks may interact with flavin-containing molecules, playing a role in redox reactions. Given our prior demonstrations of a redox mode of regulation within the FN3K family, this regulation might further extend through interactions of FN3Ks with these evolutionarily conserved redox-coupled molecules. Moreover, presence of aromatic residues such as tryptophan and tyrosine in the substrate binding pocket might further facilitate these redox reactions [42]. Although it's limited to mycobacteria, we also discovered FN3K in the same operon as Mce4 proteins which are involved in cholesterol transport [43]. This observation is particularly intriguing given our recent connection of FN3K functions to lipid metabolism, including cholesterol biosynthesis, as detailed in Chapter 3.

We observed not only a diversity in the operon structure associated with FN3K orthologs but also variations within the FN3K family, both at the sequence and structural levels (Fig. 4.3). Structurally, there are bacterial sub-groups that lack both β1 and β2, as well as the phosphate-binding P-loop situated between them. In one FN3K ortholog from *C. glyciniphilum* with the structural change, we identified a pyrophosphorylase NUDIX domain fused to the N-terminus (Fig 4.5C). These findings suggest that significant structural changes might play a pivotal role in forming new protein interfaces, paving the way for unique modes of interaction with other domains. At the sequence level, we find divergence in both ATP (*L. brevis*) and substrate binding pockets (*E. coli*) (Fig. 4.3). It's essential to highlight that while these FN3Ks are bacterial orthologs, the described divergences occur in bacterial species and groups found within the human gut microbiome [44]. Consequently, functional characterization of these divergent FN3Ks will not only reveal novel functions for the evolutionarily conserved FN3K family but also bear significant implications for human health.

In our study, we present evidence showcasing distinct modes of adenine binding in TfFN3K and highlight the utilization of the Hoogsten edge for adenine binding in the CcrZ protein (Fig. 4.7). CcrZ proteins are important cell cycle regulators that, while structurally related to ChoK, act on D-ribose and 2-deoxy-ribose rather than on choline. It's plausible that the use of the Hoogsten edge for adenine binding is a characteristic shared among certain ELKs, potentially influenced by sequence constraints surrounding the adenine binding pocket. Therefore, understanding nucleotide binding in the context of sequence and structural constraints will offer fresh perspectives for developing both selective and pan inhibitors that target the nucleotide binding pockets across various ELK families.

4.5 Materials and Methods

4.5.1 Seed alignment and profile-based identification of ELK sequences

Multiply-Aligned Profiles for Global Alignment of Protein Sequences (MAPGAPS) [45] tool using previously curated PKL profile was used to identify FN3Ks and related ELK sequences with crystal structures by querying the PDB resseq database. Subset of the crystal structures were used as input for the multiple structure alignment tool, DeepAlign [27]. The resulting alignment was manually curated to generate a seed alignment. The seed alignment tool was used to align the ELK sequences related to FN3Ks. ELK sequences were identified from the Uniprot reference proteome [46] using previously curated FN3K profile [17, 18].

4.5.2 Phylogenetic analysis and pattern-based sequence analysis

The ELK alignment sets for the different families were purged and Fasttree 2.1[32] was used to create a maximum likelihood tree. The tree was annotated and visualized using the Interactive Tree of Life (iTOL) [47] webserver. FN3K subgroups from the ELK alignment was identified using the optimal multiple category Bayesian Partitioning with Pattern Selection (omcBPPS) program [28]. Subgroups within the FN3K sequence set were identified using the same procedure. Fasttree was used to generate maximum likelihood tree for the FN3K family based on purged datasets for each of the subgroups.

4.5.3 Bacterial operon structure and AlphaFold Multimer Modeling

Bacterial operon structures were downloaded from the ProOpDB [48].Operon structures with FN3K in the operon were identified using Cluster of Orthologous Genes (COG) id COG3001. Protein complexes formed by the proteins in the operon were modelled using AlphaFold Multimer [29] on Colabfold [49].

4.5.4 Expression and purification of AtFN3K WT and TfFN3K WT for crystallization

AtFN3K plasmid inserted into pET-15b expression vector was provided by Prof. Emile Van Shaftingen (Université Catholique of Louvain, Brussels, Belgium). TfFN3K inserted into pSpeedET plasmid was bought from DNASU plasmid repository. The vectors were transformed into E. coli BL21(DE3) pLysS (Roche) cells and co-expression with GroEL/ES was performed at 16°C for 18 hours in 500 ml of Luria–Bertani medium containing 100 μg/l ampicillin and 50 μg/l kanamycin for AtFN3K and TfFN3K respectively. The bacterial extract was resuspended in 50 mM sodium phosphate buffer, pH 7.4, 300 mM NaCl, 15 % v/v glycerol (Buffer A) with 1mM Phenylmethylsulfonyl Fluoride (PMSF) and 1 mM 2-Mercaptoethanol (βME). The solution was

sonicated on ice and the lysate was centrifugated at 15,000 x g for 45 minutes at 4°C. The supernatant was then passed through 0.5 ml of Talon® resin previously equilibrated with Buffer A without βME. The column was washed with 20 ml of Buffer B (50 mM Sodium phosphate buffer, pH 7.4, 1 M NaCl, 15 % v/v glycerol) followed by a wash with 100 ml of Buffer B containing 10 mM imidazole. The protein was then eluted with Buffer A without βME containing 350 mM Imidazole. The purity of the protein was checked by SDS-PAGE. Fractions with the protein were subjected to Size Exclusion Chromatography.

4.5.5 Crystallization, phasing, and refinement

Approximately 100 mg of AtFN3K WT and TfFN3K WT purified using co-expression with GroEL/ES chaperone was passed through HiLoad 16/1600 Superdex 200 pg column at a flow rate of 1 ml/min with fractionation volume of 2 ml at 4 °C. 15 mM HEPES, pH 7.4, 100 mM NaCl, 5 % v/v glycerol was used as running buffer. The fractions with the protein were pooled and concentrated to 40 mg/ml and snap-frozen with liquid nitrogen to be stored at -80°C.

4.5.6 Crystallization, phasing, and refinement

AtFN3K WT was crystallized at room temperature using hanging drop vapor diffusion with 4 µl drops (2:2 protein to reservoir ratio). Crystals grew from a reservoir of 2 M ammonium sulfate, 0.1 M MES pH 5.5 after 4 weeks. Crystals were soaked in 5 mM ADP-Ribose, 10 mM $MgCl_2$ for 24 hours and were cryoprotected with the reservoir solution supplemented with 20% cryoprotectant (1:1:1 ratio of ethylene glycol, DMSO, and glycerol) and flash cooled in liquid nitrogen. TfFN3K complexed with AMP and magnesium were grown in 0.5 M Magnesium Acetate pH 8.0, 10% PEG 3350 with 5 mM ADP and 12 mM $MgCl_2$. The crystals grew after 10-

20 mins. TfFN3K complexed with ADP was grown in 0.9 M Na/K Tartarate, 0.5% PEG PEG5K with crystals growing in 3-4 days. The crystals were then soaked in 5 mM NAD+, 10 mM $MgCl_2$ for 24 hours and cryoprotected with the reservoir solution supplemented with 20% cryoprotectant (1:1:1 ratio of ethylene glycol, DMSO, and glycerol) and flash cooled in liquid nitrogen.

