# BARRIERS TO TRANSLOCATION:
# STABILITY, KINETICS, AND ACID DENATURATION OF AVRPTO

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jennifer Erin Dawson

August 2008

BARRIERS TO TRANSLOCATION:

STABILITY, KINETICS, AND ACID DENATURATION OF AVRPTO

Jennifer Erin Dawson, Ph.D.

Cornell University 2008

In order to infect their hosts, many gram-negative bacteria translocate agents of infection, called effector proteins, through the Type III Secretion System (TTSS) into the host cytoplasm. This process is thought to require at least partial unfolding of these agents, raising the question of how an effector protein might unfold to enable its translocation, then refold once it reaches the host cytoplasm. AvrPto is a well-studied effector protein of *P. syringae* pv tomato. The presence of a readily-observed denatured population and the observation of acid denaturation for AvrPto make it ideal for studying the kinetic and thermodynamic characteristics that facilitate translocation. Application of NMR Nzz exchange spectroscopy revealed a global, two-state folding equilibrium with 16% unfolded population, a slow folding rate of 1.8 $s^{-1}$, and an unfolding rate of 0.33 $s^{-1}$ at pH 6.1. A non-native conformation with molten-globule character was revealed using a combination of CD, Trp fluorescence, and NMR HSQC spectroscopy. These data reflect an equilibrium in which the stability changes during two separate titration events. The low pH titration is between the non-native form and the unfolded ensemble, while the higher pH titration is between the native and non-native conformations. Remarkably, the high pH titration can be simulated using the His sidechain $pK_a$ values extracted from pH titration NMR HMQC spectra. Given the acidic apoplast in which the pathogen resides and the more basic host cytoplasm, these results offer an intriguing mechanism by which the pH-dependence of stability and slow folding kinetics of AvrPto would allow efficient translocation of the unfolded form through the TTSS and refolding into its functional folded form once

inside the host. Mutants were rationally designed with the purpose of improving the overall stability or changing the pH-dependence of stability and, hence, alter the secretion efficiency. Using the CyaA translocation assay, it was determined that unstable mutant, H54P AvrPto, which had been previously hypothesized to be translocation-incompetent, could indeed be translocated into the host cell.

## BIOGRAPHICAL SKETCH

Jennifer E. Dawson was raised in Kent, OH, a liberal college town. She attended Kent State University, the same institution that her parents attended and where her grandfather worked as a mechanic. In 2002, she graduated *Summa cum laude* with a Bachelors of Science in Physics. Kent State University did not have a Biophysics concentration, so she supplemented her undergraduate education with a Chemistry minor and additional Biology coursework. From 2002 to 2008, Jennifer pursued her Ph. D. in Biophysics from Cornell University, which is located in another liberal college town: Ithaca, NY.

ACKNOWLEDGEMENTS

I would like to thank all of the people who donated generously of their time, experience, and resources during the course of my research: Dr. Pete Pascuzzi, Hahn Nguyen, Professor Gregory Martin, Kyriacos Leptos, Kevin Tan, Dr. Cynthia Kinsland, Brian Zoltowski, Dr. Lea Vacca Michel, Professor Brian Crane, Brian Kvitko, Joanne Morello, Professor Alan Collmer, Kristina Govorovska, Jolita Seckute, Aaron Oswald and Sam Schueler.

I would especially like to thank all those who supported me during this time:
My mentor and principal investigator, Professor Linda K. Nicholson
My special committee, Professors Robert E. Oswald and Lois Pollack
My labmates for fascinating conversions and their patience
Dr. Bhargavi Jayaraman and Soumya De for listening
Dr. Beth and Ronald Dawson and Brian Prochaska for their unending support

This work was supported by NSF grant MCB-0641582 and NIH Molecular Biophysics Training Grant T32 GM08267.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

# LIST OF ABBREVIATIONS

3CPro: 3c protease

AvrPto: Avirulence protein that interacts with Pto

CD: Circular Dichroism

HR: Hyper-sensitive Response

Hz: Hertz, a frequency unit

NMR: Nuclear Magnetic Resonance

ppm: parts per million, field strength independent frequency unit

pv: pathovar

TrAvrPto: Truncated Avirulence protein that interacts with Pto

TTSS: Type Three (III) Secretion System

wt: wild-type

# LIST OF SYMBOLS

I(t): Intensities of peaks from Nzz NMR spectra.

$k_{UF}$: Folding rate constant, measured in $s^{-1}$

$k_{FU}$: Unfolding rate constant, measured in $s^{-1}$

$k_{ex}$: Exchange rate constant, measured in $s^{-1}$

$R_1$: Rate constant for the return of longitudinal magnetization to equilibrium, measured in $s^{-1}$

$R_2$: Rate constant for the return of transverse magnetization to equilibrium, measured in $s^{-1}$

$\Delta\omega$: Difference in chemical shift between the conformations, defines the chemical shift timescale of an NMR experiment, measured in $s^{-1}$

$K_{eq}$: Equilibrium constant

$pK_{stab}$: The midpoint pH of a change in pH-dependent stability

$pK_a$: The pH at which half of the titratable group's population is protonated

R: $1.987*10^{-3}$ kcal $mol^{-1}$ $K^{-1}$, Gas constant

T: Temperature in Kelvin

Chapter One

Introduction

***Pathogen infection and the Type III secretion system***

The evolution of infection processes is shaped by a complex series of moves and counter-moves. Pathogens infect their hosts. Some hosts have methods of pathogen recognition and self-defense and, so, survive infection. In response, some pathogens develop methods of counter-acting these defenses. A few host strains find methods of neutralizing the new pathogen counter-agents and, so, the cycle continues.

An excellent example of this cycle is the interaction of *Pseudomonas syringae* pv tomato with its hosts. *P. syringae* pv tomato is a plant pathogen that causes disease in tomato and Arabidopsis plants. During infection, the pathogen lives in the extracellular space (apoplast) of plant leaf tissue (Jin et al. 2003). The plant cells can recognize the pathogen's presence and activate their basal defenses, including thickening the cell walls (Alfano et al. 2004). In order to bypass these defenses, *P. syringae* uses an apparatus called the Type III Secretion System (TTSS) to translocate a set of infection agents, called effector proteins, into the host cytoplasm (Galan et al. 1999) (Figure 1.1A). Many of these effector proteins are involved in counter-acting the host cell defenses. The infection of susceptible cells eventually leads to bacterial speck disease, small necrotic lesions on the leaves (Chang et al. 2000, Hauck et al. 2003, Jin et al. 2003, Shan et al. 2000, Whalen et al. 1991). Some resistant strains of tomato have a kinase Pto that can detect one of the effectors, AvrPto, within the host cytoplasm and trigger cellular defences, including the hyper-sensitive response (HR), a localized cell death to prevent bacterial proliferation within the apoplast (Scofield et al. 1996, Tang et al. 1999) (Figure 1.1B). HR is a common resistance mechanism for plants, such as *Nicotiana benthamiana* during invasion by *P. syringae* pv tomato.

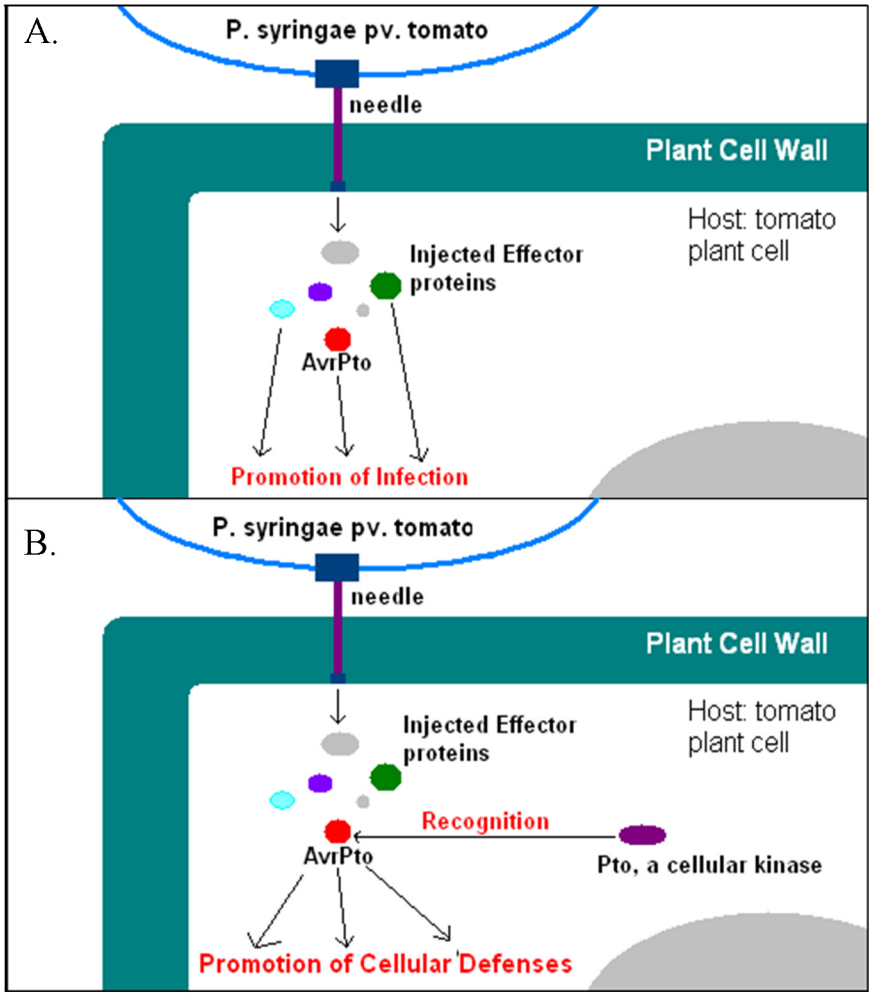



Figure 1.1: **Virulence and avirulence roles of AvrPto.** A. AvrPto, along with the other effector proteins injected into the host cytoplasm by the TTSS, promote disease (virulence). B. The avirulence role of AvrPto is involved in host cell defense. Resistant tomato strains have Pto, which interacts with AvrPto and triggers the cell defenses.

