UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Protein Dynamics and Allostery in Thrombin, Thrombomodulin, and IkappaBalpha

Permalink

https://escholarship.org/uc/item/8pf370vw

Author

Handley, Lindsey Denise

Publication Date

2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Protein Dynamics and Allostery in Thrombin, Thrombomodulin, and IkappaBalpha

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in

Chemistry

by

Lindsey Denise Handley

Committee in charge:

Professor Elizabeth A. Komives, Chair
Professor Seth Cohen
Professor Partho Ghosh
Professor Steven Gonias
Professor Patricia Jennings

2015

Copyright

Lindsey Denise Handley, 2015

All rights reserved

The dissertation of Lindsey Denise Handley is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

DEDICATION

I dedicate this work to my grandmother, Ruby Louise Foster (1924-2014).

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Å	Angstrom
ABE1	Anion Binding Exosite 1
ABE2	Anion Binding Exosite 2
AMD	Accelerated Molecular Dynamics
APC	Activated Protein C
AR	Ankyrin Repeat domain
ATIII	Antithrombin III
BME	β-mercaptoethanol
BSA	Bovine Serum Albumin
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CPMG	Carr-Purcell-Meiboom-Gill sequence
CPAP	C186P/A220P IκBα
CT	Chymotrypsin
Da	Dalton
DO	Dissolved Oxygen
EDB	Enzyme Dilution Buffer
EDTA	Ethylenediaminetetraacetic Acid
EGF	Epidermal Growth Factor-like Domain
HDX	Hydrogen-Deuterium Exchange
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence

$J(\omega)$	Spectral density function
k_a	Rate of association
k_{ex}	Rate constant for exchange
$K_{M,TM}$	Michaelis-Menten binding constant for thrombomodulin to thrombin
$K_{M,PC}$	Michaelis-Menten binding constant for protein C to thrombin
MALDI-TOF	Matrix-Assisted Laser Desorption Ionization with Time of Flight Detection
MD	Molecular Dynamics
MS	Mass Spectrometry
MWCO	Molecular Weight Cut-Off
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
p_A	Population of ground state
p_B	Population of excited state
PC	Protein C
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PES	Polyethersulfone
PEST	Protein sequence rich in proline (P), glutamic acid (E), serine (S), and threonine (T)
PMSF	Phenylmethanesulfonylfluoride
PPACK	D-Phe-Pro-Arg chloromethylketone

R_1	Longitudinal relaxation
R_2	Transverse relaxation
R_{ex}	Transverse relaxation due to chemical exchange
RDC	Residual Dipolar Coupling
RMSD	Root Mean Square Deviation
RSD	Reduced Spectral Density
S^2	Squared generalized order parameter
SPR	Surface Plasmon Resonance Spectroscopy
τ_C	Molecular rotational diffusion correlation time
τ_{CPMG}	Time between each 180° pulse in the CPMG pulse train
TAFI	Thrombin activatable fibrinolysis inhibitor
TBS	Tris-Buffered Saline
TCEP	Tris-carboxyethylphosphine
TF	Tissue Factor
Th	Thrombin
TM	Thrombomodulin
TM45	The fourth and fifth EGF-like domains of thrombomodulin (residues 346-426)
TM456	The fourth, fifth and sixth EGF-like domains of thrombomodulin (residues 345-462)
TM456m	The fourth, fifth and most of the sixth EGF-like domains of thrombomodulin (residues 346-449)

TM456t	The fourth, fifth and even more of the sixth EGF-like domain of thrombomodulin (residues 346-456)
TM56	The fifth and sixth EGF-like domains of thrombomodulin (residues 387-464)
TROSY	Transverse-Relaxation Optimized Spectroscopy
WCW	Wet Cell Weight
WT	Wild-Type
YLTA	Y254L/T257A IκBα
YNB	Yeast Nitrogen Base

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Reflecting back on my formative years, there have been many individuals who have influenced me toward the pursuit of science and education. Foremost, my parents have always been my greatest cheerleaders. I was always a kid with big aspirations, and my parents have always supported me in any pursuit I was inclined towards, whether that be becoming a phenomenal scholar (I was top 10 of my HS class) or becoming one of the best oboists in the state of Texas (I was awarded and performed in the State HS Band). My parents saw my interest in the sciences early on and guided me to take more advanced Chemistry and Physics classes. My mother especially guided me towards Trinity University, which bestowed upon me a first class education in Chemistry and research experience in protein science.

I've been lucky to have had such fantastic female mentors throughout my scientific career. My PhD advisor, Dr. Betsy Komives, is a bright and prolific scientist with a big heart. She trained me into the scientist I am today, and I'm lucky to have been welcomed into her lab for my PhD training. Before that, I had two fantastic female mentors at Trinity, Dr. Laura Hunsicker-Wang, whose research lab I worked in for four years during my undergraduate education, and Dr. Mary Konkle, who was a post-doc in Dr. Hunsicker-Wang's lab for a few years overlapping my time there. Their passion for protein science inspired me at a very formative time, and they both were fantastic role models for women juggling family life and work life.

I've always been passionate about education, and during my time in the Komives lab, I'm happy to have had the opportunity to train many young scientists,

thanks to Dr. Komives' involvement in the Academic Connections summer program at UCSD and through her personal connections to UCSD undergraduate students. All of these high school students and undergraduates ended up being helpful to me in my PhD work: Cooper Wedge, Varun Venkatesh, Ellie Jones, Avery Bonner, and Julio Hernandez. I was also very lucky to have had the assistance of Kyle Stearns, an undergraduate who had been previously trained by my labmate, Brian Fuglestad.

Finally, I want to thank my fellow labmates in the Komives lab for helping to make my years in graduate school incredibly enjoyable. A few of my old labmates helped to train me in the tools I would need to succeed in my PhD work: Brian Fuglestad trained me in NMR spectroscopy and the analysis of NMR data and helped me through my first thrombin preps; Nicholas Treuheit trained me in *Pichia pastoris* fermentations, protein C activation assays, and helped me through my first thrombomodulin preps. I'd like to thank all of my labmates though, past and present, for being such lovely folks to work beside: Holly Dembinski, Jimmy Marion, Deepa Balasubramaniam, Kristen Ramsey, Jamie Schiffer, Jonathan Parnell, Vera Alverdi, Ryan Lumpkin, Jorge Lamboy, Cesar Ramirez, Majid Ghassemian, Jesse Meyer, and Ingrid Devries.

And most importantly, I need to acknowledge my partner-in-crime, Stephen Foster. I don't know how I would have survived all of these years in graduate school without his companionship.

Chapter II, in full, is a reprint that the dissertation author was the co-principal researcher and co-author of. The material appears in the *Journal of Molecular Biology*. (**Handley, L.D.**, Cervantes, C.F., Sue, S.C., Dyson, H.J., and Komives, E.A. Long-range effects and functional consequences of stabilizing mutations in the ankyrin repeat domain of IκBα, *J. Mol. Biol.* (2013) 425(5))

VITA

2010 Bachelors of Science, Chemistry, Trinity University, San Antonio, TX

2012 Masters of Science, Chemistry, University of California, San Diego

2015 Doctor of Philosophy, Chemistry, University of California, San Diego

PUBLICATIONS

Handley, L.D.; Cervantes, C.F.; Sue, S.; Dyson, H.J.; Komives, E.A. (2013). Stabilizing consensus mutations cause long-range effects in the ankyrin repeat domain of IκBα. *Journal of Molecular Biology*, **425**, 902-913.

Handley, L.D.; Treuheit, N.A.; Venkatesh, V.J.; Komives, E.A. (2015). Thrombomodulin binding selects the catalytically active form of thrombin. (*Biochemistry*, submitted.)

Handley, L.D.; Fuglestad, B.; Stearns, K.; Tonelli, M.; Komives, E.A. (2015). Temporally segregated dynamics of functional regions in apo-thrombin. (In preparation.)

Mukhitov, N.; **Handley, L.D.**; Hamme, C.S.; Euers, L.; Piers, A.; Hunsicker-Wang, L.M. (2015). *Thermus thermophilus* Sco Cysteine 49 and Cu_A Cysteine 153 are the reactive biological pair in the formation of a mixed disulfide intermediate during the assembly of cytochrome ba3 oxidase. (In preparation.)

FIELDS OF STUDY

Major Field: Biochemistry

Studies in Biochemistry and Biophysics

Professor Elizabeth A. Komives

HONORS AND AWARDS

2013	Biophysical Society and UCSD Chemistry Grad Travel Grants
2012	Poster Award Winner, Protein Society Conference
2011	Molecular Biophysics Training Grant, trainee
2010	Urey Award, UCSD Dept. of Chemistry & Biochemistry
2009	Honorary Mention, Barry M. Goldwater Scholarship
2009	Junior Achievement Award, Trinity University Dept. of Chemistry
2008	John A. Burke Award for Achievement in Inorganic Chemistry
2008	Best Presentation, Trinity Summer Research Symposium
2008	Howard Hughes Medical Institute Summer Research Program, trainee

ABSTRACT OF THE DISSERTATION

Protein Dynamics and Allostery in Thrombin,

Thrombomodulin, and IkappaBalpha

by

Lindsey Denise Handley

Doctor of Philosophy in Chemistry

University of California, San Diego, 2015

Professor Elizabeth A. Komives, Chair

Allostery is a critical process that allows for the regulation of proteins: binding of a molecule at one distal, functional site in a protein can cause changes to a different functional site. Early on, allostery was described as a change in protein structure upon the binding of another molecule. However, a dynamic model of protein allostery is emerging, which more completely describes allosteric phenomena. This work

investigates protein dynamics under the lens of allosteric regulation for two different protein systems: (1) IkappaBalpha and (2) thrombin and its cofactor, thrombomodulin.

In Chapter II, we present results from NMR studies which probe the roles of ankyrin repeat consensus mutations in IkappaBalpha. These results reveal that consensus mutations cause long-range, allosteric effects throughout the ankyrin repeat domains, including the ordering of the C-terminal PEST sequence on the ps-ns timescale, which is a region critical for proteosomal degradation.

In Chapter III, a new construct of thrombomodulin, named TM456m, is characterized using kinetic protein C activation assays to determine whether the protein is suitable for use in future NMR experiments bound to thrombin. Here, we show that TM456m has tighter binding than the previously studied TM45 and higher stability than full-length TM456. Furthermore, we present the first HSQC of TM-bound thrombin and observe allosteric changes across the thrombin molecule.

Chapter IV presents a hydrogen-deuterium exchange mass spectrometry study that we performed in order to investigate where changes in solvent accessibility occur in the thrombin molecule when an active site substrate (PPACK) or an effector molecule (TM456m) is bound. In both cases, allosteric changes occurred across the thrombin molecule. This study reveals that N-terminus of the heavy chain of thrombin is not buried in the Ile cleft, contrary to all crystal structures of thrombin to date, and that substrate and TM binding allosterically pull the N-terminus into the Ile cleft.

In Chapter V, we assign the amide peaks of apo-thrombin and measure its NMR backbone dynamics. Comparison of PPACK-thrombin chemical shifts and

backbone dynamics with those of apo-thrombin reveals that allosteric pathways are inherently built into apo-thrombin and suggests that effector-binding might work via a conformational selection mechanism from a broad selection of conformational states.

Chapter I

Introduction

A. Protein Allostery: Structural to Dynamic Interpretations

Allostery is the process by which biological macromolecules transmit effects from one distal, functional site to another functional site upon their binding. Allostery allows for a protein's function to be regulated by other proteins or other molecules. Scientists have been studying protein allostery since the early 1960's.[1; 2] Since then, new and exciting protein regulatory mechanisms continue to emerge. Nevertheless, despite the depth of research in the field of protein allostery, its processes remain an enigma: allostery remains difficult to quantify or even predict.

Early models for understanding allostery were influenced by the experimental technologies available to protein scientists at the time. After the first protein crystal structure of sperm whale myoglobin was solved in 1958,[3] static structures of proteins emerged that began to paint the first pictures of allosteric mechanisms. Two models describing protein allostery from this early time were based on the phenomena observed in hemoglobin: that binding of oxygen to one subunit of hemoglobin enhances the ability of that hemoglobin's other subunits to bind oxygen. The "concerted" model (MWC model) proposed the existence of two states, the tensed and relaxed states, and proposed that all subunits occupy the same state at the same time; an oxygen molecule could bind to either state, but because oxygen stabilizes the relaxed state, ligand binding pushes the equilibrium of all subunits towards the higher affinity, relaxed state.[4] The other dominating model at the time was the "sequential" model (KNF model), which argued that a subunit has to change conformation in order to bind an oxygen molecule (an "induced-fit" in response to ligand binding), and this

conformational change makes it more thermodynamically favorable for other subunits to change conformation in the same way.[5]

Although both of these models were successful in explaining the phenomenon of hemoglobin subunit cooperativity by describing the interplay between two conformational states,[5; 6; 7; 8; 9; 10] neither model adequately explained how the structure of hemoglobin facilitates the conformational change between states. In 1970, Perutz first proposed a model wherein allostery could be understood through the structural changes observed in high-resolution crystal structures of the protein in different conformational states.[11] This has been the reigning model of protein allostery ever since.

Nevertheless, for decades it has been recognized that the structure-centric model of protein allostery is unable to describe all allosteric phenomena.
Since the 1970's, new experimental techniques investigating protein structure, thermodynamics, and dynamics have emerged, driving scientists to replace these static, "starting-point, ending-point" models of protein allostery with models that understand proteins as dynamic molecules. Recent discoveries are finding examples of proteins using rigid body movements,[12; 13; 14] folded yet dynamic structures,[15; 16; 17; 18] and intrinsically disordered structures[19; 20; 21] in the facilitation of protein allostery.

Many dynamic allosteric models currently rely on the conceptualization of native (folded) proteins as existing in a rugged conformational landscape made up of several hills and conformational wells.[22; 23; 24] At the bottoms of each of these wells are an ensemble of conformations which share similar structures. The rates of interconversion between different conformational ensembles (i.e. wells) within this

landscape are determined by the energy barriers between conformations (Figure 1.1). Binding of allosteric effector molecules might alter the conformational landscape by "selecting" a certain conformational ensemble, that is, making that conformational ensemble more favorable by lowering its free energy.[24; 25; 26; 27] Although these dynamic models of protein allostery are more difficult to visualize, new research is continually challenging our understanding of protein allostery, helping to lay a new framework that unifies both static & dynamic interpretation of protein allostery.

B. Methods for Probing Protein Dynamics and Allostery

1. Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR has provided some of the most significant advances in the understanding of protein dynamics and allostery, since NMR can provide atom-specific information on protein structure, dynamics, and thermodynamics.[28; 29; 30] In short, all NMR experiments involve a pure protein sample (300-600 μL, 0.1-5.0 mM) that has been isotopically labeled in a bulk magnetic field, B_0. The bulk magnetic moment of each set of NMR-active nuclei (i.e. spin ½ isotopes like ^{15}N, ^{13}C, ^{1}H, etc. which exhibit magnetic dipole moments) will align with B_0 along the z-axis. A weak magnetic field can be applied horizontally to the z-axis, rotating the bulk magnetic moment into the transverse (x,y) plane. The nuclei will then precess around B_0 at slightly different frequencies, based on small perturbations in their local magnetic fields (e.g. the proximity of charged sidechains or aromatic ring currents, bond torsion angles, hydrogen bonding, etc.). What is observed in an NMR experiment is a free induction

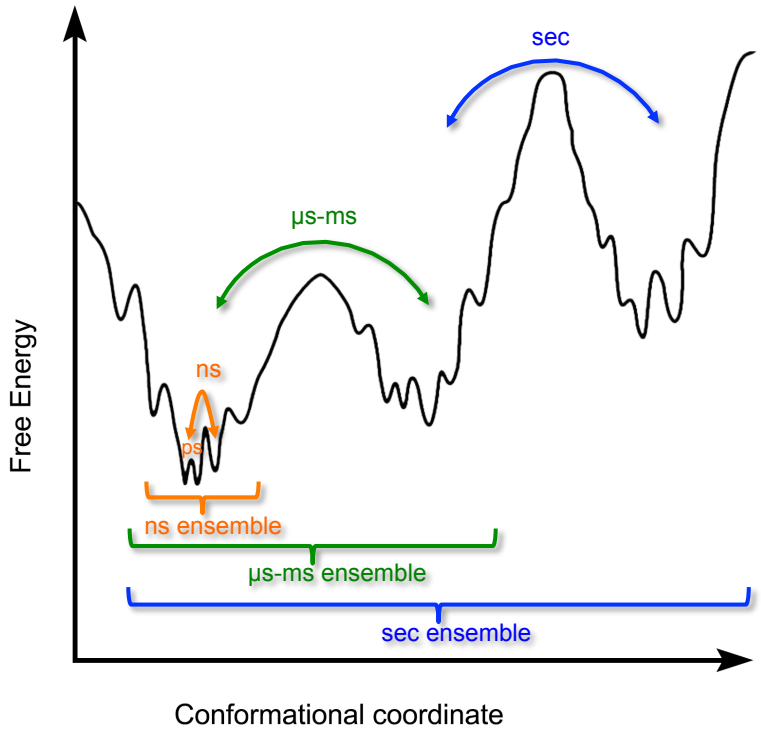


Figure 1.1. A hypothetical protein conformational landscape. The conformational landscape is rugged such that the protein exists in different conformational ensembles at different timescales. Small energy barriers define the ns-timescale conformational ensemble (orange). Larger energy barriers separate conformational ensembles that exist on the μs-ms timescale (green), and even larger energy barriers separate conformational ensembles that exist on the seconds timescale (blue). This figure was adapted from Kleckner et al.[31]

decay (FID), which is the result of all the precessing nuclei in the protein sample and which is read as a time-dependent current by the receiver coil of the NMR instrument. This FID can be Fourier transformed from the time domain to the frequency domain, yielding three observables: chemical shift (δ), intensity (I), and linewidth (λ). These three observables can be separately measured for each NMR-active nucleus in a protein. All NMR experiments manipulate this basic experiment to identify nuclei within the protein. Furthermore, NMR dynamics experiments take advantage of "relaxation" inherent in the system; these relaxation mechanisms cause nuclei to "forget" the direction in which they were oriented. R_1 relaxation, also called longitudinal or spin-lattice relaxation, is the rate of decay of nuclei as they point back in the z-direction, which is caused by an exchange of energy with their surroundings (the "lattice".) R_2 relaxation, also called transverse or spin-spin relaxation, is the decoherence of nuclei in the transverse plane, caused by minor changes in the local magnetic fields of nuclei as they precess around the z-axis. These minor changes in the local magnetic fields of nuclei can be caused by dynamic motions and conformational changes.

A growing suite of NMR experiments are capable of accessing multiple timescales that are relevant to protein motion.[25; 31; 32] NMR dynamics experiments used in our studies are shown in italics in Figure 1.2. While R_1, R_2, and ^{15}N-{^{1}H}NOE experiments measure ps-ns motions relevant to loop motions and sidechain rotations, Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments probe μs-ms timescale motions relevant to allosteric transitions and domain motions. In addition,

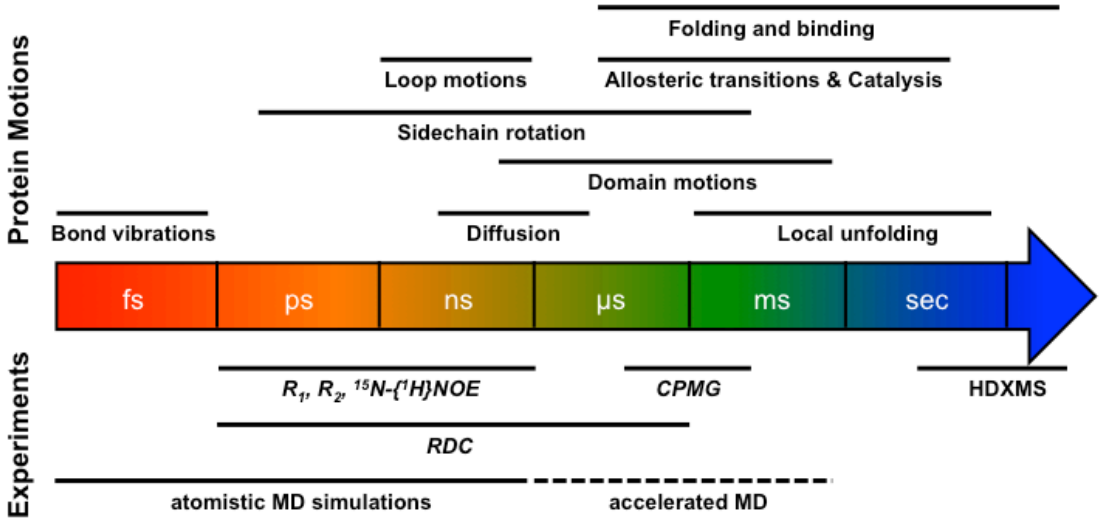


Figure 1.2. A timescale from femtoseconds to seconds is annotated with relevant protein motions (above) and experiments which can investigate those motions (below).[25; 31; 32] The timescale is colored similarly to Figure 1.1 for clarity. NMR experiments are shown in italics.

residual dipolar coupling (RDC) experiments are powerful because they provide data on the time-averaged direction of amide bonds within a protein over a wide timescale (ps-μs). Taken together, this set of NMR experiments spans a massive and significant portion of the timescale of protein motions (ps-ms). Though other NMR dynamics experiments exist, these are the experiments we chose to pursue in the exploration of protein dynamics, as they relate to allostery.

2. Molecular Dynamics (MD)

First used in 1977,[33] MD experiments allow for atomic-level understanding of the motions of molecules. In short, MD simulations calculate the trajectories of a system of atoms (such as those in a protein) by computationally solving Newton's equations of motion, where forces between the atoms and potential energy are defined by interatomic potentials or molecular mechanics force fields. During a simulation, the atoms are pushed out of equilibrium either by a temperature change or a random acceleration to a new position so that alternative conformations other than those from the starting model are explored and their energies evaluated. As a theoretical tool, MD simulations have a strong synergy with NMR experimental approaches, since they are both atomistic and they both cover a similar timescale of protein motions (Figure 1.2).[24; 34] MD simulations can provide incredible detail about the exact nature of protein dynamics and allostery. Ultimately, however, MD simulations are only as accurate as the force fields and potentials used to calculate trajectories, as well as the design constraints placed on the simulation (i.e. number of particles, time-step, length

of time simulated). Therefore, NMR and MD are often used in tandem to confirm the results of the other.

With the current state of computer technology, in order to achieve μs-ms timescales, several days to years of computational time are often necessary, making MD expensive for long timescales and large systems. Therefore, enhanced sampling methods have been designed to cut down computational time. Accelerated molecular dynamics enhances sampling by adding an energy boost to make minima in the protein conformational landscape shallower and easier to jump.[24; 35] AMD is able to inform on protein conformational ensembles on the μs-ms timescale with 2000x less computational time, but aMD is not able to inform on the precise timescale of protein dynamics.[36] However, experimental NMR techniques such as CPMG can bridge this gap in knowledge, since CPMG informs on a similar timescale precisely (Figure 1.2). We have chosen to use this tandem approach in our studies.

3. Hydrogen-Deuterium Exchange Mass Spectrometry (HDXMS)

Since its inception in the early 1990's,[37] HDXMS has become a powerful technique for investigating protein dynamics and protein allostery at longer timescales (sec-min, Figure 1.2). In short, purified protein is exposed to deuterated buffer for different lengths of time (e.g. 0 – 10 min). Backbone amides which are more accessible to solvent will exchange their hydrogens for deuteriums faster than backbone amides which are buried in the core of the protein. This exchange reaction is quenched by significantly lowering the pH of the reaction to ~2.2 and reducing the

temperature down to ~0°C. At this point, the protein can be digested into individual peptides using pepsin, and these peptides can be analyzed for their deuterium incorporation by mass spectrometry.[38] By comparing two different protein states, like a ligand-bound state vs. an unbound state, observed regions of deuterium protection can be used to identify ligand binding sites on the protein, allosteric secondary and tertiary structural changes, or allosteric changes in dynamics. [39; 40] In our study, we chose to use HDXMS to investigate dynamic allosteric changes at the sec-min timescale.

C. References

1. Monod, J. & Jacob, F. (1961). Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol* **26**, 389-401.

2. Changeux, J. P. (1961). The feedback control mechanisms of biosynthetic L-threonine deaminase by L-isoleucine. *Cold Spring Harb Symp Quant Biol* **26**, 313-8.

3. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-6.

4. Monod, J., Wyman, J. & Changeux, J. P. (1965). On the nature of allosteric transitions: a plausible model. *J Mol Biol* **12**, 88-118.

5. Koshland, D. E. J., Némethy, G. & Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365-85.

6. Monod, J., Changeux, J. P. & Jacob, F. (1963). Allosteric proteins and cellular control systems. *J Mol Biol* **6**, 306-29.

7. Cui, Q. & Karplus, M. (2008). Allostery and cooperativity revisited. *Protein Sci* **17**, 1295-307.

8. Koshland, D. E. J. (1959). Enzyme flexibility and enzyme action. *J Cell Comp Physiol* **54**, 245-58.

9. Whitley, M. J. & Lee, A. L. (2009). Frameworks for understanding long-range intra-protein communication. *Curr Protein Pept Sci* **10**, 116-27.

10. Changeux, J. P. (2012). Allostery and the Monod-Wyman-Changeux model after 50 years. *Annu Rev Biophys* **41**, 103-33.

11. Perutz, M. F. (1970). Stereochemistry of cooperative effects in haemoglobin. *Nature* **228**, 726-39.

12. Daily, M. D. & Gray, J. J. (2009). Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. *PLoS Comput Biol* **5**.

13. Swain, J. F., Dinler, G., Sivendran, R., Montgomery, D. L., Stotz, M. & Gierasch, L. M. (2007). Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. *Mol Cell* **26**, 27-39.

14. Zuiderweg, E. R., Bertelsen, E. B., Rousaki, A., Mayer, M. P., Gestwicki, J. E. & Ahmad, A. (2013;328). Allostery in the Hsp70 chaperone proteins. *Top Curr Chem*, 99-153.

15. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J. & Lee, A. L. (2009). Hidden dynamic allostery in a PDZ domain. *Proc Natl Acad Sci U S A* **106**, 18249-54.

16. Tzeng, S. R. & Kalodimos, C. G. (2012). Protein activity regulation by conformational entropy. *Nature* **488**, 236-40.

17. Popovych, N., Sun, S., Ebright, R. H. & Kalodimos, C. G. (2006). Dynamically driven protein allostery. *Nat Struct Mol Biol* **13**, 831-8.

18. Reichheld, S. E., Yu, Z. & Davidson, A. R. (2009). The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. *Proc Natl Acad Sci U S A* **106**, 22263-8.

19. Garcia-Pino, A., Balasubramanian, S., Wyns, L., Gazit, E., De Greve, H., Magnuson, R. D., Charlier, D., van Nuland, N. A. & Loris, R. (2010). Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell* **142**, 101-11.

20. Sevcsik, E., Trexler, A. J., Dunn, J. M. & Rhoades, E. (2011). Allostery in a disordered protein: oxidative modifications to α-synuclein act distally to regulate membrane binding. *J Am Chem Soc.* **133**, 7152-8.

21. Ferreon, A. C., Ferreon, J. C., Wright, P. E. & Deniz, A. A. (2013). Modulation of allostery by protein intrinsic disorder. *Nature* **498**, 390-4.

22. Ramanathan, A., Savol, A., Burger, V., Chennubhotla, C. S. & Agarwal, P. K. (2014). Protein conformational populations and functionally relevant substates. *Acc Chem Res* **47**, 149-56.

23. Ramanathan, A., Savol, A. J., Langmead, C. J., Agarwal, P. K. & Chennubhotla, C. S. (2011). Discovering conformational sub-states relevant to protein function. *PLoS One* **6**.

24. Pastor, N. & Amero, C. (2015). Information flow and protein dynamics: the interplay between nuclear magnetic resonance spectroscopy and molecular dynamics simulations. *Front Plant Sci* **6**.

25. Boehr, D. D., Dyson, H. J. & Wright, P. E. (2006). An NMR perspective on enzyme dynamics. *Chem Rev* **106**, 3055-79.

26. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638-42.

27. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. (2010). Millisecond timescale fluctuations in dihydrofolate reductase are exquisitely sensitive to the bound ligands. *Proc Natl Acad Sci U S A* **107**, 1373-8.

28. Tzeng, S. R. & Kalodimos, C. G. (2011). Protein dynamics and allostery: an NMR view. *Curr Opin Struct Biol* **21**, 62-7.

29. Sekhar, A. & Kay, L. E. (2013). NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc Natl Acad Sci U S A* **110**, 12867-74.

30. Wand, A. J. (2013). The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Curr Opin Struct Biol* **23**, 75-81.

31. Kleckner, I. R. & Foster, M. P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* **1814**, 942-68.

32. Fenwick, R. B., Esteban-Martín, S. & Salvatella, X. (2011). Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur Biophys J* **40**, 1339-55.

33. McCammon, J. A., Gelin, B. R. & Karplus, M. (1977). Dynamics of folded proteins. *Nature* **267**, 585-90.

34. Karplus, M. & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat Struct Biol* **9**, 646-52.

35. Voter, A. F. (1997). Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phy Rev Lett* **78**, 3908–3911.

36. Pierce, L. C., Salomon-Ferrer, R., Augusto F de Oliveira, C., McCammon, J. A. & Walker, R. C. (2012). Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J Chem Theory Comput* **8**, 2997-3002.

37. Zhang, Z. & Smith, D. L. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci* **2**, 522-31.

38. Balasubramaniam, D. & Komives, E. A. (2013). Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. *Biochim Biophys Acta* **1834**, 1202-9.

39. Prasannan, C. B., Artigues, A. & Fenton, A. W. (2011). Monitoring allostery in D_2O: a necessary control in studies using hydrogen/deuterium exchange to characterize allosteric regulation. *Anal Bioanal Chem* **401**, 1083-6.

40. Hentze, N. & Mayer, M. P. (2013). Analyzing protein dynamics using hydrogen exchange mass spectrometry. *J Vis Exp* **81**.

Chapter II

Long-range effects and functional consequences of stabilizing mutations in the ankyrin repeat domain of IκBα.

A. Introduction

Ankyrin repeat (AR) motifs have been observed in proteins with a wide diversity of functions, including cell cycle, transcriptional, and developmental regulation, cytoskeletal organization, and toxins.[1] AR domains, which contain as few as 3 or as many as 34 ARs, are found in the nuclei, cytoplasm, and extracellular milieu of a wide range of organisms, from mammals to bacteria and viruses.[1] Currently, over 25,000 non-redundant AR-containing proteins have been identified in nature.[2] The only common attribute shared by the members of this clearly important class of proteins is that AR domains are evolutionarily tuned to bind specific target proteins by the variation of non-consensus residues displayed on their surfaces.

The structure of an AR consists of a β-hairpin followed by two antiparallel α-helices separated by a loop (β-hairpin-helix-loop-helix). These modules stack upon each other to form a slightly twisted right-handed solenoid structure, with the β-hairpins protruding like the fingers of a cupped hand.[3] The consensus sequence of this ~33 amino acid motif has been well studied using statistical analyses.[4; 5; 6] Although certain positions are only partially conserved, the most prominent features of the consensus sequence are shown in Figure 2.1B.

Interpretation of the structural role that these consensus residues play in the stabilization of AR domains has primarily occurred through structural analysis, particularly of "idealized" AR domains containing only consensus residues,[1; 4; 6] as well as through NMR studies.[7] Most AR domain-containing proteins found in nature preserve a relatively low percentage of consensus residues, however. For example,

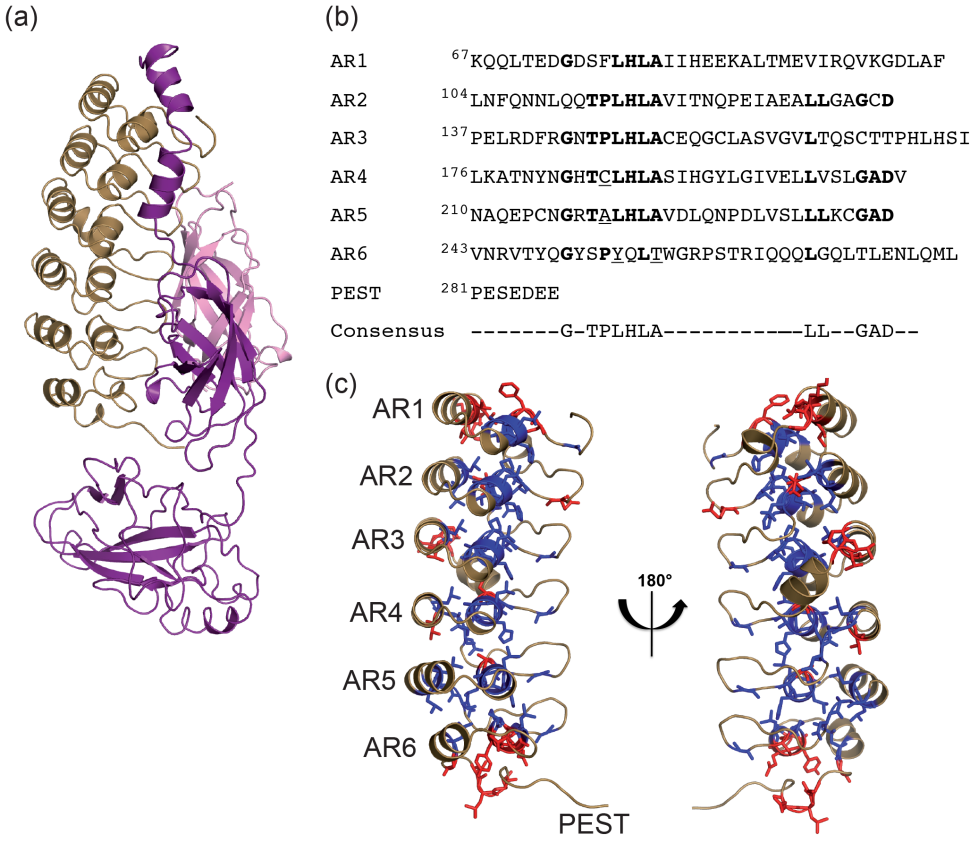


Figure 2.1. The crystal structure and sequence of IκBα compared to consensus sequence for ankyrin repeats. (A) The crystal structure of IκBα bound to p50/p65 NFκB (PDB ID: **1NFI**). The ribbon structure of IκBα is drawn in tan, p50 in pink, and p65 in purple. (B) The sequence of each ankyrin repeat of IκBα is compared to the consensus sequence for ankyrin repeats. The residues of IκBα which match the consensus sequence are in bold. In addition, the residues mutated toward the consensus in this study, C186P/A220P (CPAP) and Y254L/T257A (YLTA), are underlined. (C) A ribbon model of IκBα (PDB ID: **1NFI**) with the consensus residues drawn in sticks and color-coded as follows: consensus residues conserved at a particular position are drawn in blue, while consensus residues not conserved at a particular position are drawn in red. The image has been flipped 180° to show both sides of the protein.

