UNIVERSITY OF CALIFORNIA, SAN DIEGO

Exploring the Dynamics of Thrombin by NMR

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in

Chemistry

by

Brian A. Fuglestad

Committee in charge:

Professor Elizabeth A. Komives, Chair
Professor Michael Burkart
Professor Michael Gilson
Professor Patricia Jennings
Professor Joseph Noel

2013

The dissertation of Brian A. Fuglestad is approved,

and it is acceptable in quality and form for publication on

microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2013

DEDICATION

I dedicate this work to my family: Tina, Frank and June. Without their support, I would not have come this far.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABE1	Anion Binding Exosite 1
ABE2	Anion Binding Exosite 2
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
AMD	Accelerated Molecular Dynamics
CMD	Conventional Molecular Dynamics
CPMG	Carr Purcell Meiboom Gill sequence
Da	Dalton
D2O	^2H_2O, Deuterated Water
EDTA	Ethylenediaminetetraacetic Acid
FPLC	Fast Protein Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence
k_{cat}	Catalytic Rate Constant
MALDI-TOF	Matrix-Assisted Laser Desorption Ionization-Time of Flight Detection
MD	Molecular Dynamics
MWCO	Molecular Weight Cut-Off
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
PAR	Protease-Activated Receptor
PDB	Protein Data Bank
PEG	Polyethylene glycol
PPACK	D-Phe-Pro-Arg chloromethylketone

R_1	Longitudinal Relaxation
$R_{1\rho}$	Longitudinal Relaxation in the Rotating Frame
R_2	Transverse Relaxation
R_{ex}	Transverse Relaxation due to Chemical Exchange
RDC	Residual Dipolar Coupling
S^2	Squared Generalized Order Parameter
τ_c	Molecular Rotational Diffusion Correlation Time
TM	Thrombomodulin
TROSY	Transverse Relaxation Optimized Spectroscopy
WT	Wild-Type

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First off, I would like to thank my advisor Betsy who had provided me the resources, patience and freedom to pursue this project. She has been a huge part of my growth as a scientist. I would like to thank Tracy Handel, and Xuemei Huang for help with the magnets and training, especially during my early days as a spectroscopist. I am grateful to have a great set of colaborators who have really helped push the project forward: Marco Tonelli, Paul Gasper, Phineus Markwick and Andy McCammon. Without these folks we would not have made such interesting findings. It has been great working in the Komives group and I am very grateful to have colleagues who have become like a family: Mike, Carla, Muneera, Mela, Ingrid, Vera, Jorge, Pato, Ryan, Nick, Lindsey, Holly, Jesse, and Deepa. I have been incredibly fortunate to have a couple of stellar undergraduates that have really been instrumental in the lab: Pei-ling and Kyle. I would like to thank the research scholars program for the opportunity to work with Jett, who was a blast to work with and made me to learn more than I expected about mentoring and teaching. Perhaps the most important group of people I would like to thank is my family and friends. I have received an incredible amount of support and encouragement from these folks, anything from helping me out of a pinch, chatting about science, letting me vent about frustrations around the lab, or literally rescuing me when I was stranded on the other side of the world. I know I couldn't have made it through without this amazing group of family and friends that I'm extremely lucky to have.

Chapter IV, in full, is a reprint that the dissertation author was the principal researcher and author of. The material appears in *Biophysical Journal*. (**Fuglestad, B.,** Gasper, P.M., Tonelli, M., McCammon, J.A., Markwick, P.R. and Komives, E.A. (2012). The dynamic structure of thrombin in solution. *Biophys. J.,* **103**, 79-88)

VITA

2007 Bachelors of Science, Biochemistry, Oklahoma State University, Stillwater

2009 Masters of Science, Chemistry, University of California, San Diego

2013 Doctor of Philosophy, Chemistry, University of California, San Diego

PUBLICATIONS

Fuglestad, B., Gasper, P.M., Tonelli, M., McCammon, J.A., Markwick, P.R. and Komives, E.A. (2012). The dynamic structure of thrombin in solution. *Biophys. J.,* **103**, 79-88

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

Fuglestad, B., Gasper, P.M., Tonelli, M., Stearns, K.M., McCammon, J.A., Markwick, P.R. and Komives, E.A. Microsecond to millisecond dynamics of apo-thrombin and the affect of active site occupation (2013) *manuscript in preparation*.

FIELDS OF STUDY

Major Field: Biochemistry

Studies in Molecular Biophysics and Structural Biology

Professor Elizabeth A. Komives

HONORS AND AWARDS

2010-2012 American Heart Association Predoctoral Fellowship, fellow

2008-2010 NIH Molecular Biophysics Training Grant, trainee

2007 Lou Wentz Research Project Award, Oklahoma State University

ABSTRACT OF THE DISSERTATION

Exploring the Dynamics of Thrombin by NMR

by

Brian A. Fuglestad

Doctor of Philosophy in Chemistry

University of California, San Diego, 2013

Professor Elizabeth A. Komives, Chair

Thrombin is a tightly regulated serine protease that acts at the terminus of the blood coagulation cascade, performing proteolytic activation of platelets and converting soluble fibrinogen into insoluble fibrin. In complex with thrombomodulin, thrombin acts as an anticoagulant by activating protein C. The strict regulation of thrombin is essential to organism survival, as under-regulation of clot formation results in amplified blood loss and over-regulation leads to pathological clot formation. In human health, hemophilia, pulmonary emboli, venous thrombi, myocardial infarctions and ischemic strokes are all tied to the misregulation of thrombin. Structurally, thrombin consists of a double β-barrel core that is conserved among members of the trypsin-like protease family with extended active site loops that are not conserved. Recent evidence has suggested that the regulation of thrombin is tied much more to dynamics than previously thought and NMR is the ideal method of study.

Chapter II outlines the process of isotopic labeling, expression, refolding,

activation and purification of recombinant thrombin from *E. coli*. Once this barrier was overcome, the next hurdle in the study of thrombin by NMR was resonance assignments, discussed in Chapter III. Results from resonance assignment of thrombin with PPACK occupying its active site vs. the S195M mutant representing apo-thrombin showed major chemical shift perturbations between the forms and a large degree of resonance line broadening that abrogated the signal, an effect that was mostly abolished upon active site ligation by PPACK.

In Chapter IV, using a combined NMR and computational approach (performed by P. Gasper and P. Markwick of the McCammon lab) the dynamics of PPACK-thrombin are characterized and the solution ensemble is modeled. The results indicate that even with the active site occupied by PPACK, thrombin undergoes a large degree of structural fluctuations on timescales ranging from picoseconds all the way to milliseconds. The equivalent study was undertaken on S195M-thrombin as discussed in Chapter V. As compared to PPACK-thrombin, the apo-like S195M-thrombin, apo-thrombin displays a remarkable degree of dynamics in the μs-ms timescale. These results reinforce the paradigm shift towards an ensemble view of thrombin activity states and substrate recognition.

Chapter 1

Introduction

A. Thrombin, hemostasis and the role of allostery

Hemostasis is a tightly regulated process involving the cascade of proteolytic signaling events in response to vascular injury (Davie*, et al.*, 1991). Thrombin activation represents the terminal step in the formation of a blood clot; in humans, activated thrombin (α-thrombin) activates platelets via PAR1 and PAR4 to form the bulk of the clot (Kahn*, et al.*, 1999), and cleaves the soluble fibrinogen into insoluble fibrin. Thrombin also activates factor (f)VIII which, in turn, cleaves and cross-links fibrin to form a meshwork that further stabilizes the clot. In addition to the clot-formation activity of thrombin, positive-feedback mechanisms ensure the timely creation of a hemostatic clot. The upstream factors fV, fVIII and fXI are activated by α-thrombin, accelerating the rate of thrombin activation and thus clot formation (Figure 1.1)

Halting the propagation of and breaking down clots is also essential to hemostasis and organism survival. Thrombin, the workhorse of the procoagulation cascade, also has a major role in the down-regulation of clotting. When bound to thrombomodulin (TM), thrombin activates Protein C (PC). PC in turn inactivates the upstream fVa and fVIIIa, providing a mechanism to end the propagation of the clot. The ability of thrombin to switch activities between procoagulant and anticoagulant provides for an efficient hemostatic mechanism; when thrombin reaches a critical level in the vicinity of endothelial tethered TM, the TM-thrombin complex forms and initiates the anticoagulant activity of thrombin.

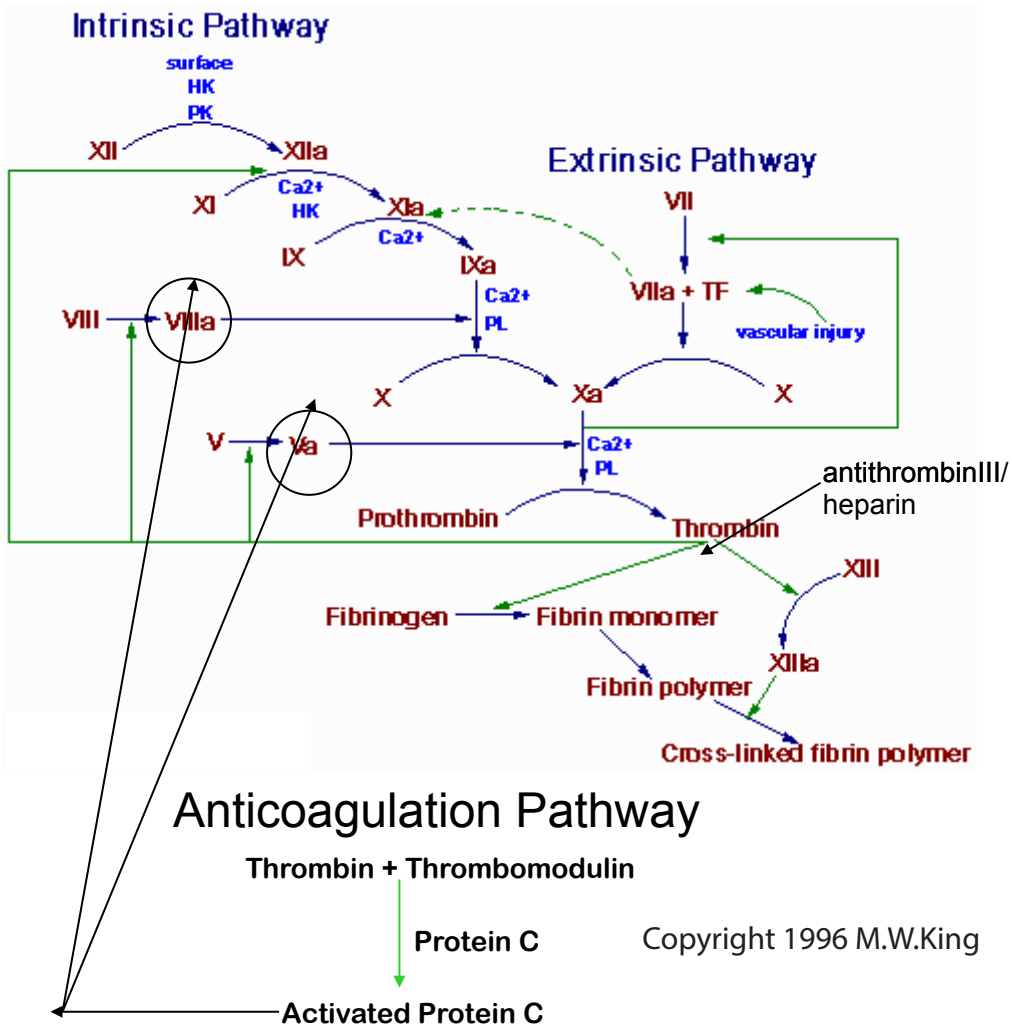


Figure 1.1 The blood coagulation cascade. With circles indicating limiting cofactors, green arrows indicating protealytic activations, and red arrows indicating inactivations. The trypsin-like core of this family of proteases consists of a double β-barrel core with surface loops extending from the β-strands (Suel, *et al.*, 2003). Compared to other members of the trypsin-like serine protease family, the active site loops in thrombin are greatly extended (Figure 1.2). The extended loop sequences in thrombin are thought to impart additional binding specificity (Paige & Di Cera, 2010). In addition, this architecture may be conducive for allostery due to the direct link of loops across the β-barrels through β-strands.

Mutagenesis studies of thrombin have distinguished the importance of several regions. Anion-binding exosite 1 (ABE1) is a cleft between the 70s and 30s surface loops that are distal to the active site. Fibrinogen, TM, and many thrombin inhibitors bind thrombin at ABE1. Alanine mutations in this region which eliminated positively charged residues severely disrupted the interaction with TM (Pineda*, et al.*, 2002). Ala mutations of Arg residues 93, 97, and 101 were shown to have decreased or no interaction with protein C (individual mutants or triple mutant) as well as a decreased interaction with heparin and the chondroitin sulfate on TM (He*, et al.*, 1997). Another site of interest on thrombin is the Na^+-binding site. This comprises residues in the 220s loop of thrombin. Mutation of several residues in this loop has led to changes in the monovalent cation specificity of thrombin from Na^+ to K^+ and to changes in the activation of protein C by thrombin in the presence of TM (Pineda*, et al.*, 2004, Pineda*, et al.*, 2004, Xu*, et al.*, 2005). Protein C activation by thrombin has further been studied by mutation of the residue Asp189 in the primary specificity pocket of thrombin. This residue is implicated in activation of protein C by thrombin in both the presence and absence of TM. When Asp189 is mutated to Ala, Asn, Gly, or Ser interaction with protein C is abolished (Prasad*, et al.*, 2004).

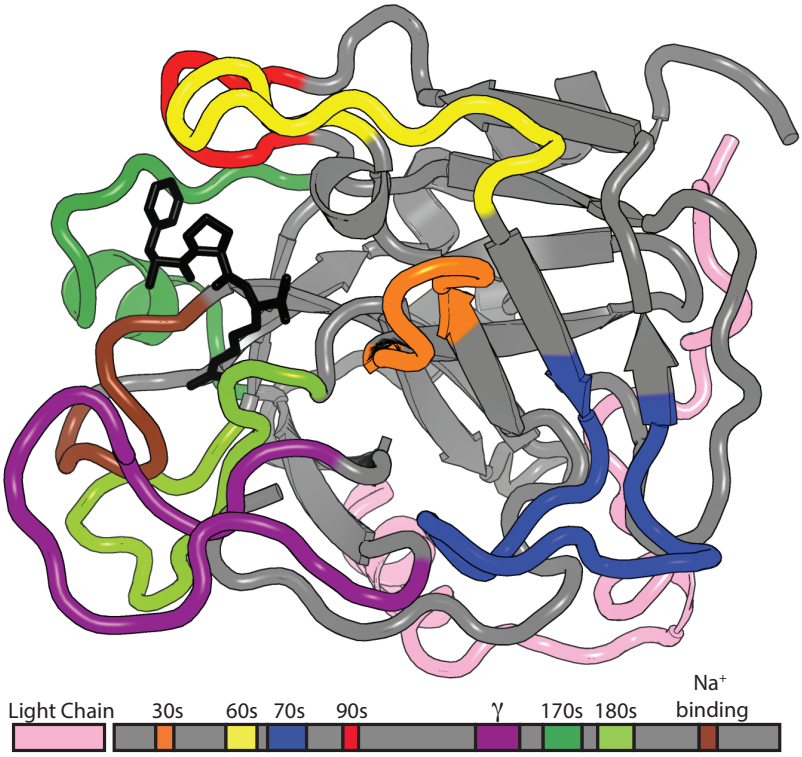


Figure 1.2 Structural representation of PPACK-thrombin from PDB 1PPB (top) and primary structure location of loops, located throughout the text as a guide for the reader. The following surface loops are labled: Light-chain (pink), 30s loop (orange), 60s loop (yellow), 70s loop (blue), 90s loop (red), γ-loop (purple), 170s loop and helix (dark green), 180s loop (light green), Na^+-binding loop (brown).

B. Thrombin dynamics in allostery

ABE1 binding affects the active site while active site binding affects ABE1. A recent study found that active site occupation in thrombin altered the thermodynamics of binding in ABE1 (Treuheit*, et al.*, 2011). The entropic penalty that needed to be overcome in order to bind ligands at ABE1 was lowered when the active site was occupied by PPACK. In addition, past studies have shown that PPACK occupation increases the thermal stability of thrombin by 20°C (Croy*, et al.*, 2004). Along with this stabilization, a decrease of amide exchange in active site distal regions was observed in ABE1 and a β-strand leading onto the γ-loop, which suggests long-range effects of active site ligation. Though thrombin allostery has been recognized as essential in regulating the clotting cascade, the exact pathway of communication across the thrombin molecule remains elusive (Suel*, et al.*, 2003, Davie & Kulman, 2006).

While the activation of protein C by the TM-thrombin complex may be partially explained by the raised affinity of protein C for the complex versus thrombin alone (Fuentes-Prior*, et al.*, 2000), the increased association rates of protein C (Xu*, et al.*, 2005) and several small thrombin inhibitors (Rezaie*, et al.*, 1995, Myles*, et al.*, 1998, Rezaie*, et al.*, 1998) suggest the role of allostery in the hemostatic regulation of thrombin activity.

C. Experiment and Simulation

The importance of allosteric regulation and the likely role of dynamics illustrates the importance of pursuing such studies. Though thrombin in various forms has been crystalized over 350 times, a thorough explanation of the mechanisms of thrombin regulation remains elusive. For example, in the case of the thrombin-thrombomodulin interaction, comparison of the crystal structures of thrombomodulin bound thrombin and apo-thrombin reveal no major differences in conformation that might explain the 1000-fold increase in k_{cat} for protein C activation when thrombomodulin is bound. While the lack of observed structural differences may be due to the presence of a covalent inhibitor in both crystal structures, dynamics may also play a role in the regulation of substrate recognition. The ideal method for studying protein dynamics and the relation to function and allsotery in solution at atomic resolution is NMR (Tzeng & Kalodimos, 2009, Lipchock & Loria, 2010, Masterson*, et al.*, 2010). Dynamics studies by NMR can probe multiple timescales of molecular motions in proteins (Figure 1.3). With the power of NMR comes some limitations. Relaxation line broadening due to motions in the μs-ms timescale often leads the total abrogation of signal and traditional measurements. Dynamic motions slower than the rotational correlation time of the molecular tumbling (~20 ns) and faster than the fast limit of chemical exchange line broadening (~50 μs) are not possible. In order to overcome these glaring blind-spots in the NMR study of thrombin, we collaborated with Phineus Markwick, and Paul Gasper of the McCammon group. This combined approach proved to be very illuminating of the dynamics in thrombin.

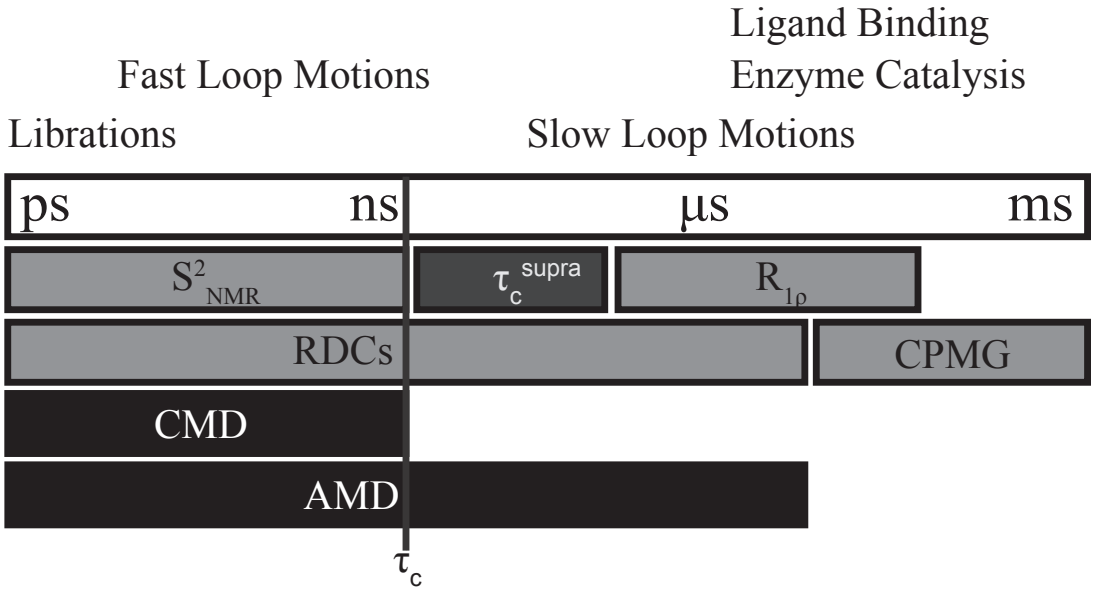


Figure 1.3 The dynamic timeline of protein motions and NMR and computational methods. In light gray are common NMR techniques and in black are two computational methods used in accessing a variety of timescales of molecular motions in proteins. In dark grey is the τ_c^{supra} blind spot in NMR that relaxation based dyanamic measurements cannot access. Above the timescale, the types of motions exhibited by proteins are depicted according to their approximate location in the timescale.

References

Croy CH, Koeppe JR, Bergqvist S & Komives EA (2004) Allosteric changes in solvent accesibility in thrombin upon active site occupation. *Biochemistry* **43**: 5246-5255.

Davie EW & Kulman JD (2006) An overview of the structure and function of thrombin. *Semin Thromb Hemost* **32 (Suppl 1)**: 3–15.

Davie EW, Fujikawa K & Kisiel W (1991) The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry.* **30**: 10363-10370.

Fuentes-Prior P, Iwanaga Y, Huber R*, et al.* (2000) Structural basis for the anticoagulant activity of the thrombin-thrombomodulin complex. *Nature* **404**: 518-525.

He X, Ye J, Esmon CT & Rezaie AR (1997) Influence of Arginines 93, 97, and 101 of thrombin to its functional specificity. *Biochemistry* **36**: 8969-8976.

Kahn ML, Nakanishi-Matsui M, Shapiro MJ, Ishihara H & Coughlin SR (1999) Protease-activated receptors 1 and 4 mediate activation of human platelets by thrombin. *J. Clin. Invest.* **103**: 879–887.

Lipchock JM & Loria JP (2010) Nanometer propagation of millisecond motions in V-type allostery. *Structure* **18**: 1596-1607.

Masterson LR, Cheng C, Yu T, Tonelli M, Kornev A, Taylor SS & Veglia G (2010) Dynamics connect substrate recognition to catalysis in protein kinase A. *Nat Chem Biol* **6**: 821-828.

Myles T, Church FC, Whinna HC, Monard D & Stone SR (1998) Role of thrombin anion-binding exosite-I in the formation of thrombin-serpin complexes. *J Biol Chem* **273**: 31203-31208.

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

Pineda AO, Cantwell AM, Bush LA, Rose T & Di Cera E (2002) The thrombin epitope recognizing thrombomodulin is a highly cooperative hot spot in exosite I. *J. Biol. Chem.* **277**: 32015-32019.

Pineda AO, Zhang E, Guinto ER, Savvides SN, Tulinsky A & Di Cera E (2004) Crystal structure of the thrombin mutant D221A/D222K: the Asp222:Arg187 ion-pair stabilizes the fast form. *Biophys Chem* **112**: 253-256.

Pineda AO, Chen ZW, Caccia S*, et al.* (2004) The anticoagulant thrombin mutant W215A/E217A has a collapsed primary specificity pocket. *J. Biol. Chem.* **279**: 39824-39828.

Prasad S, Cantwell AM, Bush LA, Shih P, Xu H & Di Cera E (2004) Residue Asp-189 controls both substrate binding and the monovalent cation specificity of thrombin. *J. Biol. Chem.* **279**: 10103-10108.

Rezaie AR, He X & Esmon CT (1998) Thrombomodulin increases the rate of thrombin inhibition by BPTI. *Biochemistry* **37**: 693-699.

Rezaie AR, Cooper ST, Church FC & Esmon CT (1995) Protein C inhibitor is a potent inhibitor of the thrombin-thrombomodulin complex. *J Biol Chem* **270**: 25336-25339.

Suel GM, Lockless SW, Wall MA & Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Mol Biol* **10**: 59-69.

Treuheit NA, Beach MA & Komives EA (2011) Thermodynamic compensation upon binding to exosite 1 and the active site of thrombin. *Biochemistry* **50**: 4590-4596.

Tzeng SR & Kalodimos CG (2009) Dynamic activation of an allosteric regulatory protein. *Nature* **462**: 368-372.

Xu H, Bush LA, Pineda AO, 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**: 7956-7961.

Xu H, Bush LA, Pineda AO, 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**: 7956-7961.

Chapter 2

Expression, Refolding and Purification of Thrombin for NMR

A. INTRODUCTION

The first step in the study of a new protein by NMR is developing a procedure to isotopically label it in milligram quantity yields. Because the signal in NMR experiments arises from differences in populations aligning with or against an external magnetic field (in cases of spin-1/2 nuclei), NMR is a famously insensitive technique due to the relatively low differences in these populations. While every single ^{1}H nucleus is NMR-active, only rare isotopes of N and C have this property and the natural isotopic abundance of NMR active heteronuclei (most commonly ^{15}N and ^{13}C) is too low to perform multi-dimensional studies of proteins. Introduction of these rare isotopes into the protein by isotopic labeling makes NMR studies possible. A high concentration is also required to obtain adequate signal, and typically a minimum of approximately 100 μM of protein is required. In the case of thrombin a common purification scheme involves isolating thrombin from blood, typically bovine blood. This technique routinely yields several milligrams of active protein, but with producing isotopically labeled proteins from bovine blood being very difficult, isotopic enrichment of NMR active nuclei is not possible using this method. *E. coli* has come to be the work-horse of protein production in the NMR field and other fields.

Using the natural protein expression machinery of *E. coli*, a cloned gene can be expressed and purified in large quantities. Isotopic labeling in *E. coli* has become straightforward; uniform ^{15}N and ^{13}C incorporation into the protein can be achieved by growth in a minimal media with single carbon and nitrogen sources (typically NH_4Cl and glucose) containing a high percentage of the desired isotope. Another isotopic enrichment that has proven invaluable to large protein NMR spectroscopists is the random fractional

deuteration of non-labile protons. This reduces the line-broadening effect due to the large number of ^{1}H nuclei surrounding the observed nucleus and is discussed further in Chapter 3. Deuterium incorporation is achieved by simply growing *E. coli* in a media dissolved in $^{2}H_2O$ as opposed to $^{1}H_2O$, however this presents an additional challenge in acclimating the *E. coli* to growing in $^{2}H_2O$.

