UNIVERSITY OF CALIFORNIA, SAN DIEGO

The Native Landscape of the Cytokine Interleukin-33: Exploring the Link Between Folding and Dynamics

A dissertation submitted in partial satisfaction of the requirements of for the degree of Doctor of Philosophy

in

Chemistry

by

Kaitlin M. Fisher

Committee in charge:

Professor Patricia Jennings, Chair
Professor Timothy Bigby
Professor Elizabeth Komives
Professor Tadeuz Molinski
Professor Jose N. Onuchic
Professor Susan Taylor

2015

The Dissertation of Kaitlin M. Fisher is approved and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

Dedication

For my Mom and Dad, the best cheerleaders in the whole world

Epigraph

“I knew that if I allowed fear to overtake me, my journey was doomed. Fear, to a great extent, is born of a story we tell ourselves, and so I chose to tell myself a different story from the one women are told. I decided I was safe. I was strong. I was brave. Nothing could vanquish me.”

-Cheryl Strayed

Wild

Table of Contents

List of Abbreviations

ANS	1-Anilinonaphthalene-8-Sulfonic Acid
ANTS	8-Aminonapthalene-1,3,6-Trisulfonic Acid
DPX	*p*-Xylene-Bis-Pyridinium Bromide
FPLC	Fast Protein Liquid Chromatography
Gdm-HCl	Guanidinium Chloride
HMQC	Heteronuclear Multiple Quantum Coherence
HSQC	Heteronuclear Single Quantum Coherence
IL-1β	Interleukin-1 Beta
IL-33	Interleukin-33
LB	Luria Broth
NMR	Nuclear Magnetic Resonance
PF	Protection Factor
SASA	Solvent Accessible Surface Area
τ_c	Tumbling Correlation Time

List of Figures

List of Tables

Acknowledgements

My graduate school experience has taught me the true meaning of fighting for a dream. I would like to thank my advisor, Patricia Jennings, for having the faith and trust in me to find my own scientific path. I am so grateful for the creative freedom I was given over the years. It is a rare thing and has allowed me to study what I have been most passionate about. It has made me proud of my work and all it took to achieve it, thank you.

I would like to thank my committee for the support over the years and for helping me grow over the course of many grueling exams. You have all lead by example and have served as wonderful scientific role models.

Thank you to the members of the Jennings lab, past and present: Melinda Roy, Kendra Hailey, Sulyman Barkho, Michael Jamros, Bob Wong, Andrea Conlan, John Zuris, Dominique Capraro, Elizabeth Baxter, Nick Tiee, D.J. Burban, Colin Lipper and Jason Stofleth. From whiskey to pipettes, you have been there for me through all of it.

Thank you to my collaborators, Jose Onuchic and Jeffrey Noel. I have learned so much from you. But even more, your patience and kindness and inclusion of me in all of your scientific discussions was invaluable.

Thank you to Ellinor Haglund. You believed in me more than anyone and your unwavering support, faith in me, and friendship got me through the hardest days. I will always be grateful to you, always.

Thank you to my family. Every day I wake up, I know I am loved. Thank you for all the love and support and laughs.

Thank you to Josh. You have made every day sunnier and brighter. Your kindness and intelligence blows me away, every day. I am so lucky to have you in my life and without you this journey would have felt like an even steeper mountain than it already was. Go team.

Chapter 3, in part, is a reprint of the material as it appears in "Geometrical frustration in Interleukin-33 decouples the dynamics of the functional element from the folding transition state ensemble" Fisher, Kaitlin M., Haglund, Ellinor, Noel, Jeffrey K., Hailey, Kendra L., Onuchic, Jose N., Jennings, Patricia A which has been submitted to JMB for publication. The dissertation author was the primary author/investigator of this paper.

Chapter 5, in part, is a reprint of the material as it appears in "Heterogeneous Dynamics and Conformational Frustration in the Binding Regions of Interleukin-33" Fisher, Kaitlin M., Fuglestad, Brian, Chan, Josh C., Jennings, Patricia A which is in preparation for submission. The dissertation author was the primary author/investigator of this paper.

Vita

2009 B.S., Biochemistry, Sonoma State University

2012 M.S., Chemistry, University of California, San Diego

2015 Ph.D., Chemistry, University of California, San Diego

Publications

1. **Fisher, Kaitlin M.,** Haglund, Ellinor, Noel, Jeffrey K., Hailey, Kendra L., Onuchic, Jose N., Jennings, Patricia A. "Geometrical frustration in Interleukin-33 decouples the dynamics of the functional element from the folding transition state ensemble." *Submitted to JMB*

2. **Fisher, Kaitlin M.,** Fuglestad, Brian, Chan, Josh C., Jennings, Patricia A. **"**Heterogeneous Dynamics and Conformational Frustration in the Binding Regions of Interleukin-33." *In preparation for submission*

Honors and Awards

2009- CSUPERB Howell Research Fellow

2009- Magna Cum Laude and Graduation with Distinction

Abstract of the Dissertation

The Native Landscape of the Cytokine Interleukin-33: Exploring the Link Between Folding and Dynamics

by

Kaitlin M. Fisher

Doctor of Philosophy in Chemistry

University of California, San Diego, 2015

Professor Patricia Jennings, Chair

The newly identified cytokine Interleukin-33 (IL-33) is currently the focus of multiple investigations into targeting pernicious inflammatory disorders. This mediator of inflammation plays a prevalent role in chronic disorders such as asthma, rheumatoid

arthritis, and progressive heart disease. To date, little has been done to characterize the biophysical properties of IL-33, leaving a void in our understanding of the mechanism of action of this important cytokine. The work presented here characterizes the full native landscape of IL-33, from classifying the folding route to understanding the dynamic behavior of the protein in the native basin. The folding of IL-33 progresses through an intermediate state whose formation is necessitated by the folding of a stable nucleus, a critical step in providing a stable scaffold for folding the frustrated functional binding region in the final folding step. Though these regions share similar contacts, similar structural elements, and similar geometry, the functional region of IL-33 folds differently, allowing for it to be responsive and malleable without unfolding the whole protein. To evaluate the malleability and responsiveness of the functional region in the native state, NMR analysis of the native state dynamics on multiple timescales was assessed. This set of experiments shows that the functional region displays significant conformational entropy and dynamic heterogeneity in comparison with the rest of the structure in the native state. Taken together, conserved frustration and functionality along the folding route and native basin ensemble is a prime example of the evolutionary pressure to balance the need for efficient folding and structural stability with the preservation of functionality. IL-33 is a protein that must engage a highly dynamic receptor, and in doing so must be highly dynamic itself. Increased dynamics that allow for the proper function necessitates the presence of a driving force, and in the case of IL-33, it is provided by conserved frustration in the functional region of IL-33 along both the folding pathway and within the native state ensemble.

Chapter 1 : General Introduction

1.1 Maintaining Balance: Inflammation

The inflammation cycle is one of the most carefully regulated immune processes in the human body. Inflammation is an innate immune response targeted against infectious pathogens and is critical for the sustained health of tissues and for overall survival [1]. In a healthy individual, the immune response is tuned to allow for the influx of blood, lymph, and immune cells through chemotaxis to a site of infection, subsequently facilitating the action of phagocytic cells such as neutrophils and macrophages [2]. The elimination of harmful agents, such as microbes, by phagocytosis is quickly followed by tissue repair. Fibroblasts are recruited to sites of inflammation to rapidly divide and produce collagen, a critical step in facilitating the repair of damaged tissue and for the final resolution of a healthy, temporary inflammatory process [3]. Together, inflammation represents a carefully and internally controlled feedback system allowing for the elimination of infection and subsequent healing of damaged tissues.

Although inflammation is a crucial process for survival from infection, inflammatory processes that become prolonged, misdirected, or overly reactive can be extremely detrimental. The prognosis for sustained and mis-regulated inflammation can include not only painful side effects associated with tissue damage, but can also have broader consequences such as the development of cancerous legions in affected tissues [4]. Autoimmune and Autoinflammatory disorders lead to sustained, mis-regulated immune responses that often produce prolonged inflammation and damaging side effects in afflicted patients. Autoimmune disorders are classified by a non-tolerance of the immune system towards the body's own proteins, leading to the production of antibodies directed against the body's own tissue

[5]. A few examples of autoimmune disorders include Lupus, Rheumatoid Arthritis, Multiple Sclerosis, and Type I Diabetes Mellitus [6-9], Figure 1-1. Autoinflammatory diseases are characterized by inflammatory processes that are chronic and non-specific, often leading to recurrent periods of inflammation that are either organ specific or systemic [10]. Inflammatory Bowel Disease, Pericarditis, Hereditary Periodic Fevers, and Type 2 Diabetes Mellitus are all examples of disorders caused by severe and chronic inflammation [11-14]. Both autoimmune and autoinflammatory disorders are examples of the detrimental effects inherent in not having the proper controls for the healthy regulation and resolution of inflammation.

The development of new, effective treatments is often difficult in the case of regulating inflammation. The intricate balance of inflammation is tightly regulated by specific mediators, creating unique challenges for tailoring treatments to specific types of disease pathologies. The primary difficulty in finding an effective treatment for mis-regulated inflammation is that the immune response cannot be completely knocked down when treating the disorders associated with excessive inflammation. A balance between too strong of an immune response and too little of a response must be achieved in order to maintain resistance to infection and to avoid development of cancers. In a healthy individual, the immune response is tuned to a perfect balance. Those individuals with inflammatory disorders, however, must rely upon treatments to restore balance in favor of only productive and healthy inflammation. In order to develop treatments to be maximally effective while still allowing for retention of a healthy immune response, it is critical to understand the mechanistic details of effectors involved in both the activation and resolution of inflammation.

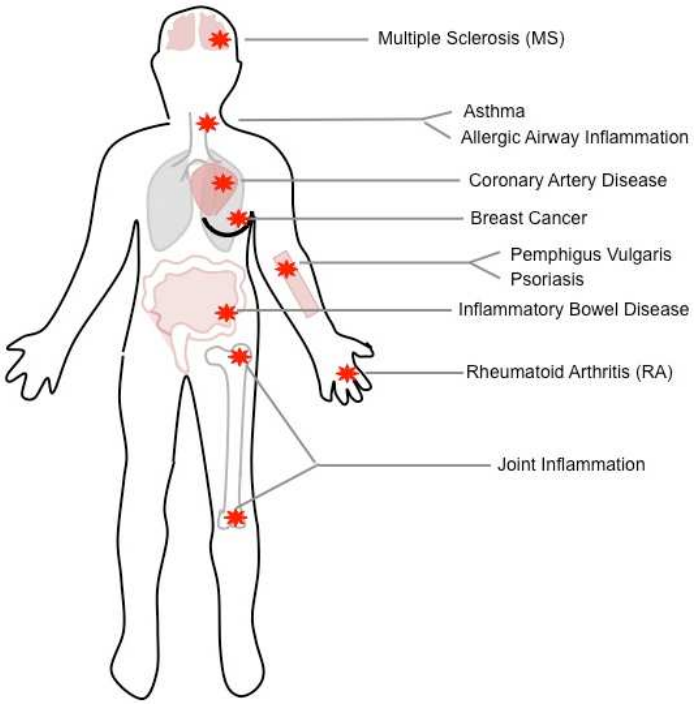

Figure 1-1: Schematic representing the systemic locations of inflammatory disorders.

Inflammation is present in most all tissue types in the human body and its mis-regulation can lead to a variety of pathologies. Presented here are just a few of the disorders associated with inflammation and tissue damage.

1.2 Regulating the Immune Response: The Role of the IL-1 family of Cytokines

Cytokines represent a broad category of small signaling proteins responsible for aiding in the regulation of immune responses [15]. They are secreted by multiple immune cell types including macrophages, monocytes, and neutrophils and can act as both local signaling mediators and as hormones through participation in long range signaling [16]. Although the definition of cytokines is broad, multiple sub families have been classified, including the Interleukin family. The Interleukin family of cytokines are secreted proteins responsible for facilitating communication between leukocytes [17]. Typically, a cytokine will receive the Interleukin designation based upon the functionality of the protein and/or the sequence homology. Currently, there are over 37 defined members of the family whose phylogeny is shown in Figure 1-2. The phylogenetic tree was generated using Clustal W to align the multiple protein sequences [18-20]. The Interleukins have diverse functions that affect the immune response and include the proliferation of T and B cells, immune cell differentiation, and apoptosis [21].

Within the Interleukin family, a further sub class has been defined as the Interleukin-1 family of cytokines. The Interleukin-1 family of cytokines is composed of 12 members who are grouped together based on their genetic homology, conservation of amino acid sequences, and secondary structure propensity [22]. The majority of the Interleukin-1 family members share homologous tertiary structure, characterized as a β-trefoil fold. The β-trefoil fold is defined as having twelve β-strands grouped

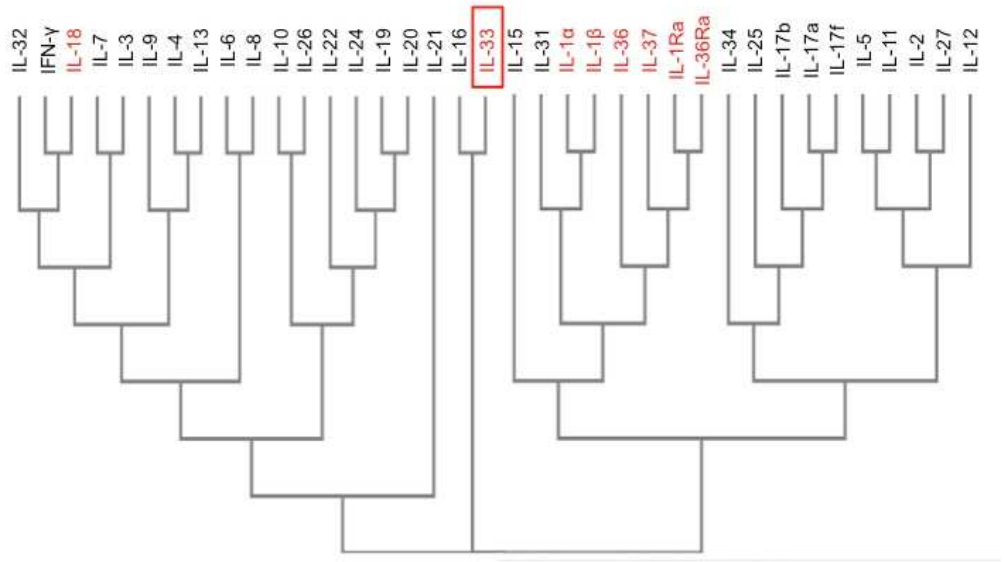

Figure 1-2: Phylogeny of Cytokines defined as Interleukin.

Presented is the phylogenetic tree of all members of the Interleukin class of proteins which are defined as small signaling proteins responsible for facilitating communication between lymphocytes. The phylogenetic tree was generated using Clustal W analysis of the protein sequence alignments. Interleukins belonging to the Interleukin-1 sub family are highlighted in red and IL-33 is presented, boxed. Interleukin-33 and Interleukin-18 have the least overlapping phylogeny within the family.

into three sets of four β-strands each, referred to as the individual trefoil units (Trefoils 1, 2, and 3, respectively). The final fold is barrel-like and is typically pseudo-symmetrical in nature. The well characterized IL-1β, IL-1Ra, and IL-1α have all been shown to adopt the characteristic β-trefoil fold in their final tertiary structure, as shown in Figure 1-3.

Within the Interleukin-1 cytokine family, members bind to conjugate extracellular receptors in order to regulate the inflammatory response. The extracellular receptors of the IL1 family are generally composed of three Immunoglobulin domains that present on the cell surface and connect to an intracellular Toll Interleukin Receptor (TIR) domain [23]. Upon binding an Interleukin-1 cytokine, the conjugate receptor will engage with a third receptor accessory protein to form an active, heterotrimeric signaling complex [24] as show in Figure 1-4. This complex allows for the downstream activation of cell proliferation, hematopoiesis, and subsequent induction of inflammation [25]. Although the exact structure of the heterotrimeric complex is still under investigation, it is understood that the signaling cascade is mediated by the initial, critical step of Interleukin binding to the primary conjugate cell surface receptor [26].

1.3 Interleukin-33: The Newest Member of An Expanding Group of Inflammation Mediators.

To date, the use of an Interleukin 1 blockade in combination with a TNF-α blockade has proven successful in the treatment of Rheumatoid Arthritis [27-29]. This

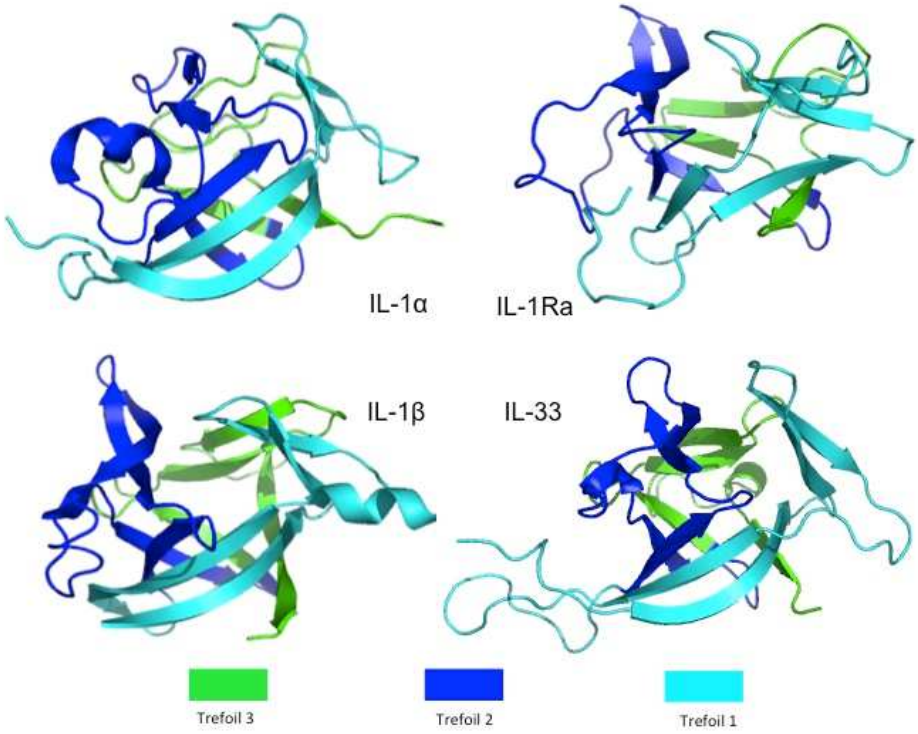

Figure 1-3: The tertiary structure of four representative members of the Interleukin-1 family.

Below are four members of the Interleukin-1 family with fold homology. IL-33 (PDB code 2KLL), IL-1β (PDB code 1I1B), IL-1α (PDB code IL1A), and IL-1Ra (PDB code 1IRP) all display the characteristic β-trefoil fold which is composed of three groups of four β-strands arranged in a β-β-β-loop-β arrangement. Each group is designated as an individual trefoil. Trefoil 1 is composed of β-strands 1-4, Trefoil 2 of β-strands 5-8, and Trefoil 3 of strands 9-12. Trefoil 1 is colored in cyan, Trefoil 2 in blue, and Trefoil 3 in green.

successful therapeutic strategy suggests that other Interleukin 1 blockade treatments targeted at specific inflammatory disorders may also prove successful. One therapeutic target of interest is the newest member of the Interleukin-1 family, Interleukin-33. The disease states associated with IL-33 are vast and systemic. Since its discovery in 2005, extensive biological investigations of IL-33 have linked its presence to disorders including but not limited to Asthma, Sepsis, Inflammatory Bowel Disease, Rheumatoid Arthritis, skin inflammation, Breast Cancer, and Coronary Artery Disease [30-36]. Although the biological implications of this small cytokine are under active investigation, very little has been done to characterize the biophysical properties of this important inflammatory modulator.

Interleukin-33 (IL-33) is primarily expressed in epithelial, endothelial, and fibroblast cells [37]. IL-33 is unique within the Interleukin-1 family of cytokines because of its dual functionality, acting as both an extracellular signaling protein and as a transcriptional modulator through direct DNA association [38]. In the absence of an inflammatory stimulus, the full length, 270 residue IL-33 protein will stay localized in the nucleus [39]. The C-terminal 160 residue “mature” domain shares the characteristic β-trefoil fold typical within the Interleukin-1 family of cytokines and is responsible for binding to the cell surface expressed receptor ST2, found primarily on the surface of T_H2 T Helper cells [40]. Although the exact mechanisms responsible for processing the pro domain into its two constitutive halves has not been fully characterized under all conditions, it is understood that the mature domain is a more potent extracellular signaler (over 30 times more potent). IL-33 is processed into its mature form by mast cell proteases allowing for activation of innate lymphoid cells through binding the cell surface expressed ST2 receptor [41,42] as shown in Figure 1-5. After IL-33 binds to ST2 it is able to engage the membrane bound Interleukin-1

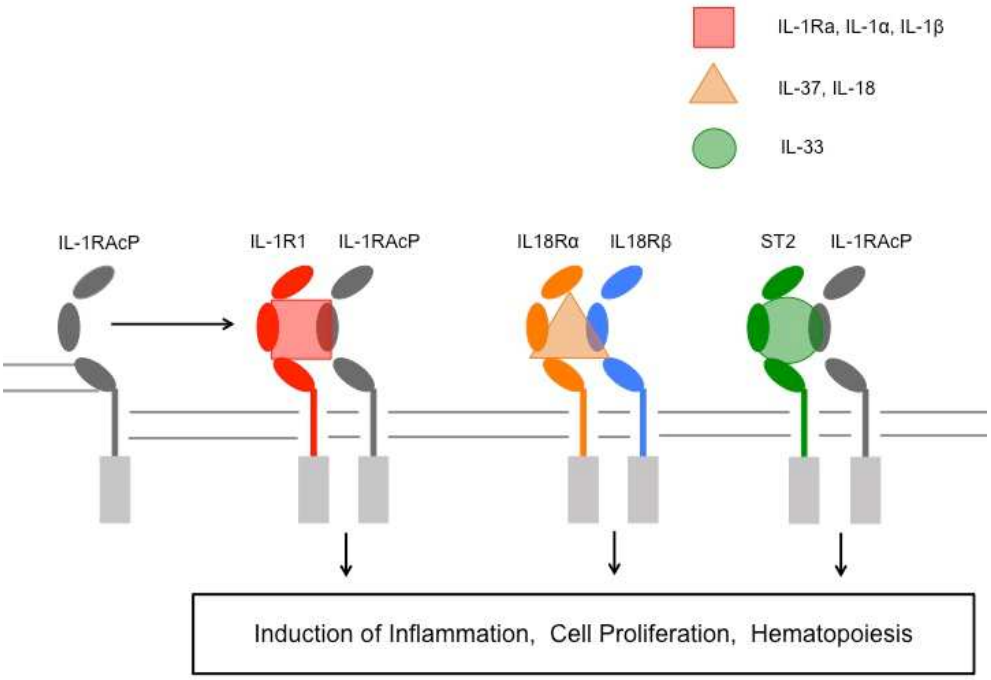

Figure 1-4: Extracellular signaling mechanism for the Interleukin-1 family of cytokines.

Displayed are representative diagrams for the formation of the heterotrimeric signaling complex for three select Interluekin-1 family members. The Interleukin-1 cytokine binds to the primary conjugate receptor and engages the receptor's three Ig domains. After binding, the Interleukin-1 Receptor Accessory receptor is recruited and binds to form the ternary signaling complex. Receptors are membrane bound and contain intracellular TIR domains whose association leads to downstream signaling.

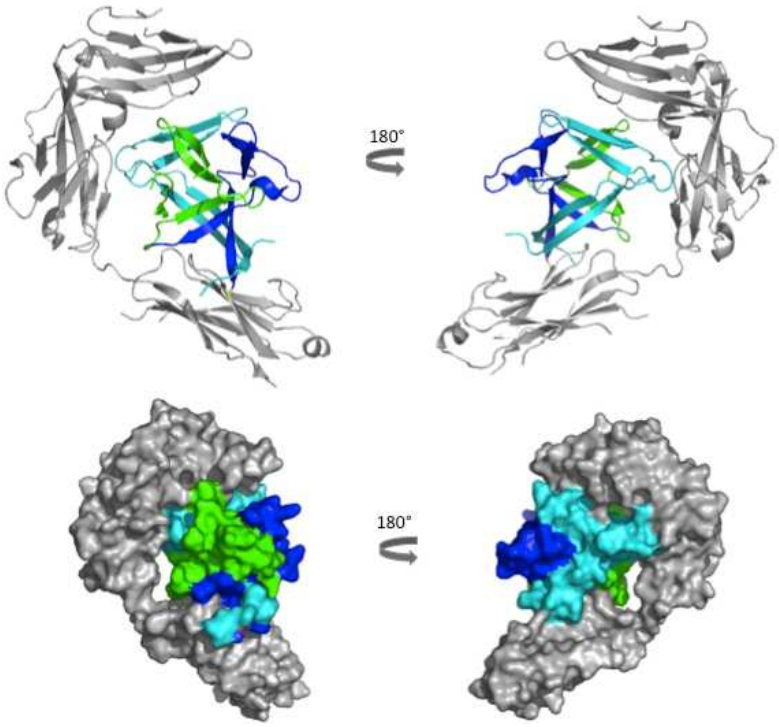

Figure 1-5: Structural detail of the Interleukin-33 and ST2 binding interaction.

The crystal structure of the Interelukin-33 and ST2 binding interaction (PDB code 4KC3) is shown above as a 180° view. The trefoils in IL-33 are highlighted as Trefoil 1 in cyan, Trefoil 2 in blue, and Trefoil 3 in green. The ST2 receptor is highlighted in grey. The space-filling model (bottom) highlights that the primary binding interface between IL-33 and ST2 is between Trefoil 1 and the receptor.

Receptor Accessory Protein to form the active heterotrimeric signaling complex responsible for activating a variety of downstream effectors leading to hematopoiesis and induction of inflammation, respectively [43,44]. Understanding the biophysical properties of IL-33 may prove valuable in gaining a more detailed insight into the functionality of the protein and the ways in which these specific properties can be exploited in targeted therapeutic intervention. In the following work the entire native landscape of IL-33 is investigated, including the folding pathway and native state dynamics, with a view towards a more detailed understanding of the biophysical properties critical in mediating the function of IL-33.

1.4 References

[1] Koyasu S, Moro K. Role of innate lymphocytes in infection and inammation. Front Immunol 2012;3.

[2] Maderna P, Godson C. Phagocytosis of apoptotic cells and the resolution of inflammation. Biochim Biophys Acta - Mol Basis Dis 2003;1639:141–51.

[3] Chesney J, Metz C, Stavitsky AB, Bacher M, Bucala R. Regulated production of type I collagen and inflammatory cytokines by peripheral blood fibrocytes. J Immunol 1998;160:419–25.

[4] Coussens LM, Werb Z. Inflammation and cancer. Nature 2002;420:860–7.

[5] Marrack P, Kappler J, Kotzin BL. Autoimmune disease: why and where it occurs. Nat Med 2001;7:899–905.

[6] Tsokos GC. Systemic lupus erythematosus. N Engl J Med 2011;365:2110–21.

[7] Choy E. Understanding the dynamics: pathways involved in the pathogenesis of rheumatoid arthritis. Rheumatology (Oxford) 2012;51 Suppl 5:v3–11.

[8] McFarland HF, Martin R. Multiple sclerosis: a complicated picture of autoimmunity. Nat Immunol 2007;8:913–9.

[9] Husebye ES, Anderson MS. Autoimmune Polyendocrine Syndromes: Clues to Type 1 Diabetes Pathogenesis. Immunity 2010;32:479–87.

[10] Masters SL, Simon A, Aksentijevich I, Kastner DL. Horror autoinflammaticus: the molecular pathophysiology of autoinflammatory disease (*). Annu Rev Immunol 2009;27:621–68.

[11] Kaser A, Zeissig S, Blumberg RS. Inflammatory bowel disease. Annu Rev Immunol 2010;28:573–621.

[12] Rapose A, Sarvat B, Sarria JC. Immune reconstitution inflammatory syndrome presenting as pericarditis and pericardial effusion. Cardiology 2008;110:142–4.

[13] Simon A, van der Meer JWM. Pathogenesis of familial periodic fever syndromes or hereditary autoinflammatory syndromes. Am J Physiol Regul Integr Comp Physiol 2007;292:R86–98.

[14] Mandrup-Poulsen T. Type 2 Diabetes Mellitus. A metabolic autoinflammatory disease. Dermatol Clin 2013;31:495–506.

[15] O'Shea JJ, Murray PJ. Cytokine Signaling Modules in Inflammatory Responses. Immunity 2008;28:477–87.

[16] Akdis M, Burgler S, Crameri R, Eiwegger T, Fujita H, Gomez E, Klunker S, Meyer N, O'Mahony L, Palomares O, Rhyner C, Quaked N, Schaffartzik A, Van De Veen W, Zeller S, Zimmermann M, Akdis CA. Interleukins, from 1 to 37, and interferon-γ: Receptors, functions, and roles in diseases. J Allergy Clin Immunol 2011;127:701–21.

[17] Dinarello C a. Immunological and inflammatory functions of the interleukin-1 family. Annu Rev Immunol 2009;27:519–50.

[18] Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–8.

[19] Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res 2010;38.

[20] McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res 2013;41.

[21] Dinarello C a. IL-1: Discoveries, controversies and future directions. Eur J Immunol 2010;40:599–606.

[22] Garlanda C, Dinarello CA, Mantovani A. The Interleukin-1 Family: Back to the Future. Immunity 2013;39:1003–18.

[23] Volpe F, Clatworthy J, Kaptein A, Maschera B, Griffin AM, Ray K. The IL1 receptor accessory protein is responsible for the recruitment of the interleukin-1 receptor associated kinase to the IL1/IL1 receptor I complex. FEBS Lett 1997;419:41–4.

[24] Thomas C, Bazan JF, Garcia KC. Structure of the activating IL-1 receptor signaling complex. Nat Struct Mol Biol 2012;19:455–7.

[25] O'Neill LAJ. The interleukin-1 receptor/Toll-like receptor superfamily: 10 Years of progress. Immunol Rev 2008;226:10–8.

[26] Lingel A, Weiss TM, Niebuhr M, Pan B, Appleton B a., Wiesmann C, Bazan JF, Fairbrother WJ. Structure of IL-33 and Its Interaction with the ST2 and IL-1RAcP Receptors-Insight into Heterotrimeric IL-1 Signaling Complexes. Structure 2009;17:1398–410.

[27] Bansard C, Lequerré T, Derambure C, Vittecoq O, Hiron M, Daragon A, Pouplin S, Daveau M, Boyer O, Tron F, Le Loët X, Salier JP. Gene profiling predicts rheumatoid arthritis responsiveness to IL-1Ra (anakinra). Rheumatology 2011;50:283–92.

[28] Quartier P, Allantaz F, Cimaz R, Pillet P, Messiaen C, Bardin C, Bossuyt X, Boutten A, Bienvenu J, Duquesne A, Richer O, Chaussabel D, Mogenet A, Banchereau J, Treluyer J-M, Landais P, Pascual V. A multicentre, randomised, double-blind, placebo-controlled trial with the interleukin-1 receptor antagonist anakinra in patients with systemic-onset juvenile idiopathic arthritis (ANAJIS trial). Ann Rheum Dis 2011;70:747–54.