Mosavi et al. found that the natural AR protein that best corresponded to their analytically determined consensus AR motif only matched with 57% sequence identity.[4] One explanation for the sparse distribution of consensus residues in naturally occurring AR proteins is that too many consensus residues results in abnormally high stability,[8] which would prevent the necessary turnover of these proteins *in vivo*. Indeed, consensus-only AR proteins are much more stable than the AR proteins found in nature.[4; 5; 6] Even single and double consensus sequence substitutions in natural proteins have significant stabilizing effects.[9] In fact, nearly every substitution to the consensus imparts measureable stabilization.[10] In order to better understand the role that the consensus sequence plays in the stabilization of AR proteins, we chose to study the AR domain of IκBα, which consists of six ARs possessing an intermediate ratio of conserved to non-conserved consensus residues typical of natural AR domain proteins.

IκBα functions as the inhibitor of NFκB, a transcription factor that regulates apoptosis and responses to stress, infection, and inflammation.[11] Crystal structures of the complex reveal that IκBα binds to NFκB through its AR domain, covering the nuclear localization sequence of NFκB and preventing the transcription factor from entering the nucleus (Figure 2.1A).[3; 12] IκBα also has a signal response region N-terminal to the AR domain and a PEST sequence C-terminal to the AR domain. The minimal region necessary for binding and dissociation of DNA-bound NFκB is IκBα(67–287), which comprises the AR domain and includes the first seven residues of the PEST.[13; 14; 15; 16] Biophysical studies demonstrated that only AR1 – AR4 fold cooperatively and show protection from amide exchange, consistent with a compact

structure, whereas AR5 and AR6 exchange their amide protons within 2 min and do not fold cooperatively with AR1 – AR4.[9; 17; 18]

The inherent flexibility of AR5 and AR6 compared to AR1 – AR4 of IκBα may result from the degree to which these ARs preserve the AR consensus sequence. Mutations of non-consensus residues toward the most prevalent signature of the AR motif, GXTPLHLA, stabilize AR domains, while mutations away from this sequence destabilize them.[9; 10; 19; 20; 21; 22] AR5 and AR6 of IκBα, particularly AR6, deviate from this consensus signature (Figure 2.1B). Mapping of the consensus residues onto the crystal structure of NFκB-bound IκBα shows that they appear to form a cylindrical core down the center of the protein. The conserved TPLHLA sequence forms the beginning of the first α-helix closer to the concave side of the protein, while the conserved LL and GAD lie, respectively, in the second helix and the loop region preceding the β-hairpin (Figure 2.1C, blue). Non-conserved consensus sequence residues (Figure 2.1C, red) are found sporadically throughout the AR domain, but are highly concentrated in AR6.

In IκBα, some of the residues that deviate from the consensus sequence involve amino acids that contact NFκB (F77, Q111, and Q255),[3; 12] many others however, do not contact NFκB and can be substituted without affecting NFκB binding.[9] Mutation of IκBα residues C186 in AR4 and A220 in AR5 to consensus proline residues resulted in a stability increase of ~1.5 kcal/mol.[9] Introduction of the consensus mutations Y254L and T257A in AR6, on the other hand, increased the stability of the protein by a more modest ~0.5 kcal/mol but caused it to have a cooperative folding transition, similar to NFκB-bound IκBα.[23]

Intriguingly, the Y254L/T257A (YLTA) mutations significantly increase the *in vivo* half-life of IκBα, while the C186P/A220P (CPAP) mutations, despite increasing the overall stability, do not.[23] We hypothesized that consensus mutations might variably alter the dynamics of protein, which might explain the different biophysical properties of wild type (WT), CPAP, and YLTA IκBα. Here we present comparisons of the backbone chemical shifts and dynamics of WT, CPAP, and YLTA IκBα by nuclear magnetic resonance (NMR). Our results reveal subtle long-range chemical shift changes throughout the AR domain due to the consensus substitutions. Importantly, the lengthened *in vivo* half-life of the YLTA mutant is explained by a long-range effect of these mutations, which decreases the flexibility of the C-terminal PEST sequence.

B. Materials and Methods

1. Preparation of protein NMR samples

WT human IκBα(67–287) was prepared and purified in the manner described previously for IκBα(67-317) and IκBα(67-206),[17; 24] with a few changes that are detailed in this section. CPAP and YLTA IκBα(67–287) mutants were prepared in a similar manner as the WT, except where indicated otherwise.

For protein samples used in backbone triple-resonance experiments, expression of [^{2}H, ^{13}C, ^{15}N]-labeled IκBα(67–287) was carried out in M9 minimal medium containing 99% D_2O, supplemented with $^{15}NH_4Cl$ (2 g/L), and [^{13}C]-glucose (8 g/L). Cells were first acclimated in 30 mL of M9ZB media for 5 h at 37°C. These cells

were then pelleted and resuspended in 10 mL M9 minimal media and grown for 1 hour at 37°C. A portion (1 mL) of this culture was then taken and diluted with 1 mL of M9 minimal media prepared in D_2O; this 2-mL culture was grown at 37°C for 2 h. M9 minimal media prepared in D_2O (18 mL) were added to this culture and grown for another 8 h. The entire 20-mL culture was then used to inoculate a 1-L culture of M9 minimal media prepared in D_2O, which had been supplemented for labeling as described above. Because the YLTA mutant in particular does not express well, 3-L of culture was grown for each NMR sample rather than 1-L.

For protein samples with specific ^{15}N-amino acid-labeling, expression was carried out in M9 minimal medium supplemented with ^{14}N-labeled amino acids, which were added to the medium in proportion to their abundance in the amino acid sequence to a final total amount of 10 g/L culture. In order to prevent scrambling of the label, we added only $1/10^{th}$ of the above amount of the ^{15}N-labeled amino acid.[25]

Purification proceeded as previously described, except that a 1 h HiLoadQ gradient ranging from 225 mM to 700 mM NaCl was used instead. The final protein samples were exchanged into buffer composed of 25 mM ^{2}H-Tris pH 7.5, 50 mM NaCl, 1 mM ethylenediaminetetraacetic acid, 5 mM 3-[(3-cholamidopropyl)dimethylammonio]propanesulfonic propanesulfonic acid, 2 mM NaN_3, 2 mM DTT, 10% D_2O, and 1% protease inhibitor cocktail. Protein for NMR experiments was concentrated down to 0.1 mM, as described previously.[24]

2. Backbone resonance experiments

Resonance assignments for IκBα free and in complex with NFκB were performed using [^{2}H, ^{13}C, ^{15}N]-labeled IκBα. HNCA,[26; 27] HN(CO)CA,[26; 27]

HN(CA)CB,[28] HN(COCA)CB,[27] and HNCO[26] spectra, or their transverse relaxation optimized spectroscopy (TROSY) equivalents for larger constructs,[29] were acquired.

NMR spectra were acquired at 20°C on a Bruker DRX600 spectrometer equipped with a cryoprobe (HNCA, HNCO, HN(CO)CA) and an Avance 800 spectrometer for TROSY spectra (HNCA, HNCB). The following parameters were used in three-dimensional experiments: for HNCA, data size = 2048 (t3) × 32 (t2) × 88 (t1) complex points, number of scans = 64; for HN(CO)CA, data size = 2048 (t3) × 32 (t2) × 88 (t1) complex points, number of scans = 24; for HNCB, data size = 2048 (t3) × 32 (t2) × 90 (t1) complex points, number of scans = 128; for HNCO, data size = 1024 (t3) × 128 (t2) × 256 (t1) complex points, number of scans = 72. The delay time between each scan was 1.2 s. Data were processed using NMRpipe[50] and analyzed using NMRView.[30]

3. RDC experiments

Spin-state selected TROSY two-dimensional [^{1}H-^{15}N] correlation experiments were collected for isotropic and Pf1 phage-aligned [^{15}N, ^{2}H]-labeled IκBα samples at 25°C on a Varian VS 800 MHz equipped with a cryoprobe.[31] Experiments were collected in an interleaved manner with 72 scans and 2048 (t2) x 128 (t1) complex data points. RDCs were extracted by calculating the differences between the TROSY and ^{1}H anti-TROSY peaks in isotropic and anisotropic data sets. In order to optimize alignment conditions, we titrated WT, CPAP, and YLTA IκBα(67–287) samples with Pf1 phage (Asla Biotech) from 5 mg/mL to 11 mg/mL in 2 mg/mL increments. Pf1 phage concentrations between 9 and 11 mg/mL provided significant alignment for all three constructs with RDCs ranging from -21.91 to 20.79 Hz. Observed RDCs were

compared to RDCs back-calculated from the crystal structure (Protein Data Bank ID: **1NFI**) using the program, PALES (Figure 2.2).[32] The correlations between observed and calculated RDCs were, as expected, weaker than those found previously for IκBα(67–206).[24] Many of the largest discrepancies were from residues in AR5 – AR6, but some were from charged residues displayed on the surface of IκBα, possibly due to interaction with the Pf1 molecules.

4. Backbone relaxation dynamics experiments

Measurements of the T_1, T_2, and $^{15}N–\{^1H\}$NOEs for uniformly [^{15}N, ^2H]-labeled WT, CPAP, and YLTA IκBα(67-287) were made at 20 °C. T_1 and T_2 values were collected at both 500 MHz and 600 MHz on a Bruker Avance 501 and Bruker DRX600 using standard Bruker programs. $^{15}N–\{^1H\}$NOEs were collected at 600 MHz also using a standard Bruker program. For $^{15}N–\{^1H\}$NOEs, independent saturated and unsaturated spectra were recorded in an interleaved manner for each sample. T_1 delays were 12 (duplicated), 45, 89, 177, 353, 705 (duplicated), 1057, 1409, 1761, and 2201 (duplicated) ms. T_2 delays were 9 (duplicated), 13, 21, 29, 37 (duplicated), 45, 53, 69, 85, and 101 (duplicated) ms. The delay time between each scan was 3 s. Data were processed using NMRPipe and analyzed using Sparky and Sparky2rate (http://xbeams.chem.yale.edu/~loria/).[33; 34]

5. RSD analysis

Backbone amide relaxation data were analyzed using RSD mapping, explained in further detail elsewhere.[35; 36; 37] In brief, the spectral density function, $J(\omega)$, describes the contributions of different frequencies to the motion of an amide

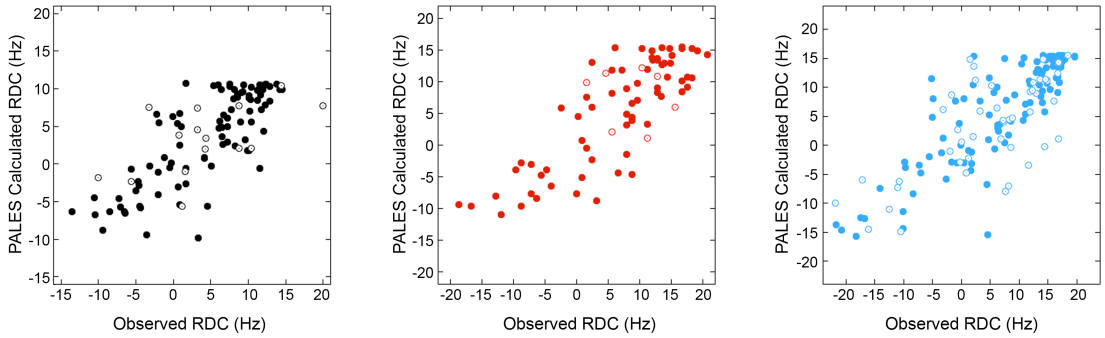


Figure 2.2. Back-calculated best-fit RDCs from a crystal structure of IκBα bound to NFκB (Protein Data Bank ID: **1NFI**). The observed RDCs were compared to the back-calculated RDCs for WT (black), CPAP (red), and YLTA (blue) IκBα(67-287). Residues from AR1-AR4 are shown as filled circles, while residues from AR5-AR6 are shown as unfilled circles.

bond vector. In a spectral density analysis, the magnitude of $J(\omega)$ at five frequencies is calculated. However, in the RSD approach, it is assumed that $dJ(\omega)/d\omega^2$ is relatively constant over the linear combinations of $(\omega_H + \omega_N)$ and $(\omega_H - \omega_N)$.[35; 36; 37; 38] Therefore, the relaxation measures can be expressed as linear combinations of only three spectral densities, $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$, which can then be rewritten in terms of the relaxation measures:

$$J(0) = (6R_2 - 3R_1 - 2.72\sigma)/(3d^2 + 4c^2)$$

$$J(\omega_N) = (4R_1 - 5\sigma)/(3d^3 + 4c^2)$$

$$J(0.87\omega_H) = 4\sigma/5d^2$$

where $\sigma = [(NOE - 1)R_1\gamma_N/\gamma_H]$ is the cross-relaxation rate of the spin pair, $d = [\mu_0 h\gamma_N\gamma_H/8\pi^2]\langle r_{NH}^{-3}\rangle$, $c = \omega_N/\sqrt{3}$, μ_o is the permeability of free space, h is Planck's constant, γ_H and γ_N are the gyromagnetic ratios of ^1H and ^{15}N nuclei, ω_H and ω_N are the Larmor frequencies of the ^1H and ^{15}N nuclei, r_{NH} is the nitrogen-hydrogen bond length (1.04Å),[39] and $\Delta\sigma$ is the chemical shift anisotropy of ^{15}N spins (-160 ppm).[40]

6. BMRB accession number

The ^1H, ^{15}N, and Cα chemical shifts assignments of WT, C186P/A220P, and Y254L/T257A IκBα(67-287) have been deposited in the BioMagResBank (accession numbers: **18756**, **18759**, **18760**, respectively).

C. Results

1. Backbone resonance assignments of YLTA IκBα(67–287) reflect folding of AR5 – AR6

The NMR spectra of YLTA, CPAP, and WT IκBα(67–287) were assigned using a combination of assignment strategies, particularly the possibility of transferring backbone assignments from the spectra of small fragments and complexes when the fragments are structurally similar to the protein of interest.[41] Assignment of the first four ARs of IκBα(67-287) was greatly facilitated by previous assignments of IκBα(67-206).[24] Prior assignment of NFκB-bound IκBα(67-287)[42] greatly facilitated assignment of YLTA IκBα(67-287), which then facilitated assignment of free WT IκBα(67-287). A schematic of this assignment transfer strategy is shown in Figure 2.3. Conventional assignment strategies as well as the use of specifically labeled WT IκBα(67-287) (in which only the G, V, L, T, or A were ^{15}N-labeled)[43] were used to verify the assignments.

We were able to assign 86% of all the residues in YLTA IκBα(67–287), but could only assign 65% of residues in CPAP and WT. This inequality in the total assigned residues between the IκBα(67–287) constructs was primarily due to a more dispersed ^{1}H,^{15}N heteronuclear single quantum coherence (HSQC) spectrum and the greater number of total cross-peaks observed for the YLTA protein. The HSQC of YLTA IκBα(67–287) showed all 208 expected cross-peaks (Figure 2.4, blue), while the HSQC of CPAP IκBα(67–287) showed 198 out of 206 expected cross-peaks, and the HSQC of WT IκBα(67–287) showed only 172 of the 208 expected cross-peaks (Figure 2.4, black).

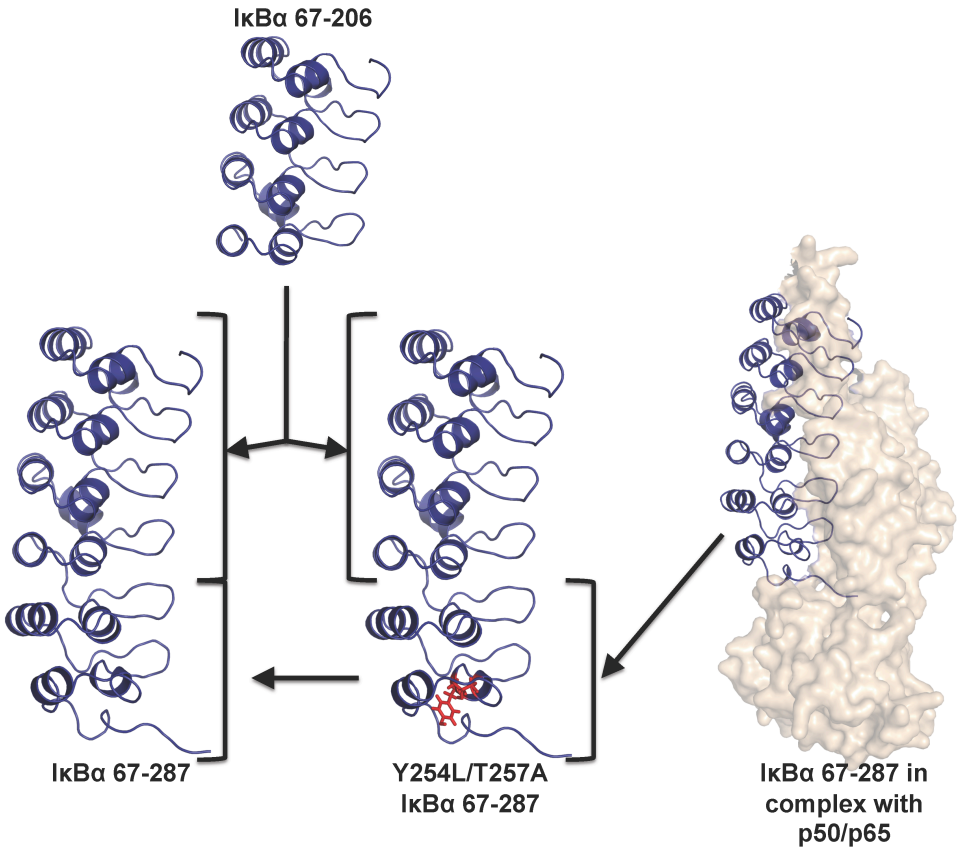


Figure 2.3. Schematic of the assignment transfer strategy used in the assignment of WT IκBα(67-287). The assignment of WT IκBα(67-287) was challenging due to the protein's proclivity toward self-association and aggregation and also the repetitiveness of its secondary structure, which results in a large degree of overlap in its NMR spectra. Rather than tackling the free WT IκBα(67-287) construct using the segmental approach alone, assignments were transferred from the truncated IκBα(67-206), as well as from the NFκB dimerization domain-bound IκBα(67-287), and the consensus mutant, YLTA. Because the HSQC spectra obtained from IκBα(67-206) corresponded well with the HSQC spectra of WT and the mutant IκBα(67-287) constructs, many assignments in AR1-AR4 were able to be transferred directly to the longer IκBα(67-287) constructs. Additionally, because the HSQC spectra of WT IκBα(67-287), YLTA IκBα(67-287), and NFκB-bound IκBα(67-287) corresponded well with one another, it was possible to transfer many resonances assigned previously in bound IκBα(67-287) to YLTA IκBα(67-287) and, furthermore, to WT IκBα(67-287). All transferred assignments were confirmed by the use of specific labeling and segmental assignment approaches.

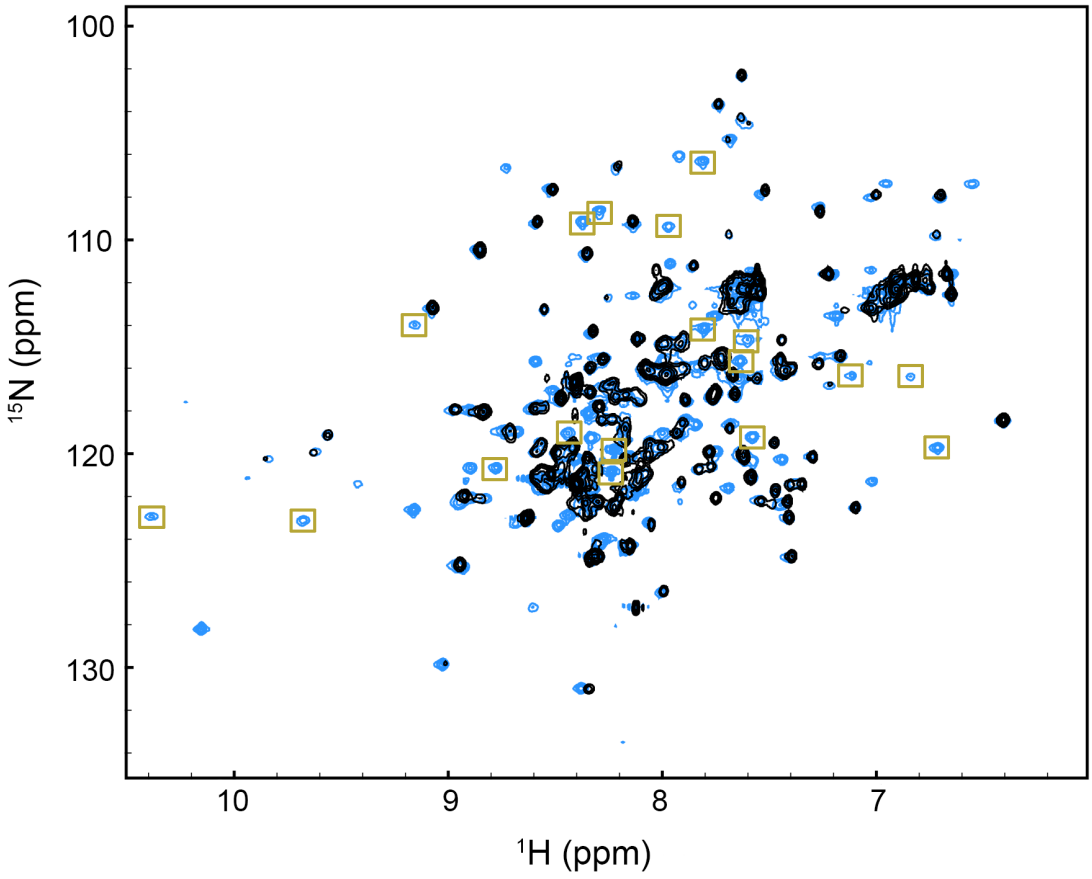


Figure 2.4. HSQC of WT IκBα(67-287) vs. YLTA IκBα(67-287). The HSQC of WT IκBα(67-287) (black) is overlaid on top of the HSQC of YLTA IκBα(67-287) (blue). The more dispersed residues of ankyrin repeats 5-6 of YLTA are boxed in gold.

Table 2.1. The number of assigned residues in each AR of WT, CPAP, and YLTA IκBα(67–287).

	# of Assigned Residues		
Ankyrin Repeat	WT	CPAP	YLTA
1	32	32	31
2	28	29	31
3	30	33	34
4	29	25	34
5	12	10	27
6	1	7	27

In almost every AR, we were able to assign more residues in the YLTA than in the WT or CPAP (Table 2.1). In particular, many of the additional crosspeaks observed in the HSQC of YLTA IκBα (Figure 2.4, gold boxes) were assigned to residues in AR5 and AR6 (Table 2.1). Specifically, residues 212-231, which make up the β-finger and first α-helix of AR5, are completely missing from the WT, but most of these residues are assigned in YLTA. Residues 244-279, which constitute nearly the entire AR motif of AR6, are missing from the WT, but are mostly assigned in YLTA. Only considering the residues in AR5 and AR6 and the PEST sequence, 78% of these residues were assigned in YLTA IκBα(67–287), while only 27% and 30%, respectively, were assigned in WT IκBα(67–287) and CPAP IκBα(67–287).

2. RDCs reveal that WT, YLTA, and CPAP IκBα(67–287) have similar time-averaged solution structures

$^{1}H,^{15}N$ residual dipolar couplings (RDCs) of WT, CPAP, and YLTA IκBα(67–287) were measured to probe for gross structural changes caused by the consensus mutations. RDCs were measured by partially aligning NMR samples with Pf1 phage and comparing the anisotropic dipolar couplings to their isotropic counterparts. Under anisotropic conditions, such as Pf1 phage, dipolar couplings report on the time-averaged direction of the amide bonds, and thus RDCs are able to describe the time-averaged solution structure of a protein.[44]

The RDCs of all three constructs emphasize the repetitive nature of IκBα, with large negative RDCs punctuating the β-finger region and large positive RDCs in the α-helical regions, indicating that the AR motif structure of AR1 – AR4 is preserved in

all three proteins (Figure 2.5). Due to the scarcity of crosspeaks from AR5 and AR6 of the WT and CPAP IκBα, little can be said about the conservation of structure in this region of the proteins. Notably, the RDCs of all three constructs deviate from the RDCs back-calculated from the crystal structure, with some of the largest discrepancies caused by residues from AR5 and AR6 (Figure 2.2).

3. Consensus mutations cause chemical shift differences located throughout the AR domain

Because chemical shifts are a sensitive probe of the environment around a nucleus, comparison of the chemical shifts of bound IκBα to those of YLTA IκBα would indicate the degree to which these proteins are structurally similar. A comparison of the Cα and amide chemical shifts of NFκB-bound IκBα versus free YLTA IκBα revealed chemical shift differences throughout the AR domain (Figure 2.6A). Cα chemical shifts which differed by more than 2 standard deviations included the following: residues 126 and 134 (AR2), 159 (AR3), 230 (AR5), and 249 and 270 (AR6). ^{1}H-^{15}N weighted average chemical shifts that differed by more than 1 standard deviation included the following: residues 83 (AR1), 187 and 193 (AR4), 226 (AR5), and 246, 264, 270-274, 276, 278, and 279 (AR6).

Comparison of Cα and amide chemical shift differences between the free WT protein and the YLTA mutant again showed differences throughout the AR domain (Figure 2.6B). Cα chemical shift differences included the following: residues 72 (AR1), 156 (AR3), 212 (AR4), and 230 (AR5). ^{1}H-^{15}N weighted average chemical

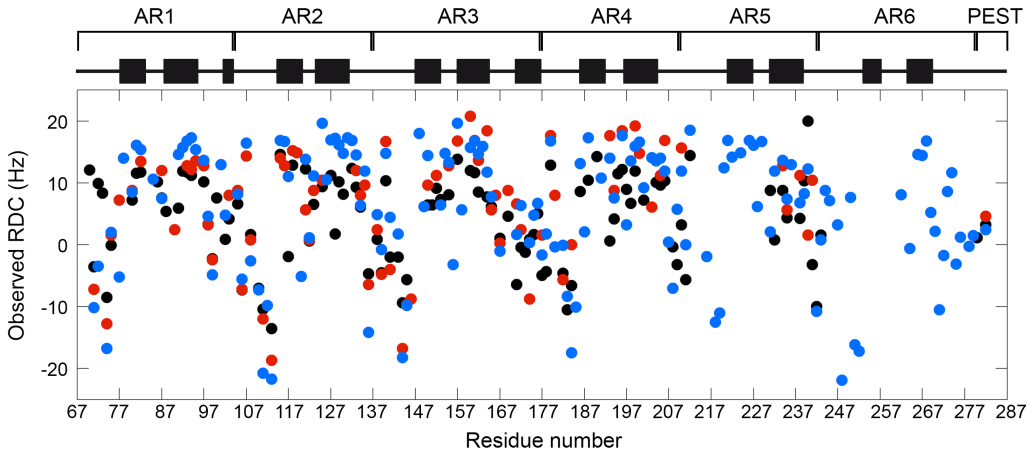
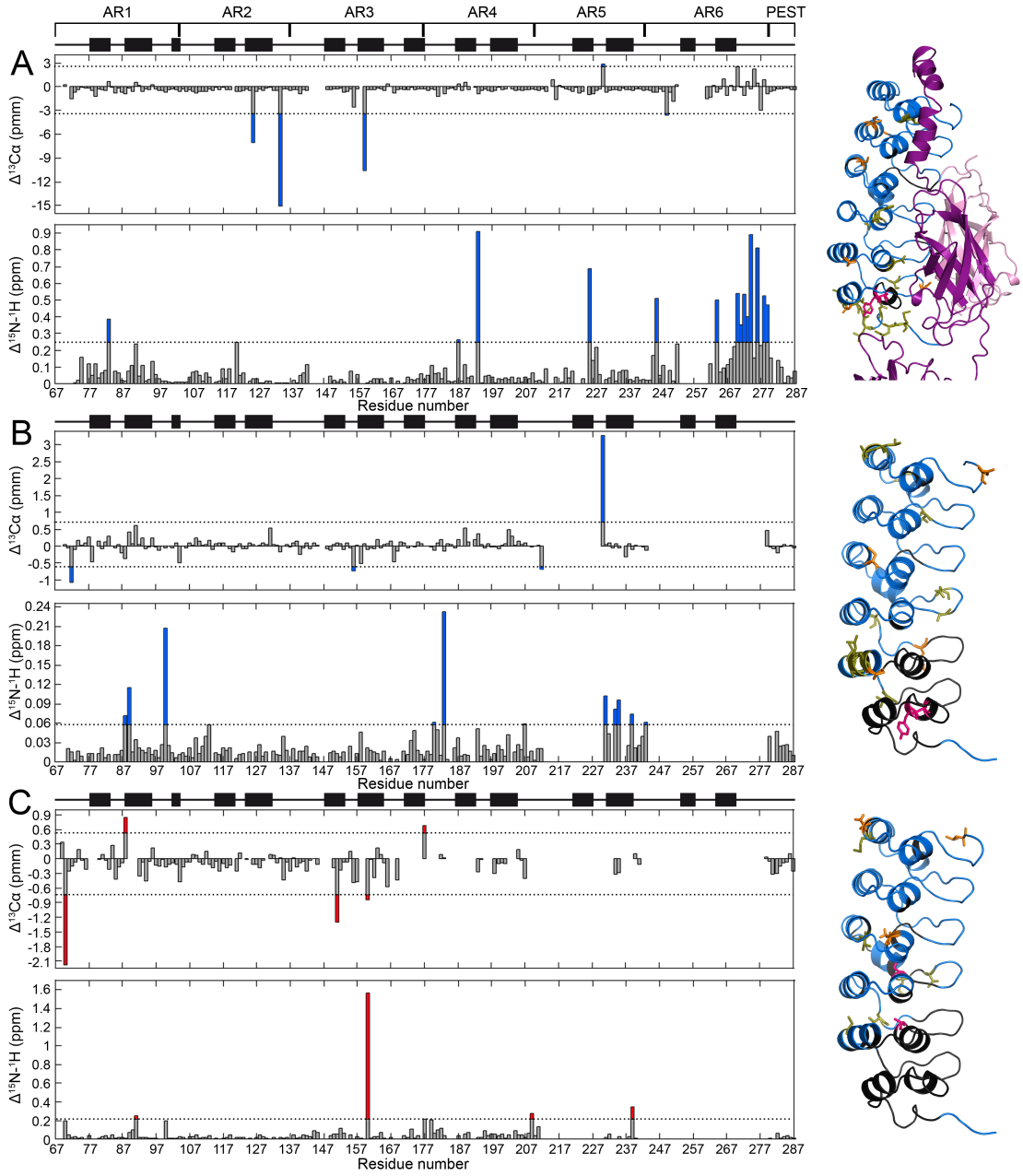


Figure 2.5. RDCs of YLTA, CPAP, and WT IκBα(67-287). Measured RDCs of YLTA (blue), CPAP (red), and WT IκBα(67-287) (black) are plotted against residue number. Diagramed above the RDCs are the positions of α-helical regions (black bars), as determined from the crystal structure (PDB ID: **1NFI**).

Figure 2.6. Cα and weighted average ^{15}N and ^{1}H chemical shift differences between different forms of IκBα. The chemical shift differences were determined for (A) bound WT – YLTA IκBα, (B) free WT – YLTA IκBα, and (C) free WT – CPAP IκBα. The top plot of (A)-(C) gives the Cα chemical shift difference for each residue. The dotted lines in these plots denote an arbitrary cut-off for significance of two standard deviations from the average chemical shift difference. The bottom plot of (A)-(C) gives the weighted average of the ^{15}N and ^{1}H amide chemical shift differences for each residue. This value was calculated using the following equation: $\left([(\Delta\delta_{HN})^2 + (\Delta\delta_{NH})^2/25]/2\right)^{1/2}$, where $\Delta\delta_{HN}$ and $\Delta\delta_{NH}$ are the chemical shift differences in the amide ^{1}H and ^{15}N, respectively. The dotted line in these plots is an arbitrary cut-off for significance of one standard deviation from the average weighted chemical shift difference. Significant changes in chemical shift between WT and YLTA have bars colored in blue, while those between WT and CPAP have bars colored in red. On the right of each set of plots, the residues with significant chemical shift differences are drawn in sticks on a blue ribbon structure of IκBα and are color-coded as follows: yellow, significant amide chemical shift difference; orange, significant Cα chemical shift difference. For residues on the ribbon structure which are colored in black, chemical shifts were not available for analysis in one or both of the proteins.


AR1
AR2
AR3
AR4
AR5
AR6
PEST
A
B
C
Δ13Cα (pmm)
Δ15N-1H (ppm)
Residue number

shift differences included the following: residues 88, 89, and 100 (AR1), 113 (AR2), 180 and 183 (AR4), and 207, 231, 234, 235, 239, and 243 (AR5).

We also compared the chemical shifts of CPAP and unbound WT in order to see if the chemical shift differences in CPAP resembled those observed in YLTA (Figure 2.6C). Similar to YLTA, the CPAP mutations cause changes throughout the AR domain, but at different positions than the YLTA. Cα chemical shift differences included the following: residues 70 and 88 (AR1), 151 and 160 (AR3), and 177 (AR4). ^{1}H-^{15}N weighted average chemical shift differences included the following: residues 91 (AR1), 160 (AR3), 177 and 179 (AR4), and 209 and 239 (AR5).

4. Reduced spectral density analysis shows a less dynamic PEST in YLTA

In order to probe changes in backbone dynamics of the AR structure as a consequence of consensus mutation, we performed NMR relaxation experiments on WT, CPAP, and YLTA IκBα(67-287). R_1 and R_2 relaxation rates were collected at two fields (500 and 600 MHz), and $^{15}N–\{^{1}H\}$ nuclear Overhauser effects (NOEs) were collected at one field (600 MHz). The data collected for YLTA are shown in Figure 2.7A-C (data for WT and CPAP are provided in Figure 2.8).