Though *E. coli* expression presents a clear solution to the protein yield and isotopic labeling problems, it is not without its own share of issues. Human proteins are notoriously difficult to produce in a well-folded soluble form from *E. coli* expression, requiring an additional procedure to refold and solubilize these proteins for study. Additionally, correct disulfide-bonding posses another great barrier in the production of human proteins. Previous work has shown that thrombin can be refolded in very low yields with the correct disulfide bonds (DiBella*, et al.*, 1995) and further screening made refolding in higher yields possible (Soejima*, et al.*, 2001). The separation and purification of correctly disulfide-bonded proteins from misfolded isomers is also a challenge. Once these challenges were overcome, NMR on thrombin became possible for the first time.

B. MATERIALS AND METHODS

1. Protein purification and sample preparation

Isotopically labeled wild-type human thrombin (^{2}H-^{15}N or ^{2}H-^{15}N-^{13}C) was produced by a procedure modified from the previously published method (Johnson*, et al.*, 2005). The procedure involves expression of prethrombin-2(+18)) in *E. coli* from a modified pET23a+ vector containing the sequence of prethrombin-2 plus residues 310-

327 of prothrombin at the N-terminus, which we showed were critical for proper refolding (prethrombin-2(+18)).

Freshly transformed *E. coli* were acclimated to D_2O by sequential growths in M9ZB (10 ml, 12 h), M9 (10 ml, 1 h), 50% D_2O-M9 (2 ml, 1 h), and 90% D_2O-M9 (10 ml, 12 h). The 10 ml 90% D_2O-M9 culture was used to inoculate 1 L of D_2O minimal M9 media with $^{15}NH_4Cl$ (2 g/L) or $^{15}NH_4Cl$ (2 g/L) and ^{13}C-glucose (8 g/L) , the culture was grown for 24 hr to $O.D._{600}$ of 0.8 and protein production was initiated by addition of 1 mM IPTG. Cells were grown for an additional 24 hrs and pellets were frozen at -80°C. Inclusion bodies were isolated from thawed cell pellets, solubilized, purified, reduced and dialyzed as previously described (Johnson*, et al.*, 2005).

Refolding of prethrombin-2(+18)) was initiated by drop wise rapid dilution into a 50x volume of buffer containing 50 mM Tris/HCl pH 8.5, 0.6 M arginine HCl, 20 mM $CaCl_2$, 10% (v/v) glycerol, 0.2% (w/v) Brij58, 50 mM NaCl followed by rapid stirring for 3 days. The protein solution was then concentrated to 500 mL using a Millipore Pellicon 10 kDa MWCO unit, the solution was brought to a final concentration of 1 mg/ml PEG8000, then the solution was dialyzed twice with 4 L of 25 mM Tris/HCl pH 7.4, 2 mM EDTA, and 0.1% (w/v) PEG 8000.

Purification of the refolded prethrombin-2(+18)) was achieved by ion exchange chromatography of the dialyzed solution diluted 1:1 with Tris/HCl pH 7.4, 1 mg/ml PEG 8000, on a SP Sepharose Fast Flow column (size 10/10, GE Healthcare Life Sciences) with a gradient of Tris/HCl pH7.4, 1 mg/ml PEG8000 from 0.0 to 1.0 M NaCl, at a flow rate of 5ml/min for a total volume of 275 ml and total time of 55 minutes. Typical yields of properly folded prethrombin-2(+18)) ranged from about 2-4 mg/l. Confirmation of

properly folded prethrombin-2(+18) was achieved by a qualitative assay of the activity towards chromogenic substrate spTh after activation by *E. carinatus* venom assay by tracking the time dependent release of para-nitroaniline from spTh (H-D-HHT-Ala-Arg-pNA.2AcOH, American Diagnostica Inc.) by thrombin cleavage under the following conditions: 10 mM HEPES/Tris pH 8.4, 50 μM spTh and 0.05 nM thrombin in 200 μl in a 96-well plate. Activation to α-thrombin was achieved by incubation of the prethrombin-2(+18)) diluted to 50 mL with 50 mM Tris/HCl pH 7.4, 1 mL 1 M $CaCl_2$, 1 mg/ml PEG8000, 5% (v/v) glycerol with 250 μg *E. carinatus* venom (Sigma-Aldrich Cat. # V8250) that was pretreated with AEBSF (Roche, Mannheim, Germany. Cat. # 11429868001;) at room temperature. Optimal activation times varied from 2-10 hr, as predetermined by a estimating the rate of thrombin activation by using MonoS chromatography to periodically track the amount of activated thrombin in an activation reaction. At the end of the activation period, the protein solution was diluted 1:1 with 50 mM Tris/HCl pH 7.4, 1 mg/ml PEG 8000 and loaded onto a MonoS cation exchange column (GE Healthcare Life Sciences). Separation of active α-thrombin from preactive (prethrombin-2, meizothrombin) and auto-protealyzed forms (β- and γ-thrombin) was efficiently achieved using MonoS chromatography (size 10/10, Figure 2.1) with a gradient of 100 mM to 500 mM NaCl in 25 mM phosphate pH 6.5 at a flow rate of 2ml/min for a total volume of 160 ml and total time of 80 minutes .

The fully active α-thrombin was confirmed by fibrinogen clotting times (activity of >3500 U/mg) and chromogenic substrate assays. MALDI-TOF mass spectrometry was

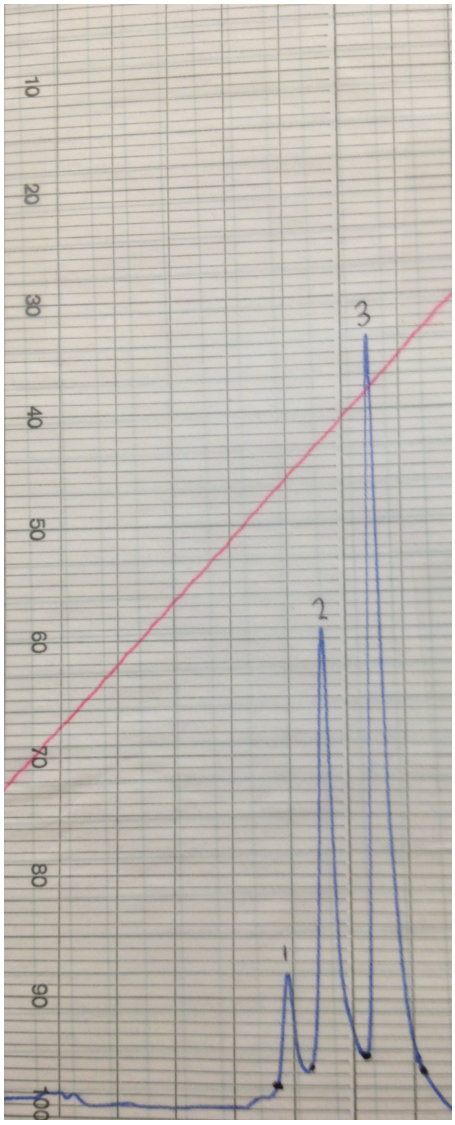


Figure 2.1 MonoS ion exchange chromatography trace showing a wild-type thrombin activation in progress. Red line indicates the percentage completion of the gradient and the y-axis is labled with % buffer A. Blue line indicates A_{280}. Peak 1 is the preactivated prethrombin-2(+18)) before any proteolysis, peak 2 is thrombin that has undergone the internal cleavage by *Echis carinatus* venom separating the light- and heavy-chains and peak 3 is fully active α-thrombin with both the internal venom cleavage and the autoprotealytic removal of the 18 amino acid N-terminal tail.

also used to track proteolysis (Figure 2.2). Samples were prepared with sinapinic acid matrix and data was collected on a Voyager DESTR MALDI-TOF. Spectra collected after full activation of reduced and oxidized samples revealed the correct molecular weight of the oxidized form of 34kD (Figure 2.2 (c)) and of the heavy chain upon reduction of 30.5 kD (including PPACK and seven acetamido groups on the cysteines, Figure 2.2(d)).

Covalent inhibition of thrombin was achieved by incubation with a 10x molar excess of H-D-Phe-Pro-Arg-chlormethylketone (PPACK, Haematologic Technologies Inc.) at room temperature for 1 hour. To separate excess PPACK from thrombin and exchange the protein into NMR buffer, the typically about 10 ml protein solution was concentrated to 2 ml and loaded onto a Superdex S75 size exclusion column (GE Healthcare Life Sciences) eluted with 25 mM sodium phosphate pH 6.5, 150 mM NaCl, 0.05% sodium azide. MALDI-TOF mass spectrometery was used to determine the 2H labeling efficieny, typically between 70% and 75%. A final PPACK-thrombin concentration of 150 μM and 10% (v/v) D_2O was used for the NMR experiments at a final volume of 300 μL in a Shigemi tube (Shigemi Inc. cat. # BMS-005).

For the production of apo-thrombin, site directed mutagenesis was performed on the pET23a+ plasmid containing the wild-type sequence of prethrombin-2(+18)) A major modification of the purification process involves the cleavage of the 18 amino acid extension. Since S195M is a catalytically inactive mutant of thrombin, wild type α-thrombin was used for the full activation of the S195M mutant. After the HiLoadS (GE Healthcare Life Sciences, Piscataway, NJ) purification of properly folded

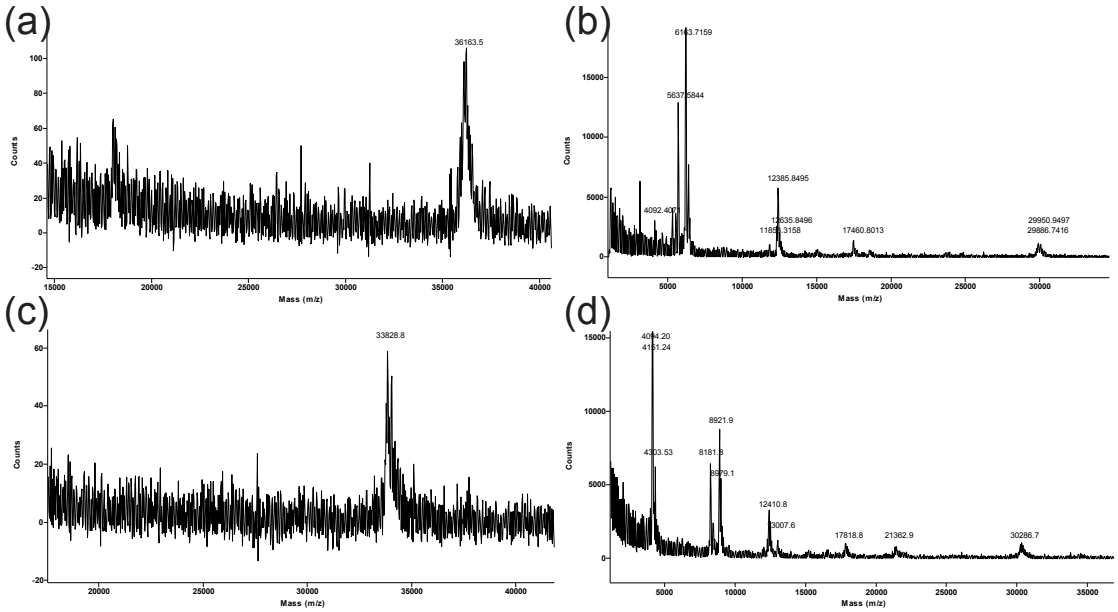


Figure 2.2 MALDI-TOF mass spectra of thrombin in various activation states. (a) Mass spectrum of peak 1 of MonoS ion exchange, with mass corresponding to Prethrombin-2(+18). (b) Mass spectrum of peak 2 under reduced conditions. Mass at about 6000 and about 29900 correspond to the molecular masses of the light and heavy chains. (c) Mass spectrum of peak 3, with molecular mass corresponding to the fully activated α-thrombin and (d) the spectrum of peak 3 after reduction with peaks corresponding to the light chain (~4000) and the heavy chain (~30300) with PPACK attached.

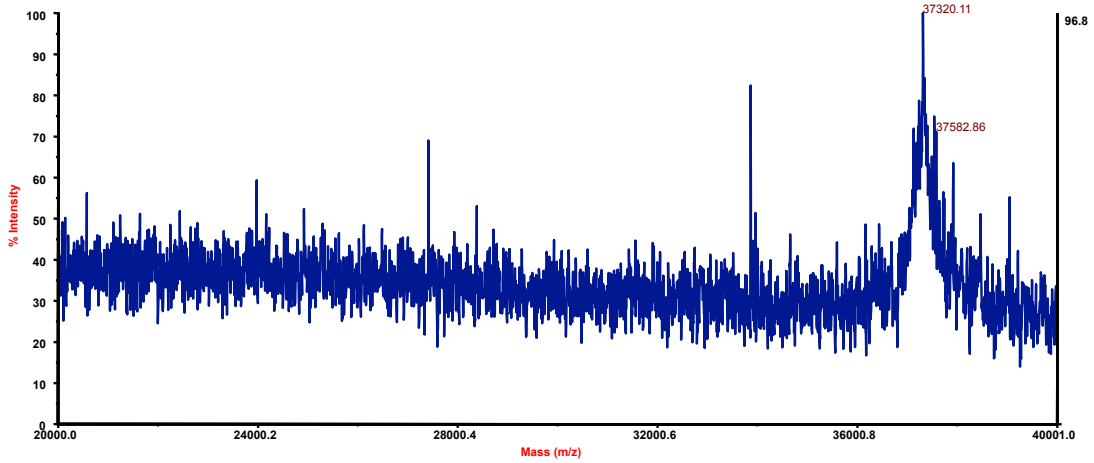


Figure 2.3 MALDI-TOF mass spectrum of activated $^{2}H,^{15}N,\ ^{13}C$ labeled S195M-thrombin. After accounting for the mass of a bound sodium and assuming 98% labeling efficiency of ^{15}N and ^{13}C, the deuteration labeling efficiency was determined to be approximately 75%.

prethrombin-2(+18), in addition to venom activation using *Echis carinatus* venom that was pretreated with AEBSF, approximately 200 μg of previously prepared WT-α-thrombin was added. After an empirically predetermined activation time, the active WT-α-thrombin was inhibited and removed by a20-min room temperature incubation with 140 μg biotinyl-PPACK (Haematologic Technologies, Essex Junction, Vermont). The now biotinylated WT-PPACK-α-thrombin was removed by 2 hr incubation at room temperature with 1ml high-capacity streptavidin immobilized resin (Thermo Scientific, Rockford, Ill.) in ~10 ml protein solution and flowed through a 1 ml empty glass chromatography column to remove resin. The differently cleaved forms of thrombin were then separated by MonoS (GE Healthcare Life Sciences, Piscataway, NJ) ion exchange chromatography as already described.. Effectiveness of this protocol to perform the double cleavage on prethrombin-2(+18) to activate to S195M-α-thrombin was confirmed by mass spectrometry The absence of WT-α-thrombin activity was confirmed using spTh assay as described above.

2. ^{2}H *E. coli* optimization

In order to shorten the time required for acclimating *E. coli* to growth in $^{2}H_2O$ for isotopic labeling of NMR samples, a strain was developed using several rounds of selection pressure to adapt the cells to faster growth times in $^{2}H_2O$. M9 minimal media is used for the production of proteins for NMR samples and to minimize the amount of $^{1}H_2O$ present in the growth medium used, M9 salts (KH_2PO_4, Na_2HPO_4, and NaCl) were dissolved in $^{2}H_2O$ then lyophilized to yield fully deuterated media components. BL21(DE3)pLysS cells (not containing the pET23 plasmid for thrombin expression) were

acclimated to ^2H_2O using the above described protocol. The cells were then plated on 99% ^2H_2O chloramphenicol plates to ensure the retention of the pLysS plasmid. After approximately 5 days (3 days for later rounds of selection), 4 colonies with the largest diameter were selected for growth in 1.5mL 99.9% ^2H_2O M9 minimal media with chloramphenicol. The OD_{600} was measured periodically to monitor the growth rates and a growth rate was calculated. To prevent the exchange of ^1H_2O from the air into the 99.9% ^2H_2O, sterile test tubes were topped with a cuvette and sealed with parafilm, allowing the measurement of OD_{600} without introduction of ^1H_2O and to maintain sterility simply by turning the tube upside-down and filling the cuvette with the culture. The cultures were allowed to grow to an OD_{600} of 0.7 to 0.9 AU and the culture with the fastest calculated growth rate was diluted to $7x10^{-4}$ AU for plating. A portion (50μl) of the diluted cell culture was plated on 99% ^2H_2O chloramphenicol plates and the process was repeated. Iterations of selection were performed until no improvement in growth rate in ^2H_2O media was observed, which occurred in round 6. The ^2H_2O-acclimated cells were made competent using a standard lab protocol for calcium chloride competent cells, with all solutions made with 99.9% ^2H_2O.

C. RESULTS AND DISCUSSION

1. **Protein expression and purification**

The Huntington lab was the first to report expression and refolding of human a-thrombin in milligram quantities (Johnson*, et al.*, 2005). Several modifications and clarifications of the Huntington lab protocol were made in order to produce human α-thrombin for NMR studies. These include the use of 50 mM NaCl in the refolding buffer,

the use of raw *E. carinatus* venom (Sigma-Aldrich Cat. # V8250) inhibited with serine protease inhibitor AEBSF and the use of MonoS ion exchange chromatography for the separation of differently proteolyzed forms of thrombin. The addition of 50 mM NaCl to the refolding buffer approximately doubled the yield of properly folded prethrombin-2(+18), from approximately 2 mg/L to 4 mg/L growth media. Thrombin is a highly dipolar molecule (Baerga-Ortiz*, et al.*, 2000Baerga-Ortiz*, et al.*, 2000) and additional ionic strength provided by the NaCl may contribute to solubility and prevent electrostatic intermolecular interactions that contribute to aggregation. In addition, thrombin specifically binds sodium, an interaction that is necessary for the proper formation of fully procoagulant α-thrombin. In the folding process, it is possible that the binding of sodium to unfolded or partially folded thrombin may nucleate further folding due to structural organization of the Na^+ binding loop.

Echis carinatus venom is used for *in vitro* activation of thrombin due to the difficulty in using the thrombin activation factors of the coagulation cascade. Initially, we used the commercially available *Echis carinatus* venom fraction that is marketed as a thrombin activator (Ecarin, Sigma-Aldrich, St. Louis, MO, Cat. # E0504), which was recommended by Soejima et al. (Soejima*, et al.*, 2001). While this proved to produce α-thrombin with reasonable yields, it was not economically feasible to scale this protocol up to NMR sample quantities. The more economical raw *Echis carinatus* venom (Sigma-Aldrich, St. Louis, MO, Cat. # V8250) ordinarily cleaves thrombin non-specifically and results in total degradation. To prevent this, the serine protease inhibitor AEBSF was used to inhibit proteases in the venom, resulting in specific activation of thrombin without losing a significant portion to degradation.

One unique challenge in protease studies is the autoproteolytic activity of these proteins. In the case of thrombin, one autoproteolytic cleavage is necessary for full activation; the N-terminus of the light chain is separated from the pro-domain by this cleavage *in vitro*. However, there are additional thrombin cleavage sites in ABE1 (to produce β-thrombin) and the γ-loop (to produce γ-thrombin) that if not carefully controlled, could lead to degradation of thrombin and misleading experimental results. To help limit this degradation, refolded thrombin was kept as dilute as possible, with the refolding buffer volume of 2-4 L. The activation reaction was also closely monitored by analytical ion exchange chromatography of small volumes of the activation mixture to ensure that the activation was stopped as soon as the majority of the protein was in the α form.

A major difference between the Huntington lab protocol and ours is that they used a heparin affinity column, whereas we used a MonoS cation exchange column for separation of multiple proteolytic products. In our experience, the heparin column captures multiple proteolytic products while the MonoS column efficiently separates these forms. Thus, MonoS ion exchange chromatography provides an analytical tool to monitor the activation reaction and yields a homogeneous sample for NMR studies.

D. CONCLUSION

NMR is a relatively insensitive technique, requiring a minimum of approximately 100 μM concentration of protein to perform dynamics studies. Without a protocol to consistently produce protein yields in the milligram range, NMR studies are not feasible. *E. coli* protein expression provides a route for isotopic enrichment and is necessary for the 3 dimensional techniques required for resonance assignments on large proteins such as thrombin. The development of an *E. coli* expression, refolding and purification scheme by the Huntington group made the study of thrombin by NMR possible for the first time. Optimization of this scheme resulted in a consistently high yielding prep and allowed for high-resolution spectra to be taken (Figure 2.4). The initial ^{1}H-^{15}N TROSY correlation experiment resulted in a well-dispersed spectrum, indicative of a well-folded protein and encouraging further study. Due to the lack of published assignments of thrombin, in order to study thrombin dynamics, resonance assignments of the ^{1}H-^{15}N TROSY spectrum were needed and were the next hurdle to be overcome in this study.

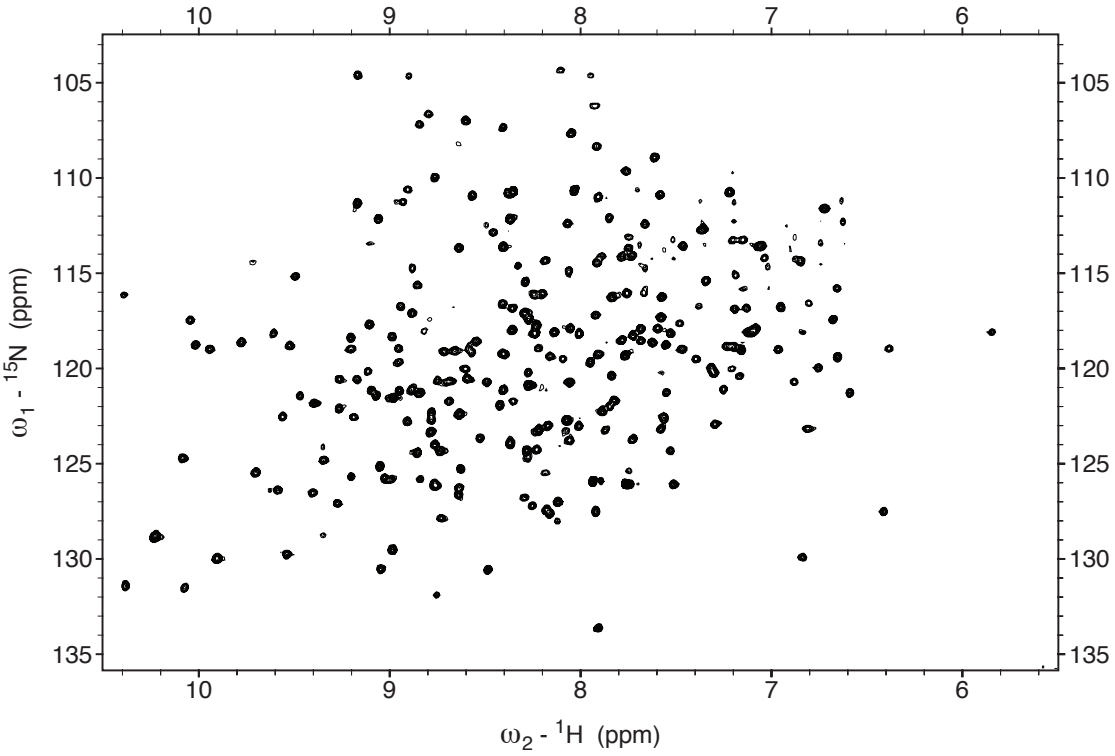


Figure 2.4 ^{1}H-^{15}N TROSY correlation spectrum of [^{2}H-^{15}N] PPACK-thrombin. Spectrum was collected at 25°C and 800MHz with 16 scans and 1024x128 real points.

References

Baerga-Ortiz A, Rezaie AR & Komives EA (2000) Electrostatic dependence of the thrombin-thrombomodulin interaction. *J Mol Biol* **296**: 651-658.

DiBella EE, Maurer MC & Scheraga HA (1995) Expression and folding of recombinant prethrombin-2 and its activation to thrombin. *J Biol Chem* **270**: 163-169.

Johnson DJ, Adams TE, Li W & Huntington JA (2005) Crystal structure of wild-type human thrombin in the Na+-free state. *Biochem J* **392**: 21-28.

Soejima K, Mimura N, Yonemura H, Nakatake H, Imamura T & Nozaki C (2001) An efficient refolding method for the preparation of recombinant human prethrombin-2 and characterization of the recombinant-derived alpha-thrombin. *J Biochem* **130**: 269-277.

Chapter 3

Resonance Assignments

A. BACKGROUND

In order to identify resonance peaks according to position within a protein molecule, connectivities between atomic nuclei must be made. There are two major types of connectivities in molecules that can be experimentally deduced by NMR: spacial proximity and covalent bonding. The presence of the nuclear Overhauser effect (NOE) indicates proximity through space (Keepers & James, 1984) and scalar coupling is present between atomic nuclei that are covalently bound (Ramsey & Purcell, 1952). Presence of either property helps determine clusters of bonded and proximal nuclei in the protein molecule. Chemical shift helps identify residue types (Keim*, et al.*, 1973, Keim*, et al.*, 1973, Keim*, et al.*, 1974), allowing the placement of clusters of connected atoms into specific positions within the known protein sequence, known as resonance assignment. Of particular importance to protein dynamics studies is the identification of the backbone NH resonances present in the ^{1}H-^{15}N HSQC (or ^{1}H-^{15}N TROSY correlation) experiment, on which the 2D separation of resonance peaks in all other dynamics experiments used in this study are based.