[29] Mertens M, Singh JA. Anakinra for rheumatoid arthritis: a systematic review. J Rheumatol 2009;36:1118–25.

[30] Stolarski B, Kurowska-Stolarska M, Kewin P, Xu D, Liew FY. IL-33 exacerbates eosinophil-mediated airway inflammation. J Immunol 2010;185:3472–80.

[31] Alves-Filho JC, Sônego F, Souto FO, Freitas A, Verri WA, Auxiliadora-Martins M, Basile-Filho A, McKenzie AN, Xu D, Cunha FQ, Liew FY. Interleukin-33 attenuates sepsis by enhancing neutrophil influx to the site of infection. Nat Med 2010;16:708–12.

[32] Sponheim J, Pollheimer J, Olsen T, Balogh J, Hammarström C, Loos T, Kasprzycka M, Sørensen DR, Nilsen HR, Küchler AM, Vatn MH, Haraldsen G. Inflammatory bowel disease-associated interleukin-33 is preferentially expressed in ulceration-associated myofibroblasts. Am J Pathol 2010;177:2804–15.

[33] Mu R, Huang H-Q, Li Y-H, Li C, Ye H, Li Z-G. Elevated serum interleukin 33 is associated with autoantibody production in patients with rheumatoid arthritis. J Rheumatol 2010;37:2006–13.

[34] Hueber AJ, Alves-Filho JC, Asquith DL, Michels C, Millar NL, Reilly JH, Graham GJ, Liew FY, Miller AM, McInnes IB. IL-33 induces skin inflammation with mast cell and neutrophil activation. Eur J Immunol 2011;41:2229–37.

[35] Jovanovic IP, Pejnovic NN, Radosavljevic GD, Pantic JM, Milovanovic MZ, Arsenijevic NN, Lukic ML. Interleukin-33/ST2 axis promotes breast cancer growth and metastases by facilitating intratumoral accumulation of immunosuppressive and innate lymphoid cells. Int J Cancer 2014;134:1669–82.

[36] Demyanets S, Speidl WS, Tentzeris I, Jarai R, Katsaros KM, Farhan S, Krychtiuk KA, Wonnerth A, Weiss TW, Huber K, Wojta J. Soluble ST2 and interleukin-33 levels in coronary artery disease: Relation to disease activity and adverse outcome. PLoS One 2014;9.

[37] Liew FY, Pitman NI, McInnes IB. Disease-associated functions of IL-33: the new kid in the IL-1 family. Nat Rev Immunol 2010;10:103–10.

[38] Ali S, Mohs A, Thomas M, Klare J, Ross R, Schmitz ML, Martin MU. The dual function cytokine IL-33 interacts with the transcription factor NF-κB to dampen NF-κB-stimulated gene transcription. J Immunol 2011;187:1609–16.

[39] Roussel L, Erard M, Cayrol C, Girard J-P. Molecular mimicry between IL-33 and KSHV for attachment to chromatin through the H2A-H2B acidic pocket. EMBO Rep 2008;9:1006–12.

[40] Martin MU. Special aspects of interleukin-33 and the IL-33 receptor complex. Semin Immunol 2013;25:449–57.

[41] Lefrancais E, Roga S, Gautier V, Gonzalez-de-Peredo a., Monsarrat B, Girard J-P, Cayrol C. IL-33 is processed into mature bioactive forms by neutrophil elastase and cathepsin G. Proc Natl Acad Sci 2012;109:1673–8.

[42] Lefrançais E, Cayrol C. Mechanisms of IL-33 processing and secretion: Differences and similarities between IL-1 family members. Eur Cytokine Netw 2012;23:120–7.

[43] Lingel A, Weiss TM, Niebuhr M, Pan B, Appleton B a., Wiesmann C, Bazan JF, Fairbrother WJ. Structure of IL-33 and Its Interaction with the ST2 and IL-1RAcP Receptors-Insight into Heterotrimeric IL-1 Signaling Complexes. Structure 2009;17:1398–410.

[44] Liu X, Hammel M, He Y, Tainer J a, Jeng U-S, Zhang L, Wang S, Wang X. Structural insights into the interaction of IL-33 with its receptors. Proc Natl Acad Sci U S A 2013;110:14918–23.

Chapter 2 : An Introduction to Protein Folding

2.1 The Importance of Protein Folding

The healthy function of a cell is dependent on the performance of folded, functional proteins in their native state. The amino acid sequences of these proteins are encoded by genes in DNA, which are subsequently transcribed into messenger RNA (Figure 2-1). The code of messenger RNA is then translated into a heteropolymeric amino acid chain with sequence specificity for a given protein [1,2]. The synthesis of the amino acid chain is the first critical step in creating a functional protein, however the sequence of the amino acids in of themselves are not enough to impart functionality. This primary structure of the protein must fold into an often complex three dimensional arrangement dictated by numerous energetic and topological driving forces, as seen for a few representative examples in Figure 2-2 [3,4]. The three dimensional arrangement, or fold, of a protein is the most critical factor in dictating the proper function of a protein [5,6]. Incorrectly folded protein leads to a variety of complications- the simplest being a loss of functionality and the most dire being disease states associated with the accumulation of mis-folded proteins [7,8]. Because nearly every physiologic function is carried out by the action of proteins, understanding how they obtain a three dimensional functional structure is of the utmost importance.

2.2 The forces and factors that lead to a folded, functional protein

Proteins can fold in a variety of ways. One of the more astounding aspects of protein folding is that they can fold independently without the aid of biological or molecular chaperones in the many cases, particularly when the cell is not undergoing stress mechanisms [9]. The folding of a protein, were it to commence through random

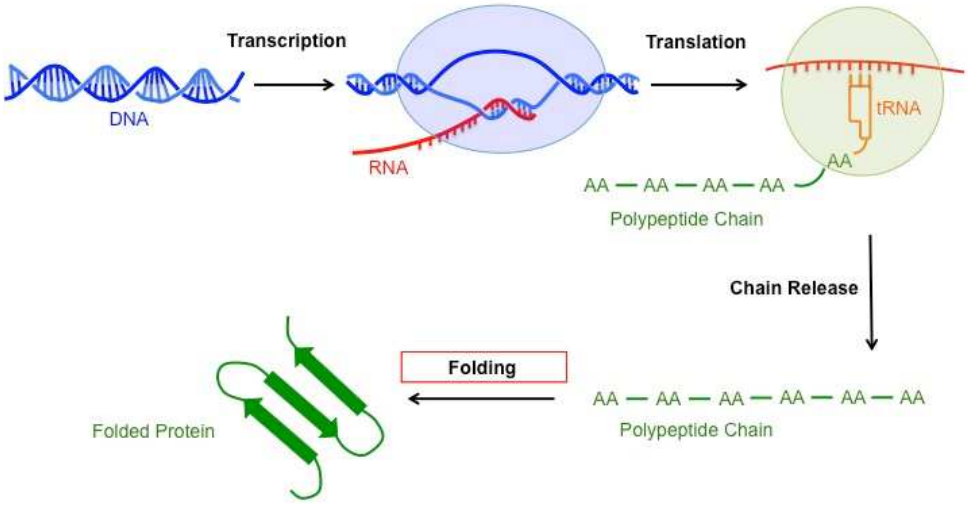

Figure 2-1: Protein Production and the Central Dogma of Biology.

The DNA chain is the center for storage of genetic information in the cell. The specific sequences of three codons present in DNA are responsible for encoding all gene products within the cell- the major of which are proteins. Upon a cell stimulus or at basal levels, RNA polymerase transcribes DNA into messenger RNA. The resulting single stranded RNA is used as a template for protein translation by a ribosomal complex and transfer RNA carrying individual amino acids. Upon the completion of protein synthesis, the nascent protein chain is released to initiate folding of the protein to its final tertiary structure. The principles underlying the folding of the protein chain will be explored in the following chapter in depth.

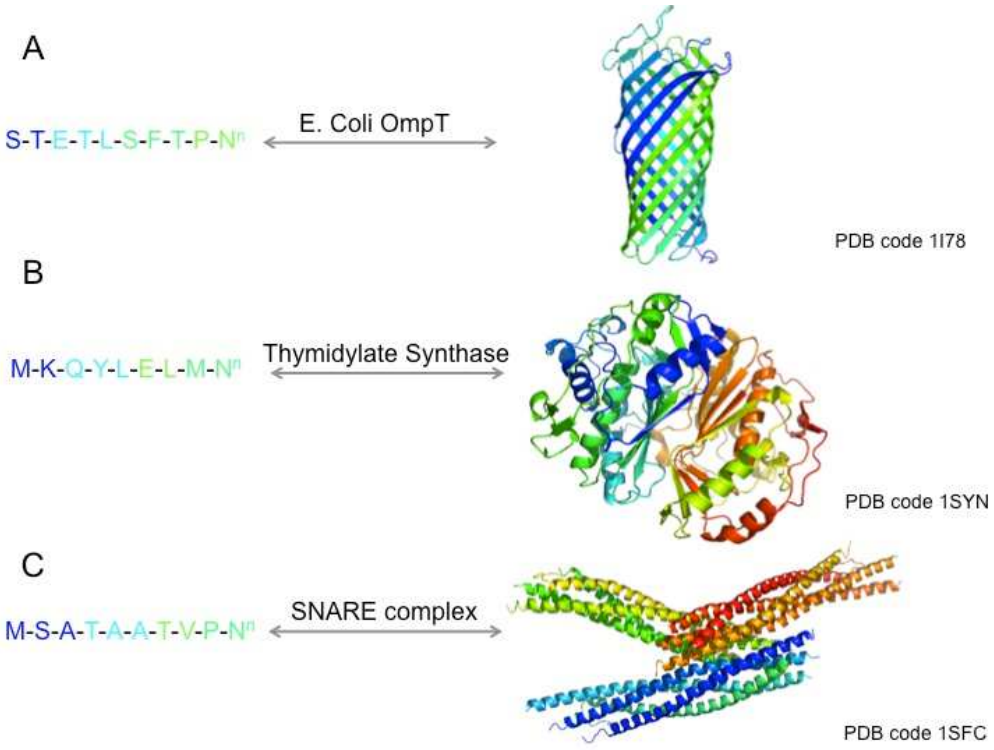

Figure 2-2: Heteropolymeric Peptide Chains are able to fold with Remarkable Diversity.

Represented above are the C-terminal end sequences and the resulting folds of A) E.Coli OmpT, an all β-sheet E.Coli membrane protein, B) Thymidylate Synthase, a mixed β-sheet and α-helical dimer, and C) a SNARE complex, composed solely of α-helices. Each of these proteins fold based on their primary amino acid sequence with a clear diversity of tertiary structure

sampling of native contacts and native topology, would progress on a time scale far too long to allow for the production of physiologically viable proteins. Levinthal's paradox outlines the extreme nature of the random sampling protein folding problem in that were random sampling of structure to occur, a small protein with 101 residues would have over 3^{100}, or $5x10^{47}$, possible conformations and would take longer than the age of the universe to sample all conformations, even if it were able to sample at an astounding rate of 10^{13} conformations per second [10]. Indeed, most proteins fold rapidly (with reactions ranging from the sub-millisecond to second timescale) to reach their final native structure. This rapid and orderly formation of native tertiary structure imposes that there must be energetic and topological factors responsible for driving the folding reaction progress forward. Although folding chaperones are in place to assist in the folding of protein or to correct instances of mis-folded proteins, the majority of proteins are able to fold independently and rapidly to their final native structure [11]. This process is guided by creating energetically favorable contacts between amino acids as well as other guiding principles such as the increased entropy involved in solvent exclusion [12], Van Der Waals forces [13], hydrogen bonding [14], salt bridges (electrostatic interactions) [15], and topological effects involved in the accessibility of certain folds [16]. A representative folding free energy reaction coordinate describing the energetic principles of the folding reaction is shown in Figure 2-3. The driving factors for protein folding are collectively best described by the funneled energy landscape theory which posits that protein folding leads to a decrease in energy, on average, as a protein fold become more native like [17]. The native fold of a protein represents an energy minimum for the native contacts and subsequent native tertiary structure. This energy minimum allows for a protein to follow a biased path to native structure through the stabilization of local conformational entropy and

through minimal frustrated energy between native contacts [17–20]. The folding along the funnel is not perfectly smooth, however. The ruggedness along the folding funnel can be attributed to competition among different contacts that can be made on the way to the stable native structure [21]. The most fundamental driving force, however, is the native bias, which allows for the rapid, independent folding of the protein to its final structure, as shown in Figure 2-4 [22]. Due to the fact that proteins need to pass through ensembles of partially folded structures to reach the native structure, the folding of proteins can be analyzed statistically through both experiment and computation.

Protein folding in its simplest form represents a trade off between energy and entropy within a system and can be described by Gibb's free energy-

3.1 $\Delta G = \Delta H - T\Delta S$

Wherein the balance between a loss in entropy as the protein becomes more ordered and the loss in energy as native contacts form favorable interactions leads to an overall favorable free energy for the folding reaction. Although the overall reaction is favorable, the free energy reaction profile is initially uphill because the initial formations of contacts are more entropically costly. However, as the folding reaction progresses forward and reaches a state that more closely resembles the native state, the favorable formation of contacts and overall decrease in the entropy becomes a favorable energetically and drives the reaction progress forward [23].

Although the folding of a protein involves a highly coordinated process of minimizing the energy associated with the native protein fold, energetics are not always the dominant factor associated with achieving the funneled landscape to the

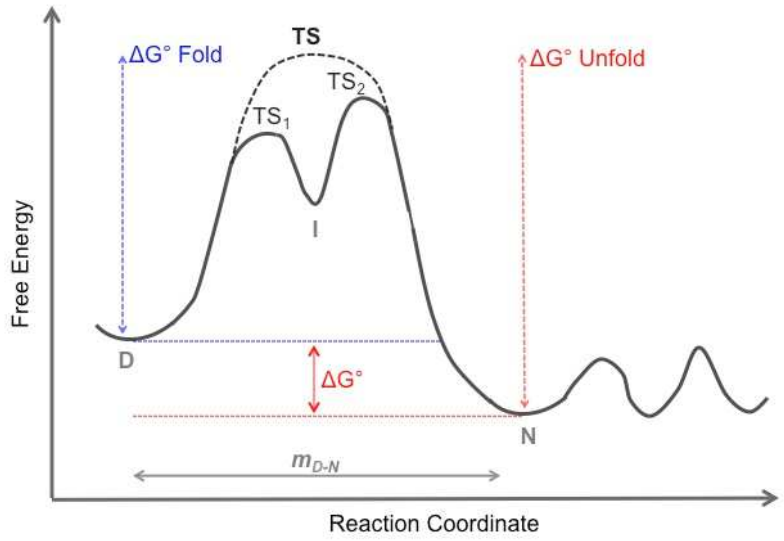

Figure 2-3: The Free Energy Reaction Coordinate for Protein Folding.

The underlying principle of protein folding is that the free energy of the folding reaction is a combination of minimizing both energy and entropy. The transition from the denatured state to the native state has an initial barrier due to the initial entropic penalties involved in initial contact formation. This is later compensated for in the reaction by the favorable energy contributions from contact formation in the native state. From the free energy reaction parameters, the rate of protein folding (k_f), the rate of unfolding (k_u), the ΔG° for the overall stability of the protein, and the $m_{D\text{-}N}$ value of the protein can all be determined. The overall folding reaction can be two-state, wherein the protein folds from the denatured basin (D) to the native basin (N) without significantly populating any transition states (TS) in between (dotted line). A protein can also fold via a more complicated three-state route wherein the denatured (D) state populates a folding intermediate state (I), necessitating two distinct transition states (TS_1 and TS_2).

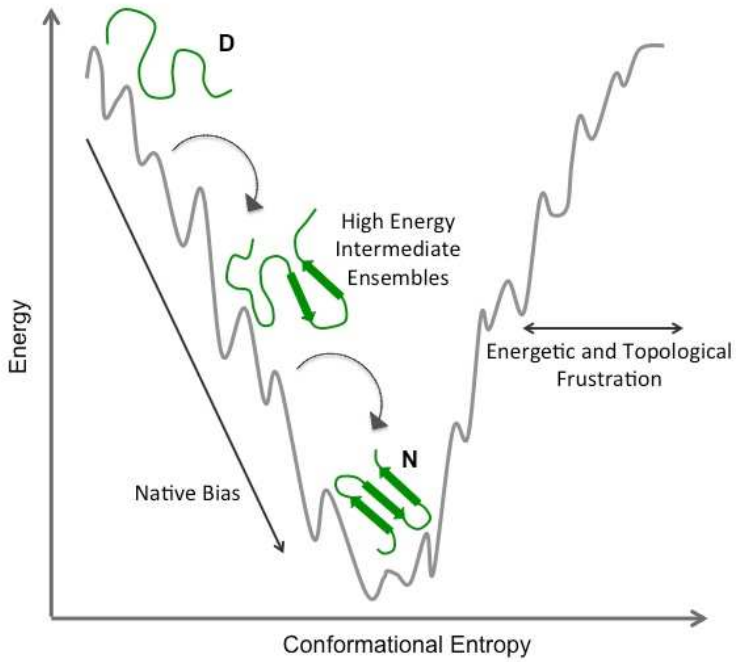

Figure 2-4: The Funneled Energy Landscape Theory of Protein Folding.

Were protein folding to commence without bias, the folding reaction would take too long to commence on a physiologically viable time scale. The funneled energy landscape theory posits that the overall energy for protein folding decreases as the protein fold becomes more native like. This bias leads to the overall funneled energy landscape. The native state represents an energetic and entropic minimum on the folding landscape. Roughness along the folding funnel is due to energetic terms involving the competition between contact formation during folding and topological frustration, which is representative of the difficulty in adopting certain conformations during folding. Folding involved traveling through multiple high-energy transition state ensembles during folding on the path to native.

native fold. Topological considerations, or the geometrical accessibility of a final fold is the strongest influence on the funneled folding landscape of a protein. In small, globular proteins the folding pathway can actually be dominated more by the topological constraints than the energetic constraints in reaching the final native fold [4,16]. The assessment of these folding routes and the influences they have on the folding reaction progress and associated folding rate can be readily analyzed through both experimental and simulation based approaches. This underlying principle of native bias and topological considerations in protein folding underlies the use of Gō-type Structure Based Models, which will be discussed more in depth later in the chapter.

2.3 Tools to Assess Folding Experimentally

2.3.1 Equilibrium Folding

An equilibrium folding experiment is a useful tool to probe the thermodynamics of a specific folding process. In an equilibrium experiment, a protein is put into varying amounts of a denaturing agent to shift it to an ensemble of states with an average resembling either a native or denatured state, Figure 2-5. For each ensemble of states, the fluorescent signature is assessed as a probe for tertiary structure and the subsequent signal is plotted as a titration curve versus the denaturant concentration. One can evaluate the thermodynamics of the system by fitting the curve according to the equation:

3.2 $\Delta G = -RT\ln K_{D\text{-}N}$ where $K_{D\text{-}N} = [D]/[N]$

By fitting the equilibrium curve the thermodynamic parameters and thus stability of the protein can be evaluated. The m value, or solvent exposed surface area (SASA),

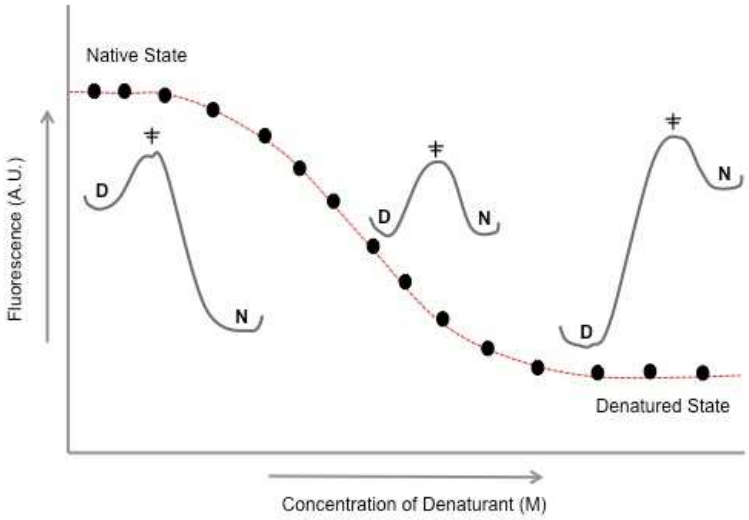

Figure 2-5: Equilibrium Folding Curve Analysis.

An example of an equilibrium folding titration is represented above. In this analysis, individual titration points of protein at varying concentrations of denaturant are plotted relative to the fluorescent signal of each sample. The variation in denaturant concentration shifts the overall population of protein from a predominantly native ensemble to a predominantly denatured ensemble. This shift can be fit as a function of the overall curve and leads to information about the midpoint, where the relative population of native and denatured protein is equal, and information about the overall stability of the protein in terms of Gibb's Free Energy (ΔG°).

present in the transition state of a protein can be determined by fitting the slope of the curve. Thermodynamic assessment of protein folding is useful for understanding the overall stability of the protein and the relative compactness of the unfolded transition state, however its limitation is that it cannot provide information on the particular folding route of a protein. To gain a more detailed understanding of the folding route, kinetic analysis is necessary.

1.1.2 Kinetics

Kinetic analysis of protein folding is useful in not only determining the parameters associated with the rates of folding, it can also shed light on the route a protein takes when it folds. For instance, the presence of a kinetic intermediate species along a particular folding route can be determined through kinetic analysis of a chevron plot. Chevron plots are constructed by monitoring the rate at which a protein of interest folds and the rate at which the protein unfolds (Figure 2-6). The two rates are extrapolated to form two halves of a chevron plot, giving it the titular chevron shape. The rates are monitored for individual titration steps, which vary depending upon the final concentration of denaturant present in the mixing reaction. The subsequent “limbs” of the chevron plot are fit to extrapolate the kinetic parameters associated with the folding reaction.

The acquisition of the individual folding rates requires following the folding reaction of a protein over time using a specific probe sensitive to changes in structure associated with folding and unfolding. Often, a fluorescent probe, such as a native Tryptophan, can be used to monitor the progress of the reaction. The folding reaction is often extremely fast, particularly when folded into native conditions and thus presents a challenge in reliably monitoring the rate, Figure 2-7. To achieve the best

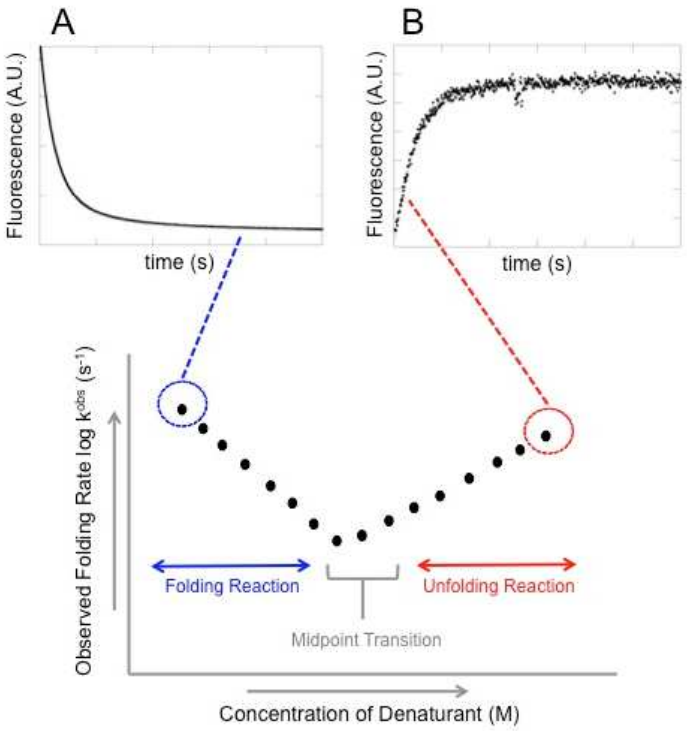

Figure 2-6: Constructing a Chevron Plot for Kinetic Analysis of a Folding Reaction.

A representative folding trace (A) and an unfolding trace (B) for IL-33 are shown as a function of fluorescent intensity versus time in seconds. The fluorescent signature of Tryptophan is sensitive to its local environment and represents a reliable probe for the native and denatured state of the protein, respectively. The traces are fit to obtain a rate of folding or unfolding, which are then plot relative to the denaturant concentration. This creates the two arms of the chevron plot.

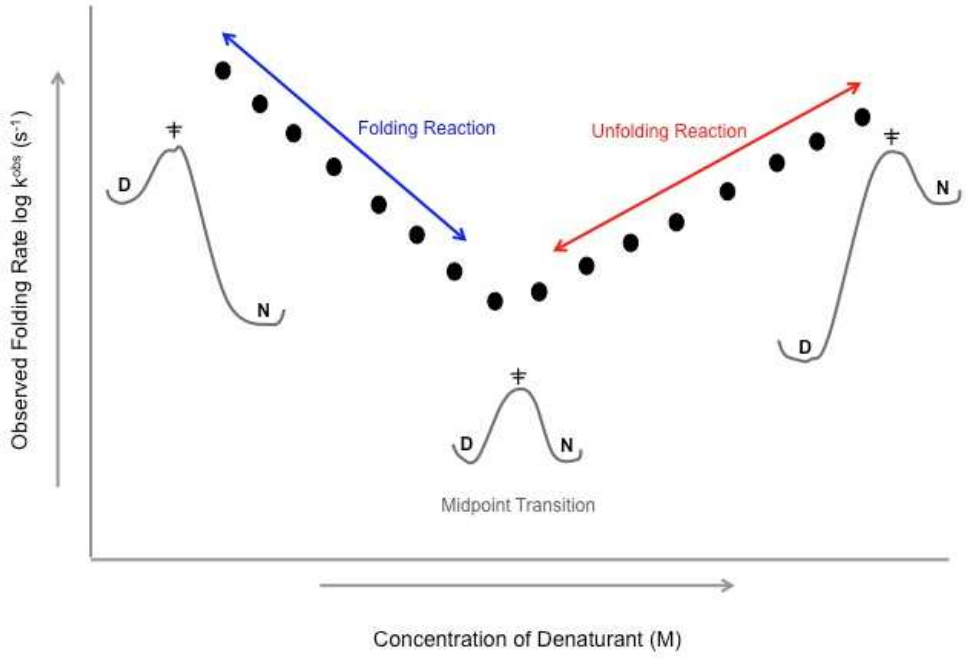

Figure 2-7: The Chevron Plot Represents Rates of the Folding Reaction Relative to the Stabilization of a particular Protein State.

As the concentration of denaturant varies, the different state of the final protein varies as well. At the beginning of the folding reaction, or the top of the folding arm, the dilution of unfolded protein into near native buffer conditions represents a shift that strongly stabilizes the native state (N) relative to the denatured state (D). This creates the fastest folding reaction relative to the subsequent titration points with a higher denaturant concentration in the refolding buffer. As the reaction moves closer to the midpoint, the native state (N) is less stabilized relative to the denatured state (D), leading to a slower rate of refolding for each assayed point. This same principle applies to the unfolding reaction except the initial, fastest reaction is established by stabilizing the denatured state (D) relative to the native state (N).

data acquisition at these faster timescales, a stopped flow instrument can be used to minimize the dead time associated with protein mixing and allows for the majority of the signal amplitude to be captured, allowing for accurate fitting of the folding rates. The individual folding and unfolding traces collected on the stopped flow are fit, typically, with a single exponential according to equation 3.3 or by higher order exponentials (such as double) according the equation 3.4:

3.3 $[N](t) = A_1 \exp^{(-kt)} + C$

3.4 $[N](t) = A_1 \exp^{(-kt)\,1} + A_2 \exp^{(-(kt)\,2} + C$

In the above exponentials, (t) is the time of the reaction, *A* is the amplitude of the signal, and *C* is the endpoint where the reaction has reached equilibrium. The observed rate constant (*k*obs) is described by the *k* term, which represents the rate of unfolding or the rate of folding, respectively These individual fits of the observed rate are then plotted versus denaturant concentration to construct the final chevron plot.

Proteins can fold through a variety of routes. The simplest case is folding through a two state route where the protein folds from a denatured to a native state without stably populating any states in between [24]. In the case of a two state folding route, the rates of folding and refolding will have a linear dependence on denaturant concentration and will subsequently give rise to a classic V-shaped chevron plot. The linear arms of these proteins may be fit to extrapolate the observed rates of folding and unfolding according to the following equations:

3.5 $\log k_f = \log k_f^{H2O} + (m_f[\text{denaturant}])$

3.6 $\log k_u = \log k_u^{H2O} + (m_u[\text{denaturant}])$

The log k_u^{H2O} and log k_f^{H2O} values represent the rates for unfolding and folding of the protein in water, independent of denaturant and are extrapolated from the fits of the folding and unfolding limbs of the chevron plot. The m_f and m_u terms represent the solvent accessible surface area (SASA) of the protein in each the unfolding and refolding reaction and are described by the associated slopes of the unfolding and refolding reactions, respectively (Figure 2-8).

The entire folding reaction represented by the chevron plot can be fit according to the following equation:

3.7 $\log k_{obs} = \log(k_f + k_u) = \log(10^{\log kf \text{ in } H2O + mf\,[GdmCl]} + 10^{\log ku \text{ in } H2O + mu\,[GdmCl]})$

Equation 3.7 allows for the parameters log k_f^{H2O}, log k_u^{H2O}, m_u, and m_f to be extrapolated as described above in equations 3.4 and 3.5. Additionally, the overall equilibrium constant K can be calculated as per the equation:

3.8 $\log K_{D\text{-}N} = \log k_f - \log k_u$

The equilibrium rate constant can then be used to calculate the overall stability of the protein as per the equation:

3.9 $\Delta G = -RT\ln k_{D\text{-}N}$

The relative packing of the structural transition state ensemble can be calculated from the kinetic parameters obtained from fitting the chevron plot with equation 3.7. The β^T (β Tanford value) describes the compactness of the transition state on a scale from 0 to 1, wherein the denatured state represents no compactness (a value of 0) and the native state represents full compactness (a value of 1). It can

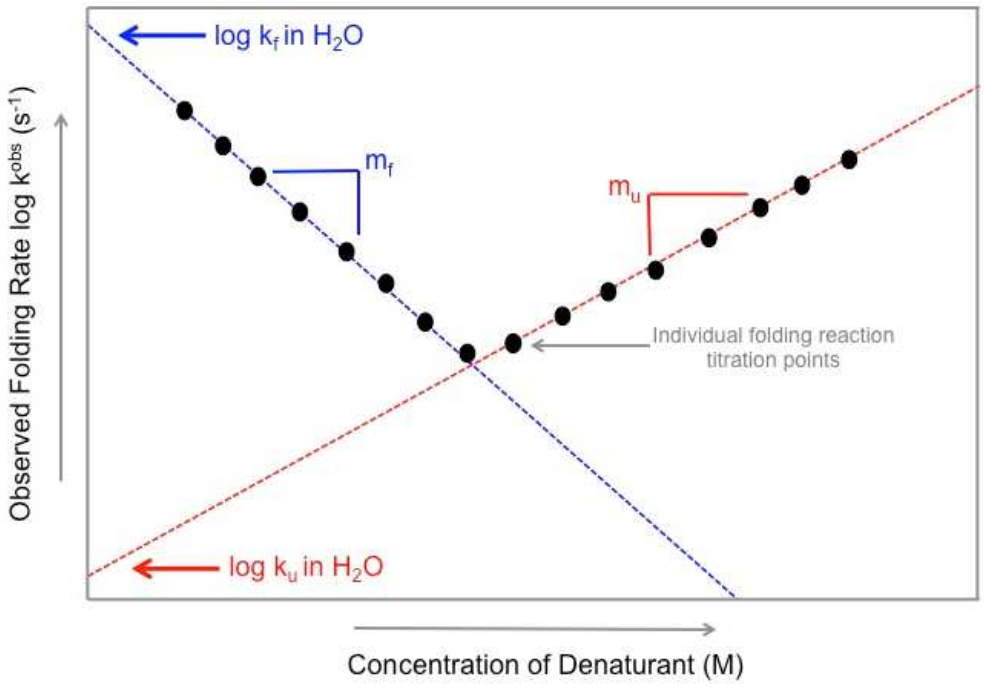

Figure 2-8: Extrapolation of Folding Kinetic Parameters from a Chevron Plot.