Backbone relaxation rates report on the picosecond-to-nanosecond timescale dynamics of amide bonds.[45] Generally, $^{15}N–\{^{1}H\}$NOEs alone reflect fast-timescale picosecond-to-nanosecond motions, since this measure is mostly unaffected by slower timescale motions like tumbling and microsecond-to-millisecond chemical exchange.[37] In Figure 2.7D, the $^{15}N–\{^{1}H\}$NOE data clearly show for all three

IκBα(67-287) constructs that the AR domain is relatively ordered compared to the PEST region, which is highly disordered on the picosecond-to-nanosecond timescale. Comparison of the $^{15}N–\{^{1}H\}$NOEs of WT, CPAP, and YLTA IκBα(67-287) reveals that the picosecond-to-nanosecond timescale dynamics of the AR domain do not change significantly upon consensus mutation (Figure 2.7D).

Generally, R_1, R_2, and $^{15}N–\{^{1}H\}$NOE relaxation data are analyzed by the Lipari-Szabo Model-Free formalism that assumes that the overall global motion of the protein is independent from the internal motions. Initial trials using Tensor2 modeling suggested that the Lipari-Szabo analysis was not suitable for the analysis of the IκBα(67-287) data. Reduced spectral density (RSD) analysis of the relaxation data was therefore employed to obtain information about the frequency content that makes up the motion of each amide bond.[35] RSD analysis of WT, CPAP, and YLTA IκBα(67-287) relaxation data confirmed that the lower and higher frequency motions of AR1 – AR6 of IκBα(67-287) do not change much upon consensus mutation (Figure 2.9 and Figure 2.10). Interestingly, the $J(0.87\omega_H)$ spectral density, which probes higher frequency motions, was significantly lower in the PEST of YLTA than in the PEST of WT (Figure 2.9, blue). This indicated that the YLTA consensus mutations cause the PEST to become more rigid on this picosecond-to-nanosecond timescale. The CPAP consensus mutations, on the other hand, did not increase the rigidity of the PEST sequence (Figure 2.9, red).

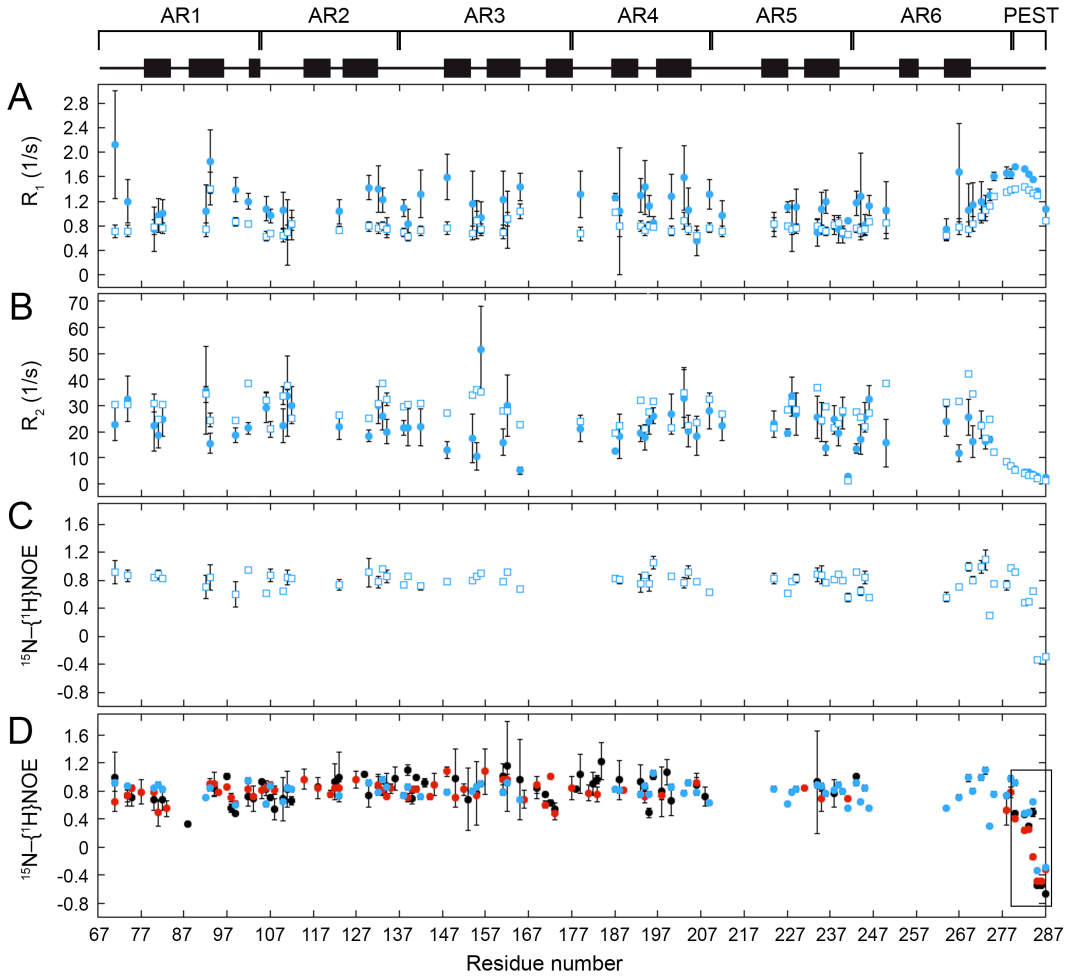
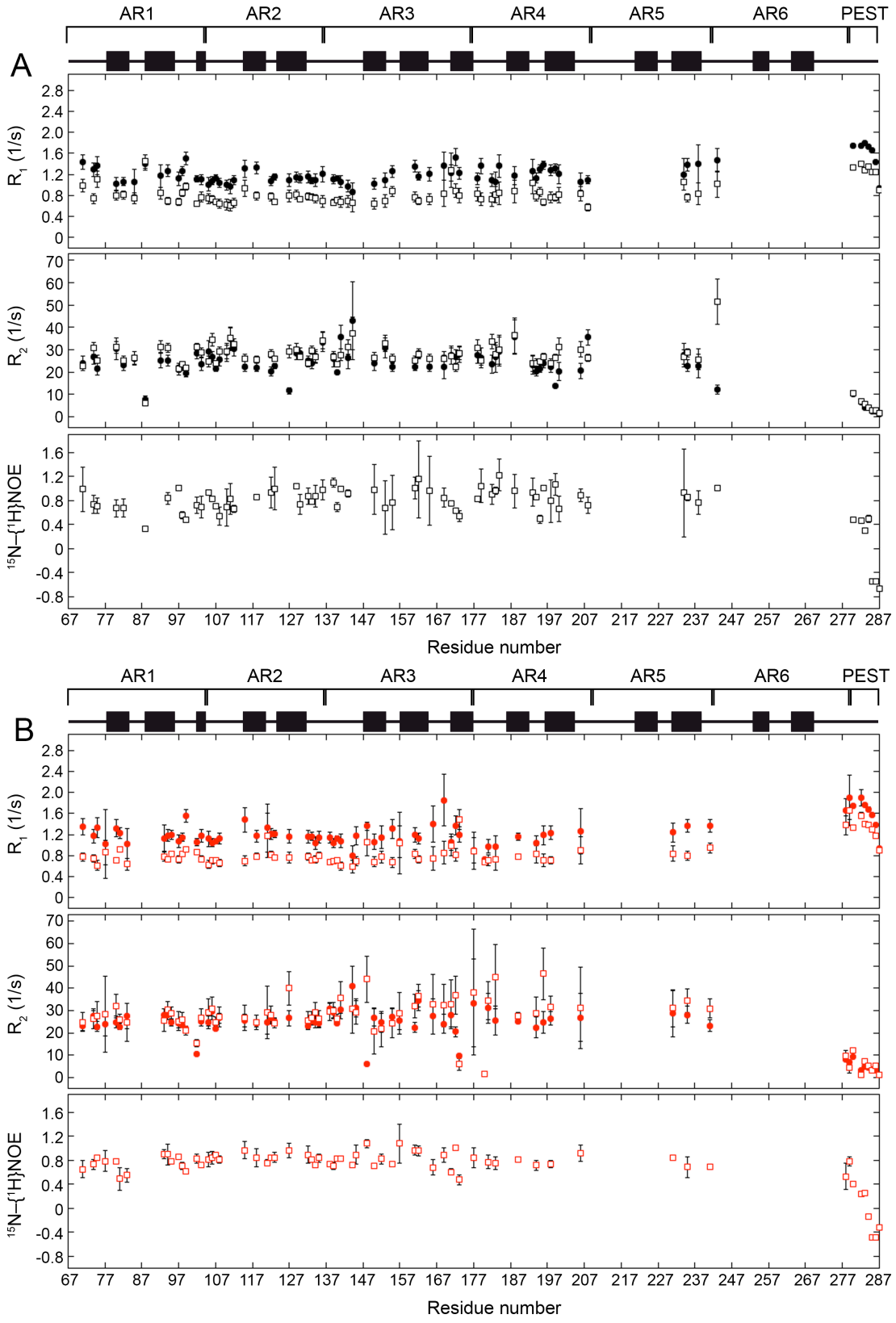


Figure 2.7. Backbone amide relaxation data collected for YLTA IκBα(67-287) at 20°C: (A) R_1, (B) R_2, and (C) $^{15}N–\{^1H\}$NOEs. R_1 and R_2 for all three IκBα constructs were collected at two magnetic fields: 500 MHz (filled circles) and 600 MHz (open squares). $^{15}N–\{^1H\}$NOE's were collected at 600 MHz (open squares). (d) Overlay of $^{15}N–\{^1H\}$NOEs for WT (black), CPAP (red), and YLTA (blue) IκBα(67-287). The PEST region is boxed for emphasis.

Figure 2.8. Relaxation parameters for WT and CPAP IκBα(67-287). Relaxation parameters collected for (A) WT (black) and (B) CPAP (red) IκBα(67-287). R_1 and R_2 measurements were collected at two fields: 500 MHz (filled circles) and 600 MHz (open squares). $^{15}N–\{^1H\}$NOEs were collected only at 600MHz (open squares).


AR1
AR2
AR3
AR4
AR5
AR6
PEST
A
B
R1 (1/s)
R2 (1/s)
15N–{1H}NOE
Residue number

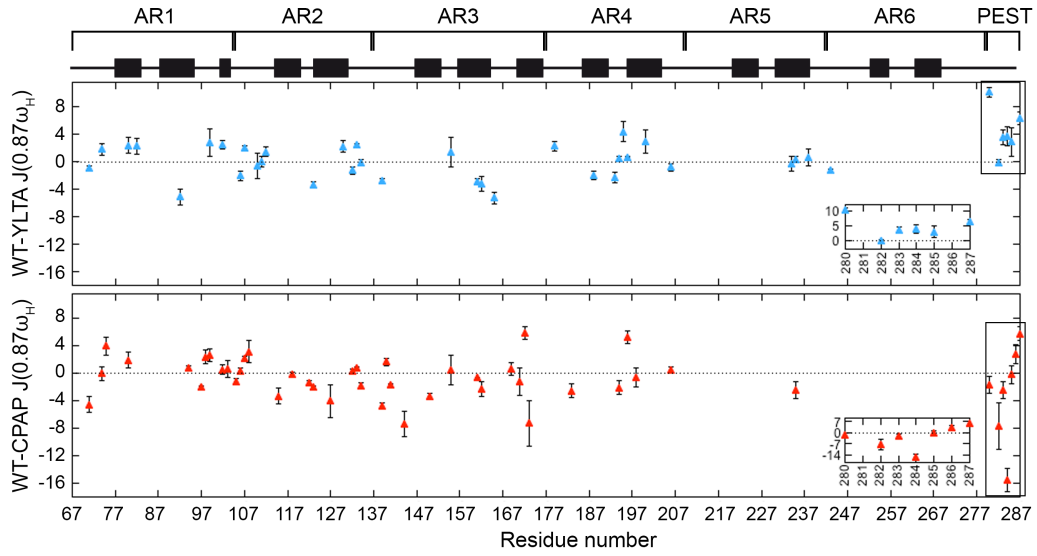


Figure 2.9. Comparison of $J(0.87\omega_H)$ reduced spectral densities for different forms of IκBα. The $J(0.87\omega_H)$ reduced spectral densities of WT were compared to the reduced spectral densities of (a) YLTA (blue) and (b) CPAP (red) IκBα(67-287). Differences in spectral density are plotted by residue only for those residues for which a complete set of relaxation data were available in both the WT and the respective consensus mutant protein. The PEST region is boxed and expanded in the inset of each plot.

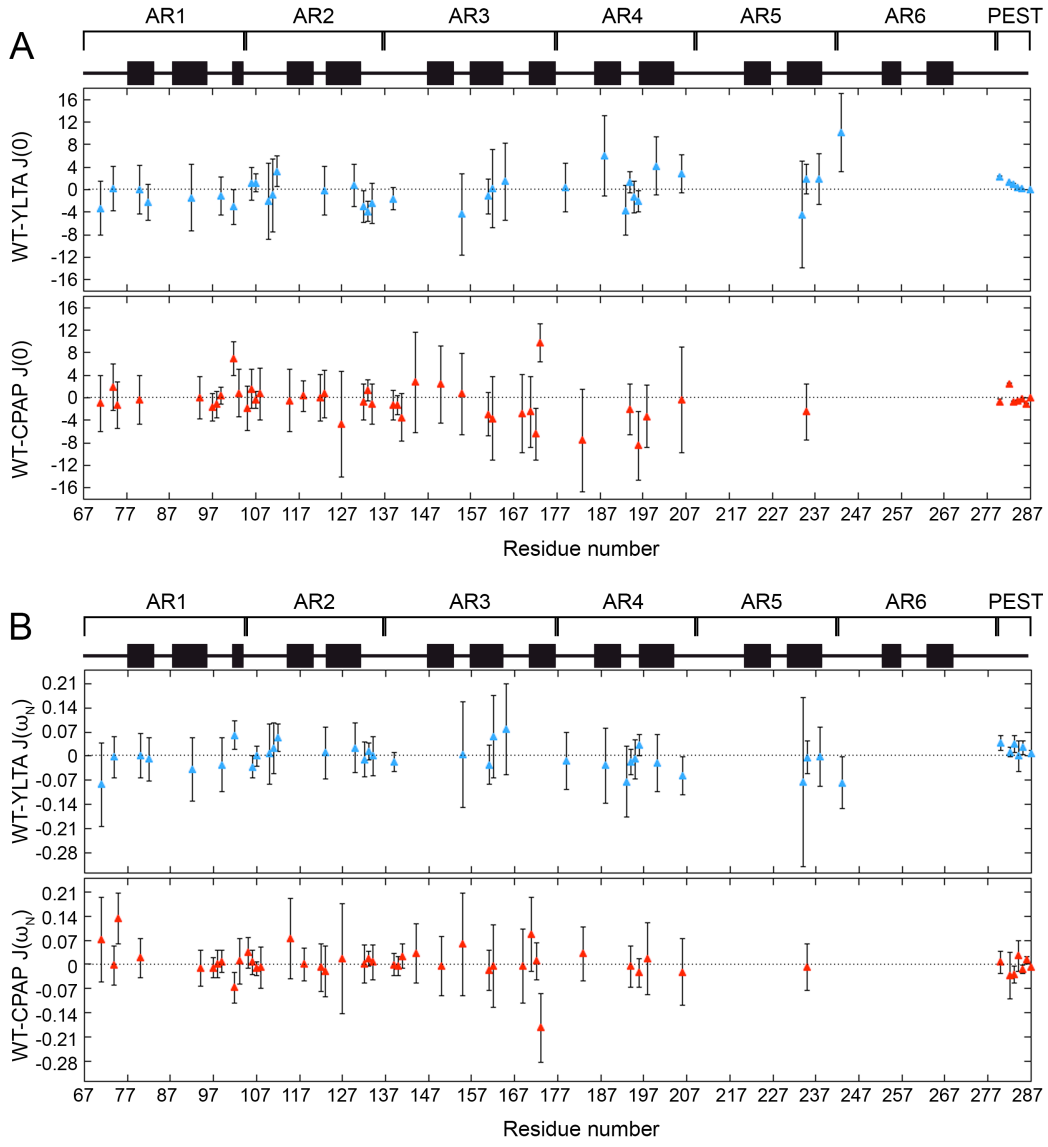


Figure 2.10. Comparison of $J(0)$ and $J(\omega_N)$ reduced spectral densities for different forms of IκBα. The (a) $J(0)$ and (b) $J(\omega_N)$ reduced spectral densities of WT were compared to the reduced spectral densities of YLTA (blue) and CPAP (red). Differences in spectral density are plotted by residue only for those residues for which a complete set of relaxation measures were available in both the WT and the respective consensus mutant protein.

D. Discussion

1. Consensus mutations do not significantly alter structural integrity of the AR domain

Most of the resonances seen in the HSQC spectra of the NFκB-bound WT, unbound WT, CPAP, and YLTA IκBα(67-287) had similar chemical shifts. Because of this, we were able to transfer many assignments from bound IκBα and YLTA to the unbound WT and CPAP IκBα(67-287) constructs (Figure 2.3). The congruity of cross peak assignments between the constructs provides a strong indication that the overall AR domain structure changed very little upon consensus mutation. Little can be said, however, about AR5 and AR6 because few assignments are available for WT and CPAP. The assignments of YLTA compared well to those of bound IκBα, providing some evidence that the free structure of YLTA was not substantially different from that of the NFκB-bound WT. In addition, the RDCs of unbound WT, CPAP, and YLTA IκBα(67-287) support this conclusion, at least for AR1 – AR4, by showing that these three constructs have very similar time-averaged solution structures (Figure 2.5). Therefore, any long-range chemical shift perturbations found in YLTA and CPAP IκBα(67-287) are likely not due to a gross change in the structure of the AR domain.

2. Consensus mutations alter long-range stacking interactions of the AR domain

Significant amide chemical shift differences between NFκB-bound IκBα and YLTA IκBα were observed near the interface where IκBα interacts with NFκB (Figure 2.6A), consistent with differences in local protein environment[46] due to the binding of

NFκB. Amide chemical shift differences are particularly prominent in AR6, hinting that the structures of AR6 in the YLTA versus NFκB-bound forms may be slightly different. In contrast, Cα chemical shift differences, indicators of differences in secondary structure,[46] are located on the face of IκBα opposite the binding interface, particularly in the less-conserved helix 2 of several ARs (Figure 2.6A). These differences in secondary structure between NFκB-bound IκBα and YLTA IκBα are located along the convex side of the proteins and mainly indicate higher helical tendency in these regions of YLTA IκBα compared to NFκB-bound IκBα (Figure 2.6A). Given that significant Cα chemical shift perturbations are seen around helix 2 of AR2, AR3, AR5, and AR6, one might surmise that binding may change the radius of curvature of the NFκB-bound IκBα. While the chemical shift differences of only a few residues in each of these repeats exceeded the arbitrary 2 σ cutoff we applied, lower cutoffs revealed the same trend, with more residues included.

Comparison of free WT and YLTA IκBα chemical shifts also shows differences throughout the AR domain (Figure 2.6B), with only one significantly perturbed residue, V243, near the sites of mutation. Differences in amide chemical shifts are observed in AR5 (once again along the less-conserved helix 2), confirming the folding of AR5 as a consequence of the YLTA mutations, which have previously been shown to integrate AR5 and AR6 with AR1 – AR4 in the cooperatively folding unit.[23] The chemical shift differences observed in AR1 – AR4 suggest that the YLTA consensus mutations have long-range effects on the interactions between AR subunits (Figure 2.6B). Although the YLTA mutations lie in helix 1 of AR6, the chemical shift differences are observed primarily in helix 2 and the β-fingers. At lower cutoffs of

significance in the chemical shift difference data, more residues with chemical shift differences confirm a pathway through the outer helices and β-fingers. Yet, the previously proposed structural roles of the consensus leucine and alanine residues do not predict these long-range effects. Rather, they focus solely on local effects such as the role of leucine in providing hydrophobic contacts within the AR subunit and the role of alanine in lengthening the first α-helix.[5] It is interesting to consider how the internal stabilization of AR6 could propagate stability through the AR domain via inter- and intra-repeat interactions. These long-range effects are consistent with previous biophysical studies showing an increase in the stability of YLTA IκBα due to the consensus mutations.[23] These results are also consistent with previous NMR studies showing that alteration of "capping" repeats can have long-range consequences on internal ARs.[47]

The chemical shift differences between CPAP and WT IκBα appear to be more localized around the sites of mutation than those observed for YLTA (Figure 2.6C). In particular, amide chemical shift differences are observed in residues in the loop region between ARs near each proline. It is likely that this interaction between the conserved proline of the TPLHLA sequence and the less-conserved loop residues across the interface of neighboring ARs improves AR stacking. Similar to the YLTA mutant, chemical shift differences are also seen in AR1, far from the sites of mutation, suggesting that the CPAP consensus mutations also promote long-range interactions between AR subunits. The previously proposed structural role for this consensus proline fails to predict these long-range interactions – the primary prediction being that the proline is necessary in order to break into helix 1.[4; 5] Unlike YLTA, few

chemical shift differences are observed in AR5 of CPAP IκBα. This result is consistent with previous biophysical studies showing that the CPAP consensus mutations of IκBα do not cause AR5 and AR6 to integrate with AR1 – AR4 in the cooperatively folding unit.

3. Consensus mutations do not significantly alter dynamics of the AR domain

Model-Free analysis of protein dynamics necessitates the knowledge of a three-dimensional structure[48]. Although crystal structures of IκBα bound to NFκB are available,[3; 12] amide exchange[18] and single-molecule fluorescence resonance energy transfer[49] analysis of free IκBα provide strong evidence that AR5 and AR6 unravel in unbound IκBα. Given the expected broad ensemble of dynamic structures, combined with its elongated and curved structure, it is not surprising that the Tensor2 Model-Free analysis program could not find an appropriate rotational diffusion tensor for unbound IκBα. In lieu of the Model-Free approach, the reduced spectral densities of WT, CPAP, and YLTA IκBα(67-287) were analyzed (Figure 2.9). This analysis revealed remarkably little difference in the dynamics of the backbone NH groups of the AR domain upon introduction of consensus mutations.

4. Ordering of PEST in YLTA slows in vivo degradation

The misregulation of NFκB signaling is implicated in several cancers, autoimmune diseases, and inflammatory disorders, including B-cell and Hodgkin's lymphoma,[50] and rapid degradation of IκBα is critical for proper signaling through the NFκB pathway.[51; 52] IκBα constructs in which the PEST sequence is deleted are

degraded five times more slowly in cells suggesting that the PEST sequence of IκBα provides an essential signal for the rapid degradation of free IκBα *in vivo*.[51] Previously, Truhlar et al. examined the half-lives of WT and YLTA IκBα in mouse embryonic fibroblast and found that the *in vivo* half-life of IκBα was increased threefold by the YLTA mutation.[23] Here, we have shown that the PEST sequence of YLTA IκBα is more ordered on the picosecond-to-nanosecond timescale than the PEST of WT IκBα. Therefore, our results suggest that the ordering of the PEST sequence by the YLTA mutations in AR6 of IκBα makes the protein almost as resistant to degradation as deletion of the PEST sequence entirely.

It is interesting to speculate whether the slowing of degradation is due to a greater fraction of the PEST sequence being folded, or whether changes in the dynamic timescales of motion of the PEST sequence actually alter access to the proteasome. Free IκBα is degraded *in vitro* by the 20S proteasome,[53] which is known to degrade proteins that are either intrinsically disordered throughout or which contain localized intrinsically disordered regions.[54] The presence of disorder in proteins that must be degraded quickly is a common regulatory mechanism and is well documented in cell functions such as signaling, cell-cycle regulation, transcription and translation.[55] It is hypothesized that the flexibility of the initiating polypeptide chain is necessary to allow physical access to the active site of the 20S proteasome, since the opening of the 20S proteasome is not very wide and because the "gate" is only open one-third of the time.[54] Our studies indicate that the YLTA consensus mutations order the PEST sequence of IκBα to the extent that it is significantly less accessible to the active site of the 20S proteasome than is the PEST of WT IκBα.[23]

Chapter II, in full, is a reprint that the dissertation author was the co-principal researcher and co-author of. The material appears in the *Journal of Molecular Biology*. (**Handley, L.D.**, Cervantes, C.F., Sue, S.C., Dyson, H.J., and Komives, E.A. Long-range effects and functional consequences of stabilizing mutations in the ankyrin repeat domain of IκBα, *J. Mol. Biol.* (2013) 425(5))

E. References

1. Sedgwick, S. G. & Smerdon, S. J. (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem Sci* **24**, 311-6.

2. Letunic, I., Doerks, T. & Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* **40**, 302-5.

3. Jacobs, M. D. & Harrison, S. C. (1998). Structure of an IkappaBalpha/NF-kappaB complex. *Cell* **95**, 749-58.

4. Mosavi, L. K., Minor, D. L. & Peng, Z. Y. (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *Proc Natl Acad Sci USA* **99**, 16029-34.

5. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J Mol Biol* **332**, 489-503.

6. Kohl, A., Binz, H. K., Forrer, P., Stumpp, M. T., Plückthun, A. & Grütter, M. G. (2003). Designed to be stable: Crystal structure of a consensus ankyrin repeat protein. *Proc Natl Acad Sci U S A* **100**, 1700-1705.

7. Yuan, C., Li, J., Mahajan, A., Poi, M. J., Byeon, I. J. & Tsai, M. D. (2004). Solution structure of the human oncogenic protein gankyrin containing seven ankyrin repeats and analysis of its structure--function relationship. *Biochemistry* **43**, 12152-61.

8. Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K. & Plückthun, A. (2008). Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. *J Mol Biol* **376**, 241-57.

9. Ferreiro, D. U., Cervantes, C. F., Truhlar, S. M., Cho, S. S., Wolynes, P. G. & Komives, E. A. (2007). Stabilizing IkappaBalpha by "consensus" design. *J Mol Biol* **365**, 1201-16.

10. DeVries, I., Ferreiro, D. U., Sánchez, I. E. & Komives, E. A. (2011). Folding kinetics of the cooperatively folded subdomain of the IκBα ankyrin repeat domain. *J Mol Biol* **408**, 163-76.

11. Ghosh, S., May, M. J. & Kopp, E. B. (1998). NF-kappa B and Rel proteins: evolutionarily conserved mediators of immune responses. *Annu Rev Immunol* **16**, 225-60.

12. Huxford, T., Huang, D. B., Malek, S. & Ghosh, G. (1998). The crystal structure of the IkappaBalpha/NF-kappaB complex reveals mechanisms of NF-kappaB inactivation. *Cell* **95**, 759-70.

13. Malek, S., Huxford, T. & Ghosh, G. (1998). Ikappa Balpha functions through direct contacts with the nuclear localization signals and the DNA binding sequences of NF-kappaB. *J Biol Chem* **273**, 25427-35.

14. Phelps, C. B., Sengchanthalangsy, L. L., Malek, S. & Ghosh, G. (2000). Mechanism of kappa B DNA binding by Rel/NF-kappa B dimers. *J Biol Chem* **275**, 24392-9.

15. Bergqvist, S., Croy, C. H., Kjaergaard, M., Huxford, T., Ghosh, G. & Komives, E. A. (2006). Thermodynamics reveal that helix four in the NLS of NF-kappaB p65 anchors IkappaBalpha, forming a very stable complex. *J Mol Biol* **360**, 421-34.

16. Bergqvist, S., Ghosh, G. & Komives, E. A. (2008). The IkappaBalpha/NF-kappaB complex has two hot spots, one at either end of the interface. *Protein Sci* **17**, 2051-8.

17. Croy, C. H., Bergqvist, S., Huxford, T., Ghosh, G. & Komives, E. A. (2004). Biophysical characterization of the free IkappaBalpha ankyrin repeat domain in solution. *Protein Sci* **13**, 1767-77.

18. Truhlar, S. M., Torpey, J. W. & Komives, E. A. (2006). Regions of IkappaBalpha that are critical for its inhibition of NF-kappaB.DNA interaction fold upon binding to NF-kappaB. *Proc Natl Acad Sci U S A* **103**, 18951-6.

19. Zweifel, M. E., Leahy, D. J., Hughson, F. M. & Barrick, D. (2003). Structure and stability of the ankyrin domain of the Drosophila Notch receptor. *Protein Sci* **12**, 2622-32.

20. Lowe, A. R. & Itzhaki, L. S. (2007). Rational redesign of the folding pathway of a modular protein. *Proc Natl Acad Sci U S A* **104**, 2679-84.

21. Tang, K. S., Guralnick, B. J., Wang, W. K., Fersht, A. R. & Itzhaki, L. S. (1999). Stability and folding of the tumour suppressor protein p16. *J Mol Biol* **285**, 1869-86.

22. Tripp, K. W. & Barrick, D. (2007). Enhancing the stability and folding rate of a repeat protein through the addition of consensus repeats. *J Mol Biol* **365**, 1187-200.

23. Truhlar, S. M., Mathes, E., Cervantes, C. F., Ghosh, G. & Komives, E. A. (2008). Pre-folding IkappaBalpha alters control of NF-kappaB signaling. *J Mol Biol* **380**, 67-82.

24. Cervantes, C. F., Markwick, P. R., Sue, S. C., McCammon, J. A., Dyson, H. J. & Komives, E. A. (2009). Functional dynamics of the folded ankyrin repeats of I kappa B alpha revealed by nuclear magnetic resonance. *Biochemistry* **48**, 8023-31.

25. Truhlar, S. M., Cervantes, C. F., Torpey, J. W., Kjaergaard, M. & Komives, E. A. (2008). Rapid mass spectrometric analysis of 15N-Leu incorporation fidelity during preparation of specifically labeled NMR samples. *Protein Sci* **17**, 1636-9.

26. Grzesiek, S. & Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson* **96**, 432-440.

27. Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. & Kay, L. E. (1994). A suite of triple-resonance NMR experiments for the backbone assignment of 15N, 13C, 2H labeled proteins with high sensitivity. *J Am Chem Soc* **116**, 11655-66.

28. Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J Magn Reson* **101**, 201-205.

29. Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wuthrich, K. (1998). TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci U S A* **95**, 13585-90.

30. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**, 313-52.

31. Ding, K. & Gronenborn, A. M. (2003). Sensitivity-enhanced 2D IPAP, TROSY-anti-TROSY, and E.COSY experiments: alternatives for measuring dipolar 15N-1HN couplings. *J Magn Reson* **163**, 208-14.

32. Zweckstetter, M. (2008). NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc* **3**, 679-90.

33. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93.

34. Goddard, T. D. & Kneller, D. G. Sparky 3, University of California, San Francisco.

35. Peng, J. W. & Wagner, G. (1995). Frequency spectrum of NH bonds in eglin c from spectral density mapping at multiple field. *Biochemistry* **34**, 16733-52.

36. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1995). Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. *Biochemistry* **34**, 868-78.

37. Ishima, R. & Nagayama, K. (1995). Protein backbone dynamics revealed by quasi spectral density function analysis of amide N-15 nuclei. *Biochemistry* **34**, 3162-71.

38. Krishnan, V. V., Sukumar, M., Gierasch, L. M. & Cosman, M. (2000). Dynamics of cellular retinoic acid binding protein I on multiple time scales with implications for ligand binding. *Biochemistry* **39**, 9119-29.

39. Ottiger, M. & Bax, A. (1998). Determination of Relative N−HN, N−C‘, Cα−C‘, and Cα−Hα Effective Bond Lengths in a Protein by NMR in a Dilute Liquid Crystalline Phase. *J Am Chem Soc* **120**, 12334-41.

40. Prieto, J. H., Sampoli Benitez, B. A., Melacini, G., Johnson, D. A., Wood, M. J. & Komives, E. A. (2005). Dynamics of the fragment of thrombomodulin containing the fourth and fifth epidermal growth factor-like domains correlate with function. *Biochemistry* **44**, 1225-33.

41. Fiaux, J., Bertelsen, E. B., Horwich, A. L. & Wuthrich, K. (2002). NMR analysis of a 900K GroEL GroES complex. *Nature* **418**, 207-11.

42. Sue, S. C., Cervantes, C., Komives, E. A. & Dyson, H. J. (2008). Transfer of flexibility between ankyrin repeats in IkappaB* upon formation of the NF-kappaB complex. *J Mol Biol* **380**, 917-931.

43. Cervantes, C. F. (2010). NMR Studies of the transcriptional Inhibitor I Kappa B alpha and its interaction with the transcription factor NF kappa B., UC San Diego.

44. Prestegard, J. H., al-Hashimi, H. M. & Tolman, J. R. (2000). NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q Rev Biophys* **33**, 371-424.

45. Palmer, A. G., III. (2004). NMR characterization of the dynamics of biomacromolecules. *Chem Rev* **104**, 3623-40.

46. Wishart, D. S. & Case, D. A. (2001). Use of chemical shifts in macromolecular structure determination. *Methods Enzymol* **338**, 3-34.

47. Wetzel, S. K., Ewald, C., Settanni, G., Jurt, S., Plückthun, A. & Zerbe, O. (2010). Residue-resolved stability of full-consensus ankyrin repeat proteins probed by NMR. *J Mol Biol* **402**, 241-58.

48. Palmer, A. G., III. (2001). NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu Rev Biophys Biomol Struct* **30**, 129-55.

49. Lamboy, J. A., Kim, H., Lee, K. S., Ha, T. & Komives, E. A. (2011). Visualization of the nanospring dynamics of the IkappaBalpha ankyrin repeat domain in real time. *Proc Natl Acad Sci U S A* **108**, 10178-83.

50. Courtois, G. & Gilmore, T. D. (2006). Mutations in the NF-kappaB signaling pathway: implications for human disease. *Oncogene* **25**, 6831-43.

51. Mathes, E., O'Dea, E. L., Hoffmann, A. & Ghosh, G. (2008). NF-kappaB dictates the degradation pathway of IkappaBalpha. *EMBO* **27**, 1357-67.

52. O'Dea, E. L., Barken, D., Peralta, R. Q., Tran, K. T., Werner, S. L., Kearns, J. D., Levchenko, A. & Hoffmann, A. (2007). A homeostatic model of IkappaB metabolism to control constitutive NF-kappaB activity. *Mol Syst Biol* **3**, 111.

53. Alvarez-Castelao, B. & Castaño, J. G. (2005). Mechanism of direct degradation of IkappaBalpha by 20S proteasome. *FEBS Lett* **579**, 4797-802.

54. Suskiewicz, M. J., Sussman, J. L., Silman, I. & Shaul, Y. (2011). Context-dependent resistance to proteolysis of intrinsically disordered proteins. *Protein Sci* **20**, 1285-97.

55. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z. & Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* **323**, 573-84.