The TTSS is used by many gram-negative bacteria to infect hosts and is involved in the propagation of some of the worst diseases: *Pseudomonas* causes various types of leaf blight, *Salmonella typhurium* infection leads to typhoid fever, and *Yersinia pestis* causes the bubonic (black) plague. The TTSS apparatus is composed of two rings, one in each bacterial membrane (Ghosh 2004). A hollow needle-like object bridges the space between bacterium and host cell through which effector proteins can enter. The TTSS is morphologically conserved across all TTSS-bearing bacteria and TTSS proteins in the inner membrane have conserved sequences (Francis et al. 1994, Ghosh 2004, Marlovits et al. 2004).

Each type of pathogen targets different hosts. Each pathovar or variant of *P. syringae* infects only one or two host species and some only a few strains of a species. For example, *P. syringae* pathovar (pv) tomato infects susceptible tomato and Arabidopsis strains. Therefore, the sets of effector proteins required for infection differ from pathogen to pathogen.

There are several requirements for secretion *via* the TTSS. The TTSS requires the presence of a conserved putative ATPase in the inner membrane ring (Ghosh 2004) and a pH gradient across the inner membrane (Wilharm et al. 2004) in order to function. The inner diameter of the needle is ~2-3 nm (Blocker et al. 2001, Cordes et al. 2003), which necessitates at least partial unfolding of effector proteins before they may pass through it. These requirements raise questions about how effector proteins are effectively secreted. Can an effector protein be too stable to be secreted by the TTSS? Can a protein fold too quickly to be secreted? Would a pH-dependence in protein stability be important to secretion efficiency?

***Probes of secretion-efficiency: AvrPto and its rationally-designed mutants***

Effector protein AvrPto is a model probe of translocation. It has been studied in both its disease-causing (virulence) and disease-preventing (avirulence) roles (Pedley et al. 2003). Some effector proteins require secretion chaperones to help them compete for access to the TTSS (Ghosh 2004, Luo et al. 2001, Wulff-Strobel et al. 2002), to unfold them (Stebbins et al. 2001), or to prevent their degradation (Losada et al. 2005). There is no known secretion chaperone for AvrPto, which simplifies the experimental system. The structure of truncated AvrPto (TrAvrPto) has already been solved (Wulf et al. 2002, Wulf et al. 2004). Both AvrPto and TrAvrPto have separate sets of peaks for the native and non-native conformations and both display acid denaturation.

The question of thermodynamic and kinetic thresholds to translocation was approached primarily from a structural biology perspective. The physical characteristics of TrAvrPto were studied to gain insight into the properties of effector proteins that may allow them to be efficiently secreted. Three techniques were used to characterize the pH-dependence of stability for wild-type (wt) TrAvrPto. Circular dichroism (CD) and Trp fluorescence spectroscopy are sensitive to the protein as a whole. CD spectroscopy is an excellent way of studying the secondary structure content, especially α-helices (Martin et al. 2008). The fluorescence emission of a Trp sidechain depends on whether it is in a non-polar environment like the protein core or is exposed to solvent (Pace 1986). Nuclear magnetic resonance (NMR) spectroscopy can resolve peaks for individual spins within the protein. For example, HSQC NMR experiments (Mulder et al. 1996) can be used to study the changes to the backbone amide groups of each amino acid residue during pH titrations. Additionally, the peaks for the native and non-native conformations are resolved from one another for

TrAvrPto. NMR Nzz exchange spectroscopy takes advantage of these discrete sets of peaks to quantify the slow folding kinetics of TrAvrPto (Farrow et al. 1994, Montelione et al. 1989). The information gained from the characterization of wt TrAvrPto was used as a guide for rational mutagenesis of proteins designed to be more stable or display different pH dependence of stability. A translocation assay (Schechter et al. 2004), developed by the Collmer lab (Cornell University, Department. of Plant Pathology and Plant-Microbe Biology), was used to determine if one of these mutants, H54P TrAvrPto, could be translocated into the host cell cytoplasm.

***Introduction to experimental methods***

Chiral centers absorb right circularly-polarized light differently than left circularly-polarized light (Martin et al. 2008). CD spectroscopy measures this difference in absorption. The peptide backbone of amino acid residues absorbs in the far-UV range, near 222 nm. The $C_\alpha$ atom of an amino acid has four substituents, each of which is a different chemical group. Therefore, the $C_\alpha$ lacks mirror symmetry and is considered a chiral center. The one exception is the amino acid Gly, which, since it has two protons on $C_\alpha$, is achiral. Additionally, all $\alpha$-helices have a right-handed twist, resulting in a strong CD signal for this secondary structural element in the far-UV range.

During fluorescence, a fluorophore absorbs UV light, exciting its electrons to the first excited electronic state (Shanker et al. 2008). While the electron is excited, it loses part of its energy as heat. When the electron subsequently relaxes back to the ground state, it emits a light photon with less energy and a longer wavelength than the photon that the fluorophore had initially absorbed. The ground state of a fluorophore is stabilized by an apolar environment, such as the hydrophobic protein core, but the

excited state is destabilized. Both absorption and emission energies are thereby increased within the protein core and, hence, their wavelengths are less than those for fluorophores exposed to polar solvent. The stability of proteins can be quantified by exploiting this difference in emission and the presence of aromatic residues, especially Trp, to act as fluorophores.

NMR spectroscopy is distinguished by its ability to resolve individual nuclear spins and its exquisite sensitivity to the spins' local environments (Cavanagh et al. 1996). Only nuclei with non-zero spins are detected by NMR spectroscopy. In a solution of identical molecules, there is an ensemble corresponding to each observable nucleus: for example, the backbone amide proton of TrAvrPto's G95 residue. At equilibrium, a nuclear spin will experience both the external magnetic field of the spectrometer and the much weaker local fields created by the spin's environment. For nuclei with spins of one-half ($^{1}H$, $^{13}C$, $^{15}N$, $^{31}P$), these local fields are dependent on the behavior of the electrons associated with the spin and its neighbors (chemical shift anisotropy), as well as dipole-dipole interactions between nuclear spins. The net magnetization of each ensemble is aligned along the external field and the spin ensembles precess at an angular frequency, $\omega$, which is proportional to $B_{eff}$, the sum of the external and local magnetic fields. A peak corresponding to this spin's ensemble is located at the chemical shift $\omega$ in NMR spectra. Since $B_{eff}$ varies from spin to spin, NMR spectra contain information on the local environments in a protein molecule.

During an NMR experiment, a series of radio-frequency pulses redistribute the spin populations, perturbing the magnetization. After the pulses, the magnetization relaxes back to equilibrium. The rate constant $R_1$ describes the return to equilibrium of the longitudinal component of the magnetization, *i.e.* the component aligned parallel with the external field. $R_2$ characterizes the magnetization component that is transverse (perpendicular) to the external field. NMR spectroscopy is a versatile

technique that has been applied using numerous pulse sequence experiments, each designed to probe different aspect of a molecule. What was most useful for the work presented here was NMR spectroscopy's sensitivity to chemical events that occur over many orders of magnitude of timescale: ns to ps (local dynamic events), μs to ms (folding events, complex formation), and ms to days (folding events, degradation).

Many of the chemical events experienced by TrAvrPto—folding, sidechain tautomerization, and pH titration—occurred in the μs to ms range. Within this range, there are three timescale regimes, each with characteristic behaviors. Consider a generic event in which a nucleus exchanges between sites A and B. The difference in the nucleus' chemical shift at A and B ($\Delta\omega = \omega_A - \omega_B$) defines the chemical exchange timescale of the event. When the exchange rate, $k_{ex}$, is slow ($\Delta\omega >> k_{ex}$), two sets of peaks will be seen, one set for the nucleus at site A and one set for when it is in site B. These two peaks will begin to broaden and move toward each other as the rate of exchange increases until, when $\Delta\omega \sim k_{ex}$, the peaks merge. This population-weighted average peak sharpens as the rate further increases into the fast exchange regime ($\Delta\omega << k_{ex}$).

During the HSQC and HMQC experiments used in this work, magnetization was selectively transferred between specific chemical groups. For a $^{15}N$-$^{1}H$ HSQC, the transfer is between covalently-bound $^{1}H$ and $^{15}N$ nuclei (Mulder et al. 1996). In a protein, these pairs are from the amide N-H bonds in the backbone and from the sidechains of Asn, Gln, Trp, Arg, and Lys. Spins that experience similar chemical environments will correspond to peaks in similar locations in the NMR spectrum. For example, peaks for an unfolded protein will tend to cluster together near the center of the HSQC. Amino acid residues within an α-helix are constrained to similar geometries and, as a result, their peaks also tend to cluster toward the center of HSQC spectra. The His sidechain amide groups generally exchange too quickly to be

observed *via* NMR. Therefore, the protonation state of the two $^{15}N$ of sidechain His were observed using a modified HMQC experiment, which was optimized to transfer magnetization between a carbon-bound proton and an $^{15}N$ nucleus two or three bonds away (Bax et al. 1990, Pelton et al. 1993).

During a Nzz exchange experiment (Farrow et al. 1994, Montelione et al. 1989), longitudinal $^{15}N$ magnetization is exchanged between slowly-exchanging species during a mixing period of varying length, T. The change in magnetization with T depends on both the forward and backward kinetic rates and on the $R_1$ relaxation rates of both species. The Nzz spectra are similar to HSQC spectra with additional cross-peaks indicating slow exchange. For a two-state system, there will be four peaks in the spectrum: two autocorrelation peaks corresponding to the slowly exchanging populations and two cross-correlation peaks. The variation of the auto- and cross-peak intensities can be modeled using a set Bloch equations modified to include the effects of chemical exchange (McConnell 1958).