The folding and unfolding arms of the chevron plot can be fit to evaluate the rate of folding in water (denaturant independent folding rate), log k_f^{H2O} and the unfolding rate, log k_u^{H2O}. The *m* values, or the m_f and m_u are the relative slopes of the two arms of the chevron plot. The *m* values represent the amount of solvent exposed surface area (SASA) a protein has as it folds or unfolds. From these basic parameters of rate and SASA, the rest of the kinetic parameters of folding including β^T, $m_{D\text{-}N}$, and ΔG can be calculated.

also be described as representing the relative position of the transition state ensemble along the reaction coordinate. The β^T value is calculated using the following equation:

3.10 $\beta^T = 1\text{-}m_u / m_{D\text{-}N}$

The $m_{D\text{-}N}$ value represents the solvent exposed surface area (SASA) of a given protein under denaturing conditions and correlates with protein size and denaturant conditions. The $m_{D\text{-}N}$ value can be calculated using the following equation:

3.11 $m_{D\text{-}N} = m_u - m_f$

The chevron plot can also be fit to determine the kinetic midpoint of the protein. The midpoint represents the denaturant concentration at which 50% of the population is folded while the other 50% is unfolded at equilibrium. The midpoint fit is represented by equation 3.12:

3. 12 $MP = (\log k_f^{H2O} - \log k_u^{H20}) / (m_u - m_f)$

Through equilibrium and kinetic measurements we can assess the folding route and stability of a given protein. The development of computational approaches in assessing the folding route of a protein adds another tool to supplement experimental folding data. Through the combined use of experimental and computational methods, it is possible to attain a more detailed view of the factors associated with the dominant folding route of a protein.

2.4 Protein Folding through Structure Based Models (SBM)

The physical principles that underlie protein folding can be computationally modeled through the use of Structure Based Models (SBM). The underlying principle of SBM is that proteins have evolved to be minimally frustrated energetically along the

folding pathway and that the formation of native contacts drives the folding reaction forward [25]. This idea that native contact formation is the dominant factor influencing folding rather than frustrated energetics from non-native interactions has been backed up by the energy landscape (folding funnel) theory and the successful use of Gō-type models [26]. Structure based models (SBM) were built upon the concept of Gō models which assumed that the folding reaction and its associated interaction potentials followed a strong bias to the native state. Proteins have evolved to fold with minimal energetic frustration, which means that the formation of native contacts and the topological factors, ie the accessibility of a certain fold, drives the folding reaction [16,21]. From this observation, the development of SBM, wherein the topology is the dominating force for folding, has proven to be a successful model to parse out folding routes of proteins [27].

Although the use of kinetic experimental methods is critical in understanding the folding route of a protein, SBM can help to further clarify some of the finer details associated with protein folding, such as the order of contacts formed by specific structural elements and folding motifs and how these specific interactions can dictate the dominant folding route. The use of SBM utilizes a native topology bias in determining which route the protein will fold and which high-energy intermediate ensembles the protein will sample on the way to achieving the native fold [28]. Structure based models allow for the native state parameters to influence the Hamiltonian. The native state parameters simplify the Hamiltonian by allowing for the

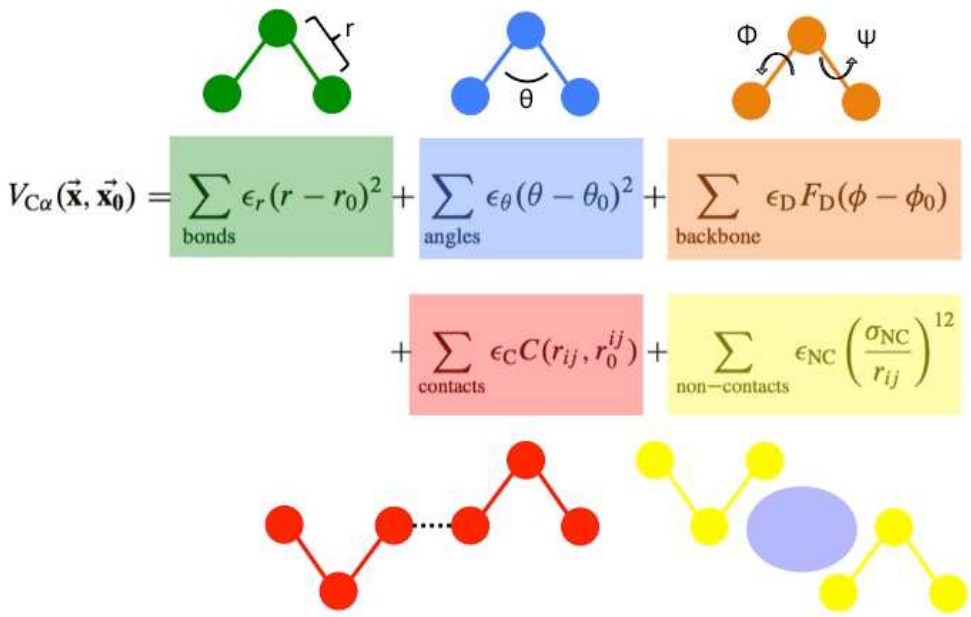

Figure 2-9: The Basic Cα Hamiltonian Parameters in a Structure Based Model (SBM).

SBM biases the folding Hamiltonian based on structural coordinates deposited from a PDB file and are represented in the Hamiltonian as **x** and $\mathbf{x}_{o,}$ respectively. The terms of the Hamiltonian take into account the bond lengths, described as ***r*** and highlighted in green. Bond angles are represented as ***θ*** and highlighted in blue, dihedrals are represented as ***Φ*** and highlighted in orange, contact distances are represented by the term r_{ij} and are highlighted in red, and the excluded volume term is represented as σ_{NC} and highlighted in yellow. Collectively, these terms represent the overall tertiary structure that a protein will achieve with a native bias and allows for the analysis of the folding route to native.

parameterization of allowable interactions in the native state [22]. The basic Hamiltonian used for the Cα model in SBM is:

$$V_{C\alpha}(\vec{\mathbf{x}}, \vec{\mathbf{x}_0}) = \sum_{\text{bonds}} \epsilon_r (r - r_0)^2 + \sum_{\text{angles}} \epsilon_\theta (\theta - \theta_0)^2 + \sum_{\text{backbone}} \epsilon_D F_D (\phi - \phi_0)$$

$$+ \sum_{\text{contacts}} \epsilon_C C(r_{ij}, r_0^{ij}) + \sum_{\text{non-contacts}} \epsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12}$$

Within the context of this Hamiltonian, each term is representative of an either attractive or repulsive force, allowing for the preservation of the tertiary structure, an important parameter implicit in maintaining the native state (Figure 2-9). **x** represents coordinates of a protein structure obtainable from a PDB file and represent the source of the native bias to be applied to the Hamiltonian. The r values represent the bond lengths between neighboring atoms, the θ values represent angles between bonds, the Φ values represent the allowable dihedrals (or torsional angles), and the r_{ij} term represents the parameters for the distance between native contacts. The final term maintains the proper excluded volume in the overall folded structure [28,29].

Through the use of a structure-based Cα model, the Cα coarse-grained simulation treats each amino acid residue as a single site centered on the alpha carbon on the amino acid chain. These interacting sites are held together along the amino acid chain by the harmonic energy associated with the Hamiltonian used to describe the angles, dihedrals, bond lengths, and distances between interacting pairs. The sites centered on the alpha carbons are assigned attractive interactions based upon their relative proximity to one another in the native state. Native contacts are

determined using a shadow map, which relates relative distances and possible steric interferences to extrapolate a reasonable cutoff for native contacts.

2.5 References

[1] Crick F. Central dogma of molecular biology. Nature 1970;227:561–3.

[2] Pukkila PJ, Carolina N. Molecular Biology : The Central Dogma. Life Sci 2001:1–5.

[3] Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. Annu Rev Biophys 2008;37:289–316.

[4] Baker D. A surprising simplicity to protein folding. Nature 2000;405:39–42.

[5] Andreeva A, Murzin AG. Evolution of protein fold in the presence of functional constraints. Curr Opin Struct Biol 2006;16:399–408.

[6] Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, Ranganathan R. Evolutionary information for specifying a protein fold. Nature 2005;437:512–8.

[7] Soto C, Estrada L, Castilla J. Amyloids, prions and the inherent infectious nature of misfolded protein aggregates. Trends Biochem Sci 2006;31:150–5.

[8] Goldberg AL. Protein degradation and protection against misfolded or damaged proteins. Nature 2003;426:895–9.

[9] Netzer WJ, Hartl FU. Protein folding in the cytosol: Chaperonin-dependent and -independent mechanisms. Trends Biochem Sci 1998;23:68–73.

[10] Zwanzig R, Szabo A, Bagchi B. Levinthal's paradox. Proc Natl Acad Sci U S A 1992;89:20–2.

[11] Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181:223–30.

[12] Harano Y, Kinoshita M. Large gain in translational entropy of water is a major driving force in protein folding. Chem Phys Lett 2004;399:342–8.

[13] Baldwin RL. Weak Interactions in Protein Folding: Hydrophobic Free Energy, van der Waals Interactions, Peptide Hydrogen Bonds, and Peptide Solvation. Protein Fold. Handb., vol. 1, 2008, p. 127–62.

[14] Bolen DW, Rose GD. Structure and energetics of the hydrogen-bonded backbone in protein folding. Annu Rev Biochem 2008;77:339–62.

[15] Strop P, Mayo SL. Contribution of surface salt bridges to protein stability. Biochemistry 2000;39:1251–5.

[16] Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 2000;298:937–53.

[17] Onuchic JN, Wolynes PG. Theory of protein folding. Curr Opin Struct Biol 2004;14:70–5.

[18] Plotkin SS, Onuchic JN. Investigation of routes and funnels in protein folding by free energy functional methods. Proc Natl Acad Sci 2000;97:6509–14.

[19] Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. Toward an outline of the topography of a realistic protein-folding funnel. Proc Natl Acad Sci U S A 1995;92:3626–30.

[20] Wolynes P, Luthey-Schulten Z, Onuchic J. Fast-folding experiments and the topography of protein folding energy landscapes. Chem Biol 1996;3:425–32.

[21] Chavez LL, Onuchic N, Clementi C. Quantifying the Roughness on the Free Energy Landscape : Entropic Bottlenecks and Protein Folding Rates 2004:8426–32.

[22] Lammert H, Schug A, Onuchic JN. Robustness and generalization of structure-based models for protein folding and function. Proteins Struct Funct Bioinforma 2009;77:881–91.

[23] Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.

[24] Alan F. Structure and mechanism in protein science, a guide to enzyme catalysis and protein folding. Lavoisier.fr 1999.

[25] Wolynes PG, Onuchic JN, Thirumalai D. Models of Protein Folding. Science (80-) 1995;268:959–71.

[26] Hills RD, Brooks CL. Insights from coarse-grained go models for protein folding and dynamics. Int J Mol Sci 2009;10:889–905.

[27] Shea JE, Brooks CL. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu Rev Phys Chem 2001;52:499–535.

[28] Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN. SMOG@ctbp: Simplified deployment of structure-based models in GROMACS. Nucleic Acids Res 2010;38.

[29] Noel JK, Whitford PC. How Simulations Reveal Dynamics, Disorder, and the Energy Landscapes of Biomolecular Function. Isr J Chem 2014;54:1093–107.

Chapter 3 : Geometric Frustration in Interleukin-33 decouples the dynamics of the functional element from the folding transition state ensemble

3.1 Abstract

Interleukin-33 (IL-33) is currently the focus of multiple investigations into targeting pernicious inflammatory disorders. This mediator of inflammation plays a prevalent role in chronic disorders such as asthma, rheumatoid arthritis, and progressive heart disease. In order to better understand the possible link between the folding free energy landscape and functional regions in IL-33, a combined experimental and theoretical approach was applied. We identified that IL-33 folds with a three-state mechanism, leading to a rollover and second slower phase in the refolding arm of its chevron plots in strongly native conditions. IL-33 is a pseudo-symmetrical protein composed of three distinct structural elements that complicate the folding mechanism due to competition for nucleation on the dominant folding route. Characterization of the intermediate state and the rate limiting steps required for folding suggests that the rollover is attributable to a moving transition state, shifting from a post- to pre-intermediate transition state in strongly native conditions. The formation of the intermediate state and subsequent moving transition state is mediated by the difficulty in folding the functional element of the protein. Significantly, the functional element of the protein is geometrically frustrated, requiring the more stable elements to fold first, acting as a scaffold for docking of the functional element to allow productive folding to the native state. Our results suggest that the requirement for the preservation of functionality within a large region of IL-33 complicates the folding landscape of this inflammatory protein.

3.2 Introduction

Chronic inflammatory disorders, to date, are very difficult to treat effectively and lead to a variety of damaging pathologies. Diseases associated with inflammation affect millions and yet the treatments currently available are primarily palliative measures [1]. Interleukin-33 (IL-33) is a pro-inflammatory cytokine predominately expressed in epithelial cells as an alarmin and is responsible for activating a variety of cells including mast cells, $CD8^+$ T cells, and hematopoietic cells [2,3]. IL-33 acts primarily as a mediator of inflammation and at this time is being actively pursued as a drug target for diseases including asthma, heart disease, cancer [4–6], and rheumatoid arthritis [7,8]. The pursuit of IL-33 as a drug target has potential to treat the inflammatory pathways responsible for causing long lasting damage from chronic inflammation more directly [9].

IL-33 has a β-trefoil fold that represents a shared motif with other members of the Interleukin-1 family (IL) such as IL-1β and IL-1ra [10,11] and with other β-trefoil proteins such as Human Fibroblast Growth Factor-1 (hFGF-1) and Hisactophilin [12]. The β-trefoil motif is characterized by a three fold pseudo-symmetrical structure. It is composed of three units of four β-strands that fold to form three β-β-β-loop-β elements (Figure 3-1). The pseudo-symmetry in the final fold may help in the propagation of allostery across the structure as seen in other beta-barrel like proteins [13,14]. Interestingly, the first pseudo-symmetrical element (Trefoil 1) constitutes the large binding interface responsible for engaging IL-33 to its conjugate cell surface receptor ST2, representing the first critical step in formation of the active heterotrimeric signaling complex [15]. The binding of IL-33 to ST2 allows for binding of a secondary

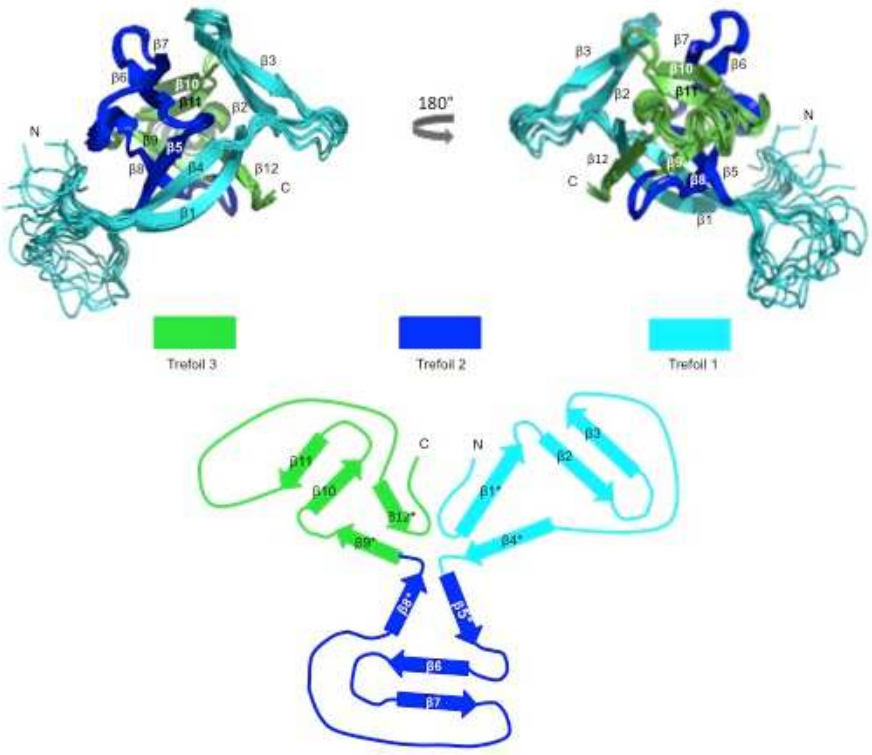

Figure 3-1: The Topology of Interleukin-33.

(Top) The 20 overlaid NMR structures of IL-33 (PDB code 2KLL) are displayed in two orientations as indicated. Trefoil 1 is highlighted in cyan, Trefoil 2 in blue, and Trefoil 3 in green. (Bottom) A representative two-dimensional splay diagram of IL-33 from a top view of the protein. Barrel strands are indicated by asterisks and β-hairpin cap strands are on the periphery of the diagram. Trefoil 1 is composed of strands 1-4, Trefoil 2 strands 5-8, and Trefoil 3 strands 9-12.

accessory receptor to form a ternary signaling complex [16]. IL-33, as a newer member within this family, has not yet been fully biophysically characterized. Understanding the folding free energy landscape is important in terms of the functionality of IL-33, specifically the influence of each trefoil on the biological activity as it pertains to drug design and targeting [17].

The folding route of a protein can illuminate the underlying causes of disease states by investigating the specific events that lead to a functional native protein. The simplest folding landscape is through a two-state mechanism, in which the protein folds directly from the denatured state (D) to the native state (N) without populating any states in between [18]. Two-state folding behavior gives rise to a characteristic V-shaped, linear chevron plot [18]. Proteins can also fold through more complicated mechanisms wherein intermediate states (I) may be populated before reaching the fully native state [19–21]. One example of a more complicated system is the well-characterized RnaseH, where the rollover in the refolding arm has been determined to be caused by the accumulation of an intermediate state along the folding pathway (17). The occurrence of kinetic rollovers in the refolding arms of chevrons have been attributed to a large variety of phenomena including but not limited to a) intermediate states (18, 19), b) protein aggregation (20), c) denaturant effects (21),and d) proline isomerization (22). The complicated folding kinetics of IL-33 involves a rollover and additional second slower phase of folding in strongly native conditions. The folding route of IL-33 required systematic evaluation to identify and characterize the exact source of the more complicated folding kinetics. We used a combined experimental and theoretical approach to investigate the folding free energy landscape in order to fully characterize the novel folding route of this mediator of inflammation.

Our combined results suggest that the rollover in IL-33 is a consequence of an intermediate state along the folding route. The rate-limiting step for folding moves from a pre-intermediate to post-intermediate transition state, causing differing denaturant dependencies and varying slopes for the refolding rates. The pseudo-symmetrical structure of IL-33 not only requires the formation of an intermediate state, it also leads to backtracking in Trefoil 1 due to competition between symmetrical folding nucleation sites. Geometrical frustration, mediated by the geometric accessibility of native contacts, in Trefoil 1 creates bottlenecks on the free energy landscape, constraining it to fold last. As the functional element, Trefoil 1 folding last imparts a unique flexibility to its structure, allowing it to be malleable without affecting the order of the rest of the protein scaffold. Our work shows that geometrical frustration and preservation of the functional element in IL-33 is a strong mediator for the complex experimental kinetics associated with the dominant folding route of this protein.

3.3 Results

3.3.1 The Structure of Interleukin-33.

IL-33, as a member of the Interleukin-1 family of proteins, has the characteristics of a β-trefoil fold. This fold is composed of twelve β-strands organized into three groups of four β-β-β-loop-β units (Trefoil1, Trefoil 2, and Trefoil 3) characterizing the overall fold. Trefoil 1 is composed of β-strands 1 through 4, Trefoil 2 of strands 5 through 8, and Trefoil 3 of strands 9 through 12 (PDB code 2KLL) in Figure 3-1 [22]. The β-strands are organized in a similar manner within each trefoil. Further highlighting the symmetry of this protein, the contact maps for IL-33 show that

the distribution of contacts within each Trefoil is essentially equivalent (208 contacts in Trefoil 1, 221 contacts in Trefoil 2, and 211 contacts in Trefoil 3). However, Trefoil 1 is slightly different in its topology than Trefoils 2 and 3 in that the β-strands are longer, making the sequence of Trefoil 1 overall longer than that of Trefoil 2 and 3. Additionally, Trefoil 1 contains a charged lysine-rich loop between β-strands 3 and 4 and has a large flexible loop bridging β-strands 4 and 5, which constitutes the interface between Trefoil 1 and 2. These unique features of Trefoil 1 make it less symmetrical in nature than the other Trefoils. The more complicated topology and less symmetrical nature of Trefoil 1 versus Trefoil 2 and 3 pairs with the functionality of IL-33: Trefoil 1 acts as the functional element within IL-33, as it has the largest binding interface with its conjugate receptor, ST2 [23].

3.3.2 Experimental characterization shows evidence for the formation of an intermediate state in folding.

In order to characterize the folding free energy landscape of IL-33, we performed thermodynamic and kinetic experiments using tryptophan fluorescence in the presence of both Guanidine Hydrochloride (Gdm-HCl) and Urea. Equilibrium titrations were evaluated (Figure 3-2A and 3-2B, bottom panel) to initially investigate the thermodynamic behavior of IL-33. The equilibrium data fits well to a two-state fit and there is no significant evidence for a thermodynamically stable intermediate state (Table 1). The relative solvent accessible surface area (SASA) can be measured experimentally and calculated as an $m_{D\text{-}N}$ value. The $m_{D\text{-}N}$ values for IL-33 are within experimental error for both Gdm-HCl and Urea (2.72 and 1.48, respectively).

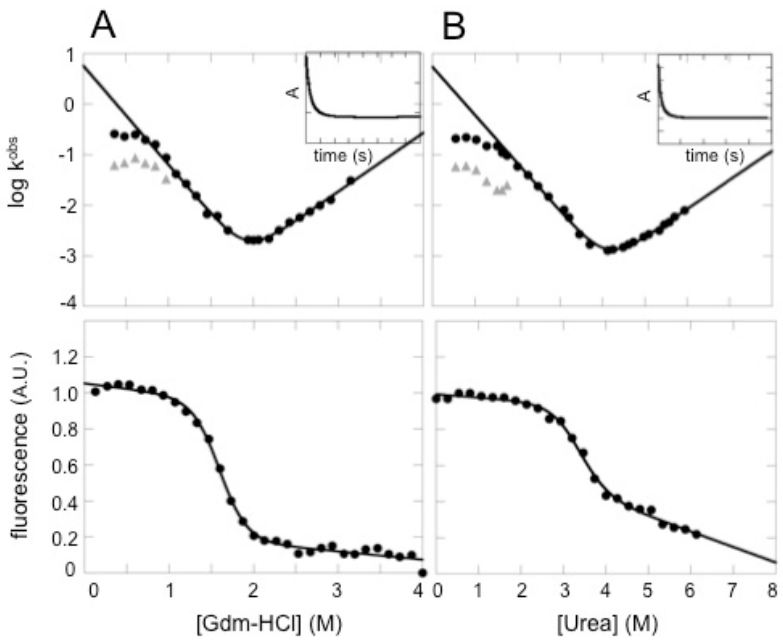

Figure 3-2: Experimental kinetic and equilibrium folding data for Interleukin-33

(*A*) A plot of the observed rate of folding and unfolding as a function of final denaturant concentration compose the chevron plot for IL-33 (upper panel) and the equilibrium titrations for IL-33 as a function of final Gdm-HCl concentration (lower panel). (*B*) The chevron plot for IL-33 (upper panel) and the equilibrium unfolding curve for IL-33 as a function of final urea concentration (lower panel) constructed as in (A). Representative exponential fits of folding for Gdm-HCl and Urea are shown as *inset* panels within the chevron plots. Grey triangles within the chevron plots represent the second fitted phase at strongly native conditions. The data are well fit by a single exponential over most denaturant concentrations except under strongly native conditions (<1 M Gdm-HCl and <2 M urea, respectively) where two exponential fits and a non-denaturant dependent phase appears. The chevron plots are fit with a two-state fit over the linear regime. The equilibrium data are fitted with a two-state fit.

To obtain a detailed view of the folding free energy landscape, the kinetics of IL-33 were evaluated. The chevron plots are representative of the rates of unfolding and refolding (log k_u and log k_f) plotted against the respective denaturant concentrations. IL-33 displays a significant rollover, representing a denaturant independent phase, in strongly refolding conditions in both denaturants along with a second slower phase (Figure 3-2A and 3-2B, top panel). The refolding traces were fit to a three-state fit in the denaturant independent phase (the rollover) and was fit based upon evaluating the residuals of the fits (Figure 3-3). The denaturant dependent phase was fit to a two-state fit. The $m_{D\text{-}N}$ values and stabilities of IL-33 (Table 2) in denaturant are in agreement with the two-state equilibrium fits (Table 1) within experimental error. Although the kinetic data, when fit to the linear regime at $\pm$ 2M from the midpoint, is in agreement with the two-state equilibrium data, the chevron plots have a distinct denaturant independent phase with an additional second phase in strongly native conditions, which is indicative of more complicated folding kinetics.

In order to fully evaluate the source of the rollover in IL-33, a thorough investigation of known and well-established sources that can lead to this kinetic behavior was performed. Protein aggregation, proline isomerization, oligomerization from disulfides were evaluated and showed no significant effects on the refolding rates of IL-33 (Figure 3-4, 27–29). As evidenced by the similar rollover in both Gdm-HCl and Urea, we can conclude that it is not an effect of electrostatics (Figure 3-2A and 3-2B, top panels).

The presence of a rollover in the refolding arm of the chevron plots of IL-33 is indicative of an intermediate state, yet there was no direct kinetic signature from

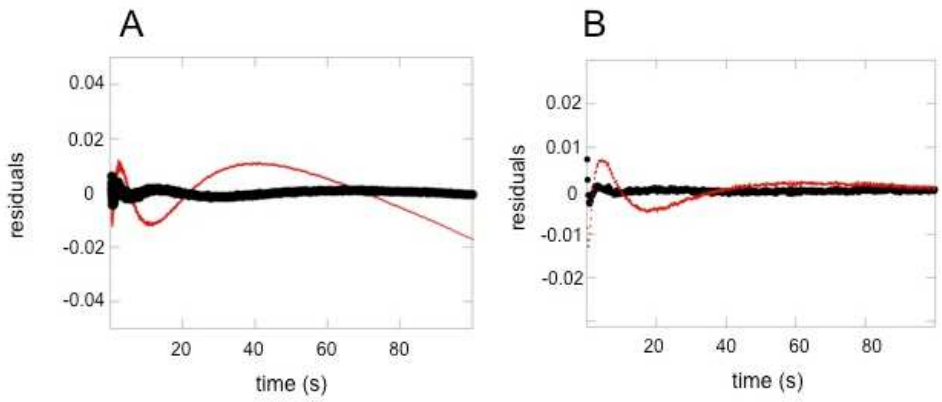

Figure 3-3: Residual fits for the refolding of IL-33 under strongly native conditions.

(A) The residual fit for the refolding of IL-33 from 4 M guanidine into native buffer. Red represents the residuals for the single exponential fit and black represents the residual fit for the double exponential fit. (B) The residual fits for refolding into urea as represented in A.

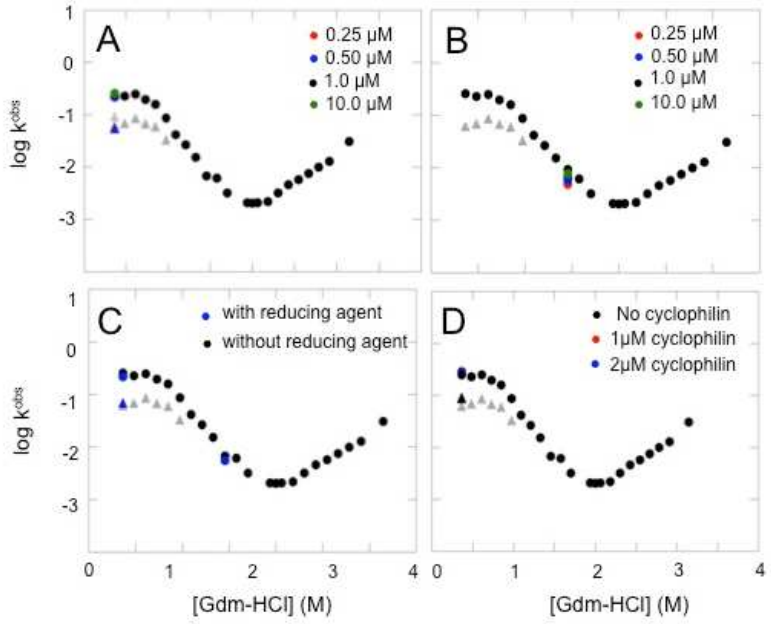

Figure 3-4: Assessing the effect of concentration, prolyl isomerase, and reducing agents on IL-33 kinetics.

(*A*) Represents the effect of protein concentration from a range of 0.25 μM to 10 μM are plotted onto the chevron plot in the denaturant independent phase. There is no significant effect of protein concentration on the rate of refolding. (*B*) Represents the concentration dependence in the same range as seen in panel A but in the denaturant dependent phase. There is no significant effect on the rate of refolding. (*C*) Represents the effect of reducing agent on the refolding rate of IL-33 in the denaturant independent phase. There is no significant effect of denaturant on the rates of refolding. (*D*) Represents the effect of Cyclophilin D on the rate of refolding in both the denaturant and denaturant independent phase. There is no effect of Cyclophilin D from the range of 0 to 2 μM on the rate of folding.

Table 3-1: Equilibrium Data

	$m_{D\text{-}N}$ (M^{-1})	MP	$\Delta G_{D\text{-}N}$
GdmCl	2.72 ± 0.30	1.60 ± 0.02	5.92
Urea	1.48 ± 0.24	3.28 ± 0.06	6.60

* Data for Urea and GdmCl were both fit with a two state fit

**Errors are reported as standard deviation

Table 3-2: Kinetic Data

	log k_f^{H2O}	m_f (M^{-1})	log k_u^{H2O}	m_u (M^{-1})	β_T	MP (M)	$m_{D\text{-}N}$ (M^{-1})	log k_{H2O}	$\Delta G_{D\text{-}N}$
GdmCl	0.50 ± 0.7	-1.78	-5.20 ± 0.1	1.16	0.38	1.93 ± 0.0	3.11	5.92	8.05
Urea	0.73 ± 0.1	-0.95	-5.50 ± 0.3	0.57	0.38	4.09 ± 0.1	1.52	6.23	8.40

* Data for Urea and GdmCl were both fit at ± 2M of the midpoint

**Errors are reported as standard deviation (SD

intrinsic fluorescence. To directly probe for the formation of an intermediate state, IL-33 was refolded in the presence of ANS (1-Anilinonaphtalene-8-Sulfonic Acid), a dye known to bind transient hydrophobic patches in the refolding of proteins [27]. IL-33 was refolded with ANS into strongly native conditions from both Gdm-HCl and Urea (Figure 3-5). Refolding traces show the formation of an intermediate state populated in the first milliseconds of folding (inset, Figure 3-5), as evidenced by the increase in fluorescent signal of ANS followed by signal decay as IL-33 folds to the native state and dye is released. The rate of intermediate species formation became slower as the folding reaction moved closer to the midpoint (Figure 3-6). Collectively, the experimental data shows that IL-33 folds with three-state kinetics mediated by the presence of an intermediate state. Because this intermediate state is difficult to isolate experimentally, coarse-grained structure-based simulations were used to provide a more detailed view of the folding free energy landscape on a molecular level.