The diffraction data were collected at the ALS beamline and processed using XDS [50]. Five percent of the data was set aside for cross validation. The structure was solved by molecular replacement in Phenix [51] using the crystal structure of AtFN3K (PDB ID 6OID) and TfFN3K (PDB ID: 3F7W). Sequence matching and model building was performed using Phenix Autobuild [51] . Automated refinement in Phenix [51] and iterative manual fitting using Coot [52] produced the final model.

Bibliography

1. Böhm, P.G.a.M., *Advanced glycation end products. Key players in skin aging?* Dermato-Endocrinology, 2012. **4**(3): p. 259-270.
2. Maria Veiga-da-Cunha, P.J., Ghislain Delpierre, Catherine Godfraind, Ivan Théate, Didier Vertommen, Frédéric Clotman, Fréderéric Lemaigre, Olivier Devuyst and Emile Van Schaftingen, *Increased protein glycation in fructosamine 3-kinase-deficient mice.* Biochemical Journal, 2006. **399**: p. 257-264.
3. Mohas, M., et al., *A polymorphism within the fructosamine-3-kinase gene is associated with HbA1c Levels and the onset of type 2 diabetes mellitus.* Exp Clin Endocrinol Diabetes, 2010. **118**(3): p. 209-12.
4. Monnier, V.M., *The fructosamine 3-kinase knockout mouse: a tool for testing the glycation hypothesis of intracellular protein damage in diabetes and aging.* Biochemical Journal, 2006. **399**: p. e11-e13.
5. S. M. A. Pascal, M.V.-d.-C., P. Gilon, E. Van Schaftingen, and J. C. Jonas, *Effects of fructosamine-3-kinase deficiency on function and survival of mouse pancreatic islets prolonged culture in high glucose or ribose concentrations.* American Journal of Physiology - Endocrinology and Metabolism, 2010. **298**: p. E586-E596.
6. Szwergold, B.S., P.J. Howell S Fau - Beisswenger, and P.J. Beisswenger, *Human fructosamine-3-kinase: purification, sequencing, substrate specificity, and evidence of activity in vivo.* (0012-1797 (Print)).
7. Benjamin S. Szwerglod, S.H., and Paul J. Beisswenger, *Human Fructosamine-3-Kinase. Purification, Sequencing, Substrate Specificity, and Evidence of Activity In Vivo.* Diabetes, 2001. **50**: p. 2139-2147.

8. Ghislain Delpierre, M.H.R., François Collard, Vincent Stroobant, Florent Vanstapel, Helena Santos, and Emile Van Schaftingen, *Identification, Cloning, and Heterologous Expression of a Mammalian Fructosamine-3-Kinase.* Diabetes, 2000. **49**: p. 1627-1634.

9. Ngo, V.A.-O. and M.L. Duennwald, *Nrf2 and Oxidative Stress: A General Overview of Mechanisms and Implications in Human Disease. LID - 10.3390/antiox11122345 [doi] LID - 2345.* (2076-3921 (Print)).

10. Sanghvi, V.R., et al., *The Oncogenic Action of NRF2 Depends on De-glycation by Fructosamine-3-Kinase.* Cell, 2019. **178**(4): p. 807-819.e21.

11. Shrestha, S., et al., *A redox-active switch in fructosamine-3-kinases expands the regulatory repertoire of the protein kinase superfamily.* Science signaling, 2020. **13**(639).

12. Taylor, S.S., et al., *Evolution of the eukaryotic protein kinases as dynamic molecular switches.* Philos Trans R Soc Lond B Biol Sci. **367**(1602): p. 2517-28.

13. Narayana, N., et al., *A binary complex of the catalytic subunit of cAMP-dependent protein kinase and adenosine further defines conformational flexibility.* Structure, 1997. **5**(7): p. 921-935.

14. Nolen, B., S. Taylor, and G. Ghosh, *Regulation of Protein Kinases: Controlling Activity through Activation Segment Conformation.* Molecular Cell, 2004. **15**(5): p. 661-675.

15. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.

16. Tony Hunter, G.M., *The Eukaryotic Protein Kinase super-family: Emergence of Receptor Tyrosine Kinases*, in *Receptor Tyrosine Kinases: Structure, Functions and Role in Human Diseases*, Y.Y. Deric L. Wheeler, Editor. 2015, Springer New York: New York. p. 1-15.

17. Kannan, N., et al., *Structural and functional diversity of the microbial kinome.* PLoS Biol, 2007. **5**(3): p. e17.

18. Oruganty, K., et al., *Identification and classification of small molecule kinases: insights into substrate recognition and specificity.* BMC Evol Biol, 2016. **16**: p. 7.

19. Hon, W.C., et al., *Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases.* Cell, 1997. **89**(6): p. 887-95.

20. Daigle, D.M., D.W. Hughes, and G.D. Wright, *Prodigious substrate specificity of AAC(6')-APH(2"), an aminoglycoside antibiotic resistance determinant in enterococci and staphylococci.* Chem Biol, 1999. **6**(2): p. 99-110.

21. Thompson, P.R., et al., *The COOH terminus of aminoglycoside phosphotransferase (3')-IIIa is critical for antibiotic recognition and resistance.* J Biol Chem, 1999. **274**(43): p. 30697-706.

22. Fong, D.H. and A.M. Berghuis, *Substrate promiscuity of an aminoglycoside antibiotic resistance enzyme via target mimicry.* The EMBO Journal, 2002. **21**(10): p. 2323.

23. Kaplan, E., et al., *Aminoglycoside binding and catalysis specificity of aminoglycoside 2''-phosphotransferase IVa: A thermodynamic, structural and kinetic study.* Biochim Biophys Acta, 2016. **1860**(4): p. 802-13.

24. Peisach, D., et al., *The crystal structure of choline kinase reveals a eukaryotic protein kinase fold.* Structure, 2003. **11**(6): p. 703-13.

25. Hudson, C.S., et al., *Kinetic and mechanistic characterisation of Choline Kinase-alpha.* Biochim Biophys Acta, 2013. **1834**(6): p. 1107-16.

26. Ku, S.Y., et al., *Structures of 5-methylthioribose kinase reveal substrate specificity and unusual mode of nucleotide binding.* (0021-9258 (Print)).

27. Wang, S., et al., *Protein structure alignment beyond spatial proximity.* (2045-2322 (Electronic)).

28. Neuwald, A.F., *A Bayesian sampler for optimization of protein domain hierarchies.* J Comput Biol, 2014. **21**(3): p. 269-86.

29. Bryant, P.A.-O., G.A.-O. Pozzati, and A.A.-O. Elofsson, *Improved prediction of protein-protein interactions using AlphaFold2.* (2041-1723 (Electronic)).

30. Kaiser, F., et al., *Fit3D: a web application for highly accurate screening of spatial residue patterns in protein structure data.* (1367-4811 (Electronic)).