The push for larger and larger size limits in NMR has led to the development of several techniques to study the ~34kDa thrombin molecule. 3-dimensional NMR was a key development in greatly improving resolution of NMR allowing separation of resonances to a much larger degree than 2-dimensional NMR (Oschkinat*, et al.*, 1988). Often, a correlation between a carbon atom (or pair of carbon atoms) to an amide N-H pair is made, allowing both identification of residue type and the establishment of scalar coupling between resonances. Overlap in the carbon dimension causes degeneracy in

chemical shift, so multiple 3-dimensional experiments are used to resolve these ambiguities and provide for more chemical shift data making residue type identification easier. The carbon atoms used most commonly and also utilized in this study are the carbonyl carbons of the amide groups (C'), and the alpha and beta carbons of the side-chains ($C^{\alpha}$and C^{β}). The coherence pathway starts on the backbone amide protons (H^N), is transferred to the backbone amide nitrogen (N), then through single or multiple steps, is transferred to the terminal carbon of the experiment, then transferred back through the same path for chemical shift encoding and detection on H^N. Due to the similar values of the one- and two-bond J-couplings between N and C^{α} ($J^1_{NC^{\alpha}}\sim 7-11Hz$ and $J^2_{NC^{\alpha}}\sim 4-9Hz$) simultaneous transfer from N to both the interresidue C^{α} and intraresidue C^{α} (C^{α}_{i-1} and C^{α}_{i}) splits the resonance signal into two peaks in the carbon dimension (Kay*, et al.*, 1990). Selective direction of the coherence transfer to the preceding residue can be achieved by transfer through C' before transferring coherence to the terminal carbon (Bax & Ikura, 1991). This allows for unambiguous identification of the i-1 resonance in the split resonance variety of experiment and also provides a much stronger signal for the preceding resonance, achieving a more complete data set.

Large proteins present a challenge in NMR; in addition to the above mentioned resolution issue, increased transverse relaxation rates (R_2) due to an increase in the molecular tumbling correlation time (τ_c) greatly diminish spectral quality as the size of the observed molecule increases. Replacement of aliphatic protons with deuterons greatly reduces R_2 relaxation effects from protons to H^N and $C^{\alpha}$and generally improves spectral

resolution and sensitivity (Yamazaki, *et al.*, 1994). In Chapter 2, the method used to achieve random fractional deuteration was described.

Another technique that helps negate the effect of increased R_2 relaxation due to large molecular size is Transverse Relaxation Optimized Spectroscopy (TROSY) (Pervushin, *et al.*, 1997). This technique uses spin-state selection to isolate and observe only those spin states with decreased R_2 relaxation. Fully coupled HSQC spectra result in four peaks; two corresponding to the splitting of the N signal by H^N and the splitting of this doublet in the ^1H dimension due to ^{15}N coupling. The resulting four peaks have differing relaxation properties due to relaxation interference from the dipole-dipole and chemical shift anisotropy relaxation mechanisms (Goldman, 1984). The signal component associated with the N^β and H^α spin states uniquely exhibits a constructive interference between dipole-dipole and chemical shift anisotropy relaxation mechanisms, and selection of this spin-state during detection results in lowered R_2 and thus higher quality spectra for large proteins at high magnetic field (Pervushin, *et al.*, 1997, Pervushin, *et al.*, 1998). Use of TROSY to select for specific spin states is also used during relaxation delays and is utilized in TROSY-selected experiments, such as the TROSY-selected Hahn-echo experiment (Igumenova & Palmer III, 2006) discussed in Chapter 4.

These relatively recent developments have greatly increased the pool of proteins that can be studied by NMR by increasing the practical limit in molecular mass. A homooctomeric protein with molecular mass of 110 kDa has been assigned using these techniques (Salzmann, *et al.*, 2000), and by using the same techniques while increasing the dimensionality of the correlation experiments to 4 dimensions, an 81 kDa monomer

has been completely assigned (Tugarinov*, et al.*, 2002), demonstrating that resolution, not relaxation effects, is currently the next major barrier in the size limit of protein NMR. These studies highlight the power of the recent developments in large protein NMR and without these developments, nor the refolding protocol discussed in Chapter 2, atomic resolution dynamics studies of thrombin would not be possible.

B. MATERIALS AND METHODS

All experiments were performed with ~150 μM [^{2}H, ^{13}C, ^{15}N] PPACK-thrombin or S195M-thrombin prepared as described in Chapter 2 in buffer containing 25 mM sodium phosphate pH 6.5, 150 mM NaCl, 0.05% sodium azide. 10% (v/v) D_2O was added as a lock solvent the NMR experiments with a final volume of 300 μL in a Shigemi tube (Shigemi, Allison Park, PA, catalog No. BMS-005).

All triple resonance experiments were performed with cold-probed spectrometers at 298°K. Experiments performed at UCSD Pharmacology using a Bruker Avance III 600 MHz spectrometer were HNCO (Kay*, et al.*, 1990) and HN(CO)CA (Bax & Ikura, 1991). HN(CA)CO (Clubb*, et al.*, 1992) and HN(COCA)CB (Yamazaki*, et al.*, 1994) were performed at NMRFAM with Varian NMR system 600iii and the TROSY-HNCA (Salzmann*, et al.*, 1998), TROSY-HN(CA)CB (performed only on PPACK-thrombin), TROSY-HN(CO)CACB (performed only on S195M-thrombin) (Salzmann*, et al.*, 1999) and NOESY-^{1}H-^{15}N-TROSY (Zhu*, et al.*, 1999) (τ_{mix}= 200 ms) were performed on the NMRFAM Varian VNS 800 MHz for PPACK-thrombin and on the UCSD Chemistry and Biochemistry Varian VS 800 MHz for S195M-thrombin. Data were processed using the NMRPipe suite (Delaglio*, et al.*, 1995) and analyzed using SPARKY (Goddard & Kneller, 2004).

C. RESULTS AND DISCUSSION

1. Resonance Assignments and Chemical Shifts

Resonances were assigned using the "backbone walk" method by searching for interresidue connectivities through the ^{13}C dimension for resonance frequency matches in the SP extension of SPARKY. The three carbon nuclei (C', C^{α} and C^{β}) provided three rungs of connectivity, effectively eliminating ambiguity in interresidue connectivities. Since the structure of PPACK-thrombin is known, the NOESY-^{1}H-^{15}N-TROSY data provided an additional rung of connectivity and confirmation of assigned residues with cross-peaks indicating proximity through space. Examples of strip plots of assigned residues are provided in order to illustrate the residue assignment process and highlight difficulties that were encountered in assigning thrombin. The first (Figure. 3.1) represents a fairly typical case in the assignment of PPACK-thrombin, with all C', C^{α} data present and most but not all interresidue C^{β} data present. For those residues lacking intraresidue C^{β} data from the HN(CA)CB experiment, assignments were made based on C' and C^{α} conectivites, confirmed by NOESY-^{1}H-^{15}N-TROSY and C^{β} chemical shifts were taken from the HN(COCA)CB experiment. In the example presented in Fig. 3.1, all C', C^{α} and C^{β} frequencies were elucidated and all of the residues form R233 (281) to V241 (289) were assigned to a peak in the ^{1}H-^{15}N TROSY spectrum (Figure 2.4). The second case (Figure 3.2) is the strip plots corresponding to the newly assigned γ-loop, a difficult stretch to assign that highlights the difficulties in assigning PPACK-thrombin and the strategy used to overcome these

Figure 3.1 Strip plot of cross peaks corresponding to residues R281 through V289. Three dimensional experiments represented are: (a) HN(COCA)CB and HN(CA)CB, (b) HN(CO)CA and HNCA, (b) HNCO and HN(CA)CO and (d) NOESY ^{15}N-^{1}H TROSY strip plot. Odd strips are from experiments that detect only interresidue ^{13}C correlations (HN(COCA)CB, HN(CO)CA, and HNCO) while even strips are experiments that transfer magnetization to inter- and intraresidue ^{13}C nuclei (HN(CA)CB, HNCA, and HN(CA)CO). ^{13}C dimensions for top three strip plots are to scale and correspond to the chemical shift dispersion of that nucleus type in PPACK-thrombin. Dashed lines represent connectivites through bonds for the top three strips and through space connectivities in the bottom strip.

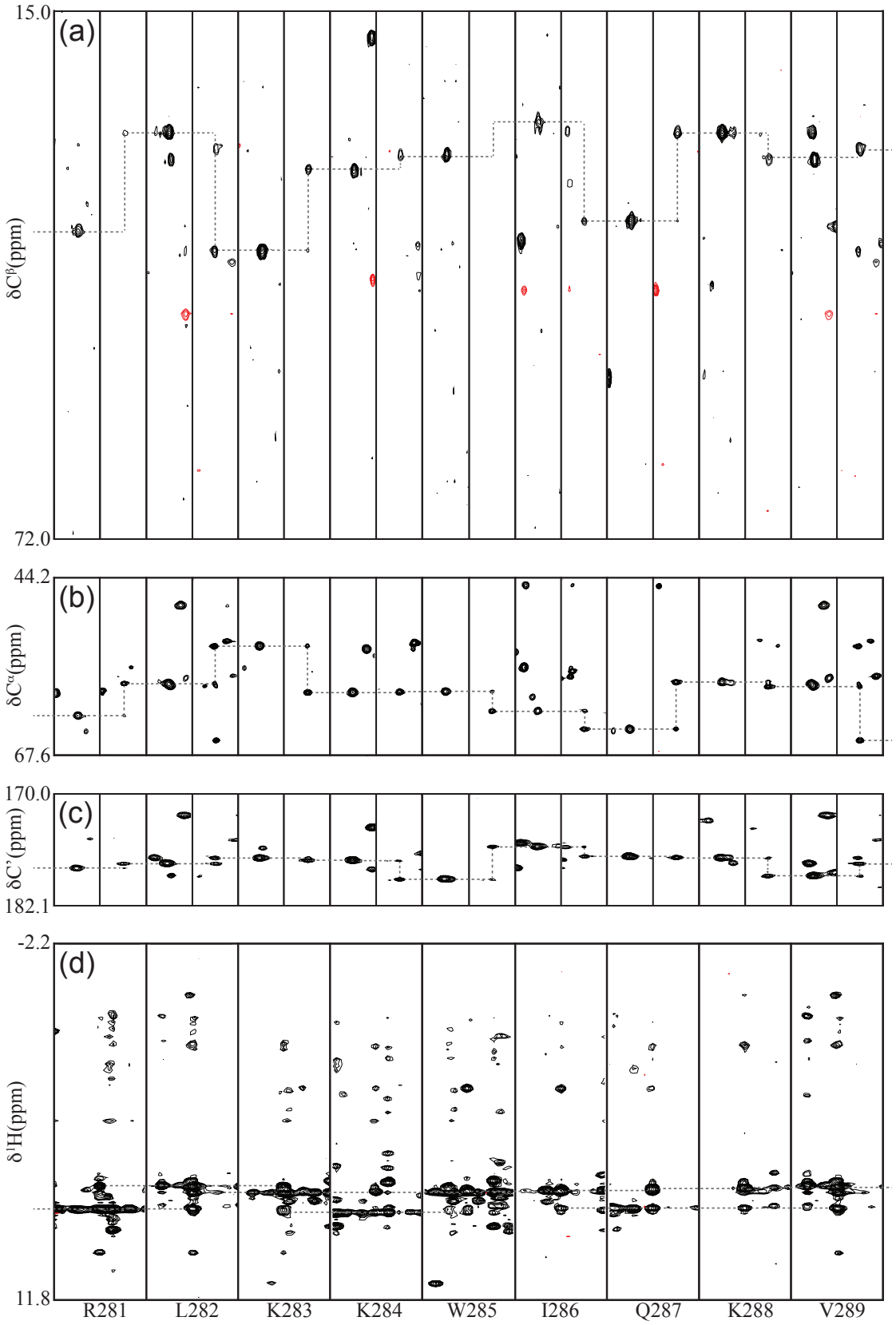
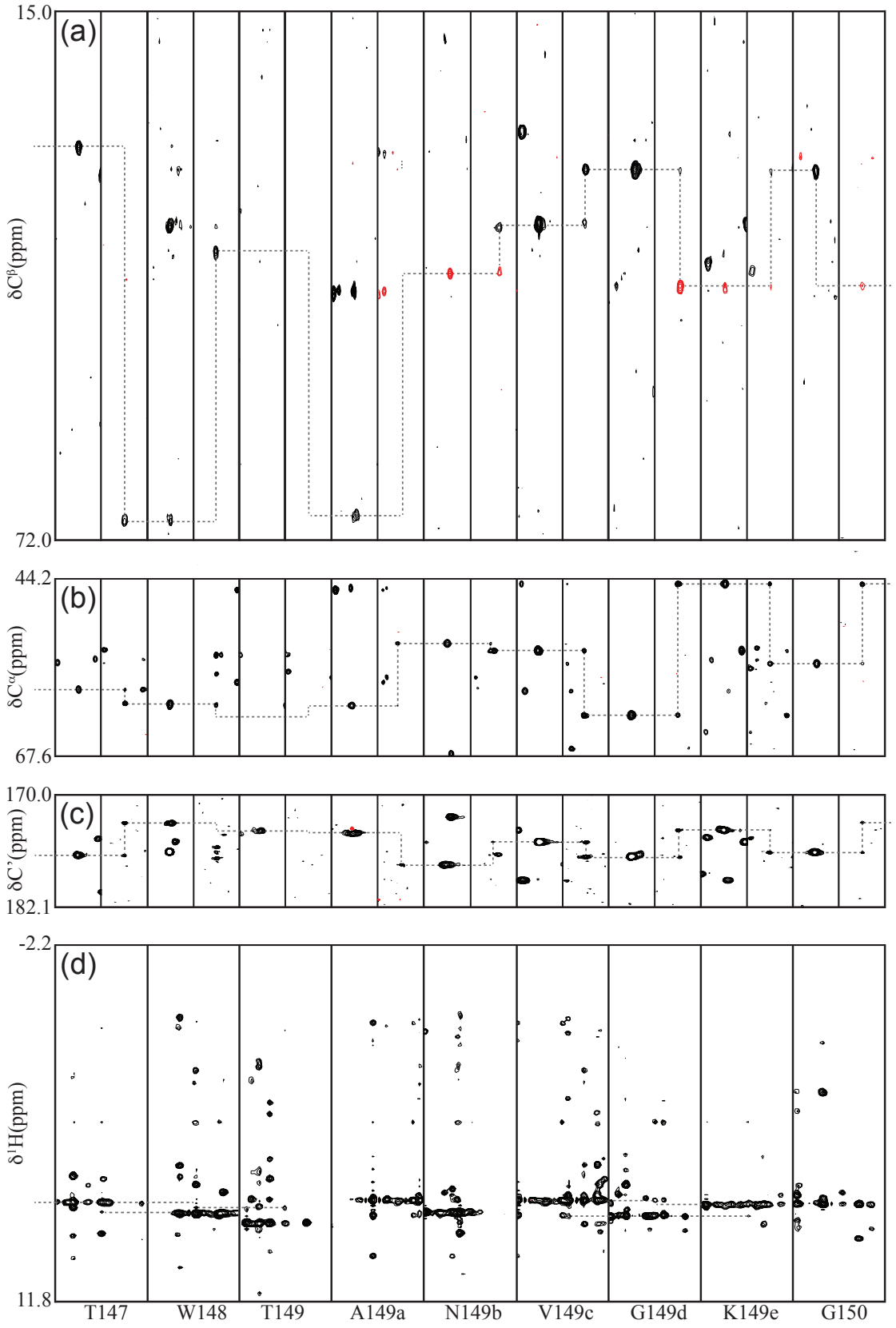

15.0
(a)
δCβ(ppm)
72.0
44.2
(b)
δCα(ppm)
67.6
170.0
(c)
δC'(ppm)
182.1
-2.2
(d)
δ1H(ppm)
11.8
R281
L282
K283
K284
W285
I286
Q287
K288
V289

Figure 3.2 Strip plots cross peaks corresponding to residues T147 through G150. Three dimensional experiments represented are: (a) HN(COCA)CB and HN(CA)CB, (b) HN(CO)CA and HNCA, (b) HNCO and HN(CA)CO and (d) NOESY ^{15}N-^{1}H TROSY strip plot. Odd strips are from experiments that detect only interresidue ^{13}C correlations (HN(COCA)CB, HN(CO)CA, and HNCO) while even strips are experiments that transfer magnetization to inter- and intraresidue ^{13}C nuclei (HN(CA)CB, HNCA, and HN(CA)CO). ^{13}C dimensions for top three strip plots are to scale and correspond to the chemical shift dispersion of that nucleus type in PPACK-thrombin. Dashed lines represent connectivites through bonds for the top three strips and through space connectivities in the bottom strip.


15.0
(a)
δC$^{\beta}$(ppm)
72.0
44.2
(b)
δC$^{\alpha}$(ppm)
67.6
170.0
(c)
δC'(ppm)
182.1
-2.2
(d)
δ^{1}H(ppm)
11.8
T147
W148
T149
A149a
N149b
V149c
G149d
K149e
G150

issues. Much of the data, T149 for example, is near the noise threshold requiring a process of elimination by assigning these residues last. More difficult regions such as this one were the last to be completed in the assignments, as the remaining unassigned resonances could only be assigned after the majority of resonances with full or nearly full data sets were unambiguously identified.

The Huntington group (Lechtenberg*, et al.*, 2010) previously assigned thrombin at 37 ^{o}C in variously liganded forms and among those forms was PPACK-thrombin (Lechtenberg*, et al.*, 2010) These assignments represented the first ever performed on thrombin and are compared to the present study performed at 25^{o}C. In this study (Fuglestad*, et al.*, 2012) 245 resonances were assigned of a possible 278 secondary backbone amides, similar to the 238 residues assigned previously. Table 3.1 lists the chemical shifts of the assigned nuclei. While the number of assigned residues was similar, some regions that were assigned in each study varied (Figure 3.3). At 25^{o}C, the C-terminus of the light chain including resonances G1f through D1a (2 through 8) were visible as was N60g (89), R126 (159), D178 (219), and K236 (284). Remarkably, the resonances for the entire γ-loop, T147 through Q151 (183 through 192), which were missing from the previously published assignments were observed and assigned in our spectra. Resonances V31 through K36, R67 through G69, I79 and E80, I88, W141, and Q156 (52, 57, 98 through 100, 111 and 112, 120, 177 and 197), which were assigned at 37^{o}C, were not observed at 25^{o}C. The temperature dependence of the line broadening is most likely due to chemical exchange in these residues and will be discussed further in Chapter 4.

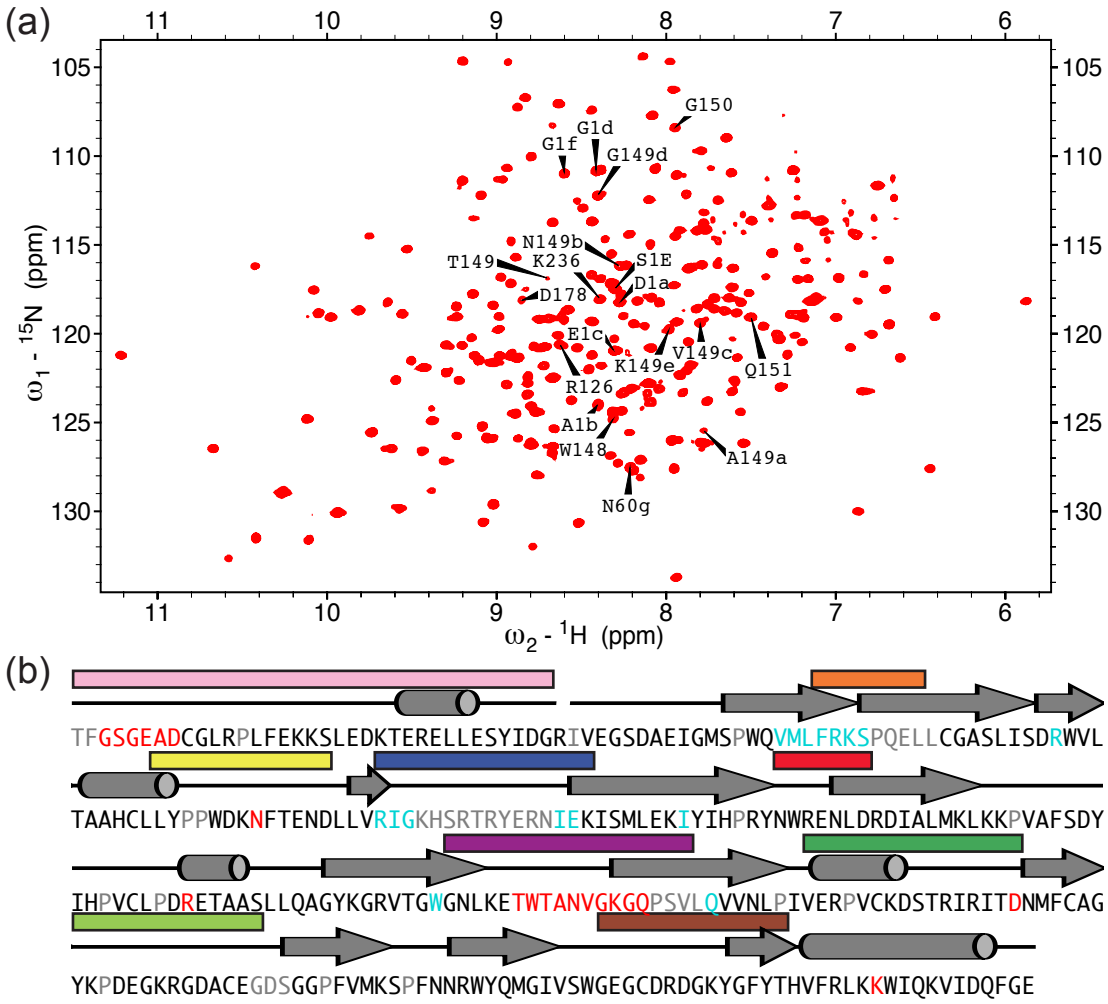


Figure 3.3 Assignment comparison between the current study and the previously published assignments (Lechtenberg, *et al.*, 2010). (a) 1H-15N TROSY correlation spectrum of PPACK-thrombin with previously unpublished assignments labled. (b) Sequence of thrombin with resonances only observable in the current study at 25°C depicted in orange, resonances which were only observable in the previous study at 37°C, resonances observed in both studies in black and resonances observed in neither study in grey. The secondary structure according to the crystal structure of PPACK-thrombin (1PPB, (Bode, *et al.*, 1989)) depicted in a cartoon above the sequence and the loop definitions provided in FIG 1.X depicted in colored rectangles above the sequence.

S195M-thrombin assignments were performed by transferring assignments from the PPACK-thrombin ^{1}H-^{15}N TROSY spectrum for the well-dispersed regions of.the spectrum with small chemical shift differences. Only some of the resonance assignments (162 or 58%) could be transferred as crowding in large portions of the spectrum precluded transfer of many assignments (Figure 3.4). As a result, 3D NMR experiments were also performed on the apo-thrombin to allow connectivities between residues to be established. The NOESY-^{1}H-^{15}N-TROSY proved useful in initial assignment transfer. In crowded regions, even in cases where there were minor chemical shift perturbations, there were often several possible resonances that were possible candidates. In addition to providing through-space connectivities, the NOESY-^{1}H-^{15}N-TROSY spectrum provided side-chain proton correlations due to partial protonation. These side-chain patters in S195M-thrombin were compared to PPACK-thrombin and allowed most of the resonances to be assigned or at least narrowed down to a small group of possibilities. To obtain a full set of assignments for all visible resonances, ^{1}H-^{15}N-^{13}C 3D correlation experiments were performed. The HN(CA)CO, HNCO, HNCA, and HN(CO)CA provided two rungs of through bond connectivity, which proved to be enough to complete assignments and confirm assignments made by spectral comparisons. In total, 212 of 278 possible resonances were assigned for S195M-thrombin, accounting for 76% of the backbone N-H pairs. In comparison, the previously published assignments from the Huntington group accounted for 67% (Lechtenberg*, et al.*, 2010). As with PPACK-thrombin, this could be attributed to differences in the peak intensities between spectra taken at 37^{o}C in the previous study and 25^{o}C in the present study. The results and

implications of variable temperature NMR experiments performed on S195M-thrombin are discussed in Chapter 5.

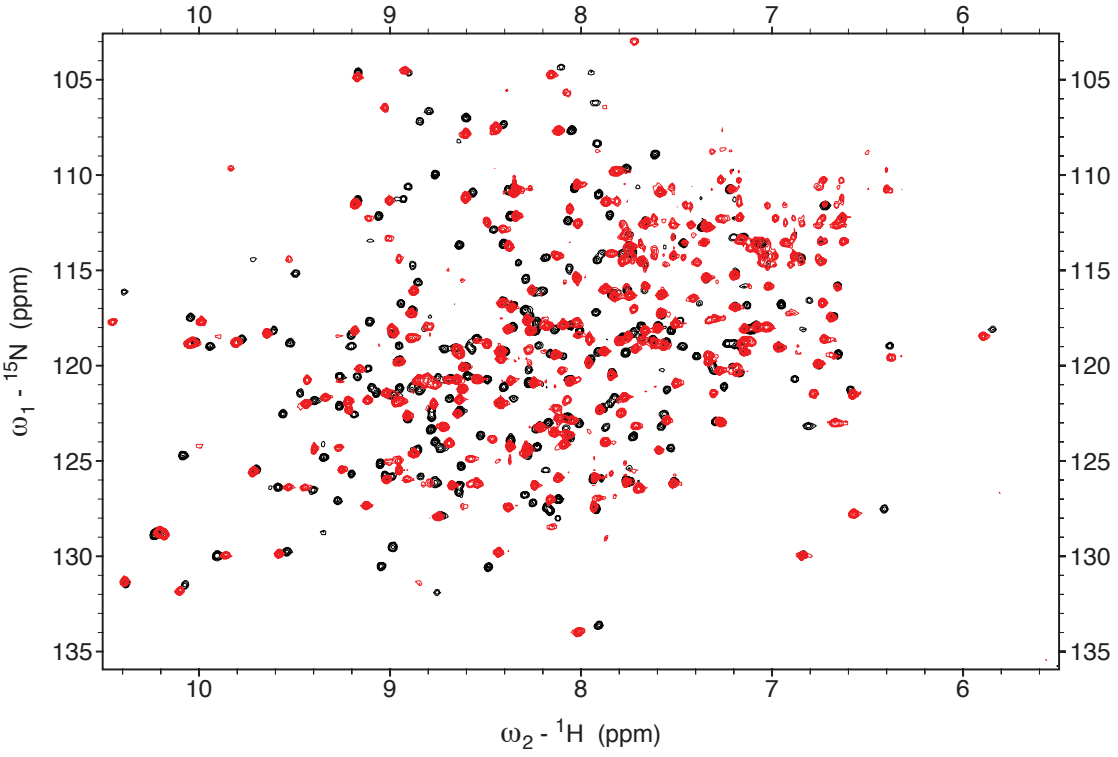


Figure 3.4 Overlay of 1H-15N TROSY correlation spectra of PPACK-thrombin (black) and S195M-thrombin (red). Both spectra were collected under identical conditions; at 25°C and 800 MHz, 16 scans and 1024x128 real points.