3.3.3 Characterizing the details of the Interleukin-33 folding landscape using structure based simulations.

Coarse-grained C-alpha (CA) structure-based models (SBM) were used to determine the molecular mechanisms behind the kinetics observed from experiments. SBM are well-suited for discerning topological aspects of the folding landscape [28]. The energy landscape of the SBM was exhaustively sampled to obtain thermodynamic information. In Figure 3-7, the free energy landscape $F(Q_{CA})/k_BT$ is plotted as a function of the overall folding reaction coordinate Q_{CA}, representing the fraction of native contacts formed. At the folding midpoint where the probability of populating the denatured and native state is equal ($T=T_f$), IL-33 has a broad barrier between $0.3< Q_{CA} <0.8$ that contains a low populated intermediate with a depth of $< 1\ k_BT$. On either

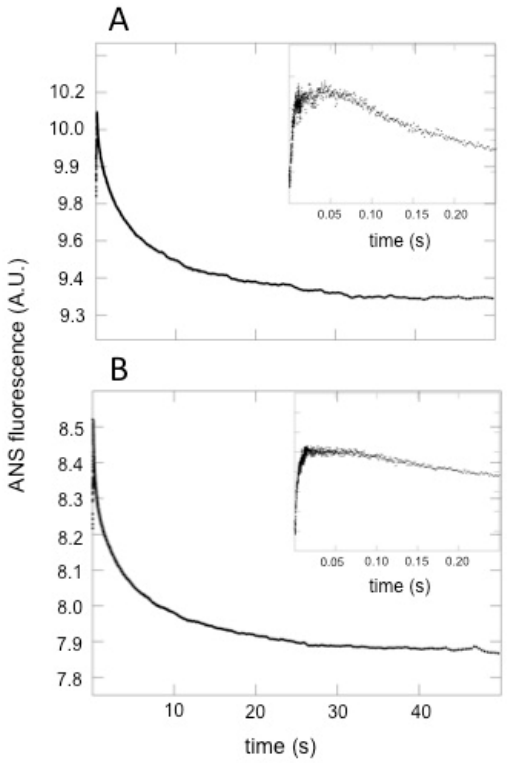

Figure 3-5: ANS binding kinetics to Interleukin-33 for the refolding reaction.

(*A*) A plot of the refolding of IL-33 in the presence of 1-Anilinonaphtalene-8-Sulfonic Acid (ANS) in 0.36 M Gdm-HCl. (*B*) Refolding of IL-33 in the presence of ANS in 0.36 M urea. Fluorescence emission of ANS was monitored upon refolding of IL-33 and shows an initial binding and release of dye (*inset*), indicative of the formation of an intermediate state.

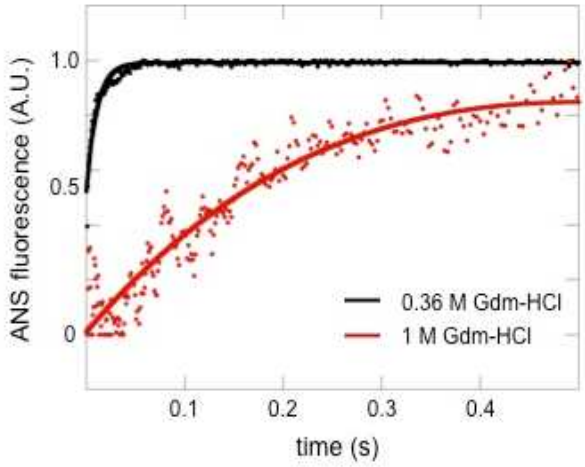

Figure 3-6: Denaturant dependent formation of the intermediate state in IL-33.

The fits of the intermediate formation in the presence of ANS are plotted as a function of time (s) and fluorescence A.U. of the dye. Black represents IL-33 refolded into 0.36 M Gdm-HCl (the rollover region) while red represents IL-33 folded into 1 M Gdm-HCl (the linear regime). Both were fit to a double exponential fit. The formation of the intermediate is slowed significantly from a dominant rate of log k_f = 0.09 when refolded into 0.36M Gdm-HCl to a rate of log k_f = 0.29 when refolded into 1M Gdm-HCl.

side of the intermediate state are two transition states (TS), TS_1 and TS_2 at a Q_{CA} of 0.4 and 0.7, respectively. At the folding midpoint, TS_2 is the rate-limiting transition with a barrier of 11 k_BT, whereas TS_1 has a barrier of 9 k_BT.

The structural pseudo-symmetry is broken in the folding mechanism, leading to a specific folding order for the trefoils. The formation of the three trefoils are separated in Figure 3-7B. The two dimensional free energy landscapes show that Trefoil 1 folds later than Trefoils 2 and 3. Comparing these 2D landscapes with the 1D landscape shows that TS_1 consists of an intermediate state composed of Trefoils 2 and 3 mostly formed while Trefoil 1 remains unfolded. TS_2 involves the final folding of Trefoil 1 onto the fully formed nucleus of Trefoil 2 and 3. Taken together, the data indicates that this symmetry breaking between the trefoils leads to an intermediate state characterized by folded Trefoils 2 and 3 and an unfolded Trefoil 1. Analyzing the structural components in more detail, Figure 3-8 shows the formation of each β-strand interface ($q_{<segment>}$) relative to the overall foldedness (Q_{CA}) of the protein. The contacts in Trefoil 2 and 3 grow monotonically with Q_{CA}, with pronounced increases near a Q_{CA} of 0.4, corresponding to TS_1. In contrast, native contact formation in Trefoil 1 grows until a Q_{CA} of 0.3, but then the contacts are broken before reaching TS_1. These broken contacts in Trefoil 1 are then reformed at a Q_{CA} of 0.6. Within Trefoil 1 there is complete backtracking of the contacts formed between β-strands 1 and 4 and partial backtracking between β-strands 2 and 3 (Figure 3-8). Even though these data are ensemble averaged, the landscape indicates that if Trefoil 1 folds early it tends to backtrack in order to follow the lowest free-energy path to the folded protein. Interestingly, since unfolding occurs on the same landscape, this means that Trefoil 1, the functional binding element, is the first to unfold coming from the native basin.

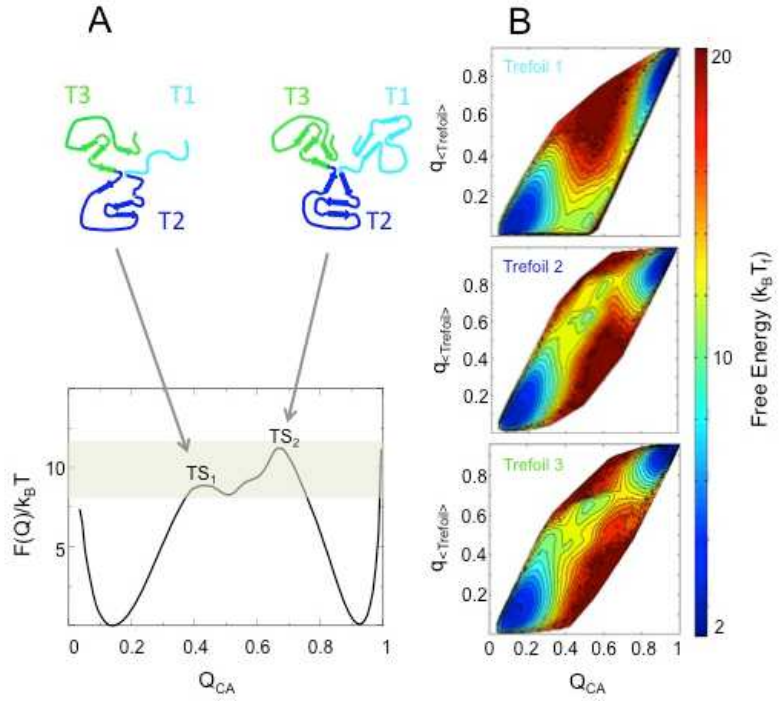

Figure 3-7: Free energy profile of Interleukin-33 and its individual trefoils determined from structure-based folding models.

(*A*) A plot of the energy as a function of Q_{CA}. The denatured basin is populated at a Q_{CA} of 0.2, the native basin is populated at a Q_{CA} of 0.9 and the transition region has two transition states (TS) at Q_{CA} of 0.4 (TS_1) and at Q_{CA} of 0.8 (TS_2). The shaded region highlights the two transition states and the intermediate state between them. Representative structures of the species present at each of the two transition regions are shown above the free energy profile plot. (*B*) Two dimensional free-energy landscapes as a function of $q_{<trefoil>}$ and Q_{CA} for Trefoils 1, 2, and 3 unit for Trefoil 1, 2 and 3 are represented from top to bottom respectively. $q_{<Trefoil>}$ includes all native contacts that a given trefoil can form and Q_{CA} includes all native contacts. The occupancy of states is represented by a color scale where red represents lowest occupancy and blue represents highest occupancy.

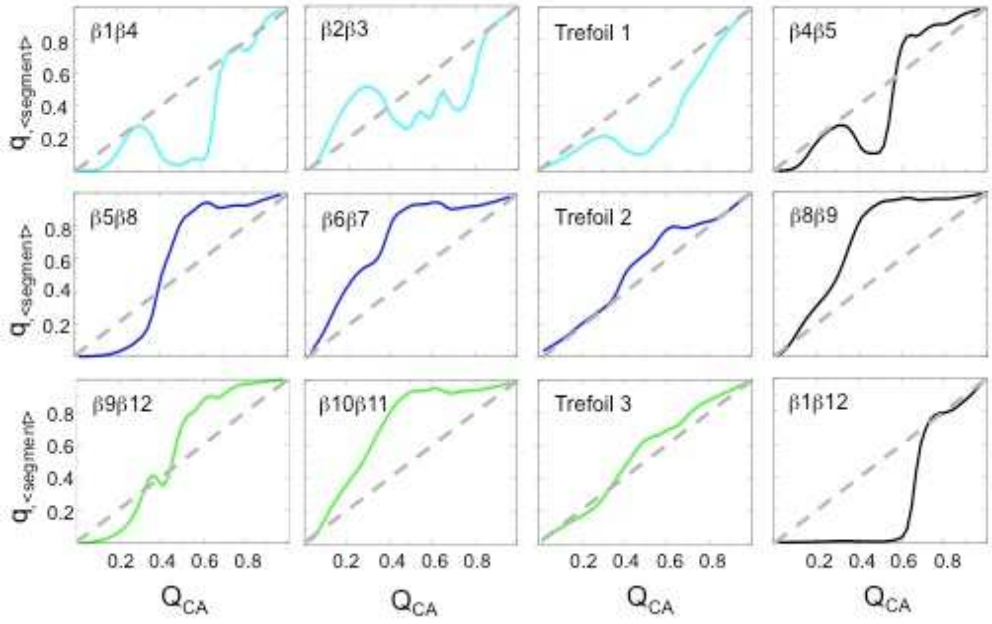
β1β4
β2β3
Trefoil 1
β4β5
β5β8
β6β7
Trefoil 2
β8β9
β9β12
β10β11
Trefoil 3
β1β12
q
0.2
0.4
0.6
0.8
QCA

Figure 3-8: The formation of individual contacts represented for each individual trefoil in IL-33.

Plots of $q_{<segment>}$ versus Q_{CA} for interactions between individual strands or within trefoil units are represented. Each plot represents the formation of contacts for the given strands ($q_{<segment>}$) versus the formation of all native contacts within IL-33 (Q_{CA}). The plots for contacts located in Trefoil 1 are highlighted in cyan, Trefoil 2 in blue, and Trefoil 3 in green. The contacts formed between trefoils are highlighted in black. The dashed lines compare the overall foldedness of the protein.

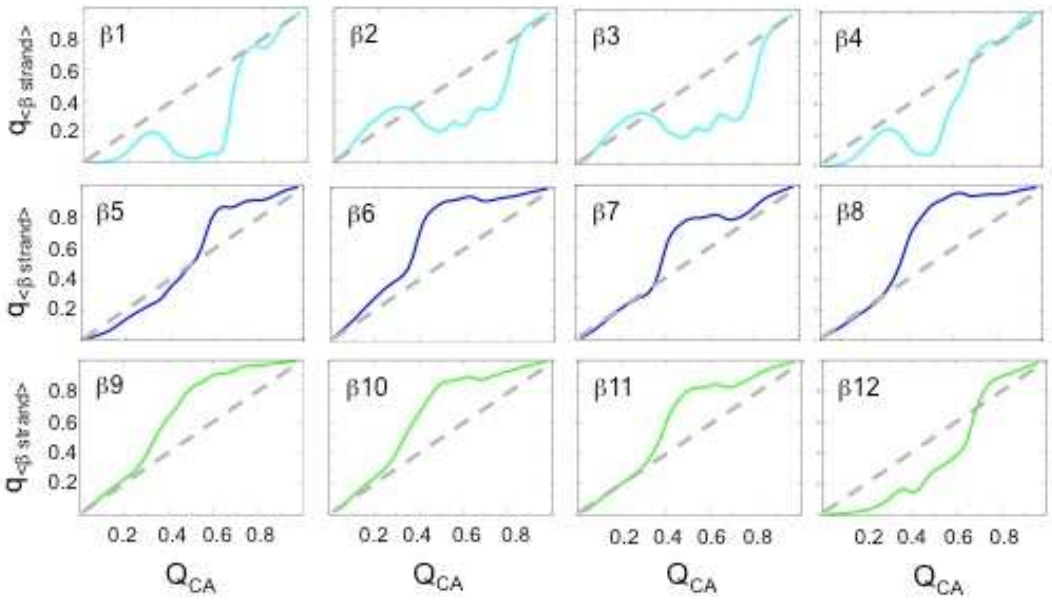

Figure 3-9: The formation of contacts represented for each individual β-strand in IL-33.

Plots of $q_{<\beta\text{-strand}>}$ versus Q_{CA} for interactions between individual β-strands are represented. Each plot represents the formation of contacts for the given strands ($q_{<\beta\text{-strand}>}$) versus the formation of all native contacts within IL-33 (Q_{CA}). The plots for contacts located in Trefoil 1 are highlighted in cyan, Trefoil 2 in blue, and Trefoil 3 in green. The dashed lines represent constant growth of elements along the free energy landscape. The β-strands composing Trefoil 3 (β-strands 9 through 11) are the first to form, with pronounced growth at a Q_{CA} of 0.3. The β-strands composing Trefoil 2 (β-strands 5 through 8) form almost concurrently with the β-strands composing Trefoil 3. The β-strands composing Trefoil 1 (β-strands 1 through 4) form and are then subsequently broken before a Q_{CA} of 0.4. The β-strands then pause their growth until after a Q_{CA} of 0.6, indicating that the β-strands in Trefoil 1 will not fold until the β-strands from Trefoils 2 and 3 are nearly folded to completion creating a scaffold for Trefoil 1. The final step in folding is the formation of contacts in β-strand 12.

Figure 3-9 shows the formation of each individual element in IL-33 and displays similar trends to those shown in Figure 3-8. This architecture could provide malleability to the binding interface, allowing it to be responsive to receptor binding without significantly perturbing the rest of the stable protein scaffold.

3.4 Discussion

Due to IL-33's prevalence in the pathology of multiple disease states [29–31], our understanding of the influence of the functional region on the overall folding landscape is useful in aiding future drug design. As a member of the β-trefoil family, IL-33 has a unique trefoil topology represented by a pseudo-symmetrical fold. The binding of the Trefoil 1 interface to ST2 is the critical initiating step in the formation of an active signaling heterotrimeric complex, therefore Trefoil 1 represents the functionally critical element of IL-33 [15]. Understanding the folding route of IL-33, specifically the folding of Trefoil 1, can highlight possible underlying sources of disease states by identifying folding events critical for the formation of the functional native state.

In evaluating the folding free energy landscape of IL-33, the chevron plots provide initial evidence for the formation of an intermediate state due to the presence of a rollover and a second slower phase in strongly native conditions. However, the rollover only provides indirect evidence for the presence of an intermediate state. The intermediate state for IL-33 proved to be difficult to isolate using intrinsic fluorescence. The refolding assay of IL-33 with ANS showed binding and release of dye in the millisecond timescales (Figure 3-5, inset), providing direct evidence for the first time of an intermediate state in the folding of IL-33. Hisactophilin and hFGF-1, two other β-trefoil proteins, exhibit rollover in their chevron plots in strongly native conditions

[32,33] attributed to an intermediate state. Additionally, the folding of the β-trefoil cytokine Interleukin-1β has also been found to fold through a three- state mechanism [20,21]. In gaining an understanding of how the formation of an intermediate state dictates the steps along the folding route and the interplay of functional elements, we can better understand the functional motions inherent to the activity of native IL-33.

3.4.1 Is the observed kinetic rollover caused by a moving rate-limiting transition state?

As seen in the simulation-derived free-energy profile of IL-33 (Figure 3-7A), there are two observable transition states, TS_1 and TS_2, along the folding route. These transition states are positioned on either side of the intermediate species. Movement of the rate-limiting TS is a canonical case of chevron rollover, as seen in U1A [34]. As temperature increases in the simulation (a proxy for denaturant in the experiment), a Hammond effect causes the rate-limiting TS to move from having little structure, a Q_{CA} of 0.35, to significant structure, a Q_{CA} of 0.65 (Figure 3-11). A transition state with low Q_{CA} means it has little change in solvent accessible surface area relative to the denatured ensemble (**Δ**SASA). Since denaturant lowers the free energy in a configuration roughly proportional to **Δ**SASA, denaturant will have less effect on an early transition state, leading to a smaller slope in the chevron plots (Figure 3-2A and 3-2B, top panel). Thus, the rollover at low denaturant is consistent with the energy landscape calculated in the simulations (Figure 3-11). The formation of the intermediate state is slowed as the protein is refolded closer to the midpoint, which we hypothesize could be a result of the switch of the rate-limiting step from TS_1 to TS_2, allowing for more accumulation of the intermediate state in the transition region as the folding reaction progresses toward the midpoint (Figure 3-6).

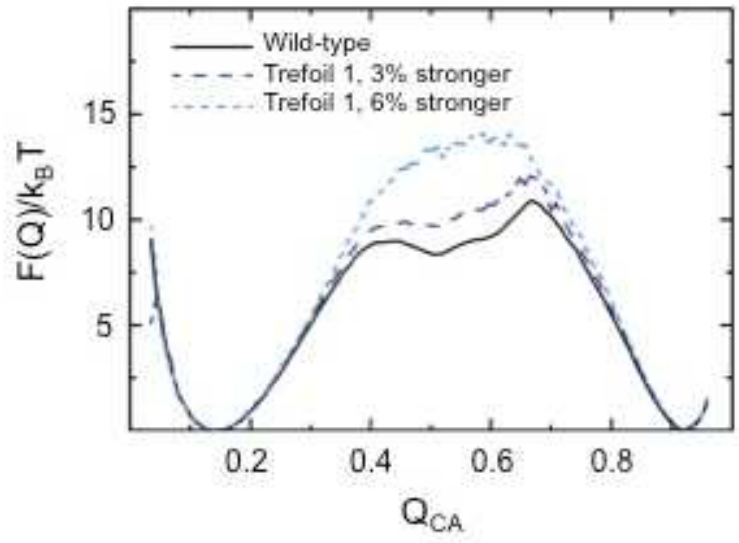

Figure 3-10: Inducing Cooperative folding through tightening contacts creates an unfavorable two-state folding behavior.

The free energy profile is represented as Q_{CA} versus $F(Q_{CA})/k_BT$. The wild-type IL-33 is represented in black and IL-33 with tighter contacts introduced in Trefoil 1 ($\varepsilon^{i,j} = \varepsilon*0.03$) is represented in dark blue while ($\varepsilon^{i,j} = \varepsilon*0.06$) is represented in light blue. IL-33 with tightened contacts in Trefoil 1 induces two-state cooperative folding as evidenced by the presence of only one transition state. Additionally, this two- state folding is unfavorable with respect to native three-state folding as evidenced by the larger barrier to fold to the native state, slowing the folding process.

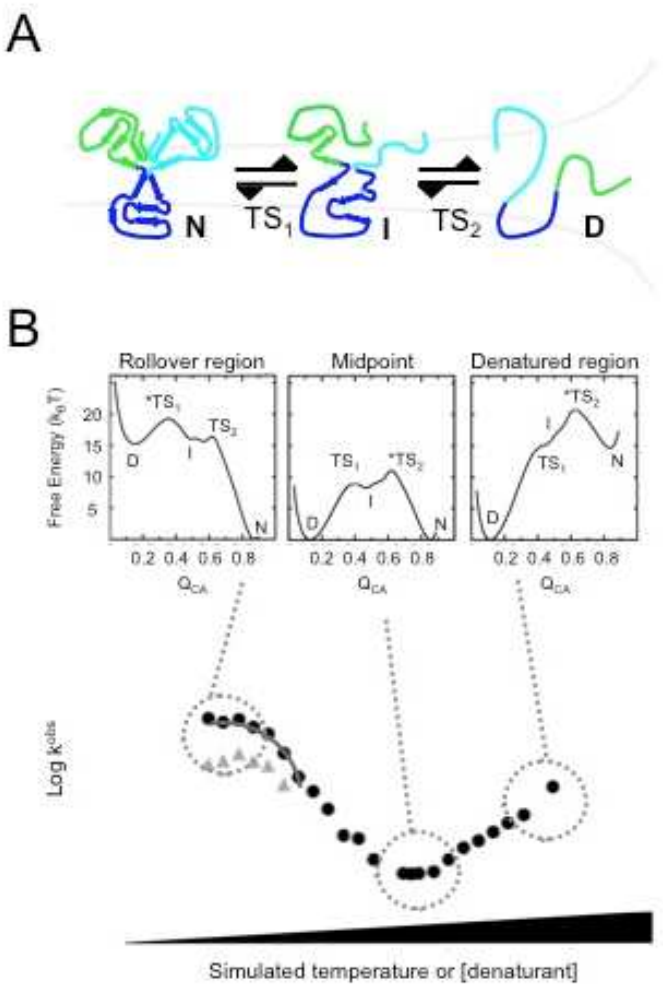

Figure 3-11: Position of rate-limiting transition state changes with protein stability.

(*A*) Representative schematic of the folding landscape of IL-33. From left to right represents the transition from native to denatured states of the protein as seen in the presence of low to high simulated temperature or experimental denaturant concentration. The intermediate state is populated on the folding route and is composed of folded Trefoils 2 and 3 while Trefoil 1 remains unfolded. (*B*) Geometric frustration on the folding landscape of IL-33 creates two bottlenecks to folding (TS_1 and TS_2) separated by an intermediate ensemble. Under folding conditions, TS_1 is the rate-limiting step, while closer to the midpoint and beyond, TS_2 is rate-limiting. The different positions of these transition states along the reaction coordinate can be differentially effected by denaturant, as the denaturant effect is strongly related to the change in solvent accessible surface area (SASA). The denatured to TS_2 transition has a larger ΔSASA than the denatured to TS_1, meaning that denaturant should have a larger effect (larger slope) in the regime where TS_2 is rate-limiting.

3.4.2 The implications of symmetry on the folding landscape.

Symmetry in proteins often leads to complicated folding kinetics and a geometrically-frustrated landscape [21,35]. The effects of pseudo-symmetry on the β-trefoil fold [20,36,37] has been studied in terms of both folding and function [28,38,39]. The balance between minimizing the frustration on the folding free energy landscape and preserving the functional regions of a protein is often complex and varied, especially concerning symmetric proteins [40–43]. Satisfying functional constraints can lead to geometric frustration that decreases the folding rate. In the case of IL-33, geometrical constraints and the interplay of the individual folding elements contribute to both the intermediate species and geometrical frustration in the folding of the functional element of the protein, Trefoil 1.

3.4.3 Backtracking of the functional element, Trefoil 1, highlights the geometrical frustration.

Backtracking is recognized as a phenomenon in protein folding wherein native contacts are made, broken, and reformed in order to reestablish productive folding along the folding route [44,45]. Here, we showed that backtracking in IL-33 is observed thermodynamically, in that the symmetry of IL-33 allows all trefoils to locally fold in the denatured state, but only Trefoils 2 and 3 are formed in TS_1. Therefore, if Trefoil 1 folds and forms native contacts early it tends to backtrack in order for IL-33 to follow the lowest free-energy route to the folded native state. The unfolding of Trefoil 1 does not have an entropic barrier (Figure 3-7B top), meaning that backtracking is fast relative to the overall folding. Thus backtracking can be eliminated as a direct

cause of the rollover seen in the chevron plots and the second slower phase in strongly native conditions.

Revisiting the topology of IL-33, the trefoils constitute three individual folding elements that combine to form the final native structure. The intermediate state is composed of folded Trefoils 2 and 3, which act as a scaffold for the folding and docking of Trefoil 1 in the final step to the native state. To test our hypothesis that Trefoil 1 creates geometric frustration, all contacts were strengthened in Trefoil 1 to evaluate how artificially enhancing the early folding of Trefoil 1 influenced the folding landscape (Figure 3-10). With a small perturbation (increasing the strength of contacts involving Trefoil 1 by 6%), IL-33 folds through a two-state route, however, the barrier increases dramatically. The increase of the folding barrier demonstrates that the folding of Trefoil 1 cooperatively with Trefoils 2 and 3 is unfavorable and much slower. Breaking the symmetry in IL-33 allows it to fold non-cooperatively and introduces backtracking and an intermediate state on the free energy landscape. Although this type of landscape often leads to slower folding of a protein, in the case of IL-33, the modular folding reduces the folding barrier allowing for both faster folding and preservation of the functional contacts.

3.5 Conclusions.

Through a combination of experimental and theory-based work we present that the modular folding and the functionality of IL-33 are intricately related. The intermediate state of IL-33 is characterized by the folding of a stable nucleus composed of Trefoils 2 and 3 whose formation allow for the subsequent folding of Trefoil 1 as the final step to reach the native state. Even though all trefoils share similar contacts, similar structural elements, and similar geometry, Trefoil 1 folds

differently and has differential effects on the functionality of the protein. Thus Trefoil 1, can respond, function, and be malleable without unfolding the protein. This further implies that Trefoil 1 can respond to changes caused by engaging the binding interface of its conjugate receptor without dramatically altering the rest of the protein scaffold.

3.6 Methods

3.6.1 Sample Preparation.

IL-33 was expressed in a pET-24 vector and grown in BL21 cells. Initial purification was achieved using osmotic shock cell lysis and final purity was attained through anion exchange chromatography. All experiments were performed in 10 mM MES, 90 mM NaCl, 1 mM EDTA at pH 6.50.

3.6.2 Equilibrium Measurements.

Equilibrium titrations were collected on a FluoroMax-4 at 25°C. Excitation was at 280 nm and emission was monitored between 350 and 450 nm. The final protein concentration was 5 μM.

3.6.3 Kinetic Measurements.

Refolding kinetics were performed on a PiStar-180 stopped flow fluorometer at 25°C. Excitation was at 280 nm and data was collected using a 320 nm cut off filter. Final protein concentration was 1 μM. Unfolding kinetics were monitored on a FluoroMax-4 using manual mixing with anti-photo bleaching due to hyper fluorescence upon unfolding. Folding was fit to a three-state fit at low denaturant concentrations

(below 1.3M for Gdm-HCl and below 1.8M for Urea) and all other date was fit to a two-state fit. We evaluated the chevron plots with a linear fit at ±2M from the midpoint to minimize denaturant dependence. ANS was excited at 365 nm and data was collected with a 420 nm cut off filter. The denatured protein was incubated with ANS at a final concentration of 100 μM and refolded into 0.36 M denaturant and 1 M denaturant, respectively.

3.6.4 Structure Based Models (SBM).

In this study we used a structure-based Cα model to investigate the folding of IL-33 [46,47]. Each residue is assigned as a single bead. These beads are assigned attractive interactions based upon their relative proximity to one another in the native state. Native contacts are determined using a shadow map [48,49], which relates relative distances and possible steric interferences to extrapolate a reasonable cutoff for native contacts. The basic Hamiltonian is:

$$V(r_{ij}) = \sum_{\text{bonds}\{ij\}} k_b(r_{ij} - r_{ij}^N)^2 + \sum_{\text{angles}\{ijk\}} k_a(\theta_{ijk} - \theta_{ijk}^N)^2 +$$
$$\sum_{\text{dihedrals}\{ijkl\}} k_d^1\left[1 - \cos(\varphi_{ijkl} - \varphi_{ijkl}^N)^2\right] + k_d^3\left[1 - \cos(3(\varphi_{ijkl} - \varphi_{ijkl}^N))^2\right] +$$
$$\sum_{\text{native contacts}} R(r_{ij}) + G_{ij}(r_{ij}) + R(r_{ij})G_{ij}(r_{ij}) + \sum_{\text{nonnative contacts}} R(r_{ij})$$
$$R(r_{ij}) = \varepsilon\left(\frac{\sigma}{r_{ij}}\right)^{12}$$
$$G_{ij}(r_{ij}) = -\varepsilon\exp\left(-\left(r_{ij} - r_{ij}^N\right)^2 \Big/ 2w^2\right)$$

The webserver SMOG (http://smog-server.org/) was used to create the input files for our simulations. All simulations were performed using GROMACS 4.5.3 software package [50]. Integration steps of t =0.005 were used in all simulations and

all results are presented with reduced units. For sufficient sampling of the transition state, IL-33 was simulated using umbrella sampling along the coordinate Q_{CA} [51]. In order to analyze the correct folding mechanism, (to create $F(Q_{CA})$ for the free energy profiles) we used the Weighted Histogram Analysis Method (WHAM) [52-57]. The results are reported as units of ε, representative of the energy gain in forming a native contact. In the case of the tightened contacts in Trefoil 1, the contacts between *i* and *j* were tightened by $\varepsilon^{i,j} = \varepsilon^{*}1.03$ and 1.06 and the rest of the contacts were homogenously reduced to keep the total stabilizing energy constant.

Chapter 3, in part, is a reprint of the material as it appears in "Geometrical frustration in Interleukin-33 decouples the dynamics of the functional element from the folding transition state ensemble" Fisher, Kaitlin M., Haglund, Ellinor, Noel, Jeffrey K., Hailey, Kendra L., Onuchic, Jose N., Jennings, Patricia A which has been submitted to JMB for publication. The dissertation author was the primary author/investigator of this paper.

3.7 References

[1] Buckley CD, Gilroy DW, Serhan CN, Stockinger B, Tak PP. The resolution of inflammation. Nat Rev Immunol 2013;13:59–66.

[2] Lefrancais E, Roga S, Gautier V, Gonzalez-de-Peredo a., Monsarrat B, Girard J-P, Cayrol C. IL-33 is processed into mature bioactive forms by neutrophil elastase and cathepsin G. Proc Natl Acad Sci 2012;109:1673–8.