Chapter III

TM456m: A Thrombomodulin Construct Suitable for NMR.

A. Introduction

When the endothelium of a blood vessel becomes damaged, tissue factor is exposed to the blood causing a cascade of protease activations leading to coagulation (Figure 3.1.) Thrombin is the terminal protease in this coagulation cascade. Thrombin carries out several procoagulant functions, including cleaving fibrinogen to form fibrin, stimulating its own generation by activating upstream coagulation factors (Factor XI, Factor VIII, and Factor V), activating XIII that catalyzes the formation of covalent bonds between fibrin molecules, and promoting the activation of platelets (not shown in Figure 3.1).[1]

Thrombin is thus able to rapidly promote clot formation and prevent blood loss. However, it is critical that coagulation be kept in check; if thrombin is able to incessantly promote coagulation, pathological clotting can result in life-threatening conditions such as myocardial infarction, pulmonary embolism, and stroke. Combined, these conditions involving pathological clot formation account for one-third of all deaths in the United States.[2] The body must therefore regulate thrombin.

This regulation is performed by the inhibitor, antithrombin III, as well as by thrombomodulin (TM), a protein which, when bound to thrombin, changes thrombin's function from procoagulant (clot forming) to anticoagulant (clot inhibiting). With TM bound, thrombin activates protein C.[1] Protein C then shuts down the coagulation cascade by inactivating coagulation factors upstream of thrombin (Factor Va and Factor VIIIa).[3] A soluble version of TM is used clinically in other countries to treat sepsis, but it has so far not proven cost-effective enough for use here in the US.[4; 5]

Human TM is a transmembrane glycoprotein which is expressed on the endothelial surface of all blood vessels.[6] The mature, full-length form of TM in humans is 557 amino acids long and consists of 5 domains, diagramed in Figure 3.2. The N-terminal domain of TM (residues 1-222) shares homology to C-type lectins and appears to exercise anti-inflammatory, cell adhesion, and cell proliferation functions.[6; 7; 8; 9; 10] Electron microscopy studies of TM suggest that this domain, which makes up about ~40% of TM's total mass, is globular and is situated furthest from the membrane.[11; 12] The next domain of TM, discussed in greater detail below consists of six epidermal growth factor-like (EGF) repeats and is critical to activation of protein C and thrombin activatable fibrinolysis inhibitor (TAFI) by thrombin.[13; 14; 15; 16] The third domain is serine/threonine rich with four potential sites for O-linked glycosylated chondroitin sulfates.[17] TM terminates in a well-conserved membrane spanning domain, followed by a short cytoplasmic tail. The cytoplasmic tail contains a single cysteine, which appears to affect multimerization of TM molecules.[18]

Although TM is a large, multi-domain protein, previous studies have homed in on the domains of TM with critical functions in coagulation. Early on, EGFs 3-6 were shown to be the minimum active fragment for thrombin activation of TAFI,[16; 19; 20] whereas only EGFs 4-6 (TM456) are necessary for thrombin activation of protein C.[21] It was discovered that EGFs 5-6 (TM56) are sufficient to bind the anion binding exosite 1 (ABE1) of thrombin, while EGF 4 is required in addition to promote the anticoagulant activity of thrombin.[22; 23] The X-ray crystal structure of the thrombin-TM456 complex shows that EGF 4, despite its critical role in altering thrombin activity, is remarkably distant from the thrombin active site (Figure 3.3).[24]

Furthermore, a comparison of crystal structures shows that the structure of thrombin is essentially unchanged whether thrombin is bound to TM456 or free (< 0.6 Å RMSD).[24; 25] Lacking evidence of a conformational change upon binding, some have suggested that TM provides a docking site positioning protein C near the thrombin active site.[26] However, while docking is undoubtedly part of the story, docking alone cannot explain the 1000-fold increase in the k_{cat} of thrombin toward protein C activation.[27] A large increase in catalytic turnover without an apparent conformational change is suggestive of a dynamic mechanism for the regulation of thrombin activity by TM.

Although several experiments indirectly provide evidence of this dynamic mechanism,[28;29;30] a recent accelerated molecular dynamics (aMD) analysis of the thrombin-TM complex revealed dynamic coupling between the motions of TM and thrombin.[31] In these simulations, loops surrounding the thrombin active site exhibited slow-timescale motions that were correlated with motions in EGFs 4 and 5. These correlated motions were present largely in the TM456:thrombin complex, and not the TM56:thrombin complex, strongly suggesting that these motions play an important role in the anticoagulant activity of TM:thrombin complex.

Unfortunately, because of the nature of the aMD method (see Introduction), the aMD study was not able to provide specific numbers describing the rates or timescale of these correlated motions. In order to determine these rates of motion, a thorough investigation of the complex's dynamics by nuclear magnetic resonance (NMR) spectroscopy would be necessary. However, an NMR sample of a TM construct in complex with thrombin would first need to be prepared.

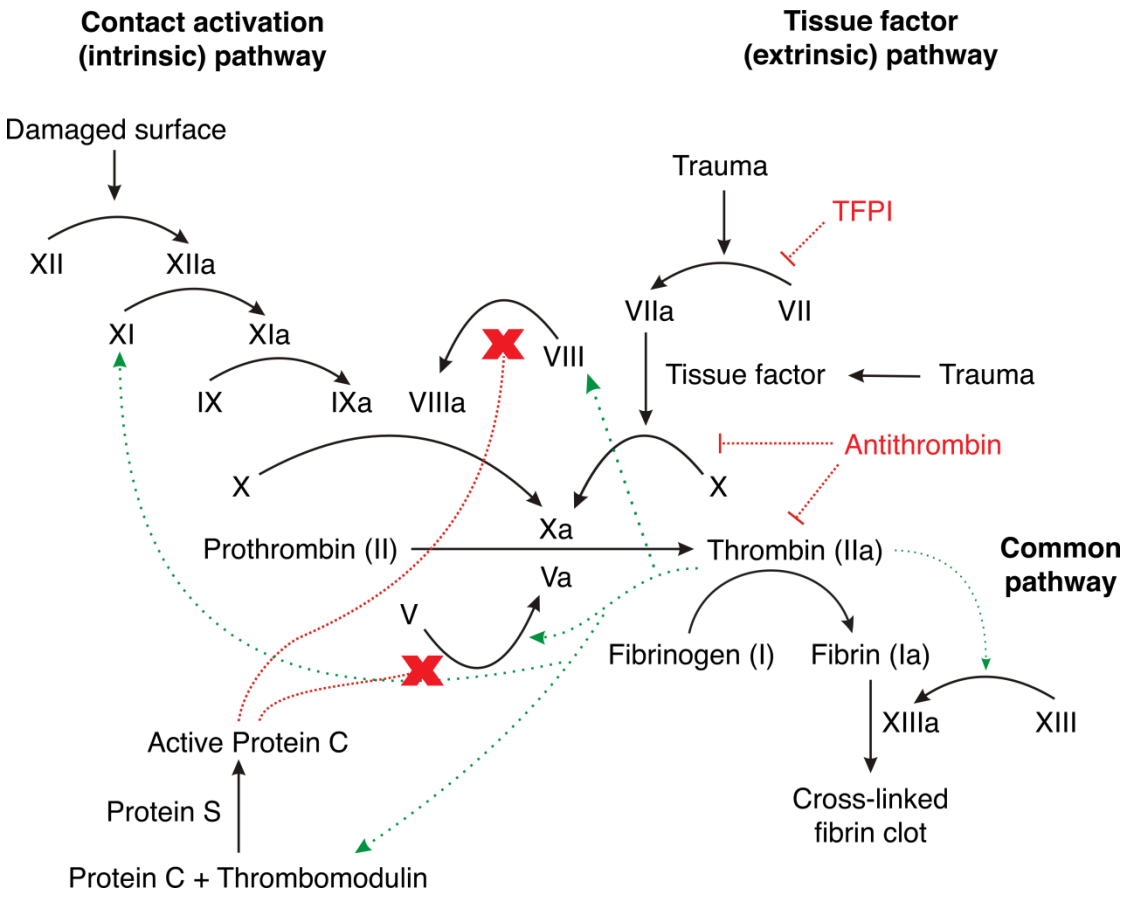


Figure 3.1. A schematic of the coagulation cascade. Thrombin is the terminal protease that cleaves fibrinogen. This figure was adapted from www.wikipedia.org.

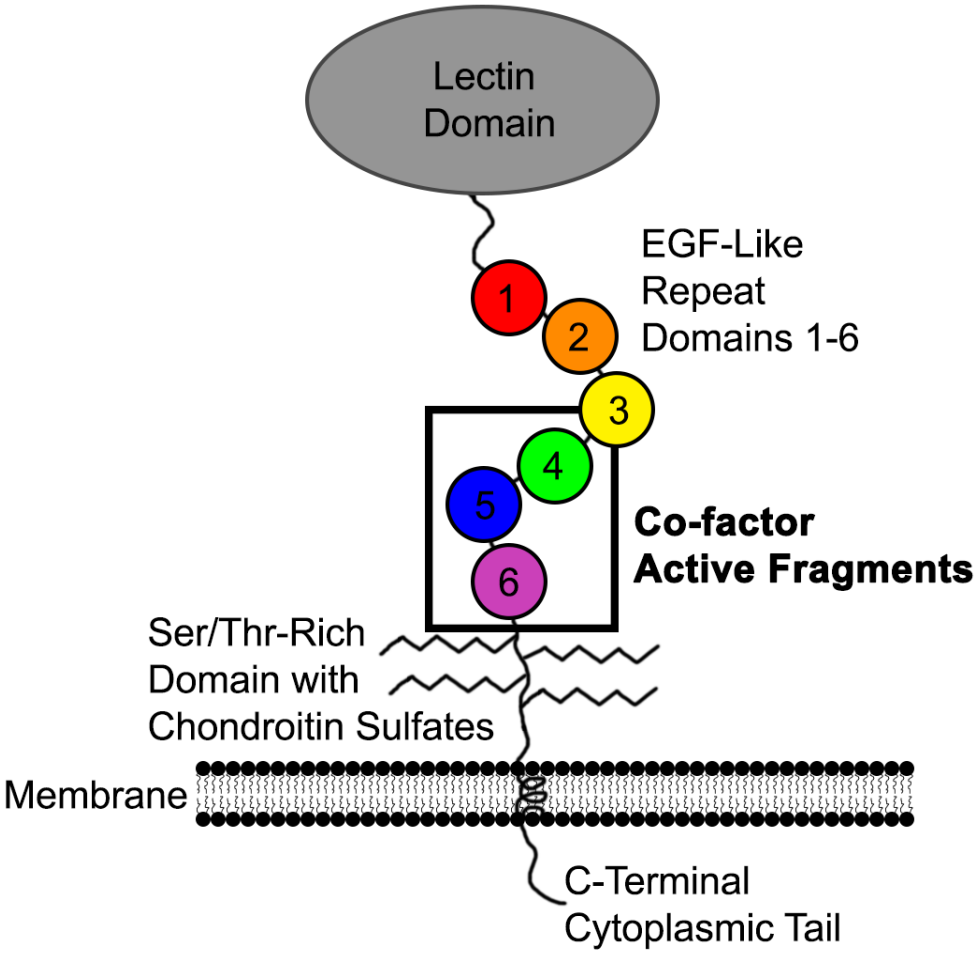


Figure 3.2. Representation of full-length thrombomodulin expressed on an epithelial cell membrane.

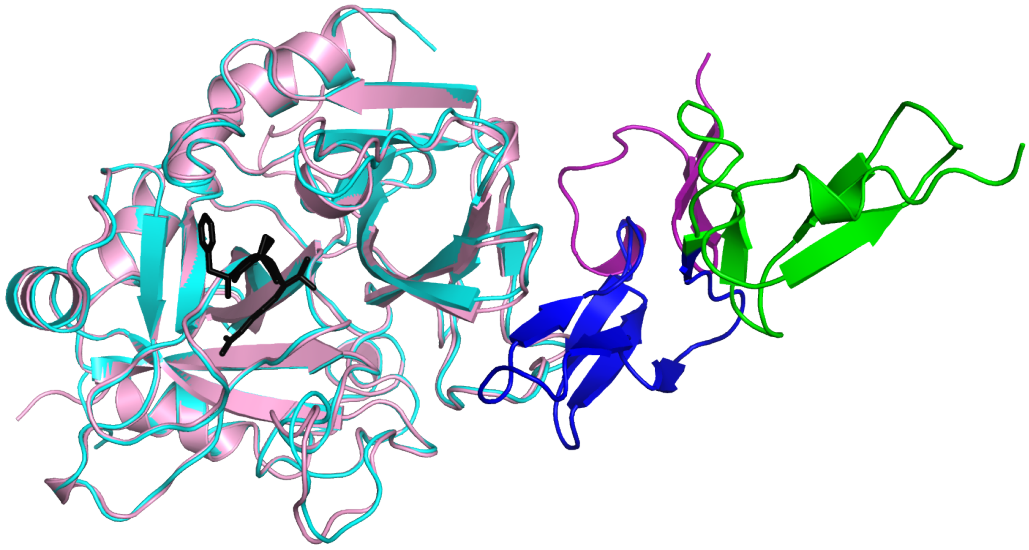


Figure 3.3. Crystal structure of TM456-bound thrombin (1DX5) overlaid on the crystal structure of PPACK-thrombin (1PPB). Thrombin from 1DX5 is colored in cyan, with EGF-like repeats 4, 5, and 6 of TM456 colored in green, blue, and purple, respectively. PPACK-thrombin from 1PPB is colored in pink with the PPACK molecule drawn in black sticks.

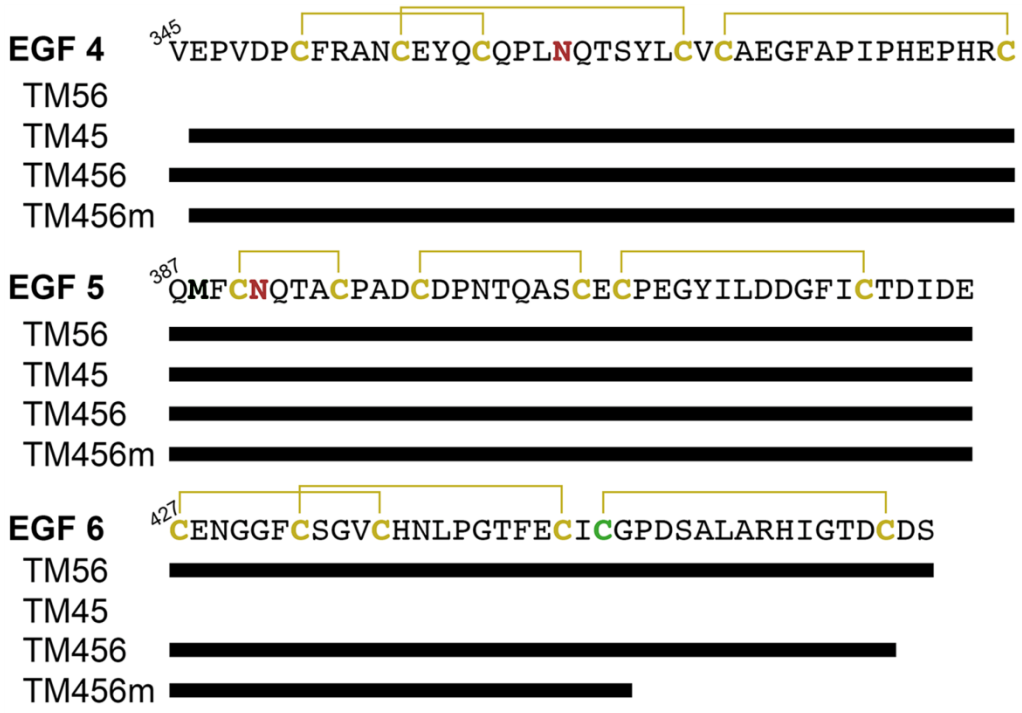


Figure 3.4. Amino acid sequences of relevant TM fragments. Disulfide bridges are shown between cysteines in yellow. Glycosylation sites are shown in red. The C448S mutation is shown in green; this mutation is necessary since Cys_{448} lacks its disulfide partner in TM456m.

Of all the TM constructs, the most obvious candidate for use in NMR studies would be full-length TM456 since it is the minimum active fragment for thrombin-binding and thrombin-altering activity. However, early on it was shown that expression of full-length TM456 in *Pichia pastoris* yields very little protein, and that the construct is prone to aggregation, making it unsuitable for NMR.[22] For this reason, our lab has previously focused on characterizing the interaction of thrombin with TM45. However, TM45 has approximately one-tenth the thrombin-binding capacity of full-length TM456,[22] making studies of the thrombin:TM45 complex by NMR impossible considering the concentrations at which thrombin is still soluble in the NMR tube (> 0.15 mM).[32]

Therefore, our lab has designed a new TM456 truncation, called TM456m, which can be produced in large quantities in *Pichia pastoris* and which encompasses most of EGF6, a critical region for thrombin-binding.[33] Figure 3.4 diagrams the differences between TM456m and other constructs previously studied by our lab. TM456m is identical to the previously studied TM456 construct, except that it lacks Val_{345} and the last 13 amino acids of EGF6 (Pro_{450}-Cys_{462}) and introduces a Cys_{448} to Ser mutation (Figure 3.4).

In order to determine whether TM456m would be suitable for future NMR experiments complexed with thrombin, we performed kinetic protein C activation assays to determine the thrombin-altering activity of TM456m and its ability to bind thrombin. Furthermore, because expression in *Pichia pastoris* yields little more than a single NMR sample worth of thrombin, we explored other methods of producing TM456m, namely, refolding TM456m from the inclusion bodies of *E. coli* and

purifying TM456m from the periplasmic space of *E. coli*. Finally, we prepared the first apo-thrombin:TM456m NMR sample and present the first HSQC of (and assignments for) TM-bound thrombin.

B. Materials and Methods

1. TM456m Expression: Fermentations in *P. pastoris*

All TM456m fermentations were carried out in a BioFlow 3000 fermentor (New Brunswick Scientific, NJ) equipped with a 3 L bioreactor containing 1.8 L of Basal Salts Medium (13.3 mL/L 85% phosphoric acid, 2.3 g/L calcium sulfate dihydrate, 14.3 g/L potassium sulfate, 11.7 g/L magnesium sulfate heptahydrate, 3.9 g/L potassium hydroxide, 40 mL/L glycerol – freshly prepared), including an additional 8 mL PTM Salts (2 g/L cupric sulfate pentahydrate, 0.08 g/L sodium iodide, 3 g/L manganese sulfate, 0.2 g/L sodium molybdate dihydrate, 0.02 g/L boric acid, 0.5 g/L cobalt chloride, 7 g/L zinc chloride, 22 g/L ferrous sulfate heptahydrate, 5 mL/L sulfuric acid – sparged with N_2 gas, filter-sterilized , and stored at 4°C) and 8 mL 1 mM biotin. Media within the bioreactor was adjusted to pH 5.0 and autoclave-sterilized before use. A portion (10 μL) of a cell freeze was used to inoculate 10 mL of BMGY medium (1 x YNB [3.4 g/L yeast nitrogen base, 10 g/L $(NH_4)_2SO_4$], 1% glycerol, 1% casamino acids, 500 μL 1 mM biotin, pH 6.0) in a 25 x 150 mm culture tube, and allowed to shake at 30°C and 300 rpm for 1 day. After 1 day, the 10 mL culture was used to inoculate 200 mL BMGY media in a sterile Erlenmeyer and this was allowed to shake at 30°C and 300 rpm for 1 day. The entire 200 mL culture was

used to inoculate the bioreactor. Throughout the fermentation, the pH of the vessel was maintained at 5.3 by a controlled feed of 30% NH_4OH and the dissolved oxygen (DO) was maintained at 30% by a mixed feed of filtered air and pure O_2 supplied to the fermenter and adjusted automatically. After 20-22 hours of batch growth, the batch glycerol was depleted, indicated by a sudden spike in the measured DO. If a 1 mL sample of fermentation is taken at this point and spun down at 3000 g for 1 min, and the Wet Cell Weight (WCW) is determined, the WCW at this point in the fermentation is approximately 120 mg cell weight/mL culture. To start the glycerol fed-batch process, autoclave-sterilized 50% glycerol medium containing 6 mL PTM Salts and 6 mL 1 mM biotin was fed into the bioreactor at a feed rate of 10 mL/L/hr. During the fed-batch process, the WCW is increased to over 300 mg cell weight/mL culture. For optimal protein production, the induction phase should be started at a WCW of 320 mg cell weight/mL culture. To begin the first induction phase, 50% methanol medium containing 6 mL PTM Salts and 6 mL 1 mM biotin was fed into the bioreactor starting at 4 mL/L/hr; this feed rate was ramped up by 1 mL/L/hr every hour for 7-8 hours until a feed rate of 11-12 mL/L/hr is reached. Then, the second induction phase is begun by switching the 50% methanol medium with 100% methanol medium containing 6 mL PTM Salts and 6 mL 1 mM biotin, and the feed rate was set to 4 mL/L/hr. This feed rate was ramped up by 1 mL/L/hr every hour for 2 hours and left at 6 mL/L/hr for an additional 16 hours. The next day the feed rate was ramped up by 1 mL/L/hr every half hour until the feed rate was at 12 mL/L/hr. The fermentation was left at this feed rate for an additional 2 hours. After this second induction phase, the supernatant from the fermentation was collected by centrifugation at 3000 g (15

minutes, 4°C) in 1 L bottles. Usually, only protein from half of the supernatant was purified at a time. Before purification, ethylenediaminetetraacetic acid (EDTA) was added to the supernatant to a final concentration of 10 mM, and then the supernatant was split into two portions: one half continuing was purified as described below, the other half was frozen at -80°C and purified in the future.

2. TM456m Purification

The supernatant was first diluted to a conductivity near that of buffer A (50 mM MES, 1 mM EDTA pH 6.5) and loaded overnight onto an anion-exchange chromatography column containing 16 g QAE Sephadex resin equilibrated with buffer A at 4^{o}C. Protein was eluted from the column with a 2L 0 – 1 M NaCl gradient in buffer A. Fractions containing TM activity were pooled and purified by a second, higher-resolution anion-exchange chromatography column, a HiLoad 26/10 Q Sepharose column, equilibrated with 50 mM MES, 1 mM EDTA pH 7.5. Impurities were eluted with a 12% B step gradient (50 mM MES, 1 mM EDTA, 1 M NaCl pH 7.5) and TM was eluted from the column with a 30% B step gradient. Next, the protein was purified by reverse-phase C-18 HPLC using an acetonitrile gradient from 10 – 50% in 20mM NH_4OAc pH 5.65. The protein was lyophilized, reconstituted in H_2O, and purified by a final chromatography step: HiLoad 16/60 Superdex 75 size-exclusion chromatography column (Amersham/GE Healthcare) in 50 mM Tris, 150 mM NaCl pH 7.4. Protein C activation assays to determine specific activity of TM (described below) were used at each step in the purification strategy to determine the highest activity fractions to be taken from step to step.

3. Protein C Activation Assay for Specific Activity

The following stock solutions were prepared: 1x TBS (20 mM Tris, 100 mM NaCl pH 7.4), BSA/Ca^{2+} (55 mg BSA, 27.5 mM $CaCl_2$ in 10 mL 1x TBS), and Enzyme Dilution Buffer (EDB) (2 mL BSA/Ca^{2+} in 9 mL 1x TBS.) A solution of active thrombin at a final concentration of 0.004 mg/mL (108 nM) was made immediately before beginning the assay by adding 4 μL 0.2 mg/mL α-thrombin, stored at -80°C, with 196 μL EDB. TM solutions were made with concentrations ranging from 0.001-0.01 mg/mL depending upon expected activity. After determining how many wells would be used for the assay, 20 μL of BSA/Ca^{2+} was added to each well of a 96-well plate, followed by 90 μL of 1x TBS, and then 15 μL of thrombin solution and 10 μL of each TM solution being tested for activity. In general, a negative control, using 100 μL of 1x TBS in place of 10 μL of TM solution, and a positive control, using 10 μL of 0.01 mg/mL pure WT TM45, were included in the 96-well plate for each protein C activation assay. After allowing 10 min of incubation at 25°C to allow TM to bind thrombin, 20 μL of 60 μg/mL human protein C (Haematologic Technologies, Essex Junction, VT, diluted in EDB immediately before use) was added to each well. After incubation for an additional 20 min at 25°C, the aPC production was quenched by the addition of 40 μL heparin-antithrombin III solution (70 μg/mL ATIII (Haematologic Technologies, Essex Junction, VT) with 82 μg/mL heparin and 2.7 μg/mL BSA in 1x TBS). After another 10 min incubation to completely inactivate thrombin, the pH was adjusted in each well by the addition of 20 μL 100 mM Tris pH 8. Finally, the amount of aPC was determined with the addition of 15 μL of aPC-specific chromogenic substrate, S-2366 (12.5 mg/mL in water, stored frozen at -20°C until use), to each well. The cleavage of the chromogenic substrate results in a linear

increase in absorbance at 405 nm over time, which is proportional to the amount of aPC in the well. The data can be converted to nmol aPC produced per minute by comparing the results to a standard curve of aPC and chromogenic substrate versus time previously generated by White et al.[22]

4. Kinetic Protein C Assays

The kinetic mechanism of a protein C activation assay is described by the scheme of an enzymatic process [thrombin (Th) activating protein C (PC)] requiring an activator molecule (TM):

Th + TM ⇌ Th•TM + PC ⇌ Th•TM•PC ⇌ Th•TM•aPC ⇌ Th•TM + aPC,

where thrombin and TM do not need to dissociate before another round of catalysis. The kinetic equations which described this mechanism, originally described by Segal,[34] have been used previously by our lab.[22] These equations are as follows:

$$v = \frac{V_{\max}[TM]}{K_{M,TM}(1+K_{M,PC})/[PC] + [TM](1+K_{M,PC})/[PC]} \qquad \text{Equation 1}$$

$$V_{\max} = V_{\max,app}(TM)(1+\frac{K_{M,PC}}{[PC]}) \qquad \text{Equation 2}$$

and

$$v = \frac{V_{\max}[PC]}{K_{M,PC}(1+K_{M,TM})/[TM] + [PC](1+K_{M,TM})/[TM]} \qquad \text{Equation 3}$$

$$V_{\max} = V_{\max,app}(PC)(1+\frac{K_{M,TM}}{[TM]}) \qquad \text{Equation 4}$$

Modifications were made to the specific activity assay above in order to determine kinetic constants for TM binding to thrombin and for thrombin activating protein C,

using these equations. An enzymatically limiting amount of thrombin (1 nM final stock concentration, compared to 108 nM in the specific activity assays) was used in the following two experiments to make thrombin the limiting reagent in the reactions: (1) The Michaelis-Menten constant for TM binding to thrombin ($K_{M,TM}$) was determined by using a range of [TM] concentrations (1-40 μM stock solution concentration) at a fixed PC concentration (7.75 μM stock solution concentration); the $K_{M,TM}$ was calculated as the TM concentration at $V_{max,app}/2$ in the corresponding Michaelis-Menten curve. (2) The Michaelis-Menten constant for PC binding to thrombin ($K_{M,PC}$) was determined using a range of [PC] concentratinos (0.3-2 μM in stock solution) at a fixed TM concentration (the $K_{M,TM}$ calculated from experiment 1 was used as the TM concentration); the $K_{M,PC}$ was calculated as the TM concentration at $V_{max,app}/2$ in the corresponding Michaelis-Menten curve. To determine the true V_{max}, the apparent V_{max} was obtained from Equations 1 and 3, and then used to calculate the true V_{max} in Equations 2 and 4, using the measured $K_{M,TM}$ and $K_{M,PC}$. This true V_{max} was used to calculate the k_{cat} of the reaction.

5. TM Subcloning for *E. coli* Experiments

Previous to the TM456m construct, our lab explored another construct called TM456t, which contained an additional 7 amino acids of EGF6 compared to TM456m. We found these additional 7 amino acids did not improve the thrombin-binding capacity or activity of TM456t over TM456m;[33] in future experiments we pursued TM456m instead of TM456t, but in many of our early *E. coli* purification experiments, described here, TM456t was used instead. The TM456t gene (residues 346-456) was cloned into a His8/TrpLeader vector from a pUC57 vector using

BamHI/EcoRI restriction sites (synthesized by Bio Basic, Inc.) This new vector places a His8-tag + TrpLeader sequence + thrombin cleavage site immediately before the TM456t gene. The TrpLeader sequence targets TM456t to inclusion bodies.

Because *E. coli* does not have the machinery to glycosylate the two glycosylation sites of TM456t (N364 and N391), we explored mutating these glycosylation sites in order to stabilize these regions of the protein, since previous work had shown TM456 to be less stable without these glycosylations. To prepare TM456 for crystallization in complex with thrombin, Fuentes-Prior et al. purified a minor glycoform of TM456m N-glycosylated only at N364 and treated it with N-glycosidase F, generating an Asp 364 analogue.[24] In hopes of stabilizing an unglycosylated form of TM456t, site-directed mutagenesis was used to introduce the following mutations: N364D, N391D, and N364D/N391D.

6. Screening of Refolding Buffer for TM456t

The cell pellet from a 100 mL growth of TM456t *E. coli* was resuspended in 27 mL 50 mM Tris pH 7.4, 1 mM BME. Phenylmethanesulfonylfluoride (PMSF) was added to a final concentration of 1 mM. Immediately, the cell resuspension was lysed by sonication (5 min sonication time, 10s bursts with 10s breaks, power level 7.) Then, 3 mL 10% triton x-100 and another 1 mM PMSF were added, before letting the mixture incubate on ice for 15 min.

Inclusion bodies were pelleted by centrifuging at 12,000 rpm for 30 min at 4°C. This pellet was then resuspended in 30 mL 50 mM Tris pH 7.4, 1 mM BME, 10 mM imidazole. The inclusion body resuspension was sonicated for another 2.5 minutes (10s sonication bursts with 10s break, power level 7).

Inclusion bodies were pelleted by centrifuging at 12,000 rpm for 30 min at 4°C. This pellet was resuspended in 8 mL resolubilizing buffer (8M urea, 50mM Tris pH 8.2, 150 mM NaCl, 1 mM BME). The pellet was rocked on a nutator at room temperature until it was completely solubilized (~2+ hours.)

HisPur Ni-NTA resin (333 µL portions) were added to 24 different epp tubes. The Ni resin was pelleted in these epp tubes at 700xg for 2 min at room temperature. The supernatant was carefully disposed of. The resin was equilibrated by added 1 mL solubilizing buffer to each of the epp tubes. The resin was mixed with the solubilizing buffer by flicking & inverting the epp tubes. The Ni resin was pelleted at 700xg for 2 min. The supernatant was carefully disposed of and these equilibration steps were repeated one more time.

After the inclusion bodies were solubilized, the inclusion body solution was centrifuged for 5 min. at 12000rpm to remove any insoluble portion. Then, 333 uL of solubilized inclusion bodies were added to each of the 24 epp tubes, followed by 500 uL of solubilizing buffer containing 20 mM imidazole. The epp tubes were slowly rocked at room temperature on a shaker for 30 minutes to ensure complete binding of all His-tagged TM456t to the Ni resin. Then, the resin was centrifuged at 700xg for 2 min at room temperature. The supernatant was disposed of carefully.

Next, a 1 mL portion of the refolding buffer was added to each epp tube. The tubes were mixed briefly by flicking, and then the resin was pelleted as before, and the supernatant was carefully disposed of. An additional 1 mL portion of refolding buffer was added to each epp tube, and the epp tubes were allowed to gently mix on a shaker at 4°C overnight.

After overnight incubation at 4°C with continuous mixing, the resin in the epp tubes was pelleted by centrifuging as previously. The supernatant was replaced with yet another 1.2 mL of fresh refolding buffer and then left to mix for another 2 days at 4°C. The resin was pelleted as before, supernatant removed, and 1 mL of TBS (25 mM Tris pH 7.4, 150 mM NaCl), 25 mM imidazole was added to each epp tube. The epp tubes were gently mixed, centrifuged as before, and supernatant removed. A 1mL portion of TBS 250 mM imidazole was then added to each epp tube to elute the refolded protein. The epp tubes were mixed gently and then the resin was pelleted as before. The supernatant was transferred to a fresh epp tube and screened for activity using a PC activation assay (see above).

This method allowed screening of 24 different refolding conditions for TM456t at a time, and hundreds of conditions over a period of weeks. We tried a multitude of conditions, including: volume of inclusion body solution to use, refolding temperature, pH, buffer, chilled vs. room temperature refolding buffer, and varying concentrations of NaCl, $CaCl_2$, D-Sucrose, Glycerol, CHAPS, Arginine, Triton X-100, n-Dodecyl-β-D-maltoside, Reduced Glutathione, Oxidized Glutathione, Protein Disulfide Isomerase, and Sarkosyl.

7. NMR Sample of TM456m-Bound Apo-thrombin

A 3K MWCO Vivaspin PES concentrator (Sartorius, Inc.) was pre-rinsed by spinning NMR buffer (25 mM phosphate, 150 mM NaCl, 0.05% sodium azide, pH 6.5) at 6000 g for 10 minutes. A 1:1.5 sample of ^{15}N,D-S195M thrombin:unlabeled TM456m was made by adding the appropriate amount of unlabeled TM456m and ^{15}N,D-S195M thrombin (see Ch. 5 Materials & Methods for preparation) in a dilute 10

mL volume. The NMR sample was concentrated down to a 170 μL 112 μM ^{15}N,D-S195M thrombin:168 μM TM456m sample by spinning in 20 minute intervals at 6000g and 4°C. TROSY 2D [^{1}H-^{15}N] correlation spectra were collected on a Varian VS 800 MHz equipped with a cryoprobe (UCSD Chemistry): 128 scans and 2048x128 complex data points at 298 K. All NMR data were processed with NMRPipe and visualized using Sparky. The HSQC of ^{15}N,D-S195M thrombin:unlabeled TM456m was overlaid with a previously collected HSQC of ^{15}N,D-S195M thrombin. Assignments were transferred from ^{15}N,D-S195M thrombin, where possible, in order to calculate chemical shift differences between TM456m-bound and unbound thrombin.