31. Scheeff, E.D. and P.E. Bourne, *Structural evolution of the protein kinase-like superfamily.* PLoS Comput Biol, 2005. **1**(5): p. e49.

32. Price, M.N., A.P. Dehal Ps Fau - Arkin, and A.P. Arkin, *FastTree 2--approximately maximum-likelihood trees for large alignments.* (1932-6203 (Electronic)).

33. Ermolaeva, M.D., S.L. White O Fau - Salzberg, and S.L. Salzberg, *Prediction of operons in microbial genomes.* (1362-4962 (Electronic)).

34. Hodgman, T.C., *A historical perspective on gene/protein functional assignment.* (1367-4803 (Print)).

35. Narunsky, A.A.-O., et al., *On the evolution of protein-adenine binding.* (1091-6490 (Electronic)).

36. Wozniak, K.A.-O., et al., *Structure and kinase activity of bacterial cell cycle regulator CcrZ.* (1553-7404 (Electronic)).

37. Rita Gemayel, J.F., Rim Rzem, Didier Vertommen, Maria Veiga-da-Cunha and Emile Van Schaftingen, *Many fructosamine 3kinase homologues in bacteria are*

ribulosamine/erythrulosamine 3-kinase potentially involved in protein deglycation. The FEBS Journal, 2007. **274**: p. 4360-4374.

38. Fortpied, J., et al., *Identification of protein-ribulosamine-5-phosphatase as human low-molecular-mass protein tyrosine phosphatase-A.* (1470-8728 (Electronic)).

39. Becker, D.F., M.A. Zhu W Fau - Moxley, and M.A. Moxley, *Flavin redox switching of protein functions.* (1557-7716 (Electronic)).

40. Massey, V., *The chemical and biological versatility of riboflavin.* (0300-5127 (Print)).

41. Forneris, F., et al., *New roles of flavoproteins in molecular cell biology: histone demethylase LSD1 and chromatin.* (1742-4658 (Electronic)).

42. Tommos, C., *Insights into the Thermodynamics and Kinetics of Amino-Acid Radicals in Proteins.* (1936-1238 (Electronic)).

43. Rank, L., L.E. Herring, and M.A.-O. Braunstein, *Evidence for the Mycobacterial Mce4 Transporter Being a Multiprotein Complex. LID - 10.1128/JB.00685-20 [doi] LID - e00685-20.* (1098-5530 (Electronic)).

44. Almeida, A.A.-O., et al., *A unified catalog of 204,938 reference genomes from the human gut microbiome.* (1546-1696 (Electronic)).

45. Neuwald, A.F., *Rapid detection, classification and accurate alignment of up to a million or more related protein sequences.* Bioinformatics, 2009. **25**(15): p. 1869-75.

46. UniProt Consortium, T., *UniProt: the universal protein knowledgebase.* Nucleic Acids Research, 2018. **46**(5): p. 2699-2699.

47. Letunic, I.A.-O. and P. Bork, *Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees.* (1362-4962 (Electronic)).

48. Taboada, B., et al., *ProOpDB: Prokaryotic Operon DataBase.* (1362-4962 (Electronic)).

49. Mirdita, M.A.-O., et al., *ColabFold: making protein folding accessible to all.* (1548-7105 (Electronic)).

50. Kabsch, W., *XDS.* Acta Crystallographica Section D, 2010. **66**(2): p. 125-132.

51. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallographica Section D, 2010. **66**(2): p. 213-221.

52. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

CHAPTER 5

DISCUSSION AND CONCLUDING REMARKS

5.1 Achievement of goals

Using a combination of experimental and computational techniques, I have addressed the questions in the first chapter by making major contributions towards our understanding of the evolutionarily conserved yet understudied FN3K family. Unlike the highly regulated EPKs, ELKs were believed to be single domain and constitutively active protein families. By solving the crystal structure of AtFN3K, we showed the structural basis for a novel mode of redox regulation. We further showed that redox regulation is conserved in humans and identified metabolic and transcriptional changes when the human enzyme (HsFN3K) is knocked out. By leveraging the CRISPR knockout of HsFN3K in HepG2 cells, our multi-omics study provided new links for FN3Ks to lipid and cofactor metabolism pathways. Based on enrichment of NAD(P) binding proteins, we showed specific binding of HsFN3K to the different NAD molecules and provide predictions for the binding mode using molecular docking and mutagenesis. In the final work, we performed deep evolutionary analysis and identified sequence constraints distinctive of the FN3K family as compared to the related ELK members. We computationally identified bacterial subgroups that are divergent within the nucleotide and substrate binding pockets. Using recent advances in machine learning, we leveraged AlphaFold multimer tool to predict complexes formed by bacterial FN3Ks with other proteins in the operon. Finally, we structurally characterize

nucleotide binding in FN3K orthologs and show for the first time the ability to bind adenine in two distinct modes.

5.1.1 Structural basis for the novel mode of redox mode in FN3Ks

Unlike EPKs, ELKs were believed to be single domain and constitutively active protein families. However, by solving the crystal structure of AtFN3K homodimer, we not only identified a novel mode of redox regulation within the FN3K family but also, for the first time, demonstrated oligomerization within the family. "Using biochemical assays and mutagenesis, we identified that the P-loop cysteine is critical for forming the disulfide-linked dimer.. By following the conservation of the cysteine within FN3K sequences, we identified the equivalent cysteine in HsFN3K and through further *in vitro* and *in vivo* work, we found more evidence that supported the conservation of redox mode regulation in humans as well.

5.1.2 New links between FN3Ks with lipid and co-factor metabolism through its interactions with NAD related molecules.

While the role of FN3Ks in deglycation is well characterized, the role of FN3Ks in a larger pathway context lacked clear understanding. Considering identification of redox regulation in the enzyme family, the relevance for the enzyme to be regulated in such manner was not understood. By leveraging CRISPR KO of HsFN3K in HepG2 cells and performing an integrative multi-omics approach, we identified new links of FN3Ks to various pathways including lipid metabolism, co-factor metabolism, and oxidative stress response. Given the enrichment of NAD(P) binding proteins and nicotinamide metabolism observed in our integrative networks, we hypothesized and tested binding of HsFN3K to different metabolites including NAD molecules. Through

biochemical assays, we demonstrated that HsFN3K binds specifically, albeit variably, to different NAD compounds.

5.1.3 Distinct mode of nucleotide binding

By building on the previous work that derived evolutionary relationships between the PKL members [1, 2], I defined and extended the core shared by FN3K with related ELKs. Through deep evolutionary analysis using phylogenetics and pattern-based classification on the extended core sequence alignment, I defined sequence constraints distinctive of the FN3K family. Notably, these constraints were located within the nucleotide and substrate binding pockets. Using a similar approach, I looked deeper into the FN3K family and identified several divergent sub-groups of bacterial FN3Ks that require further functional characterization. I also co-crystallized bacterial FN3Ks, TfFN3K and AtFN3K, with adenine-containing nucleotides. This provided new insights into how FN3K-specific sequence constraints can facilitate a unique mode of nucleotide binding.