With the exception of ABE1, most residues distal to the active site are visible in both PPACK-thrombin and S195M-thrombin. Regions in S195M that became invisible due to increased line-broadening in comparison to PPACK-thrombin are (in addition to the unobserved regions in PPACK-thrombin): V17(38) which is adjacent to the N-terminus of the heavy chain; Q30(51) at the β-sheet leading to the 30s loop; K81(113) at the base of the 60s loop, E97a-L99(130-132) of the 90s loop; T139(175), G140(176), G142(178) through A149a(186) and Q151(192) of the γ-loop; T177(218) of the 170s loop; G188(234) through E192(238) of the 180s loop; G196(242) which is adjacent to the catalytic S195(241), the active site adjacent W215(263); C220(267) through R221(269) of the Na^+-binding loop and the adjacent Y225(273) and G226(274). These regions are mostly in the functional loops surrounding the active site (Figure 3.5). This suggests that ligation of the active site reduces motions in the μs-ms regime and may induce the formation of a functional active site, as discussed previously(Lechtenberg*, et al.*, 2010).

One striking result of the chemical shift assignments is the large degree of chemical shift perturbations experienced by some residues upon occupation of the active site by PPACK. At first glance, it is apparent that the ^1H-^{15}N TROSY spectra of the two forms of the protein are quite dissimilar. Calculation of the overall differences in chemical shifts between multi-dimensional spectra can be performed by a weighted average of the chemical shift differences(Grzesiek*, et al.*, 1996), in this case the ^1H-^{15}N TROSY correlation chemical shifts of PPACK- and S195M-thrombin:

$$\Delta\delta = \sqrt{\left(\Delta\delta_H{}^2 + \frac{\Delta\delta_N{}^2}{25}\right)/2}$$

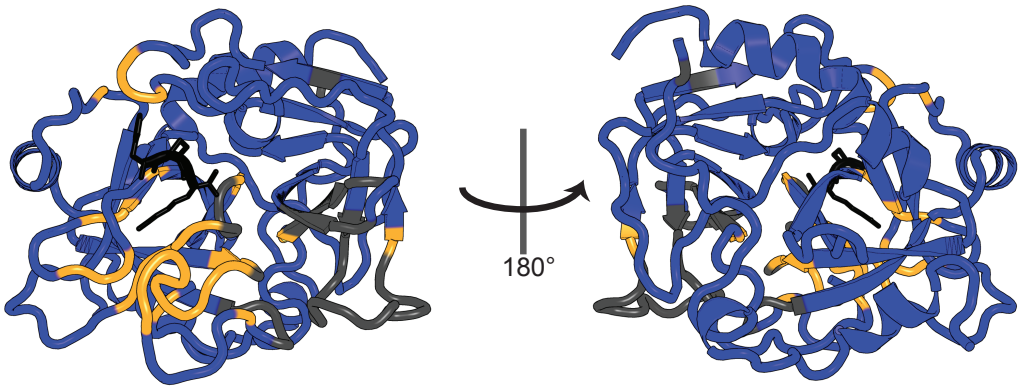


Figure 3.5 Structural representation of residues without resonance signal in PPACK-thrombin (black) and both PPACK-thrombin and S195M-thrombin (Orange) with the covalent inhibitor PPACK depicted in black sticks. A 180° z-axis rotation is provided to illustrate the lack of a large degree of resonance signal loss in this some active-site proximal regions.

where $\Delta\delta$ is the chemical shift difference between PPACK-thrombin and S195M-thrombin of either ^{1}H or ^{15}N nuclei. Residues with the largest chemical shift perturbations ($\Delta\delta > 0.075$ ppm, Figure 3.6) are: 21 (42), 24 (45), 25 (46) and 27 (48) in the N-termainal region of the heavy chain; 42-45 (64-67), and 54 (76) in the N-terminal β-barrel; the α-helix 54-58 (76-80) that contains the His that is part of the catalytic triad; 60 (82) and 60e-60h (87-90) in the 60s loop; 64 (95) in the β-strand connecting the 60s and 70s loops; 88 (120) and 90 (122) in the β-strand connecting the 70s and 90s loops; 95 (127) and 97 (129) of the 90s loop; the C-terminal β-barrel residues 100-105 (133-138) 197 (243), 199 (245), 200 (246), 202 (248) and 203 (249); 130 (166); 157 (198), 159 (200), 160 (201) 210 (258), 212-214 (260-262), and 227-230 (275-278); 163 (204), 169 (210) and 174-176 (215-217) of the 170s loop and helix; 180-185 (221-227) in the β-strand leading to the 180s loop; 216-219 (264-266) and 222 (270) of the Na^{+} binding loop; and 233 (281), 237 (285) and 238 (286) of the C-terminal helix. As expected from mutation and covalent modification, a large degree of chemical shift perturbations are in close proximity to the PPACK binding, however regions distal to the active site also experience chemical shift perturbations (Figure 3.6(b)). Interestingly, a large portion of the C-terminal b-barrel experiences a large degree of chemical shift perturbations, as well as portions of the N-terminal β-barrel that contact the C-terminal β-barrel.

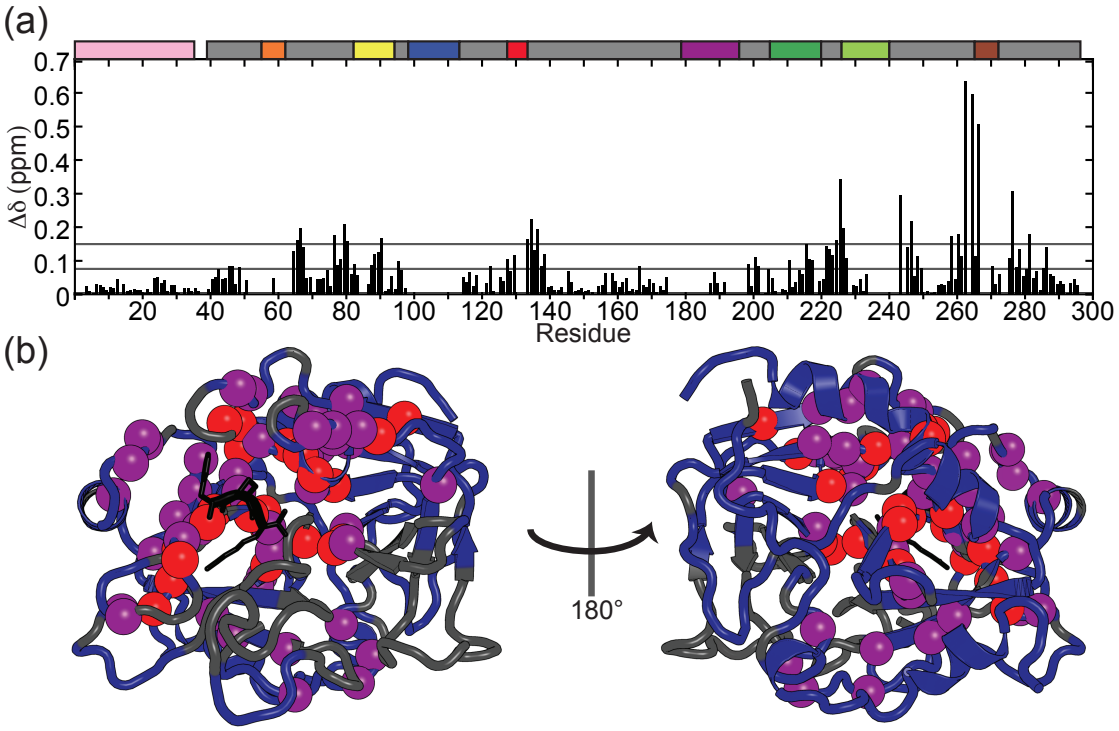


Figure 3.6 Change in chemical shift upon active site occupation of thrombin. (top) Difference in chemical perturbations between PPACK- and S195M-thrombin calculated as a weighted average of the ^{1}H and ^{15}N chemical shift differences. Lines indicate threshold of structural mapping of chemical shift perturbaions (bottom) with perturbations above 0.15 ppm mapped in red spheres on a cartoon representation of the PPACK-thrombin crystal structure (1PPB, (Bode*, et al.*, 1989)), perturbations between 0.075 ppm and 0.15 ppm mapped in purple spheres and resonances with perturbations under 0.075 ppm depicted in blue on the cartoon. Resonances that were unassigned in either or both forms are depicted in grey on the cartoon. A view after a 180° z-axis rotation is depicted on the bottom right.

D. CONCLUSIONS

Using modern NMR techniques such as ^{2}H protein labeling, 3D spectroscopy and transverse relaxation optimization, 88% of PPACK-thrombin and 76% of S195M-thrombin were assigned. The differences between the two forms of thrombin suggest line-broadening from intermediate motions on the NMR timescale, in the μs-ms range. The large degree of resonance broadening indicates that functional dynamics in the active site loops may drive the highly specific recognition of various substrates and highlights the importance of pursuing further experiments to probe the role of these dynamics. Chemical shift perturbations upon occupation of the active site are sensitive to changes in the average magnetic environment surrounding atomic nuclei. Upon occupation of the active site by PPACK, most of the nearby functional loops experienced significant chemical shift perturbations, as well as the C-terminal β-barrel and regions in the N-terminal β-barrel that interfaces with the C-terminal barrel. Previous work from our lab found that binding PPACK to the active site greatly increased the thermal stability of thrombin (the midpoint of thermal denaturation increased by nearly $20^{o}C$) and decreased solvent accessibility in regions distal to the active site (Croy*, et al.*, 2004). The results of resonance assignment and chemical shift perturbations agree with this study; active site occupation stabilizes thrombin and abrogates motions in not only active site adjacent regions, but also regions distal to the active site. Further study must be undertaken in order to elucidate the role of these motions and the functional impact of active site occupation.

References

Bax A & Ikura M (1991) An efficient 3D NMR technique for correlating the proton and 15 N backbone amide resonances with the α-carbon of the preceding residue in uniformly 15 N/13 C enriched proteins. *J Biomol NMR* **1**: 99-104.

Bode W, Mayr I, Baumann U, Huber R, Stone SR & 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.

Clubb R, Thanabal V & Wagner G (1992) A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C'chemical shifts in 15N-13C-labeled proteins. *J Magn Reson* **97**: 213-217.

Croy CH, Koeppe JR, Bergqvist S & Komives EA (2004) Allosteric changes in solvent accesibility in thrombin upon active site occupation. *Biochemistry* **43**: 5246-5255.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J & Bax AD (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. Journal of biomolecular NMR. *J Biomol NMR* **6**: 277-293.

Fuglestad B, Gasper PM, Tonelli M, McCammon JA, Markwick PR & Komives EA (2012) The dynamic structure of thrombin in solution. *Biophys J* **103**: 79-88.

Goddard TD & Kneller DG (2004) SPARKY 3. University of California, San Francisco.

Goldman M (1984) Interference effects in the relaxation of a pair of unlike spin-1/2 nuclei. Journal of magnetic resonance. *J Magn Reson* **60**: 437-452.

Grzesiek S, Stahl SJ, Wingfield PT & 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.

Igumenova TI & Palmer III AG (2006) Off-Resonance TROSY-Selected R1ρ Experiment with Improved Sensitivity for Medium-and High-Molecular-Weight Proteins. *J Am Chem Soc* **128**: 8110-8111.

Kay LE, Ikura M, Tschudin R & Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson* **89**: 496-514.

Keepers JW & James TL (1984) A theoretical study of distance determinations from NMR. Two-dimensional nuclear Overhauser effect spectra. *J Magn Reson* **57**: 404-426.

Keim P, Vigna RA, Marshall RC & Gurd FRN (1973) Carbon 13 nuclear magnetic resonance of pentapeptides of glycine containing central residues of aliphatic amino acids. *J Biol Chem* **248**: 6104-6113.

Keim P, Vigna RA, Morrow JS, Marshall RC & Gurd FRN (1973) Carbon 13 nuclear magnetic resonance of pentapeptides of glycine containing central residues of serine, threonine, aspartic and glutamic acids, asparagine, and glutamine *J Biol Chem* **248**: 7811-7818.

Keim P, Vigna RA, Nigen AM, Morrow JS & Gurd FR (1974) Carbon 13 nuclear magnetic resonance of pentapeptides of glycine containing central residues of methionine, proline, arginine, and lysine. *J Biol Chem* **249**: 4149-4156.

Lechtenberg BC, Johnson DJ, Freund SM & Huntington JA (2010) NMR resonance assignments of thrombin reveal the conformational and dynamic effects of ligation. *Proc Natl Acad Sci U S A* **107**: 14087-14092.

Oschkinat H, Griesinger C, Kraulis PJ, Sørensen OW, Ernst RR, Gronenborn AM & Clore GM (1988) Three-dimensional NMR spectroscopy of a protein in solution. *Nature* **332**: 374-376.

Pervushin K, Riek R, Wider G & Wüthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**: 12366-12371.

Pervushin K, Riek R, Wider G & Wüthrich K (1998) Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in 13C-labeled proteins. *J Am Chem Soc* **120**: 6394-6400.

Ramsey NF & Purcell EM (1952) Interactions between Nuclear Spins in Molecules. *Phys Rev* **85**: 143–144.

Salzmann M, Pervushin K, Wider G, Senn H & Wüthrich K (1998) TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci U S A* **95**: 13585-13590.

Salzmann M, Wider G, Pervushin K, Senn H & Wüthrich K (1999) TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *J Am Chem Soc* **121**: 844-848.

Salzmann M, Pervushin K, Wider G, Senn H & Wüthrich K (2000) NMR assignment and secondary structure determination of an octameric 110 kDa protein using TROSY in triple resonance experiments. *J Am Chem Soc* **122**: 7543-7548.

Tugarinov V, Muhandiram R, Ayed A & Kay LE (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. *J Am Chem Soc* **124**: 10025-10035.

Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR & Kay LE (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**: 11655-11666.

Zhu G, Xia Y, Sze KH & Yan X (1999) 2D and 3D TROSY-enhanced NOESY of 15N labeled proteins. *J Biomol NMR* **14**: 377-381.

Chapter 4

Dynamics of PPACK-Thrombin

A. BACKGROUND

The structure of thrombin has mainly been analyzed by x-ray crystallography, and the many available structures reveal thrombin surface loops "trapped" in different conformations under crystallographic conditions. Di Cera's group recently captured the same mutant form of thrombin in two different conformational states in the same crystals (Niu*, et al.*, 2011). Hydrogen-deuterium exchange followed by mass spectrometry suggested that most of the surface loops are dynamic based on their rapid amide exchange behavior (Koeppe*, et al.*, 2005). In the first report of NMR resonance assignments of variously liganded forms the resonances for exosite 1 (both 70s and 30s loops) were missing unless a ligand was bound directly to this loop and resonances for the γ-loop were not observed under any conditions (Lechtenberg*, et al.*, 2010). Some resonances for the light chain were also missing. Although the authors of this work asserted that the absence of resonances is indicative of μsec-msec dynamics, many other reasons including proteolysis of surface loops, or weak association of the proteins at high NMR concentrations could also account for such observations.

In this study we provide a detailed quantitative analysis of the dynamic properties of thrombin across a broad range of time-scales using a combination of NMR experiments and state-of-the-art molecular dynamics simulation. In the past two decades, NMR has emerged as the method of choice for investigating protein dynamics across a broad spectrum of time scales (Markwick & Nilges, 2011, Salmon*, et al.*, 2011). In the present study, we have employed nuclear spin relaxation (R_1, R_2, heteronuclear NOE), residual dipolar coupling (RDC) experiments and relaxation due to chemical exchange

(R_{ex}) measurements to study the dynamic properties of PPACK-inhibited thrombin on multiple time scales. Although both nuclear spin relaxation and RDCs are associated with magnetic dipolar coupling, they probe protein dynamics on very different time scales. Nuclear spin relaxation results from the time-dependent stochastic modulation of physical interactions between spin-active nuclei, inducing random field fluctuations, which relax the excited spin state back to equilibrium. These experiments are now routinely employed to probe local dynamic fluctuations up to the characteristic rotational correlation time of the system, which is nearly 20 ns for thrombin. In the framework of the well-known Lipari-Szabo model-free analysis, the temporal reorientation fluctuation of an inter-nuclear vector is described using three parameters: the internal and total correlation times and an order parameter, S^2, which characterizes the spatial restriction of the inter-nuclear vector.

Residual Dipolar Couplings (RDCs) have long been recognized as very powerful tools for structure determination, particularly when combined with other distance and dihedral angle restraints derived from NOEs and scalar J-couplings respectively (Tjandra & Bax, 1997). However, it is now well accepted that RDCs can also be applied to the study of protein dynamics (Prestegard*, et al.*, 2000). RDCs are averaged over all orientations of the magnetic dipolar interaction vector sampled up to a time scale defined by the inverse of the alignment-induced coupling, and therefore report on time-averaged motions up to the millisecond range. This early work focused on rigid body domain motions (Tolman*, et al.*, 1997, Braddock*, et al.*, 2001). More recently attention has focused on the study of slow backbone conformational dynamics primarily in ubiquitin using both motional models and MD simulation strategies: The methods include Self-

Consistent RDC-based Model free approach (SCRM), (Lakomek*, et al.*, 2006), the Direct Interpretation of Dipolar Couplings (DIDC) method (Tolman, 2002), the Structure-Free Gaussian Axial Fluctuation model (SF-GAF) approach (Salmon*, et al.*, 2009), Ensemble Refinement with Orientational Restraints (EROS) (Lange*, et al.*, 2008) and the Accelerated Molecular Dynamics (AMD) approach (Markwick*, et al.*, 2009) employed in the present work. The agreement between the RDCs back calculated from the AMD ensemble was statistically significantly better than the match to any crystal structure. Even more dramatic improvements were observed when the same approach was applied to the first four ankyrin repeats of IκBα (Cervantes*, et al.*, 2009).

Measurements of R_{ex} are an indicator of structural fluctuations on the μs to msec timescales (Salmon*, et al.*, 2011). These measurements are a result of the nuclear probe (backbone nitrogens in this study) undergoing exchange between states of distinct chemical shifts on the NMR timescale. The effect of temperature on line broadening was also measured. With increasing temperature, peak intensities are expected to increase as the tumbling of the molecule becomes faster. Deviations from this expected behavior are indicative of exchange processes and may give insight into the characteristics of these exchange processes.

The experimental NMR studies were complemented by a theoretical investigation using both conventional molecular dynamics (CMD) simulations and AMD, an enhanced conformational space sampling algorithm. We report here the first direct measurements of the backbone dynamics of D-Phe-Pro-Arg chloromethylketone (PPACK)-inactivated human α-thrombin. The results reveal an ensemble in which much of the surface of thrombin remains surprisingly dynamic even in the active-site liganded form.

B. THEORY OF DYNAMICS BY NMR

Order parameters (S^2) describe the motion of a bond vector. S^2 is modeled as diffusion in a cone, independent of the rotational diffusion of the molecule. It makes for a powerful analysis of protein motions with no need to make assumptions about dynamics and also provides a convenient parameter in the comparison to computational models of protein dynamics. The Lipari-Szabo model-free formalism allows the calculation of order parameters for protein amide bond vectors using backbone Nitrogen R_1, R_2 and ^{15}N-{^{1}H}NOE relaxation measurements (Lipari & Szabo, 1982). The spectral density function ($J(\omega)$) is sampled using the following:

$$R_1 = \frac{1}{4}d^2[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2J(\omega_N)$$

$$R_2 = \frac{1}{8}d^2[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)]$$

$$+ \frac{1}{6}c^2[4J(0) + 3J(\omega_N)] + R_{ex}$$

$$NOE = 1 + \frac{1}{4}d^2T_1(\gamma_H/\gamma_N)[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]$$

and,

$$d = (\mu_o h\gamma_H\gamma_N/8\pi^2\langle r_{NH}^3\rangle)$$

$$c = \omega_N\Delta\sigma/\sqrt{3}.$$

ω_H and ω_N are the angular Larmor frequencies for ^{1}H and ^{15}N nuclei; μ_0 is the permeability of free space; h is Planck's constant; γ_H and γ_N are the gyromgnetic ratios

of ^{1}H and ^{15}N nuclei; r_{NH} is the length of the N-H bond vector; $\Delta\sigma$ is the chemical shift anisotropy of the ^{15}N nucleus.

The spectral density is fit using the relaxation calculated points and is related to the order parameter by assuming uncoupled rotational diffusion of the protein and internal motions of the bond vector:

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_c}{1+(\tau_c\omega)^2} + \frac{(1-S^2)\tau}{1+(\tau\,\omega)^2}\right)$$

Where S^2 is the squared generalized order parameter (referred herein as the order parameter), τ_c is the molecular rotational correlation time, $\tau^{-1} = \tau_c^{-1} + \tau_e^{-1}$ and τ_e is the effective correlation time of the bond vector under study. This model assumes isotropic motion and a single internal order parameter with a correlation time faster than that of the rotational diffusion of the molecule. Additional models of internal motion can be applied accounting for multiple modes of motion (Clore*, et al.*, 1990). The above (model 2) can be simplified to describe a bond vector with motions too fast to characterize ($\tau_c \gg \tau_e$) by removing the term containing the internal correlation time (τ), known as model 1. Model 3 represents a bond vector undergoing model 1 motions with an additional contribution to transverse relaxation arising from chemical exchange (R_{ex}). Similarly, model 4 represents model 2 motions with an R_{ex} term contributing to additional R_2. Model 5 represents internal motions on two time scales, one that is very fast compared to the molecular correlation time, and to other that is slow enough to be characterized by an internal correlation time (τ_e).

In addition to internal motions, the overall motion of the molecule must be carefully approximated as artificial models or inaccurate calculation of motional parameters may be introduced (Schurr*, et al.*, 1994). In addition to isotropy, full anisotropy and axial symmetry may be modeled and accuracy can be determined according to a fit with the experimental relaxation data (Blackledge*, et al.*, 1998). A fully anisotropic model incorporates three unequal axes of rotational diffusion; the molecule rotates at different rates along each of the principle axes of the fitted diffusion tensor ($D_{zz} \neq D_{yy} \neq D_{xx}$, (Woessner, 1962)). Axial symmetry is defined by two equal transverse axes that are unequal to the longitudinal axis ($D_{zz} = D_{\parallel}, D_{yy} = D_{xx} = D_{\perp}$), with a prolate spheroid having a longer longitudinal axis ($D_{\parallel} > D_{\perp}$) and an oblate spheroid having a shorter longitudinal axis ($D_{\parallel} < D_{\perp}$). The best fitting model of rotational diffusion is incorporated into the spectral density analysis for the most accurate characterization of motional parameters.

Protein dynamics exist in a continuum of timescales. For nuclei undergoing exchange in the time regime of approximately 50 μs to 10 ms, contributions to transverse relaxation are:

$$R_2 = R_2^0 + R_{ex}$$

Where R_2^0 is the transverse relaxation assuming no contribution from conformational fluctuations and R_{ex} is the additional contribution due to this exchange. The above model-free analysis with extended motional models does account for R_{ex}, but these terms

are notoriously inaccurate. In cases where the exchange is defined by excursions from a highly populated state, p_a, to a lowly populated state, p_b, and the R_2^0 of the two states is approximately equal, R_{ex} is defined by:

$$R_{ex} \approx \frac{p_a p_a k_{ex}}{1 + (k_{ex}/\Delta\omega^2)}$$

Where k_{ex} is the rate of exchange between the two states and $\Delta\omega$ is the difference in the chemical shift of the two states. This opens an avenue to precisely study protein motions in the μs-ms timescale, motions that may be important to chemical catalysis, ligand recognition and binding, folding and allostery. A comprehensive review of theory, methodology, and biological applications of protein dynamics studies is available (Palmer, 1994).

B. MATERIALS AND METHODS

1. NMR experiments

Relaxation experiments were performed at 600 MHz on a ^{15}N, ^{2}H-labeled thrombin sample at 150μM. ^{15}N–{^{1}H}NOEs were measured with presaturated and unsaturated experiments that were recorded interleaved and in duplicate to estimate error based on standard deviation between each duplicate ^{15}N–{^{1}H}NOE value. Experiments were collected with 256 scans in a 1024x120 data matrix with a recycle delay of 4.5 s. The ^{15}N–{^{1}H}NOE values were calculated from the ratio of peak intensities from the presaturated to the unsaturated experiments determined in SPARKY. ^{15}N longitudinal

(T_1) and transverse (T_2) relaxation experiments were performed at 600 MHz in an interleaved manner with 2 duplicate points for error estimation. Experiments were collected with 64 scans, 1024x200 complex data points, T_1 delays of 1, 50, 100(x2), 150, 350, 700(x2), 1000 and 1700 ms, and T_2 delays of 0, 17.6, 35.2(x2), 52.8, 70.4, 88.0(x2), 123.2, and 176.0 ms. Peak intensities were measured in Sparky and errors were linearly interpolated from duplicate measurement errors by the perl script sparky2rate (http://xbeams.chem.yale.edu/~loria/sparky2rate, P. J. Loria, Yale University) before non-linear least squares fitting with the program CurveFit (http://www.palmer.hs.columbia.edu/software/curvefit.html, A. G. Palmer III, Columbia University).

PPACK-thrombin dynamic analysis was performed using the program TENSOR2 (Dosset*, et al.*, 2000). To estimate the overall rotational diffusion, we used only those residues with an R_2/R_1 ratio within 1 standard deviation of the average, and low errors. These data were best fit by the axially symmetric tensorial model for rotational diffusion of PPACK-thrombin and yielded a τ_c value of 16.3 ns, consistent with the molecular weight of thrombin and $D_{\parallel}/D_{\perp}$ of 0.868. S^2 values were calculated using the Lipari-Szabo model free approach (Lipari & Szabo, 1982, Mandel*, et al.*, 1995). When necessary, extended models accounting for chemical exchange and motions on two timescales were employed (Clore*, et al.*, 1990).

^1H-^{15}N backbone Residual Dipolar Couplings (RDCs) were measured by comparing changes in the chemical shift differences between the ^1H TROSY and anti-TROSY peaks in an interleaved spin-state selected TROSY-anti-TROSY experiment (Ding & Gronenborn, 2003). Aligned samples of 100 μM ^{15}N, ^2H-labeled thrombin were

prepared by using ultracentrifugation to exchange Pf1 phage (Asla Biotech) into the buffer described above (including 10% D_2O), then adding Pf1 to a total concentration of 3mg/ml. Experiments on the aligned (isotropic) sample at 800 MHz had 128 scans (64 scans), with 2048 x 256 complex data points collected over about 20 hours (10 hours).