C. Results

1. TM456m is highly stable and has high specific activity and thrombin-binding capacity

Protein C activation assays were performed on a purified 1.4 mM sample of TM456m over a period of several months while the protein was stored at 4°C in 50 mM Tris, 150 mM NaCl, pH 7.4. This long-term study showed that TM456m retains full specific activity and remains soluble over several months. In addition, protein C kinetic assays were performed on both TM456m and TM45 to compare the binding properties and activity of both proteins. Compared to TM45, TM456m includes an additional 23 residues of EGF6, terminating at G449 (Figure 3.4); EGF6 is known to improve thrombin-binding. The specific activity of TM456m was measured to be 3

times higher than TM45, and the $K_{M,TM}$ was measured to be 5 times lower (tighter) than the $K_{M,TM}$ of TM45 (Figure 3.5, Table 3.1). The $K_{M,PC}$ and k_{cat} was measured to be approximately the same for both TM45 and TM456m. (Figure 3.6, Table 3.1)

2. TM456t refolding produces active protein, but not at large-scale and not efficiently

The most challenging feat to refolding TM456t was expected to be correctly pairing the cysteines of all 8 disulfide bonds in the construct. Refolding of TM456t was expected to create incorrectly disulfide-bonded multimers (dimers, trimers, etc.) of TM456t, in addition to monomers of the correctly or incorrectly disulfide-bonded protein. Early evidence supported this notion: we found that refolding less, more dilute TM456t on a Ni column greatly improved refolding. After some preliminary screening of optimal refolding conditions for WT TM456t, we did a small-scale study comparing the refolding of WT TM456t and the glycosylation site mutations: N364D TM456t, N391D TM456t, and N364D/N391D TM456t. Size chromatography was used to see which construct produced the fewest high-order oligomers after refolding. N391D bound poorly to Ni columns compared to the other three constructs, suggesting this construct aggregates significantly; therefore, this construct was no longer pursued as an option for refolding. Of the remaining three constructs, N364D produced significantly fewer higher-order multimers, as assessed by size-exclusion chromatography of the purified, refolded protein. Whereas N364D/N391D TM456t and WT TM456t eluted primarily as a trimer or tetramer from the size-exclusion column, N364D appeared to elute primarily as a dimer or trimer. Therefore, it

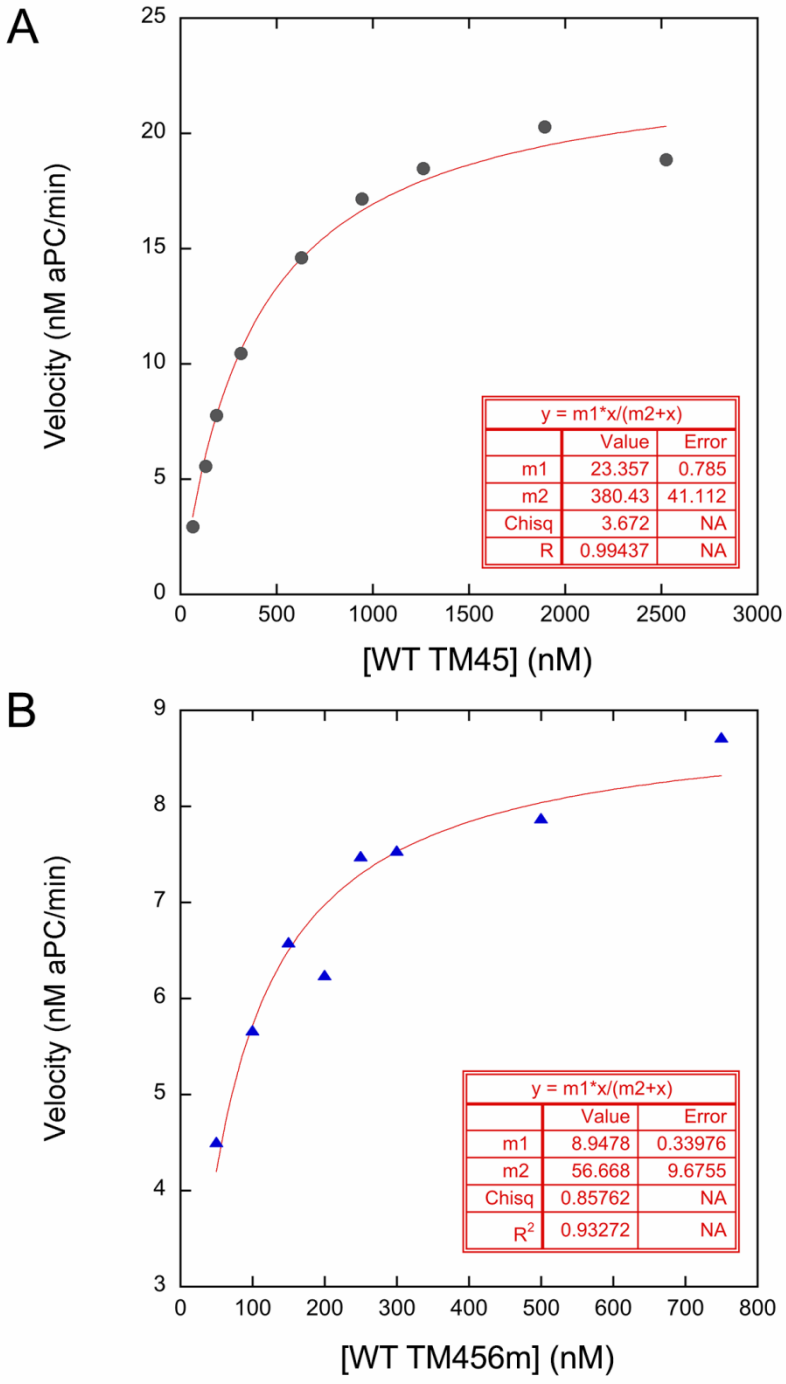


Figure 3.5. Representative plots for kinetic protein C assays to determine $K_{M,TM}$ for WT TM45 (A) and WT TM456m (B). Both plots show reaction velocity on the y-axis plotted again TM concentration.

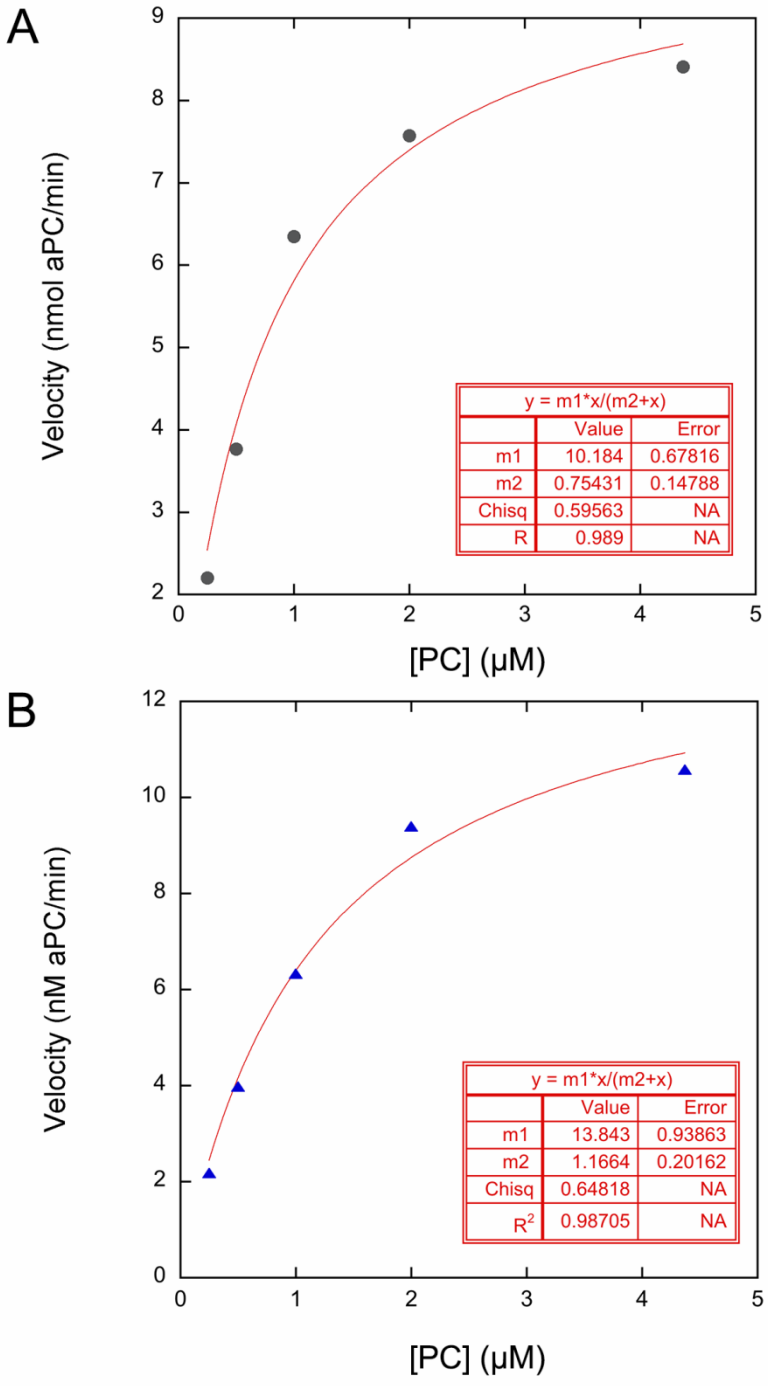


Figure 3.6. Representative plots for kinetic protein C assays to determine $K_{M,PC}$ for WT TM45 (A) and WT TM456m (B). Both plots show reaction velocity on the y-axis plotted again PC concentration.

Table 3.1. The Michaelis-Menten parameters for WT TM45 and WT TM456m, as determined from kinetic protein C activation assays.

TM Construct	$K_{M,TM}$ (nM)	$K_{M,PC}$ (μM)	k_{cat} (s^{-1})	$k_{cat}/K_{M,PC}$ (s^{-1} μM^{-1})	Specific Activity
WT TM45	330 ± 50	0.8 ± 0.1	1.4 ± 0.1	1.7	1.7×10^5
WT TM456m	60 ± 10	1.2 ± 0.2	0.89 ± 0.03	0.7	5.8×10^5

appeared that Asp mutation at the N364 glycosylation site facilitates TM456t refolding.

Next, small-scale refolding screens were used to test a wide variety of refolding buffers. Several dozen screens were performed, testing two different chemicals at a time: usually 6 different concentrations of "Chemical 1" and 4 different concentrations of "Chemical 2", or 24 conditions at a time. A small amount of resolubilized inclusion bodies was added to each refolding buffer in an epp tube containing Ni resin. After refolding for a few days, the refolded, His-tagged protein was eluted from the Ni resin using imidazole, and the specific activity of the refolded protein was tested using a protein C activation assay. The results of a representative refolding screen are shown in Table 3.2.

Most chemicals tested in refolding buffers had little effect on the refolding of TM456t. Only the most significant results of our screening experiments are explained below. We tested a wide range of concentrations of reduced and oxidized glutathione because this chemical has been suggested to improve the refolding of other disulfide-bonded proteins by facilitating the formation of mixed-disulfides (in fact, this chemical is used in the refolding of thrombin from *E. coli*). Of all conditions tested, 1 mM / 0.1 mM reduced/oxidized glutathione was found to be the most productive. Of all the buffers tested, Tris pH 8.2 appeared to be the most suitable. The most successful chemical tested that improved TM456t folding was sarkosyl (6-18% w/v), an ionic surfactant. The results of a sarkosyl screen are shown in Table 3.2.

Even with refolding improvements using sarkosyl, size-exclusion chromatography still showed the vast majority of TM456t from refolding preps was

non-monomeric (often tetrameric or trimeric), indicating inappropriate disulfide bonding between monomers. This was interesting since the protein eluted from the Ni columns after refolding had very detectable specific activity, indicating even in a mis-folded or multimeric form, many TM molecules were able to bind to thrombin and alter its activity.

3. TM456m-binding produces chemical shift differences across apo-thrombin

Many resonances were missing from the HSQC spectrum of TM456m-bound thrombin both because more residues fall in the intermediate timescale invisible to NMR and because some residues have chemical shift differences too great to assign by assignment transfer (Figure 3.7). Nevertheless, 178 amides were assigned by transferring assignments from the apo-thrombin HSQC spectrum alone (which has 213 amide assignments). This allowed us to calculate weighted average chemical shift differences for the amides in TM456m-bound thrombin vs. apo-thrombin (Figure 3.8A and 3.8B). Residues which had amide chemical shift differences greater than 1 standard deviation were: E18-G19(39-40) at the N-terminus; A22(43), G25(46), S27(48), and W29(50) in the peptide between the N-terminus and the 30s loop which makes up part of the TM-binding site; S45-L46(67-68) and T54-A55(76-77) of the N-terminal β-barrel, the β-barrel closest to the TM-binding site; T60i(91) and N62(93) in the 60s loop, close to the TM-binding site; H91(123), R93(125), and N95(127), preceding the 90s loop; L105-M106(138-139), L108(141), and A113(146) in the N-terminal β-barrel and a projecting loop near the TM-binding site; K149e-G150(190-

Table 3.2. Representative raw refolding screen data from a TM456t refolding screen for Tris-HCl concentration vs. Sarkosyl concentration. Table shows raw data from microtitre plate reader before converting to specific activity. Refolding conditions with the most active TM456t (> 190) are shown in bold and indicate that higher concentrations of Sarkosyl and Tris are beneficial for folding.

		% Sarkosyl (w/v)					
		0%	3%	6%	9%	15%	18%
Tris-HCl pH 8.2	50 mM	87.8	72.5	74.4	76.2	142.1	119.1
	100 mM	126.1	148.8	181.4	**192.4**	**192.6**	**199.1**
	150 mM	138.5	138.5	151.2	**201.6**	141.12	138.2
	200 mM	102.0	103.1	134.4	**198.9**	139.8	**198.5**

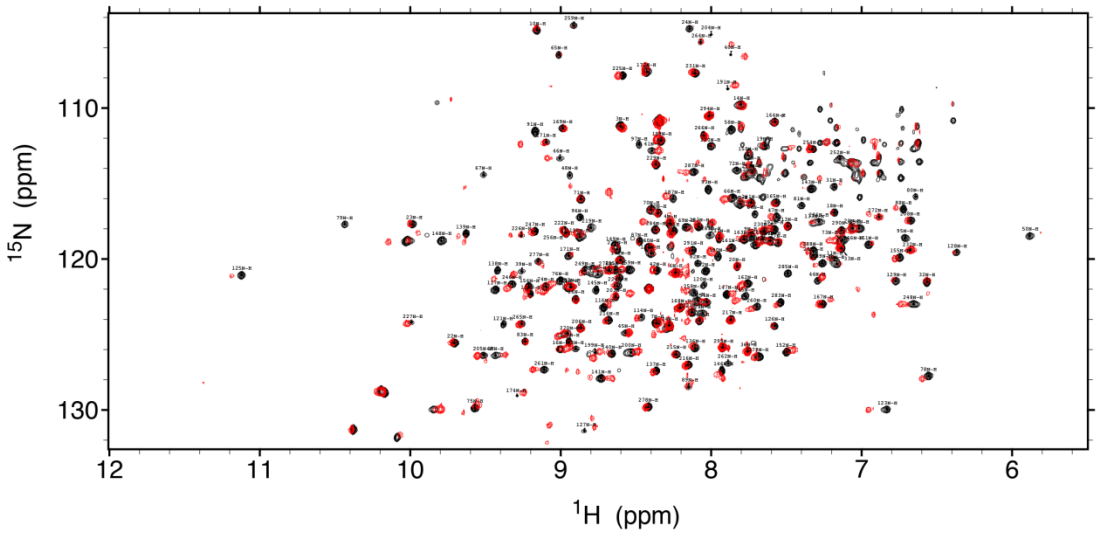


Figure 3.7. The HSQC of TM456m-bound S195M-thrombin (red) is overlaid on top of the HSQC of S195M-thrombin (black).

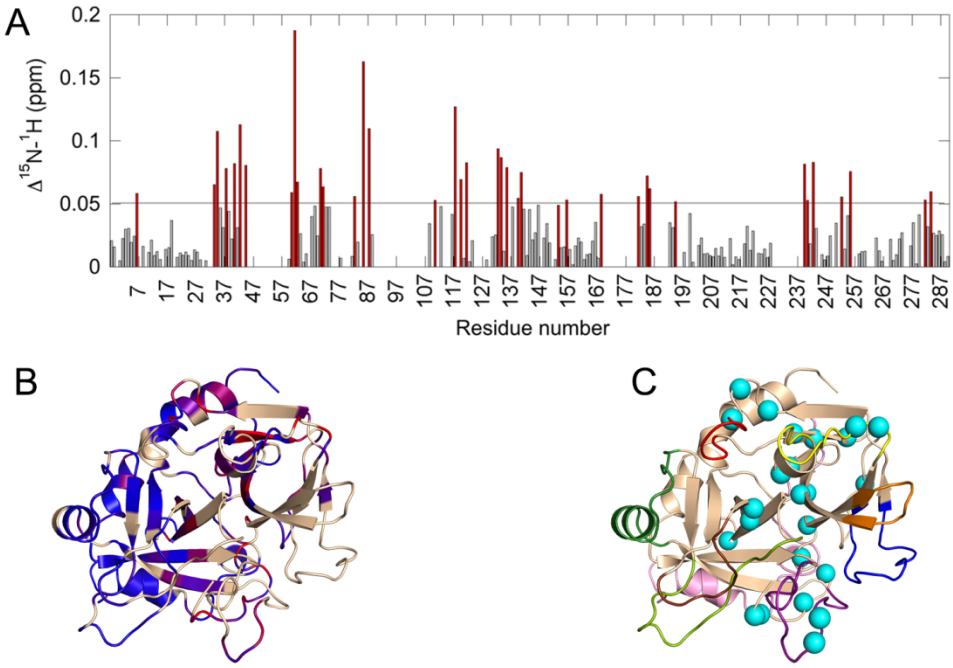


Figure 3.8. Weighted average chemical shift differences between TM456m-bound S195M-thrombin and unbound S195M-thrombin. (A) A plot shows the weighted average chemical shift differences between TM456m-bound S195M-thrombin and unbound S195M-thrombin. The average weighted chemical shift difference for each residue was calculated using the following equation: $\left([(\Delta\delta_{HN})^2 + (\Delta\delta_{NH})^2/25]/2\right)^{1/2}$, where $\Delta\delta_{HN}$ and $\Delta\delta_{NH}$ are the chemical shift differences in the amide ^{1}H and ^{15}N, respectively. (B) A visualization of the chemical shift differences is shown on a crystal structure of thrombin (1PPB). Residues colored in red have higher chemical shift differences. Residues colored in blue have no chemical shift differences. Residues colored in tan are missing data from one or both sets of proteins. (C) A visualization of those residues with chemical shift differences greater than 1 standard deviation above the average (cyan spheres) shown on a crystal structure of thrombin (1PPB) with loops colored as follows: The light-chain is colored pink, the 30s loop is colored orange, the 60s loop is colored yellow, the 70s loop is colored blue, the 90s loop is colored red, the γ-loop is colored purple, the 170s loop is colored green, the 180s loop is colored light green, the Na^{+}-binding loop is colored brown.

191) in the γ-loop; F199(245), K202(248), and V213(261) in the C-terminal β-barrel (Figure 3.8C). (Note: thrombin residues are numbered here using the chymotrypsin numbering scheme followed by the sequential numbering scheme, used in all plots, in parentheses).

D. Discussion

1. A Suitable TM Construct for NMR

A K_d of 2 nM for full-length TM456 was measured by previous SPR experiments in our lab,[35] and previous protein C activation assay results show that full-length TM456 has a $K_{M,TM}$ of around 20 nM.[22] Based on our protein C activation assay work here and previously, we know that the presence of EGF6 does not affect the affinity of protein C for thrombin. Therefore, the increase in $K_{M,TM}$ for TM456m is primarily a result of a change in the on or off rate of thrombin binding to TM456m, not the on or off rate of thrombin binding to protein C. This makes it likely that TM456m has a K_d close to 6 nM (a K_d 10-times less than the $K_{M,TM}$, as seen in full-length TM456). Preliminary SPR results for TM456m confirm this approximate measurement (Handley, unpublished). Considering a K_d of 6 nM for the complex and a thrombin concentration of 112 μM and a TM456m concentration of 168 μM, the complex is expected to be over 99% bound in our NMR experiments.

Thus, TM456m has the activity, thrombin-binding capacity, and stability to make it a suitable candidate for use in complex with thrombin in future NMR

dynamics experiments. It is clear that the addition of the first 23 amino acids of EGF6, a domain long known to be important for thrombin binding, is responsible for this improvement over TM45. However, it is unclear why the removal of the last 13 amino acids of EGF6 so greatly improves the stability of TM456m over TM456. A cursory examination of the predicted disordered regions in TM456m vs. full-length TM456 using the webtools, DisEMBL and GlobPlot (http://www.expasy.org/), offers some semblance of an explanation. Both tools found that not only were the additional 13 amino acids of full-length TM456 highly disordered, but also their addition caused higher disorder in $Phe_{443} - G_{449}$ compared to the shorter TM456m. This disorder might be an explanation for higher degree of aggregation that occurs for full-length TM456.

2. Refolding of TM456m from *E. coli* is not efficient enough for preparation of NMR samples

^{15}N-labeling of proteins for NMR is much more straightforward in *E. coli* than in *P. pastoris*. In a regular unlabeled *P. pastoris* fermentation, NH_4OH is used to control pH, but also doubles as a nitrogen source. In a ^{15}N-labeled fermentation, however, because the cost of $^{15}NH_4OH$ is prohibitively expensive, $(^{15}NH_4)_2SO_4$ is used as the sole nitrogen source and NH_4OH is replaced with a mixture of KOH/NaOH to control the pH. This substitution unfortunately leads to the build up of salts (primarily K_2SO_4) in the fermentation media overtime. *P. pastoris* will cease growing under these high ionic strength conditions,[36] unless the media is replaced part way through the fermentation. For this reason, not only are these ^{15}N-labeled fermentations more tedious, but they are also more unstable. This was the primary motivation behind our

attempts to produce TM456m in *E. coli*, which has a much simpler and quicker isotopic labeling strategy.

Previous alanine-scanning mutagenesis work done on TM456 was done by expressing TM456 mutants in the *E. coli* periplasm and assaying the activity of periplasmic extracts.[37] This indicated that TM might be a viable candidate for purification from *E. coli*, despite the organism's lack of glycosylation machinery. Early on, our lab attempted periplasmic expression of TM456 in *E. coli* (subcloning TM456t into pET26b(+)), but saw almost no expression of the protein or activity, even when using a co-expression plasmid which promotes periplasmic folding of disulfide-bonded proteins, pTUM4.[38]

It is likely that the periplasmic extracts of Nagashima et al. did not necessarily contain TM molecules which were all properly folded, but contained some fraction of TM molecules that retained activity, allowing them to still assay the effects of various mutations. Their results, which identify residues in TM that are specific for its activity, are widely accepted in the thrombomodulin field. Our refolding from *E. coli* inclusion bodies approach appears to be an alternative approach to testing various thrombomodulin mutations for their activity (similar to the approach of Nagashima et al.), but does not appear to be a viable option for the expression of TM456 for use in NMR, unfortunately. TM456m expression from *Pichia pastoris* fermentations was therefore the expression procedure used for producing TM456m for HDXMS studies (see Chapter 4) and NMR (see below and Chapter 5).

3. Allostery Evident in Thrombin upon TM-binding

The HSQC of TM456m-bound thrombin is the first of its kind and reveals significant chemical shift differences far beyond the TM-binding site. These chemical shift differences would often be interpreted as changes in structure, but, from previous work, it is clear that TM456 does not change the structure of thrombin, since the crystal structure of thrombin remains unchanged upon TM-binding.[24] Instead, TM seems to alter the slow-timescale dynamics of loops around the thrombin active site, as indicated by aMD simulations.[31] It is well known that chemical shift differences are reflective of a change in the local chemical environment surrounding the measured nucleus and can be caused by changes in torsion angle, hydrogen bonding, secondary structural elements, or changes in local electrical charges or currents.[39] Chemical shift differences can also be due to a change in dynamics, however. A change in dynamics will essentially add to, reduce, or shift the conformational ensemble that the protein occupies, which can cause a change in chemical environment for a nucleus.

Interestingly, many of the residues which show significant chemical shift differences upon TM456m-binding also show μs-ms dynamics in apo-thrombin: E18(39) at the N-terminus; A22(43) and S27(48) in the peptide between the N-terminus and the 30s loop which makes up part of the TM-binding site; S45(67) of the N-terminal β-barrel; K149e-G150(190-191) in the ϒ-loop; F199(245) in the C-terminal β-barrel (see Chapter 5). In particular, the region around the N-terminus of the heavy chain and γ-loop is of particular interest. In our HDXMS studies, we saw long-distance deuterium protection at the N-terminus when TM was bound to ABE1 (see Chapter 4). And in our complete NMR dynamics characterization (see Chapter 5), we were able to fit the μs-ms timescale dynamics present in the region around the N-

terminus to a two-state exchange model. Taken together, it appears that TM-binding allosterically promotes the burial of the N-terminus of the heavy chain, which is known to be necessary for proteolytic activity in serine proteases.[40]

E. References

1. Bode, W. (2006). Structure and interaction modes of thrombin. *Blood Cells, Molecules, and Diseases* **36**, 122-130.

2. (2013). Deaths: Final Data for 2013. CDC/National Center for Health Statistics.

3. Esmon, C. T. (1989). The roles of protein C and thrombomodulin in the regulation of blood coagulation. *J Biol Chem* **264**, 4743-6.

4. Ito, T. & Maruyama, I. (2011). Thrombomodulin: protectorate God of the vasculature in thrombosis and inflammation. *J Thromb Haemost*, 168-73.

5. van Iersel, T., Stroissnig, H., Giesen, P., Wemer, J. & Wilhelm-Ogunbiyi, K. (2011). Phase I study of Solulin, a novel recombinant soluble human thrombomodulin analogue. *Thromb Haemost* **105**, 302-12.

6. Conway, E. M. (2012). Thrombomodulin and its role in inflammation. *Semin Immunopathol* **34**, 107-25.

7. Conway, E. M., Van de Wouwer, M., Pollefeyt, S., Jurk, K., Van Aken, H., De Vriese, A., Weitz, J. I., Weiler, H., Hellings, P. W., Schaeffer, P., Herbert, J. M., Collen, D. & Theilmeier, G. (2002). The lectin-like domain of thrombomodulin confers protection from neutrophil-mediated tissue damage by suppressing adhesion molecule expression via nuclear factor kappaB and mitogen-activated protein kinase pathways. *J Exp Med* **196**, 565-77.

8. Geudens, N., Van de Wouwer, M., Vanaudenaerde, B. M., Vos, R., Van De Wauwer, C., Verleden, G. M., Verbeken, E., Lerut, T., Van Raemdonck, D. E. & Conway, E. M. (2008). The lectin-like domain of thrombomodulin protects against ischaemia-reperfusion lung injury. *Eur Respir J* **32**, 862-70.

9. Van de Wouwer, M., Plaisance, S., De Vriese, A., Waelkens, E., Collen, D., Persson, J., Daha, M. R. & Conway, E. M. (2006). The lectin-like domain of thrombomodulin interferes with complement activation and protects against arthritis. *J Thromb Haemost* **4**, 1813-24.

10. Huang, H. C., Shi, G. Y., Jiang, S. J., Shi, C. S., Wu, C. M., Yang, H. Y. & Wu, H. L. (2003). Thrombomodulin-mediated cell adhesion: involvement of its lectin-like domain. *J Biol Chem* **278**, 46750-9.

11. Villoutreix, B. & Dahlback, B. (1998). Molecular model for the C-type lectin domain of human thrombomodulin. . *J Mol Model* **4**, 310–322.

12. Weisel, J. W., Nagaswami, C., Young, T. A. & Light, D. R. (1996). The shape of thrombomodulin and interactions with thrombin as determined by electron microscopy. *J Biol Chem* **271**, 31485-90.

13. Zushi, M., Gomi, K., Yamamoto, S., Maruyama, I., Hayashi, T. & Suzuki, K. (1989). The last three consecutive epidermal growth factor-like structures of human thrombomodulin comprise the minimum functional domain for protein C-activating cofactor activity and anticoagulant activity. *J Biol Chem* **264**, 10351-3.

14. Suzuki, K., Hayashi, T., Nishioka, J., Kosaka, Y., Zushi, M., Honda, G. & Yamamoto, S. (1989). A domain composed of epidermal growth factor-like structures of human thrombomodulin is essential for thrombin binding and for protein C activation. *J Biol Chem* **264**, 4872-6.

15. Kurosawa, S., Stearns, D. J., Jackson, K. W. & Esmon, C. T. (1988). A 10-kDa cyanogen bromide fragment from the epidermal growth factor homology domain of rabbit thrombomodulin contains the primary thrombin binding site. *J Biol Chem* **263**, 5993-6.

16. Wang, W., Nagashima, M., Schneider, M., Morser, J. & Nesheim, M. (2000). Elements of the primary structure of thrombomodulin required for efficient thrombin-activable fibrinolysis inhibitor activation. *J Biol Chem* **275**, 22942-7.

17. Bourin, M. C., Lundgren-Akerlund, E. & Lindahl, U. (1990). Isolation and characterization of the glycosaminoglycan component of rabbit thrombomodulin proteoglycan. *J Biol Chem* **265**, 15424-31.

18. Conway, E. M., Nowakowski, B. & Steiner-Mosonyi, M. (1994). Thrombomodulin lacking the cytoplasmic domain efficiently internalizes thrombin via nonclathrin-coated, pit-mediated endocytosis. *J Cell Physiol* **158**, 285-98.

19. Nesheim, M., Wang, W., Boffa, M., Nagashima, M., Morser, J. & Bajzar, L. (1997). Thrombin, thrombomodulin and TAFI in the molecular link between coagulation and fibrinolysis. *Thromb Haemost* **78**, 386-91.

20. Wu, C., Kim, P. Y., Manuel, R., Seto, M., Whitlow, M., Nagashima, M., Morser, J., Gils, A., Declerck, P. & Nesheim, M. E. (2009). The roles of

selected arginine and lysine residues of TAFI (Pro-CPU) in its activation to TAFIa by the thrombin-thrombomodulin complex. *J Biol Chem* **284**, 7059-67.

21. Tsiang, M., Lentz, S. R. & Sadler, J. E. (1992). Functional domains of membrane-bound human thrombomodulin. EGF-like domains four to six and the serine/threonine-rich domain are required for cofactor activity. *J Biol Chem* **267**, 6164-70.

22. White, C. E., Hunter, M. J., Meininger, D. P., White, L. R. & Komives, E. A. (1995). Large-scale expression, purification and characterization of small fragments of thrombomodulin: the roles of the sixth domain and of methionine 388. *Protein Eng* **8**, 1177-1187.

23. Kurosawa, S., Stearns, D. J., Jackson, K. W. & Esmon, C. T. (1988). A 10-kDa cyanogen bromide fragment from the epidermal growth factor homology domain of rabbit thrombomodulin contains the primary thrombin binding site. *J Biol Chem* **263**, 5993-5996.

24. Fuentes-Prior, P., Iwanaga, Y., Huber, R., Pagila, R., Rumennik, G., Seto, M., Morser, J., Light, D. R. & Bode, W. (2000). Structural basis for the anticoagulant activity of the thrombin-thrombomodulin complex. *Nature* **404**, 518-525.

25. Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R. & Hofsteenge, J. (1989). The refined 1.9 A crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J* **8**, 3467-3475.

26. Yang, L. & Rezaie, A. R. (2003). The fourth epidermal growth factor-like domain of thrombomodulin interacts with the basic exosite of protein C. *J Biol Chem* **278**, 10484-90.

27. Esmon, C. T., Esmon, N. L. & Harris, K. W. (1982). Complex formation between thrombin and thrombomodulin inhibits both thrombin-catalyzed fibrin formation and factor V activation. *J Biol Chem* **257**, 7944-7.

28. Koeppe, J. R., Seitova, A., Mather, T. & Komives, E. A. (2005). Thrombomodulin tightens the thrombin active site loops to promote protein C activation. *Biochemistry* **44**, 14784-91.

29. Ye, J., Esmon, N. L., Esmon, C. T. & Johnson, A. E. (1991). The active site of thrombin is altered upon binding to Thrombomodulin. Two distinct structural changes are detected by fluorescence, but only one correlates with protein C activation. *J Biol Chem* **266**, 23016-23021.

30. Treuheit, N. A., Beach, M. A. & Komives, E. A. (2011). Thermodynamic compensation upon binding to exosite 1 and the active site of thrombin. *Biochemistry* **50**, 4590-6.

31. Gasper, P. M., Fuglestad, B., Komives, E. A., Markwick, P. R. & McCammon, J. A. (2012). Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proc Natl Acad Sci U S A* **109**, 21216-21222.

32. Fuglestad, B., Gasper, P. M., Tonelli, M., McCammon, J. A., Markwick, P. R. L. & Komives, E. A. (2012). The dynamic structure of thrombin in solution. *Biophys J* **103**, 1-10.

33. Treuheit, N. A. (2013). Solution Biophysical Characterization of the Thrombin-Thrombomodulin Interaction, UC San Diego.

34. Segal, I. H. (1975). *Enzyme kinetics: Behavior and analysis of rapid equilibrium and steady-state enzyme systems*, Wiley-Interscience, New York.

35. Baerga-Ortiz, A., Bergqvist, S., Mandell, J. G. & Komives, E. A. (2004). Two different proteins that compete for binding to thrombin have opposite kinetic and thermodynamic profiles. *Protein Sci* **13**, 166-76.

36. Wood, M. J. & Komives, E. A. (1999). Production of large quantities of isotopically labeled protein in Pichia pastoris by fermentation. *J Biomol NMR* **13**, 149-59.

37. Nagashima, M., Lundh, E., Leonard, J. C., Morser, J. & Parkinson, J. F. (1993). Alanine-scanning mutagenesis of the epidermal growth factor-like domains of human thrombomodulin identifies critical residues for its cofactor activity. *J Biol Chem* **268**, 2888-92.

38. Schlapschy, M., Grimm, S. & Skerra, A. (2006). A system for concomitant overexpression of four periplasmic folding catalysts to improve secretory protein production in Escherichia coli. *Protein Eng Des Sel* **19**, 385-90.

39. Kleckner, I. R. & Foster, M. P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* **1814**, 942-68.

40. Bode, W. (1979). The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding: II. The binding of the pancreatic trypsin inhibitor and of isoleucine-valine and of sequentially related peptides to trypsinogen and to p-guanidinobenzoate trypsinogen. *J Mol Biol* **127**, 357–374.

Chapter IV

Thrombomodulin binding selects the catalytically active form of thrombin.