5.2 Future directions

The findings and analyses presented here not only provide new insights into the FN3K family but also opportunities and hypotheses for future investigations which are highlighted below.

5.2.1 Trace the evolution of dimerization in FN3Ks.

Both the crystal structure and solution studies of AtFN3K provided evidence for dimerization in the FN3K family. Previously solved crystal structures of FN3K orthologs from *Thermobifida fusca* (PDB ID: 3F7W) and *Haemophilus somnus* (PDB ID: 3JR1) are both predicted to be monomers in solution by PISA [3]. Solution studies on the AtFN3K triple cysteine mutant

(C32A/C236A/196A) which includes the two cysteines (C32 and C236) forming the disulfides showed presence of both the dimer and monomer peaks suggesting the presence of a dimer interface independent of the cysteine residues. Based on mutagenesis, we identified the importance of the P-loop cysteine (C32 in AtFN3K) for disulfide linked dimer formation. However, the underlying sequence basis for dimerization itself is not clearly understood.

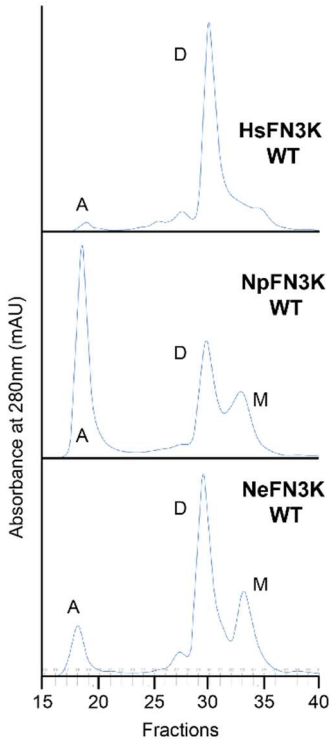


Figure 5.1: Size Exclusion Chromatography (SEC) profiles of FN3K orthologs.

SEC was performed using S200 10/300 GL column at a flow rate of 0.5 ml/ml with 0.5 ml fractions collected. A: Aggregates; D: Dimer; M: Monomer. Dimer and monomer peaks were identified by comparing with the SEC profile of ladder proteins.

Given the importance of P-loop cysteine for disulfide linked dimer formation, I have gathered preliminary data on FN3K orthologs that conserve the P-loop cysteine from different domains of life. Using SEC, I looked at the oligomeric states of different FN3K orthologs in solution. We first performed SEC on HsFN3K WT (Fig. 5.1). It showed that HsFN3K is

predominantly a dimer in solution, unlike AtFN3K WT that was comprised of both dimeric and monomeric species. Next, we performed SEC on two bacterial FN3K orthologs from proteobacteria *Nitrosomonas europaea* (NeFN3K) and cyanobacteria *Nostoc punctiforme* (NpFN3K) as both conserve the P-loop cysteine as well. Like AtFN3K, both bacterial FN3Ks are present as both dimeric and monomeric forms (Fig. 5.1).

Based on the deep evolutionary analysis of the FN3K family and the operon structures, we also find evidence for two copies of FN3Ks in the same bacterial genome (Fig. 4.5B). Modeling the operon structure suggested dimerization mode like AtFN3K crystal structure. The results suggest prevalence of oligomerization within the FN3K family and understanding the structural and sequence basis of dimerization will provide insights into FN3K function and regulation in these different species.

5.2.2 Trace the loss of β3 lysine in FN3K orthologs.

Pattern based Bayesian classification of the FN3K sequences revealed a subset of Firmicutes FN3K orthologs including the ortholog from *L. casei* (LcFN3K) that diverged in the ATP binding pocket and conserved a leucine residue (L42 in LcFN3K) in place of the invariant positively charged β3 lysine (K72 in PKA, PDB ID: 1ATP) found in EPKs (Fig. 5.2A). In EPKs, the β3 lysine coordinates the α and β phosphates of the ATP molecule and has been shown to be important for ATP positioning, stabilization of the active conformation, and for catalytic mechanism [4, 5]. Bioinformatically, sequences missing the β3 lysine (K72, PKA numbering) are classified as a pseudo-kinase [6-8] and are predicted to have lost their catalytic ability. The loss of the β3 lysine is not new and has previously been seen in multiple human EPKs including WNK

kinases [9], where the lysine is replaced by a cysteine residue. Despite the loss of the lysine, human WNK1 was shown to be active. Structural studies later showed that another lysine (K233) emanating from β2 was acting as the catalytic lysine [10]. Another study found the role of WNK1 kinases in chloride ion sensing through direct binding and suggested that repositioning of the catalytic lysine enabled the chloride sensing function [11].

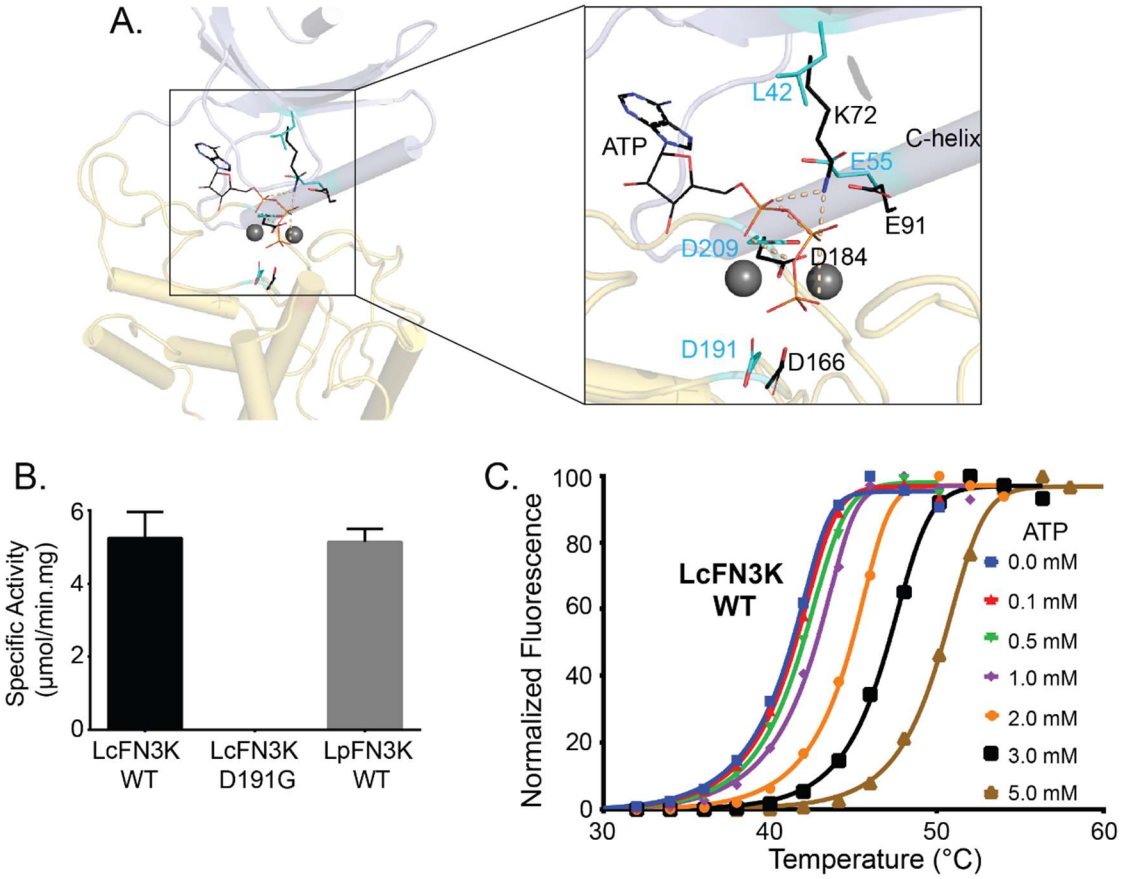


Figure 5.2: Biochemical characterization of pseudo-kinase LcFN3K.