REFERENCES


Alfano, J.R. and Collmer, A. 2004. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu.Rev.Phytopathol.* **42:** 385-414.

Bax, A., Ikura, M., Kay, L.E., Torchia, D.A. and Tschudin, R. 1990. Comparison of different modes of two-dimensional reverse-correlation NMR for the study of proteins. *J Magn Reson* **86:** 304-318.

Blocker, A., Jouihri, N., Larquet, E., Gounon, P., Ebel, F., Parsot, C., Sansonetti, P. and Allaoui, A. 2001. Structure and composition of the Shigella flexneri "needle complex", a part of its type III secreton. *Mol. Microbiol.* **39:** 652-663.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G.I. and Skelton, N.J. 1996. *Protein NMR spectroscopy: Principles and practice,* 243-299. Academic Press, San Diego, CA.

Chang, J.H., Rathjen, J.P., Bernal, A.J., Staskawicz, B.J. and Michelmore, R.W. 2000. AvrPto enhances growth and necrosis caused by Pseudomonas syringae pv. tomato in tomato lines lacking either Pto or Prf. *Mol. Plant-Microbe Interact.* **13:** 568-571.

Cordes, F.S., Komoriya, K., Larquet, E., Yang, S., Egelman, E.H., Blocker, A. and Lea, S.M. 2003. Helical structure of the needle of the type III secretion system of Shigella flexneri. *J. Biol. Chem.* **278:** 17103-17107.

Farrow, N.A., Zhang, O., Forman-Kay, J.D. and Kay, L.E. 1994. A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR* **4:** 727-734.

Francis, N.R., Sosinsky, G.E., Thomas, D. and DeRosier, D.J. 1994. Isolation, characterization and structure of bacterial flagellar motors containing the switch complex. *J. Mol. Biol.* **235:** 1261-1270.

Galan, J.E. and Collmer, A. 1999. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284:** 1322-1328.

Ghosh, P. 2004. Process of protein transport by the type III secretion system. *Microbiol. Mol. Biol. Rev.* **68:** 771-795.

Hauck, P., Thilmony, R. and He, S.Y. 2003. A Pseudomonas syringae type III effector suppresses cell wall-based extracellular defense in susceptible Arabidopsis plants. *Proc. Natl. Acad. Sci. U.S.A.* **100:** 8577-8582.

Jin, Q., Thilmony, R., Zwiesler-Vollick, J. and He, S.Y. 2003. Type III protein secretion in Pseudomonas syringae. *Microb. Infect.* **5:** 301-310.

Losada, L.C. and Hutcheson, S.W. 2005. Type III secretion chaperones of Pseudomonas syringae protect effectors from Lon-associated degradation. *Mol.Microbiol.* **55:** 941-953.

Luo, Y., Bertero, M.G., Frey, E.A., Pfuetzner, R.A., Wenk, M.R., Creagh, L., Marcus, S.L., Lim, D., Sicheri, F., Kay, C., Haynes, C., Finlay, B.B. and Strynadka, N.C. 2001. Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. *Nat. Struct. Biol.* **8:** 1031-1036.

Marlovits, T.C., Kubori, T., Sukhan, A., Thomas, D.R., Galan, J.E. and Unger, V.M. 2004. Structural insights into the assembly of the type III secretion needle complex. *Science* **306:** 1040-1042.

Martin, S.R. and Schilstra, M.J. 2008. Circular dichroism and its application to the study of biomolecules. *Methods Cell Biol.* **84:** 263-293.

McConnell, H.M. 1958. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* **28:** 430-431.

Montelione, G.T. and Wagner, G. 1989. 2D chemical-exchange NMR-spectroscopy by proton-detected heteronuclear correlation. *J. Am. Chem. Soc.* **111:** 3096-3098.

Mulder, F.A.A., Spronk, C.A.E.A., Slijper, M., Kaptein, R. and Boelens, R. 1996. Improved HSQC experiments for the observation of exchange broadened signals. *J. Biomol. NMR* **8:** 223-228.

Pace, C.N. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131:** 266-280.

Pedley, K.F. and Martin, G.B. 2003. Molecular basis of Pto-mediated resistance to bacterial speck disease in tomato. *Annu. Rev. Phytopathol.* **41:** 215-243.

Pelton, J.G., Torchia, D.A., Meadow, N.D. and Roseman, S. 1993. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Sci.* **2:** 543-558.

Schechter, L.M., Roberts, K.A., Jamir, Y., Alfano, J.R. and Collmer, A. 2004. Pseudomonas syringae type III secretion system targeting signals and novel effectors studied with a Cya translocation reporter. *J.Bacteriol.* **186:** 543-555.

Scofield, S.R., Tobias, C.M., Rathjen, J.P., Chang, J.H., Lavelle, D.T., Michelmore, R.W. and Staskawicz, B.J. 1996. Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. *Science* **274:** 2063-2065.

Shan, L., Thara, V.K., Martin, G.B., Zhou, J.M. and Tang, X. 2000. The Pseudomonas AvrPto protein is differentially recognized by tomato and tobacco and is localized to the plant plasma membrane. *Plant Cell* **12:** 2323-2338.

Shanker, N. and Bane, S.L. 2008. Basic aspects of absorption and fluorescence spectroscopy and resonance energy transfer methods. *Methods Cell Biol.* **84:** 213-242.

Stebbins, C.E. and Galan, J.E. 2001. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* **414:** 77-81.

Tang, X., Xie, M., Kim, Y.J., Zhou, J., Klessig, D.F. and Martin, G.B. 1999. Overexpression of Pto activates defense responses and confers broad resistance. *Plant Cell* **11:** 15-29.

Whalen, M.C., Innes, R.W., Bent, A.F. and Staskawicz, B.J. 1991. Identification of Pseudomonas syringae pathogens of Arabidopsis and a bacterial locus determining avirulence on both Arabidopsis and soybean. *Plant Cell* **3:** 49-59.

Wilharm, G., Lehmann, V., Krauss, K., Lehnert, B., Richter, S., Ruckdeschel, K., Heesemann, J. and Trulzsch, K. 2004. Yersinia enterocolitica type III secretion depends on the proton motive force but not on the flagellar motor components MotA and MotB. *Infect. Immun.* **72:** 4004-4009.

Wulf, J., Pascuzzi, P.E., Martin, G.B. and Nicholson, L.K. 2002. 1H, 15N and 13C chemical shift assignments of the structured core of the pseudomonas effector protein AvrPto. *J. Biomol. NMR* **23:** 247-248.

Wulf, J., Pascuzzi, P.E., Fahmy, A., Martin, G.B. and Nicholson, L.K. 2004. The solution structure of type III effector protein AvrPto reveals conformational and dynamic features important for plant pathogenesis. *Structure* **12:** 1257-1268.

Wulff-Strobel, C.R., Williams, A.W. and Straley, S.C. 2002. LcrQ and SycH function together at the Ysc type III secretion system in Yersinia pestis to impose a hierarchy of secretion. *Mol.Microbiol.* **43:** 411-423.

Chapter Two

# Folding kinetics and thermodynamics of *Pseudomonas syringae* effector protein AvrPto provide insight into translocation *via* the Type III secretion system[1]

## *Introduction*

The Type III Secretion System (TTSS) is observed in a broad range of gram-negative bacterial genera, including *Salmonella*, *Shigella*, *Yersinia*, and *Pseudomonas* (Preston et al. 1995). This syringe-like supramolecular assembly is composed of multiple copies of over 20 different proteins (Fivaz et al. 1999, Galan et al. 1999) and can transport effector proteins from the bacterium in a process dependent upon both a putative ATPase in the inner membrane ring (Eichelberg et al. 1994, Tamano et al. 2000, Woestyn et al. 1994) and a proton gradient across the inner bacterial membrane (Wilharm et al. 2004) (Figure 2.1). Although effector proteins themselves vary widely among pathogens, the translocation machinery of the TTSS is a common feature found among many important pathogens (Jin et al. 2003).

Partial or complete unfolding of effector proteins is thought to be necessary for translocation through the TTSS into the host (Johnson et al. 2005). A hollow, needle-like structure called the Hrp pilus extends from the basal body of the TTSS and has an inner diameter of 2-3 nm (Blocker et al. 2001, Cordes et al. 2003). It is known from immunogold-labeling experiments that the extracellular components of the TTSS and the effector proteins are conducted along the needle (Brown et al. 2001, Jin et al. 2001) and exit through the needle's tip (Li et al. 2002). In the current models for both TTSS needle elongation and effector translocation, these proteins pass through the inside of the needle in at least a partially unfolded form. This mechanism is analogous

[1] Chapter Two is based on a manuscript published Courtesy of *Protein Science* (Dawson et al. 2008), which was accepted before the existence of the intermediate conformation was discovered (Chapter Three). The "unfolded" NMR peaks at pH≥6 are proportional to this intermediate population.