R_{ex} was measured at 800 MHz by using the TROSY based Hahn-echo pulse sequence as previously described (Wang*, et al.*, 2003) with a spin-echo delay of $2\tau = 2/J_{NH}$ = 21.6 ms. The experiment was performed interleaved and in triplicate with 64 scans and 1024x256 complex data points. Peak intensities of each spin state relaxation experiment (I^{α}, I^{β} and I^{2HzNz}) were used to calculate R_{ex} according to the relations (Wang*, et al.*, 2003):

$$R_{ex} = R_2^{\alpha} - R_1^{2HzNz}/2 - \eta_{xy}(\kappa - 1) + R_1^N/2$$

$$R_2^{\alpha} - R_1^{2HzNz}/2 = \frac{\ln[I^{\alpha}(2\tau)/I^{2HzNz}(\tau)]}{-2\tau}$$

$$-\eta_{xy} = \frac{\ln[I^{\beta}(2\tau)/I^{\alpha}(2\tau)]}{4\tau}$$

with κ calculated as the trimmed mean of $1 + (R_2^{\alpha} - R_1^{2HzNz}/2)/\eta_{xy}$ for all non-exchanging residues. R_1^N is negligible for the 34 kDa thrombin molecule.

Variable temperature experiments were recorded with ^{1}H-^{15}N TROSY 2D [^{1}H-^{15}N] correlation experiments (64 scans and 2048x256 complex data points) at 283°K, 293°K, 298°K, 303°K, and 308°K on a sample of 100 μM ^{15}N, ^{2}H-labeled PPACK-

thrombin. Peak intensities were measured using Sparky as described by others working with similarly large proteins (Masterson*, et al.*, 2011).

2. Computation

All computational work was performed in the McCammon group by Paul Gasper and Phineus Markwick. Initially a set of five standard conventional MD (CMD) simulations was performed using an in-house modified version of the AMBER 10 code (Case*, et al.*, 2008). For each of these simulations, a different random seed generator for the Maxwellian distribution of atomic velocities was employed and, after standard energy minimization and equilibration procedures, a 20 ns production run MD simulation was performed under periodic boundary conditions with a time-step of 2 fs. Bonds involving protons were constrained using the SHAKE algorithm. Electrostatic interactions were treated using the Particle Mesh Ewald (PME) method with a direct space sum limit of 10 Å. With the exception of the PPACK inhibitor, for which an in-house gaff force field was generated, the ff99SB force field was used for the solute residues and the TIP3P water force field was employed for the solvent molecules. These initial five 20 ns CMD simulations acted as a control set and were used as the starting point for the AMD simulations. These simulations also provided the average (unbiased) dihedral angle energy, $<V_0(dih)>$ and average total potential energy, $<V_0(tot)>$, used to define the acceleration parameters in the AMD simulations.

In the AMD approach a reference or 'boost energy', E_b, is defined which lies above the minimum of the potential energy surface. At each step in simulation, if the instantaneous potential energy, V(r), lies below the boost energy a continuous, non-negative bias

potential, ΔV(r), is added to the actual potential. If the potential energy is greater than the boost energy, it remains unaltered. The application of the bias potential results in a raising and flattening of the potential energy surface (PES), decreasing the magnitude of the energy barriers and thereby accelerating the exchange between low energy conformational states, while still maintaining the essential details of the potential energy landscape. Explicitly, the modified potential, V*(r), on which the system evolves during an AMD simulation is given by:

$$V^*(r) = V(r), \qquad V(r) \geq E_b$$

$$V^*(r) = V(r) + \Delta V(r), \qquad V(r) < E_b ,$$

and the bias potential, ΔV(r), is defined as:

$$\Delta V(r) = \frac{(E_b - V(r))^2}{E_b - V(r) + \alpha} ,$$

The extent of acceleration (i.e., how aggressively the conformational space is sampled) is determined by the choice of the boost energy, E_b, and the acceleration parameter, α. Conformational space sampling can be enhanced by either increasing the boost energy, or decreasing α. During the course of the AMD simulation, if the potential energy is modified, the forces on the atoms are recalculated for the modified potential and the use of the bias potential defined above ensures that the derivative of the modified

potential will not become discontinuous at points where $V(r)=E_b$. In the present work, we have implemented a "dual boost" AMD approach (Hamelberg*, et al.*, 2007), in which two acceleration potentials are applied simultaneously to the system: The first acceleration potential is applied to the torsional terms only, and a second, weaker acceleration is applied across the entire potential. The background, total acceleration potential enhances the translational-rotational diffusion properties of the solvent molecules, thereby facilitating slow diffusive motions in the solute. As such, the dual boost AMD protocol represents a unified approach to efficiently sample both the torsional degrees of freedom and the diffusive motions of system.

One of the favorable characteristics of the AMD approach is that it yields a canonical average of an ensemble, so that thermodynamic and other equilibrium properties can be accurately determined. The corrected canonical ensemble average of the system is obtained by re-weighting each point in the configuration space on the modified potential by the strength of the Boltzmann factor of the bias energy, $\exp[\beta\Delta V(r_{t(i)})]$ at that particular point.

A series of five "dual boost" AMD simulations (which also used the AMBER 10 code) were performed for 10,000,000 steps (the equivalent of 20 ns) at increasing levels of acceleration. For the torsional acceleration, the acceleration parameter, α(dih) was fixed at 240 kcal/mol and the torsional boost energy, E_b(dih) for the four acceleration levels was set at 480, 720, 960 and 1200 kcal/mol above the average dihedral angle energy, $<V_0(\text{dih})>$ estimated from the unbiased 20 ns CMD simulations. In all AMD simulations, a fixed background total acceleration potential was employed with a boost energy, E_b(tot), defined such that $[E_b(\text{tot})-V_0(\text{tot})]$ was equal to 0.16 kcal/mol times the

number of atoms in the simulations cell (NASC), and the acceleration parameter, α(tot) was also set to 0.16 kcal/mol*(NASC). The physical conditions, force fields and all other simulation parameters employed in the AMD simulations were identical to those described for the CMD trajectories.

For each AMD simulation, a corrected canonical ensemble was obtained by performing the Boltzmann free energy re-weighting protocol described above using the bias potential block averaging method (Markwick & McCammon, 2011) to remove statistical noise errors. In this way, five representative free-energy weighted molecular ensembles were generated at each of the four acceleration levels, along with the five unbiased CMD simulations. Each free energy weighted molecular ensemble was divided into molecular clusters using an in-house RMSD-based QR factorization clustering algorithm. In a series of initial tests, we varied the clustering parameters until optimal convergence in the computed RDC values was obtained. Under optimal clustering conditions we found that each member of a given cluster had a backbone RMSD ranging up to 0.6-0.7 Å with respect to the average structure for the cluster. A Singular Value Decomposition (SVD) analysis (Losonczi*, et al.*, 1999) was performed to determine the optimal alignment tensor for each cluster, and N-H^N RDCs were calculated. A weighted average of the calculated RDC values was obtained using the relative number of structures in each cluster. Finally the RDCs for each free energy weighted ensemble associated with a given acceleration level were averaged.

In previous studies (Markwick*, et al.*, 2009), we assumed that a single alignment tensor was sufficient to describe the orientation of the molecule in the alignment medium. In the present work, we employed a multiple alignment tensor model due to the fact that

the PPACK-thrombin system exhibits considerable flexibility, particularly on slower time-scales. Whereas for the previous study of ubiquitin the backbone RMSD across the AMD trajectory with respect to a randomly chosen structure for this system varied up to 1.4 Å; the backbone RMSD varied up to 2.9 Å for PPACK-thrombin. Large backbone RMSD values may represent a significant change in the preferential orientation of the molecule in the alignment medium. As the exchange between conformational states occurs on time-scales considerably slower than the rotation diffusion coefficient (~20 ns according to Stokes theory), the system has ample time to adopt a different preferential orientation as the system evolves from one conformational state to the next. The introduction of a multiple alignment tensor analysis presented in this work essentially couples the slow intra-molecular conformational dynamics of the system to the preferential orientation of the molecule, while still maintaining the appropriate time- and ensemble averaging properties of the RDC observables. The agreement between the theoretical and experimental RDCs was monitored using the R-factor:

$$R_{factor} = \sqrt{\frac{\sum_{i=1}^{N}(X(i)_{theory} - X(i)_{\exp})^2}{2\sum_{i=1}^{N}(X(i)_{\exp})^2}},$$

where $X(i)_{theory}$ and $X(i)_{exp}$ are the theoretically determined and experimental observables respectively. The "optimal" torsional acceleration level (and hence the "optimal" conformational space sampling) for the reproduction of the experimental RDCs was found to be at [$E_b(dih)$-$V_0(dih)$=960 kcal/mol, $\alpha(dih)$= 240 kcal/mol]. The R-factors for

the agreement of the experimental RDCs were 0.37 to the x-ray structure, 0.32 to the CMD ensemble, and 0.20 to the optimal AMD ensemble. Without the use of the multiple alignment tensor approach, the agreement was 0.23.

The internal dynamics present in the different CMD and AMD simulations of PPACK-thrombin were assessed by calculating the S^2 values relevant for the Lipari-Szabo type analysis (Lipari & Szabo, 1982) of ^{15}N auto-relaxation data and for the S^2 values derived from the AMD ensemble that best fits the RDC data, respectively. In all cases, molecular ensembles generated from the standard CMD simulations and the free-energy weighted molecular ensembles generated from the AMD trajectories were superposed onto the backbone atoms (N, C^{α}, C') of all heavy chain residues for the appropriate average structure. Order parameters, S^2 were calculated as:

$$S^2 = \frac{1}{2}\left[3\sum_{i=1}^{3}\sum_{j=1}^{3} < \mu_i \mu_j >^2 - 1\right],$$

where μ_i are the Cartesian coordinates of the normalized inter-nuclear vector of interest. Others have shown that S^2 values calculated from standard MD simulations in this way were in excellent agreement with experimental S^2 values calculated using the Lipari-Szabo autocorrelation function approach (Genheden*, et al.*, 2010).

3. Residual local frustration analysis

To analyze the crystal structure of PPACK-inhibited thrombin, we used an algorithm for determining residual local frustration, *i.e.* whether a contact between amino acid residues is energetically favorable or not in the folded state (Ferreiro*, et al.*, 2007). This algorithm assesses residue-residue interactions by systematically perturbing the identity of individual residues and evaluating the resulting total energy change. The amino acids forming a particular contact are changed to other amino acids generating a set of decoys for which the total energy of the protein is recomputed. After constructing a histogram of the energy of the decoys and comparing the distribution to the native energy, cut-offs are implemented to identify minimally frustrated or highly frustrated residues. Energetically favorable contacts between residues are depicted by green lines and highly frustrated or energetically unfavorable contacts are red.

D. RESULTS AND DISCUSSION

1. **Residual dipolar couplings**

RDCs were measured using spin-state selected TROSY 2D [^{1}H-^{15}N] correlation experiments to measure chemical shift differences between the TROSY and ^{1}H anti-TROSY peaks in both isotropic and Pf1 phage aligned samples (Ding & Gronenborn, 2003). Surprisingly, PPACK-thrombin aligned in only 3 mg/ml Pf1 phage and RDC values under these alignment conditions ranged from -39.07 to 48.68 (Figure 4.1 (a)). Molecular dipole calculations indicate that the thrombin molecule is highly polarized (Baerga-Ortiz*, et al.*, 2000), and it is likely that the alignment is electrostatically-driven. RDCs could be measured for 209 of the 245 assigned residues. The RDCs calculated from the x-ray crystal structure (PDB ID: 1PPB, (Bode*, et al.*, 1992)) using the program, PALES (Zweckstetter, 2008), matched poorly with the experimental values (R^2=0.72, Figure 4.1 (b)). Specifically, of 209 measured RDCs, 55 were more than 10 Hz different from those calculated for the best-fit alignment tensor for the crystal structure. The poor agreement was substantially worse than was observed for the more rigid protein, ubiquitin (Markwick*, et al.*, 2009) and somewhat worse than was observed for IκBα (Cervantes*, et al.*, 2009). The largest discrepancies were found in the loop regions.

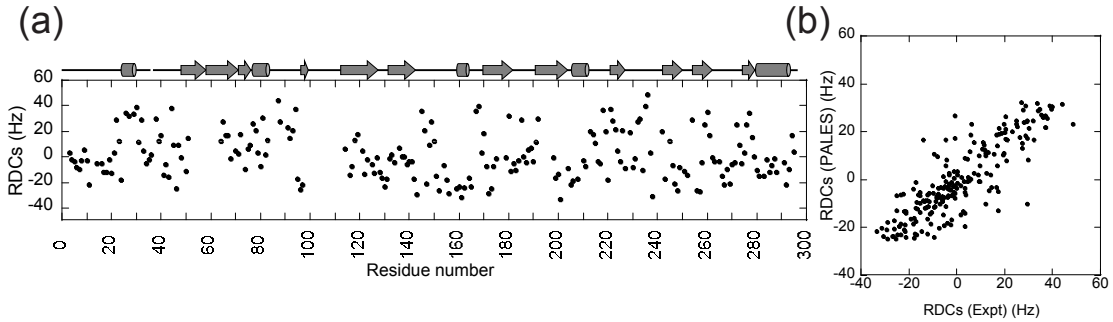


FIGURE 4.1 (a) Residual dipolar coupling (RDC) values measured by TROSY-antiTROSY experiment collected at 800 MHz and 25°C using 3 mg/ml Pf1 phage for alignment. A schematic of the secondary structure is depicted above the plot. A break in the secondary structure diagram indicates where the light chain ends and the heavy chain starts. (b) Plot of measured RDCs vs. those calculated from the x-ray structure (PDB code 1PPB) using PALES (Zweckstetter, 2008).

2. Backbone dynamics experiments

R_1, and R_2 relaxation rates and ^{15}N-{^1H}NOEs were measured for PPACK-[^2H, ^{15}N] thrombin at 600 MHz (Figure 4.2). Much of the protein appeared rigid as indicated by high ^{15}N-{^1H}NOEs and consistent R_2/R_1 ratios, but some regions had remarkably lower ^{15}N-{^1H}NOE values (Figure 4.2 (c)). In particular, the C-terminus of the light chain (residues 14g-14m; 29-36) the γ-loop (residues 149-152; 185-193), and the 180s loop (residues 186a-186d; 229-233) had ^{15}N-{^1H}NOE values lower than 0.4, similar to the disordered termini of the protein. Such low ^{15}N-{^1H}NOE values strongly indicate motions on ps-ns time scales. These regions also had lower than expected R_2 values (Figure 4.2 (b)) indicative of heterogeneous dynamics.

The relaxation data was interpreted within the framework of the Lipari-Szabo model-free analysis using the program, TENSOR2. Initially, the R_1 and R_2 relaxation rates were used to calculate the overall rotational correlation time. A comparative analysis of isotropic, axially symmetric and fully anisotropic diffusion tensors revealed that the axially symmetric model provided the best representation of the rotational diffusion as identified by a statistical Monte Carlo analysis (Schurr*, et al.*, 1994, Dosset*, et al.*, 2000). The rotational correlation time was 16.3 ns, consistent with the molecular weight of thrombin, and $D_\parallel/D_\perp$ of 0.868. Order parameters (S^2) were calculated for 192 of the thrombin resonances (Figure 4.2 (d)). The program TENSOR2 offers a variety of motional models of increasing complexity to fit the relaxation data (Mandel*, et al.*, 1995). Over half of the residues (a total of 101 residues) could be fit with the simplest motional model (model 1) and nearly all of

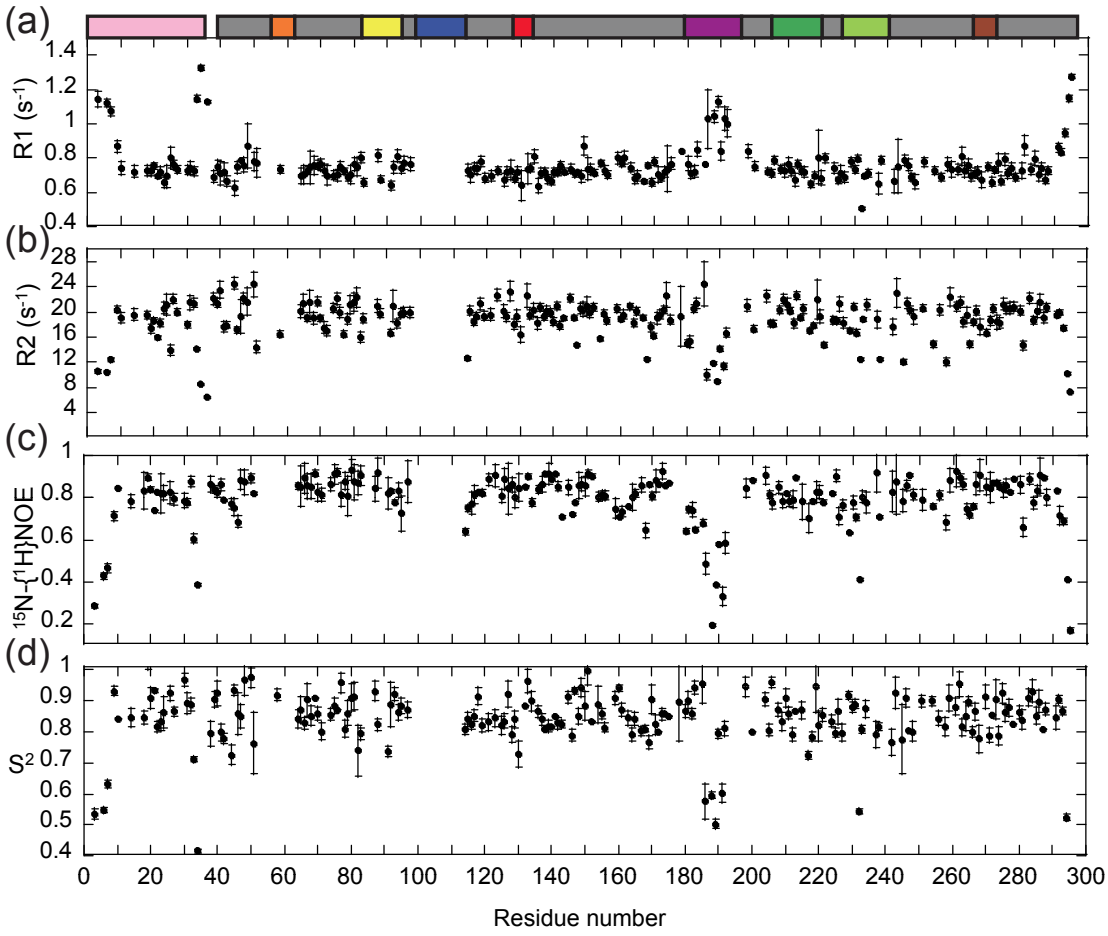


FIGURE 4.2 Backbone relaxation data ((a) R_1, (b) R_2 and (c) heteronuclear NOE) collected at 600 MHz, 25°C for [^{2}H, ^{15}N] PPACK-inhibited human α-thrombin as described in the Supplementary Methods section. (d) S^2 values derived from the model-free analysis of the relaxation data using TENSOR2. The schematic of the loops is depicted above the plot.

these had S^2 values above 0.8 and were in the core of the structure. Many residues (a total of 40) required fitting with a more complex motional model (model 5) and most of these had S^2 values below 0.8. These were located in the N-terminus and C-terminus of the light chain, the C-terminus of the heavy chain, the 30s loop (residues 30 and 36A; 51 and 58), and the γ-loop (residues 149A-150; 186-191). Residues in the 60s loop (60-60A and 60I; 82-83 and 91), the 90s loop (96 and 97A; 128 and 130), and the 180s loop (residues 184,185, 186D and 191; 225, 227, 232 and 237) also had S^2 values below 0.8 (Figure 4.2 (a)). The fact that so many resonances required model 5 fitting was strong evidence of heterogeneous fast time scale dynamics. In order to extract the contribution of chemical exchange to spin-spin relaxation, TROSY Hahn-echo experiments were performed (Wang, *et al.*, 2003). Of the 225 residues, only 15 had significant line broadening due to chemical exchange ($R_{ex} > 6\ s^{-1}$, Figure 4.4 (b)). These residues were: 1F and 3 (3 and 11) of the light chain; 23 and 27 (44 and 48); 66 (97) leading up to the 70s loop; 87 (119) in the surface strand that connects the 70s loop to the 90s loop; 144, 147, 149a, 149c, and 151 (180, 183, 186, 188 and 192) of the γ-loop; 138, 157, and 199 (174, 198, and 245) in the core β-sheet; and 233 (281) that forms a kink in the C-terminal helix. Of these residues, only 23 (44) and 138 (174) required an R_{ex} term to properly fit a Lipari-Szabo model to the relaxation data. These direct measurements of R_{ex} are much more accurate than the estimates obtained from the TENSOR2 analysis of the R_1, R_2, and ^{15}N-$\{^1H\}$NOE data.

3. Variable temperature experiments

Resonances for most of the 30s loop and the 70s loop were neither observed in previous experiments (at 37°C) nor in our experiments (at 25°C) (cf. Figure 3.3). To further probe the temperature dependence of the thrombin backbone resonance intensities, we carried out variable temperature ^{1}H-^{15}N TROSY-HSQC experiments. These experiments revealed complex motions for several of the thrombin surface loops. For example, resonances in the γ-loop were not observed at 37°C (Lechtenberg*, et al.*, 2010), but are observed at lower temperatures (Figure 4.4 (c)). Resonances at the C-terminus of the light chain show similar behavior. In both cases, the decreased signal intensity at higher temperature indicates the region is undergoing exchange processes that are contributing to relaxation (McConnell, 1958).

4. Relation between dynamics and local residual frustration

Over half of the residues in thrombin have high S^2 values, and formed a rigid core. This is consistent with analysis of the results of a residual local frustration analysis (Ferreiro*, et al.*, 2007), which shows a strong network of minimally frustrated contacts throughout the core of the protein (Figure 4.5 (b)). It should be noted that this part of the molecule is also the part that is evolutionally conserved (Suel*, et al.*, 2003). Adjacent to the active site is Ile 16 (36), which makes a very large number of minimally frustrated contacts, and this corresponds to the new N-terminus generated by proteolytic activation that inserts into the core of the protein. Cross-peaks for the oxyanion hole, residues 193-196 (239-242) are not observed, and it is interesting that these residues are engaged in highly frustrated contacts. This is the only region of thrombin with frustrated contacts

located in the core of the protein. Conversely, the dynamic surface loops have many highly frustrated contacts (Figure 4.5 (b)).

5. Molecular Dynamics Simulations

CMD simulations were carried out for 20 ns to reach the same time scale as the rotational diffusion time of the thrombin molecule (Jarymowycz & Stone, 2006, Salmon*, et al.*, 2011). The S^2 values calculated from the CMD simulations matched well with those determined from the R_1, R_2 and ^{15}N-$\{^1H\}$NOE experiments (Figure 4.4 (a)). This result is a testament to the quality of contemporary force field used in this study, ff99SB, and is particularly remarkable in light of the many assumptions that have to be made when interpreting spin relaxation data (Case, 2002, Showalter*, et al.*, 2007).

The excellent agreement between experimental and simulated S^2 values encouraged us to extend our theoretical study to investigate slower time-scale motions using an enhanced conformational space sampling algorithm; AMD (Hamelberg*, et al.*, 2004). AMD is an extended biased potential MD approach that allows for the efficient study of biomolecular systems up to time scales several orders of magnitude greater than accessible using standard CMD approaches, yet maintains a fully atomistic representation of the system. AMD has already been employed with great success to study the dynamics and conformational behavior of a variety of biomolecular systems including polypeptides, folded and natively unstructured proteins (Markwick & McCammon, 2011).

A series of five dual-boost AMD simulations were performed at increasing levels of acceleration (see Supplementary Methods). For each AMD simulation a corrected

canonical ensemble was obtained by performing a Boltzmann free energy reweighting protocol. Successively larger acceleration levels sample greater conformational space and the optimal conformational space sampling for the reproduction of the RDC data was obtained using singular value decomposition (SVD), following a protocol similar to that previously applied to ubiquitin and IκBα (Cervantes*, et al.*, 2009, Markwick*, et al.*, 2009). A multiple alignment tensor model was implemented due to the fact that the PPACK-thrombin system exhibited considerable flexibility, particularly on slower time scales (see details in Methods). The slower motions of the surface loops resulted in slight changes in the molecular alignment tensor such that the use of the multiple alignment tensors improved the agreement of the RDCs calculated from the AMD ensemble with the measured RDCs. This should not be confused with the molecular tumbling analysis implemented in TENSOR2 that determines whether the residues with relaxation parameters within the trimmed mean fit a particular tumbling model. Using this procedure, the optimal acceleration level, and hence the optimal conformational space sampling for the reproduction of the experimental RDCs was found to be [E_b(dih)-V_0(dih)=960 kcal/mol, α(dih)= 240 kcal/mol] (see Methods). This optimal AMD ensemble recapitulated the measured RDC values much better (R^2 =0.92) (Figure 4.3 (b)) than both the x-ray crystal structure (R^2 =0.72) (Figure 4.1 (b)) and the CMD ensemble (R^2 =0.80) (Figure 4.3 (a)).