A. Introduction

Many coagulation proteases are activated by binding to cofactors, and a growing body of evidence supports the concept that these cofactors induce allosteric changes in their cognate enzymes.[1] Whereas most coagulation proteases exist only in an inactive or active state, thrombin is a dual-action protease, with its cofactor, TM, modulating between two of its active states. In the absence of TM, thrombin acts as a procoagulant, cleaving fibrinogen to fibrin. However, when in complex with TM, thrombin acts as an anticoagulant by cleaving and thereby activating protein C.[2; 3] For both procoagulant and anticoagulant activity, thrombin must first be converted from its zymogen, prothrombin, to the active protease α-thrombin. As for all serine proteases, prothrombin is converted to α-thrombin by proteolysis of an internal peptide bond, exposing a new N-terminal Ile-NH_3. Previous work on serine proteases has shown that the newly generated N-terminus inserts in a pocket, known as the "Ile cleft",[4] where the new N-terminal amino group forms a salt bridge with a highly conserved aspartic residue.[5] This salt bridge is thought to promote the formation of a correctly assembled S1 pocket of the active site and thereby full proteolytic activity. Using accelerated molecular dynamics (aMD) simulations, we recently showed that the N-terminus of trypsin inserts from a trypsinogen-like state into this pocket and remains in the pocket once inserted, indicating that the N-terminus is more stable inside the pocket than out.[6] All crystal structures of α-thrombin show the N-terminus in this inserted conformation (rcsb.org).

TM contains six EGF-like domains; the 5th domain interacts directly with thrombin at the fibrinogen binding site, anion binding exosite 1 (ABE1). It has been shown that a TM fragment consisting of only the 5th and 6th EGF-like domains (TM56) is sufficient to bind thrombin and inhibit fibrinogen cleavage, but the additional presence of the 4th EGF-like domain of TM (TM456) is critical to induce the anticoagulant activity of thrombin.[7; 8] TM456 binding to thrombin significantly increases the association rate, k_a, of a variety of active site-directed inhibitors of thrombin,[9; 10; 11; 12] and the k_a for protein C binding is 1000-fold higher for the thrombin:TM456 complex compared to thrombin alone.[13] TM binding dramatically improves the k_{cat} for protein C activation while only modestly affecting the K_M.[8]

The mechanism by which TM456 enhances protein C cleavage is not yet fully understood and is particularly intriguing. The TM binding site in ABE1 is distal to the active site and the essential TM4 domain makes no direct contact with thrombin whatsoever.[14] The simplest explanation for the dramatically increased association rates is that TM456 allosterically alters the conformation of the loops that surround the active site. However, a comparative analysis of the X-ray crystal structures of thrombin (1PPB)[15] and the thrombin-TM456 complex (1DX5)[14] revealed no significant structural differences in the thrombin active site loops. TM4 also forms an extended binding surface for protein C, providing optimal alignment for insertion into the active site and subsequent cleavage.[14] Critical electrostatic interactions between protein C and residues in TM4, including Glu382 in the β4-β5 loop, Asp398, and Glu357, as well as hydrophobic interactions with the aromatic residues Tyr358 and Phe376, were identified. The importance of these residues for the activation of protein

C has been empirically validated by alanine-scanning mutagenesis.[16] However, the proposed 'docking and optimal alignment' mechanism is insufficient to explain the 1000-fold increase in observed association rates, particularly for smaller substrates and other inhibitors, which may enter the active site of thrombin without directly interacting with the TM4 domain.

Several studies including fluorescence,[17] MALDI-TOF amide hydrogen/deuterium exchange mass spectrometry (HDXMS),[18] and isothermal titration calorimetry[19] have identified changes in the thrombin active site region that occur upon binding to different constructs of TM in the absence of protein C. These studies suggest that the presence of TM4 may affect the dynamic properties of both the active site and the loops surrounding it by dynamic allostery.[20; 21; 22] A very recent NMR/aMD study on thrombin bound to the inhibitor D-Phe-Pro-Arg-chloromethylketone (PPACK) identified significant dynamic motions in the active site loops across timescales ranging from picoseconds to tens of microseconds, even with inhibitor bound.[23]

The previous MALDI-TOF HDXMS studies were carried out using a small fragment of TM, TM45 (residues 346-426), which could be expressed and purified in large quantities. However, lacking EGF6, this fragment did not have full thrombin-binding capabilities. In addition, only 50% of the thrombin sequence could be covered using the MALDI-TOF instrumentation. In the present work we report HDXMS experiments using a newly engineered TM construct, TM456m (residues 346-449), which has full thrombin-binding capability. In addition, the experiments were carried out using a Waters Synapt G2Si instrument, yielding an unprecedented 90% sequence

coverage of thrombin. The results provide exciting new evidence of how TM activates thrombin for protein C cleavage.

B. Materials and Methods

1. Design and production of TM456m

TM456m was subcloned from the TM456 gene, which had been previously synthesized using *E. coli*-optimized codons.[8] PCR mutagenesis was used to introduce a stop codon to terminate TM456m at G449 and to change Cys448 to Ser. Details of the expression vectors, transformation, and transformant selection of *Pichia pastoris* clones were all carried out as previously described.[8] In brief, freshly purified TM456m pPic9K expression plasmid was linearized using BglII, and introduced into spheroplasts of *P. pastoris* protease resistant strain, SMD1168, to achieve multicopy insertion of the expression plasmid into the chromosomal DNA resulting in stable transformants. The multicopy transformants were selected by replica plating on G418, a kanamycin analog that penetrates *P. pastoris*. The SMD1168 transformants that produced the highest levels of TM expression were stored as glycerol freezes at -80°C.

2. Preparation of TM456m

Fermentations of *P. pastoris* strains containing multiple copies of the TM expression vectors were carried out in a Bioflow 3000 fermentor (New Brunswick Scientific, N.J.) equipped with a 3L bioreactor. Both TM constructs (TM45 and TM456m) were expressed and purified from *P. pastoris* fermentations, as described previously.[8] The TM proteins are secreted into the media, and initially captured by

anion-exchange chromatography using a QAE Sephadex A-50 column. The protein is then further purified on a HiLoad 26/10 Q Sepharose High Performance column followed by reverse-phase C-18 HPLC using an acetonitrile gradient from 10 – 50% in 20mM NH_4OAc pH 5.65. The protein was lyophilized before being reconstituted in DI H_2O and purified by Superdex 75 size-exclusion chromatography in 50 mM Tris 150 mM NaCl pH 7.4. Protein C activation assays (described below) were used at each step in the purification strategy to determine the highest activity fractions to be taken to the next step.

3. Preparation of Thrombin

All human α-thrombin was expressed and refolded from *E. coli* inclusion bodies, as previously described.[23] Aliquots of α-thrombin (250 μg/mL) were stored at -80°C for no longer than 1 month before use. PPACK-thrombin was prepared by incubating α-thrombin with a 10X-molar excess of PPACK (Haematologic Technologies, Inc.) for 30 minutes at room temperature. The protein was then purified by Superdex 75 size-exclusion chromatography in 25 mM phosphate 100 mM NaCl pH 6.5. Aliquots of PPACK-thrombin were stored at -80°C for no longer than 1 month before use.

4. Protein C Activation Assays

Protein C activation assays were performed as previously described.[8] Briefly, a discontinuous assay for TM activation of thrombin towards protein C cleavage was performed by first incubating TM with thrombin before adding protein C. Following a 20 min incubation with protein C, the thrombin was inactivated by addition of

heparin/antithrombin-III, and the activated protein C (aPC) was assayed by addition of the aPC-specific chromagenic substrate, S-2366 (Diapharma, West Chester, OH). The cleavage of the chromogenic substrate results in a linear increase in absorbance at 405 nm over time, which is proportional to the amount of aPC. Kinetic constants (k_{cat}, $K_{M,TM}$, and $K_{M,PC}$) were determined for each TM construct as discussed previously.[8] $K_{M,TM}$ and k_{cat} were measured at constant TM concentration and varying amounts of protein C. All protein C assays used to calculate kinetic constants were repeated in triplicate; values agreed within 10%.

5. Hydrogen-Deuterium Exchange Mass Spectrometry (HDXMS)

PPACK-thrombin and α-thrombin samples were prepared for HDXMS from frozen aliquots (~7.4 μM) using pre-rinsed 0.5 mL 3K MWCO Amicon concentrators, spinning in a microcentrifuge at 14,000 g in 5 min intervals at room temperature. The appropriate amount of each protein was concentrated to make 400 μL with a final concentration of 10 μM in 25 mM phosphate 100 mM NaCl pH 6.5. A portion, 50 μL, of each protein was removed for peptide identification analysis, and the remaining 350 μL was diluted to 5 μM for the deuterium exchange experiments. The TM456m:α-thrombin complex was prepared for HDXMS by first concentrating the TM456m to ~40 μM and fully exchanging it into 25 mM phosphate 100 mM NaCl pH 6.5 using a pre-rinsed 15 mL 3K MWCO Amicon concentrator. The appropriate amount of α-thrombin and TM456m were mixed and concentrated (as above) in a pre-rinsed 0.5 mL 3K MWCO Amicon to make 5 μM α-thrombin:150 μM TM456m in a final

volume of 100 μL. This ratio results in a thrombin sample that is 99.97% bound with TM456m.

HDXMS was performed using a Waters Synapt G2Si system with HDX technology (Waters Corporation). Deuterium exchange reactions were prepared using a LEAP HDX PAL autosampler (Leap Technologies, Carrboro, NC). D_2O buffer was prepared by lyophilizing 10 mL of 25 mM phosphate 100 mM NaCl pH 6.5 and resuspending in 99.96% D_2O immediately before use. Each deuterium exchange time point (0 min, 30 s, 1 min, 2 min, 5 min, 10 min) was measured in triplicate. For each deuteration time point, 5 μL of protein (or complex) was mixed with 55 μL of D_2O buffer at 25°C and allowed to incubate for 0-10 min. The deuterium exchange was quenched for 1 min at 1°C by combining 50 μL of the deuteration reaction with 50 μL of 3M Guanidine HCl, 0.1% Formic Acid, 250 mM TCEP pH 2.66. The quenched sample was then injected into the 50 μL sample loop, followed by digestion on an in-line pepsin column (immobilized pepsin, Pierce, Inc.) at 15°C. The resulting peptides were captured on a BEH C18 Vanguard pre-column, separated by analytical chromatography (Acquity UPLC BEH C18, 1.7 μM, 1.0 X 50 MM, Waters Corporation) using a 7-85% acetonitrile in 0.1% formic acid over 7.5 min, and electrosprayed into the Waters SYNAPT G2Si quadrupole time-of-flight mass spectrometer. The mass spectrometer was set to collect data in the Mobility, ESI+ mode; mass acquisition range of 200–2,000 (m/z); scan time 0.4 s. Continuous lock mass correction was accomplished with infusion of leu-enkephalin (m/z = 556.277) every 30 s (mass accuracy of 1 ppm for calibration standard). For peptide identification, the mass spectrometer was set to collect data in MS^E, ESI+ mode

instead. The peptides were identified from triplicate analyses of 10 μM α-thrombin, and data were analyzed using PLGS 2.5 (Waters Corporation). Peptides masses were identified using a minimum number of 250 ion counts for low energy peptides and 50 ion counts for their fragment ions; the peptides also had to be larger than 1500 Da. The following cut-offs were used to filter peptide sequence matches: minimum products per amino acid of 0.2, minimum score of 8, maximum MH+ error of 3 ppm, a retention time RSD of 5%, and the peptides had to be present in all three ID runs collected. The peptides identified in PLGS were then analyzed in DynamX 3.0 (Waters Corporation). The relative deuterium uptake for each peptide was calculated by comparing the centroids of the mass envelopes of the deuterated samples with the undeuterated controls following previously published methods.[24]

C. Results

1. Properties of TM456m

Unlike full-length TM456 (residues 345-462), TM456m (residues 346-449) is soluble, stable, and can be concentrated without aggregating. Long-term measurement of specific activity of a purified 1.4 mM TM456m sample stored at 4°C showed that TM456m retained full activity and remained soluble over several months. Compared to the previously studied TM45,[8] TM456m includes an additional 23 residues of EGF6, terminating at G449. The measured specific activity of TM456m was 3 times higher than TM45 and the measured $K_{M,TM}$ was 5 times lower (tighter) than the $K_{M,TM}$

of TM45. The measured $K_{M,PC}$ and k_{cat} for TM456m was approximately the same as for TM45 and TM456.[8]

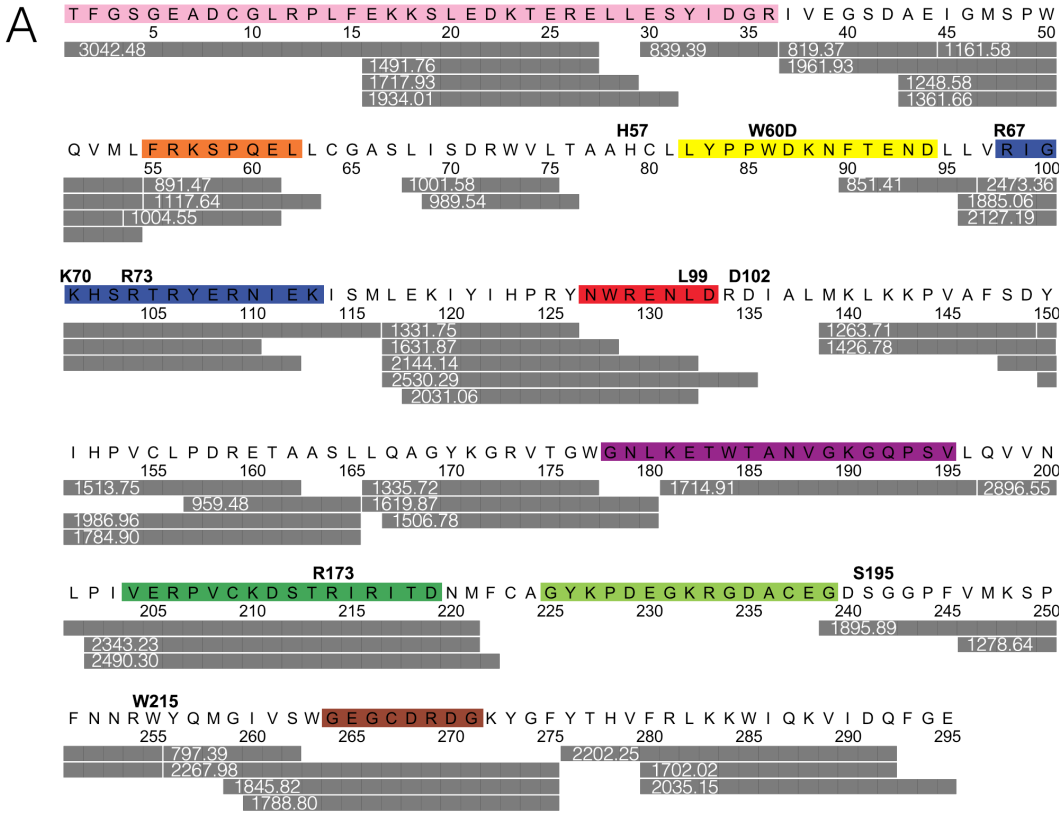
2. HDXMS Coverage

Stringent filtering of the peptides identified by PLGS resulted in 46 peptides that covered 87.8% of the α-thrombin sequence with a redundancy of 2.47 (Figure 4.1). Despite inclusion of high concentrations of TCEP in the quench solution, three of the regions missing in the coverage map (residues 64-67, 77-89, and 223-238) correspond to regions in and around disulfide bonds. Another region (residues 136-138) is surrounded by a large number of peptic cleavage sites, likely resulting in peptides that were too short to be unambiguously identified.

3. Decreased solvent accessibility of the PPACK binding site on α-thrombin

Several peptides that showed significantly decreased deuterium incorporation upon PPACK binding were classified as directly binding because they contained residues with amides within 5 Å of the PPACK molecule itself (as determined from the 1PPB crystal structure of PPACK-bound thrombin). This was a conservative determination of the amides that are physically protected by the PPACK molecule, and therefore have decreased solvent accessibility. Three peptides with decreased deuterium incorporation covered residues Asn131-Leu132, which are within 5 Å of PPACK: residues 117-132 (residues 85-99_{CT}, MH+ mass of 2144.140), residues 117-135 (residues 85-102_{CT}, MH+ mass of 2530.295), and residues 118-132 (residues 86-99_{CT}, MH+ mass of 2031.056). The deuterium uptake plot for the peptide

Figure 4.1. HDXMS coverage map for α-thrombin. (A) HDXMS coverage map for α-thrombin. Peptides are marked in grey bars underneath the amino acid sequence and sequential numbering of α-thrombin. Peptide masses are printed in white within each grey bar. The amino acid sequence is highlighted in different colors to mark important regions of α-thrombin: the light-chain is colored pink, the 30s loop is colored orange, the 60s loop is colored yellow, the 70s loop is colored blue, the 90s loop is colored red, the γ-loop is colored purple, the 170s loop is colored green, the 180s loop is colored light green, and the Na^{+}-binding loop is colored brown. The chymotrypsin numbering is given for some important residues in bold over the amino acid sequence. (B) A crystal structure of PPACK-thrombin (1PPB) with loops colored, as identified in (A), and PPACK drawn in black sticks. A crystal structure of thrombin bound to TM456 (1DX5) was used to align TM456m with the PPACK-thrombin structure to indicate the TM binding site. TM456m is shown in cyan.


A
T F G S G E A D C G L R P L F E K K S L E D K T E R E L L E S Y I D G R I V E G S D A E I G M S P W
5 10 15 20 25 30 35 40 45 50
3042.48
839.39
819.37
1161.58
1491.76
1961.93
1717.93
1248.58
1934.01
1361.66
H57
W60D
R67
Q V M L F R K S P Q E L L C G A S L I S D R W V L T A A H C L L Y P P W D K N F T E N D L L V R I G
55 60 65 70 75 80 85 90 95 100
891.47
1001.58
851.41
2473.36
1117.64
989.54
1885.06
1004.55
2127.19
K70
R73
L99
D102
K H S R T R Y E R N I E K I S M L E K I Y I H P R Y N W R E N L D R D I A L M K L K K P V A F S D Y
105 110 115 120 125 130 135 140 145 150
1331.75
1263.71
1631.87
1426.78
2144.14
2530.29
2031.06
I H P V C L P D R E T A A S L L Q A G Y K G R V T G W G N L K E T W T A N V G K G Q P S V L Q V V N
155 160 165 170 175 180 185 190 195 200
1513.75
1335.72
1714.91
2896.55
959.48
1619.87
1986.96
1506.78
1784.90
R173
S195
L P I V E R P V C K D S T R I R I T D N M F C A G Y K P D E G K R G D A C E G D S G G P F V M K S P
205 210 215 220 225 230 235 240 245 250
1895.89
2343.23
1278.64
2490.30
W215
F N N R W Y Q M G I V S W G E G C D R D G K Y G F Y T H V F R L K K W I Q K V I D Q F G E
255 260 265 270 275 280 285 290 295
797.39
2202.25
2267.98
1702.02
1845.82
2035.15
1788.80
B

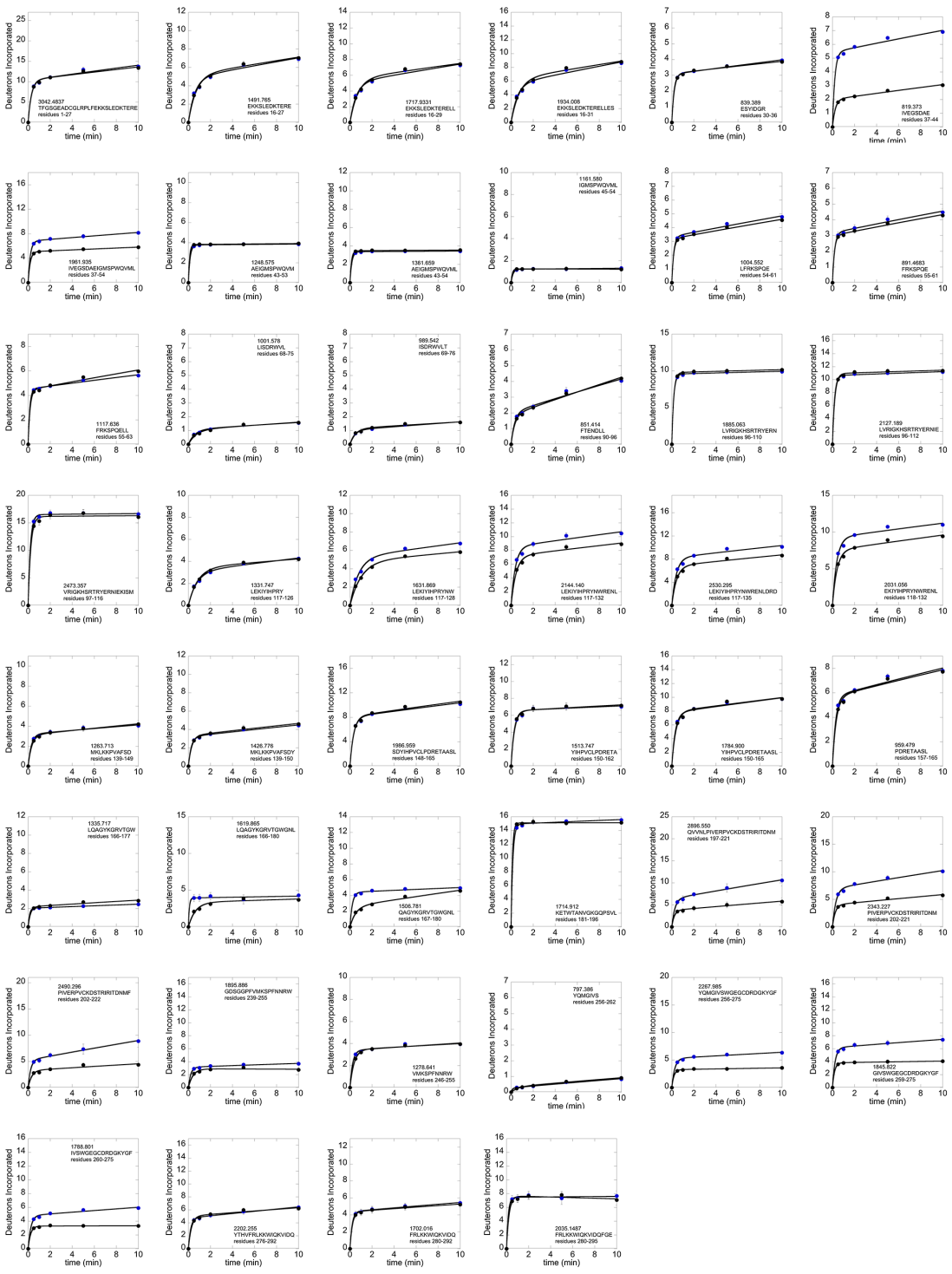


Figure 4.2. All relative deuterium incorporation plots for PPACK-thrombin versus α-thrombin. PPACK-thrombin data is in black, and α-thrombin data is in blue. The max number of amides that could theoretically be incorporated by the peptide is the max number on each y-axis.

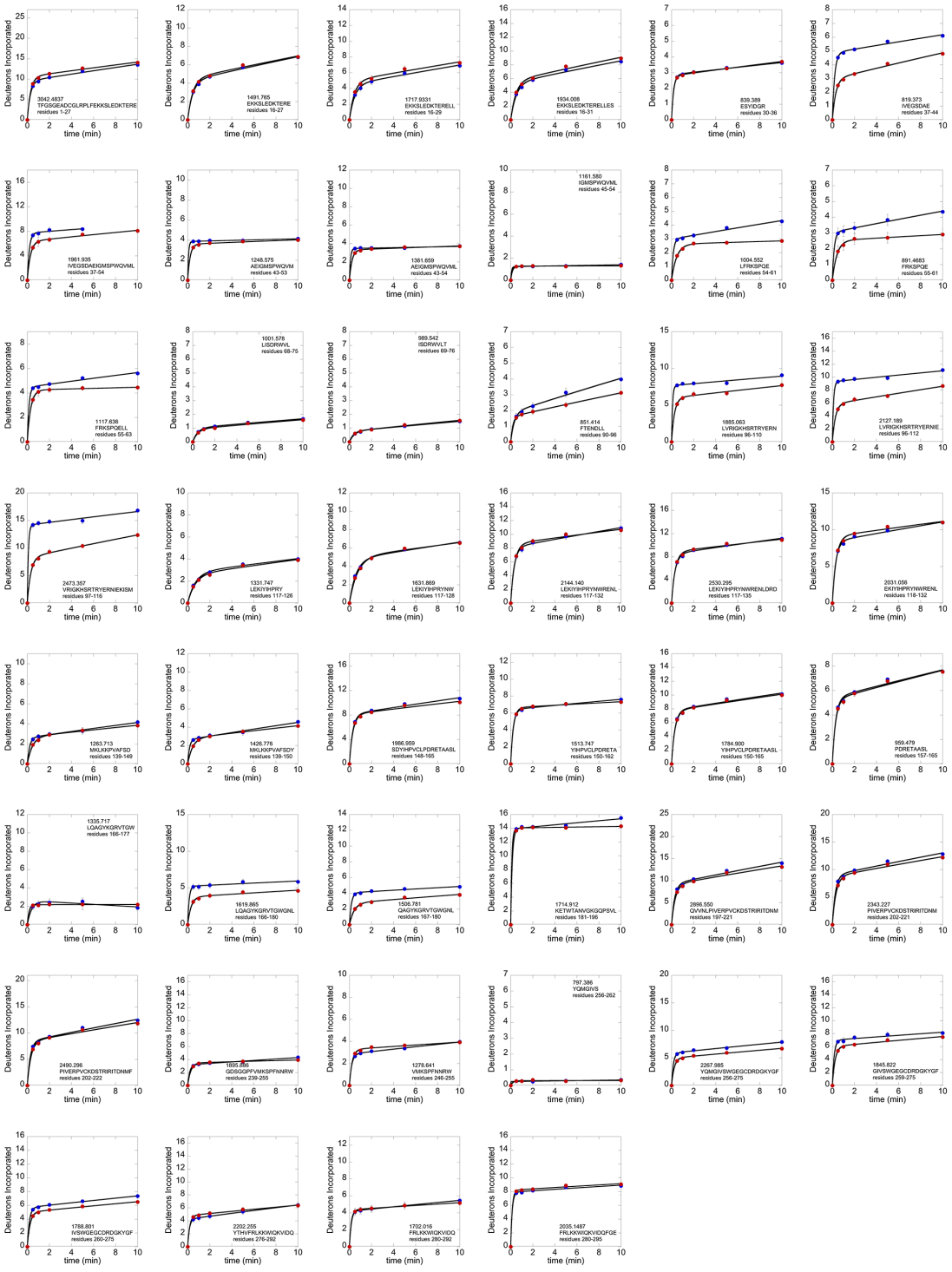


Figure 4.3. All relative deuterium incorporation plots for TM456m-bound thrombin versus α-thrombin. TM456m-bound thrombin data is in red, and α-thrombin is in blue. The max number of amides that could theoretically be incorporated by the peptide is the max number on each y-axis.

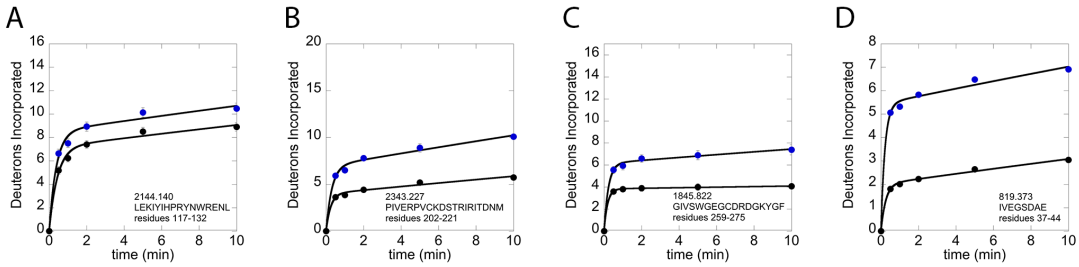


Figure 4.4. Selected PPACK-thrombin versus α-thrombin deuterium uptake plots. Relative deuterium uptake plots for peptides in PPACK-thrombin in black versus unbound α-thrombin in blue. (A) The 90s loop, which directly contacts PPACK, residues 117-132 (residues 85-99_{CT}, MH+ mass of 2144.140), (B) the 170s loop, residues 202-221 (residues 161-180_{CT}, MH+ mass of 2343.277), (C) the Na^{+}-binding loop, residues 259-275 (residues 211-227_{CT}, MH+ mass of 1845.822) which also directly contacts PPACK, and (D) the N-terminus of the heavy chain, residues 37-44 (residues 16-23_{CT}, MH+ mass of 819.373).

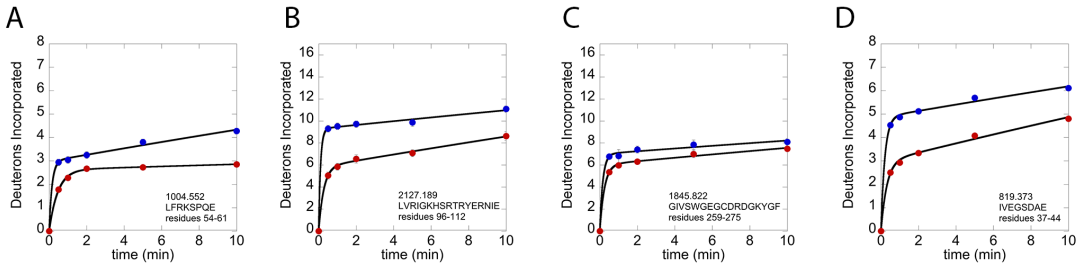


Figure 4.5. Selected TM456m-thrombin versus α-thrombin deuterium uptake plots. Relative deuterium uptake plots for peptides in TM456m-bound α-thrombin in red versus unbound α-thrombin in blue. (A) The 30s loop, residues 54-61 (residues 33-39_{CT}, MH+ mass of 1004.552), (B) the 70s loop in ABE1, residues 96-112 (residues 65-80_{CT}, MH+ mass of 2127.189), (C) the Na^{+}-binding site, residues 259-275 (211-227_{CT}, MH+ mass of 1845.822), and (D) the N-terminus of the heavy chain, residues 37-44 (residues 16-23_{CT}, MH+ mass of 819.373).

corresponding to residues 117-132 is shown in Figure 4.4A, and the rest of the uptake plots are shown in Figure 4.2. Also within 5 Å of PPACK are residues Asp240-Ser241, which were covered by one peptide with decreased deuterium incorporation: residues 239-255 (residues 193-207_{CT}, MH+ mass of 1895.886). All residues with amides within 5 Å of PPACK are highlighted in yellow in Figure 4.6A.

4. Decreased solvent accessibility of the TM456m binding site on α-thrombin

Similar to what was observed in previous work with TM45,[25] TM456m binding was observed to protect amides in ABE1, the TM binding site, which includes the 30s loop and the 70s loop.[14] Three peptides covered the 30s loop: residues 54-61(residues 33-39_{CT}, MH+ mass of 1004.552), residues 55-61 (residues 34-39_{CT}, MH+ mass of 891.468), and residues 55-63 (residues 33-41_{CT}, MH+ mass of 1117.636). The 70s loop was also covered by three peptides: residues 96-110 (residues 65-78_{CT}, MH+ mass of 1885.063), residues 96-112 (residues 65-80_{CT}, MH+ mass of 2127.189), and residues 97-116 (residues 66-84_{CT}, MH+ mass of 2473.357). All peptides within ABE1 showed significantly lower deuterium incorporation when TM456m was bound; the deuterium uptake plot for the peptide corresponding to residues 54-61 is shown in Figure 4.5A and the deuterium uptake plot for the peptide corresponding to residues 96-112 is shown in Figure 4.5B. These peptides are highlighted in yellow in Figure 4.6B.

5. Differences between HDXMS results in studies of TM45 and TM456m

We previously reported slight decreases in amide exchange of the 90s loop and the C-terminal helix of thrombin upon TM45 binding,[18; 25] which were attributed to allosteric changes in the thrombin active site upon TM45 binding. These decreases in amide exchange were not observed when thrombin was bound to the new, higher affinity TM456m, even though the same peptides that were observed before were also observed in the current studies. The 90s loop and the C-terminal helix form part of ABE2, which, like ABE1, is a positively charged patch on the surface of thrombin. The experiments presented here were performed at much lower protein concentrations (5 μM) than the previous experiments (100 μM). It is possible that the fifth EGF-like domain, which is highly negatively charged and was fully exposed in TM45, may have been interacting non-specifically with ABE2. We suspect that the addition of most of the sixth EGF-like domain prevents this non-specific interaction.

6. Decreased amide exchange in distal regions within α-thrombin when PPACK is bound

Regions further than 5 Å from the PPACK molecule were also observed to have decreased deuterium incorporation when PPACK was bound. One such region contains residues 117-128 (residues 85-96_{CT}, MH+ mass of 1631.869). In addition, significant decreases in deuterium incorporation were seen in the 170s loop upon PPACK binding. This region was covered by three peptides: residues 197-221 (residues 156-180_{CT}, MH+ mass of 2896.550), residues 202-221 (residues 161-180_{CT}, MH+ mass of 2343.227), and residues 202-222 (residues 161-181_{CT}, MH+ mass of

2490.296). The associated deuterium uptake plots can be found in Figure 4.2, and a representative plot of the data for residues 202-221 is shown in Figure 4.4B.

7. Decreased amide exchange in the sodium-binding loop upon PPACK binding

Residues Trp263-Glu265, Cys267-Asp268, and Gly274-Phe275, all found within the Na+-binding loop, have amides within 5 Å of PPACK. Three peptides covering this region all showed lower deuterium incorporation in PPACK-bound thrombin: residues 256-275 (residues 208-227_{CT}, MH+ mass of 2267.985), residues 259-275 (residues 211-227_{CT}, MH+ mass of 1845.822), and residues 260-275 (residues 212-227_{CT}, MH+ mass of 1788.801). The deuterium uptake plot for the peptide corresponding to residues 259-275 is shown in Figure 4.4C and the rest of the uptake plots are shown in Figure 4.2.

8. Decreased amide exchange in the N-terminus of the thrombin heavy chain upon PPACK binding

Several important regions of thrombin were not covered in the previous HDXMS studies. These included the N-terminus of the thrombin heavy chain for which a dramatic decrease in amide exchange was observed in PPACK-bound thrombin as compared to apo-thrombin. The decreased exchange was observed in two different N-terminal peptides, one of which covered residues 37-44 (residues 16-23_{CT}, MH+ mass of 819.373) and the other covered residues 37-54 (residues 16-33_{CT}, MH+ mass of 1961.54). The plot showing deuterium uptake in residues 37-44 in apo-

thrombin versus PPACK-thrombin is shown in Figure 4.4D and the uptake plot for residues 37-54 is shown in Figure 4.2.