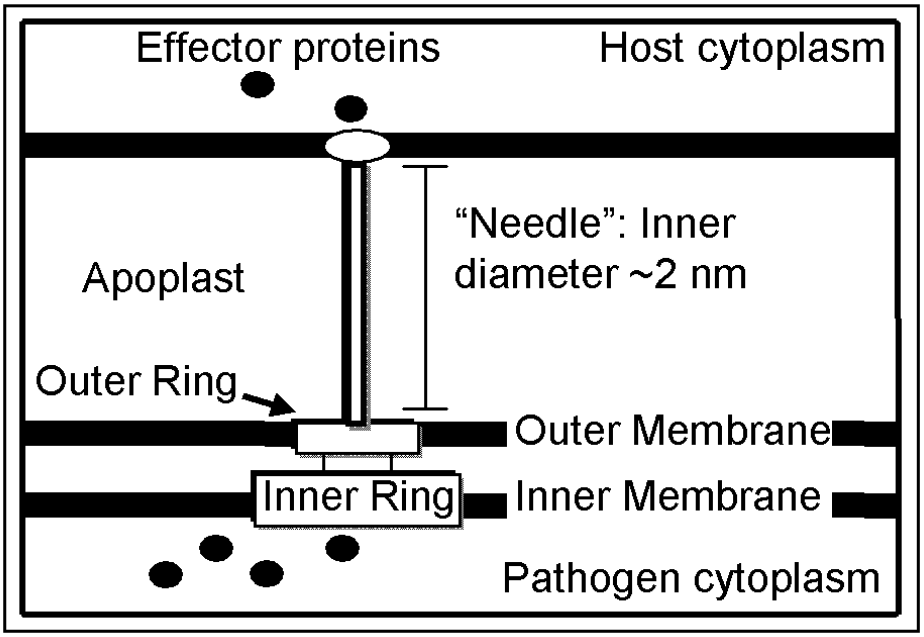



**Figure 2.1: The Type III Secretion System provides a pathogen-host conduit.** Schematic diagram of the TTSS used by pathogens to transport effector proteins from the bacterial cytoplasm into the host cytoplasm. These proteins are thought to pass through the hollow, needle-like structure (pilus) of the TTSS. Since the inner diameter of the pilus is ~2-3 nm, most proteins must at least partially unfold before secretion.

to the elongation model for the evolutionarily-related bacterial flagellar assembly (Emerson et al. 1970, Yonekura et al. 2003).

The stability and folding kinetics of an effector protein have important effects on its ability to be secreted. These dependences are not simple, however. Fusion of ubiquitin to the C-terminus of YopE, a *Yersinia* effector protein, prevented secretion of the fusion protein, whereas mutational destabilization of the ubiquitin fusion partner resulted in secretion (Lee et al. 2002). Conversely, the RhoGAP domain of YopE and YopH effector proteins have reasonable stabilities of 5 to 6 kcal/mol (Ghosh 2004, Zhang et al. 1992), yet they are secreted. The importance of folding kinetics is also implicated by the finding that YopE secretion was prevented by fusion to rapidly-folding domains such as dihydrofolate reductase (Sorg et al. 2005). Some, but not all, effector proteins associate with secretion chaperones that remain in the bacterium after effector secretion (Cheng et al. 1999, Frithz-Lindsten et al. 1995, Rosqvist et al. 1994), which adds another dimension to the complexity of the secretion process. Proposed roles of secretion chaperones include assisting conjugate effector proteins to unfold (Stebbins et al. 2001) or to compete for access to the TTSS (Ghosh 2004, Luo et al. 2001, Wulff-Strobel et al. 2002). The emerging picture is that the secretion efficiency of an effector protein appears to depend on some combination of its folding thermodynamics and kinetics.

The *Pseudomonas syringae* pv tomato effector protein AvrPto, which has no known chaperone and has been shown to have marginal stability (Wulf et al. 2004), offers a simple model system to probe the dependence of secretion and translocation on effector protein stability and kinetics. Upon delivery into the host cytoplasm, AvrPto has roles in promoting the infection of tomato leaf cells (Bogdanove et al. 2000, Chang et al. 2000, Hauck et al. 2003, Shan et al. 2000) and, ironically, in the plant's defense against the pathogen (Pedley et al. 2003, Scofield et al. 1996, Tang et

al. 1999). The NMR structure of a truncated 105-residue version of AvrPto (TrAvrPto), in which the unstructured N- and C-termini were removed, reveals a ~5 nm long and ~2.5 nm wide three-helix bundle with a short orthogonal helix, and a large Ω loop (Wulf et al. 2004). These dimensions would necessitate at least partial unfolding for passage through the Hrp pilus.

Interestingly, both AvrPto and TrAvrPto have a significant unfolded population in slow exchange with the folded form in aqueous solution at pH 6 and 25°C (Wulf et al. 2004). The ability of AvrPto to unfold and refold *in vitro* without chaperone assistance, and the presence of an NMR-observable unfolded population, enables us to study relationships between the kinetics and thermodynamics of effector protein folding and chaperone-independent secretion. Of particular interest are the residue-specific folding and unfolding rates and their potential influence on the viability of a protein for TTSS secretion.

In this work, the equilibrium kinetics and thermodynamics of TrAvrPto folding were characterized using two-dimensional NMR techniques (Farrow et al. 1994a, Farrow et al. 1994b, Montelione et al. 1989, Mulder et al. 1996). Nzz exchange spectroscopy is sensitive to chemical exchange processes that occur at rates on the order of 0.1 to 10 $s^{-1}$ (Palmer et al. 2001). This NMR technique has the advantage of providing a multiple-site perspective that enables the elucidation of independent folding regions, if present, or the accurate extraction of kinetic and thermodynamic parameters for a global two-state folding process. Here, Nzz exchange spectroscopy was applied to determine the folding and unfolding rates and stability of TrAvrPto at pH 6.1. Additional application of HSQC and longitudinal relaxation ($T_1$) experiments showed that TrAvrPto stability increases with increasing pH and provided confirmation of the folding equilibrium parameters obtained from Nzz exchange spectroscopy. Given the acidic apoplast environment of the pathogen and the more

basic host cytoplasm, the results presented here offer an intriguing mechanism by which the pH-dependence of stability and slow folding kinetics of AvrPto would allow efficient translocation through the TTSS and refolding into its functional folded form once inside the host.

***Results***

*The unfolding of TrAvrPto proceeds on the slow chemical exchange timescale across the entire protein.*

Nzz exchange spectroscopy (Farrow et al. 1994b, Montelione et al. 1989) was used to identify individual residues in slow chemical exchange and to quantify their folding and unfolding rate constants. For each spin in a slow two-state folding reaction, four peaks are expected: two auto-correlation peaks for the folded, FF, and unfolded, UU, resonances and two exchange cross-correlation peaks, FU and UF (Figure 2.2, inset). The time-dependent changes in intensity of these four peaks — $I_{FF}(t)$, $I_{FU}(t)$, $I_{UF}(t)$, and $I_{UU}(t)$ (Figure 2.2)- are modeled by the two-state solution of the Bloch-McConnell exchange equations (Cavanagh et al. 1996, McConnell 1958), which depends on the kinetic rate constants, $k_{FU}$ and $k_{UF}$, and the rate of longitudinal relaxation in the absence of exchange for each conformation, $R_{1F}$ and $R_{1U}$.

The presence of crosspeaks in the Nzz exchange spectroscopy spectra of TrAvrPto immediately demonstrates that folding occurs on the slow chemical exchange timescale (*circa* $(k_{FU}+k_{UF}) \leq 100\ s^{-1}$). Ninety-five potential unfolded TrAvrPto resonances were observed, suggesting that the entire 105-residue protein exchanges between its folded and unfolded conformations on this slow exchange timescale. The folded autopeaks of 41 N-H groups were linked to at least one of their corresponding crosspeaks and, thereby, were identified as being in slow exchange (Figure 2.3, Appendix Table 2.1). Some of these residues (e.g. L37) displayed their

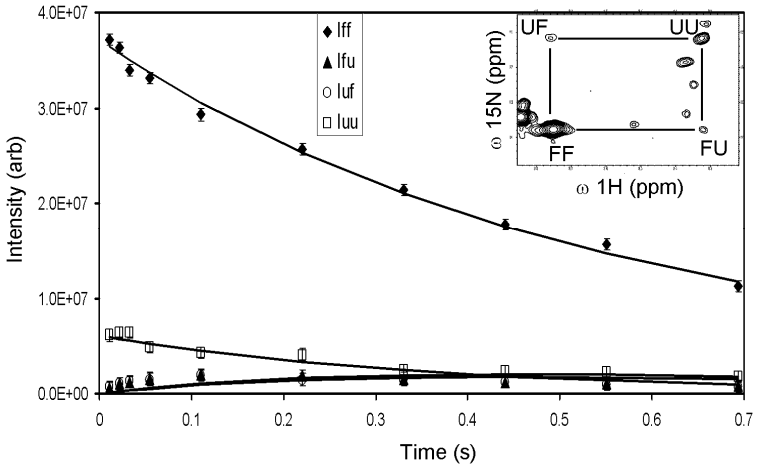



**Figure 2.2: Time-dependence of G95 peak intensities during Nzz experiment**. The experimental data (symbols) and two-state exchange model best fits (solid lines) for G95 auto- and cross-correlation peak intensity are shown. Auto peak intensities for the folded (FF, filled diamonds) unfolded (UU, open squares) conformations, cross-correlation intensities, FU (filled triangles) and UF (open circles), are plotted *vs.* Nzz exchange spectroscopy mixing time. *Inset*: G95 auto- and cross-correlation peaks at a mixing time of 0.1101 seconds.