[3] Villarreal DO, Weiner DB. Interleukin 33: a switch-hitting cytokine. Curr Opin Immunol 2014;28:102–6.

[4] Lloyd CM. IL-33 family members and asthma - bridging innate and adaptive immune responses. Curr Opin Immunol 2010;22:800–6.

[5] Willems S, Hoefer I, Pasterkamp G. The role of the Interleukin 1 receptor-like 1 (ST2) and Interleukin-33 pathway in cardiovascular disease and cardiovascular risk assessment. Minerva Med 2012;103:513–23.

[6] Sun P, Ben Q, Tu S, Dong W, Qi X, Wu Y. Serum Interleukin-33 levels in patients with gastric cancer. Dig Dis Sci 2011;56:3596–601.

[7] Rong C, Hu W, Wu F-R, Chen F-H. Targeting interleukin-33 in rheumatoid arthritis. Rheumatol Int 2013;33:1377–8.

[8] Mizutani N, Nabe T, Yoshino S. Interleukin-33 and alveolar macrophages contribute to the mechanisms underlying the exacerbation of IgE-mediated airway inflammation and remodelling in mice. Immunology 2013;139:205–18.

[9] Pei C, Barbour M, Fairlie-Clarke KJ, Allan D, Mu R, Jiang HR. Emerging role of interleukin-33 in autoimmune diseases. Immunology 2014;141:9–17.

[10] Priestle JP, Schär HP, Grütter MG. Crystal structure of the cytokine interleukin-1 beta. EMBO J 1988;7:339–43.

[11] Stockman BJ, Scahill TA, Strakalaitis NA, Brunner DP, Yem AW, Deibel MR. Solution structure of human interleukin-1 receptor antagonist protein. FEBS Lett 1994;349:79–83.

[12] Habazettl J, Gondol D, Wiltscheck R, Otlewski J, Schleicher M, Holak TA. Structure of hisactophilin is similar to interleukin-1 beta and fibroblast growth factor. Nature 1992;359:855–8.

[13] Huntington J a., Esmon CT. The Molecular Basis of Thrombin Allostery Revealed by a 1.8 Å Structure of the “Slow” Form. Structure 2003;11:469–79.

[14] Hailey KL, Capraro DT, Barkho S, Jennings P a. Allosteric switching of agonist/antagonist activity by a single point mutation in the interluekin-1 receptor antagonist, IL-1Ra. J Mol Biol 2013;425:2382–92.

[15] Lingel A, Weiss TM, Niebuhr M, Pan B, Appleton B a., Wiesmann C, Bazan JF, Fairbrother WJ. Structure of IL-33 and Its Interaction with the ST2 and IL-1RAcP Receptors-Insight into Heterotrimeric IL-1 Signaling Complexes. Structure 2009;17:1398–410.

[16] Liu X, Hammel M, He Y, Tainer J a, Jeng U-S, Zhang L, Wang S, Wang X. Structural insights into the interaction of IL-33 with its receptors. Proc Natl Acad Sci U S A 2013;110:14918–23.

[17] Schug A, Onuchic JN. From protein folding to protein function and biomolecular binding by energy landscape theory. Curr Opin Pharmacol 2010;10:709–14.

[18] Alan F. Structure and mechanism in protein science, a guide to enzyme catalysis and protein folding. Lavoisier.fr 1999.

[19] Kern G, Handel T, Marqusee S. Characterization of a folding intermediate from HIV-1 ribonuclease H. Protein Sci 1998;7:2164–74.

[20] Heidary DK, Gross LA, Roy M, Jennings PA. Evidence for an obligatory intermediate in the folding of interleukin-1 beta. Nat Struct Biol 1997;4:725–31.

[21] Finke JM, Jennings P a. Interleukin-1?? folding between pH 5 and 7: Experimental evidence for three-state folding behavior and robust transition state positions late in folding. Biochemistry 2002;41:15056–67.

[22] Went HM, Benitez-Cardoza CG, Jackson SE. Is an intermediate state populated on the folding pathway of ubiquitin? FEBS Lett 2004;567:333–8.

[23] Kathuria S V, Day IJ, Wallace L a, Matthews CR. Kinetic traps in the folding of beta alpha-repeat proteins: CheY initially misfolds before accessing the native conformation. J Mol Biol 2008;382:467–84.

[24] Silow M, Oliveberg M. Transient aggregates in protein folding are easily mistaken for folding intermediates. Proc Natl Acad Sci U S A 1997;94:6084–6.

[25] Otzen DE, Oliveberg M. Salt-induced detour through compact regions of the protein folding landscape 1999;1999:0–5.

[26] Texter FL, Spencer DB, Rosenstein R, Matthews CR. Intramolecular catalysis of a proline isomerization reaction in the folding of dihydrofolate reductase. Biochemistry 1992;31:5687–91.

[27] Lingel A, Fairbrother WJ. NMR assignments of the human cytokine interleukin-33. Biomol NMR Assign 2009;3:223–5.

[28] Liu X, Hammel M, He Y, Tainer J a, Jeng U-S, Zhang L, Wang S, Wang X. Structural insights into the interaction of IL-33 with its receptors. Proc Natl Acad Sci U S A 2013;110:14918–23.

[29] Gittelman MS, Matthews CR. Folding and stability of trp aporepressor from Escherichia coli. Biochemistry 1990;29:7011–20.

[30] Mann CJ, Shao X, Matthews CR. Characterization of the slow folding reactions of trp aporepressor from Escherichia coli by mutational analysis of prolines and catalysis by a peptidyl-prolyl isomerase. Biochemistry 1995;34:14573–80.

[31] Jones BE, Jennings PA, Pierre RA, Matthews CR. Development of Nonpolar Surfaces in the Folding of Escherichia coli Dihydrofolate Binding ? Reductase Detected by 1 -Anilinonaphthalene-8-sulfonate 1994:15250–8.

[32] Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 2000;298:937–53.

[33] Kurowska-Stolarska M, Hueber A, Stolarski B, McInnes IB. Interleukin-33: A novel mediator with a role in distinct disease pathologies. J. Intern. Med., vol. 269, 2011, p. 29–35.

[34] Lefrançais E, Cayrol C. Mechanisms of IL-33 processing and secretion: Differences and similarities between IL-1 family members. Eur Cytokine Netw 2012;23:120–7.

[35] Pushparaj PN, Li D, Komai-Koma M, Guabiraba R, Alexander J, Mcsharry C, Xu D. Interleukin-33 exacerbates acute colitis via interleukin-4 in mice. Immunology 2013;140:70–7.

[36] Liu C, Gaspar JOEA, Wong HJ, Meiering EM. Conserved and nonconserved features of the folding pathway of hisactophilin , a ␣ -trefoil protein 2002:669–79.

[37] Samuel D, Kumar TKS, Balamurugan K, Lin WY, Chin DH, Yu C. Structural Events during the Refolding of an All ??-Sheet Protein. J Biol Chem 2001;276:4134–41.

[38] Otzen DE, Kristensen O, Proctor M, Oliveberg M. Structural changes in the transition state of protein folding: Alternative interpretations of curved chevron plots. Biochemistry 1999;38:6499–511.

[39] Plotkin SS, Onuchic JN. Investigation of routes and funnels in protein folding by free energy functional methods. Proc Natl Acad Sci 2000;97:6509–14.

[40] Onuchic N, Jennings PA, Capraro DT, Roy M. Backtracking on the folding landscape of the beta -trefoil 2008.

[41] Longo LM, Kumru OS, Middaugh CR, Blaber M. Article Evolution and Design of Protein Structure by Folding Nucleus Symmetric Expansion 2014:1–8.

[42] Broom A, Doxey AC, Lobsanov YD, Berthin LG, Rose DR, Howell PL, McConkey BJ, Meiering EM. Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein. Structure 2012;20:161–71.

[43] Smith MTJ, Meissner J, Esmonde S, Wong HJ, Meiering EM. Energetics and mechanisms of folding and flipping the myristoyl switch in the β -trefoil protein , hisactophilin 2010.

[44] Baxter EL, Jennings P a, Onuchic JN. Interdomain communication revealed in the diabetes drug target mitoNEET. Proc Natl Acad Sci U S A 2011;108:5266–71.

[45] Noel JK, Schug A, Verma A, Wenzel W, Garcia AE, Onuchic JN. Mirror images as naturally competing conformations in protein folding. J Phys Chem B 2012;116:6880–8.

[46] Andrews BT, Capraro DT, Sulkowska JI, Onuchic JN, Jennings PA. Hysteresis as a marker for complex, overlapping landscapes in proteins. J Phys Chem Lett 2013;4:180–8.

[47] Gosavi S. Understanding the folding-function tradeoff in proteins. PLoS One 2013;8:e61222.

[48] Tripathi S, Makhatadze GI, Garcia AE. Backtracking due to residual structure in the unfolded state changes the folding of the third fibronectin type III domain from tenascin-C. J Phys Chem B 2013;117:800–10.

[49] Hills RD, Brooks CL. Subdomain Competition, Cooperativity, and Topological Frustration in the Folding of CheY. J Mol Biol 2008;382:485–95.

[50] Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, Onuchic JN. An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. Proteins Struct Funct Bioinforma 2009;75:430–41.

[51] Andrews BT, Gosavi S, Finke JM, Onuchic JN, Jennings PA. The dual-basin landscape in GFP folding. Proc Natl Acad Sci U S A 2008;105:12283–8.

[52] Noel JK, Whitford PC, Onuchic JN. The shadow map: A general contact definition for capturing the dynamics of biomolecular folding and function. J Phys Chem B 2012;116:8692–702.

[53] Noel JK, Whitford PC, Sanbonmatsu KY, Onuchic JN. SMOG@ctbp: Simplified deployment of structure-based models in GROMACS. Nucleic Acids Res 2010;38.

[54] Van der Spoel D, Lindahl E, Hess B, Kutzner C, van Buuren AR, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM, Feenstra KA, van Drunen R, Berendsen HJC. Gromacs User Manual version 4.0. Optimization 2005:308.

[55] Lammert H, Noel JK, Onuchic JN. The Dominant Folding Route Minimizes Backbone Distortion in SH3. PLoS Comput Biol 2012;8.

[56] Ferrenberg A, Swendsen R. New Monte Carlo technique for studying phase transitions. Phys Rev Lett 1988;61:2635–8.

[57] Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data analysis. Phys Rev Lett 1989;63:1195–8.

Chapter 4 : An Introduction to Protein Dynamics

4.1 Beyond the Structure: How Dynamics Influence the Function of Proteins.

A primary amino acid sequence in of itself does not impart functionality to a protein. An amino acid sequence becomes a fully functional protein by folding into secondary structural elements that then collectively form a specific tertiary structure. For many years the native tertiary fold of a protein was considered the primary determinant of functionality for a specific protein, however, the tertiary structure does not fully represent the mediating factors associated with function. This is due to the fact that proteins do not, in fact, remain in a static tertiary structure once folded [1,2]. At any given time proteins will unfold and refold in the native basin and the population of these unfolded states will correlate with the global stability of the protein [3–5]. Proteins can also have diverse motions within the native basin that are often dictated by energetics and the favorable population of different accessible conformations [6,7]. Rather than remaining static in the native basin, proteins move on various timescales ranging from picosecond-nanosecond conformational fluctuations to millisecond and longer motions which represent local unfolding events [8]. Indeed, the native basin is not accurately represented by a single basin, but instead by an ensemble of basins that together encompass the different states sampled by the protein structure [9]. These ensembles can be characterized by the variation in motional dynamics, or the degree and range of timescales inherent to the particular native fold of a protein.

An entropic change associated with a dynamic transition in the native basin can be an important and fundamental step in functional regulation of a protein system. An example is the regulation of binding interfaces and sequential binding events. Often, a dynamic transition upon ligand binding will allow for longer exposure of binding interface residues or stabilization of dynamic domains, allowing for more

effective docking of co-proteins or receptors [10,11]. In the past the regulation of these types of binding events have been analyzed from more of a structural perspective, however, it is becoming more widely accepted that allosteric regulation through dynamic pathways can dictate protein function independent of any large structural changes [12–15]. Much as certain aspects of protein structures are conserved through evolution to retain functionality, protein dynamic motions are also conserved through evolution to retain functionality [16,17].

Dynamic motions in proteins can exist on a variety of different timescales ranging from motions that represent conformational flexibility (picosecond to nanosecond timescales) to local regions of partial unfolding (milliseconds and up) and many diverse timescales in between (Figure 4-1). The entropic, allosteric, and conformational transitions inherent in processes such as catalysis, ligand binding, and receptor binding are important in transferring information along a protein structure, subsequently aiding in the effective modulation of final activity. While there are many ways to analyze the dynamic motions of proteins, NMR spectroscopy is a particularly useful tool because it allows for site-specific information on dynamics across a large range of timescales, from fast motions on the ps-ns timescale to seconds and longer to probe local unfolding events [18–20]. Understanding the dynamic fluctuations in a protein is often the key to uncovering hidden functional complexity that may not be readily apparent from the static picture of protein structure [18,21]. The ability to access the range of motions that span not just a protein structure, but also the full range of timescales of dynamic motion is optimal for analysis through NMR spectroscopy.

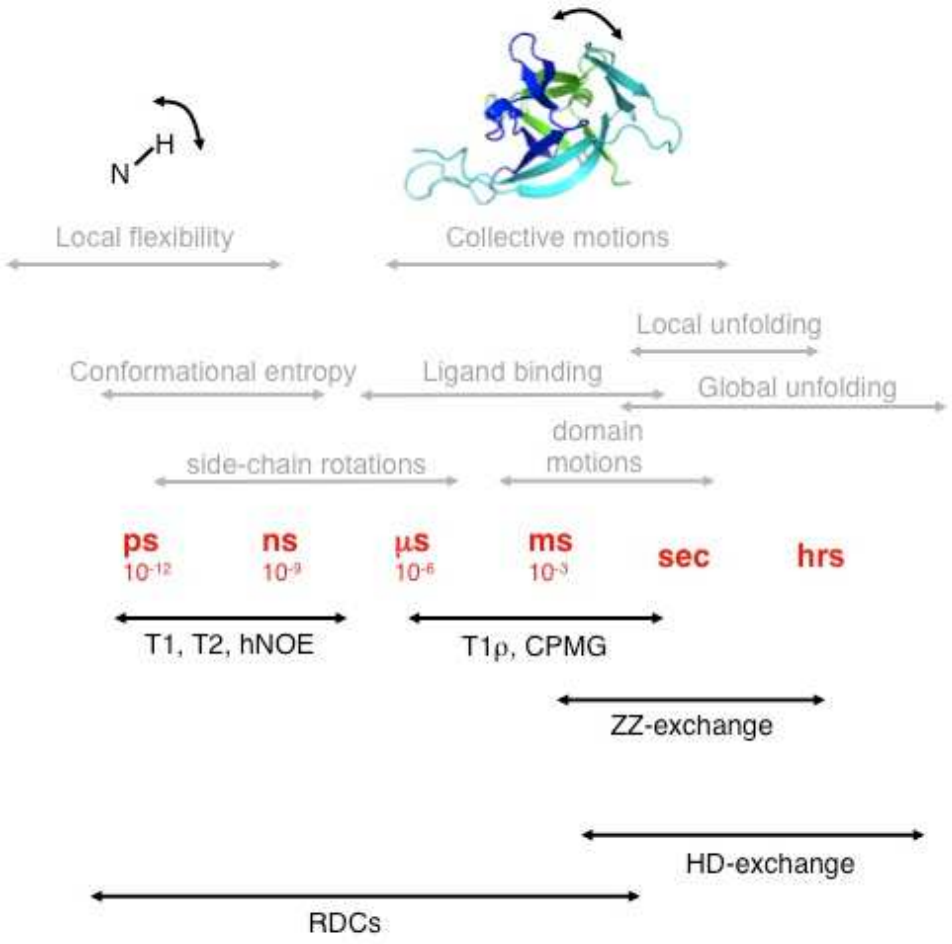

Figure 4-1: Dynamic Timescales for Protein Motions.

Protein motions can exist on a variety of different timescales ranging from picoseconds to seconds. The above figure shows the timescales of protein motions (red), the biophysical and biological phenomena associated with each timescale motion (grey), and the NMR approaches used to assess each particular timescale (black).

4.2 Long timescale motions: Hydrogen/Deuterium Exchange to probe local unfolding events.

Hydrogen/Deuterium Exchange Spectroscopy by NMR allows for the site-specific analysis of long timescale dynamics in a protein structure. This technique utilizes the differential magnetic properties of a hydrogen and deuterium atom. The backbone amide hydrogen within a protein sequence creates an observable signal through its spin ½ state that the NMR is tuned to detect. As Deuterium with a spin 1 state is incorporated into the backbone amide hydrogen position, the signal coherence is lost. Thus, the variation in the observed signal frequency enhancement allows for a direct probe of signal "decay" as deuterium is incorporated into the protein backbone. As a protein fluctuates within the native basin, certain regions may sample locally unfolded states more often than other regions within the protein. This variation in local unfolding will lead to differential incorporation of deuterium into the structure and acts as a probe of the variation of states populated within the native basin.

Exchange of deuterium into a protein backbone can occur through two different mechanisms and distinguishing between the two requires knowledge of the global stability of the protein of interest. An EX1 mechanism of exchange represents exchange to deuterium when the entire protein unfolds to allow exchange. Under an EX1 mechanism, the proton of interest does not exchange unless the protein samples a fully unfolded conformation. This mechanism can be identified because it occurs for residues where the protection factor is higher than the global stability of the protein. This typically results in a probe whose signal persists over the time course of the experiment instead of experiencing instantaneous or a steady decay. An EX2 mechanism of exchange is more common and represents the incorporation of

deuterium due to either direct solvent exposure of particular amide proton or from local unfolding and fluctuations of a specific region within the protein structure. If a protein has residues that exchange via an EX2 mechanism, the protection factor associated with that residue will be lower than the global stability of the protein (Figure 4-2).

The EX2 mechanism of exchange is reliant upon not only the local unfolding events of a protein, but also extended mediating factors [22,23]. The rates of exchange are related to the intrinsic rate of exchange of an amide proton in the disordered peptide, the frequency at which a specific region of the protein samples an open conformation, or the fraction occupancy of the open state, and the pH of the exchange reaction [24,25]. The occupancy of an open state is specific to each regional probe in the protein structure. The balance between the intrinsic exchange rates and fractional occupancy of the open state can be denoted by the equation:

4.1 $K_{ex}^{obs} = k_{ex}^{int}([F_n^{open}]/[F_n^{open}] + [F])$

Which can be simplified to:

4.2 $K_{ex}^{obs} = k_{ex}^{int} k_n^{open}$ when $k_n^{open} << 1$

The protection factor, therefore, corresponds to $-\log k_n^{open}$ which takes into account both the observed and intrinsic exchange rates of a given amino acid within a protein sequence [26]. Protection factors allow for the normalization of rates of exchange, taking into account the primary sequence and its effect on the intrinsic rate of exchange, the pH of the reaction, and the observed rates of exchange. Protection factors, therefore, represent a scalable metric to allow for the direct comparison of

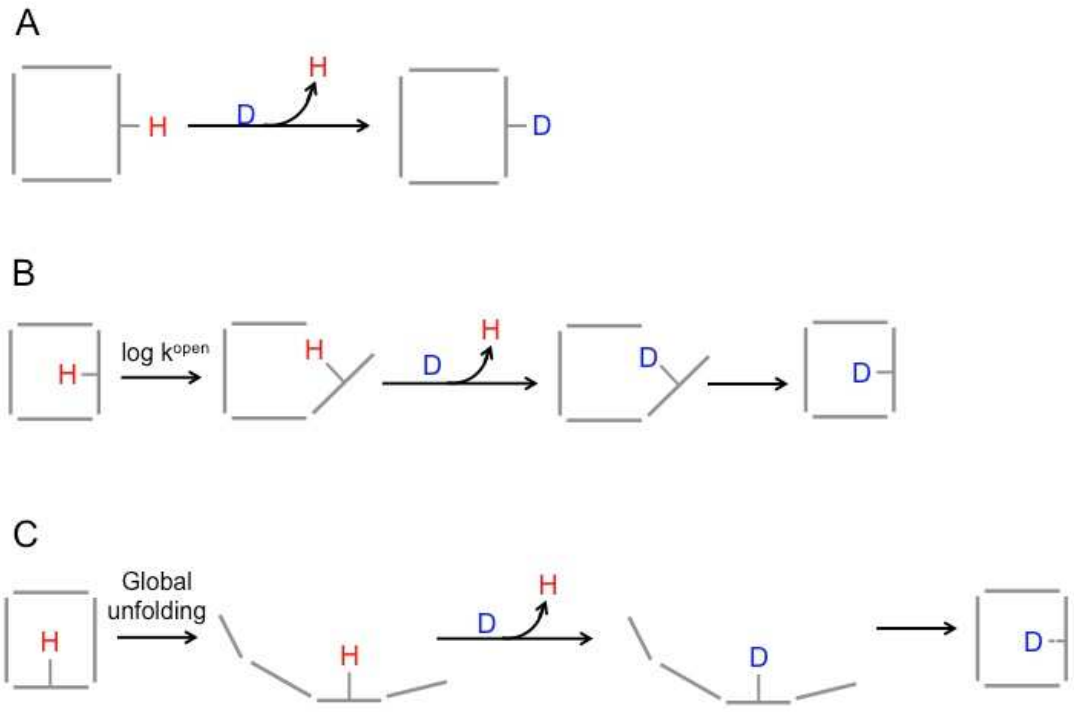

Figure 4-2: Mechanisms of Hydrogen/Deuterium Exchange in Proteins

The exchange of amide protons with Deuterium can proceed through three generalized mechanisms of exchange. (A) Solvent exposed protons will exchange to deuterium quickly and without the need for any perturbations of the protein structure. (B) Local unfolding of protein structure can occur through a rate of log k^{open} = PF and refers to the rate at which the local structure unfolds and allows for exchange of Deuterium. These probes exchange with protection factors lower than the global stability of the protein. (C) Amide protons may also require full unfolding, or global unfolding, of the protein to exchange. This corresponds to residues that exchange to deuterium with protection factors higher than the global stability of the protein.

exchange parameters between two different proteins analyzed in two different reaction conditions.

4.3 Assessing Fast Timescale Dynamics: The R1, R2, and {H^1}-^{15}N NOE suite of experiments

Fast timescale dynamics refer to dynamics within the ps-ns timescale and represent motions that are fairly ubiquitous across the proteome. Although these fast timescale dynamics are commonly present in the final fold of a protein, the degree to which they exist and how they localize across the protein structure can provide valuable information about the functionality of the protein. These motions can be probed through the application of a set of three ^{15}N backbone relaxation experiments: R1, R2, and {H^1}-^{15}N NOE.

The T1 and T2 experiments both involve perturbing the nuclear spins of the samples from equilibrium and then monitoring their return back to the ground state. This return to equilibrium is then plotted as a function of time to obtain the rate at which each nulcei regains equilibrium. For the T1 and T2 experiments, the subsequent extracted rates are represented as R1 and R2, which represent the inverse of the relaxation rates, respectively. The rate of repopulating equilibrium is highly dependent on the local magnetic fields surrounding the nuclei, the external magnetic field (B_o), and the degree of motions that the nuclei samples. Fast time scale motions and degrees of conformational flexibility within the structure of the protein will have a strong influence on the rates of relaxation, and thus the T1 and T2 experiments allow for quantification of dynamics for each specific amide bond within the protein backbone.

Specifically, the T1 experiment probes the spin-lattice relaxation rate, or the rate at which energy is transferred from the spin on the amide proton to the lattice of

neighboring molecules. The signal decay plotted as a function of time gives an observable rate at which the spin population is restored back to the Boltzmann distribution of the spin states at equilibrium. T1 relaxation occurs along the Z vector plane (which is oriented in the same vector plane as the external magnetic field B_o) and represents bulk magnetization return along the static magnetic field.

The T2 experiment probes the transverse relaxation, or the precession of the magnetic vector along the x and y plane. The loss of coherence is faster than for the T2 experiment and is also more sensitive to molecular motions. The sensitivity to molecular motions has to do with the fact that T2 relaxation is highly responsive to local magnetic fields and variations in the chemical environment, and thus will be highly influenced by the motional fluctuation that the vector experiences as the transverse relaxation proceeds. Collectively, the R1 and R2 relaxation rates represent perturbation of the spin vector for each amide bond along all three magnetic coordinates, x, y, and z and allow for analysis of how the perturbation is effected by local motions on fast timescales.

Heteronuclear NOE experiments utilize the Nuclear Overhauser Effect to monitor spin transfer from one spin population to another via cross-relaxation. This spin flip transfer process is sensitive to both the dynamic motions of the protein as well as the molecular tumbling of the protein, and therefore encompass both the dynamics and structure of the protein being studied. The suite of T1, T2, and heteronuclear NOE experiments allow for the identification of the effects of dipolar coupling as well as monitoring how perturbed spin states relax back to equilibrium, all of which will be influenced by the relative dynamics and order within a protein structure. In this way, the parameters for conformational entropy associated with fast timescale motions can

be extracted from incorporating the three experimental data sets into the model-free order parameter calculations developed by Lipari and Szabo.

4.4 Order parameters: The Lipari-Szabo Formalism for Extracting Comparable Values of Conformational Entropy

The Lipari-Szabo formalism allows for the input of NMR relaxation data to quantify the average position of an amide bond vector in space. This formalism allows for the assessment of conformational entropy and seeks to neutralize the specific effects of both magnetic field strengths of data collection. The model-free calculations allow extraction of parameters for the amide bonds that takes into account the effect of the specific rotational correlation of a protein in space based on its size as well as the dynamic input from the relaxation experiments. Because order parameter assessments help to scale the effects of magnetic field contributions and specific factors for the unique protein data set, it represents a comparable metric for different proteins independent of variable data collection and the incorporated effect of the rotational motion of the protein in solution. The general equation for extracting the order parameter (S^2) is:

4.3 $J(\omega) = 2/5 ((S^2_{\tau c} / 1+(\tau_c)^2) + ((1-S^2)_\tau / 1+(\tau\omega)^2)$

Where $J(\omega)$ represents the spectral density function of the system, S^2 represents the order parameters, and τ_c represents the rotational correlation function of the given protein [27,28].

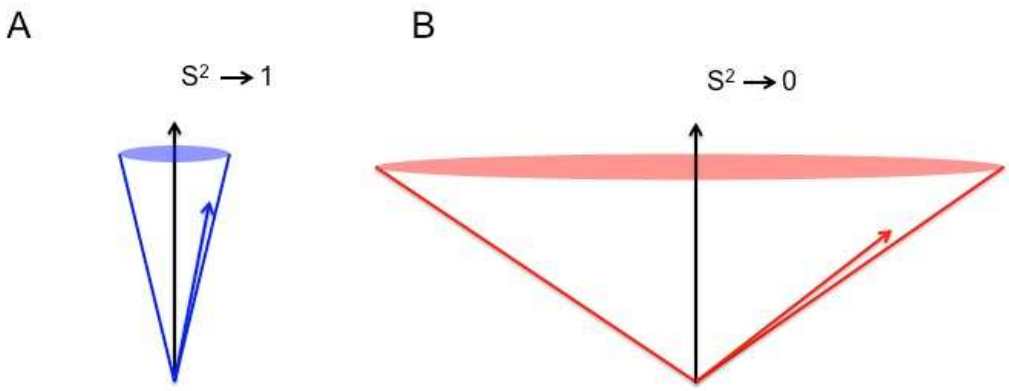

Figure 4-3: Order Parameters (S^2) are representative of dynamics as a probability of vector orientation.

Order parameter of S^2 derived from the model free analysis by Lipari-Szabo is on a scale from 0 to 1 where the value represents the probability of finding a magnetic vector in the same starting position at any given time. (A) An S^2 value of 1 represents a 100% probability of finding a vector in the same position at any given time and thus represents the most rigid an amide bond can be because it is unlikely to travel far in space. (B) An S^2 value of 0, conversely, represents a 0% probability of finding a magnetic vector in the same position at any given time and is thus representative of a highly dynamic amide bond because of the unlikeliness of finding the vector in the same position at any given time.

The order parameter of S^2 is on a scale from 0 to 1 wherein the value associated represents the probability of finding a magnetic vector in the same starting position. An S^2 value of 1 represents a 100% probability of finding a vector in the same position at any given time and thus represents the most rigid an amide bond can be because it is unlikely to travel far in space. An S^2 value of 0, conversely, represents a 0% probability of finding a magnetic vector in the same position at any given time and is thus representative of a highly dynamic amide bond because of the unlikeliness of finding the vector in the same position at any given time. Residues will fall within this range of order parameters and the degree to which their probability fluctuates reflects the relative order of the amide bond for a particular residue. Typically, residues with an S^2 value above 0.8 are considered rigid whereas residues with an S^2 value below about 0.6 and 0.5 are considered flexible and have a significant amount of conformational entropy.

4.5 References

[1] Tokuriki N, Tawfik DS. Protein dynamism and evolvability. Science 2009;324:203–7.

[2] Smock RG, Gierasch LM. Sending signals dynamically. Science 2009;324:198–203.

[3] Pace CN, Treviño S, Prabhakaran E, Scholtz JM. Protein structure, stability and solubility in water and other solvents. Philos Trans R Soc Lond B Biol Sci 2004;359:1225–34; discussion 1234–5.

[4] Daggett V. Molecular dynamics simulations of the protein unfolding/folding reaction. Acc Chem Res 2002;35:422–9.

[5] Matouschek A. Protein unfolding - An important process in vivo? Curr Opin Struct Biol 2003;13:98–109.

[6] Robertson AD, Murphy KP. Protein Structure and the Energetics of Protein Stability. Chem Rev 1997;97:1251–68.

[7] Naganathan AN, Doshi U, Fung A, Sadqi M, Muñoz V. Dynamics, energetics, and structure in protein folding. Biochemistry 2006;45:8466–75.

[8] Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature 2007;450:964–72.

[9] Shortle D, Simons KT, Baker D. Clustering of low-energy conformations near the native structures of small proteins. Proc Natl Acad Sci U S A 1998;95:11158–62.

[10] Tzeng S-R, Kalodimos CG. Protein activity regulation by conformational entropy. Nature 2012;488:236–40.

[11] Tsai CJ, del Sol A, Nussinov R. Allostery: Absence of a Change in Shape Does Not Imply that Allostery Is Not at Play. J Mol Biol 2008;378:1–11.

[12] Tzeng S-R, Kalodimos CG. Allosteric inhibition through suppression of transient conformational states. Nat Chem Biol 2013;9:462–5.

[13] Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL. Hidden dynamic allostery in a PDZ domain. Proc Natl Acad Sci U S A 2009;106:18249–54.

[14] Popovych N, Sun S, Ebright RH, Kalodimos CG. Dynamically driven protein allostery. Nat Struct Mol Biol 2006;13:831–8.

[15] Kern D, Zuiderweg ERP. The role of dynamics in allosteric regulation. Curr Opin Struct Biol 2003;13:748–57.

[16] Marianayagam NJ, Jackson SE. Native-state dynamics of the ubiquitin family: implications for function and evolution. J R Soc Interface 2005;2:47–54.

[17] Frauenfelder H, Chen G, Berendzen J, Fenimore PW, Jansson H, McMahon BH, Stroe IR, Swenson J, Young RD. A unified model of protein dynamics. Proc Natl Acad Sci U S A 2009;106:5129–34.