9. Decreased amide exchange in the γ-loop upon PPACK binding

Three peptides corresponding to the γ-loop residues were analyzed. The peptide covering residues 166-177 (residues 130-141$_{CT}$, MH+ mass of 1335.717) (Figure 4.7A) and the peptide covering residues 181-196 (residues 145-155$_{CT}$, MH+ mass of 1714.912) (Figure 4.7C) did not show differences between α-thrombin and PPACK-thrombin. However, the peptide corresponding to residues 167-180 (residues 131-144$_{CT}$, mass of 1506.781 Da) (Figure 4.7B) showed lower amide exchange in PPACK-bound thrombin. This trend was also observed in the peptide corresponding to residues 166-180 (residues 130-144$_{CT}$, MH+ mass of 1619.865), shown in Figure 4.2. Taken together, these plots indicate decreased amide exchange in the γ-loop is isolated to a region containing only three residues: Gly178-Leu180.

10. Decreased amide exchange in the sodium-binding loop upon TM456m-binding

Deuterium uptake into the Na+-binding loop was lower in TM456m-bound thrombin than in apo-thrombin. Although the differences weren't as large as in the PPACK-thrombin versus apo-thrombin experiment, the difference was observed in three different peptides, all of which covered this loop. The uptake plot for the peptide covering residues 259-275 (residues 211-227$_{CT}$, MH+ mass of 1845.822) is shown in Figure 4.5C and the uptake plot for the peptides covering residues 256-275 (residues

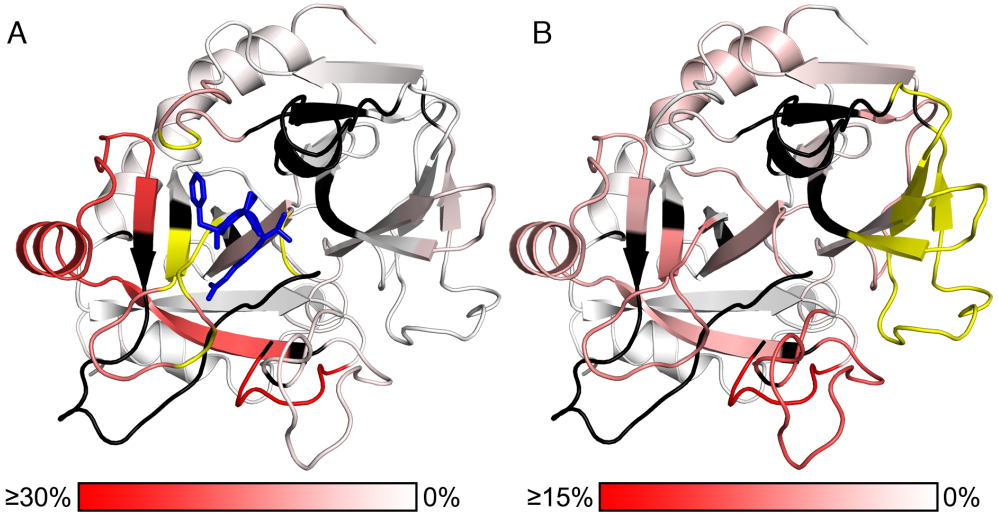


Figure 4.6. Differences in deuterium exchange for different forms of thrombin. Differences in deuterium exchange between (A) PPACK-thrombin versus α-thrombin and (B) TM456m-bound α-thrombin versus unbound α-thrombin displayed on the thrombin structure (1PPB). These differences in deuterium protection were calculated by comparing the deuterium uptake after 10 min incubation with deuterium. Residues are colored with a gradient of white to red, with residues colored in red showing the highest degree of deuterium protection upon binding and residues colored in white showing no deuterium protection upon binding. Residues colored black were not covered in the experiment. In (A), PPACK is shown in blue sticks, and all residues, covered in the HDXMS experiment, which have amides within 5Å of PPACK are colored in yellow. In (B), residues constituting the TM binding site are colored in yellow.

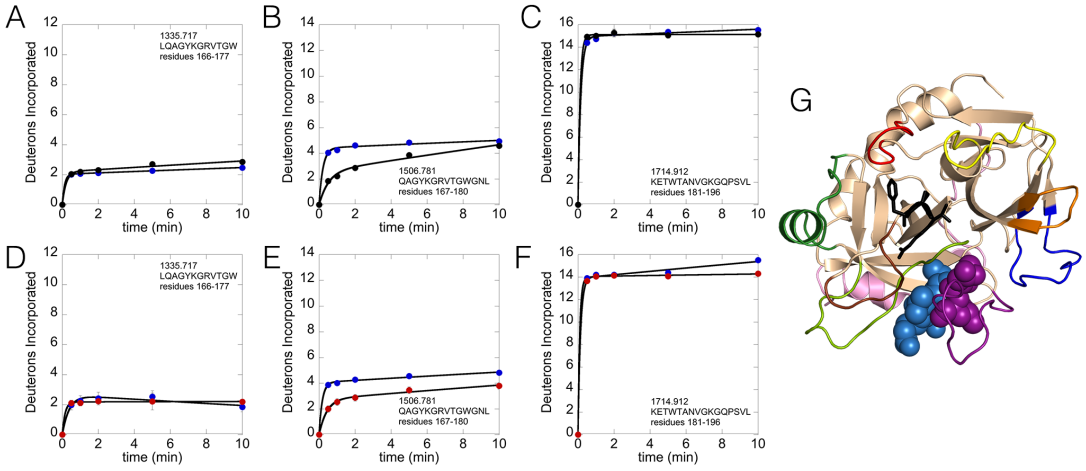


Figure 4.7. Comparison of deuterium uptake plots in regions covering the γ-loop in different forms of thrombin. Comparison of the deuterium uptake in three regions covering the γ-loop in α-thrombin versus PPACK-thrombin (A-C) and α-thrombin versus TM456m-thrombin (D-F): (A, D) residues 166-177 (residues 130-141$_{CT}$, MH+ mass of 1335.717), (B, E) residues 167-180 (residues 131-144$_{CT}$, MH+ mass of 1506.781), and (C, F) residues 181-196 (residues 145-155$_{CT}$, MH+ mass of 1714.912). (G) Structure of thrombin (1PPB) showing the interaction between the residues Gly178-Leu180 (residues 142-144$_{CT}$) of the γ-loop (purple spheres) with residues Ile37-Gly40 (residues 16-19$_{CT}$) of the N-terminus of the heavy chain (blue spheres). The same loop colorings from Figure 4.1B are used.

208-227$_{CT}$, MH+ mass of 2267.985) and residues 260-275 (residues 212-227$_{CT}$, MH+ mass of 1788.801) are shown in Figure 4.3. It is important to note that the PPACK moiety is within 5Å of several residues within this loop, so some of the decreased exchange observed for PPACK-thrombin is likely due to direct alterations of solvent accessibility due to PPACK occlusion of this region. This region is very far from TM456 in the crystal structure, suggesting that TM affects the Na+- binding loop by an allosteric mechanism.

11. Decreased amide exchange in the N-terminus of the thrombin heavy chain upon TM456m binding

Similar to PPACK binding, the N-terminus of the thrombin heavy chain showed significantly decreased deuterium uptake in TM456m-bound thrombin as compared to apo-thrombin. The same two N-terminal peptides in which decreased amide exchange was observed upon PPACK-binding also showed decreases in amide exchange upon TM456m-binding. The uptake plot for the peptide corresponding to residues 37-44 (residues 16-23$_{CT}$, MH+ mass of 819.373) is shown in Figure 4.5D and the uptake plot for residues 37-54 (residues 16-33$_{CT}$, MH+ mass of 1961.54) is shown in Figure 4.3. Again, whereas PPACK occupies the active site and may be expected to promote burial of the N-terminus of the heavy chain, the TM456 binding region is far from the active site. Thus, the decreased amide exchange in the N-terminus of the heavy chain must be caused allosterically.

12. Decreased amide exchange in the N-terminus of the thrombin heavy chain upon TM456m binding

TM456m binding caused similar decreases in deuterium uptake into the γ-loop as were observed for PPACK-binding. Comparative analysis of the deuterium uptake of several overlapping peptides [Figure 4.7D-F and Figure 4.3 for residues 166-180 (residues 130-144$_{CT}$, MH+ mass of 1619.865)], showed that the same region affected by PPACK-binding was also most affected by TM456m-binding. Interestingly, the decreases in amide exchange upon TM456m-binding appeared to be sustained over much longer times as compared to the change caused by PPACK-binding (compare Figure 4.7B with Figure 4.7E).

D. Discussion

A newly designed TM456 analog, TM456m, which contains all of the residues of TM456 except residues P_{450} through C_{462} (the majority of the last loop of EGF6), was prepared. This analog has higher activity, better thrombin binding, and improved stability against aggregation compared to full-length TM456. Using this new analog, we have been able to perform amide HDXMS analysis of the effects of TM binding on thrombin at lower concentrations of the two proteins. As previously reported with TM45, TM456m causes reduced amide exchange throughout the thrombin molecule.[18; 25] Direct binding of TM456m reduces exchange in the 30s and 70s loops of ABE1 in accord with the TM binding site observed in the crystal structure of the thrombin-TM456 complex.[14] Although we previously reported decreased amide exchange in the 90s loop and C-terminal helix of ABE2, this was not observed with

TM456m. These differences can be attributed to possible non-specific interactions of the exposed negative charges in EGF5 in the previously studied TM45 construct.

Here we report, for the first time, that TM binding causes burial of the N-terminus of the heavy chain, a region of thrombin that was not covered in previous MALDI-TOF-based HDXMS studies.[18; 25] The stabilization of the N-terminus of the heavy chain in the Ile cleft is observed in PPACK-bound thrombin, as well as in TM456m-bound thrombin. The first conclusion one can draw from these results is that in apo-thrombin, the N-terminus of the heavy chain must not be properly positioned. This terminal segment may be entering and exiting the Ile cleft, or it may be mostly exposed, as indicated by the high level of deuterium incorporation observed for residues 37-44 (16-23_{CT}) of apo-thrombin (Figure 4.4D and 4.5D). No crystal structures of thrombin, including those of apo-thrombin (PDB code 1JOU)[26] show the N-terminus exposed, probably because the protein would not crystallize in this form. The exchange kinetics also suggest that the burial of the N-terminus may be a slow event, consistent with the fact that we were able to observe this insertion event in trypsin only in aMD simulations and not in MD simulations.[6] Like thrombin, trypsin also preferentially crystallizes with the N-terminus of the heavy chain buried (rcsb.org). Once inserted, it seems the N-terminus is highly stabilized; unsuccessful attempts have been made to observe the N-terminus exiting the Ile cleft for both α-thrombin and trypsin in aMD simulations (Markwick, PRL, unpublished observations).[6] Taken together, our results indicate that the N-terminus of the heavy chain is not properly positioned in the Ile cleft in apo-thrombin, contrary to what is

observed in thrombin crystal stuctures. It will be interesting to see whether further solution analyses, perhaps by NMR, can reveal the dynamics of this critical region.

It could be expected that covalent substrate-analog binding (like the PPACK binding explored in our study) would stabilize the N-terminus of the heavy chain in the Ile cleft, which is the catalytically active form of thrombin. However, it was wholly unexpected to observe such a phenomenon upon TM binding at ABE1, which is distant from the active site and in no way connected to it. The HDXMS results provide an additional clue as to how TM may be altering the stability of the bound N-terminus of the heavy chain. Decreases in deuterium uptake were also observed in the γ-loop upon both PPACK-binding and TM456m-binding. Analysis of several overlapping peptides allowed us to localize the difference in uptake to three residues, Gly178-Leu180, which directly abut the N-terminus of the heavy chain (Figure 4.7G). The differences in deuterium uptake were only observed at measurement times under 5 min for PPACK-thrombin, but were sustained beyond 10 min for TM456m-bound thrombin. It is likely that PPACK binding organizes the active site effectively "pulling" the N-terminus of the heavy chain into its proper location. It may be that once the N-terminus is in the Ile cleft, the γ-loop residues Gly178-Leu180 adjacent to it are somewhat stabilized. We can hypothesize that in the case of TM binding, the opposite process occurs. We previously reported that TM binding increases correlated motions in the γ-loop as observed in aMD simulations.[27] Thus, TM may stabilize the γ-loop, which, in turn may stabilize the N-terminus of the heavy chain. A conformational selection process is likely given the dynamic nature of apo-thrombin.[23; 28]

The TM-induced allosteric stabilization of the N-terminus into the Ile cleft provides a mechanistic explanation for the dramatic increase in k_a for protein C binding and k_{cat} for protein C cleavage.[8; 13] Decreases in exchange in the N-terminus of coagulation factor VIIa were also observed upon Tissue Factor (TF) binding at ABE2.[29; 30] Tissue factor also caused decreased amide exchange throughout the factor VIIa molecule but the changes that could be correlated with the activity enhancement upon TF binding were the burial of the N-terminus of the heavy chain and stabilization of the 170s loop. It is interesting that TM binding to thrombin causes only very modest stabilization of the 170s loop whereas PPACK binding dramatically stabilizes the 170s loop (Figure 4.4B and Figures 4.2 and 4.3). Our results suggest that two separable conformationally selected stabilizations must occur for maximal coagulation protease activity; insertion of the N-terminus of the heavy chain and 170s loop stabilization. It is fascinating to surmise that fibrinogen, which binds to both ABE1 and the active site, can conformationally select the active form of thrombin since its interaction with ABE1 would stabilize the N-terminus into the Ile cleft (similar to TM binding) and its interaction with the active site would stabilize the 170s loop (similar to PPACK binding). In contrast, since protein C only interacts with the active site, it requires TM to interact with ABE1 and select for the "N-terminus buried within the Ile cleft" conformation.

E. References

1. Furie, B. & Furie, B. C. (1988). *Cell* **53**, 505–518.

2. Bode, W. (2006). Structure and interaction modes of thrombin. *Blood Cells, Molecules, and Diseases* **36**, 122-130.

3. Di Cera, E. (2008). Thrombin. *Mol Aspects Med* **29**, 203-254.

4. Bode, W. (1979). The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding: II. The binding of the pancreatic trypsin inhibitor and of isoleucine-valine and of sequentially related peptides to trypsinogen and to p-guanidinobenzoate trypsinogen. *J Mol Biol* **127**, 357–374.

5. Huber, R. & Bode, W. (1978). Structural basis of the activation and action of trypsin. . *Acc Chem Res* **11**, 114–122.

6. Boechi, L., Pierce, L., Komives, E. A. & McCammon, J. A. (2014). Trypsinogen activation as observed in accelerated molecular dynamics simulations. *Protein Sci* **23**, 1550-1558.

7. Kurosawa, S., Stearns, D. J., Jackson, K. W. & Esmon, C. T. (1988). A 10-kDa cyanogen bromide fragment from the epidermal growth factor homology domain of rabbit thrombomodulin contains the primary thrombin binding site. *J Biol Chem* **263**, 5993-5996.

8. White, C. E., Hunter, M. J., Meininger, D. P., White, L. R. & Komives, E. A. (1995). Large-scale expression, purification and characterization of small fragments of thrombomodulin: the roles of the sixth domain and of methionine 388. *Protein Eng* **8**, 1177-1187.

9. Rezaie, A. R., Cooper, S. T., Church, F. C. & Esmon, C. T. (1995). Protein C inhibitor is a potent inhibitor of the thrombin-thrombomodulin complex. *J Biol Chem* **270**, 25336-25339.

10. Rezaie, A. R., He, X. & Esmon, C. T. (1998). Thrombomodulin increases the rate of thrombin inhibition by BPTI. *Biochemistry* **37**, 693-699.

11. Myles, T., Church, F. C., Whinna, H. C., Monard, D. & Stone, S. R. (1998). Role of thrombin anion-binding exosite-I in the formation of thrombin-serpin complexes. *J Biol Chem* **273**, 31203-31208.

12. van de Locht, A., Bode, W., Huber, R., Le Bonniec, B. F., Stone, S. R., Esmon, C. T. & Stubbs, M. T. (1997). The thrombin E192Q-BPTI complex reveals gross structural rearrangements: implications for the interaction with antithrombin and thrombomodulin. *EMBO J* **16**, 2977-84.

13. Xu, H., Bush, L. A., Pineda, A. O., Caccia, S. & Di Cera, E. (2005). Thrombomodulin changes the molecular surface of interaction and the rate of complex formation between thrombin and protein C. *J Biol Chem* **280**, 7956-61.

14. Fuentes-Prior, P., Iwanaga, Y., Huber, R., Pagila, R., Rumennik, G., Seto, M., Morser, J., Light, D. R. & Bode, W. (2000). Structural basis for the anticoagulant activity of the thrombin-thrombomodulin complex. *Nature* **404**, 518-525.

15. Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R. & Hofsteenge, J. (1989). The refined 1.9 A crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J* **8**, 3467-3475.

16. Nagashima, M., Lundh, E., Leonard, J. C., Morser, J. & Parkinson, J. F. (1993). Alanine-scanning mutagenesis of the epidermal growth factor-like domains of human thrombomodulin identifies critical residues for its cofactor activity. *J Biol Chem* **268**, 2888-2892.

17. Ye, J., Esmon, N. L., Esmon, C. T. & Johnson, A. E. (1991). The active site of thrombin is altered upon binding to Thrombomodulin. Two distinct structural changes are detected by fluorescence, but only one correlates with protein C activation. *J Biol Chem* **266**, 23016-23021.

18. Koeppe, J. R., Seitova, A., Mather, T. & Komives, E. A. (2005). Thrombomodulin tightens the thrombin active site loops to promote protein C activation. *Biochemistry* **44**, 14784-91.

19. Treuheit, N. A., Beach, M. A. & Komives, E. A. (2011). Thermodynamic compensation upon binding to exosite 1 and the active site of thrombin. *Biochemistry* **50**, 4590-6.

20. Cooper, A. & Dryden, D. T. F. (1984). Allostery without conformational change. A plausible model. *Eur Biophys J* **11**, 103-109.

21. Long, D. & Bruschweiler, R. (2012). Structural and Entropic Allosteric Signal Transduction Strength via Correlated Motions. *Journal of Physical Chemistry Letters* **3**, 1722-1726.

22. Tzeng, S. R. & Kalodimos, C. G. (2011). Protein dynamics and allostery: an NMR view. *Curr Opin Struct Biol* **21**, 62-67.

23. Fuglestad, B., Gasper, P. M., Tonelli, M., McCammon, J. A., Markwick, P. R. L. & Komives, E. A. (2012). The dynamic structure of thrombin in solution. *Biophys J* **103**, 1-10.

24. Wales, T. E., Fadgen, K. E., Gerhardt, G. C. & Engen, J. R. (2008). High-speed and high-resolution UPLC separation at zero degrees Celsius. *Anal Chem* **80**, 6815-20.

25. Mandell, J. G., Baerga-Ortiz, A., Akashi, S., Takio, K. & Komives, E. A. (2001). Solvent accessibility of the thrombin-thrombomodulin interface. *J Mol Biol* **306**, 575-589.

26. Huntington, J. A. & Esmon, C. T. (2003). The molecular basis of thrombin allostery revealed by a 1.8 A structure of the "slow" form. *Structure* **11**, 469-479.

27. Gasper, P. M., Fuglestad, B., Komives, E. A., Markwick, P. R. & McCammon, J. A. (2012). Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proc Natl Acad Sci U S A* **109**, 21216-21222.

28. Fuglestad, B., Gasper, P. M., McCammon, J. A., Markwick, P. R. & Komives, E. A. (2013). Correlated motions and residual frustration in thrombin. *J Phys Chem B* **117**, 12857-12863.

29. Rand, K. D., Jørgensen, T. J., Olsen, O. H., Persson, E., Jensen, O. N., Stennicke, H. R. & Andersen, M. D. (2006). Allosteric Activation of Coagulation Factor VIIa Visualized by Hydrogen Exchange. *J Biol Chem* **281**, 23018–23024,.

30. Rand, K. D., Andersen, M. D., Olsen, O. H., Jørgensen, T. J., Ostergaard, H., Jensen, O. N., Stennicke, H. R. & Persson, E. (2008). The origins of enhanced activity in factor VIIa analogs and the interplay between key allosteric sites revealed by hydrogen exchange mass spectrometry. *J Biol Chem* **283**, 13378-13387.

Chapter V

Temporally segregated dynamics of functional regions in thrombin.

A. Introduction

Serine proteases are found ubiquitously in both eukaryotes and prokaryotes, and they comprise the largest of all of the peptidase families with 27,328 sequences in the MEROPS database and 109 structures in the protein data bank.[1] Their critical functions include intestinal digestion (trypsin and chymotrypsin), IgA-mediated immune response (tryptase, chymase, and the granzymes), control of dorsal-ventral signaling pathways of *Drosophila*, blood coagulation, and complement activation. All serine proteases share a Ser, His, Asp catalytic triad and greek-key or double β-barrel structure held together by disulfide bridges.[2] Sequence alignments with trypsin and chymotrypsin revealed that the coagulation proteases have non-conserved insertions at the base of several surface loops.[3] It has been suggested that these longer loops are responsible for an increase in substrate specificity; in the case of the coagulation proteases, which cleave after arginine, the longer loops might allow each coagulation protease to have stringent specificity towards the few protein substrates that each protease cleaves.

The terminal protease in the coagulation cascade, human α-thrombin,[4; 5] cleaves twelve known substrates causing fibrin clot formation, platelet activation, and thrombomodulin (TM)-mediated feedback inhibition of coagulation via inactivation of cofactors Va and VIIIa through activation of protein C. The allosteric regulation of thrombin specificity by its five effector molecules and towards its many substrates is not yet understood.[6; 7] We recently completed computational studies which revealed that the motion of the thrombin loops are largely uncorrelated (loops corresponding to

different "communities" are colored differently in Fig. 5.1A). Binding of TM to anion-binding exosite 1 (ABE1) caused the motion of several of the different "communities" of residues to apparently become correlated and this effect extended from the TM binding site into the active site loops.[8; 9] These computational studies suggest that effector and/or substrate binding alters the conformational ensemble of thrombin. However, NMR experiments, which could directly reveal such re-distributions of states,[10] have been slow to emerge due to the difficulties of preparing thrombin samples for NMR study.[11; 12]

Many x-ray crystallographic structures of thrombin are available and these reveal thrombin surface loops "trapped" in different conformations.[13] However, comparison of the core structure of thrombin with different effectors and/or with different substrates (or no substrate) bound, do not reveal significant conformational changes (rcsb.org). Hydrogen-deuterium exchange followed by mass spectrometry showed rapid exchange of many of the backbone amides in thrombin[14] and NMR experiments showed missing resonances[11] both suggesting that thrombin may be dynamic.

We recently characterized the dynamics of active site-liganded thrombin (D-Phe-Pro-Arg chloromethylketone (PPACK)-thrombin) using a combination of NMR experiments and state-of-the-art molecular dynamics simulations.[12] Dynamics on the ps-ns time scale were determined from R_1, and R_2 relaxation rates and ^{15}N-$\{^1H\}$NOE measurements analyzed by the Model-free approach. Measurements of R_{ex} using the TROSY-Hahn-Echo experiment revealed structural fluctuations on the 100 μs to ms timescales. Conventional molecular dynamics simulations accurately recapitulated the

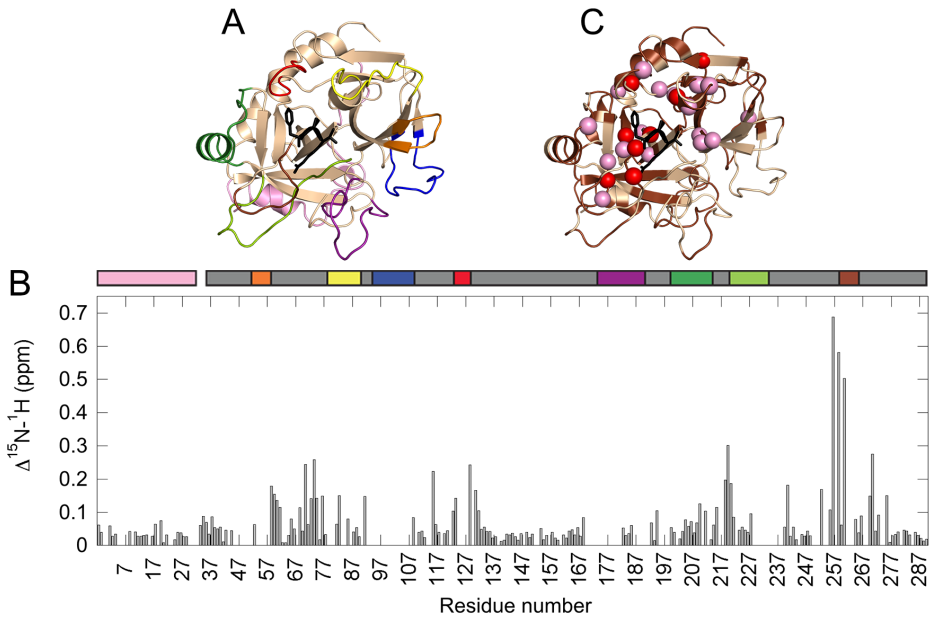


Figure 5.1. Weighted average chemical shift differences between apo-thrombin and PPACK-thrombin. (A) A crystal structure of thrombin (1PPB) is colored to highlight important communities of residues. (The light-chain is colored pink, the 30s loop is colored orange, the 60s loop is colored yellow, the 70s loop is colored blue, the 90s loop is colored red, the γ-loop is colored purple, the 170s loop is colored green, the 180s loop is colored light green, the Na^+-binding loop is colored brown. PPACK is shown in black sticks. (B) A plot showing weighted average chemical shift differences between apo-thrombin and PPACK-thrombin, with the important loops from (A) shown in schematic above. The weighted average chemical shift difference for each residue was calculated using the following equation: $\left([(\Delta\delta_{HN})^2 + (\Delta\delta_{NH})^2/25]/2\right)^{1/2}$, where $\Delta\delta_{HN}$ and $\Delta\delta_{NH}$ are the chemical shift differences in the amide ^1H and ^{15}N, respectively. (C) A visualization of the chemical shift differences plotted in B, with residues with chemical shift differences more than 1 standard deviation above the average shown in pink balls, and residues with more than 2 standard deviations above the average shown in red balls. Residues with data missing from either PPACK-thrombin and/or apo-thrombin are colored in tan.

ps-ns motions, and accelerated molecular dynamics (AMD) simulations predicted motions on the μs timescale. The AMD results revealed a structural ensemble in which several surface loops on thrombin remain surprisingly dynamic, even in the active-site liganded form.

Here we present the complete backbone dynamics profile of apo-thrombin as mimicked by the S195M active site mutant form. This mutation renders the enzyme incapable of autolytic cleavage. The results of our study demonstrate that much of apo-thrombin exhibits motions on the μs-ms time scale, which are largely dampened upon PPACK-binding. Thus, as was seen for dihydrofolate reductase, although the structural differences between the substrate-bound and apo state are very small, substrate binding substantially dampens protein dynamics.[15] As was also observed for the enzyme, PIN1,[16] the relaxation dispersion curves could not be globally-fit with a single two-state model, most likely reflecting the complex protease mechanism as well as the many sub-states of apo-thrombin that must be selected by binding of various allosteric regulators.

B. Materials and Methods

1. Expression and purification of S195M-thrombin

Details of the protein expression, refolding, and purification are as described previously,[12; 17] with a few modifications. Due to the lack of autoproteolytic activity of the S195M-thrombin mutant, additional steps were required for activation of the mutant to the α-thrombin form. In brief, the S195M mutation was achieved by quick-

change mutagenesis in pET23(+) containing the thrombin sequence with an 18 amino acid extension on the N-terminus (prethrombin-2(+18)) that is necessary for folding. $^{2}H,^{15}N$- or $^{2}H,^{15}N,^{13}C$-labeled S195M-thrombin was expressed in *E. coli* in minimal media containing the appropriate combinations of $^{2}H_2O$, $^{15}NH_4Cl$, and ^{13}C-D-glucose. Protein inclusion bodies containing S195M-thrombin were isolated, and the protein was refolded and purified as previously described.[12] Prethrombin-2(+18) was activated using 1 mg of *Echis carinatus* venom (Sigma-Aldrich) to cleave between the light and heavy chains and 700 μg of wild type α-thrombin (prepared previously and frozen at 4°C in 350 μg/mL aliquots) was added to remove the N-terminal 18 amino acids of prothrombin. After activation (typically 18-24 hrs), the wild type α-thrombin was inhibited with biotinyl-PPACK (Haematologic Technologies) followed by addition of streptavidin resin (Thermo Scientific), and the biotinyl-PPACK-α-thrombin complex was removed via centrifugation. The α-S195M-thrombin was purified by MonoS ion exchange chromatography. Proper proteolytic activation and isotope incorporation were confirmed by MALDI-TOF. Samples were buffer exchanged into NMR buffer: 25 mM sodium phosphate pH 6.5, 150 mM sodium chloride, and 0.05% sodium azide, with 10% v/v D_2O added as a lock solvent. The final protein concentration in NMR samples was 0.15 mM.

2. NMR resonance assignments and dynamics measurements

All NMR experiments were performed at 298K on spectrometers equipped with cryogenic probes. Details of the experimental procedures for resonance assignments are outlined in Fuglestad et al.[12] Experiments performed for resonance assignment were: HNCO, and HN(CO)CA at UCSD Pharmacology on a Bruker

Avance III 600 MHz, TROSY-HN(CA)CO at NMRFAM with a Varian NMR system 600, TROSY-HN(CO)CACB and TROSY-HNCA at NMRFAM with a Varian VNS 800, and NOESY-^{1}H,^{15}N-TROSY with the UCSD Chemistry and Biochemistry Varian 800. Some assignments were transferred from the previously assigned PPACK-thrombin[12] and additional assignments were made. Assignment transfers were confirmed with the 3D experimental data.

Forty residues for which N-H peaks were visible in PPACK-bound thrombin did not have visible peaks in the TROSY spectrum of apo-thrombin. These included E13(21) of the light chain; V17(38) which is adjacent to the N-terminus of the heavy chain; Q30(51) at the b-sheet leading to the 30s loop; E61(92) of the 60s loop; S83(115) and K87(119) of the strand connecting ABE1 and the 90s loop; E97a-D100(130-133) of the 90s loop and the catalytic aspartic acid, D102(135); T139-G140(175-176), G142(178), K145-A149A(181-186) and Q151(192) of the γ-loop; L160(201) and T177(218) on either side of the 170s loop; C182(223) of a disulfide bridge; G188-E192(234-238) of the 180s loop; G196(242) next to the catalytic serine; Y208-Q209(256-257); the active site adjacent W215(263) in the β-strand preceding the Na^{+}-binding loop; C220-R221(267-269) of the Na^{+}-binding loop; and Y225-G226(273-274), F232(280), and I238(286) of the C-terminus helix.

To calculate chemical shift differences, a weighted average approach was used to combine the ^{15}N and ^{1}H differences, as described previously.[18] Residues with a weighted average chemical shift difference over 1 standard deviation (> 0.11 ppm) were C42-A44(64-66) and T54(76) in between the 30s and 60s loops; the α-helix A56-C58(78-80) which includes the catalytic triad residue His57; L60(82) and K60f(88) of

the 60s loop; V66(97) at the base of the 70s loop; Y89(121) in the β-strand connecting the 70s and 90s loops; R97(129) of the 90s loop; R101(134) and I103(136) following the 90s loop; I174(215) of the 170s loop; A183-Y184a(224-226) at the base of the 180s loop; V200(246) and M210(258) of the C-terminal β-barrel; S214(262), G216(264), and G219(266) of the Na^+-binding loop; and F227-T228(275-276) and R233(281) near the base of the C-terminal helix.

NMR dynamics experiments for the calculation of order parameters (R_1, R_2, and ^{15}N-{^1H}NOE experiments) were performed at UCSD Pharmacology on a Bruker Avance III 600 MHz and analyzed as described previously.[12] For comparisons between the apo- and PPACK-bound thrombin, a consistent set of "rigid" residues was selected for the R_2/R_1 analysis using TENSOR2[19] with a random snapshot from the MD simulation performed on apo-wild-type thrombin used as the structural model to fit the relaxation data to a rotational diffusion model. As with the previous analysis on PPACK-thrombin, no significant differences were observed for the isotropic vs. anisotropic diffusion model. The smaller set of residues selected for τ_c determination yielded a slightly larger τ_c than that previously reported.[12]

3. Measurement of R_{ex} by TROSY-Hahn Echo

To determine which residues in apo-thrombin were undergoing R_{ex}, a TROSY Hahn-Echo experiment was performed on the UCSD Chemistry and Biochemistry Varian 800 as described previously.[20; 21] We previously reported 15 residues with R_{ex} in PPACK-thrombin using an R_{ex} of > 6 s^{-1} as a cut-off for significance.[12] The 37 apo-thrombin residues showed significant R_{ex} using the same cut-off were: G1d(5) and L14f(28) of the light-chain; E18(39), A22(43), E23(44) and S27(48) of the N-terminus

of the heavy chain; C42(64) and S45(67) in the N-terminal β-barrel that forms an interface with the C-terminal β-barrel; the α-helix H57-L59(79-81) which includes catalytic triad His57; K60f(88) within the 60s loop; V66(97) and I82(114) on either side of the 70s loop; I88-Y89(120-121) in the β-strand connecting the 70s and 90s loops; R97(129) of the 90s loop; the Asp of the catalytic triad, D102(135); V112(145) in a loop adjacent to both the light chain and the 70s loop; V149c(188) and K149e(190) of the γ-loop; N159-L160(200-201), F181(222), and T229(277) of the C-terminal β-barrel; C168(209), T172(213), and I176(217) of the 170s helix and loop; D186a-E186b(229-230) and R187(233) of the 180s loop; E217-G219(265-266) and D222-G223(270-271) of the Na^+ binding loop; and R233(281) and I242(290) of the C-terminal helix.

4. R_2 Relaxation Dispersion Experiments

The effective relaxation rates ($R_{2,eff}$) due to contributions from conformational exchange on intermediate timescales were evaluated with Carr-Purcell-Meiboom-Gill (CPMG) experiments collected at NMRFAM on a Varian VNS 600 and a Varian VNS 800 using a TROSY-CPMG pulse sequence.[22] Two-dimensional data sets with 2048 x 96 complex points, 32 scans, were acquired at $1/t_{cp}$ of 50, 100, 200, 300, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, and 2000 Hz and a constant t_{CPMG} = 40 ms with duplicates collected at $1/t_{cp}$ of 50, 200, 1600, and 2000 Hz in order to estimate errors.