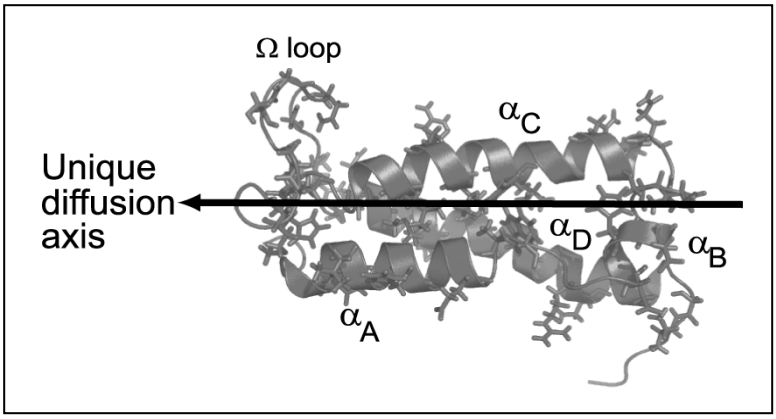



**Figure 2.3: Prolate shape and global folding process of TrAvrPto.** TrAvrPto is a prolate molecule, whose unique diffusion axis (shown as an arrow) is aligned with helices $\alpha_A$, $\alpha_C$, and $\alpha_D$. Forty-one slowly-exchanging amide groups were identified and are shown mapped onto the structure of TrAvrPto as sidechains. They are distributed across all secondary structural elements of TrAvrPto, implying that folding occurs globally: Q35, L37, H41, and E45 of helix $\alpha_A$; A47 and G48 of loop-AB; D52, H54, and E55 of $\alpha_B$; S58, S59, and A61 of loop-BC; Q63, S64, N67, L72, Y73, T76, R78, and L80 of $\alpha_C$; Q86, H87, M90, T91, G92, S94, G95, N97, G99, and L101 of the Ω loop; H103, E104, N105, M109, R110, A112, W116 ε1, R120, and E121 of $\alpha_D$; and G128 and I129 on the C-terminal tail. Image rendered with PyMOL (DeLano. 2002) using NMR structure (PDB file 1R5E) (Wulf et al. 2004), ensemble-averaged *via* MOLMOL (Koradi et al. 1996).

folded autopeak and a single resolved crosspeak. For other residues, three peaks (e.g. N105 and A47) or all four peaks (e.g. G92) were resolved. The unfolded autopeaks were well-resolved for a subset of 17 of these 41 residues (Figure 2.4). Limited spectral resolution prevents further identification of backbone amides in slow exchange, especially in the center of the spectrum where resonances of unfolded TrAvrPto and α-helices congregate (Figure 2.4). The distribution of residues in slow exchange across the structure (Figure 2.3) and the good sampling of the well-resolved peaks (Figure 2.4), indicate that exchange between folded and unfolded states occurs protein-wide.

*TrAvrPto is characterized by global two-state unfolding.*

A series of Nzz spectra were acquired, each with a different time interval, t, during which the $^{15}N$ longitudinal magnetization decayed towards equilibrium and TrAvrPto exchanged between its folded and unfolded conformations. At each time *t*, the variation of $I_{FU}(t)$ and $I_{UF}(t)$ across TrAvrPto is comparable to the experimental uncertainty in peak volume, σ. Conversely, at a given time point, the autopeak intensities show more variation between residues than can be accounted for by σ. The variation in $I_{FF}(t)$ was an order of magnitude greater than σ and, at early times, that of $I_{UU}(t)$ was approximately twice σ. To understand this difference in behavior between the autopeak and crosspeak intensities across the protein, it is informative to derive linear approximations of the Bloch-McConnell equations (Cavanagh et al. 1996, McConnell 1958) at short time intervals ($t << 1$) using the Taylor series relation,

$$\exp(-\lambda * t) \approx 1 - \lambda * t . \qquad \text{Eq. 1}$$

Using this approach, the intensities associated with each amide group in slow chemical exchange are given by,

$$I_{FF}(t) \propto 1 - (R_{1F} + k_{FU}) * t \qquad \text{Eq. 2}$$

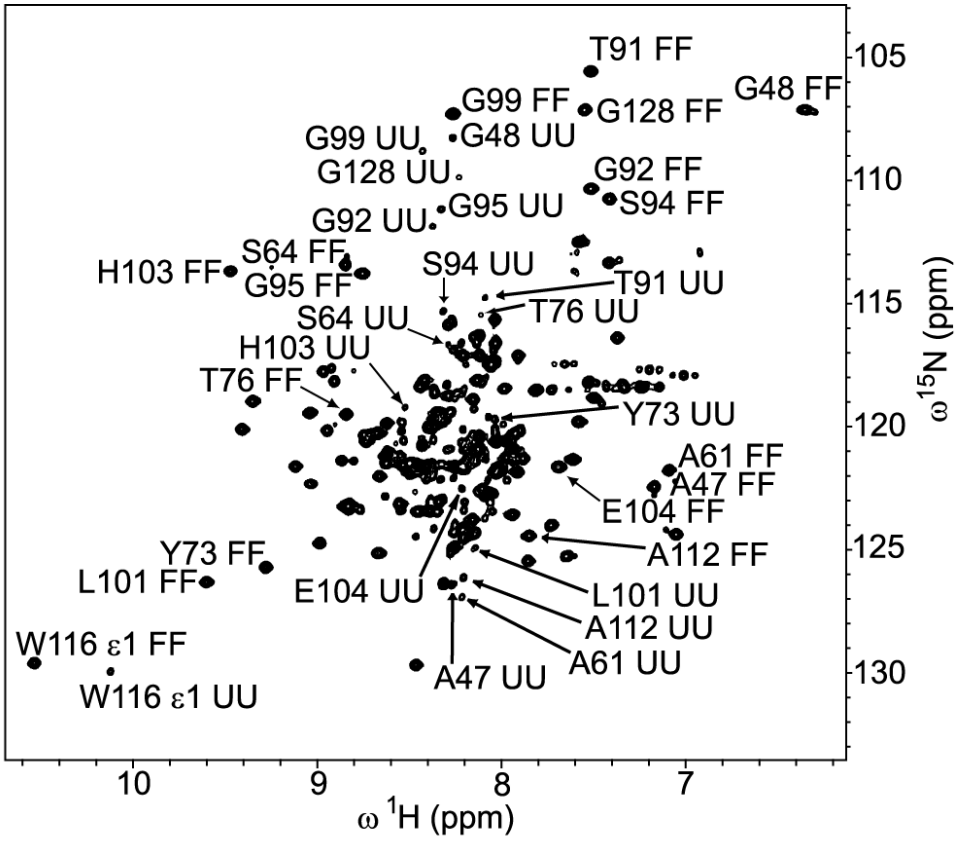



**Figure 2.4: Folded and unfolded TrAvrPto populations are detected at pH 6.1.** Two discrete sets of peaks are observed in a fully-relaxed Nzz spectrum collected with zero mixing time, corresponding to the folded and unfolded conformations of TrAvrPto. The peaks for both unfolded TrAvrPto and the three parallel α-helices of folded TrAvrPto cluster near the center of the spectrum. The FF and UU peaks are labeled for the seventeen residues with assigned UU resonances (A47, G48, A61, S64, Y73, T76, T91, G92, S94, G95, G99, L101, H103, E104, A112, W116 ε1, and G128). Nine of these UU peaks had been assigned previously (Wulf et al. 2004).

$$I_{FU}(t) \propto k_{FU} * t \qquad \text{Eq. 3}$$

$$I_{UF}(t) \propto k_{UF} * t \qquad \text{Eq. 4}$$

$$I_{UU}(t) \propto 1 - (R_{1U} + k_{UF}) * t\,. \qquad \text{Eq. 5}$$

These simplified equations illustrate that, at early time points, the crosspeak intensities are sensitive to the kinetic rate constants only, while the autopeak intensities are sensitive to both the relaxation rates ($R_{1F}$ and $R_{1U}$) and the kinetic rate constants ($k_{FU}$ and $k_{UF}$). The observation that residue-to-residue variations of $I_{FU}(t)$ and $I_{UF}(t)$ are indistinguishable from experimental noise indicates that a single set of global kinetic rate constants, $k_{FU}$ and $k_{UF}$, apply for all residues. Therefore, the larger variation in autopeak intensities is due to differences in $R_{1F}$ and $R_{1U}$ from residue to residue. As discussed in more detail below, folded TrAvrPto is a prolate molecule (Figure 2.3), whose backbone amide $R_{1F}$ are *expected* to vary based on their placement in the protein. The smaller but significant variation in $I_{UU}(t)$ implies that there is some intramolecular variation in $R_{1U}$, perhaps due to residual structure or to the amide group's proximity to the ends of the polypeptide chain.

The above analysis motivates the modeling of TrAvrPto conformational exchange as a two-state process with global $k_{FU}$ and $k_{UF}$ and residue-specific $R_{1F}$ and $R_{1U}$. The resulting global kinetic rate constants (extracted as described in Experimental methods using data from the 17 residues with assigned $I_{UU}$ peaks) are $k_{FU} = 0.33 \pm 0.04\ s^{-1}$ and $k_{UF} = 1.8 \pm 0.3\ s^{-1}$, corresponding to $\Delta G = -RT * \ln(k_{UF}/k_{FU}) = -1.0 \pm 0.1$ kcal/mol. Compared to the ambient thermal energy, RT = -0.6 kcal/mol, TrAvrPto is only modestly stable at pH 6.1. The population in the unfolded ensemble, $p_U$, is

$$p_U = 1/\left[1 + \left(k_{UF} / k_{FU}\right)\right] = 0.16 \pm 0.03\,, \qquad \text{Eq. 6}$$

where $p_U + p_F = 1$ is the total population of TrAvrPto. The peak intensities for each of the 41 slowly-exchanging residues were then individually fitted for $R_{1F}$ and $R_{1U}$, using

the global $k_{FU}$ and $k_{UF}$ as set constants. These fits yielded chi square probability values, Q (Press et al. 1988), where the criterion for the rejection of a model was Q<0.001. Thirty-five of the 41 residues had acceptable fits (Table 2.1; $R_{1F}$ values are plotted in Figure 2.5A). The remaining residues (S59, L72, N97, G99, H103, and R110) had peaks that suffered from overlap, resulting in unacceptable fits. Although G99 and H103 were included in the global kinetic rate constants analysis, subsequent re-fitting without these residues yielded statistically equivalent rates. In general, with $k_{FU}$ and $k_{UF}$ as set constants, $R_{1F}$ and $R_{1U}$ could be acceptably fitted using, at minimum, well-resolved peaks including the folded autopeak and one of the crosspeaks.

*Longitudinal relaxation and HSQC data provide an independent view of TrAvrPto folding kinetics and pH-dependence of stability.*

The stability of TrAvrPto is pH-dependent with greater stability at pH 7.0 than at pH 6.1 (Figure 2.6). As determined from the Nzz data of 0.63 mM TrAvtPto, 16% of the population is unfolded at pH 6.1. At higher pH and higher concentration (pH 7.0 and 1.2 mM TrAvrPto), the unfolded population (quantified using a fully-relaxed HSQC spectrum) decreased to 2%. As discussed in the next section, the increased relative intensity of the folded peaks at pH 7.0 is due purely to the pH-dependence of TrAvrPto stability, not to self-association at the higher protein concentration.