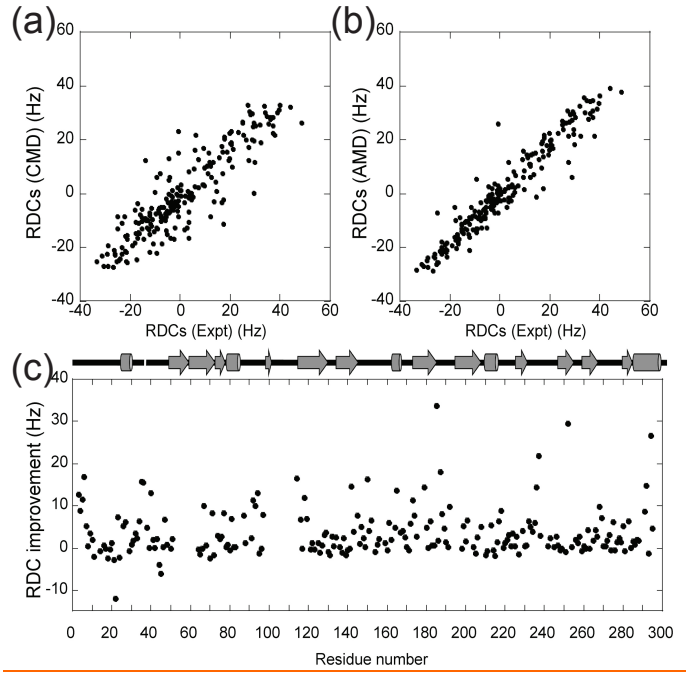


Figure 4.3 (*a*) Plot of measured RDCs versus those calculated from the CMD ensemble (R^2 = 0.80). (*b*) Plot of measured RDCs versus the RDCs calculated from the best-fit AMD ensemble (R^2 = 0.92). (*c*) Plot of the RDC improvement as defined by $|(RDC_{expt}-RDC_{x\text{-}ray})|-|(RDC_{expt}-RDC_{AMD})|$. The schematic of the secondary structure is depicted above the plot.

The most dramatic improvement in the agreement with the experimental RDCs was found for residues in the loop regions (Figure 4.3 (c)). S^2 values were calculated from the standard MD simulation (see Methods), and the results from the CMD ensemble were in excellent agreement with the experimental values calculated from the Lipari-Szabo autocorrelation function analysis, with a few small deviations for the most flexible residues (Figure 4.4 (a)). A temporally heterogeneous distribution of dynamics is revealed in the AMD ensemble (as well as in the experiments), particularly in the light chain (residues 14G-14M; 29-36), the γ-loop, residues 149A-150 (186-191), the 180s loop, residues 184,185, 186D and 191 (225, 227, 232 and 237), and the Na^+ binding site, residues 216–223 (264-271). The loops do not appear to hinge, but rather they appear to be appendages that extend from the rigid core in such a way that within the span of a few residues the S^2 values transition from completely ordered to highly disordered with motions on multiple time scales (Figures (a) and (c)).

The γ-loop remains remarkably dynamic in the inhibited form. The resonances of this loop were missing from all liganded states in previous studies that were carried out at 37^{o}C (Lechtenberg*, et al.*, 2010), and at first we were convinced that the previous studies had been inadvertently carried out with proteolyzed protein. The entire γ-loop was visible in our studies carried out at 25^{o}C, but it showed remarkably low S^2 values, some of the lowest reported for a region other than the terminus of a protein this size. These results help explain the proteolytic sensitivity of the γ-loop, which is also called the autolysis loop, because it can be cleaved by thrombin itself.

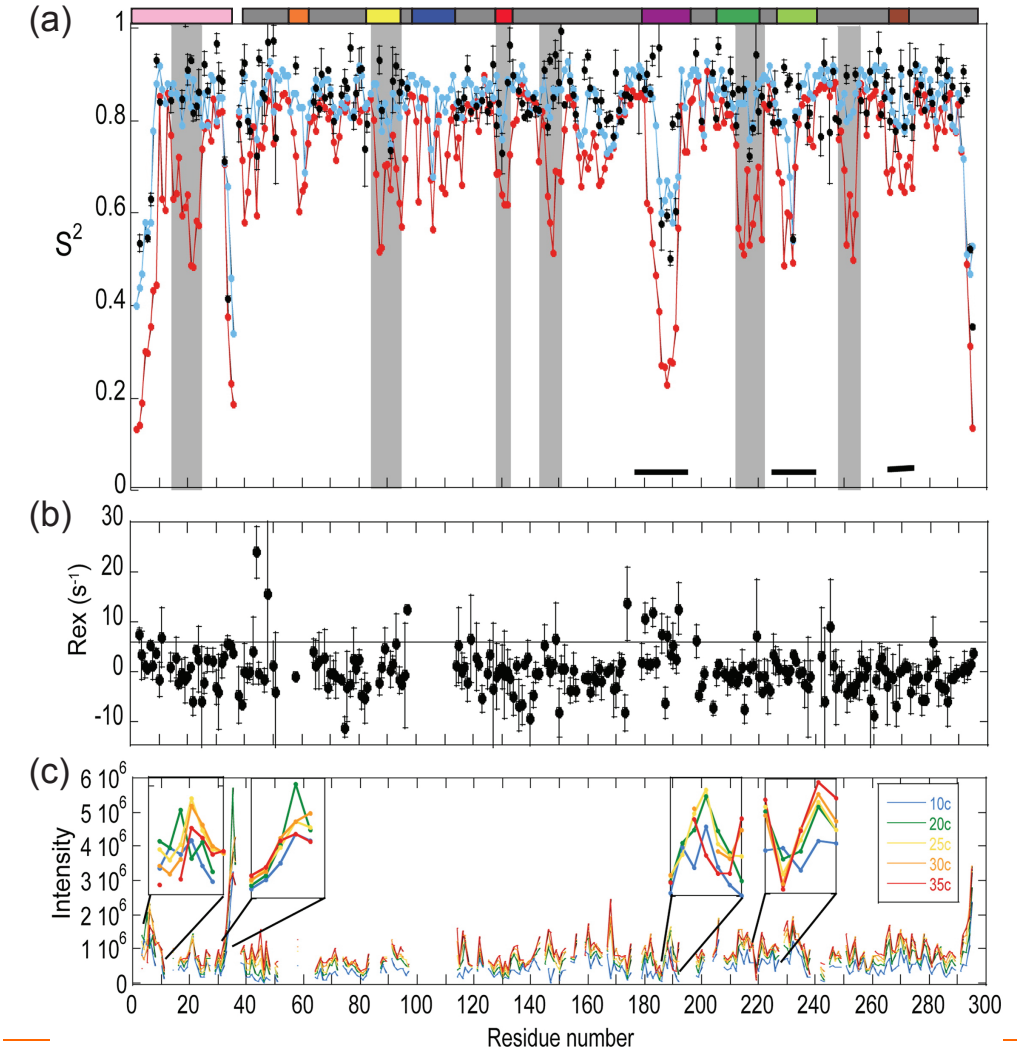


Figure 4.4 (*a*) Plot of the S^2 values calculated from the relaxation data using TENSOR2 (*black*), compared to the S^2 values calculated from the CMD ensemble (*blue*), and compared to the S^2_{RDC} values calculated from the RDC-optimized AMD ensemble (*red*). Regions of the heavy chain comprising the activation domain that were thought to become less dynamic in α-thrombin as compared to prethrombin-2 are marked with horizontal bars. Regions with significant S^2_{RDC} values that show little motional contribution from relaxation calculated S^2 values and measured R_{ex} are boxed in gray. (*b*) R_{ex} values calculated from the TROSY Hahn-echo experiment. Horizontal line at $R_{ex} = 6$ s^{-1} is the threshold for significant contributions to relaxation from chemical exchange. (*c*) Plot of the intensity of the TROSY peak for each resonance at different temperatures from 35°C down to 10°C. For most of the protein, the resonance intensity increases with temperature, however for some regions (*boxed and enlarged*) the relationship between intensity and temperature is indicative of complex slower timescale dynamics. The schematic of the loops is depicted above the plot.

The fact that the γ-loop does not become ordered even in the active site-inhibited form is remarkable.

Resonances in the 70s loop in anion-binding exosite 1 were only observed in previous work when a ligand was bound at this exosite and it was thus thought to be dynamic (Lechtenberg*, et al.*, 2010). We also did not observe resonances for this loop at 25°C. Surprisingly, however, the S^2_{RDC} values calculated from the AMD ensemble (which represents dynamics out to milliseconds) revealed that the 70s loop is no more dynamic than other visible dynamic regions of the protein. Typically the reason that resonances are not observed is that they are transitioning between conformational states with different chemical shifts on an intermediate time scale in the NMR experiment. It is interesting that the AMD simulations do not show any evidence of large conformational fluctuations in this loop, in fact, it appears uniformly structured (Figure 4.5 (c), 70s loop shown in blue). The results of the AMD simulations demonstrate that the absence of peaks in the 70s loop is not due wholly to dynamics as previously suggested (Lechtenberg*, et al.*, 2010) and illustrates the necessity of quantitative study of dynamics in thrombin.

D. CONCLUSION

Due to its size and complex folding pathway, direct measurements of backbone dynamics in thrombin have not been reported previously. In this first report, we wished to explore the solution structure of active human α-thrombin with a small inhibitor bound at the active site. Parts of the light chain, the entire γ-loop, and tips of the 60s and 90s loops and the Na^+-binding site had order parameters lower than 0.8 indicative of substantial

dynamics on the ps-ns timescale. TROSY Hahn-echo experiments revealed a substantial number of residues undergoing relaxation due to chemical exchange including residues in the light chain, the beginning of the heavy chain, one residue in the 90s loop, one residue in the strand that connects the 70s to the 90s loop, much of the γ-loop, two residues in regions proximal to the light chain, and three residues in the β-sheet core. The emerging picture was thus of a very dynamic molecule.

To obtain a more complete quantitative description of the dynamic properties of the system across multiple time scales, MD simulations were employed. AMD simulations were calibrated using RDC data and these extended MD simulations provided a more complete picture of temporally heterogeneous dynamic behavior of observed residues. Together the experiments and simulations yield an unprecedented molecular ensemble view of thrombin that truly represents the solution structure, revealing that PPACK-thrombin possesses a largely rigid core corresponding to the evolutionarily conserved protease with highly dynamic and less conserved loop appendages. The order parameters obtained from the AMD simulations (S^2_{RDC} values) were often much lower than the S^2 values calculated from the relaxation data revealing the true extent of thrombin's motions on longer timescales. Many residues had high S^2 values and much lower S^2_{RDC} values, but did not appear to be undergoing R_{ex} (Figure 4.5 (c)). These residues may be undergoing motions slower than the slow limit of the relaxation calculated S^2 values (τ_c, ~20ns) but faster than the fast limit of the chemical exchange regime (~100μs). This is strong evidence of motions in the supra-τ_c regime (Lange*, et al.*, 2008).

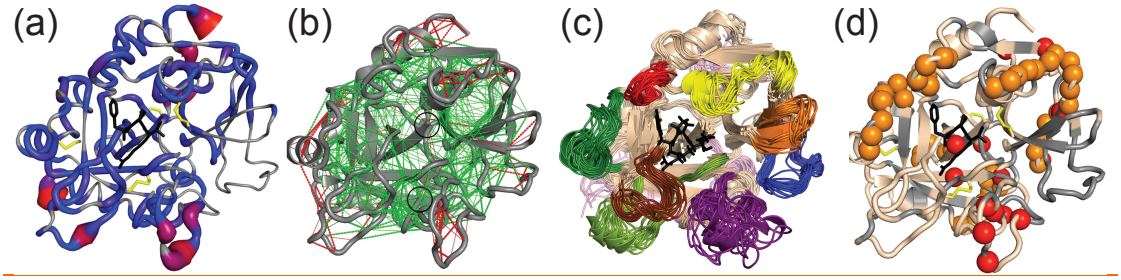


Figure 4.5 (*a*) Structure of PPACK-inhibited human α-thrombin depicting the ps-ns motions as determined from the backbone relaxation experiments. S^2 values are depicted as a backbone sausage model where the scale was red ($S^2 \leq 0.5$) to blue ($S^2 \geq 0.8$) with unobserved residues in gray. (*b*) Structure of PPACK-inhibited human α-thrombin showing the minimally frustrated contacts (*green*) and the highly frustrated contacts (*red*); the highly frustrated oxyanion hole and the minimally frustrated N-terminus of the heavy chain are circled. (*c*) Structural ensemble of PPACK-inhibited human α-thrombin calculated using AMD that best matched the RDC data. The loops are colored: light chain (*pink*), 30s loop (orange) 60s loop (*yellow*), 70s loop (*blue*), 90s loop (red), γ-loop (*purple*), 170s loop (dark green) 180s loop (*light green*), Na^+-binding site (*brown*). (*d*) Structure of PPACK-thrombin depicting residues with measured Rex > 6 s^{-1} (*red*), those with predicted supra-τ_c motions (*orange*), and unobservable residues (*gray*).

In addition, the AMD simulations revealed the dynamic behavior of the residues that were invisible in the NMR experiments.

Previous dynamic studies of enzymes have mostly observed that loops surrounding the active site display a hinging motion on the millisecond time scale between an open and closed state. Our combined relaxation experiments and simulations revealed multiple time scale dynamics of the thrombin loops, thus differing from this paradigm. The complex temperature dependence of resonance peak intensities in these regions is further evidence of the temporal heterogeneity of the loop motions. The loops do not appear to hinge, but rather they appear to be flexible appendages that extend from the rigid core. One could say that thrombin resembles an anemone in the sense that its active site is surrounded by loops that exhibit both fast and long time scale dynamics.

It has long been proposed that one feature of zymogen activation is the further structuring of four structural segments that were observed to be deformed and flexible in crystal structures of the zymogen form: residues light chain up to residue 19, the γ-loop (residues 142–152; 178-195), the 180s loop (residues 184–193; 225-239), and the Na^+ binding site (residues 216–223; 264-271), which together form the activation domain (Huntington, 2012). These regions also display temporally heterogeneous dynamics. Our results show that, in fact, zymogen activation does not result in ordering of these regions (marked with horizontal bars in Figure 4.4 (a)) since they remain highly dynamic even in the active form albeit at longer timescales. The tracks of residues undergoing motions on the supra-τ_c timescale appear to surround a track of residues undergoing R_{ex} (Figure 4.5 (d)). These apparent pathways of dynamic residues may be particularly important in the

allosteric communication between the anion binding exosites and the active site (Koeppe*, et al.*, 2005, Treuheit*, et al.*, 2011).

References

Baerga-Ortiz A, Rezaie AR & Komives EA (2000) Electrostatic dependence of the thrombin-thrombomodulin interaction. *J Mol Biol* **296**: 651-658.

Blackledge M, Cordier F, Dosset P & Marion D (1998) Precision and uncertainty in the characterization of anisotropic rotational diffusion by 15N relaxation. *J Am Chem Soc* **120**: 4538-4539.

Bode W, Turk D & Karshikov A (1992) The refined 1.9-A X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Sci* **1**: 426-471.

Braddock DT, Cai M, Baber JL, Huang Y & Clore GM (2001) Rapid identification of medium- to large-scale interdomain motion in modular proteins using dipolar couplings. *J Am Chem Soc* **123**: 8634-8635.

Case DA (2002) Molecular dynamics and NMR spin relaxation in proteins. *Accts Chem Res* **35**: 325-331.

Case DA, Darden TA, Cheatham I, T.E. *, et al.* (2008) AMBER 10. *University of California, San Francisco* **AMBER10**.

Cervantes CF, Markwick PRL, Sue SC, McCammon JA, Dyson HJ & Komives EA (2009) Functional dynamics of the folded ankyrin repeats of IkappaB alpha revealed by nuclear magnetic resonance. *Biochemistry* **48**: 8023-8031.

Clore GM, Driscoll PC, Wingfield PT & Gronenborn AM (1990) Analysis of the Backbone Dynamics of Interleukin- 1 ,6 Using Two-Dimensional Inverse Detected Heteronuclear 15N-'H NMR Spectroscopy. *Biochemistry* **29**: 7387-7401.

Ding K & Gronenborn AM (2003) Sensitivity-enhanced 2D IPAP, TROSY-anti-TROSY, and E.COSY experiments: alternatives for measuring dipolar ^{15}N-^{1}HN couplings. *J Magn Reson* **163**: 208-214.

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

Ferreiro DU, Hegler JA, Komives EA & Wolynes PG (2007) Localizing frustration in native proteins and protein assemblies. *Proc Natl Acad Sci U S A* **104**: 19819–19824.

Genheden S, Diehl C, Akke M & Ryde U (2010) Starting-condition dependence of order parameters derived from molecular dynamics simulations. *J Chem Theory Comput* **6**: 2176–2190.

Hamelberg D, Mongan J & McCammon JA (2004) Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys* **120**: 11919-11929.

Hamelberg D, de Oliveira CA & McCammon JA (2007) Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *J Chem Phys* **127**: 155102.

Huntington JA (2012) Thrombin plasticity. *Biochim Biophys Acta* **1824**: 246-252.

Jarymowycz VA & Stone MJ (2006) Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev* **106**: 1624-1671.

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

Lakomek NA, Carlomagno T, Becker S, Griesinger C & Meiler J (2006) A thorough dynamic interpretation of residual dipolar couplings in ubiquitin. *J Biomol NMR* **34**: 101-115.

Lange OF, Lakomek NA, Farès C*, et al.* (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**: 1471-1475.

Lechtenberg BC, Johnson DJ, Freund SM & Huntington JA (2010) NMR resonance assignments of thrombin reveal the conformational and dynamic effects of ligation. *Proc Natl Acad Sci U S A* **107**: 14087-14092.

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.

Losonczi JA, Andrec M, Fischer MW & Prestegard JH (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J Magn Reson* **138**: 334-342.

Mandel AM, Akke M & Palmer AG (1995) Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* **246**: 144-163.

Markwick PR, Bouvignies G, Salmon L, McCammon JA, Nilges M & Blackledge M (2009) Toward a unified representation of protein structural dynamics in solution. *J Am Chem Soc* **131**: 16968-16975.

Markwick PRL & McCammon JA (2011) Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Phys Chem Chem Phys* **13**: 20053-20065.

Markwick PRL & Nilges M (2011) Computational approaches to the interpretation of NMR data for studying protein dynamics. *Chem Phys* **396**: 124-134.

Masterson LR, Yu T, Shi L, Wang Y, Gustavsson M, Mueller MM & G. V (2011) cAMP-dependent protein kinase A selects the excited state of the membrane substrate phospholamban. *J Mol Biol* **412**: 155-164.

McConnell HM (1958) Reaction rates by nuclear magnetic resonance. *J Chem Phys* **28**: 430-431.

Niu W, Chen Z, Gandhi PS*, et al.* (2011) Crystallographic and kinetic evidence of allostery in a trypsin-like protease. *Biochemistry* **50**: 6301–6307.

Palmer AG (1994) NMR characterization of the dynamics of biomacromolecules. *Chem Rev* **104**: 3623-3640.

Prestegard JH, Al-Hashimi HM & Tolman JR (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Quart Rev Biophys* **33**: 371-424.

Salmon L, Bouvignies G, Markwick P & Blackledge M (2011) Nuclear magnetic resonance provides a quantitative description of protein conformational flexibility on physiologically important time scales. *Biochemistry* **50**: 2735-2747.

Salmon L, Bouvignies G, Markwick P*, et al.* (2009) Protein conformational flexibility from structure-free analysis of NMR dipolar couplings: quantitative and absolute determination of backbone motion in ubiquitin. *Angew Chem Int Ed Engl* **48**: 4154-4157.

Schurr JM, Babcock HP & Fujimoto BS (1994) A test for the model-free formulas. Effects of anisotropic rotational diffusion and dimerization. *J Magn Reson* **Series B 105**: 211-224.

Showalter SA, Johnson E, Rance M & Brüschweiler R (2007) Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. *J Am Chem Soc* **129**: 14146-14147.

Suel GM, Lockless SW, Wall MA & Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Mol Biol* **10**: 59-69.

Tjandra N & Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**: 1111-1114.

Tolman JR (2002) A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy. *J Am Chem Soc* **124**: 12020-12030.

Tolman JR, Flanagan JM, Kennedy MA & Prestegard JH (1997) NMR evidence for slow collective motions in cyanometmyoglobin. *Nat Struct Biol* **4**: 292-297.

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

Wang C, Rance M & Palmer AG (2003) Mapping chemical exchange in proteins with MW > 50 kD. *J Am Chem Soc* **125**: 8968-8969.

Woessner DE (1962) Nuclear spin relaxation in ellipsoids undergoing rotational Brownian motion. *J Chem Phys* **37**: 647-654.

Zweckstetter M (2008) NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc* **3**: 679-690.

Chapter IV, in full, is a reprint that the dissertation author was the principal researcher and author of. The material appears in *Biophysical Journal*. (**Fuglestad, B.,** Gasper, P.M., Tonelli, M., McCammon, J.A., Markwick, P.R. and Komives, E.A. (2012). The dynamic structure of thrombin in solution. *Biophys. J.,* **103**, 79-88)

Chapter 5

Dynamics of Apo-Thrombin

A. BACKGROUND

It has been suggested that the thrombin active site is malleable and a series of interactions is required to shuttle the active site loops to their active conformations (Di Cera, 2008). The most direct route to locking the thrombin active site in an active conformation is the ligation of the active site (Huntington & Esmon, 2003). Activation of the zymogen form, prothrombin, to the active form, thrombin, involves proteolytic separation of the light chain from the heavy chain. Crystallography studies have suggested that the newly formed N-terminus of the heavy chain buries itself into the protein, changing the conformation of the active site (Pozzi*, et al.*, 2011). Another key player in the proper formation of a procoagulant active site is sodium. Without sodium, thrombin exists in a marginally active "slow" form, while sodium binding, like zymogen activation, helps the active site achieve a conformation that can recognize and cleave procoagulant substrates. In contrast, sodium binding does not influence the catalytic activity of thrombin towards protein C, it's anticoagulant substrate.

The results of crystallography studies have interpreted the formation of an the active site due to activation or ligand binding as a change in conformation from an inactive state to an active state. In contrast, recent solution studies have suggested that the inactive states of thrombin are highly dynamic (Lechtenberg*, et al.*, 2010). Even upon activation, apo-thrombin remains highly dynamic and zymogen like. As described in Chapter 4, NMR and computational studies have demonstrated that even a fully active, sodium bound, and active site occupied thrombin still exhibits a large degree of structural fluctuations spanning the dynamic timescale from picoseconds all the way to

milliseconds (Fuglestad*, et al.*, 2012). These results indicate that the long-held view, that thrombin as an inactive molecule until the correct conformation is formed upon activation or ligand binding, is incomplete. Rather, thrombin seems to exist in a state of flux, with active site occupation inducing an active conformation of the surrounding loops, and proteolytic activation and sodium binding funneling the conformational fluctuations towards members of the ensemble that can bind procoagulant substrates.

The switch of activity from procoagulant to anticoagulant is also dependent on the interaction of thrombin and the cofactor thrombomodulin. Crystallography with active site inhibited thrombin (PDB ID IDX5, (Fuentes-Prior*, et al.*, 2000)) in complex with thrombomodulin shows little difference in the thrombin structure as compared to PPACK-thrombin. A recent computational study revealed networks of correlated motions on approximately μs timescales that directly connect the anti-coagulant activity dependent EGF4 domain of thrombomodulin to the active site of thrombin, revealing a pathway of allostery that is dependent on dynamics (Gasper*, et al.*, 2012). The paradigm of a conformational switch mechanism of thrombin activity and specificity seems to be switching to one that takes a broader view of thrombin being in a state of flux between different ensemble members with different properties (Huntington, 2012, Lechtenberg*, et al.*, 2012). With this in mind, I undertook an NMR study of apo thrombin in which the active site serine was mutated to methionine. Comparison of the results of NMR studies of the apo protein with the results already obtained on active site-liganded thrombin reveal the effects and implications of this type of dynamic mechanism.

B. MATERIALS AND METHODS

Details of the protein expression, refolding and purification were performed as described in Chapter 2 and Fuglestad *et. al* (Fuglestad*, et al.*, 2012). In brief, the catalytically inactive S195M mutation was achieved by quick-change mutagenesis in pET23(+). ^{2}H-^{15}N or ^{2}H-^{15}N-^{13}C labeled S195M-thrombin was expressed by *E. coli* in minimal media containing the appropriate isotopes. Protein inclusion bodies containing S195M-thrombin were isolated from soluble genomic proteins by centrifugation, washing the inclusion body pellets by resuspension in wash buffer, then repeating twice more. Protein was solubilized by a buffer containing guanidine-HCl and DTT. Glutathione mediated mixed disulfide bonds were formed after removal of DTT by Sephadex G25 chromatography. The protein was refolded by rapid dilution, then the misfolded isomers were separated from properly folded prethrombin-2(+18) by HiLoadS ion exchange chromatography. Prethrombin-2(+18) was activated by *Echis carinatus* venom, with performed the internal cleavage separating the light and heavy chains, and by previously prepared α-thrombin, which proteolytically removed the 18 amino acid tail. Differently activated thrombin species were separated by MonoS ion exchange chromatography. Isotope incorporation was tracked by MALDI-TOF and proper proteolytic activation was confirmed on reduced sample by MALDI-TOF. Samples were buffer exchanged into NMR buffer: 25 mM sodium phosphate, 150 mM sodium chloride, and 0.05% sodium azide, with 10% v/v D_2O added as a lock solvent.

All NMR experiments were performed on spectrometers equipped with cold-probes and at 25°C unless otherwise stated. Details of the experimental procedures for

resonance assignments are outlined in Chapter 3 and Fuglestad *et. al*. Experiments performed for resonance assignment were: HNCO, and HN(CO)CA at UCSD Pharmacology on a Buker Avance III 600 MHz, HN(CA)CO at NMRFAM with a Varian NMR system 600iii, TROSY-HN(CO)CACB and TROSY-HNCA were performed at NMRFAM with a Varian VNS 800 MHz and NOESY-^{1}H-^{15}N-TROSY was performed with the UCSD Chemistry and Biochemistry Varian 800 MHz. Some assignments were transferred from PPACK-thrombin and additional assignments were made and assignment transfers were confirmed with the 3D experimental data.

Dynamics experiments performed for the calculation of order parameters were R_1, R_2, and ^{15}N-{^{1}H}NOE as detailed in Chapter 4 and Fuglestad *et. al*. Relaxation rates were calculated by CurveFit with the front-end module SPARKY2RATE. S^2 values were calculated with TENSOR2 with a random snapshot of the CMD simulation performed on apo-wild-type thrombin used as the structural model to fit relaxation data to a rotational diffusion model. R_{ex} was measured with the TROSY Hahn-echo experiment (Wang*, et al.*, 2009).

Molecular dynamics simulations were performed and analyzed by Phineus Markwick and Paul Gasper as detailed in Chapter 4, Fuglestad *et. al.* (Fuglestad*, et al.*, 2012)and Gasper *et. al*(Gasper*, et al.*, 2012). CMD and AMD simulations were performed on wild-type apo-thrombin. AMD boost parameters that best recapitulated RDC data on PPACK-thrombin (Fuglestad*, et al.*, 2012), which have been proven in previous studies to be dependent only on the size of the system{Cervantes, 2009 #2;Markwick, 2009 #10}.