[18] Baldwin AJ, Kay LE. NMR spectroscopy brings invisible protein states into focus. Nat Chem Biol 2009;5:808–14.

[19] Mittermaier AK, Kay LE. Observing biological dynamics at atomic resolution using NMR. Trends Biochem Sci 2009;34:601–11.

[20] Wand AJ. Dynamic activation of protein function: a view emerging from NMR spectroscopy. Nat Struct Biol 2001;8:926–31.

[21] Kempf JG, Loria JP. Protein Dynamics from Solution NMR. Cell Biochem Biophys 2002;37:187–212.

[22] Bai Y, Milne JS, Mayne L, Englander SW. Protein stability parameters measured by hydrogen exchange. Proteins 1994;20:4–14.

[23] Bai YW, Milne JS, Mayne L, Englander SW. Primary Structure Effects on Peptide Group Hydrogen-Exchange. Proteins-Structure Funct Genet 1993;17:75–86.

[24] Best RB, Vendruscolo M. Structural interpretation of hydrogen exchange protection factors in proteins: characterization of the native state fluctuations of CI2. Structure 2006;14:97–106.

[25] Maity H, Lim WK, Rumbley JN, Englander SW. Protein hydrogen exchange mechanism: local fluctuations. Protein Sci 2003;12:153–60.

[26] Haglund E, Lindberg MO, Oliveberg M. Changes of protein folding pathways by circular permutation: Overlapping nuclei promote global cooperativity. J Biol Chem 2008;283:27904–15.

[27] Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. J Am Chem Soc 1982;104:4559–70.

[28] Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. J Am Chem Soc 1982;104:4559–70.

Chapter 5 : Heterogeneous Dynamics and Conformational Frustration in the Binding Regions of Interleukin-33.

5.1 Abstract

Interleukin-33 (IL-33) is currently the focus of numerous investigations into targeting chronic inflammatory disorders. As a mediator of inflammation, IL-33 plays a prevalent role in disorders such as asthma, rheumatoid arthritis, and breast cancer. IL-33 targets and binds the extracellular receptor ST2 which then forms a hetero-trimeric signaling complex by engaging a third cellular receptor. This hetero-trimeric signaling complex allows for the activation a variety of downstream effectors responsible for modulating inflammation. The ST2 receptor represents a highly dynamic receptor, with three Ig domains connected by highly flexible linkers and upon binding, IL-33 must span all three of the flexible Ig domains to bind successfully. In this work we show that the regions of IL-33 responsible for engaging ST2 are highly dynamic on a range of timescales and that these dynamic regions in IL-33 are conformationally frustrated. The regions of IL-33 that do not engage the ST2 receptor are much less dynamic and create a rigid, stable core to anchor the functional dynamic regions of the protein. The strong correlation between residual conformational frustration and heterogeneous dynamics on multiple timescales within the functional region of IL-33 suggests that both the dynamics and frustration may be a conserved feature important in modulating the functionality of the protein. Specifically, the highly dynamic and frustrated character of the receptor binding regions in IL-33 may allow for effective engagement of the highly dynamic and mobile binding interface along the three Ig domains of ST2.

5.2 Introduction

The action of small signaling cytokines such as Interleukin-33 (IL-33) and its role in regulating the immune response is critical in maintaining inflammatory homeostasis [1]. Inflammation is a finely tuned process responsible for establishing resistance to infection and for facilitating the healing of damaged tissues [1,2]. Aberrant inflammation, however, can lead to a variety of damaging pathologies and can potentially cause chronic disorders [3,4]. One of the difficulties in treating diseases typified by instances (both sustained and systemic) of aberrant inflammation is treating these disorders directly and effectively without causing detrimental side effects [5,6]. It is not feasible to completely eliminate inflammation because doing so invariably leads to the development of cancers and dangerous infections [7–9]. The effective treatment of inflammatory disorders requires a more detailed understanding of the mechanistic details that govern the inflammatory response. In such an approach, specific regions of interest on IL-33 can be identified for the development of more tailored treatments. The specific action of IL-33 is influenced by the biophysical properties of the protein and are of particular interest due to IL-33's emerging role in disorders ranging from Rheumatoid Arthritis to Asthma [10,11].

IL-33 folds into a conservedβ-Trefoil fold that is shared among other members of the interleukin-1 family such as IL-1β, IL-1Ra and IL-36 [12]. This fold is characterized by a pseudo-symmetrical structure composed of three units of four β-sheets that fold to form a 12 stranded "β-barrel' like structure. Each of these three units of four β-strands are designated as Trefoil 1, 2, and 3, respectively, in order from the N-terminus to the C-terminus (Figure 5-1, A and B). As a member of the Interleukin-1 family of cytokines, Interluekin-33 (IL-33) not only shares a homologous

β-trefoil fold, but also acts as a small signaling cytokine in a manner similar to other members of the family [13,14]. IL-33 targets and binds its conjugate extracellular receptor, ST2 allowing the complex to engage a third extracellular receptor, Interleukin 1 receptor accessory protein [15]. This active hetero-trimeric signaling complex activates a variety of downstream effectors responsible for modulating inflammation in different tissue types, including mucosal and epithelial tissues [6,16,17].

The Interleukin-33 binding interface, responsible for engaging the Ig domains of ST2, constitutes the entirety of the first 5 β-strands in IL-33 (Trefoil 1) consisting of residues 1 to 70 and the final terminal end of the protein, composed of β-strands 11 and 12, residues 140 to 150 [15,18]. This extensive interface on IL-33 is able to engage the three separate Ig domains of ST2 simultaneously (Figure 5-1, B and C). Of the Ig domains composing ST2, Ig2 and Ig3 are connected by a flexible linker region, allowing for a large range of dynamics motions within the receptor [15,19]. The influence of dynamics and entropy within the fold of IL-33 on the effective, simultaneous engagement of the three Ig domains of ST2 has not been investigated. In this work we show that the regions of IL-33 responsible for engaging the Ig domains of ST2 are dynamic on a range of timescales and that these dynamic regions correlate with a high degree of conformational frustration. The region of IL-33 that does not engage the ST2 receptor is much less dynamic, creating a rigid and stable core to anchor the dynamic, functional regions of the protein. Our previous work showed that folding the functional domain (Trefoil 1) was complicated due to geometric frustration along the free energy profile. Due to the geometric frustration in the functional region during folding, we predicted subsequent malleability and greater dynamics within this

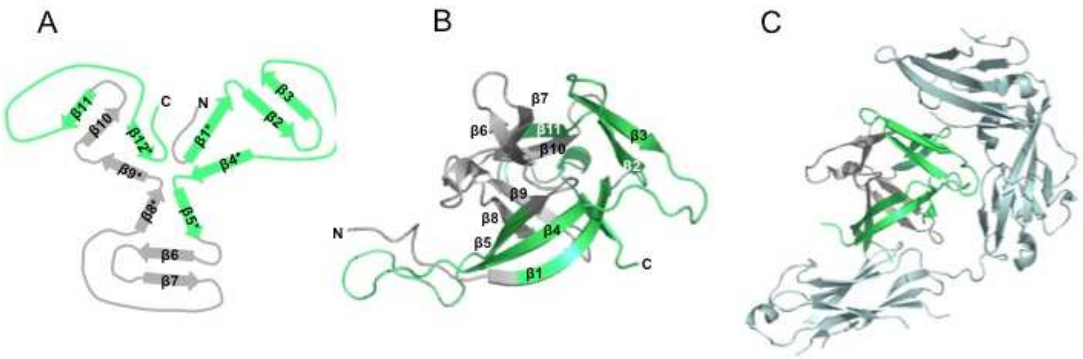

Figure 5-1: Representation of the Interleukin-33 β-trefoil fold and ST2 binding interface.

(A) A two-dimensional splay diagram representing IL-33 from a top view of the protein. Asterisks indicate strands composing the "barrel" portion of the protein while the strands of the β-hairpin cap are on the periphery. Trefoil 1 is composed of β-strands 1-4, Trefoil 2 of strands 5-8, and Trefoil 3 of strands 9-12. The regions that constitute the binding interface with ST2 are colored green while the non-interacting regions are colored grey. (B) The 20 overlaid NMR structures of IL-33 (PDB code 2KLL) are displayed with the ST2 binding interface colored green as in A. (C) The crystal structure of the complex between IL-33 and ST2 (PDB code 4KC3) is displayed with the strands of IL-33 responsible for interacting with ST2 colored green

region of IL-33. Conformational frustration and heterogeneous dynamics in the native basin along Trefoil 1 is likely a conserved feature in the final fold of IL-33 that aids in the engagement of the flexible Ig domains in ST2. Further, the differential dynamics of the functional region of the protein, particularly Trefoil 1, may allow for increased responsiveness to receptor binding without significantly perturbing the rest of the stable protein scaffold.

5.3 Materials and Methods

5.3.1 Protein Purification and Sample Preparation.

The uniformly labeled ^{15}N extracellular signaling domain of IL-33, residues 110-270, was expressed in a pET-24vector and grown in ^{15}N enriched minimal media. Initial purification was achieved using osmotic shock cell lysis. Final purity was attained through anion exchange FPLC using a Sodium Chloride gradient. All NMR samples were run in the same experimental buffer of 20 mM Sodium Phosphate at pH 6.50, 100 mM Sodium Chloride, and 10% Deuterium Oxide.

5.3.2 NMR Experiments and Data Analysis.

The hydrogen/deuterium exchange time course was collected on a Bruker Avance III 600.112 MHz spectrometer equipped with a CryoProbe. All spectral acquisition was run at 25°C. Final NMR samples of IL-33 were buffered in 20 mM sodium phosphate, 100 mM sodium chloride, and 10 % deuterium at a final pH of 6.50. The sample was split into two aliquots of equal volume and with equal concentrations of 400 μM IL-33. Acquisition parameters were optimized using the first aliquot containing 10 % deuterium. The second aliquot was lyophilized overnight until dry

followed by the addition of 100 % deuterium before collecting time course data on the spectrometer. The dead time before the first spectral acquisition was 130 seconds. Spectra were each collected with 2048 x 124 complex points. The SO-FAST HMQC protocol was implemented for collection of data and optimized to collect HMQC spectra every 59 seconds. The time course was run for a total of 72,000 seconds (20 hours). The signal intensity of backbone ^{1}H-^{15}N cross peak heights were monitored as a function of time.

HMQC spectra were analyzed by fitting the ^{1}H-^{15}N cross peak intensity heights for each residue in IL-33 as a single exponential function according to the equation $I(t) = I_o \exp^{(-tk)} + I_{ss}$ where I_o represents the signal intensity, k represents the rate of deuterium incorporation, *t* represents the time of the exchange reaction, and I_{ss} represents the signal intensity when it has reached steady state, or equilibrium (Figure 5-2). Spectra were analyzed with nmrPipe and Sparky software. The signal intensity decays from each spectrum were fit with a single exponential decay to extract an observed rate of deuterium incorporation using Kaleidagraph software. Rates of deuterium incorporation were used to calculate both intrinsic exchange rates and protection factors using software at http://hx2.med.upenn.edu [20,21]

R1 and R2 experiments were recorded on a Varian 500.18 MHz spectrometer and a Varian 801.45 MHz spectrometer equipped with a CryoProbe. All experiments were run at 25°C. Values of the longitudinal relaxation rate (R1) were collected for the backbone amide resonances at delay times of t= 100, 300, 700, 1000, 1500, 1700, 2100, 2300, 3000, 3800, and 4600 ms. Values of the transverse relaxation rate (T2) were collected at t= 10, 30, 50, 70, 110, 130, 150, 170, 190, 210, and 230 ms. All relaxation delay times were chosen based on evaluation of the 1D signal decays. All

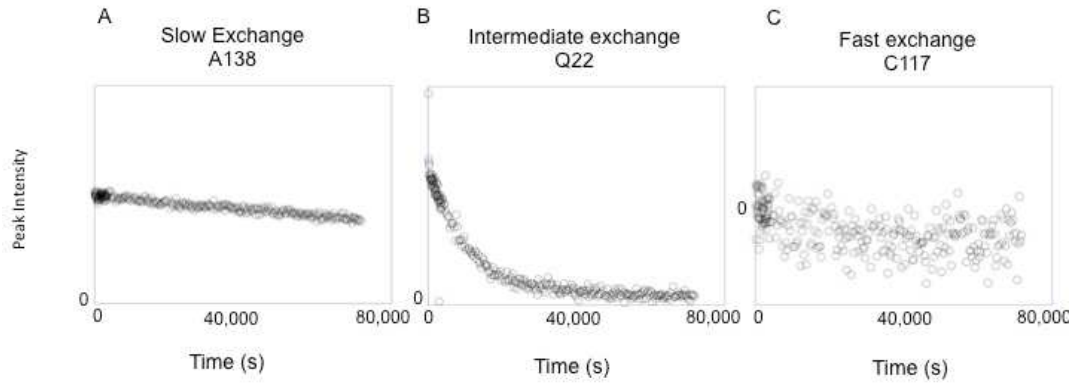

Figure 5-2: Representative curves of signal decay in the Hydrogen /Deuterium exchange time course.

Three representative time courses are shown for the decay of signal upon Deuterium incorporation. (A) Shows the decay of signal intensity for the residue 138. The decay is monitored as a function of signal intensity versus time in seconds. Residue 138 is representative of slow exchange due to the persistence of the signal intensity over time. (B) Shows the decay of signal intensity for residue 22 as in A. Residue 22 is representative of intermediate exchange due to achieving exchange equilibrium after a decay time of intermediate time. (C) Shows the decay of signal intensity for residue 117 as in A. Residue 117 is representative of fast exchange due to the total loss of signal intensity within the dead time of the

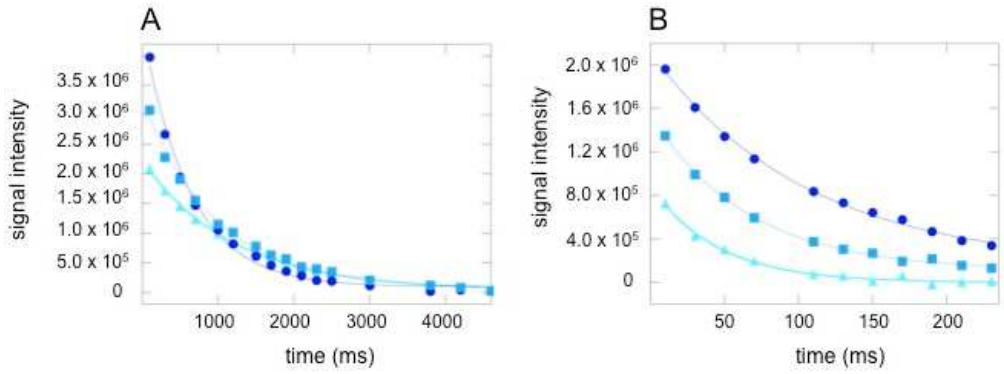

Figure 5-3: Representative Decay curves for T1 and T2 exchange rates.

(A) Representative plots of the fits of T1 decay rates is shown where the y-axis represents the signal intensity and the x-axis represents the delay times for each data point collection. The three curves are three representative residues within IL-33. The circles (dark blue) represent the decay of Glycine 66, the squares (blue) represent the decay of Valine 37, and the triangles (cyan) represent the decay of Histidine 111. The associated curves represent the single exponential fit of each decay curve for the individual residues. (B) A representation of the T2 decay curves represented as in A.

spectra were collected with 2048 x 124 complex points and 16 steady state scans. Spectra were processed using the nmrPipe and Sparky software suites. Data for each residue was acquired through a single exponential fit of the decay curves of each relaxation delay time point for the ^{1}H-^{15}N cross peak intensity described by $I(t) = I_o \exp^{(-tR)} + I_{ss}$ where I_o represents the signal intensity, R represents the rate of signal decay, *t* represents the variable relaxation delay for each peak, and I_{ss} represents the signal intensity when it reaches a steady state for the reaction progress (Figure 5-3). Errors for the decay fits were calculated through a repeated fit with gaussian randomization of the decay curve for each residue.

{H^1}-^{15}N NOE experiments were recorded on a Varian VS 500.18 MHz spectrometer and a Varian VS 801.45 MHz spectrometer equipped with a CryoProbe. All experiments were run at 25°C. Both a saturated and unsaturated spectra were recorded with a relaxation delay of 5 s. Each NOE experiment was run in duplicate to evaluate reproducibility of signal intensities. Experiments were run with 2048 x 124 complex points and 64 steady state scans for optimal signal-to-noise ratios. Spectra were processed using the nmrPipe and Sparky software suites. NOE values were calculated by taking the peak intensity ratio between the saturated and unsaturated cross peak intensities $=(I_{saturated}/I_{unsatruated})$. Errors are representative of the standard deviation between two different acquisitions of the {H^1}-^{15}N NOE data.

Model free analysis and reduced spectral density mapping were calculated with the Lipari-Szabo formalism [22,23] using the eMF software [24] and inputting the R1, R2 and {H^1}-^{15}N NOE values and errors from the 500 and 800 MHz acquisition runs. An axially symmetric diffusion tensor with a τ_c of 10.7 was calculated using the NMR structure for IL-33 (PDB code 2KLL) with selected rigid residues with order parameter values above 0.80 corresponding to the most ordered regions of the protein,

composed of β-strands in the middle of the protein structure. The best model for each residue was determined by a Bayesian criterion after excluding models with $S^2 > 1$ or $\tau_c < 0$ [25–27]. Reduced spectral densities were determined assuming that $J(\omega)$ is proportional to $1/\omega^2$ [28]

Frustration data was analyzed using the Frustratometer webserver using the PDB file for Interleukin-33 (PDB code 2KLL) [29]. The associated algorithm uses the funneled energy landscape theory of folding to quantify the energy that spans the protein structure. It then evaluates each native pair interaction with respect to alternate contacts and alternate compact structures that the protein may sample. Energetically favorable contacts are denoted as minimally frustrated and less favorable contacts are denoted as frustrated.

5.4 Results

5.4.1 The topology of Interleukin-33 and its functional binding interface.

Interleukin-33 shares a homologous fold with the Interleukin-1 family of proteins [12]. This fold is called the β-trefoil fold and is characterized by a pseudo-symmetrical arrangement consisting of three groups of four β-strands. The three groups of four β-strands arrange into a "β-barrel" like fold [30–32]. The trefoils are designated as Trefoils 1, 2, and 3 and span from the N-terminus to the C-terminus, respectively. Trefoil 1 is composed of strands 1-4, Trefoil 2 of strands 5-8, and Trefoil 3 of strands 9-12. The primary region of IL-33 responsible for engaging the Ig domains of the ST2 receptor consists of the entire span of IL-33 along Trefoil 1 and the C-terminal end of the protein along β-strands 11 and 12 [13,15]. Interestingly, the

binding interface is diffuse and composes all of Trefoil 1 whose contacts contribute in an equal, additive fashion to the binding affinity to ST2 (Figure 5-1).

5.4.2 Hydrogen/Deuterium Exchange shows evidence for regions of differential dynamics.

Hydrogen/Deuterium exchange of IL-33 shows that overall the protein is fast to exchange and that there are few regions of protection in the structure. Protection factors were calculated using the Englander lab software described in materials and methods. 44% of the residues in the structure exchange within the dead time of the exchange time course and are denoted as having a log k^{obs} of -1 and typically correspond to solvent exposed amide protons (Table 1). Exchange on intermediate timescales typically corresponds to local unfolding events in the protein sequence, and are classified as having an EX2 mechanism of exchange [20,33]. The rates of exchange are related to both the intrinsic rate of exchange of an amide proton in the disordered peptide as well as the frequency at which the protein samples a local, open conformation and the subsequent fraction occupancy of the open state [34,35]. This can be denoted by the equation:

$$k_{ex}^{obs} = k_{ex}^{int}([F_n^{open}]/[F_n^{open}] + [F])$$

Which can be simplified to:

$$k_{ex}^{obs} = k_{ex}^{int} k_n^{open} \text{ when } k_n^{open} << 1$$

The protection factor therefore corresponds to $-\log k_n^{open}$ which takes into account both the observed and intrinsic exchange rates of exchange of a given amino acid within a protein sequence [36]. Only 12 residues have protection factors above 6, indicating that they are slow to exchange to deuterium and that their signal

intensities persist over the exchange time course. These residues include V37, M73, V74, L76, N112, V118, I130, G131, L137, A138, I140, C149. Often residues where the proton signal persists over the exchange time course and have high protection factors are indicative of exchange due to global unfolding of the protein rather than local conformational fluctuations or local unfolding events. This type of exchange mechanism is categorized as an EX1 exchange mechanism. To verify the exchange as EX1, the protection factors were compared to the stability of the unfolding reaction, or the ΔG of unfolding. The ΔG of unfolding (or the global unfolding stability) of IL-33 is 7.07 kcal, indicating that residues exchanging due to global unfolding have protection factors that correspond to rates slower than the overall rate of global unfolding. Therefore, residues that require global unfolding of the protein, and thus correspond to an EX1 mechanism, are M73, V118, I130, L137, A138, and C149, all of which are located in Trefoils 2 and 3, respectively (Figures 5-4 and 5-5).

5.4.3 Backbone Dynamics of IL-33 in the fast time regime.

R1 and R2 relaxation rates and $\{H^1\}$-^{15}N NOEs were measured at 801.45 MHz as shown in Figure 5-6 C. There is evidence for heterogeneous dynamics within IL-33. In particular, the $\{H^1\}$-^{15}N NOE values vary widely across the sequence of IL-33. As shown in Figure 5-6, the regions with $\{H^1\}$-^{15}N NOE values above 0.8 are found primarily in β-strands 1 and 2 (residues 14-28), β-strand 3 (residues 40-45), and β-strands 6-11 (residues 80-140), indicating regions with order and rigidity. These strands compose all of Trefoils 2 and 3 and part of Trefoil 1, respectively. Low $\{H^1\}$-^{15}N NOE values below 0.4 are found in regions spanning a large portion of the structure and include the N-terminal residues, the loop between β-strands 2 and 3 (residues 25-

Table 5-1: Hydrogen/Deuterium Exchange Parameters

Amino Acid	$logK_{ex}^{obs}$	K_{ex}^{int}, s^{-1}	PF	Amino Acid	$logK_{ex}^{obs}$	K_{ex}^{int}, s^{-1}	PF	Amino Acid	$logK_{ex}^{obs}$	K_{ex}^{int}, s^{-1}	PF	Amino Acid	$logK_{ex}^{obs}$	K_{ex}^{int}, s^{-1}	PF
S1	-----	-----	-----	E44	-1.00	0.47	2.45	N87	-1.22	2.66	3.42	I130	-6.00	0.19	7.05
S2	-----	-----	-----	K45	-1.00	0.59	2.55	N88	-1.00	5.52	3.52	G131	-4.40	0.94	6.15
I3	-1.00	0.32	2.28	K46	-1.00	1.03	2.79	K89	-1.00	1.63	2.99	V132	-1.00	0.25	2.18
T4	-1.00	0.44	2.42	D47	-1.02	1.47	2.97	E90	-1.00	0.92	2.74	K133	-1.61	0.57	3.14
G5	-1.61	2.54	3.79	K48	-1.00	0.54	2.51	H91	-1.00	3.25	3.29	D134	-1.00	1.45	2.94
I6	-1.89	0.24	3.04	V49	-2.47	0.23	3.60	S92	-1.00	11.27	3.83	N135	-1.00	1.83	3.04
S7	-1.00	1.18	2.85	L50	-2.77	0.17	3.77	V93	-1.00	0.34	2.31	H136	-1.36	9.08	4.09
P8	-----	-----	-----	L51	-----	-----	-1.07	E94	-2.61	0.50	4.08	L137	-7.10	1.26	8.97
I9	-2.31	0.09	3.05	S52	-2.32	1.23	4.19	L95	-1.00	0.17	2.01	A138	-5.92	0.44	7.42
T10	-1.00	0.44	2.42	Y53	-2.33	0.92	4.07	H96	-1.57	2.68	3.78	L139	-2.43	0.23	3.57
E11	-1.00	1.10	2.82	Y54	-2.58	0.02	2.67	K97	-1.59	4.34	4.00	I140	-6.00	12.20	6.77
Y12	-2.31	0.35	3.63	E55	-1.00	1.52	2.96	C98	-2.33	4.73	4.78	K141	-1.80	0.49	3.25
L13	-1.00	0.26	2.19	S56	-1.00	1.96	3.07	E99	-1.00	2.47	3.17	V142	-1.00	0.24	2.13
A14	-2.53	0.55	4.05	Q57	-1.00	N.D.	N.D.	K100	-2.56	0.59	4.11	D143	-1.61	0.81	3.30
S15	-1.00	2.02	3.08	H58	-1.00	N.D.	N.D.	P101	-----	-----	-----	S144	-1.00	1.39	2.92
L16	-1.00	0.46	2.44	P59	-----	-----	-----	L102	-1.00	0.13	1.90	S145	-1.00	3.99	3.38
S17	-2.41	1.26	4.29	S60	-----	-----	-----	P103	-----	-----	-----	E146	-1.00	0.68	2.92
T18	-2.46	1.47	4.41	N61	-1.00	N.D.	N.D.	D104	-1.00	0.63	2.48	N147	-1.57	2.00	3.65
Y19	-1.25	0.73	2.89	E62	-1.00	1.45	2.94	Q105	-1.00	0.68	2.61	L148	-1.00	0.48	2.46
N20	-1.00	2.96	3.25	S63	-1.00	1.52	2.96	A106	-3.60	1.38	5.51	C149	-7.22	2.22	9.34
D21	-1.19	2.31	3.33	G64	-1.68	3.17	3.96	F107	-2.55	0.50	4.03	T150	-2.46	2.63	4.66
Q22	-2.18	0.68	3.79	D65	-1.00	1.63	2.99	F108	-2.45	0.59	3.99	E151	-1.00	1.10	2.82
S23	-1.00	3.18	3.28	G66	-1.00	1.10	2.82	V109	-1.00	0.20	2.07	N152	-1.00	2.00	3.08
I24	-1.86	0.32	3.14	V67	-1.00	0.25	2.18	L110	-----	-----	-----	I153	-1.31	0.34	2.62
T25	-1.81	0.44	3.23	D68	-1.00	0.80	2.68	H111	-1.62	2.68	3.83	L154	-2.92	0.13	3.83
F26	-1.32	0.79	3.00	G69	-1.63	1.08	3.45	N112	-3.88	14.70	6.82	F155	-3.57	0.31	4.83
A27	-1.00	1.00	2.78	K70	-1.00	1.15	2.84	M113	-1.00	1.75	3.02	K156	-3.41	0.91	5.14
L28	-1.33	0.23	2.46	M71	-1.00	1.10	2.82	H114	-1.00	5.65	3.53	L157	-1.00	0.30	2.26
E29	-1.00	0.43	2.41	L72	-1.00	0.29	2.25	S115	-----	-----	-----	S158	-1.00	1.23	2.87
D30	-1.00	0.84	2.70	M73	-6.22	0.09	7.72	N116	-1.76	5.30	4.27	E159	-1.62	1.37	3.54
E31	-1.23	0.47	2.68	V74	-5.25	0.22	6.37	C117	-1.00	7.45	3.65	T160	-1.00	0.01	0.72
S32	-1.75	1.51	3.70	T75	-3.44	0.53	4.95	V118	-6.39	0.60	7.95				
Y33	-1.00	0.92	2.74	L76	-2.23	0.36	6.37	S119	-1.34	1.47	3.28				
E34	-1.00	0.78	2.67	S77	-3.15	1.26	4.95	F120	-1.00	1.00	2.78				
I35	-1.00	0.12	1.86	P78	-----	-----	3.56	E121	-2.46	0.79	4.13				
Y36	-1.57	0.27	2.78	T79	-1.96	0.43	5.03	C122	-3.49	2.71	5.70				
V37	-5.32	0.19	6.38	K80	-1.00	1.24	-----	K123	-2.06	2.81	4.29				
E38	-1.05	0.50	2.52	D81	-1.00	1.45	3.37	T124	-----	-----	-----				
D39	-2.17	0.83	3.87	F82	-2.45	0.35	2.87	D125	-----	-----	-----				
L40	-1.00	0.16	1.97	W83	-2.90	0.39	2.94	P126	-----	-----	-----				
K41	-1.00	0.48	2.46	L84	-3.05	0.18	3.77	G127	-1.00	0.92	2.74				
K42	-1.00	1.03	2.79	H85	-2.84	2.59	4.26	V128	-3.04	0.25	4.22				
D43	-1.00	1.45	2.94	A86	-2.24	4.76	4.08	F129	-1.64	0.36	2.98				

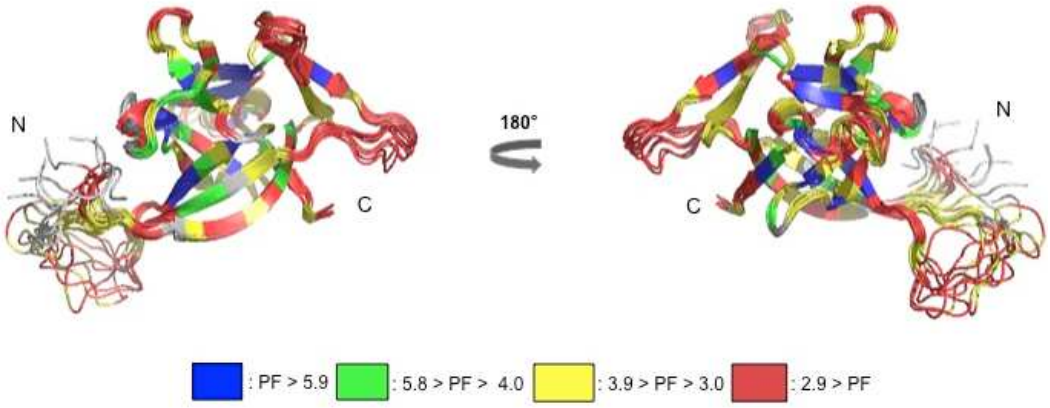

Figure 5-4: Map of the protection factors on the structure of IL-33.

The 20 overlaid NMR structures of IL-33 (PDB code 2KLL) are displayed with the associated protection factors mapped onto the structure. Protection factors above 5.9 are displayed in blue, protection factors between 5.8 and 4 are displayed in green, protection factors between 3.9 and 3.0 are displayed in yellow, and protection factors below 2.9 are displayed in red. Blue represents the most protection, green and yellow represent intermediate protection, and red represents low protection. Overall, the region spanning β-strands 1 and 5 show the lowest overall span of protection in the entire structure. Probes that could not be assigned rates are in grey.

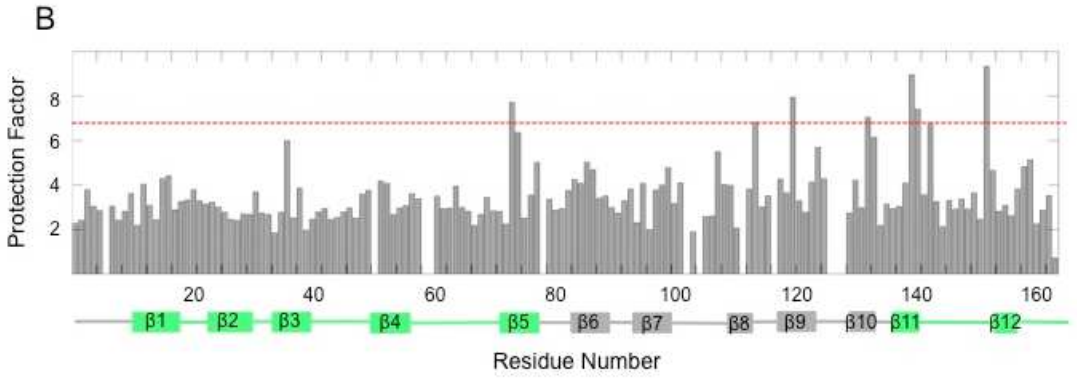

Figure 5-5: Bar diagram representation of protection factor distribution along the structure of IL-33.