In contrast to the $R_{1F}$ values extracted from Nzz data ($R_{1F}^{Nzz}$), which are the relaxation rates in the absence of exchange, the effective relaxation rates constants extracted from $^{15}N$ $T_1$ experiments ($R_{1F}$(eff)) can contain contributions from the kinetic unfolding rates. Consequently, $R_{1F}$(eff) values measured for TrAvrPto should reflect the folding equilibrium and be pH-dependent. $R_{1F}$(eff) values at pH 6.1 are offset to greater values than those at pH 7.0 (Figure 2.5B). The conserved profile of

Table 2.1: **$R_{1F}$ and $R_{1U}$ extracted from pH 6.1 Nzz data**

| Residue | $R_{1F}$ ($s^{-1}$) | $R_{1U}$($s^{-1}$) | Q [a] | Residue | $R_{1F}$ ($s^{-1}$) | $R_{1U}$ ($s^{-1}$) | Q[a] |
|---|---|---|---|---|---|---|---|
| Q35 | 1.11±0.06 | 3.3±0.8 | 0.503 | Q86 | 1.26±0.07 | 2.9±1.2 | 0.253 |
| L37 | 1.17±0.06 | 3.0±0.9 | 0.354 | H87 | 1.18±0.07 | 3.5±1.3 | 0.597 |
| H41 | 1.10±0.05 | 3.9±0.8 | 0.380 | M90 | 1.20±0.07 | 3.2±1.1 | 0.221 |
| E45 | 1.16±0.06 | 4.2±1.2 | 0.666 | T91 | 1.34±0.07 | 3.7±1.4 | 0.201 |
| A47 | 1.11±0.06 | 3.0±0.7 | 0.056 | G92 | 1.40±0.06 | 2.1±0.8 | 0.365 |
| G48 | 1.18±0.06 | 1.4±0.5 | 0.012 | S94 | 1.36±0.05 | 3.1±0.8 | 0.118 |
| D52 | 1.47±0.08 | 2.9±1.2 | 0.402 | G95 | 1.48±0.07 | 1.1±0.5 | 0.035 |
| H54 | 1.34±0.05 | 4.3±0.9 | 0.180 | L101 | 1.39±0.06 | 2.1±0.5 | 0.015 |
| E55 | 1.55±0.10 | 2.6±1.3 | 0.156 | E104 | 1.44±0.07 | 2.1±0.7 | 0.408 |
| S58 | 1.52±0.06 | 3.7±1.0 | 0.505 | N105 | 1.13±0.07 | 3.2±0.7 | 0.306 |
| A61 | 1.29±0.05 | 2.5±0.6 | 0.400 | M109 | 1.21±0.06 | 3.4±1.0 | 0.364 |
| Q63 | 1.32±0.08 | 2.4±0.7 | 0.070 | A112 | 1.12±0.04 | 2.1±0.6 | 0.390 |
| S64 | 1.25±0.07 | 2.2±0.6 | 0.499 | W116 ε1 | 1.17±0.06 | 1.2±0.5 | 0.005 |
| N67 | 1.17±0.06 | 3.5±1.1 | 0.526 | R120 | 1.12±0.06 | 2.7±1.0 | 0.501 |
| Y73 | 1.14±0.07 | 2.7±0.7 | 0.086 | E121 | 1.25±0.05 | 4.2±1.2 | 0.021 |
| T76 | 1.22±0.08 | 2.1±0.7 | 0.115 | G128 | 1.35±0.06 | 2.3±0.8 | 0.219 |
| R78 | 1.28±0.09 | 2.3±1.0 | 0.049 | I129 | 1.34±0.05 | 5.4±1.6 | 0.749 |
| L80 | 1.09±0.06 | 3.0±1.0 | 0.197 | | | | |

[a]Q-statistic reflects the quality of the 2-parameter { $R_{1F}$, $R_{1U}$ } Nzz fitting using global kinetic rate constants, $k_{FU}$ and $k_{UF}$. A model is rejected using the criterion, Q<0.001.

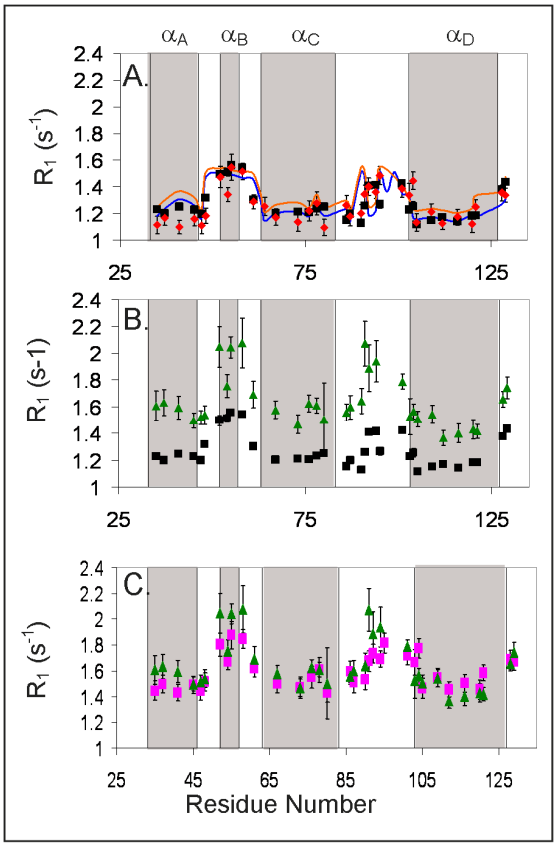



**Figure 2.5: The effect of slow chemical exchange and global anisotropy on $R_1$.** TrAvrPto $R_1$ relaxation rates are plotted *vs.* residue number (the four alpha helices are shaded in grey boxes for reference) for three different comparisons: A. *Relaxation in the absence of TrAvrPto folding.* $R_1$ values for folded TrAvrPto in the absence of chemical exchange extracted from Nzz data measured at pH 6.1 ($R_{1F}^{Nzz}$, red diamonds) are comparable to those obtained from a $T_1$ experiment at pH 7.0 ($R_{1F}$(eff), black squares), where TrAvrPto is nearly completely folded. The $R_{1F}$ profiles were simulated by modeling TrAvrPto as a prolate monomer (blue line) and as a prolate monomer/oblate dimer equilibrium (orange line). B. *The pH-dependence of $R_{1F}$(eff) from $T_1$ experiments:* $R_{1F}$(eff) values obtained from $T_1$ experiments at pH 6.1 (green triangles) and at pH 7.0 (black squares) show nearly constant offset, with $R_{1F}$(eff) at pH 6.1 shifted to greater values than at pH 7.0. C. *Nzz-extracted parameters accurately predict $R_{1F}$(eff):* Calculated $R_{1F}$(eff) values (pink squares), obtained using Eq. 7 and parameters $R_{1F}^{Nzz}$ and $k_{FU}$ extracted from the pH 6.1 Nzz data, are comparable to $R_{1F}$(eff) values obtained from $T_1$ experiments at pH 6.1 (green triangles). The error bars reflect experimental uncertainties for $R_{1F}$ extracted from Nzz and $T_1$ data and propagated uncertainties for the calculated $R_{1F}$(eff).

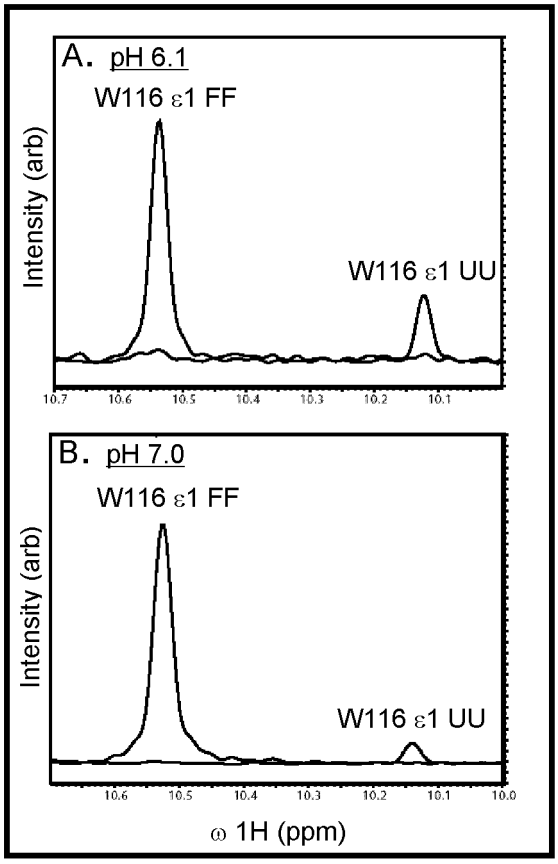



**Figure 2.6: TrAvrPto is more stable at pH 7.0 than at pH 6.1.** The relative volumes of peaks associated with folded and unfolded protein states depend on the stability of the protein. The TrAvrPto W116 ε1 sidechain amide group gives rise to well-resolved peaks in both folded ($I_{FF}$) and unfolded ($I_{UU}$) states. One-dimensional lineshapes of the W116 ε1 $I_{FF}$ and $I_{UU}$ peaks extracted from a 2D $^{15}N$-$^{1}H$ HSQC spectrum at A. pH 6.1 and B. pH 7.0 illustrate the larger unfolded TrAvrPto population at pH 6.1 than at pH 7.0. Lineshapes were generated using NMRDraw (Delaglio et al. 1995) and were overlaid using Adobe Illustrator.

the $R_{1F}$(eff) *versus* residue number curves indicates that the molecular shape of TrAvrPto does not dramatically change with changing pH. Rather, the offset between $R_{1F}$(eff) profiles at pH 6.1 and 7.0 can be explained by the dependence of $R_{1F}$(eff) on the rate of unfolding ($k_{FU}$). For a single residue that interconverts between its folded and unfolded states on the slow exchange timescale, if we assume $R_{1U}>>R_{1F,}$ $p_F>>p_U$ (Leigh 1971), and $R_{1U}>>k_{UF}$, then the estimated effective longitudinal relaxation rate for the folded state will be

$$R_{1F}(eff) \approx R_{1F} + k_{FU} = R_{1F} + p_U k_{ex}, \qquad \text{Eq. 7}$$

where $k_{ex}(=k_{FU}+k_{UF})$ is the exchange rate constant and $R_{1F}$ is the longitudinal relaxation rate of the folded state independent of exchange. It is known from the Nzz fittings that $R_{1U}$ is generally greater than $R_{1F}$ (Table 2.1), and the population $p_F > 5p_U$. The final approximation ($R_{1U}>>k_{UF}$) is an oversimplification; however, the uncertainties in the fitted $R_{1U}$ prohibit the use of a more complete model that depends on this parameter and, as demonstrated below, the fits of the data to Eq. 7 justify the use of this model.