(A) Cartoon representation of LcFN3K homology model built using SWISS-MODEL47 with PDB ID: 3F7W as template. The key catalytic residues are shown in sticks and carbon atoms colored in cyan. The model was aligned with PKA structure (PDB: 1ATP) using PyMOL 2.3.2. Only the catalytic residues, ATP, and divalent ions from PKA structure are shown. The catalytic residues and ATP molecules are shown as lines with carbon atoms colored in black. Magnesium and manganese ions are shown as grey spheres. **(B)** Comparison of the specific activity of LcFN3K to closely related canonical LpFN3K. Data are means ± SE of at least three independent experiments. **(C)** Thermal shift showing concentration dependent shift in thermal stability with increasing ATP concentration. The concentration of MgCl2 were 3 times the concentration of ATP.

Preliminary characterization of the pseudo kinase LcFN3K ortholog showed both activity towards the ketosamine substrate (Fig. 5.2B) as well as an ability to bind ATP (Fig. 5.2C). Pattern based classification further identified sequence constraints distinctive of the subgroups consisting of LcFN3K with K266 protruding from α8 as one of the co-conserved residues (Fig. 5.3A). Interestingly, phylogenetic tree consisting of only FN3K orthologs from Firmicutes suggested multiple instances of loss of the β3 lysine (Fig. 5.3B). The characterization of the pseudo kinase ortholog will provide insights into adaptations within the FN3K family for novel functions since species including *L. casei* belonging to Firmicutes are prevalent in the human gut microbiome.

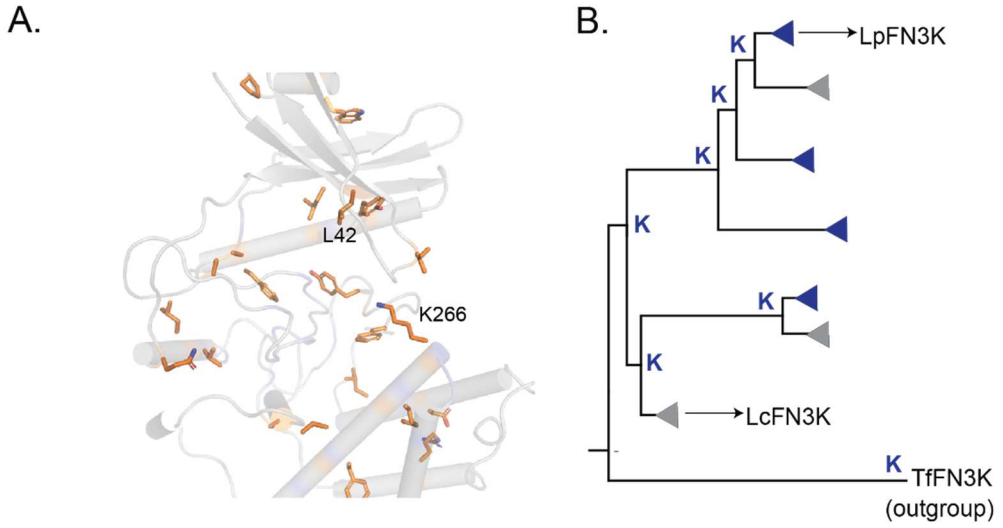


Figure 5.3: Sequence analyses of FN3K orthologs lacking β3 lysine.

(A) Sequence constraints on the clusters conserving a Leucine at the β3 lysine position (includes LcFN3K). The cluster specific constraints were identified using the Bayesian pattern-based classification method. **(B)** Ancestral sequence reconstruction on FN3K sequences from firmicutes. The ancestral amino acid at β3 lysine positions (L42 in LcFN3K) are shown. TfFN3K was used as an outgroup. Ancestral sequences were generated using MEGA X Version 10.1.8. Collapsed clades as represented by triangles. Blue and grey triangles represent a lysine and a leucine at the β3 position, respectively.

5.2.3 Probe binding of different nucleotides to FN3K orthologs

Based on the multi-omics study in Chapter 3, we identified links between HsFN3K and NAD molecules. We further showed the ability of HsFN3K to bind NAD related compounds. In

Chapter 4, through structural characterization we showed the ability of TfFN3K to bind adenine moiety in two distinct modes whereas we saw weak density for binding of ADP-ribose to AtFN3K. Taken together, the data suggests FN3Ks ability to bind various nucleotide molecules.

Preliminary binding data for the different orthologs using DSF assays suggest differences in binding to NAD related compounds by different FN3K orthologs (Fig. 5.4). Notably, pseudo kinase FN3K ortholog, LcFN3K showed higher changes in melting temperature for NADP+ unlike AtFN3K and TfFN3K. However, reverting the β3 Lysine (L42K) increased stability with ADP and NADH in metal dependent manner (Fig. 5.4).