C. RESULTS AND DISCUSSION

The effects of removing PPACK from the active site of thrombin on the observable resonances and how the resonances of the apo-thrombin were assigned are discussed in Chapter 3.

1.Fast timescale dynamics

In order to elucidate general flexibility of thrombin with an unoccupied active site, relaxation experiments were performed and S^2 values were calculated (Figure 5.1). Order parameters calculated from NMR relaxation experiments are limited to motions that are faster than the tumbling time of the molecule, τ_c, which is approximately 20 ns for thrombin. The order parameters for S195M calculated from the NMR relaxation data compared to those for PPACK thrombin were qualitatively similar (Figure 5.1, black circles for apo-thrombin, red circles for PPACK-thrombin). For S^2 values calculated on PPACK- thrombin and apo-thrombin from CMD simulations (Figure 5.1, black line for apo-thrombin, red line for PPACK-thrombin), a comparison revealed that the only significant differences were observed in the 30s and 60s loops with residues 57-60 (36-38), excluding P59 (37). Both of these loops had higher S^2 values in the PPACK-liganded form than in the apo form of thrombin with S^2 values in the 30s loop being higher by 0.10 - 0.14 and in the 60s loop being higher by 0.05 - 0.16. The results suggest that on the fast time scale, the effect of active site ligation is minimal, except for the 30s and the 60s loops which exhibit some degree of rigidification in this timescale.

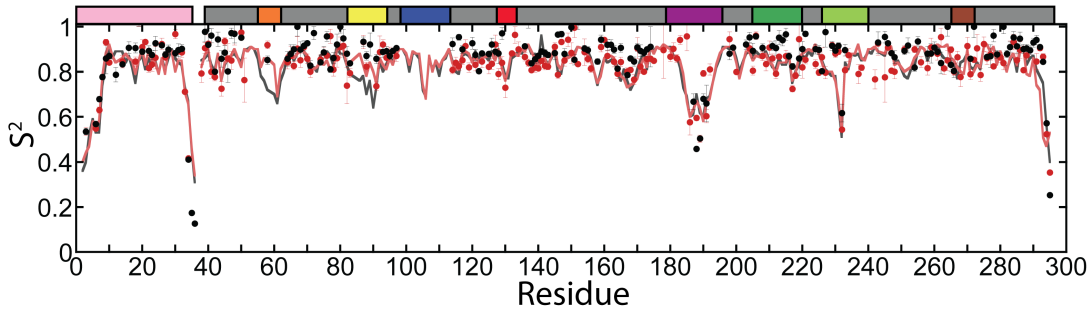


Figure 5.1 Order parameter comparison of PPACK-thrombin and apo-thrombin. Order parameters calculated from NMR relaxation experiments for S195M-thrombin mutant representing the apo form are depicted in black circles and those for PPACK-thrombin are depicted as red circles. Order parameters calculated from CMD simulations for apo-thrombin are depicted in a black line and those for PPACK-thrombin in a red line.

2. **Accelerated Molecular Dynamics**

Residual dipolar couplings (RDCs), which are ensemble averaged up to the chemical shift coalescence limit, provide useful NMR experimentally-derived data for determining the ensemble of structures that best represents the dynamic ensemble of a protein {Salmon, 2009 #33;Cervantes, 2009 #8}. When the experimentally derived RDCs are compared to RDCs that are back calculated from ensembles of structures generated from AMD simulations, the acceleration level that provides the most realistic representative structural ensemble can be identified. Using this approach, we have previously obtained an ensemble of PPACK-thrombin structures that accurately reproduced experimental RDCs {Fuglestad, 2012 #2}. Order parameters (S^2) for N-H bond vectors calculated from this ensemble are representative of motions on longer time scales than are accessible from NMR backbone relaxation measurements, which are limited by the molecular tumbling time {Fuglestad, 2012 #2}. A comparison of order parameters obtained from AMD simulations of apo-thrombin (Gasper*, et al.*, 2012) to those obtained previously for PPACK-thrombin {Fuglestad, 2012 #2} is shown in Figure 5.2. Most of the active site loops in both forms are highly flexible, yet a marked decrease in flexibility is observed upon active site ligation with PPACK (Figure 5.2 (a)). As expected, PPACK ligation caused significant ordering ($\Delta S^2>0.1$) of residues that directly contact the PPACK Arg side chain: residues 192 (238), 215 (263) and 216 (264). The loops that surround the active site also experience significant ordering ($\Delta S^2>0.1$) including the 60s loop residues 60D-60H (86-90), the γ-loop residues 141, 143, 146-148, 149A, 149B, 149E-151, and 153 (177, 179, 182-184, 186, 187, 190-192, and 194), the 180s loop residues 185 and 186D (227 and 232).

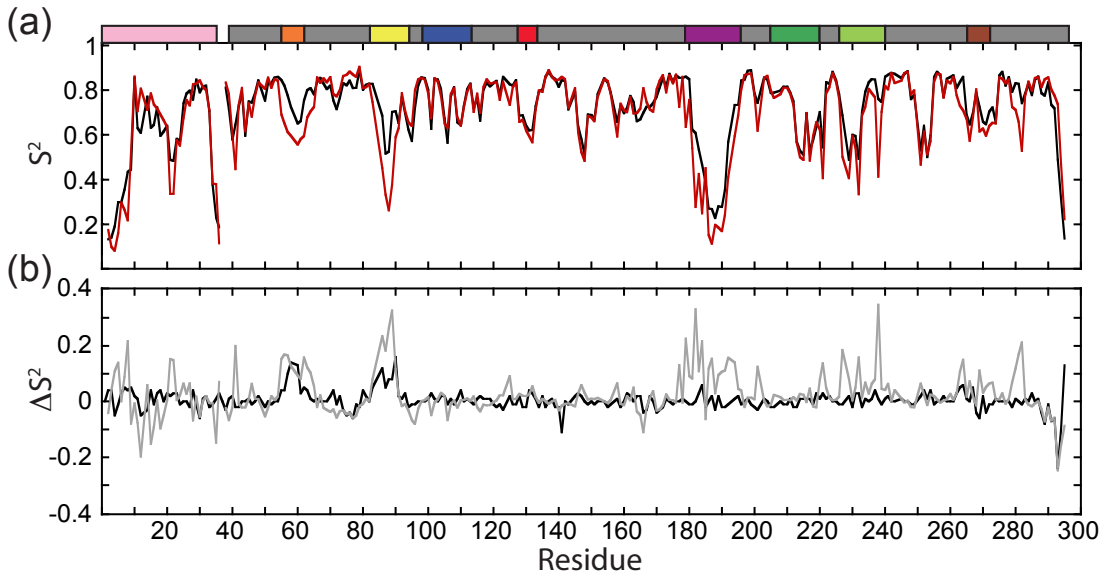


Figure 5.2 (a) Order parameters calculated from the AMD simulations for Apo-thrombin are depicted in red and those for PPACK-thrombin are in black. (b) Differences between apo-thrombin order parameters and PPACK order parameters calculated from CMD simulations in black and from AMD simulations in gray.

Some regions distal to the PPACK also showed significant ordering (ΔS^2>0.1) upon PPACK ligation including the light-chain residues 1E, 1D, 1A, 13 and 14 (4, 5, 8, 21 and 22); the 30s loop residues 34-36A, and 40-42 (55-58 and 62-64); and residues 232-234 (280-282) in the C-terminal helix. Other residues that showed significant ordering (ΔS^2>0.1) upon PPACK ligation included residue 20 (41) near the N-terminus of the heavy chain, residue 160 (201) in the strand leading to the 170s loop and 180 (221) in the 170s loop, and residue 221 (269) of the sodium binding loop. Interestingly, the 90s loop that directly contacts the Phe of the PPACK did not become more ordered upon PPACK ligation. The differences between PPACK-thrombin and apo-thrombin were calculated for the CMD calculated order parameters and the AMD calculated order parameters (FIG 5.2 (b)). These differences show that for 30s loop, the majority of the rigidification upon active site occupation occurs on the short timescale defined by the CMD simulation time (20 ns). For the 60s loop, rigidification occurs in both the ps-ns timescale as well as the μs-ms timescale accessed by the AMD simulations. Remarkably, the region adjacent to the N-terminus, the γ-loop, the 170s loop, the 180s loop, the and the residues leading into the C-terminal helix all become more rigid exclusively on the longer timescale. This suggests that active site occupation mostly limits fluctuations on timescales above 20 ns.

3. Correlated Motion Analysis

To identify residues undergoing correlated motions on intermediate time scales, AMD simulations were performed by Paul Gasper and Phineus Markwick of the McCammon group and were optimized based on previous work comparing experimental

RDCs to those back calculated from AMD simulations carried out at different acceleration levels {Fuglestad, 2012 #2;Gasper, 2012 #7}. The analysis of apo-thrombin revealed correlated motions between the active site loops, exosite 1, and other distal sites. In particular, the entire γ-loop (178-192) appears to undergo strongly correlated motions with the light-chain residues 1H-1D and 12-14C (1-5 and 20-25), the catalytic triad; H57, D102, S195 (79, 135, 241), the 60s loop (82-90), the 70s loop (100-110), the 90s loop (128-135), the surface strand under the 70s loop (145-151), the 170s loop (205-220), and the 180s loop (225-240) (Fig. 5.3, lower triangle). Whereas these correlated motions appeared to involve the entire γ-loop in apo-thrombin, only the tip of the γ-loop (residues 146-149E (182-190)) appears to be undergoing the same set of motions in PPACK-thrombin (Fig. 5.3 upper triangle). In apo-thrombin, the 170s and to a lesser extent the 180s loop, are strongly correlated to the γ-loop, are also correlated with the light-chain and catalytic residues. Upon PPACK ligation, most of these correlated motions are lost. In apo-thrombin, the 30s loop and the 60s loop are weakly correlated, but the 30s loop is not correlated to the 70s and 90s loops. Correlated motions between the 30s and 60s loops are stronger in PPACK thrombin and these extend to the 70s and 90s loops (Fig. 5.3).

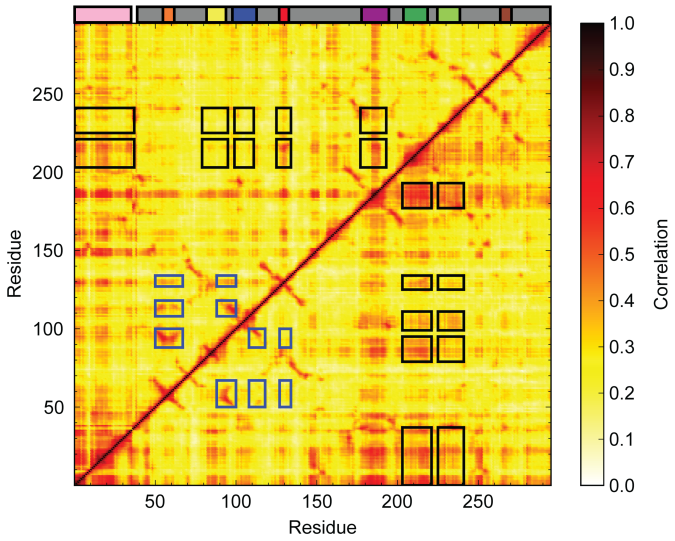


Figure 5.3 Correlated motions plot for apo-thrombin in the lower triangle and for PPACK-thrombin in the lower triangle. Loop regions with a large degree of correlated motions are boxed.

4. Chemical exchange dynamics

As discussed in Chapter 3, the thrombin mutant S195M, representing the apo form of the enzyme, undergoes a larger degree of resonance line broadening than PPACK-thrombin, suggesting that the occupation of the thrombin active site by PPACK greatly decreases motions in the NMR intermediate time scale, μs-ms. In order to measure R_{ex} in visible residues and further map residues undergoing structural fluctuations in this timescale, R_{ex} measurements were performed using the TROSY Hahn-Echo experiment at 800MHz {Wang, 2003 #9} (Fig. 5.4). Measurements on PPACK-thrombin revealed residues with significant R_{ex} (>6 s^{-1}) discussed in Chapter 4: 1F and 3 (3 and 11) of the light-chain; 23 and 27 (44 and 48); 66 (97) leading up to the 70s loop; 87 (119) in the surface strand that connects the 70s loop to the 90s loop; 144, 147, 149a, 149c, and 151 (180, 183, 186, 188 and 192) of the γ-loop; 138, 157, and 199 (174, 198, and 245) in the core β-sheet; and 233 (281) that forms a kink in the C-terminal helix {Fuglestad, 2012 #2}. Residues in the apo S195M thrombin having significant R_{ex} (>6 s^{-1}) included: 1D (5) and 14F (28) of the light-chain; 18 (39), 22 (43), 23 and 27 (44 and 48) of the region C-terminal of the activation proteolysis site; 42 and 45 (64 and 67) in the N-terminal β-barrel that forms an interface with the C-terminal β-barrel, 57- 59 (79- 81) near the 60s loop; 60f (88) within the 60s loop; 66 and 82 (97 and 114) near the 70s loop; 88 and 89 (120 and 121) in the b-strand connecting the 70s and 90s loops; 97 (129) of the 90s loop; the Asp that is part of the catalytic triad, 102 (135), 112 (145); 149c, and 149e (188, and 190) of the γ-loop; 159, 160, 181, and 229 (200, 201, 222, and 277) of the C-terminal β-barrel; 168, 172 and 176 (209, 213 and 217) of the 170s helix and loop; 186a, 186b

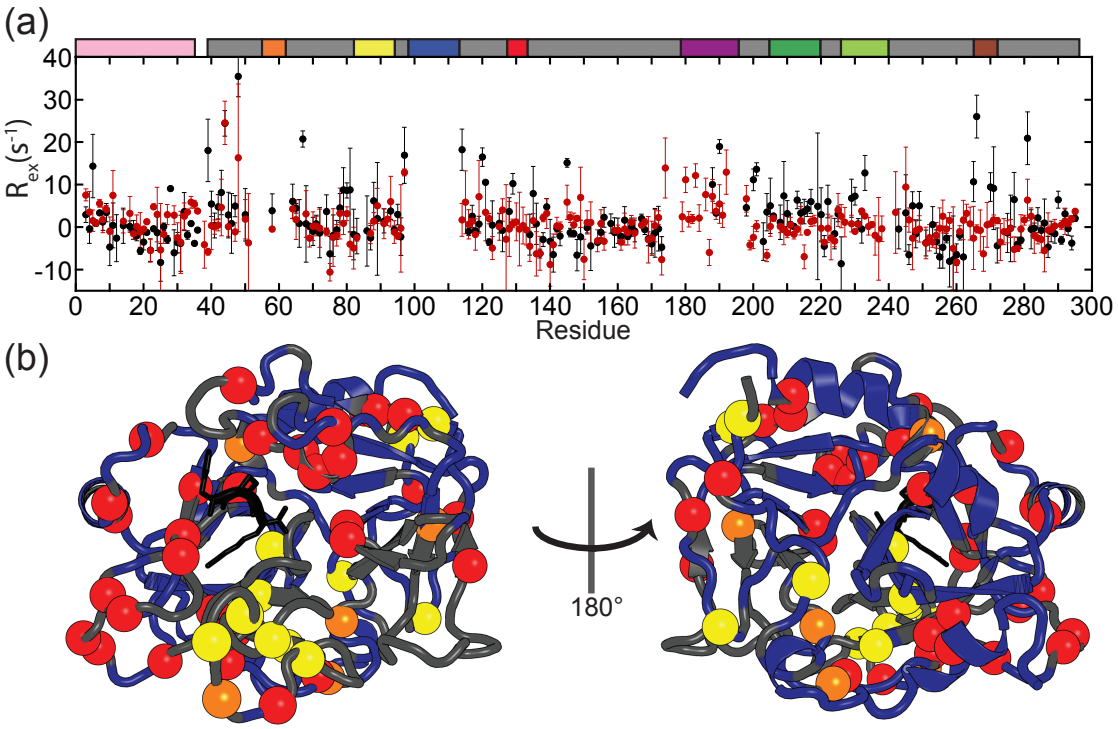


Figure 5.4 Comparison of R_{ex} measurements for PPACK-thrombin and apo-thrombin. (a) R_{ex} values calculated from the TROSY Hahn-echo experiment with apo-thrombin in black and PPACK-thrombin in red. (b) Structural representation of residues undergoing $R_{ex} > 6\ s^{-1}$. Depicted in red are residues with R_{ex} in apo-thrombin only, in yellow are residues with Rex in PPACK-thrombin only, and in orange are residues that undergo R_{ex} in both forms.

and 187 (229, 230 and 233) of the 180s loop; 217, 219, 222, 223 (265, 266, 270, and 271) of the Na^+ binding loop; and 242 (290) of the C-terminal helix. Regions that show significant R_{ex} only in the apo form include the 60s loop, the 90s loop, the Na+ binding loop, and residues 82, 89 and 97 (114,121 and 129) in the surface beta strand connecting the 90s loop to the 70s loop in exosite 1.

D. Conclusions

Analyses of dynamic motions in apo-thrombin across the timescale are presented here. From NMR and CMD calculated order parameters, both with a limit of approximately 20 ns, the structural fluctuations of apo-thrombin on this short timescale seem to match well to thrombin with the active site occupied by PPACK. RDC tuned AMD simulations accessed order parameters on a longer timescale and showed a large degree of motions appearing specifically on that time scale. With the inoccupation of the active site, motions were shown to be correlated between functional loops by correlated motion analysis. R_{ex} measurements provided an experimental measure of residues fluctuating on the μs-ms timescale. Along with the resonance line-broadening results presented in Chapter 3, these results suggest that thrombin samples a wide ensemble on μs-ms timescales that may be narrowed by ligand binding. This stands in stark contrast of the paradigm that thrombin exists in specific conformations that determine whether it is active and which substrates it will recognize; these results reinforce the paradigm shift that states that dynamics provide a mechanism for shuttling thrombin between various activity and substrate recognition states.

E. FUTURE DIRECTIONS

In the near term, the nature of the μs-ms fluctuations in thrombin will be characterized by relaxation dispersion measurements by NMR (Loria*, et al.*, 2008). TROSY-selected $R_1\rho$ (Igumenova & Palmer, 2006) and TROSY-CPMG (Weininger*, et al.*, 2012) are promising experimental avenues for these measurements and with rigorous analyses (and luck) the rate of conformational fluctuations and the population differences between ground and excited states may be elucidated.

References

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

Fuentes-Prior P, Iwanaga Y, Huber R*, et al.* (2000) Structural basis for the anticoagulant activity of the thrombin-thrombomodulin complex. *Nature* **404**: 518-525.

Fuglestad B, Gasper PM, Tonelli M, McCammon JA, Markwick PR & Komives EA (2012) The dynamic structure of thrombin in solution. *Biophys J* **103**: 79-88.

Gasper PM, Fuglestad B, Komives EA, Markwick PR & McCammon JA (2012) Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proc Natl Acad Sci U S A* **109**: 21216-21222.

Huntington JA (2012) Thrombin plasticity. *Biochim Biophys Acta* **1824**: 246-252.

Huntington JA & Esmon CT (2003) The molecular basis of thrombin allostery revealed by a 1.8 A structure of the "slow" form. *Structure* **11**: 469-479.

Igumenova TI & Palmer AG (2006) Off-resonance TROSY-selected R 1rho experiment with improved sensitivity for medium- and high-molecular-weight proteins. *J Am Chem Soc* **128**: 8110-8111.

Lechtenberg BC, Freund SM & Huntington JA (2012) An ensemble view of thrombin allostery. *Biol Chem* **393**: 889-898.

Lechtenberg BC, Johnson DJ, Freund SM & Huntington JA (2010) NMR resonance assignments of thrombin reveal the conformational and dynamic effects of ligation. *Proc Natl Acad Sci U S A* **107**: 14087-14092.

Loria JP, Berlow RB & Watt ED (2008) Characterization of enzyme motions by solution NMR relaxation dispersion. *Acc Chem Res* **41**: 214-221.

Pozzi N, Chen Z, Zapata F, Pelc LA, Barranco-Medina S & Di Cera E (2011) Crystal structures of prethrombin-2 reveal alternative conformations under identical solution conditions and the mechanism of zymogen activation. *Biochemistry* **50**: 10195-10202.

Wang Y, Berlow RB & Loria JP (2009) Role of loop-loop interactions in coordinating motions and enzymatic function in triosephosphate isomerase. *Biochemistry* **48**: 4548-4556.

Weininger U, Respondek M & Akke M (2012) Conformational exchange of aromatic side chains characterized by L-optimized TROSY-selected ^{13}C CPMG relaxation dispersion. *J Biomol NMR* **54**: 9-14.

Appendix

Appendix 1.1: PPACK-thrombin assignment table

#	Type	Chym#	$\delta\ N$	$\delta\ H^N$	$\delta\ C'$	$\delta\ C^\alpha$	$\delta\ C^\beta$
1	T	1H					
2	F	1G			176.378	59.038	38.863
3	G	1F	110.919	8.568	174.584	44.749	
4	S	1E	117.42	8.272	175.087	58.559	63.314
5	G	1D	110.788	8.389	174.315	45.026	
6	E	1C	120.882	8.273	177.717	57.134	29.197
7	A	1B	123.922	8.37	178.185	53.528	17.832
8	D	1A	118.148	8.24	176.525	52.856	40.43
9	C	1	118.266	7.726	175.089	54.947	41.997
10	G	2	104.591	9.168	172.386	47.74	
11	L	3	120	7.206	175.026	52.465	42.335
12	R	4	120.714	8.701	176.527	54.287	26.775
13	P	5			176.992	64.813	33.065
14	L	6	109.639	7.76	174.505	53.883	39.92
15	F	7	113.579	7.062	176.574	58.672	39.484
16	E	8	125.819	8.992	180.279	62.877	28.375
17	K	9	113.697	7.75	176.642	58.362	30.443
18	K	10	116.891	7.193	174.732	54.71	32.583
19	S	11	112.417	7.663	173.11	58.744	60.95
20	L	12	120.38	7.837	176.393	52.735	43.844
21	E	13	121.932	8.423	177.431	54.39	31.411
22	D	14	125.442	9.7	176.821	53.085	40.203
23	K	14A	117.466	10.046	178.651	59.159	32.605
24	T	14B	104.328	8.104	175.833	60.627	69.602
25	E	14C	123.272	8.079	178.028	59.693	28.927
26	R	14D	122.793	8.908	177.504	58.635	27.749
27	E	14E	118.76	7.552	179.847	57.983	28.293
28	L	14F	117.954	7.092	178.555	57.375	38.261
29	L	14G	120.39	7.166	180.078	56.97	40.054
30	E	14H	116.857	8.358	177.821	57.723	28.339
31	S	14I	115.1	7.189	174.079	59.675	63
32	Y	14J	121.282	6.589	176.238	53.504	34.642
33	I	14K	119.035	7.158	176.155	60.877	38.208
34	D	14L	124.327	8.283	176.8	54.172	40.781
35	G	14M	110.703	8.352	173.506	45.322	
36	R	15	126.052	7.764	181.252	57.19	30.202
37	I	16			171.593	56.593	36.345
38	V	17	130.546	9.047	174.966	60.064	32.969
39	E	18	120.569	9.264	175.472	57.119	25.217

40	G	19	106.218	7.922	172.899	43.749	
41	S	20	112.877	8.459	172.919	56.038	65.928
42	D	21	121.132	8.406	177.021	54.353	39.512
43	A	22	125.802	8.841	177.611	51.311	18.22
44	E	23	121.093	7.251	176.898	55.659	31.474
45	I	24	125.272	8.628	177.618	61.992	35.358
46	G	25	113.444	9.105	175.347	45.126	
47	M	26	117.32	7.577	175.777	56.509	33.561
48	S	27	114.749	8.882			
49	P	28			174.223	65.108	29.275
50	W	29	118.1	5.844	177.409	53.802	27.375
51	Q	30	125.49	8.185	176.662	57.843	
52	V	31					
53	M	32					
54	L	33					
55	F	34					
56	R	35					
57	K	36			176.733	59.397	32.732
58	S	36A	110.989	7.907	172.836	54.638	64.052
59	P	37					
60	Q	38					
61	E	39					
62	L	40					
63	L	41			177.147	55.106	43.439
64	C	42	116.014	7.668	174.27	55.734	54.377
65	G	43	106.659	8.796	170.959	43.341	
66	A	44	117.194	7.921	175.518	50.85	19.54
67	S	45	114.427	9.716	172.404	55.551	65.835
68	L	46	126.543	9.404	175.616	54.453	43.948
69	I	47	118.097	8.137	175.066	60.917	38.501
70	S	48	116.618	8.408	171.782	57.142	63.678
71	D	49	115.608	8.854	176.046	55.326	39.228
72	R	50	114.116	7.89	173.529	53.89	32.019
73	W	51	118.873	7.199	174.459	56.413	32.09
74	V	52	121.153	9.096	173.9	59.043	35.026
75	L	53	129.755	9.539	176.334	53.242	45.292
76	T	54	120.149	9.113	172.818	60.318	69.614
77	A	55	123.03	8.008	177.042	50.06	18.161
78	A	56	127.519	6.412	180.413	55.51	15.521
79	H	57	116.126	10.392	174.389	58.055	26.297
80	C	58	116.582	6.803	173.127	59.285	44.253

81	L	59	116.743	7.38	174.625	54.847	42.119
82	L	60	119.512	8.093	174.686	53.223	43.04
83	Y	60A	125.674	9.202	171.284	57.876	36.639
84	P	60B					
85	P	60C			176.857	65.314	29.221
86	W	60D	113.579	7.062	174.835	52.582	30.128
87	D	60E	119.156	8.569	174.218	54.837	38.858
88	K	60F	115.801	6.656	175.574	53.673	31.384
89	N	60G	127.44	8.178	173.291	52.35	37.097
90	F	60H	122.409	8.635	175.526	57.382	40.917
91	T	60I	111.327	9.171	175.724	59.439	71.434
92	E	61	118.768	10.018	177.041	60.326	26.98
93	N	62	114.912	8.063	175.89	54.048	38.331
94	D	63	116.254	7.836	176.322	55.451	43.579
95	L	64	118.105	6.837	173.846	53.632	46.43
96	L	65	116.745	8.944	175.28	54.097	44.609
97	V	66	112.473	8.494	174.669	58.408	
98	R	67					
99	I	68					
100	G	69					
101	K	70					
102	H	71					
103	S	72					
104	R	73					
105	T	74					
106	R	75					
107	Y	76					
108	E	77					
109	R	77A					
110	N	78					
111	I	79					
112	E	80			175.461	54.859	32.453
113	K	81	121.548	9.016	174.974	53.338	34.913
114	I	82	123.667	8.528	176.029	59.425	38.705
115	S	83	120.627	8.751	173.024	57.623	65.27
116	M	84	123.349	8.786	176.401	53.514	32.557
117	L	85	121.822	9.396	175.172	53.49	41.905
118	E	86	122.567	9.189	175.45	56.486	31.12
119	K	87	114.208	7.038	173.882	54.461	35.943
120	I	88			173.872	59.919	40.119
121	Y	89	124.109	9.351	174.551	56.93	39.719
122	I	90	125.777	9.027	176.122	59.92	38.383