The exchange process between folded and unfolded TrAvrPto explains the difference at pH 6.1 between $R_{1F}^{Nzz}$ and $R_{1F}$(eff) values. The $R_{1F}^{Nzz}$ values are theoretically equivalent to the $R_{1F}$ in Eq. 7. Experimental $R_{1F}$(eff) can be simulated using the Nzz parameters ($R_{1F}^{Nzz}$ and $k_{FU}$) and Eq. 7. The resulting simulated $R_{1F}$(eff) values accurately predict the experimental $R_{1F}$(eff) values measured at pH 6.1 (Figure 2.5C, Appendix Table 2.2). As a further validation test, $R_{1F}^{Nzz}$ values should be approximately equal to $R_{1F}$(eff) values measured at pH 7.0, where the effects of the conformational exchange are minimal due to the negligible (2%) unfolded population. Indeed, these values are in general within experimental uncertainty of each other (Figure 2.5A). These results demonstrate that the variation of relaxation at different

pH values is fully explained by the global folding exchange process and, importantly, provide independent verification of the Nzz-derived kinetic parameters.

*The concentration-independent $R_{1F}$ rates of TrAvrPto are consistent with a prolate folded monomer*

The exchange-independent $R_{1F}^{Nzz}$ values differ across residues in a manner that can be explained by the prolate shape of folded TrAvrPto (Figure 2.5A), where alpha helices $\alpha_A$, $\alpha_C$, and $\alpha_D$ are aligned along its unique diffusion axis and the Ω loop and $\alpha_B$ are approximately perpendicular to this axis (Figure 2.3). A TrAvrPto monomer can rotate more rapidly about the long, unique axis than it can tumble end-over-end, a phenomenon that changes the relaxation rates and, so, an N-H bond has different $R_1$ values depending on how it is oriented on the molecule. For folded TrAvrPto, $R_{1F}$ can be simulated using previously established theory (Lipari et al. 1982, Schurr et al. 1994, Woessner D.E. 1962) and the N-H bond vectors from the known structure of the protein (PDB file 1R5E) (Wulf et al. 2004). Using the monomeric, prolate TrAvrPto structure without a hydration layer yields a $R_{1F}$ profile that matches the fitted $R_{1F}^{Nzz}$ profile, but is shifted to higher values, reflecting a smaller molecule (data not shown). Expanding the effective dimensions of TrAvrPto to 6.1 nm long and 3.1 nm wide aligns the simulated $R_{1F}$ profile with the $R_{1F}^{Nzz}$ curve (Figure 2.5A). Although not necessarily a unique solution, the alignment of the simulated and experimental data is striking. The larger effective size could correspond to a hydration layer of roughly 1.5 water molecules, drag from the loops and tails, or an internally-hydrated and less compact average folded structure. A water-swollen structure is not implausible; the presence of internal waters has been demonstrated previously both experimentally and in MD folding simulations (Bezsonova et al. 2006, Nishiguchi et al. 2007, Shea et al. 2002). Since $R_{1F}^{Nzz}$ rates are independent of conformational exchange, these values

should be dominated by global tumbling. The above analysis demonstrates that the $R_{1F}^{Nzz}$ values can in fact be reproduced based solely on the TrAvrPto monomer structure and further validates the global kinetic rates used in extraction of these values.

Previous work from this lab hypothesized that TrAvrPto is a transient dimer at pH 6 (Wulf et al. 2004). Indeed, it is also possible to reproduce the $R_{1F}^{Nzz}$ values extracted from 0.63 mM Nzz data (Figure 2.5A) as the population-weighted average $R_{1F}$ from monomer and side-by-side (oblate) dimer conformations in fast chemical exchange. Oblate dimers were modeled as 2.6 nm by 7.7 nm oblate ellipsoids and monomers were modeled as 5.6 nm by 2.6 nm prolate ellipsoids, dimensions that include one hydration layer. A reasonable fit of the $R_{1F}^{Nzz}$ values was obtained by weighting the monomer population by $p_M$=0.7 and the oblate dimer subunit population by $p_D$=(1-$p_M$)=0.3. These populations correspond to weak self-association, described by the dissociation constant

$$K_d = 2*p_M*p_M*[T]/p_D = 2 \text{ mM}. \qquad \text{Eq. 8}$$

In this simulation, [T] is 0.63 mM, the total TrAvrPto concentration during the Nzz experiments.

The monomer-dimer model fails to correctly predict the behavior of $R_{1F}$(eff) data at other concentrations, however. Additional $R_{1F}$(eff) values measured at pH 6.1 were obtained for total concentrations of 0.34 and 0.87 mM. Given $K_d$= 2 mM and the previously estimated dimensions of the monomers and oblate dimers, the model predicts $p_M$=0.79 at 0.34 mM and $p_M$=0.65 at 0.87 mM. The predicted $R_{1F}$ values for 0.34 mM TrAvrPto were offset by an average of +0.18 $s^{-1}$ relative to those for 0.87 mM TrAvrPto (Appendix Table 2.3). Despite the contribution of $k_{FU}$ to $R_{1F}$(eff) expected at pH 6.1 (see Eq. 7), the average separation between experimental $R_{1F}$(eff) data at 0.34 and 0.87 mM should be equivalent to +0.18 $s^{-1}$ since $k_{FU}$ is a constant.

However, the experimental 0.34 mM and 0.87 mM $R_{1F}$(eff) data have negligible average separation, -0.04 $s^{-1}$, with a standard deviation on the order of the experimental uncertainty (Appendix Table 2.2). This inequality demonstrates that, while the 0.63 mM $R_{1F}^{Nzz}$ profile can be coincidentally fit using a monomer-dimer equilibrium model, the observed concentration-independence of $R_{1F}$(eff) at pH 6.1 rules out this model. Additionally, measurements at pH 6.8 (0.92 mM) and 7.0 (1.2 mM) yielded $R_{1F}$(eff) values within uncertainty of each other (Appendix Table 2.2), further refuting the dimer model. The "dimers" observed in the previous DLS and native-gel studies most likely were unfolded protein, which would have a larger radius of gyration than the folded conformation (Piaggio et al. 2007). Altogether, these results argue against self-association and suggest that the dominant folded form of TrAvrPto is monomeric.

***Discussion***

The kinetic and thermodynamic properties of an effector protein may have important consequences on its ability to be efficiently translocated by the TTSS, a process requiring at least partial unfolding of the protein (Johnson et al. 2005). AvrPto is a well-characterized effector protein of *P. syringae* pv tomato and an excellent model system to probe requirements for efficient effector translocation. NMR spectroscopy allowed the folding of TrAvrPto to be probed at multiple locations across the protein. The multiple probes enrich the analysis, demonstrating that the entire TrAvrPto molecule, not just the core or the long Ω loop, folds cooperatively with two-state kinetics. The low stability (16% unfolded population at pH 6.1), slow folding rate ($k_{UF}$=1.8±0.3 $s^{-1}$), and correspondingly long lifetime of the unfolded state ($\tau=1/k_{UF}$=0.6±0.1 s) of TrAvrPto would allow a pool of readily secretable protein to

be maintained, especially given that the full-length effector protein, AvrPto, is even less stable than TrAvrPto at pH 6 (Wulf et al. 2004).

After translocation into the host cell cytoplasm, an effector must be able to spontaneously fold into its functional form. As shown here, TrAvrPto is almost completely folded at pH 7.0, where the observed 2% unfolded population corresponds to a moderate folding free energy of -2.3 kcal/mol. Therefore, TrAvrPto should be predominately folded within the cytoplasm of the host cell.

The pH-dependence of TrAvrPto potentially explains at least part of the observed difference in full-length AvrPto's secretion efficiency under different culturing conditions. Expression of the TTSS and its associated proteins, including AvrPto, can be induced in *P. syringae* pv tomato using a *hrp*-inducing minimal medium (Rahme et al. 1992, Xiao et al. 1992). In this medium, AvrPto is expressed when the external pH, $pH_{ext}$, is 6 or 7; however, AvrPto is only secreted at $pH_{ext}$=6 and not at pH 7 (van Dijk et al. 1999). The proton concentration inside a pathogen, $pH_{int}$, can differ from $pH_{ext}$. Indeed, the rotation of the *E. coli* flagellium in the evolutionally-related flagellar apparatus (Jin et al. 2003) depends on the proton gradient, $\Delta pH$, between the cytoplasm and the periplasm (Minamino et al. 2003). $\Delta pH$ is maintained even when $pH_{ext}$ is varied between 5.5 and 8.0 with $pH_{int}$ remaining greater than $pH_{ext}$ (Minamino et al. 2003). In *Yersinia enterocolitica*, secretion through the TTSS has also been found to be dependent on $\Delta pH$, where again $pH_{int}$ is greater than $pH_{ext}$ (Wilharm et al. 2004). Assuming that secretion by *P. syringae* requires a similar $\Delta pH$, then $pH_{int}$ would exceed 7 if $pH_{ext}$ is neutral. Our results indicate that TrAvrPto would be predominantly folded inside the pathogen under these conditions, providing an explanation for the lack of AvrPto secretion in *hrp*-inducing minimal medium at neutral pH (van Dijk et al. 1999).