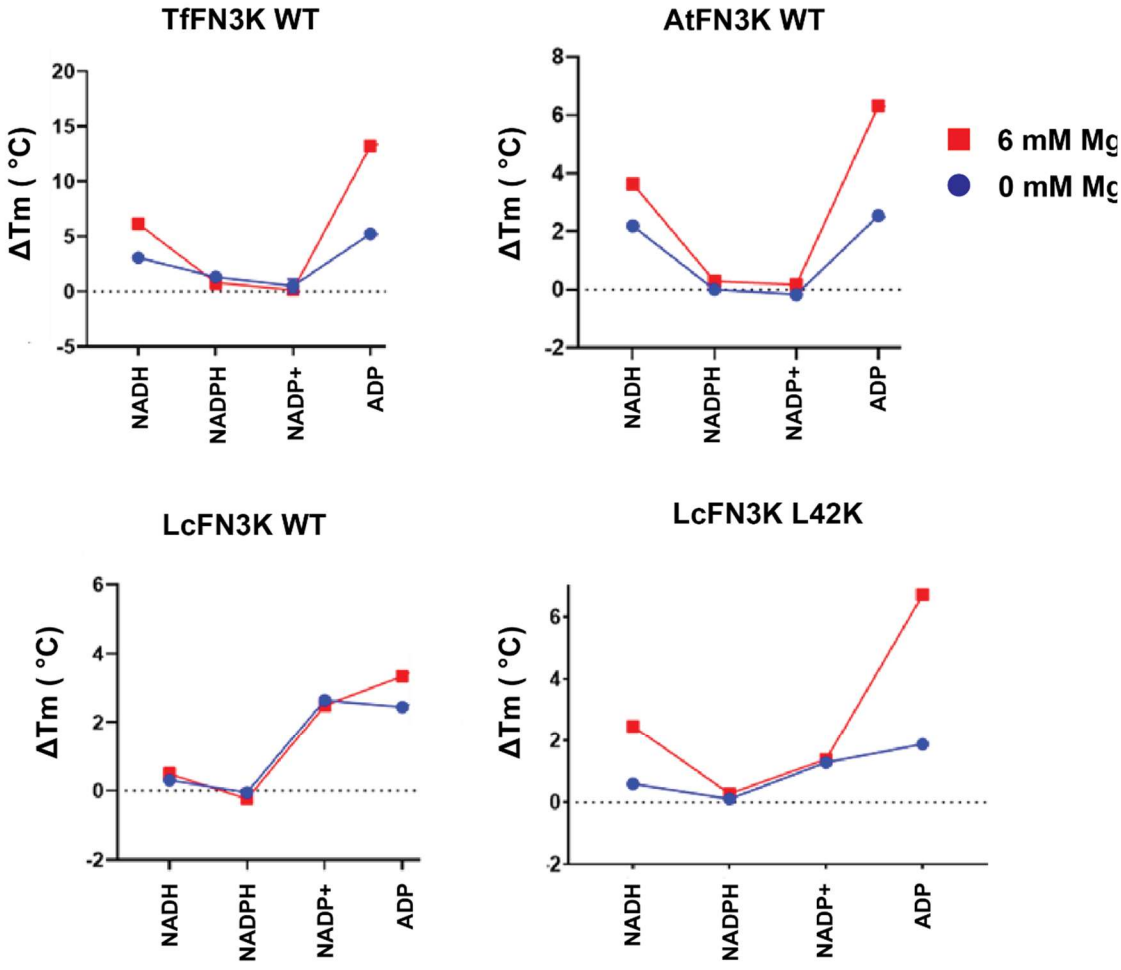


Figure 5.4: FN3K orthologs binding to NAD related molecules.

Thermal shift showing changes in melting temperature (Tm) for different NAD compounds and ADP in the presence (red square) and absence (blue circle) of 6 mM $MgCl_2$. Tm changes are normalized to apo form. 5 mM of NAD compounds and ADP were used.

Bibliography

1. Oruganty, K., et al., *Identification and classification of small molecule kinases: insights into substrate recognition and specificity.* BMC Evol Biol, 2016. **16**: p. 7.
2. Kannan, N., et al., *Structural and functional diversity of the microbial kinome.* PLoS Biol, 2007. **5**(3): p. e17.
3. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.
4. Iyer, G.H., et al., *Catalytic independent functions of a protein kinase as revealed by a kinase-dead mutant: study of the Lys72His mutant of cAMP-dependent kinase.* J Mol Biol, 2005. **351**(5): p. 1110-22.
5. Meharena, H.S., et al., *Decoding the Interactions Regulating the Active State Mechanics of Eukaryotic Protein Kinases.* PLoS Biol, 2016. **14**(11): p. e2000127.
6. Kwon, A., et al., *Tracing the origin and evolution of pseudokinases across the tree of life.* Science Signaling, 2019. **12**(578): p. eaav3810.
7. Eyers, P.A., K. Keeshan, and N. Kannan, *Tribbles in the 21st Century: The Evolving Roles of Tribbles Pseudokinases in Biology and Disease.* Trends Cell Biol, 2017. **27**(4): p. 284-298.
8. Kornev, A.P. and S.S. Taylor, *Pseudokinases: functional insights gleaned from structure.* Structure, 2009. **17**(1): p. 5-7.
9. Vitari, A.C., et al., *The WNK1 and WNK4 protein kinases that are mutated in Gordon's hypertension syndrome phosphorylate and activate SPAK and OSR1 protein kinases.* Biochem J, 2005. **391**(Pt 1): p. 17-24.

10. Min, X., et al., *Crystal structure of the kinase domain of WNK1, a kinase that causes a hereditary form of hypertension.* (0969-2126 (Print)).

11. Piala, A.T., et al., *Chloride sensing by WNK1 involves inhibition of autophosphorylation.* (1937-9145 (Electronic)).

APPENDIX A

EXTENDED RESULTS

Addition to the results in the dissertation chapters 2-4, I have contributed to several collaborative projects during my Ph.D. that have resulted in multiple first and co-authors publications. The abstracts for the publications are listed below.

Cataloguing the dead: breathing new life into pseudokinase research

Pseudoenzymes are present within many, but not all, known enzyme families and lack one or more conserved canonical amino acids that help define their catalytically active counterparts. Recent findings in the pseudokinase field confirm that evolutionary repurposing of the structurally defined bilobal protein kinase fold permits distinct biological functions to emerge, many of which rely on conformational switching, as opposed to canonical catalysis. In this analysis, we evaluate progress in evaluating several members of the 'dark' pseudokinome that are pertinent to help drive this expanding field. Initially, we discuss how adaptions in erythropoietin-producing hepatocellular carcinoma (Eph) receptor tyrosine kinase domains resulted in two vertebrate pseudokinases, EphA10 and EphB6, in which co-evolving sequences generate new motifs that are likely to be important for both nucleotide binding and catalysis-independent signalling. Secondly, we discuss how conformationally flexible Tribbles pseudokinases, which have radiated in the complex vertebrates, control fundamental aspects of cell signalling that may be targetable with covalent small molecules. Finally, we show how species-level adaptations in the duplicated canonical kinase protein serine kinase histone (PSKH)1 sequence have led to the appearance of the pseudokinase PSKH2, whose physiological role remains mysterious. In conclusion, we show how the patterns we discover are selectively conserved within specific pseudokinases, and that when they are modelled alongside closely related canonical kinases, many are found to be located in functionally important regions of the conserved kinase fold. Interrogation of these patterns will be useful for future evaluation of these, and other, members of the unstudied human kinome.

Shrestha S, Byrne DP, Harris JA, Kannan N, Eyers PA. Cataloguing the dead: breathing new life into pseudokinase research. FEBS J. 2020 Oct;287(19):4150-4169. doi: 10.1111/febs.15246. Epub 2020 Mar 10. PMID: 32053275; PMCID: PMC7586955.