^{15}N relaxation dispersion data measured at two or more frequencies is most often fit to the Richard-Carver equations,[23] which describes conformational exchange between two conformational states, A & B:

$$R_2(1/\tau_{CPMG}) = R_2^0 + \frac{1}{2}\left[k_{ex} - \frac{1}{\tau_{CPMG}}\cosh^{-1}\left[D_+\cosh(\eta_+) - D_-\cos(\eta_-)\right]\right]$$

where,

$$D_\pm = \frac{1}{2}\left[\pm 1 + \frac{\psi + 2\Delta\omega^2}{(\psi^2 + \zeta^2)^{1/2}}\right]^{1/2},$$

$$\eta_\pm = \frac{\tau_{CPMG}}{2}\left[\pm\psi + (\psi^2 + \zeta^2)^{1/2}\right]^{1/2}.$$

In these equations, τ_{CPMG} is the time between each 180° pulse in the CPMG pulse train, R_2^0 is the R_2 relaxation rate in the absence of conformational exchange, $\psi = k_{ex}^2 - \Delta\omega^2$, $\zeta = -2\Delta\omega k_{ex}(p_A - p_B)$, $\Delta\omega$ is the chemical shift difference between conformational states A and B, and p_A and p_B are the populations of conformational state A (ground state) and B (excited state), respectively, where $p_A + p_B = 1$.

Initially, the 600 MHz and 800 MHz R_2 relaxation dispersion data for all apo-thrombin residues with significant R_{ex} (36 residues) were fit globally to the Richard-Carver equation by using the software package, GLOVE.[24] The same global fitting method was attempted for PPACK-thrombin residues with significant R_{ex} (11 residues). GLOVE minimizes global and local parameters alternately, and incorporates a Monte-Carlo minimization method to allow fitting parameters to pass through local minima. The following grid space was used in the Marquardt search: $\Delta\omega$ = [100, 2500] (10 steps), k_{ex} = [5, 4000] (10 steps), p_Ap_B = [0.01, 0.1] (10 steps). Global fits yielded poor results for both the Richard-Carver equations and the Luz and Meiboom equation (which is optimized for motions exhibiting fast exchange, $k_{ex}/2 >> \Delta\omega$),[25] so, instead, subsets of residues were fit globally to the Richard-Carver equation. Analysis

converged on three groups of residues in different thrombin "communities". Group 1 contained nine residues in and near the N-terminus of thrombin: E18(39), S20(41), D21(42), A22(43), N149b(187), V149c(188), G149d(189), K149e(190), and G150(191). The k_{ex} of this group was determined to be 933 s^{-1} with a p_B of 2.6%. Group 2 contained four residues in the 170s loop: K169(210), T172(213), I174(215), and I176(217). The k_{ex} of this group was determined to be 585 s^{-1} with a p_B of 1.2%. Group 3 contained five residues from the Na^+-binding loop: G216(264), E217(265), G219(266), D222(270), and G223(271). The k_{ex} of this group was determined to be 1892 s^{-1} with a p_B of 2.1%. All final fits had χ_2/DoF < 2.0.

C. Results

1. Resonance assignment and chemical shift perturbations

Because wild type apo-thrombin undergoes autocatalytic proteolysis, the wild type protein is not suitable for NMR. To generate an apo-form of human α–thrombin suitable for NMR, we prepared the S195M mutant (hereafter referred to as apo-thrombin). As with PPACK-thrombin studied previously,[12] some backbone amides did not show peaks in the ^1H-^{15}N TROSY HSQC experiment, probably due to line broadening from intermediate time scales. All observed resonances were assigned, accounting for 76% of the backbone amides in apo-thrombin. In comparison, the previously published assignments of the S195A mutant accounted for 68% of possible backbone amide resonances.[11] As with PPACK-thrombin,[12] assignments were

accomplished at both 37°C and 25°C, since some resonances were only visible at one temperature or the other.

Forty residues for which amide peaks were visible in PPACK-bound thrombin did not have visible peaks in the 25°C TROSY spectrum of apo-thrombin. Disappearing resonances in apo-thrombin included five residues in the 90s loop, nine in the γ-loop, two around the 170s loop, five in the 180s loop, a residue next to the catalytic serine, and three in the Na^+-binding loop (residue numbers and more details are given in the Materials & Methods section). These residues were mostly in the functional loops surrounding the active site demonstrating that PPACK ligation reduces both the conformational and temporal range of dynamic motions, as suggested previously.[11] In addition, a large number of residues showed significant chemical shift differences between apo-thrombin and PPACK-thrombin. Residues with large chemical shift differences (> 0.11 ppm) (Fig. 5.1B), were both near the active site where PPACK binding occurs, but also distal to the active site, particularly in the C-terminal β-barrel and portions of the N-terminal β-barrel that contact the C-terminal β-barrel (Fig. 5.1C, residue numbers and more details are given in the Methods section). The fact that chemical shifts change throughout thrombin upon active site ligation is consistent with allosteric coupling throughout thrombin from the active site to effector binding sites.[26; 27; 28]

2. Backbone dynamics

Picosecond-nanosecond motions are reflected in amide bond vector order parameters (S^2) calculated from R_1 and R_2 relaxation rates and ^{15}N-{^1H}NOEs measured for a ^2H,^{15}N-apo-thrombin NMR sample. The Lipari-Szabo model-free

analysis,[29; 30] implemented in TENSOR2,[19] was used to calculate S^2 from the NMR relaxation data and the results were compared to those previously measured for PPACK-thrombin. The majority of the residues in apo-thrombin could be fit to model 1 (58%) or model 2 (69 %, with Models 1 and 2 combined). Few differences were observed in the ps-ns motions between apo and PPACK-thrombin, with apo-thrombin on average slightly more flexible on this timescale (Fig 5.2A).

To explore motions on the ms-ms time scale, R_{ex} was measured using TROSY Hahn-Echo NMR experiments.[20] In apo-thrombin, thirty-seven residues scattered throughout the protein were found to have significant R_{ex} (Fig. 5.2B), including residues near and in the 60s, 70s, and 90s loop (including a residue of the catalytic triad, Asp102), residues in the γ-loop, many residues in the C-terminal β-barrel, residues in the 170s, 180s, and the Na^+ binding loops, and the C-terminal helix (residue numbers and more details listed in Materials & Methods section). Previous measurements of PPACK-thrombin revealed only fifteen residues with significant R_{ex} ($> 6\ s^{-1}$) located in the light-chain, near the 70s loop, the 90s loop, and in the γ-loop.[12]

To determine the rates of ms motions in apo-thrombin, relaxation dispersion (CPMG) experiments were performed on both apo-thrombin and PPACK-thrombin at 800 MHz and 600 MHz.[31] To initially evaluate the μs-ms dynamics, we calculated the R_{ex} for each observed resonance from the CPMG data [$R_{ex} = R_2^{obs}(1/\tau_{CPMG}^{0Hz}) - R_2^{obs}(1/\tau_{CPMG}^{\infty})$]. In PPACK-thrombin, two main regions remain dynamic: the N-terminus of the heavy chain and the region in and around the γ-loop (Fig 5.3A). In the apo-thrombin, the N-terminus of the heavy chain is even more dynamic (Fig. 5.3B). Although peaks for residues in the middle of γ-loop are completely missing in

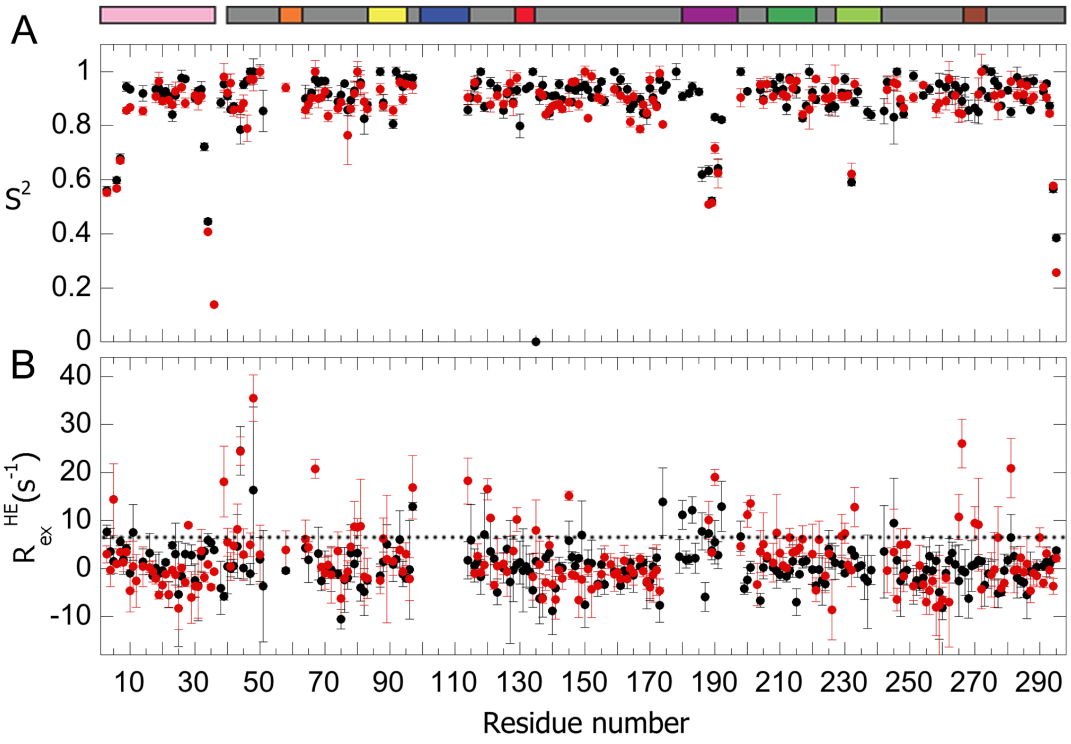


Figure 5.2. Order parameters and TROSY-Hahn Echo results for PPACK-thrombin and apo-thrombin. (A) The order parameters for PPACK-thrombin (black) and apo-thrombin (red) determined from R_1, R_2, and ^{15}N-$\{^1H\}$NOE measurements analyzed by the Model-free approach implemented in TENSOR2. (B) TROSY-Hahn Echo experimental results showing which residues have significant R_{ex}^{HE} (i.e. those with R_{ex}^{HE} above the dotted line representing 6 s^{-1}) in PPACK-thrombin (black) and apo-thrombin (red).

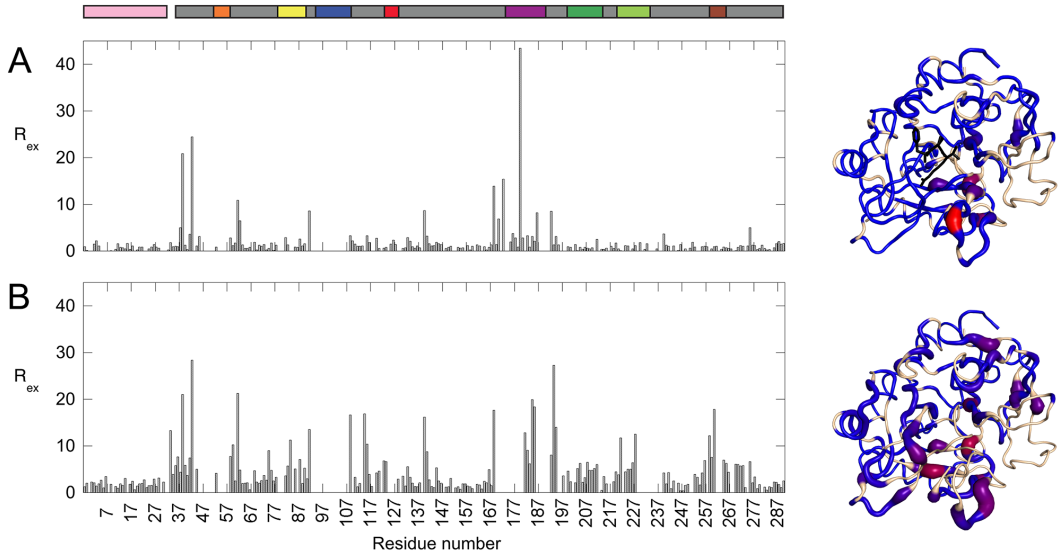
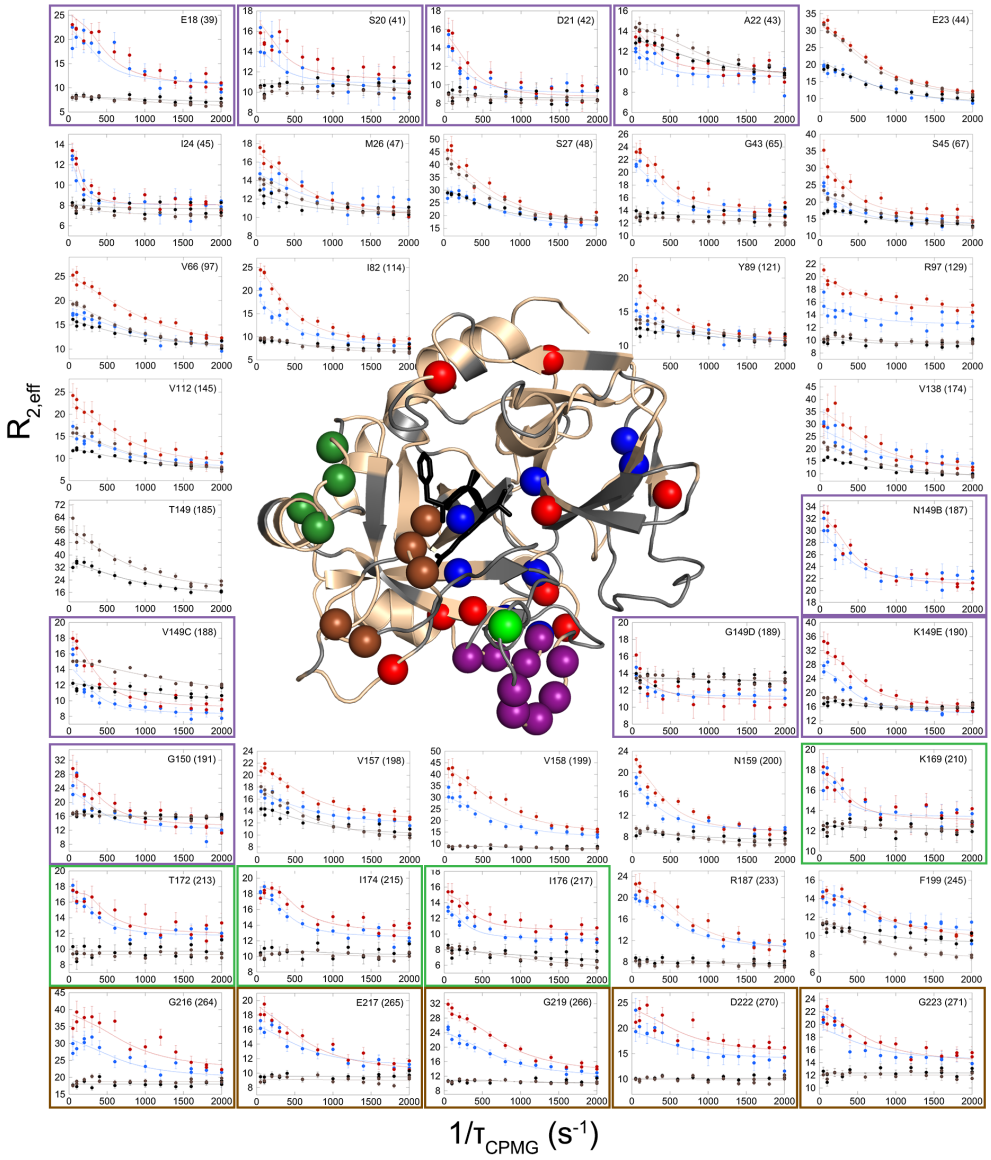


Figure 5.3. R_{ex} values for PPACK-thrombin and apo-thrombin. R_{ex} values for PPACK-thrombin and apo-thrombin are plotted by residue number in (A) and (B), respectively. On the right-hand side of (A) & (B), a sausage model of thrombin shows the degree of R_{ex} in a gradient of blue (R_{ex} = 0 s^{-1}) to red (R_{ex} = 40 s^{-1}). Residues for which CPMG NMR data was missing are colored in tan

Figure 5.4. CPMG curves for apo-thrombin and PPACK-thrombin. CPMG curves for apo-thrombin (800 MHz in red, 600 MHz in blue) and PPACK-thrombin (800 MHz in brown, 600 MHz in black). A visualization of thrombin (1PPB) at the center shows spheres for each residue for which a CPMG plot is shown. Groups of residues for which apo-thrombin CPMG data could be globally fit are shown in purple (the N-terminal region of the heavy chain), forest green (the 170s loop), and brown spheres (the Na^+-binding loop). Blue spheres represent residues for which both apo-thrombin and PPACK-thrombin have significant R_{ex} (defined as > 6 s^{-1}). Red spheres represent residues for which only apo-thrombin, not PPACK-thrombin, has significant R_{ex}. The light green sphere represents T149 (185) which is not assigned in apo-thrombin, but which shows significant R_{ex} in PPACK-thrombin. CPMG plots for the globally fit residues are boxed in the color designating their group. On all CPMG plots, individual GLOVE fits are shown unless the plot is boxed, in which case the global fit is shown for the apo-thrombin data.


E18 (39)
S20 (41)
D21 (42)
A22 (43)
E23 (44)
I24 (45)
M26 (47)
S27 (48)
G43 (65)
S45 (67)
V66 (97)
I82 (114)
Y89 (121)
R97 (129)
V112 (145)
V138 (174)
T149 (185)
N149B (187)
V149C (188)
G149D (189)
K149E (190)
G150 (191)
V157 (198)
V158 (199)
N159 (200)
K169 (210)
T172 (213)
I174 (215)
I176 (217)
R187 (233)
F199 (245)
G216 (264)
E217 (265)
G219 (266)
D222 (270)
G223 (271)
R2,eff
1/τCPMG (s-1)

apo-thrombin, residues on either side of the loop are highly dynamic. Overall, apo-thrombin exhibits greater R_{ex} on average across the molecule (Fig. 5.3A vs. Fig. 5.3B).

Thirty-six residues in apo-thrombin had significant relaxation dispersion curves indicating backbone amide motions on the μs-ms time scale (Fig. 5.4). In contrast, only eleven residues in PPACK-thrombin had significant backbone amide motions on the μs-ms time scale. One residue, Thr149(185) in the γ-loop, was not observed in apo-thrombin, but was highly mobile in PPACK-thrombin.

Initial analyses of the relaxation dispersion data obtained for apo-thrombin and for PPACK-thrombin showed a wide range of exchange rates and populations. All attempts to globally fit the data for each protein state using several different fitting programs including GLOVE[24] were unsuccessful. Attempts were also made to cluster residues based on whether they showed significant chemical shift differences in the PPACK-bound form or the TM-bound form (unpublished data, Handley and Komives), but the relaxation data for these clusters of residues also did not fit a two-state model. Finally, attempts were made to globally fit residues in apo-thrombin based on the community network analysis we previously reported (see Fig. 5.1).[8] Residues within three communities could be fit together globally (details in Materials & Methods). One group contained nine residues in and near the N-terminus of thrombin (purple spheres in Fig. 5.4): E18(39), S20(41), D21(42), A22(43), N149b(187), V149c(188), G149d(189), K149e(190), and G150(191). The k_{ex} of this group was determined to be 933 s^{-1} with a p_B of 2.6%. The second group contained four residues in the 170s loop (forest green spheres in Fig. 5.4): K169(210), T172(213), I174(215), and I176(217). The k_{ex} of this group was determined to be 585

s^{-1} with a p_B of 1.2%. The third group contained five residues in the Na^+-binding loop (brown spheres in Fig. 5.4): G216(264), E217(265), G219(266), D222(270), and G223(271). The k_{ex} of this group was determined to be 1892 s^{-1} with a p_B of 2.1%. The remaining residues exhibiting significant R_{ex} in apo-thrombin, along with all the residues in PPACK-thrombin, were unable to be globally fit to a two-state model.

D. Discussion

The backbone dynamics of thrombin in the substrate-free apo-thrombin and the substrate-bound PPACK-thrombin represent the first NMR study of backbone dynamics in a serine protease. Indeed, they represent the first such study of a secreted, disulfide-bonded enzyme. It is tempting to speculate that the wide distribution of residues showing μs-ms motions throughout thrombin may be possible due to the different way in which disulfide bonds stabilize the folded structure, as in basic pancreatic trypsin inhibitor,[32] as opposed to intracellular proteins which must be stabilized by secondary structure. Such speculations require more studies of other disulfide-bonded proteins.

Substrate binding is known to dramatically stabilize thrombin towards thermal denaturation.[33] Comparison of the dynamics of substrate-bound to substrate-free thrombin revealed that substrate binding dramatically dampens μs-ms motions of the loops surrounding the active site while leaving the ps-ns motions essentially unchanged. Similar observations have been made for several other enzymes.[10; 15; 16; 34; 35; 36]

The large number of residues in the active site loops moving on the μs-ms timescale appears much larger than has been observed in other enzymes, which typically have a single loop that closes over the active site (cf references cited earlier). Residue insertions that are not present in trypsin and chymotrypsin are found at the base of the most dynamic loops,[3] which are also sites of considerable residual local frustration[9] suggesting evolutionary pressure for a wide range of dynamic motions in these loops. The active site loops retain much of the ps-ns motion even after substrate binding,[12] but motions in μs-ms timescale are largely quenched, suggesting that μs-ms motions may be involved in recognition of the various thrombin substrates.[37]

Interestingly, PPACK-thrombin retains a cluster of core residues with motions in the μs-ms time scale, and these appear to connect a major allosteric site, ABE1, with the active site (Fig. 5.4). This result strongly suggests that the motions correlated with function and allostery in thrombin, occur in the μs-ms time regime. Despite the clear clustering of these residues, however, their motions could not be fit to a two-state model hinting that even in the substrate-bound form, thrombin can sense binding of different effectors at ABE1.

The observation that apo-thrombin has a very large number of backbone amide groups showing μs-ms motions, which did not fit to a global two-state motional model, was difficult to interpret. We attempted several different approaches to the analysis, including grouping residues based on chemical shift perturbation upon substrate binding and thrombomodulin binding, as well as grouping based on populations and exchange rates from individual fits. None of these approaches yielded statistically significant global fitting results. Finally, based on our previous community

network analysis of AMD simulation results, we attempted to globally fit groups of residues based on community. It is important to note that the relaxation dispersion results, which seem to indicate uncorrelated motions in apo-thrombin are consistent with our previous community network analysis of apo-thrombin, which resulted in many different "communities" of residues with uncorrelated motions.[8]

Using the community analysis as a guide, we were able to uncover three groups of residues with motions that could be described as two-state. Two of these groups are regions of thrombin previously implicated in substrate specificity. The 170s loop, which is highly frustrated,[9] showed μs-ms dynamics in apo-thrombin that were quenched in PPACK-thrombin. We can speculate that the 170s loop of apo-thrombin may be exchanging between an inactive conformation and a conformation that promotes proteolytic activity and that substrate binding selects the more stable, proteolytically active conformation. Indeed, stabilization of the 170s loop enhances activity in Factor VIIa.[38] Similarly, the Na^+-binding loop, also called the substrate specificity loop, was observed to undergo ms timescale motions that could be fit by a two-state exchange model. The Na^+-binding loop has also previously been implicated in defining substrate specificity of thrombin and in other blood coagulation proteases.[39; 40]

The region around the N-terminus of the heavy chain of apo-thrombin, including residues from the adjacent γ-loop, exhibits strong μs-ms motions that as a group could be described by a two-state exchange model. Serine proteases are all activated by proteolysis of the zymogen at an internal peptide bond, exposing a new N-terminal Ile-NH_3^+ which subsequently fits into a specific pocket called the "Ile

cleft", forming a conserved salt bridge with an aspartic acid residue, correctly assembling the S1 pocket for proteolytic activity.[41; 42] Although all crystal structures of thrombin (both apo-thrombin and substrate-bound) show the N-terminus of the heavy chain already buried in this Ile cleft, our recent HDXMS studies showed that this region in apo-thrombin has very little deuterium protection indicating it must be very dynamic (Handley 2015, submitted). The relaxation dispersion data indicated that this region was not dynamic In PPACK-thrombin. However, in apo-thrombin, the μs-ms motions of the N-terminus of the heavy chain could be globally fit with those of the resonances from γ-loop using a two-state model. In the HDXMS study, we showed that binding of thrombomodulin causes decreased exchange in both the N-terminus of the heavy chain as well as in the γ-loop. The results presented here strongly suggest that even in apo-thrombin, the γ-loop motions are coupled to those of the N-terminus and that substrate binding results in allosteric conformational selection of the catalytically active form of thrombin.

The complete NMR dynamics profile we present here reveals that apo-thrombin exists as a broad ensemble of malleable states that move on multiple time scales. The widespread dynamic motions could result from the architecture of thrombin, with many and larger surface loops protruding from the protease core, and perhaps also to the disulfide bonds, which stabilize the protein without the need for a tightly packed core. Nearly 12% of the thrombin residues show motions in the μs-ms time regime, and the motions of these residues only fit to two-state kinetic models when separated into the same "communities" previously discovered by MD simulations.[8] The network of temporally correlated residues in apo-thrombin, which

spans the entire molecule, indicates that the information transfer networks, or allosteric pathways, are inherently built into the thrombin molecule. Our results suggest that the specific functional activity of thrombin may occur via a conformational selection mechanism from a broad and highly dynamic apo state as a result of allosteric effector binding. It will be interesting to see whether other serine proteases have similarly dynamic apo states from which specific catalytic forms can be selected by effector binding.

Through our HDXMS and NMR dynamics studies of thrombin, a new picture of thrombin allostery is emerging. Allosteric N-terminal insertion by TM appears to be the most compelling mechanism by which TM increases the k_{cat} of thrombin towards protein C 1000-fold, as N-terminal insertion is known to be a conserved activating mechanism in serine proteases. The dynamic nature of the N-terminus of the heavy chain in apo-thrombin has never before been seen, and previously the N-terminus was assumed to be buried, even in the apo-form. Furthermore, thrombin exhibits significant quenching of μs-ms timescale dynamics when substrates bind; this quenching of dynamics was evident in the same regions in both our HDXMS and NMR studies. Taken together, our dynamics studies of thrombin suggest apo-thrombin exists on a broad conformational landscape, likely sampling several different effector-bound or substrate-bound conformations. When one of the 5 known effectors or 12 known substrates of thrombin binds to the protease, a narrower set of conformational ensembles is selected; for example, when TM binds, a conformational ensemble is selected that positions the N-terminus of the heavy chain within the Ile cleft. This conformational selection by effector molecules hypothesis is reminiscent of the well-

studied conformational selection mechanisms in the dihydrofolate reductase system.{Boehr, 2006 #49} It will be interesting to see in future dynamics studies of thrombin what conformational ensembles other thrombin effector molecules select from this diverse conformational landscape.

E. References

1. Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. (2014). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **42**, D503-D509.

2. Paige, M. J. & Di Cera, E. (2010). Combinatorial enzyme design probes allostery and cooperativity in the trypsin fold. *J Mol Biol* **399**, 306-319.

3. Suel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Mol Biol* **10**, 59-69.

4. Macfarlane, R. G. (1964). An enzyme cascade in the blood clotting mechanism, and its function as a biochemical amplifier. *Nature* **202**, 498–499.

5. Davie, E. W. & Ratnoff, O. D. (1964). Waterfall sequence for intrinsic blood clotting. *Science* **145**, 1310–1312.

6. Huntington, J. A. (2005). Molecular recognition mechanisms of thrombin. *J Thromb Haemost* **3**, 1861–1872.

7. Davie, E. W. & Kulman, J. D. (2006). An overview of the structure and function of thrombin. *Semin Thromb Hemost* **32 (Suppl 1)**, 3–15.

8. Gasper, P. M., Fuglestad, B., Komives, E. A., Markwick, P. R. & McCammon, J. A. (2012). Allosteric Networks in Thrombin Distinguish Procoagulant vs. Anticoagulant Activities. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21216-21222.

9. Fuglestad, B., Gasper, P. M., McCammon, J. A., Markwick, P. R. & Komives, E. A. (2013). Correlated motions and residual frustration in thrombin. *J Phys Chem B.* **117**, 12857-63.

10. Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M. & Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913-6. .

11. Lechtenberg, B. C., Johnson, D. J., Freund, S. M. & Huntington, J. A. (2010). NMR Resonance Assignments of Thrombin Reveal the Conformational and Dynamic Effects of Ligation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14087-14092.

12. Fuglestad, B., Gasper, P. M., Tonelli, M., McCammon, J. A., Markwick, P. R. L. & Komives, E. A. (2012). The Dynamic Structure of Thrombin in Solution. *Biophys. J.* **103**, 1-10.

13. Niu, W., Chen, Z., Gandhi, P. S., Vogt, A. D., Pozzi, N., Pelc, L. A., Zapata, F. & Di Cera, E. (2011). Crystallographic and kinetic evidence of allostery in a trypsin-like protease. *Biochemistry* **50**, 6301–6307.

14. Koeppe, J. R., Seitova, A., Mather, T. & Komives, E. A. (2005). Thrombomodulin tightens the thrombin active site loops to promote protein C activation. *Biochemistry* **44**, 14784-14791.

15. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. (2010). Millisecond timescale fluctuations in dihydrofolate reductase are exquisitely sensitive to the bound ligands. *Proc Natl Acad Sci U S A* **107**, 1373-8.

16. Labeikovsky, W., Eisenmesser, E. Z., Bosco, D. A. & Kern, D. (2007). Structure and dynamics of pin1 during catalysis by NMR. *J Mol Biol* **367**, 1370-1381.

17. Johnson, D. J., Adams, T. E., Li, W. & Huntington, J. A. (2005). Crystal structure of wild-type human thrombin in the Na+-free state. *Biochem J* **392**, 21-8.

18. Grzesiek, S., Stahl, S. J., Wingfield, P. T. & Bax, A. (1996). The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR. *Biochemistry* **35**, 10256-10261.

19. Dosset, P., Hus, J. C., Blackledge, M. & Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. . *J Biomol NMR* **16**, 23-28.

20. Wang, C., Rance, M. & Palmer, A. G., III (2003). Mapping chemical exchange in proteins with MW > 50 kD. *J Am Chem Soc* **125**, 8968–8969.

21. Wang, Y., Berlow, R. B. & Loria, J. P. (2009). Role of loop-loop interactions in coordinating motions and enzymatic function in triosephosphate isomerase. *Biochemistry* **48**, 4548 - 4556.

22. Loria, J. P., Rance, M. & Palmer, A. G., III (1999). Transverse-relaxation-optimized (TROSY) gradient-enhanced triple-resonance NMR spectroscopy. *J Magn Reson* **141**, 180-4.

23. Carver, J. P. & Richards, R. E. (1972). A general two-site solution for the chemical exchange produced dependence of T2 upon the carr-Purcell pulse separation. *J Magn Reson* **6**, 89-105.

24. Sugase, K., Konuma, T., Lansing, J. C. & Wright, P. E. (2013). Fast and accurate fitting of relaxation dispersion data using the flexible software package GLOVE. *J Biomol NMR* **56**, 275-283.

25. Luz, Z. & Meiboom, S. (1963). Nuclear magnetic resonance study of protolysis of trimethylammonium ion in aqueous solution – order of reaction with respect to solvent. *J Chem Phys* **39**, 366–370.

26. Treuheit, N. A., Beach, M. A. & Komives, E. A. (2011). Thermodynamic compensation upon binding to exosite 1 and the active site of thrombin. *Biochemistry* **50**, 4590-4596.

27. Kamath, P., Huntington, J. A. & Krishnaswamy, S. (2010). Ligand binding shuttles thrombin along a continuum of zymogen- and proteinase-like states. *J Biol Chem* **285**, 28651-8. .

28. Malovichko, M. V., Sabo, T. M. & Maurer, M. C. (2013). Ligand binding to anion-binding exosites regulates conformational properties of thrombin. *J Biol Chem* **288**, 8667-78.

29. Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* **104**, 4546-4559.

30. Mandel, A. M., Akke, M. & Palmer, A. G., III (1995). Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* **246**, 144-63.

31. Loria, J. P., Rance, M. & Palmer, A. G., III (1999). A relaxation-compensated Carr–Purcell–Meiboom–Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J Am Chem Soc* **121**, 2331–2332.

32. Grey, M. J., Wang, C. & Palmer, A. G., III (2003). Disulfide bond isomerization in basic pancreatic trypsin inhibitor: multisite chemical exchange quantified by CPMG relaxation dispersion and chemical shift modeling. *J Am Chem Soc* **125**, 14324-35.

33. Croy, C. H., Koeppe, J. R., Bergqvist, S. & Komives, E. A. (2004). Allosteric changes in solvent accessibility observed in thrombin upon active site occupation. *Biochemistry* **43**, 5246-55.

34. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313**, 1638-1642.

35. Kern, D. & Zuiderweg, E. R. P. (2003). The role of dynamics in allosteric regulation. *Curr Op Struct Biol* **13**, 748–757.

36. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E. & Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117-121.

37. Lange, O. F., Lakomek, N. A., Farès, C., Schröder, G. F., Walter, K. F., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C. & de Groot, B. L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. . *Science* **320**, 1471-1475.

38. Rand, K. D., Andersen, M. D., Olsen, O. H., Jørgensen, T. J., Ostergaard, H., Jensen, O. N., Stennicke, H. R. & Persson, E. (2008). The origins of enhanced activity in factor VIIa analogs and the interplay between key allosteric sites revealed by hydrogen exchange mass spectrometry. *J Biol Chem* **283**, 13378-13387.

39. Wells, C. M. & Di Cera, E. (1992). Thrombin is a Na(+)-activated enzyme. *Biochemistry* **31**, 11721-11730.

40. Griffona, N. & Di Stasio, E. (2001). Thermodynamics of Na+ binding to coagulation serine proteases. *Biophys Chem* **90**, 89–96.

41. Bode, W. (1979). The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding: II. The binding of the pancreatic trypsin inhibitor and of isoleucine-valine and of sequentially related peptides to trypsinogen and to p-guanidinobenzoate trypsinogen. *J Mol Biol* **127**, 357–374.

42. Huber, R. & Bode, W. (1978). Structural basis of the activation and action of trypsin. *Acc Chem Res* **11**, 114–122.