123	H	91	129.911	6.838	175.552	57.416	30.133
124	P	92			178.437	64.059	31.52
125	R	93	121.103	11.186	176.03	53.637	28.831
126	Y	94	124.325	7.53	175.068	56.369	38.066
127	N	95	131.898	8.754	174.172	49.874	38.159
128	W	96	118.956	6.383	176.066	55.579	27.306
129	R	97	120.71	6.881	176.545	58.14	29.548
130	E	97A	115.832	7.147	176.462	56.571	30.828
131	N	98	113.263	7.203	171.463	51.894	38.257
132	L	99	110.635	7.702	176.865	55.116	38.686
133	D	100	116.829	7.131	176.311	55.449	41.385
134	R	101	119.132	8.716	174	57.052	24.634
135	D	102	114.369	6.849	175.025	52.754	36.9
136	I	103	126.766	8.294	169.531	60.265	41.362
137	A	104	127.218	8.254	174.572	50.362	23.268
138	L	105	121.436	9.471	176.217	52.583	47.019
139	M	106	118.144	9.607	172.689	53.71	34.814
140	K	107	126.237	8.635	176.443	53.423	33.883
141	L	108	127.86	8.73	177.053	54.411	41.034
142	K	109	120.739	8.056	175.574	58.691	32.084
143	K	110	115.398	7.344	172.645	52.188	33.551
144	P	111			176.55	61.748	31.169
145	V	112	122.352	8.782	173.682	59.2	32.865
146	A	113	127.521	7.92	177.356	50.511	18.246
147	F	114	122.24	7.885	176.021	54.108	37.02
148	S	115	118.63	9.777	174.985	56.815	66.409
149	D	116	119.066	8.66	174.051	56.191	38.902
150	Y	117	113.097	7.749	174.469	55.978	40.158
151	I	118	119.01	6.964	173.664	60.771	39.991
152	H	119	126.075	7.513	172.526		34.49
153	P	120			176.09	62.274	31.424
154	V	121	122.729	8.073	173.039	59.73	32.204
155	C	122	119.963	6.753	174.915	53.499	40.956
156	L	123	122.108	9.266	175.055	51.378	40.5111
157	P	124					
158	D	125	118.053	7.101	174.972	51.198	41.895
159	R	126	120.522	8.595	178.191	59.32	29.039
160	E	127	119.249	8.403	179.229	58.955	28.016
161	T	128	119.275	7.907	175.877	66.6	67.169
162	A	129	121.685	7.824	178.556	54.865	16.504
163	A	129A	118.513	7.78	179.932	53.909	17.446
164	S	129B	110.891	7.584	175.466	60.778	63.392

165	L	129C	116.245	7.573	176.412	55.043	41.349
166	L	130	118.846	7.234	175.63	52.775	37.785
167	Q	131	122.927	7.296	174.028	52.708	30.001
168	A	132	123.244	8.219	177.827	52.833	17.166
169	G	133	111.249	8.93	174.135	44.699	
170	Y	134	123.771	8.059	176.238	55.665	35.884
171	K	135	119.682	8.953	175.359	55.547	35.678
172	G	136	107.352	8.406	170.57	43.496	
173	R	137	120.019	8.603	174.509	53.837	34.125
174	V	138	128.746	9.349	173.747	60.468	33.907
175	T	139	111.599	6.722	175.42	56.288	71.04
176	G	140	106.661	8.037			
177	W	141			173.425		
178	G	142	108.21	8.642	172.416	43.286	
179	N	143	116.084	8.199	176.398	53.599	38.26
180	L	144	120.217	8.273	176.32	54.516	43.073
181	K	145	113.577	7.464	175.358	53.529	36.482
182	E	146	121.421	9.073	176.513	58.758	29.566
183	T	147	112.105	7.85	173.061	60.618	69.94
184	W	148	124.713	8.277	173.881	61.913	28.944
185	T	149	116.807	8.665	174.143	60.853	69.392
186	A	149A	125.384	7.749	177.543	52.651	18.085
187	N	149B	116.121	8.243	175.097	53.59	37.847
188	V	149C	119.319	7.767	176.685	62.157	32.024
189	G	149D	112.158	8.371	173.825	44.78	
190	K	149E	119.684	7.95	176.199	55.329	32.166
191	G	150	108.354	7.916	172.943	44.825	
192	Q	151	119.003	7.467	173.758	52.429	28.772
193	P	152					
194	S	153					
195	V	154					
196	L	155					
197	Q	156			175.376	52.217	33.131
198	V	157	118.944	8.953	173.296	58.524	35.225
199	V	158	126.134	8.758	170.611	60.291	34.942
200	N	159	126.611	8.641	175.394	50.874	39.006
201	L	160	121.161	8.95	173.752	51.034	46.602
202	P	161			175.897	60.784	30.763
203	I	162	122.409	8.632	176.601	62.177	36.528
204	V	163	133.626	7.907	174.036	61.607	32.763
205	E	164	126.368	9.586	177.348	56.841	28.952
206	R	165	124.431	8.856	174.911	61.659	26.9

207	P	166			178.686	65.692	30.447
208	V	167	117.436	6.678	179.427	64.82	30.106
209	C	168	118.529	7.684	178.737	56.64	37.754
210	K	169	121.721	8.689	177.731	59.298	31.834
211	D	170	116.054	7.757	177.283	55.417	40.079
212	S	171	114.12	7.783	172.477	59.862	64.519
213	T	172	118.167	7.527	171.241	59.623	69.005
214	R	173	123.992	8.764	178.059	55.025	29.354
215	I	174	127.009	8.118	174.753	61.335	35.593
216	R	175	127.623	8.162	175.115	56.019	29.435
217	I	176	123.712	7.726	176.129	59.629	38.534
218	T	177	110.735	7.22	175.396	58.629	71.972
219	D	178	118.052	8.821	176.138	55.208	39.542
220	N	179	112.392	8.07	175.085	53.539	37.427
221	M	180	124.257	8.23	174.254	54.199	36.288
222	F	181	117.097	8.884	171.862	56.356	40.411
223	C	182	121.3	8.845	171.846	54.785	41.327
224	A	183	120.726	8.491	176.303	50.934	23.219
225	G	184	109.976	8.763	170.838	43.943	
226	Y	184A	118.799	9.525	177.465	59.295	38.499
227	K	185	124.715	10.084	177.783	54.726	30.893
228	P	186			176.738	65.152	31.075
229	D	186A	113.606	8.405	177.12	53.941	39.454
230	E	186B	117.933	7.686	177.895	57.385	30.687
231	G	186C	107.649	8.048	173.375	46.068	
232	K	186D	119.41	6.654	173.067	54.468	33.301
233	R	187	118.175	8.008	174.484	54.229	31.793
234	G	188	106.982	8.602	173.677	44.131	
235	D	189	117.744	8.233	176.315	52.584	43.783
236	A	190	124.34	8.736	172.659	51.707	20.37
237	C	191	115.174	9.495	174.471	50.232	36.541
238	E	192	123.013	8.172	175.531	60.709	27.685
239	G	193					
240	D	194					
241	S	195			170.784	58.077	62.659
242	G	196	110.606	8.906	173.828	45.689	
243	G	197	104.606	7.946		43.591	
244	P	198			174.072	62.84	30.594
245	F	199	123.262	7.871	177.231	55.571	39.764
246	V	200	122.535	9.561	173.906	58.227	34.546
247	M	201	118.385	9.203	173.128	54.921	38.611
248	K	202	123.164	6.81	174.944	53.845	30.772

249	S	203	121.126	8.885		54.281	64.357
250	P	204			176.017	62.359	30.721
251	F	204A	118.648	7.623	176.813	57.936	38.213
252	N	204B	113.24	7.149	175.097	51.432	38.499
253	N	205	114.105	7.732	171.741	54.551	38.213
254	R	206	112.707	7.362	176.514	54.58	31.93
255	W	207	120.641	8.683	174.726	56.585	30.53
256	Y	208	118.334	8.987	175.765	55.856	42.892
257	Q	209	121.542	8.981	172.677	55.341	25.899
258	M	210	121.718	8.353	178.909	53.246	31.544
259	G	211	104.667	8.901	173.458	44.81	
260	I	212	122.004	7.847	176.315	59.962	41.345
261	V	213	127.086	9.274	175.912	65.19	31.169
262	S	214	122.591	7.564	174.317	59.453	63.316
263	W	215	114.453	7.915	172.573	56.138	32.126
264	G	216	107.194	8.844	171.139	44.777	
265	E	217	124.82	9.348	175.587	55.924	30.888
266	G	219	113.674	8.637	174.482	43.757	
267	C	220	117.693	9.106	173.044	55.121	45.632
268	D	221A	126.387	10.651	175.355	54.96	42.496
269	R	221	117.077	8.293	179.015	55.516	30.069
270	D	222	125.139	9.052	177.594	55.124	39.355
271	G	223	112.143	9.06	171.84	44.903	
272	K	224	116.787	6.951	174.949	51.244	32.973
273	Y	225	118.99	9.203	175.236	57.022	43.036
274	G	226	108.926	7.612	172.918	45.504	
275	F	227	118.588	8.548	177.168	57.424	39.458
276	Y	228	122.723	8.782	174.384	55.685	40.4
277	T	229	120.587	9.171	173.678	63.258	68.469
278	H	230	130.569	8.486	173.682	55.579	31.374
279	V	231	126.07	7.745	177.047	67.648	30.86
280	F	232	119	9.943	178.069	62.281	38.797
281	R	233	118.927	8.224	177.563	58.068	28.185
282	L	234	119.985	7.312	176.962	53.118	40.847
283	K	235	123.174	7.579	177.202	59.232	32.127
284	K	236	117.98	8.358	179.228	59.119	30.629
285	W	237	121.278	7.55	175.738	61.686	26.973
286	I	238	117.649	7.482	176.78	64.071	37.634
287	Q	239	114.349	8.186	176.942	57.845	28.197
288	K	240	119.503	7.397	178.896	58.287	31.027
289	V	241	120.227	7.302		65.602	30.031
290	I	242	118.082	7.088	178.557	64.388	36.328

291	D	243	119.362	8.159	178.019	55.93	39.997
292	Q	244	117.917	7.595	177.125	57.1	28.103
293	F	245	117.911	8.053	176.535	57.686	38.627
294	G	246	110.666	8.033	173.214	45.596	
295	E	247	125.928	7.936	181.282	57.849	30.199

Appendix 1.2: S195M-thrombin assignment table

#	Type	Chym#	$\delta\ N$	$\delta\ H^N$	$\delta\ C'$	$\delta\ C^\alpha$	$\delta\ C^\beta$
1	T	1H					
2	F	1G			176.517	58.915	39.018
3	G	1F	111.177	8.605	174.658	44.978	
4	S	1E	117.625	8.285	175.157	58.661	
5	G	1D	110.899	8.4	174.364	45.035	41.304
6	E	1C	120.902	8.248	177.761	57.061	29.298
7	A	1B	124.257	8.374	178.233	53.621	17.977
8	D	1A	118.172	8.268	176.598	52.921	40.539
9	C	1	118.455	7.751	175.086	54.98	41.816
10	G	2	104.872	9.17	172.321	47.775	
11	L	3	120.071	7.192	175.061	52.477	42.267
12	R	4	121.02	8.764	176.616	54.362	
13	P	5			177.063	64.89	33.252
14	L	6	109.787	7.814	174.541	53.918	40.071
15	F	7	113.712	7.083	176.638	58.739	39.577
16	E	8	125.985	9.017	180.336	62.947	28.64
17	K	9	113.763	7.742	176.682	58.373	30.62
18	K	10	116.94	7.194	174.761	54.789	32.729
19	S	11	112.5	7.66	173.157	58.802	60.874
20	L	12	120.507	7.843	176.423	52.813	43.864
21	E	13	121.993	8.418	177.467	54.44	31.554
22	D	14	125.546	9.716	176.936	53.163	40.575
23	K	14A	117.687	9.989	178.663	59.267	32.808
24	T	14B	104.753	8.156	175.906	60.798	70.09
25	E	14C	123.621	8.072	178.072	59.781	29.193
26	R	14D	122.634	8.911	177.527	58.672	27.936
27	E	14E	118.943	7.569	179.905	58.083	28.436
28	L	14F	117.942	7.077	178.544	57.456	38.531
29	L	14G	120.334	7.184	180.225	57.074	40.112
30	E	14H	116.97	8.367	177.816	57.739	28.621
31	S	14I	115.24	7.197	174.115	59.693	63.025
32	Y	14J	121.508	6.579	176.306	53.319	34.592
33	I	14K	119.25	7.148	176.221	60.969	38.255
34	D	14L	124.409	8.288	176.834	54.213	40.839
35	G	14M	110.739	8.348	173.745	45.401	33.101
36	R	15	126.129	7.767	181.285	57.122	
37	I	16					
38	V	17			175.489	60.103	33.101

39	E	18	120.805	9.261	175.009	56.998	
40	G	19	106.449	7.873	172.888	43.843	
41	S	20	112.845	8.407	172.985	56.143	65.836
42	D	21	120.794	8.368	176.957	54.449	39.632
43	A	22	125.975	8.914	177.617	51.207	18.527
44	E	23	121.473	7.308	176.792	55.773	31.913
45	I	24	124.905	8.578	177.542	62.421	36.059
46	G	25	113.305	9.012	175.441	45.181	
47	M	26	117.242	7.583	175.807	56.375	
48	S	27	114.417	8.954			
49	P	28			174.415	65.151	29.301
50	W	29	118.465	5.894	177.258	53.967	
51	Q	30					
52	V	31					
53	M	32					
54	L	33					
55	F	34					
56	R	35					
57	K	36			177.063	59.33	32.71
58	S	36A	111.392	7.872	172.836	54.588	
59	P	37					
60	Q	38					
61	E	39					
62	L	40					
63	L	41			177.666	54.963	43.47
64	C	42	117.032	7.723	174.03	55.955	
65	G	43	106.468	9.026	170.555	43.384	
66	A	44	115.964	7.875	175.432	50.891	
67	S	45	114.415	9.529	172.594	55.795	65.654
68	L	46	126.397	9.449	175.722	54.489	44.234
69	I	47	117.905	8.179	175.227	60.903	38.736
70	S	48	116.717	8.413	171.854	57.238	63.491
71	D	49	116.058	8.877	176.1	55.444	39.257
72	R	50	114.121	7.844	173.648	53.782	32.093
73	W	51	118.819	7.154	174.402	56.434	32.321
74	V	52	121.805	9.116	173.913	59.115	34.998
75	L	53	129.864	9.584	176.432	53.43	45.507
76	T	54	121.433	9.02	172.915	59.812	69.68
77	A	55	122.779	8.111	177.709	50.467	18.422
78	A	56	127.768	6.57	180.644	55.546	15.675
79	H	57	117.721	10.452	175.415	58.31	26.619
80	C	58	115.901	6.659	173.476	59.994	44.748

81	L	59	116.465	7.413	175.001	54.836	42.588
82	L	60	120.283	8.104	174.41	53.359	42.98
83	Y	60A	125.467	9.252	171.55	57.253	
84	P	60B					
85	P	60C			176.866	65.415	30.108
86	W	60D	113.92	7.066	174.997	52.999	29.588
87	D	60E	118.854	8.492	174.114	54.948	38.986
88	K	60F	116.714	6.737	175.487	54.47	
89	N	60G	128.464	8.156	173.499	52.313	37.22
90	F	60H	122.002	8.436	175.745	57.528	40.733
91	T	60I	111.535	9.186	175.773	59.511	71.773
92	E	61	118.849	10.045	177.093	60.166	27.048
93	N	62	115.369	8.027	175.777	54.158	38.552
94	D	63	116.375	7.825	176.164	55.382	43.833
95	L	64	118.597	6.727	174.172	53.686	46.887
96	L	65	117.272	8.888	175.23	54.353	44.944
97	V	66	112.484	8.492	174.761	58.082	
98	R	67					
99	I	68					
100	G	69					
101	K	70					
102	H	71					
103	S	72					
104	R	73					
105	T	74					
106	R	75					
107	Y	76					
108	E	77					
109	R	77A					
110	N	78					
111	I	79					
112	E	80					
113	K	81			175.06	53.35	34.984
114	I	82	123.873	8.467	176.174	59.75	38.98
115	S	83	120.813	8.813	172.899	57.678	65.068
116	M	84	123.206	8.722	176.512	53.493	32.836
117	L	85	122.017	9.438	175.173	53.518	42.024
118	E	86	122.311	9.216	175.09	56.477	31.372
119	K	87	114.329	7.049	173.833	54.443	36.145
120	I	88	121.787	8.07	173.896	59.89	40.251
121	Y	89	124.407	9.398	174.426	57.036	40.119
122	I	90	125.487	8.954	176.036	59.789	38.65

123	H	91	129.973	6.844	175.664	57.264	
124	P	92			178.478	64.055	31.676
125	R	93	121.057	11.134	176.116	53.755	29.058
126	Y	94	124.428	7.594	175.197	57.053	
127	N	95	131.423	8.845	173.722	50.288	38.624
128	W	96	119.577	6.381	176.303	56.043	
129	R	97	121.479	6.783	176.788	58.111	
130	E	97A					
131	N	98					
132	L	99			176.169		38.595
133	D	100	117.633	7.331	176.183	54.823	42.173
134	R	101	119.646	8.42	173.828	57.286	24.667
135	D	102	114.596	7.046	175.068	53.007	37.613
136	I	103	125.878	8.12	169.745	60.339	41.546
137	A	104	127.431	8.383	174.664	50.427	23.523
138	L	105	120.762	9.433	176.313	52.504	47.239
139	M	106	118.316	9.645	172.866	53.733	34.937
140	K	107	126.277	8.676	176.573	53.446	34.132
141	L	108	127.92	8.745	177.081	54.514	41.364
142	K	109	120.814	8.056	175.583	58.721	32.343
143	K	110	115.377	7.348	172.7	52.157	
144	P	111			176.661	61.804	31.274
145	V	112	122.013	8.773	173.774	59.256	32.947
146	A	113	127.4	7.933	177.412	50.592	18.354
147	F	114	122.326	7.901	176.143	54.188	37.106
148	S	115	118.768	9.806	175.006	56.843	66.557
149	D	116	119.118	8.65	174.07	56.255	39.061
150	Y	117	113.171	7.768	174.52	55.914	40.242
151	I	118	119.06	6.97	173.767	60.809	40.01
152	H	119	126.172	7.518	172.534	54.607	
153	P	120			176.109	62.367	31.657
154	V	121	122.875	8.048	173.152	59.722	32.382
155	C	122	119.903	6.761	174.888	53.497	41.191
156	L	123	121.882	9.223	175.466	51.306	
157	P	124			175.183	63.015	31.435
158	D	125	117.989	7.031	175.075	51.216	42.019
159	R	126	120.709	8.548	178.292	59.322	29.273
160	E	127	119.235	8.421	179.254	59.025	28.15
161	T	128	119.243	7.88	175.854	66.69	
162	A	129	121.649	7.774	178.64	54.894	16.843
163	A	129A	118.704	7.799	179.985	53.961	17.483
164	S	129B	110.916	7.587	175.522	60.811	63.853

165	L	129C	116.292	7.583	176.423	55.098	41.296
166	L	130	118.726	7.139	175.699	52.764	37.991
167	Q	131	122.967	7.271	174.029	52.741	30.185
168	A	132	123.236	8.215	177.95	52.994	17.363
169	G	133	111.357	9.003	174.182	44.798	
170	Y	134	124.118	8.091	176.274	55.877	36.043
171	K	135	119.792	8.956	175.41	55.625	
172	G	136	107.598	8.444	170.704	43.554	
173	R	137	120.073	8.617	174.657	54.084	
174	V	138	129.157	9.311	173.768		
175	T	139					
176	G	140					
177	W	141					
178	G	142					
179	N	143					
180	L	144					
181	K	145					
182	E	146					
183	T	147					
184	W	148					
185	T	149					
186	A	149A			177.629	52.738	
187	N	149B	116.03	8.259	175.104	53.759	38.078
188	V	149C	119.085	7.719	176.67	62.142	32.308
189	G	149D	112.163	8.341	173.757	44.771	
190	K	149E	119.865	7.963	176.186	55.375	
191	G	150	108.733	7.914	173.002		
192	Q	151					
193	P	152					
194	S	153					
195	V	154					
196	L	155					
197	Q	156			175.478	52.333	33.386
198	V	157	118.563	8.892		58.584	
199	V	158	126.147	8.791	170.624	60.624	
200	N	159	126.159	8.546	175.308	50.998	39.222
201	L	160	121.895	8.966	173.776	50.999	
202	P	161			175.829	60.819	30.887
203	I	162	122.525	8.645	176.765	62.238	36.62
204	V	163	133.978	8.015	174.247	61.647	32.97
205	E	164	126.394	9.531	177.43	56.967	29.177
206	R	165	124.592	8.876	175.071	61.675	

207	P	166			178.79	65.733	30.622
208	V	167	117.471	6.692	179.424	64.902	30.281
209	C	168	118.694	7.676	178.779	56.8	37.983
210	K	169	121.239	8.624	177.849	59.311	31.827
211	D	170	116.312	7.747	177.314	55.425	40.282
212	S	171	114.519	7.783	172.415	59.959	64.561
213	T	172	117.815	7.506	171.554	59.508	69.167
214	R	173	124.065	8.691	177.93	55.338	29.545
215	I	174	126.28	8.245	174.894	60.358	35.98
216	R	175	127.027	8.163	174.888	55.892	29.161
217	I	176	123.999	7.875	176.499	59.368	
218	T	177			175.472	58.916	
219	D	178	117.929	8.808	176.145	55.376	39.68
220	N	179	112.553	8.018	174.612	53.545	37.599
221	M	180	123.513	8.139	174.403	54.249	36.162
222	F	181	118.058	8.993	171.685	56.326	40.582
223	C	182	120.665	8.801	172.135	54.826	41.972
224	A	183	121.783	8.636	175.821	51.016	23.01
225	G	184	107.842	8.604	171.301	43.962	
226	Y	184A	118.441	9.274	177.387	59.314	
227	K	185	124.253	9.998	177.982	54.717	
228	P	186			176.648	65.138	31.286
229	D	186A	113.748	8.378	177.141	53.913	39.603
230	E	186B	118.157	7.665	177.845	57.431	30.846
231	G	186C	107.679	8.119	173.352	46.031	
232	K	186D	119.508	6.676	173.075	54.438	33.464
233	R	187	118.607	7.955	174.17	54.379	
234	G	188					
235	D	189					
236	A	190					
237	C	191					
238	E	192					
239	G	193					
240	D	194					
241	S	195					
242	G	196			174.086	45.74	
243	G	197	103.011	7.722	170.486	43.508	
244	P	198			174.333	63.407	31.093
245	F	199	122.496	7.792	177.275	55.811	40.309
246	V	200	121.661	9.337	174.039	58.134	34.756
247	M	201	118.164	9.186	173.093	55.001	38.871
248	K	202	122.998	6.673	174.95	53.826	30.941

249	S	203	120.752	8.854	175.012	54.306	
250	P	204			176.204	62.437	30.822
251	F	204A	118.774	7.619	176.885	57.968	38.223
252	N	204B	113.393	7.161	175.106	51.49	38.655
253	N	205	114.281	7.745	171.765	54.596	38.317
254	R	206	112.739	7.341	176.486	54.593	32.105
255	W	207	120.724	8.652	174.782	56.684	30.539
256	Y	208	118.29	8.981	175.836	55.925	43.215
257	Q	209	121.874	8.949	172.755	55.378	25.817
258	M	210	122.261	8.131	178.674	53.215	31.365
259	G	211	104.504	8.925	173.098	44.692	
260	I	212	123.18	7.712	175.704	59.926	41.259
261	V	213	127.343	9.125	175.613	64.309	
262	S	214	126.891	7.905		59.524	
263	W	215					
264	G	216	105.666	8.08	172.939	44.39	
265	E	217	124.291	9.268	175.49	55.171	29.319
266	G	219	111.757	8.059	173.227	43.434	
267	C	220					
268	D	221A					
269	R	221					
270	D	222	125.024	8.955	177.699	55.218	39.534
271	G	223	112.276	9.112	171.761	44.966	
272	K	224	117.243	6.906	174.611	51.309	
273	Y	225					
274	G	226			173.196	45.479	
275	F	227	119.372	8.638	176.092	57.076	39.65
276	Y	228	120.746	8.696	175.147	55.467	40.731
277	T	229	120.162	9.16	173.947	63.297	68.66
278	H	230	129.803	8.431	173.661	55.351	31.126
279	V	231	126.47	7.698	177.224	67.72	31.23
280	F	232	118.773	10.03	177.933	62.251	
281	R	233	118.416	8.016	177.63	58.249	
282	L	234	119.779	7.337	176.973	53.222	40.97
283	K	235	122.867	7.551	177.184	59.163	32.378
284	K	236	118.087	8.383	179.326	59.166	30.318
285	W	237	120.949	7.504	175.724	61.676	27
286	I	238	117.523	7.301	176.86	64.074	37.801
287	Q	239	114.248	8.127	176.909	57.896	
288	K	240	119.459	7.338	178.923	58.5	31.19
289	V	241	120.272	7.275	177.606	65.58	30.139
290	I	242	118.056	7.146	178.651	64.467	36.473

291	D	243	119.409	8.132	178.052	56.016	40.118
292	Q	244	118.039	7.6	177.197	57.223	28.228
293	F	245	117.866	8.094	176.554	57.652	38.667
294	G	246	110.493	8.022	173.233	45.698	
295	E	247	125.878	7.931	181.322	57.852	