The distribution of protection factors along the sequence of IL-33 are represented as a bar diagram where the y-axis represents the magnitude of the protection factor. The dashed line represents the cutoff where residues must have protection factors above the line to exchange from global unfolding. The sequence of IL-33 is shown below the x-axis where squares represent the β-strands within the structure. The high protection factors are distributed mainly in the regions of β-strands 5 through 12, which are located in Trefoils 2 and 3, respectively. The majority of the N-terminal end of the protein, between β-strands 1 and 4 (Trefoil 1) is populated by low protection factors, indicating a region of fast exchange.

35), the loop between β-strands 3 and 4 (residues 40-47), the loop between β-strands 4 and 5 (residues 55-70), and the loop between β-strands 11 and 12 (residues 140-150). These regions are flexible, and in particular the region between β-strands 4 and 5 has $\{H^1\}$-^{15}N NOE values that are low enough that they behave similarly to an unstructured peptide in solution. Further, the low $\{H^1\}$-^{15}N NOE values across the structure strongly indicate motions on the ps-ns timescale. This is consistent with low R2 values along the sequence that correlate with the $\{H^1\}$-^{15}N NOE values below 0.6, further indicating the presence of ps-ns dynamics spanning IL-33 (Figure 5-6 top and 5-6 middle).

The R1, R2, and $\{H^1\}$-^{15}N NOE relaxation data sets at 801.45 and 500.18 MHz were further analyzed using Lipari-Szabo model-free analysis to evaluate the conformational entropy associated with the protein structure [37]. Using the software of eMF [24] we estimated a rotational correlation time (τ_c) of IL-33 through the use of an axially symmetric model and iterative Bayesian analysis of the tensor fits. The τ_c, or rotational correlation time, calculated to 10.7 was consistent with the observed dynamic behavior and molecular weight of IL-33 (Figure 5-7) [38–41]. The resulting order parameters (S^2) showed distinct regions of conformational entropy described by regions with order parameters much lower than 0.8, the cutoff for rigid or ordered residues. The regions with low order parameters include the N-terminus as well as unstructured loops that link together the β-strands of IL-33. These dynamic loops span between β-strands 2 and 3 (residues 28-35), β-strands 3 and 4 (residues 40-48), β-strands 4 and 5 (residues 55-70), and β-strands 11 and 12 (residues 140-150). Although β- strands typically display more order than loops, the region spanning β-strands 2 and 3 display overall lower S^2 values across the sequence (S^2 less than 0.6) indicative of conformational flexibility (Figure 5-8).

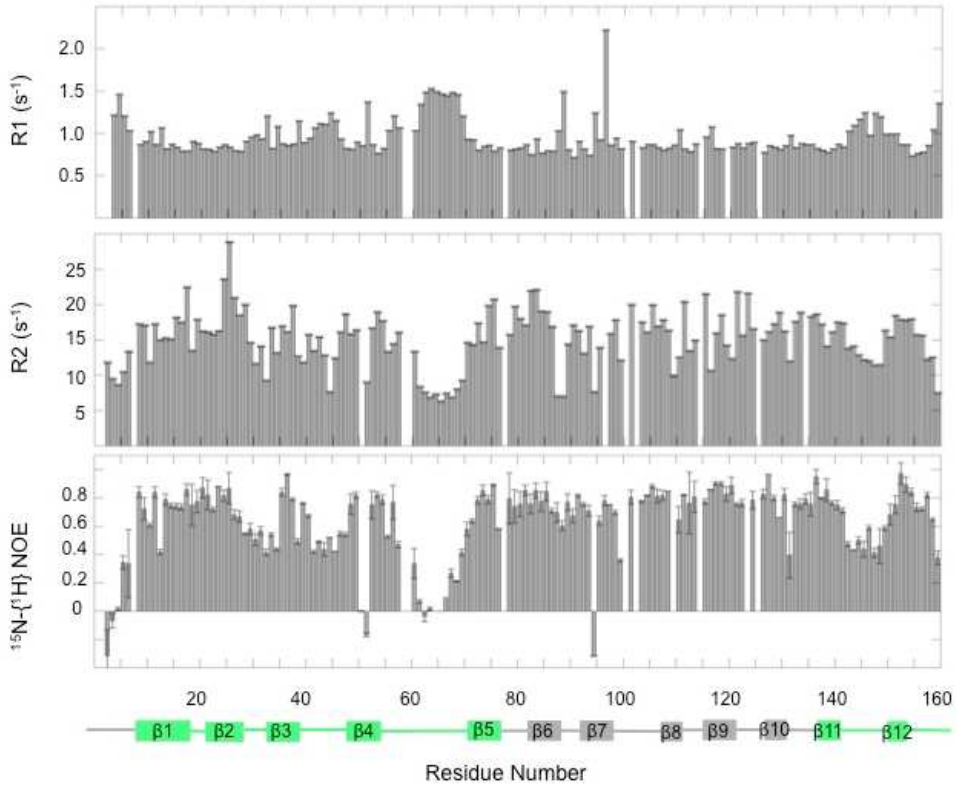

Figure 5-6: Backbone Relaxation Data.

(Top) R1, (Center) R2, and (Bottom) $\{H^1\}$-^{15}N NOE data for IL-33 collected at 801.45 MHz, 25°C as described in Methods. The structure of IL-33 is represented below the charts where the squares represent the β-strands in the sequence and the binding interface is colored in green.

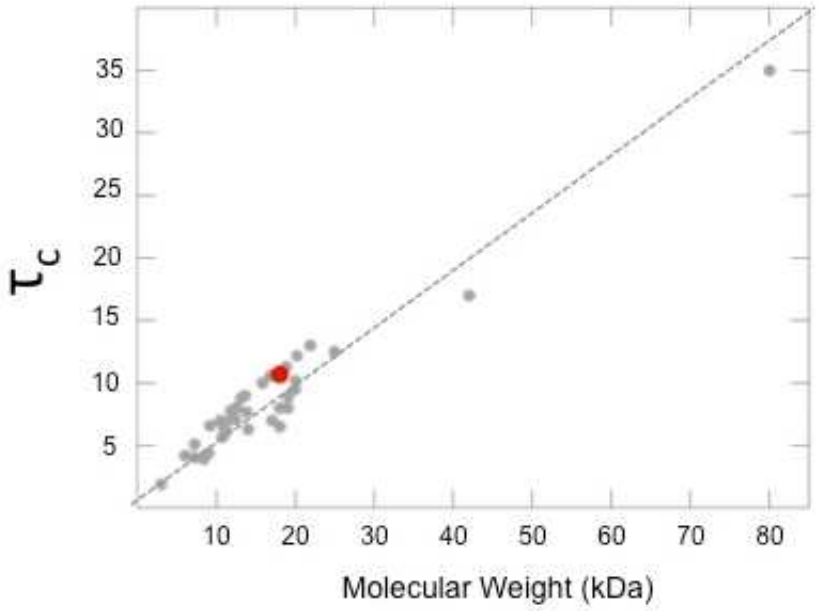

Figure 5-7: The Rotational Correlation Time of IL-33 versus the Molecular Weight.

A plot of the rotational correlation times (τ_c) of assorted proteins (y-axis) versus the molecular weights in terms of kDa (x-axis) is represented above. The rotational correlation time of IL-33 versus the molecular weight is represented as a red circle along the trend line of the correlation between the τ_c and the molecular weights. The τ_c of IL-33 relative to the molecular weight and observed dynamic behavior is in good agreement with the trend expected.

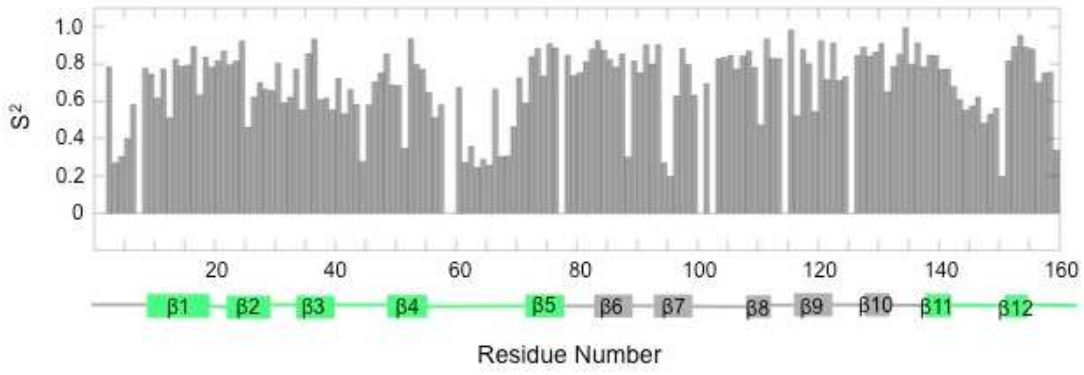

Figure 5-8: Backbone Order Parameters.

The order parameters for IL-33 (S^2) are represented relative to the sequence of IL-33. The structure is represented below the chart where the squares represent the β-strands in the sequence and the binding interface is colored in green. Missing points are from unassigned residues.

5.4.4 Conformational Flexibility in Interleukin-33 corresponds to conformational frustration in the protein.

Based upon the distinct regions of conformational entropy along the protein sequence, frustration analysis was performed using the frustratometer software found at (http://www.frustratometer.tk/). The frustration of a given structure is evaluated as the distribution of energy across the protein structure with an algorithm designed around the funneled energy landscape of protein folding [29]. The conformational frustration is a metric of the energy specific contacts make compared to the energetics these contacts could make in spatially proximal, alternate compact structures [44] . Within IL-33, there are multiple regions with distinct conformational frustration (Figure 5-9). These regions show strong correlation with regions that have low order parameters (S^2 less than 0.6), $p < 0.001$ by a paired t-test indicating a link between conformational frustration and conformational entropy (Figure 5-9). These regions include the loop between β-strands 2 and 3 (residues 25-35), the loop between β-strands 4 and 5 (residues 55-70), and the loop between β-strands 11 and 12 (residues 140-150).

5.5 Discussion

In analyzing the protection factors associated with the structure of IL-33, it is clear that regions of protection from deuterium exchange are minimal overall. Of the 160 residues that make up the β-trefoil fold, only 12 residues exchange to deuterium on long timescales, and of those 12 only 6 have protection factors that correspond to global unfolding of the protein, or exchange by an EX1 mechanism. These residues

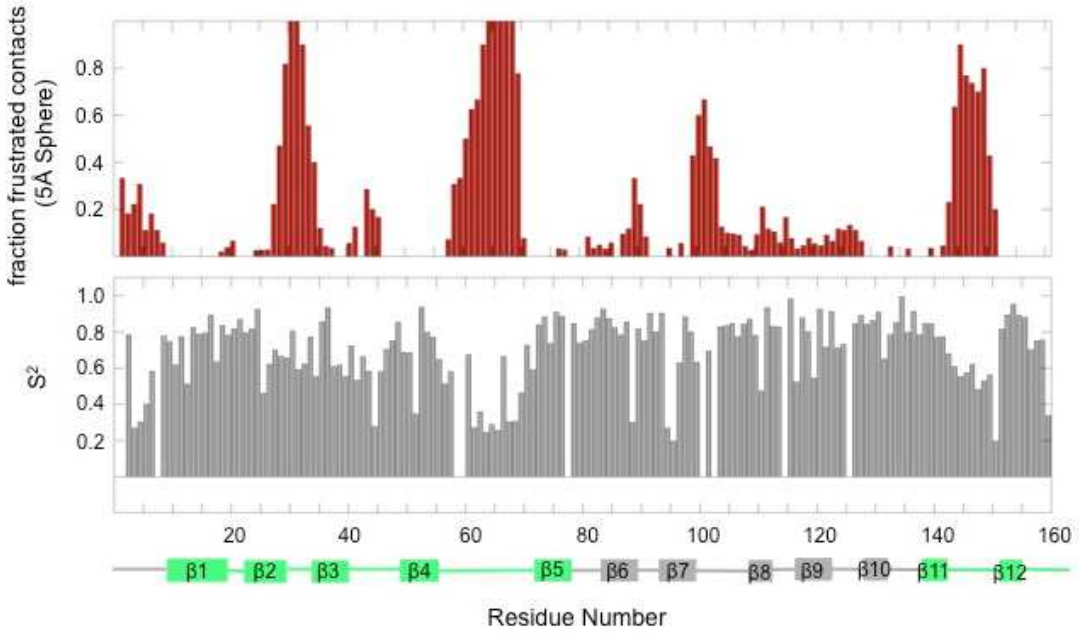

Figure 5-9: Correlation between order parameters and the conformational frustration in IL-33.

(Top) Represents the conformational frustration on a scale from 0 to 1 where 1 represents maximum frustration of a residue. (Bottom) Represents the order parameters where on a scale from 0 to 1, 1 represents the most ordered or rigid and 0 represents the most entropy, or motion. The regions of conformational frustration and configurational entropy are correlated with $p < 0.001$*
*using paired t-test

include M73, V118, I130, L137, A138, and C149. Notably, these residues are only found within the structure in regions spanning Trefoils 2 and 3 and do not represent any regions within Trefoil 1 (Figure 5-4 and 5-5) which shows that there are more local unfolding events in Trefoil 1, meaning that it represents a more malleable portion of the protein structure overall. As we suggested in previous work, along the folding route of IL-33, Trefoil 1 is more difficult to fold due to geometric frustration. This subsequent frustration may be a factor in allowing for more malleability and local unfolding events within Trefoil 1.

To evaluate the fast timescale dynamics in the native state of IL-33 in addition to the long-range time scale dynamics, we directly probed for the configurational entropy across the structure of IL-33. Further, we sought to evaluate a possible link between dynamics and conformational frustration patterns among functional regions of the protein. Through the suite of R1, R2, and $\{H^1\}$-^{15}N NOE experiments, we found multiple regions of heterogeneous relaxation values and $\{H^1\}$-^{15}N NOE values, suggesting the presence of dynamic motions across the structure on the ps-ns timescale (Figure 5-6). These regions include the loops between β-strands 2 and 3, the loop between β-strand 3 and 4, and the loop between β-strands 4 and 5. These loops are all located in Trefoil 1 and together span a significant portion of the ST2 binding interface. Additionally, there is a second region of heterogeneous dynamics at the C-terminal end of the protein between β-strands 11 and 12 (residues 140 to 150) that represents an additional region implicated in engaging to the ST2 receptor. By evaluating this set of experiments with the Lipari-Szabo formalism we were able to show that the large span of Trefoil 1 displays significant conformational entropy and dynamic heterogeneity in comparison with the rest of the structure (Figure 5-8 and 5-10). Frequently, this region sampled S^2 values well below 0.6, indicative of

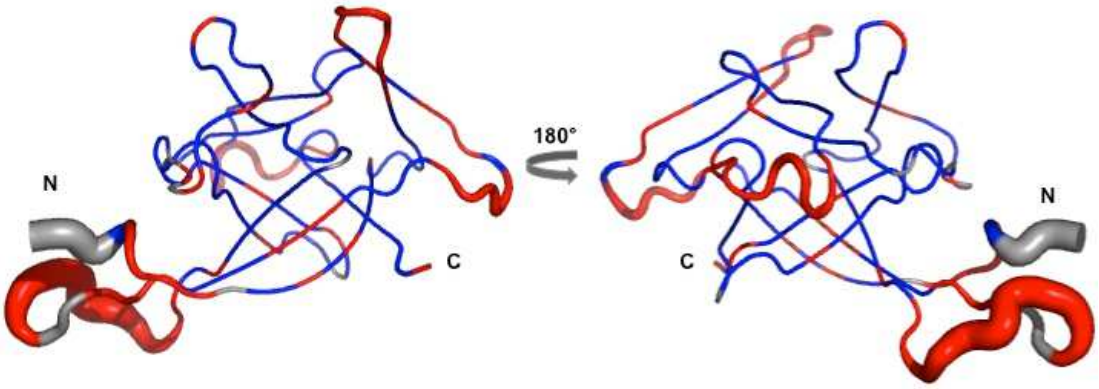

Figure 5-10: Order parameters mapped onto the structure of IL-33.

The order parameters are mapped onto a sausage model of IL-33 (PDB code 2KLL) where residues with S^2 values less than 0.6 are mapped in red and S^2 values above 0.6 are mapped in blue. Unobserved residues are colored grey.

conformational entropy necessitated by ps-ns motions. This region of IL-33 with conformational entropy is critical for engaging the Ig domains of ST2. The Ig2 and Ig3 domains of ST2 is connected by a flexible linker and in of themselves represent a highly dynamic interface that IL-33 is responsible for engaging. IL-33 must engage all three Ig domains of ST2 through the whole of Trefoil 1 and the terminal end of Trefoil 3. In order to engage the highly dynamic Ig domains in ST2, IL-33 may be dynamic on multiple timescales to facilitate malleability within the binding interface to effectively engage the highly entropic receptor Ig domains.

The dynamic functional region of IL-33 may represent an evolutionarily conserved region necessary for engagement of the equally dynamic ST2 receptor. In order to probe for this link, we compared the order parameters with the configurational frustration index to assess a possible correlation. The regions with low order parameters (S^2), and thus high configurational entropy correlate closely with regions of configurational frustration, $p < 0.001$ by paired t-test (Figure 5-9). In particular, Trefoil 1 is predicted to be highly frustrated on the configurational landscape as is the C-terminal end between β-strands 11 and 12. Significantly the conformational entropy of the order parameters along the structure, the configurational frustration, and the local unfolding events all track to the functional regions of the protein, in particular the large binding interface spanning Trefoil 1 (Figure 5-10). The combination of residual configurational frustration and dynamics on multiple timescales within the functional regions of IL-33 suggests that the dynamics and frustration is a conserved feature necessary for the functionality of the protein.

Chapter 5, in part, is a reprint of the material as it appears in **“**Heterogeneous Dynamics and Conformational Frustration in the Binding Regions of Interleukin-33”

Fisher, Kaitlin M., Fuglestad, Brian, Chan, Josh C., Jennings, Patricia A which is in preparation for submission. The dissertation author was the primary author/investigator of this paper.

5.6 References

[1] Dinarello C a. Immunological and inflammatory functions of the interleukin-1 family. Annu Rev Immunol 2009;27:519–50.

[2] Buckley CD, Gilroy DW, Serhan CN, Stockinger B, Tak PP. The resolution of inflammation. Nat Rev Immunol 2013;13:59–66.

[3] Marrack P, Kappler J, Kotzin BL. Autoimmune disease: why and where it occurs. Nat Med 2001;7:899–905.

[4] Maderna P, Godson C. Phagocytosis of apoptotic cells and the resolution of inflammation. Biochim Biophys Acta - Mol Basis Dis 2003;1639:141–51.

[5] Coussens LM, Werb Z. Inflammation and cancer. Nature 2002;420:860–7.

[6] Pei C, Barbour M, Fairlie-Clarke KJ, Allan D, Mu R, Jiang HR. Emerging role of interleukin-33 in autoimmune diseases. Immunology 2014;141:9–17.

[7] Cavallo F, De Giovanni C, Nanni P, Forni G, Lollini PL. 2011: The immune hallmarks of cancer. Cancer Immunol. Immunother., vol. 60, 2011, p. 319–26.

[8] Deretic V, Saitoh T, Akira S. Autophagy in infection, inflammation and immunity. Nat Rev Immunol 2013;13:722–37.

[9] Shi C, Pamer EG. Monocyte recruitment during infection and inflammation. Nat Rev Immunol 2011;11:762–74.

[10] Smith DE. IL-33 : a tissue derived cytokine pathway involved in allergic inflammation and asthma Experimental Allergy 2010:200–8.

[11] Mu R, Huang H-Q, Li Y-H, Li C, Ye H, Li Z-G. Elevated serum interleukin 33 is associated with autoantibody production in patients with rheumatoid arthritis. J Rheumatol 2010;37:2006–13.

[12] O'Neill LAJ. The interleukin-1 receptor/Toll-like receptor superfamily: 10 Years of progress. Immunol Rev 2008;226:10–8.

[13] Wang D, Zhang S, Li L, Liu X, Mei K, Wang X. Structural insights into the assembly and activation of IL-1β with its receptors. Nat Immunol 2010;11:905–11.

[14] Garlanda C, Dinarello CA, Mantovani A. The Interleukin-1 Family: Back to the Future. Immunity 2013;39:1003–18.

[15] Lingel A, Weiss TM, Niebuhr M, Pan B, Appleton B a., Wiesmann C, Bazan JF, Fairbrother WJ. Structure of IL-33 and Its Interaction with the ST2 and IL-1RAcP Receptors-Insight into Heterotrimeric IL-1 Signaling Complexes. Structure 2009;17:1398–410.

[16] Jovanovic IP, Pejnovic NN, Radosavljevic GD, Pantic JM, Milovanovic MZ, Arsenijevic NN, Lukic ML. Interleukin-33/ST2 axis promotes breast cancer growth and metastases by facilitating intratumoral accumulation of immunosuppressive and innate lymphoid cells. Int J Cancer 2014;134:1669–82.

[17] Villarreal DO, Weiner DB. Interleukin 33: a switch-hitting cytokine. Curr Opin Immunol 2014;28:102–6.

[18] Liu X, Hammel M, He Y, Tainer J a, Jeng U-S, Zhang L, Wang S, Wang X. Structural insights into the interaction of IL-33 with its receptors. Proc Natl Acad Sci U S A 2013;110:14918–23.

[19] Liu X, Hammel M, He Y, Tainer J a, Jeng U-S, Zhang L, Wang S, Wang X. Structural insights into the interaction of IL-33 with its receptors. Proc Natl Acad Sci U S A 2013;110:14918–23.

[20] Bai YW, Milne JS, Mayne L, Englander SW. Primary Structure Effects on Peptide Group Hydrogen-Exchange. Proteins-Structure Funct Genet 1993;17:75–86.

[21] Connelly GP, Bai Y, Jeng MF, Englander SW. Isotope effects in peptide group hydrogen exchange. Proteins 1993;17:87–92.

[22] Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. J Am Chem Soc 1982;104:4559–70.

[23] Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. J Am Chem Soc 1990;112:4989–91.

[24] Bae SH, Legname G, Serban A, Prusiner SB, Wright PE, Dyson HJ. Prion proteins with pathogenic and protective mutations show similar structure and dynamics. Biochemistry 2009;48:8120–8.

[25] Bae SH, Dyson HJ, Wright PE. Prediction of the rotational tumbling time for proteins with disordered segments. J Am Chem Soc 2009;131:6814–21.

[26] d'Auvergne EJ, Gooley PR. The use of model selection in the model-free analysis of protein dynamics. J Biomol NMR 2003;25:25–39.

[27] d'Auvergne EJ, Gooley PR. Model-free model elimination: A new step in the model-free dynamic analysis of NMR relaxation data. J Biomol NMR 2006;35:117–35.

[28] Farrow NA, Zhang O, Szabo A, Torchia DA, Kay LE. Spectral density function mapping using 15N relaxation data exclusively. J Biomol NMR 1995;6:153–62.

[29] Jenik M, Parra RG, Radusky LG, Turjanski A, Wolynes PG, Ferreiro DU. Protein frustratometer: A tool to localize energetic frustration in protein molecules. Nucleic Acids Res 2012;40.

[30] Gosavi S, Chavez LL, Jennings PA, Onuchic JN. Topological frustration and the folding of interleukin-1β. J Mol Biol 2006;357:986–96.

[31] Gosavi S. Understanding the folding-function tradeoff in proteins. PLoS One 2013;8:e61222.

[32] Finke JM, Jennings P a. Interleukin-1?? folding between pH 5 and 7: Experimental evidence for three-state folding behavior and robust transition state positions late in folding. Biochemistry 2002;41:15056–67.

[33] Bai Y, Milne JS, Mayne L, Englander SW. Protein stability parameters measured by hydrogen exchange. Proteins 1994;20:4–14.

[34] Best RB, Vendruscolo M. Structural interpretation of hydrogen exchange protection factors in proteins: characterization of the native state fluctuations of CI2. Structure 2006;14:97–106.

[35] Maity H, Lim WK, Rumbley JN, Englander SW. Protein hydrogen exchange mechanism: local fluctuations. Protein Sci 2003;12:153–60.

[36] Haglund E, Lindberg MO, Oliveberg M. Changes of protein folding pathways by circular permutation: Overlapping nuclei promote global cooperativity. J Biol Chem 2008;283:27904–15.

[37] Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. J Am Chem Soc 1982;104:4559–70.

[38] Raman S, Lange OF, Rossi P, Tyka M, Wang X, Aramini J, Liu G, Ramelot TA, Eletsky A, Szyperski T, Kennedy MA, Prestegard J, Montelione GT, Baker

D. NMR structure determination for larger proteins using backbone-only data. Science 2010;327:1014–8.

[39] Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 1989;28:8972–9.

[40] Cavanagh J. Protein NMR Spectroscopy: Principles and Practice. vol. 2nd. 2007.

[41] Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry 1994;33:5984–6003.

[42] Palmer AG. Probing molecular motion by NMR. Curr Opin Struct Biol 1997;7:732–7.

[43] Lefevre JF, Dayie KT, Peng JW, Wagner G. Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions. Biochemistry 1996;35:2674–86.

[44] Ferreiro DU, Hegler JA, Komives EA, Wolynes PG. Localizing frustration in native proteins and protein assemblies 2007;104.

Chapter 6 : IL-33 interacts with lipids through enriched regions of charged residues.

6.1 Abstract

Interleukin-33 (IL-33), a member of the Interleukin-1 family, is secreted from the cell through an unidentified mechanism broadly characterized as a Non-Classical Secretory pathway. In order to initiate signaling for varied immune responses, IL-33 must move through different cellular compartments in order to bind to its extracellular receptor. The mechanism of Non-Classical Secretion through which IL-33 is secreted from the cell is an open question. It has been shown that IL-33 can localize to vesicular bodies in high concentrations along the membrane upon cellular stimulation. We sought to identify whether IL-33 can interact with vesicles as a model of membrane localization and identify regions of the IL-33 that could potentially mediate this lipid interaction. Vesicle perturbation was found through a dosage dependent addition of IL-33 to vesicles. The specific regions of IL-33 interacting with the lipid interface represent highly charged residues within localized regions of the protein structure. The localization and interaction of IL-33 with the membrane could help facilitate the secretion of the protein from the cell. The physiological implications of charge lability and membrane interactions of IL-33 may lend the protein a sensitivity to factors that influence charge in the cell such as oxidative stress and specific disease states that influence cellular pH.

6.2 Introduction

The Interleukin-1 family of cytokines represents a unique subset of 11 proteins ubiquitously expressed in multiple cell and tissue types, displaying wide-ranging activities as mediators of the immune response [1]. Interleukin-33 (IL-33) is a more recently discovered member of the Interleukin-1 family of cytokines and plays a key

role in regulating the innate immune response across a wide variety of tissues including lung, smooth muscle, and endothelial cells [2–4]. IL-33, like another well-studied member of the family, IL-1β, lacks an N-terminal leader sequence and as a result bypasses the traditional extra-cellular secretion pathway of processing through the ER and Golgi Apparatus [5,6]. It has also been demonstrated that the C-terminal, β-trefoil region of IL-33 is necessary and sufficient to illicit strong, active signaling through the ST2 receptor complex [7]. Unique to the Interleukin-1 family of proteins, however, the N-terminal pro sequence of IL-33 has a DNA binding motif and histone binding region which has been shown to up-regulate NFκB expression [8,9]. The mechanism of IL-33 processing into its two constitutive halves and the pathway of secretion of IL-33 from the cell to initiate paracrine signaling is still an open question.

IL-33 and IL1β are secreted from the cell through an unidentified mechanism broadly characterized as a Non-Classical Secretory pathway. Proteins that are classically secreted contain an N-terminal leader sequence, facilitating targeted processing through the Endoplasmic Reticulum and Golgi apparatus for cellular export [10,11]. Proteins that lack the hydrophobic, N-terminal leader sequence are secreted from the cell through an alternate mechanism that in many cases remains uncharacterized. IL-1β and IL-33 both lack an N-terminal leader sequence and are able to bypass the classical secretory pathway [12,13]. The mechanism by which these cytokines are secreted from the cell without an N-terminal leader sequence is not currently understood.

In order to initiate signaling for varied immune responses, both IL-33 and IL-1β must move through different cellular compartments in order to bind to their respective extra cellular receptors. It has been shown that both IL-1β and IL-33 localize to

vesicular bodies in high concentrations along the membrane upon cellular stimulation [10,14–16]. The mechanism of IL-33 secretion from the cell is not currently understood, however many of the proposed mechanisms involve localization to the cellular membrane. The proposed mechanisms of secretion include encapsulation in multivesicular bodies, microvesicle shedding, the use of secretory lysosomes, or through a specific transporter (Figure 6-1) [13]. In addition to localizing similarly in vesicles, IL-33 and IL-1β share structural homology [17]. Taken in combination, we sought to identify whether IL-33 can interact with model membranes and the regions of the protein that could mediate this interaction. Given the ubiquitous presence of IL-33 in many tissue types and its large range of effects on inflammatory responses, a better understanding of IL-33 transport to signaling targets is of importance. In the case of IL-33, passing across both the outer cell membrane and the nuclear membrane is necessary due to the dual functionality of IL-33 in DNA binding and extra cellular signaling [9]. Regulation of the Interleukin-1 family of cytokines is partly mediated by tightly controlled cellular localization, processing, and secretion. Thus, cellular regulation of the transport of IL-1β and IL-33 plays a key role in controlling the relative effects of these cytokines with respect to initiating the immune response.

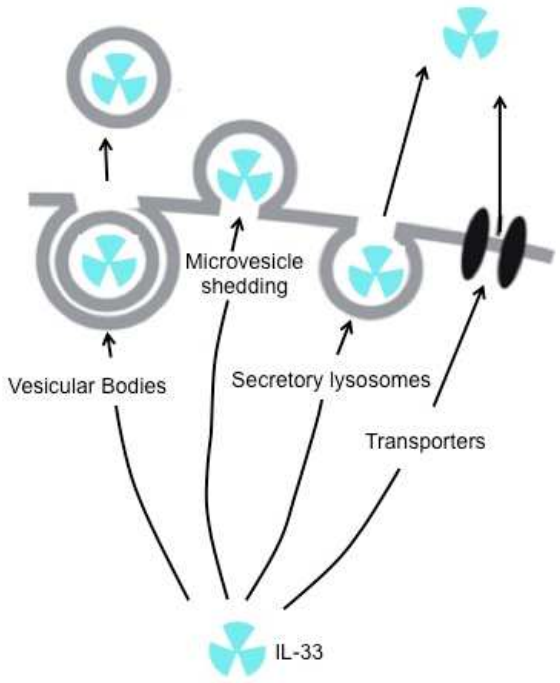

Figure 6-1: Proposed Mechanisms of IL-33 secretion from the cell.

The exact mechanism of IL-33 secretion from the cell is not currently understood, however many of the proposed mechanism involve localizing to the cellular membrane. These mechanisms include through encapsulation in multivesicular bodies, microvesicle shedding, through secretory lysosomes, or through a transporter that is yet to be identified.