Evolutionary and cellular analysis of the 'dark' pseudokinase PSKH2

Pseudokinases, so named because they lack one or more conserved canonical amino acids that define their catalytically active relatives, have evolved a variety of biological functions in both prokaryotic and eukaryotic organisms. Human PSKH2 is closely related to the canonical kinase PSKH1, which maps to the CAMK family of protein kinases. Primates encode PSKH2 in the form of a pseudokinase, which is predicted to be catalytically inactive due to loss of the invariant catalytic Asp residue. Although the biological role(s) of vertebrate PSKH2 proteins remains unclear, we previously identified species-level adaptions in PSKH2 that have led to the appearance of kinase or pseudokinase variants in vertebrate genomes alongside a canonical PSKH1 paralog. In this paper we confirm that, as predicted, PSKH2 lacks detectable protein phosphotransferase activity, and exploit structural informatics, biochemistry and cellular proteomics to begin to characterise vertebrate PSKH2 orthologues. AlphaFold 2-based structural analysis predicts functional roles for both the PSKH2 N- and C-regions that flank the pseudokinase domain core, and cellular truncation analysis confirms that the N-terminal domain, which contains a conserved myristoylation site, is required for both stable human PSKH2 expression and localisation to a membrane-rich subcellular fraction containing mitochondrial proteins. Using mass spectrometry-based proteomics, we confirm that human PSKH2 is part of a cellular mitochondrial protein network, and that its expression is regulated through client-status within the HSP90/Cdc37 molecular chaperone system. HSP90 interactions are mediated through binding to the PSKH2 C-terminal tail, leading us to predict that this region might act as both a cis and trans regulatory

Byrne DP*, Shrestha S*, Daly LA, Marensi V, Ramakrishnan K, Eyers CE, Kannan N, Eyers PA. Evolutionary and cellular analysis of the 'dark' pseudokinase PSKH2. Biochem J. 2023 Jan 31;480(2):141-160. doi: 10.1042/BCJ20220474. PMID: 36520605; PMCID: PMC9988210.

element, driving outputs linked to the PSKH2 pseudokinase domain that are important for functional signalling.

Aurora A regulation by reversible cysteine oxidation reveals evolutionarily conserved redox control of Ser/Thr protein kinase activity

Reactive oxygen species (ROS) are physiological mediators of cellular signaling and play potentially damaging roles in human diseases. In this study, we found that the catalytic activity of the Ser/Thr kinase Aurora A was inhibited by the oxidation of a conserved cysteine residue (Cys290) that lies adjacent to Thr288, a critical phosphorylation site in the activation segment. Cys is present at the equivalent position in ~100 human Ser/Thr kinases, a residue that we found was important not only for the activity of human Aurora A but also for that of fission yeast MAPK-activated kinase (Srk1) and PKA (Pka1). Moreover, the presence of this conserved Cys predicted biochemical redox sensitivity among a cohort of human CAMK, AGC, and AGC-like kinases. Thus, we predict that redox modulation of the conserved Cys290 of Aurora A may be an underappreciated regulatory mechanism that is widespread in eukaryotic Ser/Thr kinases. Given the key biological roles of these enzymes, these findings have implications for understanding physiological and pathological responses to ROS and highlight the importance of protein kinase regulation through multivalent modification of the activation segment.

Byrne DP, **Shrestha S**, Galler M, Cao M, Daly LA, Campbell AE, Eyers CE, Veal EA, Kannan N, Eyers PA. Aurora A regulation by reversible cysteine oxidation reveals evolutionarily conserved redox control of Ser/Thr protein kinase activity. Sci Signal. 2020 Jul 7;13(639):eaax2713. doi: 10.1126/scisignal.aax2713. PMID: 32636306.

Covalent inhibitors of EGFR family protein kinases induce degradation of human Tribbles 2 (TRIB2) pseudokinase in cancer cells

A major challenge associated with biochemical and cellular analysis of pseudokinases is a lack of target-validated small-molecule compounds with which to probe function. Tribbles 2 (TRIB2) is a cancer-associated pseudokinase with a diverse interactome, including the canonical AKT signaling module. There is substantial evidence that human TRIB2 promotes survival and drug resistance in solid tumors and blood cancers and therefore is of interest as a therapeutic target. The unusual TRIB2 pseudokinase domain contains a unique cysteine-rich C-helix and interacts with a conserved peptide motif in its own carboxyl-terminal tail, which also supports its interaction with E3 ubiquitin ligases. We found that TRIB2 is a target of previously described small-molecule protein kinase inhibitors, which were originally designed to inhibit the canonical kinase domains of epidermal growth factor receptor tyrosine kinase family members. Using a thermal shift assay, we discovered TRIB2-binding compounds within the Published Kinase Inhibitor Set (PKIS) and used a drug repurposing approach to classify compounds that either stabilized or destabilized TRIB2 in vitro. TRIB2 destabilizing agents, including the covalent drug afatinib, led to rapid TRIB2 degradation in human AML cancer cells, eliciting tractable effects on signaling and survival. Our data reveal new drug leads for the development of TRIB2-degrading compounds, which will also be invaluable for unraveling the cellular mechanisms of TRIB2-based signaling. Our study highlights that small molecule–induced protein down-regulation through drug "off-targets" might be relevant for other inhibitors that serendipitously target pseudokinases.

Foulkes DM, Byrne DP, Yeung W, **Shrestha S**, Bailey FP, Ferries S, Eyers CE, Keeshan K, Wells C, Drewry DH, Zuercher WJ, Kannan N, Eyers PA. Covalent inhibitors of EGFR family protein kinases induce degradation of human Tribbles 2 (TRIB2) pseudokinase in cancer cells. Sci Signal. 2018 Sep 25;11(549):eaat7951. doi: 10.1126/scisignal.aat7951. PMID: 30254057; PMCID: PMC6553640.

Nucleotide Binding, Evolutionary Insights, and Interaction Partners of the Pseudokinase Unc-51-like Kinase 4

Unc-51-like kinase 4 (ULK4) is a pseudokinase that has been linked to the development of several diseases. Even though sequence motifs required for ATP binding in kinases are lacking, ULK4 still tightly binds ATP and the presence of the co-factor is required for structural stability of ULK4. Here, we present a high-resolution structure of a ULK4-ATPγS complex revealing a highly unusual ATP binding mode in which the lack of the canonical VAIK motif lysine is compensated by K39, located N-terminal to αC. Evolutionary analysis suggests that degradation of active site motifs in metazoan ULK4 has co-occurred with an ULK4-specific activation loop, which stabilizes the C helix. In addition, cellular interaction studies using BioID and biochemical validation data revealed high confidence interactors of the pseudokinase and armadillo repeat domains. Many of the identified ULK4 interaction partners were centrosomal and tubulin-associated proteins and several active kinases suggesting interesting regulatory roles for ULK4.

Preuss F, Chatterjee D, Mathea S, **Shrestha S**, St-Germain J, Saha M, Kannan N, Raught B, Rottapel R, Knapp S. Nucleotide Binding, Evolutionary Insights, and Interaction Partners of the Pseudokinase Unc-51-like Kinase 4. Structure. 2020 Nov 3;28(11):1184-1196.e6. doi: 10.1016/j.str.2020.07.016. Epub 2020 Aug 18. PMID: 32814032.