6.3 Materials and Methods

6.3.1 Protein Expression, Cloning, and Purification:

Recombinant human IL-33 was expressed and purified as described previously (Chapter 3, materials and methods). Recombinant IL-1β was expressed and purified as described previously [18–20]. All experiments were run with protein in 20 mM HEPES, 100 mM NaCl, pH 7.2.

6.3.2 Vesicle Leakage Assay:

Large unilamillar vesicles comprised of a 70:30 ratio of L-α-phosphatidylcholine (PC) and L-α-phosphatidyl-DL-glycerol (PG) at a 30mM concentration were doped with an equimolar concentration (50mM) of the fluorescent dye 8-aminonapthalene-1,3,6 trisulfonic acid (ANTS) and the quencher p-xylene-bis-pyridinium bromide (DPX). Vesicles were incubated with protein at varying concentrations in the μM range in 20mM HEPES, 100mM NaCl, pH 7.2 for 1 hour to 24 hours. Perturbation was assessed using the fluorescence emission in the range of 450nm to 580nm after excitation at 360nm on a Horiba Jobin Yvon fluoromax-4 fluorimeter. Change in fluorescence from ANTS was normalized to vesicles in the presence of trypsin and leptin, non-lipid interacting proteins of similar molecular weight to IL-33 and IL-1β.

6.3.3 NMR Spectroscopy:

^{1}H-^{15}N HSQCs of IL-33 in the presence and absence of bicelles was assessed on a Varian 500 MHz spectrometer. Bicelles were prepared with a Q ratio of 0.25

using 6-OPC (1,2-di-*O*-hexyl-*sn*-glycero-3-phosphocholine) and 14-OPC (1,2-di-*O*-tetradecyl-*sn*-glycero-3-phosphocholine) lipids. Several freeze thaw cycles of the lipid mixture was undertaken to allow for self-assembly under the correct phase conditions. An ^{15}N sample of IL-33 was prepared and split equally into a sample with 50 % added buffer or 50 % added bicelles. NMR samples were reassessed upon addition of all samples to ensure pH consistency. Data was processed using the nmrPipe and Sparky software suites.

6.4 Results and Discussion:

6.4.1 IL-33 leads to pronounced leakage of dye from vesicles:

Both IL-1β and IL-33 are shown to localize along the outer cell membrane in vesicular like bodies upon cell stimulation. To assess the possibility of a direct interaction between the IL-33 and the lipid membrane, we utilized a vesicle leakage assay to assess the degree of lipid interaction between IL-33 and IL-1β using the fluorescent dye ANTS as a reporter combined with DPX quencher. When encapsulated together within the core of large unilamillar vesicles, the DPX quencher is in close enough proximity to the ANTS fluorescent dye to quench the fluorescent signal. Upon a significant enough perturbation of the vesicle, the ANTS dye may leak, resulting in dilution to the surrounding buffer [21]. The loss in close proximity to the DPX results in the loss of the quenching effect and leads to an increased signature fluorescent emission profile of the dye ANTS.

The relative stability of the doped vesicles was assessed and found to be stable over a 36-hour period. Doped vesicles were then incubated in addition to detergent to assess the relative increase in ANTS fluorescence. As seen in Figure 6-

2, the addition of Triton-X resulted in an increase in fluorescent signature above the base line of vesicles alone. This was indicative that the DPX/ANTS pair was encapsulated in vesicles and that perturbation of the lipid membrane did in fact result in an increase in fluorescence.

Upon establishing the encapsulation within vesicles and effective quenching of ANTS/DPX, IL-33 and IL-1β was incubated with vesicles and subsequently assessed for their relative perturbation of the vesicle wall with respect to Triton-X. Three negative controls for membrane interaction were assessed: Interleukin-1 Receptor antagonist, Leptin, and Trypsin. Interleukin-1 Receptor antagonist is an Interleukin-1 family member that has a homologous β-trefoil structure to IL-33 and IL-1β [22]. IL-1Ra contains an N-terminal leader sequence and is processed by the ER and Golgi apparatus and does not utilize a non-classical secretion mechanism [23–25]. Leptin is a lipid-associated protein that does not have the β-trefoil fold shared by IL-33 and IL-1β and is a classically secreted protein [26,27]. Finally, Trypsin is a non- lipid associated protein, does not have a β-trefoil fold, and is classically secreted [28,29].

As shown in Figure 6-3A, both IL-33 and IL-1β are both able to induce the leakage of ANTS dye from vesicles. IL-33 shows the most potency, indicating the highest degree of perturbation to the vesicular membrane. Although to a lesser degree than IL-33, IL-1β is also able to induce the leakage of ANTS from vesicles, indicating that it too interacts with vesicles. Notably, IL-1Ra, an interleukin-1 family member with a homologous β-trefoil fold to both IL-33 and IL-1β was unable to produce increased

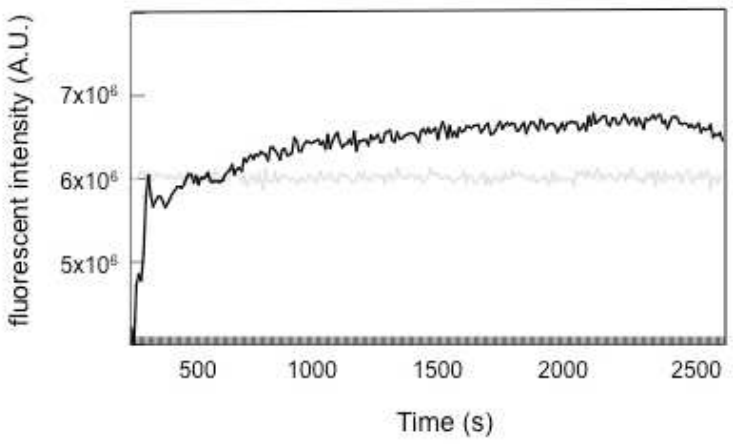

Figure 6-2: Vesicle Stability over time versus Vesicle leakage induced by Triton X detergent.

In order to verify the successful formation of vesicles encapsulating ANTS and DPX dye and quencher, the relative fluorescent signatures were monitored in the presence of Triton X (black trace) and no addition of detergent (grey trace). The vesicles do not experience an increased fluorescent signature until the addition of Triton X, indicative of successful dye encapsulation.

fluorescence in the sample, indicating a limited to non-existent interaction with vesicles.

As seen in Figure 6-3B, the trend for leakage potency remains consistent over a 24-hour time course for IL-33 and IL-1β. IL-33 is able to induce the most fluorescent signature and IL-1β maintains its efficacy. IL-1ra shows potency in line with all other negative protein controls used. Slight perturbations from surfactant effects may result in small amounts of vesicle leakage seen in controls. Taken together, the vesicle leakage data shows that both IL-33 and IL-1β are able to interact with vesicles and that this interaction is significant enough to cause the structural integrity of the membrane to be compromised.

6.4.2 NMR analysis shows chemical shift perturbations in charged regions of IL-33.

Based on the leakage of dye caused by the dosage dependent addition of IL-33 to vesicles, we wanted to evaluate the specific regions of IL-33 mediating the interaction with the lipid interface. To achieve this we evaluated the differences in ^{15}N backbone resonances through the use of ^{1}H-^{15}N HSQCs in the presence of bicelles (Figure 6-4). Analysis of chemical shift perturbations show that regions with the most significant shifts upon the addition of bicelles are centered on two specific regions encompassing residues 45-48 and residues 85-96, respectively (Figure 6-5). With a cutoff for chemical shift significance set above a $\Delta\delta$ of 0.4, the select residues that experience a significant chemical shift perturbation are Thr 4, Gly 5, Lys 45, Lys 46, Asp 47, Lys 48, His 85, Glu 90, Ser 92, and His 96. With the exception of the residues found on the C-terminal end of the sequence and Ser 92, these residues are all highly charged. That charge is a potential mediating factor for interaction with a lipid interface

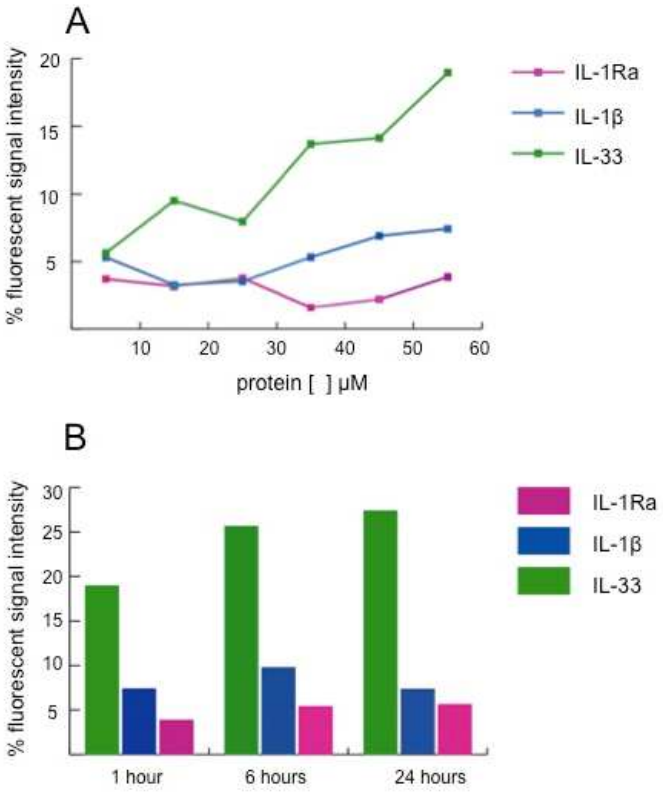

Figure 6-3: IL-33 incubation with vesicles containing ANTS and DPX shows potent leakage of dye, suggesting significant interaction with lipids.

(A) Incubation of IL-33 at concentrations ranging from 10 to 60 µM shows a concentration dependent increase in fluorescent signature (green) versus IL-1β (blue) and IL-1Ra (purple). Values of fluorescent signature are represented as a percent increase over the addition of the negative control, Trypsin. (B) Incubation of IL-33 with vesicles at time points of 1 hour, 6 hours, and 24 hours shows persistence of an increased fluorescent signature (green) versus IL-1β (blue) and IL-1Ra (purple). The induction of vesicle leakage is much higher over time than that induced by both IL-1β and IL-1Ra. Values of fluorescent signature are again represented as a percent increase over the addition of the negative control, Trypsin.

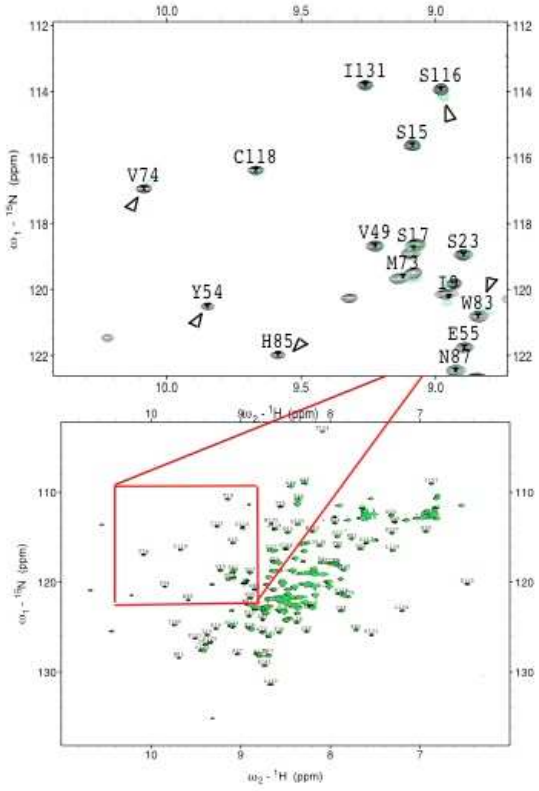

Figure 6-4: Overlay of the ^{1}H-^{15}N HSQC spectra of IL-33 in the presence and absence of Bicelles.

The ^{1}H-^{15}N HSQC spectra of IL-33 is shown in black overlayed with the ^{1}H-^{15}N HSQC spectra of IL-33 with bicelles shown in green. Only a fraction of the assignments are displayed for clarity. A close up section of the spectra is shown to highlight regions of peak differences between the spectra with and without bicelles.

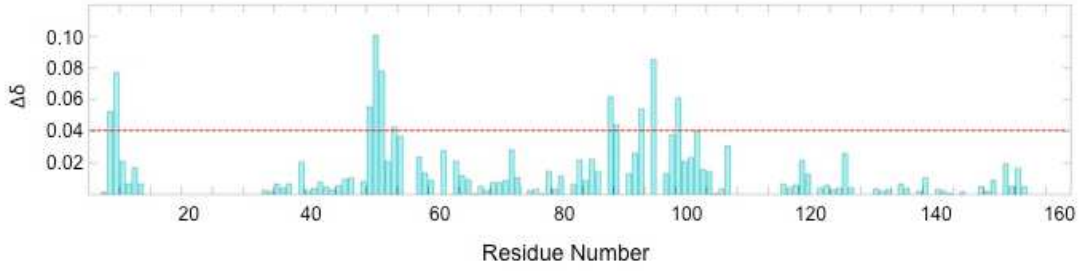

Figure 6-5: Chemical Shifts in IL-33 in the presence of Bicelles show significant shifts centered around charged regions.

Chemical shift differences are plotted for IL-33 in the presence of bicelles. Residues with a Δδ above 0.4 indicate residues with significant perturbation upon the addition of bicelles. These residues include Thr 4, Gly 5, Lys 45, Lys 46, Asp 47, Lys 48, His 85, Glu 90, Ser 92, and His 96 and represent mostly charged residues.

is likely in the case of IL-33 as it is a highly charged protein, with an enrichment of both Lysines and Histidines. IL-33 has more histidines than most other β-trefoil proteins and is second only to Hisactophilin, a protein largely composed of Histidines known to interact with membranes through pockets of charged residue enrichment similar to those seen in IL-33 (Lys, His, and Asp) [22,30,31]. It may be that the enrichment of charged residues in IL-33 allow for electrostatic partitioning of IL-33, aiding in its ability to localize to the membrane interface.

The current data collected suggests that the inherent structure of IL-33 allows for direct interaction with lipid bi-layers through an electrostatic interaction. This interaction may be part of the inherent ability of IL-33 to localize to the membrane interface. Both the dosage dependent perturbation of the model membranes in combination with the chemical shift perturbations upon bicelle addition shows that IL-33 is interacting with the model membrane surfaces. Chemical shift perturbations monitored with NMR in the presence of bicelles shows that the shifts occur in areas that are highly charged, composed of lysines, histidines, and aspartic acid in particular (Figure 6-6). The membrane leakage caused by IL-33 appears to be an effect of electrostatics and partitioning of the protein to the charge potential at the surface of lipid membranes. The ability of IL-33 to localize to the membrane and interact through charge could facilitate the secretion of the protein through a still uncharacterized non-classical secretory mechanism. The physiological implications of charge lability and membrane interactions for IL-33 may be numerous beyond the ability of IL-33 to be secreted. The charge states of the residues that interact with lipids will be dependent on pH. The pH of a cell is very dependent on factors such as oxidative stress and certain disease states that influence the environment of different cellular

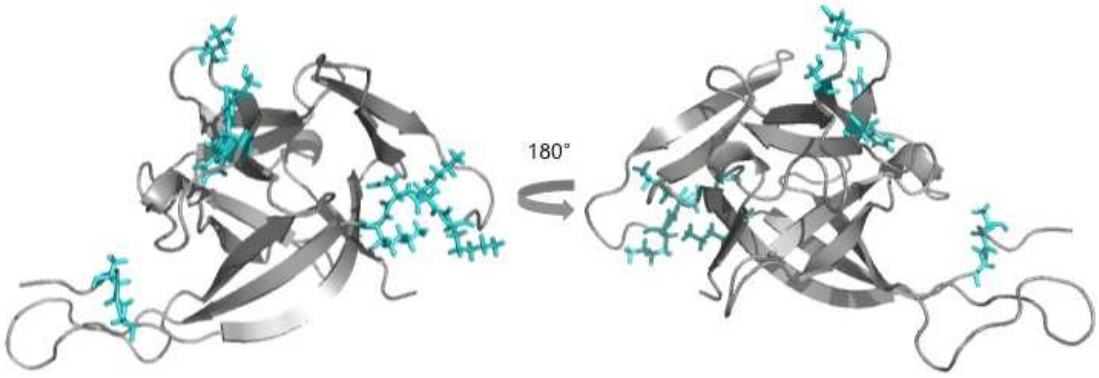

Figure 6-6: Charged Residues with perturbations upon addition of lipids show clustering of charged residue regions.

Residues with a Δδ above 0.4 are shown on the structure of IL-33 (Protein Data Bank code 2KLL) in cyan. All other residues are shown in grey.

compartments, which could subsequently affect the charge state of IL-33 and its ability to localize to the cellular membrane.

6.5 References

[1] Dinarello CA. Immunological and inflammatory functions of the interleukin-1 family. Annu Rev Immunol 2009;27:519–50.

[2] Pei C, Barbour M, Fairlie-Clarke KJ, Allan D, Mu R, Jiang HR. Emerging role of interleukin-33 in autoimmune diseases. Immunology 2014;141:9–17.

[3] Mu R, Huang H-Q, Li Y-H, Li C, Ye H, Li Z-G. Elevated serum interleukin 33 is associated with autoantibody production in patients with rheumatoid arthritis. J Rheumatol 2010;37:2006–13.

[4] Kurowska-Stolarska M, Hueber A, Stolarski B, McInnes IB. Interleukin-33: A novel mediator with a role in distinct disease pathologies. J. Intern. Med., vol. 269, 2011, p. 29–35.

[5] Zhao W, Hu Z. The enigmatic processing and secretion of interleukin-33. Cell Mol Immunol 2010;7:260–2.

[6] Ali S, Nguyen DQ, Falk W, Martin MU. Caspase 3 inactivates biologically active full length interleukin-33 as a classical cytokine but does not prohibit nuclear translocation. Biochem Biophys Res Commun 2010;391:1512–6.

[7] Lefrancais E, Roga S, Gautier V, Gonzalez-de-Peredo a., Monsarrat B, Girard J-P, Cayrol C. IL-33 is processed into mature bioactive forms by neutrophil elastase and cathepsin G. Proc Natl Acad Sci 2012;109:1673–8.

[8] Villarreal DO, Weiner DB. Interleukin 33: a switch-hitting cytokine. Curr Opin Immunol 2014;28:102–6.

[9] Ali S, Mohs A, Thomas M, Klare J, Ross R, Schmitz ML, Martin MU. The dual function cytokine IL-33 interacts with the transcription factor NF-κB to dampen NF-κB-stimulated gene transcription. J Immunol 2011;187:1609–16.

[10] Nickel W. The mystery of nonclassical protein secretion. Eur J Biochem 2003;270:2109–19.

[11] Hasegawa H, Thomas HJ, Schooley K, Born TL. Native IL-32 is released from intestinal epithelial cells via a non-classical secretory pathway as a membrane-associated protein. Cytokine 2011;53:74–83.

[12] MacKenzie A, Wilson HL, Kiss-Toth E, Dower SK, North RA, Surprenant A. Rapid secretion of interleukin-1?? by microvesicle shedding. Immunity 2001;15:825–35.

[13] Lloubes R, Bernadac A, Houot L, Pommier S. Non classical secretion systems. Res Microbiol 2013;164:655–63.

[14] Kakkar R, Hei H, Dobner S, Lee RT. Interleukin 33 as a mechanically responsive cytokine secreted by living cells. J Biol Chem 2012;287:6941–8.

[15] Prudovsky I, Mandinova A, Soldi R, Bagala C, Graziani I, Landriscina M, Tarantini F, Duarte M, Bellum S, Doherty H, Maciag T. The non-classical export routes : FGF1 and IL-1 α point the way 2003.

[16] Pathway I-I-, Weber A, Wasiliew P, Kracht M. Interleukin-1 (IL-1) Pathway 2012;1.

[17] Garlanda C, Dinarello CA, Mantovani A. The Interleukin-1 Family: Back to the Future. Immunity 2013;39:1003–18.

[18] Heidary DK, Gross LA, Roy M, Jennings PA. Evidence for an obligatory intermediate in the folding of interleukin-1 beta. Nat Struct Biol 1997;4:725–31.

[19] Finke JM, Jennings P a. Interleukin-1?? folding between pH 5 and 7: Experimental evidence for three-state folding behavior and robust transition state positions late in folding. Biochemistry 2002;41:15056–67.

[20] Capraro DT, Gosavi S, Roy M, Onuchic JN, Jennings P a. Folding circular permutants of IL-1β: Route selection driven by functional frustration. PLoS One 2012;7:e38512.

[21] Ladokhin AS, Wimley WC, White SH. Leakage of membrane vesicle contents: determination of mechanism using fluorescence requenching. Biophys J 1995;69:1964–71.

[22] Habazettl J, Gondol D, Wiltscheck R, Otlewski J, Schleicher M, Holak TA. Structure of hisactophilin is similar to interleukin-1 beta and fibroblast growth factor. Nature 1992;359:855–8.

[23] Stockman BJ, Scahill TA, Strakalaitis NA, Brunner DP, Yem AW, Deibel MR. Solution structure of human interleukin-1 receptor antagonist protein. FEBS Lett 1994;349:79–83.

[24] Dinarello CA. Interleukin-1, interleukin-1 receptors and interleukin-1 receptor antagonist. Int Rev Immunol 1998;16:457–99.

[25] Arend WP, Malyak M, Guthridge CJ, Gabay C. Interleukin-1 receptor antagonist: role in biology. Annu Rev Immunol 1998;16:27–55.

[26] Zhou Y, Rui L. Leptin signaling and leptin resistance. Front Med 2013:1–16.

[27] Zhang F, Basinski MB, Beals JM, Briggs SL, Churgay LM, Clawson DK, DiMarchi RD, Furman TC, Hale JE, Hsiung HM, Schoner BE, Smith DP, Zhang XY, Wery JP, Schevitz RW. Crystal structure of the obese protein leptin-E100. Nature 1997;387:206–9.

[28] Lohman RJ, O'Brien TJ, Cocks TM. Protease-activated receptor-2 regulates trypsin expression in the brain and protects against seizures and epileptogenesis. Neurobiol Dis 2008;30:84–93.

[29] Huber R, Kukla D, Bode W, Schwager P, Bartels K, Deisenhofer J, Steigemann W. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 A resolution. J Mol Biol 1974;89:73–101.

[30] Hanakam F, Gerisch G, Lotz S, Alt T, Seelig A. Binding of hisactophilin I and II to lipid membranes is controlled by a pH-dependent myristoyl-histidine switch. Biochemistry 1996;35:11036–44.

[31] Behrisch A, Dietrich C, Noegel AA, Schleicher M, Sackmann E. The actin-binding protein hisactophilin binds in vitro to partially charged membranes and mediates actin coupling to membranes. Biochemistry 1995;34:15182–90.

Chapter 7 : Conclusions and Future Directions

To understand the tie between the biophysical properties of a protein and the function of a protein, it is imperative to understand the native landscape. The native landscape encompasses the entire trajectory of a protein from the denatured state to the final, functional state. The folding route that a protein takes from a denatured amino acid chain to a native tertiary structure and the subsequent final functional ensemble of states that fluctuate between multiple, small native basins are integral in dictating the correct function of a given protein. Our full investigation of the native landscape of the cytokine IL-33 has shown, for the first time, that the folding route is intimately connected to the function of the protein and that the disparate dynamics within the native basin correlate not only to the functional regions of IL-33, but also back to the folding route.

IL-33 is a highly dynamic protein whose folding pathway is dictated by the geometrical frustration present in the functional region of the protein, Trefoil 1. The geometric frustration along the folding route lends malleability to Trefoil 1, a useful feature we hypothesize could allow for engagement of the highly flexible Ig2 and Ig3 domains in ST2. As we predicted from the folding route, Trefoil 1 has a high degree of conformational frustration in the native state. This frustration in the native state of IL-33 corresponds to the large amount of conformational entropy within the dynamic, ST2 binding regions. Together, the malleability imparted by increased local unfolding events, the conformational entropy, and the conformational frustration in Trefoil 1 may allow for the effective engagement, or pinning, of the separate and dynamic Ig2 and Ig3 domains in ST2. A dynamic transition between free and bound IL-33 may account for the need to preserve the geometric frustration and subsequent disparate dynamics inherent in Trefoil 1, despite the secondary effect of complicating the folding pathway and imparting conformational frustration within the native fold. The balance between

efficient folding and stable structure in IL-33 must be balanced with the preservation of dynamic motions within the functional region of Trefoil 1. Trefoil 1 must facilitate both effective binding of the separate Ig2 and Ig3 domains in ST2 and establish the basis of the dynamic transition upon binding that may be critical in modulating the functionality of the receptor complex.

The presence of conformational, or geometric, frustration in Trefoil 1 appears to be an evolutionarily retained feature along the folding route and functional, native basin ensemble. The implications of geometric frustration in Trefoil 1 is key in the selection of the three-state folding route, which we identified is necessary to counter balance the difficulty in folding the frustrated element of the protein. In order to further analyze the significance of the relationship between route selection and folding frustration, analyzing the folding kinetics of a group of permutations of IL-33 would be advantageous. The first useful variant of IL-33 would be a deletion in which the entire geometrically frustrated span of Trefoil 1 is eliminated in order to evaluate whether or not the stable core composed of Trefoil 2 and 3 is able to fold and maintain structure independently of Trefoil 1. The stable scaffold region of IL-33 would most likely fold quickly and through a two-state mechanism and would prove to be a useful test to verify that eliminating the entire geometrically frustrated region can lead to two-state folding behavior. Conversely, expressing Trefoil 1 alone to assess the folding kinetics may also prove useful. If the region spanning Trefoil 1 is as frustrated as we predict, then it would suggest that this region of the protein would be unable to fold independently of the stable scaffold composed of Trefoils 2 and 3. Without the stable scaffold that Trefoil 1 typically uses to dock during the modular folding mechanism, it is unlikely that Trefoil 1 would be able to overcome the geometric frustration to obtain a final fold or to maintain the final native fold in the native basin.

Although isolating entire trefoils would be useful in experimentally identifying the main arbiter of the three-state folding mechanism, changing the order of the trefoils within the sequence would be particularly useful in assessing the potential wide-ranging effects of the conserved geometric frustration in Trefoil 1. A permutation wherein Trefoil 1 is flanked by Trefoils 2 and 3 and resides in the middle of the sequence rather than being present on the N-terminus would be particularly useful to assess differential effects of frustration on the folding route. Because Trefoil 1 is so geometrically frustrated, its native position on the N-terminus allows for it to behave in an independent manner from the stable folding nucleus. This native position is advantageous because it allows for the rest of the protein to fold independently while the geometrically frustrated Trefoil 1 can "wait" until the stable scaffold created by Trefoils 2 and 3 is formed. However, if the geometrically frustrated Trefoil 1 is positioned in the middle of Trefoils 2 and 3, this could have a variety of consequences for the folding route of the protein. If positioned in the center of the protein, Trefoil 1 can no longer fold independently of the rest of the protein scaffold. Further, if Trefoil 1 is in the center of the protein sequence, the formation of the stable scaffold will inherently be disrupted. Likely, this permutation will manifest in a far slower folding rate for the protein overall and may lead to a two-state folding mechanism. In our previous work we showed that IL-33 can be forced into a cooperative folding mechanism by strengthening the contacts in Trefoil 1, although with a much higher folding barrier. Although switching the order of the Trefoils is not directly analogous to strengthening the contacts, it would eliminate many of the same folding considerations that we eliminated through contact strengthening. This includes limiting the ability of Trefoil 1 to backtrack and increasing the need for cooperative folding by the eliminating the ability of Trefoil 2 and 3 to form a continuous stable

scaffold. If this circular permutant can be successfully folded, it would be interesting to assess the subsequent signaling activity of the permutant wherein the dynamics of Trefoil 1 would be dampened and its position re-oriented. This could aid in verifying whether the disparate dynamics in Trefoil 1 are playing a role in successful engagement of the dynamic Ig2 and Ig3 domains in ST2.

Our previous work showed that the contacts formed between β-strands 1 and 4 experience the most extreme degree of backtracking if Trefoil 1 happens to fold first. Further, we showed that Trefoils 2 and 3 act as the stable manifold for the subsequent folding of Trefoil 1. In order to understand the most fundamental contacts that facilitate the folding of IL-33, phi value analysis would be a good strategy. In this thesis, we discovered and verified experimentally that IL-33 folds through an intermediate state. We were able to directly show the formation and resolution of the intermediate state within an extremely short time frame and we verified the structure of the intermediate state through the use of Structure Based Models (SBM). To probe the structure of the intermediate state experimentally, the use of phi value analysis would allow for direct observation of the individual contact formation in Trefoil 2 and 3. Phi value analysis directed at analyzing the formation of the transition state would be useful with respect to understanding the most stabilizing contacts necessary to counter-balance the geometric frustration in Trefoil 1 and the subsequent dynamics within the native basin. Unfortunately, the backtracking in Trefoil 1 occurs as a fluctuation in the denatured basin and therefore would be difficult to capture with the experimental methods currently available.

The final, native fold of IL-33 is strongly influenced by the geometric and conformational frustration in Trefoil 1, leading to disparate and multi-time scale native state dynamics within the functional regions of the protein. These increased regions

of dynamics track to Trefoil 1, the critical region that allows for IL-33 to bind the ST2 receptor-binding interface. The role of these dynamics can be linked to the importance in allowing for efficient engagement of the flexible Ig2 and Ig3 domains in ST2. The dynamic transition of the IL-33/ST2 complex may be further probed through the use of NMR. Thanks to the work of the Fairbrother lab, the full assignments of IL-33 bound to ST2 receptor are available for use. These assignments can be used to evaluate the order parameters of bound IL-33 to ST2 and can be used to probe not only the differences in conformational entropy in IL-33 upon binding, but also to evaluate any pockets of allosteric effects upon the binding and subsequent dynamic transition. IL-33 and ST2 have only been able to be crystalized when bound to each other, suggesting that the complex is overall far more ordered than the individual components. As a framework, this suggests a strong dynamic transition that can be probed at a residue specific level on IL-33 through the use of NMR.

The analysis of the native state dynamics through NMR done in this thesis focused on ^{15}N backbone dynamic measurements. Although we have identified that there is a large degree of conformational entropy and long time scale malleability in Trefoil 1, the exact origins of these dynamics can still be elucidated further. In particular, the dynamics in the native basin may be strongly influenced by side chain interactions and motions within the frustrated and functional regions. The availability of the side chain NMR assignments and the well resolved spectra of IL-33 would make the analysis of side chain dynamics extremely amenable, particularly for tracking the side chain dynamics within the regions of conformational entropy seen in Trefoil 1.

In addition to being a fascinating small signaling molecule with wide-ranging inflammatory consequences, IL-33 is a protein with a fascinating native landscape. The interplay between conserved frustration and functionality in both the folding route

and native basin is a prime example of the evolutionary pressure to balance the need for efficient folding and structural stability with the preservation of functionality. IL-33 is a protein that must engage a highly dynamic receptor, and in doing so must be highly dynamic itself. The high degree of dynamics necessitates the presence of a driving force, and in the case of IL-33, the driving force is provided by frustration along the folding pathway and frustration within the native state. The fine-tuning and adjustment of both the folding pathway and the ensemble of native basins is an astounding example of balancing evolutionary pressure in a protein with the critical preservation of functionality.