A Chemical Probe for Dark Kinase STK17B Derives Its Potency and High Selectivity through a Unique P-Loop Conformation

STK17B is a member of the death-associated protein kinase family and has been genetically linked to the development of diverse diseases. However, the role of STK17B in normal and disease pathology is poorly defined. Here, we present the discovery of thieno[3,2-d] pyrimidine SGC-STK17B-1 (11s), a high-quality chemical probe for this understudied "dark" kinase. 11s is an ATP-competitive inhibitor that showed remarkable selectivity over other kinases including the closely related STK17A. X-ray crystallography of 11s and related thieno[3,2-d]pyrimidines bound to STK17B revealed a unique P-loop conformation characterized by a salt bridge between R41 and the carboxylic acid of the inhibitor. Molecular dynamic simulations of STK17B revealed the flexibility of the P-loop and a wide range of R41 conformations available to the apo-protein. The isomeric thieno[2,3-d]pyrimidine SGC-STK17B-1N (19g) was identified as a negative control compound. The >100-fold lower activity of 19g on STK17B was attributed to the reduced basicity of its pyrimidine N1.

Picado A, Chaikuad A, Wells CI, **Shrestha S**, Zuercher WJ, Pickett JE, Kwarcinski FE, Sinha P, de Silva CS, Zutshi R, Liu S, Kannan N, Knapp S, Drewry DH, Willson TM. A Chemical Probe for Dark Kinase STK17B Derives Its Potency and High Selectivity through a Unique P-Loop Conformation. J Med Chem. 2020 Dec 10;63(23):14626-14646. doi: 10.1021/acs.jmedchem.0c01174. Epub 2020 Nov 20. PMID: 33215924; PMCID: PMC7816213.

Granulovirus PK-1 kinase activity relies on a side-to-side dimerization mode centered on the regulatory αC helix

The life cycle of Baculoviridae family insect viruses depends on the viral protein kinase, PK-1, to phosphorylate the regulatory protein, p6.9, to induce baculoviral genome release. Here, we report the crystal structure of Cydia pomenella granulovirus PK-1, which, owing to its likely ancestral origin among host cell AGC kinases, exhibits a eukaryotic protein kinase fold. PK-1 occurs as a rigid dimer, where an antiparallel arrangement of the αC helices at the dimer core stabilizes PK-1 in a closed, active conformation. Dimerization is facilitated by C-lobe:C-lobe and N-lobe:N-lobe interactions between protomers, including the domain-swapping of an N-terminal helix that crowns a contiguous β-sheet formed by the two N-lobes. PK-1 retains a dimeric conformation in solution, which is crucial for catalytic activity. Our studies raise the prospect that parallel, side-to-side dimeric arrangements that lock kinase domains in a catalytically-active conformation could function more broadly as a regulatory mechanism among eukaryotic protein kinases.

Oliver MR, Horne CR, **Shrestha S**, Keown JR, Liang LY, Young SN, Sandow JJ, Webb AI, Goldstone DC, Lucet IS, Kannan N, Metcalf P, Murphy JM. Granulovirus PK-1 kinase activity relies on a side-to-side dimerization mode centered on the regulatory αC helix. Nat Commun. 2021 Feb 12;12(1):1002. doi: 10.1038/s41467-021-21191-7. PMID: 33579933; PMCID: PMC7881018.

Computational tools and resources for pseudokinase research

Pseudokinases regulate diverse cellular processes associated with normal cellular functions and disease. They are defined bioinformatically based on the absence of one or more catalytic residues that are required for canonical protein kinase functions. The ability to define pseudokinases based on primary sequence comparison has enabled the systematic mapping and cataloging of pseudokinase orthologs across the tree of life. While these sequences contain critical information regarding pseudokinase evolution and functional specialization, extracting this information and generating testable hypotheses based on integrative mining of sequence and structural data requires specialized computational tools and resources. In this chapter, we review recent advances in the development and application of open-source tools and resources for pseudokinase research. Specifically, we describe the application of an interactive data analytics framework, KinView, for visualizing the patterns of conservation and variation in the catalytic domain motifs of pseudokinases and evolutionarily related canonical kinases using a consistent set of curated alignments organized based on the widely used kinome evolutionary hierarchy. We also demonstrate the application of an integrated Protein Kinase Ontology (ProKinO) and an interactive viewer, ProtVista, for mapping and analyzing primary sequence motifs and annotations in the context of 3D structures and AlphaFold2 models. We provide examples and protocols for generating testable hypotheses on pseudokinase functions both for bench biologists and advanced users.

O'Boyle B, **Shrestha S**, Kochut K, Eyers PA, Kannan N. Computational tools and resources for pseudokinase research. Methods Enzymol. 2022;667:403-426. doi: 10.1016/bs.mie.2022.03.040. Epub 2022 Apr 8. PMID: 35525549; PMCID: PMC9733567.

Mechanistic and evolutionary insights into isoform-specific 'supercharging' in DCLK family kinases

Catalytic signaling outputs of protein kinases are dynamically regulated by an array of structural mechanisms, including allosteric interactions mediated by intrinsically disordered segments flanking the conserved catalytic domain. The Doublecortin Like Kinases (DCLKs) are a family of microtubule-associated proteins characterized by a flexible C-terminal autoregulatory 'tail' segment that varies in length across the various human DCLK isoforms. However, the mechanism whereby these isoform-specific variations contribute to unique modes of autoregulation is not well understood. Here, we employ a combination of statistical sequence analysis, molecular dynamics simulations and in vitro mutational analysis to define hallmarks of DCLK family evolutionary divergence, including analysis of splice variants within the DCLK1 sub-family, which arise through alternative codon usage and serve to 'supercharge' the inhibitory potential of the DCLK1 C-tail. We identify co-conserved motifs that readily distinguish DCLKs from all other Calcium Calmodulin Kinases (CAMKs), and a 'Swiss-army' assembly of distinct motifs that tether the C-terminal tail to conserved ATP and substrate-binding regions of the catalytic domain to generate a scaffold for auto-regulation through C-tail dynamics. Consistently, deletions and mutations that alter C-terminal tail length or interfere with co-conserved interactions within the catalytic domain alter intrinsic protein stability, nucleotide/inhibitor-binding, and catalytic activity, suggesting isoform-specific regulation of activity through alternative splicing. Our studies provide a detailed framework for investigating kinome–wide regulation of catalytic output through cis-regulatory

Venkat A, Watterson G, Byrne DP, O'Boyle B, **Shrestha S**, Gravel N, Fairweather EE, Daly LA, Bunn C, Yeung W, Aggarwal I, Katiyar S, Eyers CE, Eyers PA, Kannan N. Mechanistic and evolutionary insights into isoform-specific 'supercharging' in DCLK family kinases. eLife. 2023;12:RP87958. doi: 10.7554/eLife.87958.2

events mediated by intrinsically disordered segments, opening new avenues for the design of mechanistically-divergent DCLK1 modulators, stabilizers, or degraders.

APPENDIX B

SUPPLEMENTARY INFORMATION

In addition to the figures and data presented here in the main dissertation, there are some additional supplementary figures and tables for the published Science signaling paper that can be accessed on the website here:

https://www.science.org/doi/10.1126/scisignal.aax6313