The structure and low stability of TrAvrPto may have additional implications for its ability to compete for passage through the TTSS without chaperone assistance. One of the conserved proteins in the TTSS inner-membrane ring is a putative ATPase (Ghosh 2004) required for translocation (Wilharm et al. 2004). The TTSS ATPases are unfoldases (Gorbalenya et al. 1993, Lupas et al. 2002) whose functional similarity to the AAA+ ATPases has been suggested (Akeda et al. 2005). For the AAA+ ATPase domain of the ClpX bacterial protease, the rate of catalyzed unfolding by the hexameric ring decreased with increasing substrate stability (Kenniston et al. 2003). The identity of the secondary structural element immediately after a degradation signal also affects the degradation efficiency (Lee et al. 2001). The greatest enhancement in efficiency occurs when this element is a surface loop or α-helix that can be removed relatively easily and whose removal destabilizes the protein core (Lee et al. 2001). By analogy, AvrPto appears to possess the correct architecture and stability that would allow for efficient action of the putative TTSS ATPase and subsequent translocation into the host. AvrPto's secretion signal is located on its N-terminus, as in all effector proteins (Anderson et al. 1997, Sory et al. 1994), and the first secondary structural element is helix $\alpha_A$. The core of TrAvrPto is only slightly stable and the removal of $\alpha_A$ from the bundle would be expected to destabilize the fold. Therefore, an additional consideration for secretion and translocation efficiency may be the direction-specific mechanical force exerted on the N-terminal region of the effector protein structure by the TTSS ATPase (Matouschek 2003).

Two real-time confocal microscopy studies on the translocation rates for pathogen effector proteins into the host *via* the TTSS provide an *in vivo* functional context for our TrAvrPto folding kinetic rates. For *Shigella flexneri*, it was found that 50% of the populations of effector proteins IpaB and IpaC was translocated out of the bacterium in $t_{1/2}$~240s (Enninga et al. 2005). Assuming a mono-exponential decay

curve, the translocation rate is $k = \ln 2/t_{1/2} \sim 0.003\ s^{-1}$. For *Salmonella enterica,* the translocation rate was estimated by the build-up of effector protein SipA within the host cell over time using a straight-line approximation, $y(t)=mt$, where the observed slope was m=7-60 molecules/s and the average total number of molecules translocated was $y_0 = 6*10^3 \pm 3*10^3$ molecules (Schlumberger et al. 2005). This is equivalent to a linear Taylor-series approximation of an exponential build-up, where $y(t) = y_0*\{1-\exp(-kt)\}$ is approximated by $y(t)=y_0*\{1-(1-kt)\}=y_0kt$. Hence, the corresponding mono-exponential translocation rate, k (estimated as $m/y_0$), was on order of $k \sim 0.01\ s^{-1}$. Based on these two microscopy studies, a rough estimate of the general TTSS translocation rate is on order of $10^{-3}$ to $10^{-2}\ s^{-1}$. Therefore, the $k_{FU}=0.33\pm0.04\ s^{-1}$ unfolding rate of TrAvrPto observed here in solution would not represent a rate-limiting step of the secretion process.

Altogether, these results offer important insights into the folding and unfolding of TrAvrPto and, ultimately, into how the thermodynamics and kinetics of effector proteins affect their translocation efficiency. The folding kinetics, even while viewed in isolation without the TTSS or any putative ATPase unfolding machinery, are sufficient to sustain a translocation-competent pool of TrAvrPto. This observation is consistent with observed ability of AvrPto to be translocated through the TTSS (van Dijk et al. 1999). This application of NMR spectroscopy to detect and measure the pH-dependent stability and folding kinetics of TrAvrPto lays the groundwork for future studies of TrAvrPto mutants to determine the mechanism of a possible pH-triggered folding switch and the thermodynamic and kinetic thresholds that govern the translocation efficiency of effector proteins.

***Experimental Methods***

A. NMR sample preparation

The TrAvrPto expression vector encodes a fusion protein consisting of a N-terminal 6-His tag, an HRV 3c protease (3CPro) cut site (Walker et al. 1994), and TrAvrPto residues 29 – 133 cloned into the ampicillin-resistant plasmid pQEP (Qiagen) (gift from Pete Pascuzzi and Greg Martin, Cornell University). The protein was expressed in kanamycin-resistant M15 [pREP4] *E. coli* cells (Qiagen) and cultured in M9 minimal media with $^{15}NH_4Cl$ (Isotec Inc.), 1 mM thiamine HCl (SIGMA), 100 µg/ml ampicillin (SIGMA, sodium salt), and 30 µg/ml kanamycin monosulfate (SIGMA). All cultures were grown at 37°C with shaking at ~120 rpm. One L cultures were started with a 50-ml overnight miniculture and grown to $O.D._{600}$~0.8. Protein expression was induced with 1 mM IPTG (Acros Organics), the cells were grown for an additional two hours, and then pelleted. Pellets were washed in binding buffer {50 mM $NaH_2PO_4 \bullet H_2O$ (Mallinckrodt), 300 mM NaCl (SIGMA), 10 mM imidazole (Sigma), pH 8}, re-pelleted, and then resuspended in lysis buffer (20 ml binding buffer + 50 µl protease inhibitor) on ice. The cells were lysed with 1 mg/ml lysozyme (SIGMA) and 0.75 mg/ml deoxylcholic acid (SIGMA) and two rounds of sonication. After high-speed centrifugation, supernatant was filtered through a 0.8 µm syringe filter (Corning Inc.) onto a Ni-NTA column (Qiagen) with a 3-5 ml bed volume. The column was washed with 20 bed volumes of binding buffer, then 20 bed of volumes wash buffer {50 mM $NaH_2PO_4 \bullet H_2O$, 300 mM NaCl, 20 mM imidazole, pH 8}. Cleavage of the 6His-TrAvrPto fusion protein to yield tag-free TrAvrPto was accomplished by adding ~1 mg 6His-3CPro fusion protein (prepared as described elsewhere: (Jayaraman et al. 2007)) to the Ni-NTA column and tumbling at 4°C overnight. The desired TrAvrPto was eluted with wash buffer and dialyzed into NMR buffer {10 mM $NaH_2PO_4 \bullet H_2O$, 5 mM $KH_2PO_4$ (SIGMA), 225 mM NaCl, pH

6.9}. The final concentration was estimated *via* UV absorbance at 280 nm using the theoretical extinction coefficient of 9530 $M^{-1}cm^{-1}$ for TrAvrPto (Gasteiger E. et al. 2005). The pH values of the NMR samples were adjusted using HCl and NaOH.

B. NMR Spectroscopy

The NMR experiments were performed at 25°C on a Varian Inova 600 MHz spectrometer using a {H,C,N} z-axis gradient probe. The spectra were processed using NMRpipe and NMRDraw (Delaglio et al. 1995). The data for all experiments were apodized using a shifted sine-bell function and then zero-filled. The peak volumes were determined using nlinLS (Frank Delaglio, NIH/NIDDK) or Sparky (Goddard et al. ), assuming a Gaussian lineshape.

The pH 6.1 experiments were performed on $^{15}$N-labeled TrAvrPto samples at concentrations of 0.63 mM for the Nzz spectra (Farrow et al. 1994b, Montelione et al. 1989) and 0.34 mM and 0.87 mM for $T_1$ experiments (Farrow et al. 1994a). The $T_1$ experiments at 0.34 mM were taken at the timepoints 0.01, 0.01, 0.04, 0.08, 0.16, 0.32, 0.64, 1.0, 1.0, and 2.0 seconds. At 0.87 mM, the $T_1$ experimental timepoints were 0.01, 0.01, 0.02, 0.04, 0.08, 0.08, 0.16, 0.32, 0.64, 1.0, and 2.0 seconds. $T_1$ experiments (with timepoints at 0.01, 0.77, and 0.77 seconds) were recorded for a 1.2 mM, pH 7.0 $^{15}$N-labeled sample. Additionally, a set of $T_1$ data for pH 6.8, 0.92 mM TrAvrPto was collected for timepoints 0.01, 0.77, and 0.77 seconds. In the $T_1$ experiments at each concentration, the timepoints were recorded out of sequential order to avoid bias.

The pH-dependence of TrAvrPto stability was evaluated using $^{1}$H-$^{15}$N 2D NMR spectroscopy. At pH 6.1, a fully-relaxed (10 second delay interval) Nzz spectrum with zero mixing time was used. It was subsequently found that a 5 second delay was sufficient for complete recovery of equilibrium magnetization. The stability at pH 7.0

was determined using the fully-relaxed (5 second delay interval) HSQC spectrum (Mulder et al. 1996) of 1.2 mM TrAvrPto.

C. Nzz exchange spectroscopy

Nzz exchange spectroscopy (Farrow et al. 1994b, Montelione et al. 1989) was performed on a 0.63 mM TrAvrPto solution at pH 6.1 and 25°C. Spectra were taken with mixing times of 0.01101, 0.02202, 0.03303, 0.04404, 0.05505, 0.1101, 0.2202, 0.3303, 0.4404, 0.5505, and 0.69363 seconds and replicate spectra at 0.01101, 0.02202, 0.1101, and 0.4404 seconds. As with the $T_1$ experiments, the timepoint spectra were recorded out of sequential order to avoid bias. A delay interval of 3 seconds was used in all spectra.

Identification of sets of auto- and cross-peaks was aided by the judicious use of overlaid spectra. The intensities in the frequency domain of all Nzz spectra were added together using NMRPipe ('addNMR' command, (Delaglio et al. 1995)), creating a summed spectrum that contains all auto- and cross-peaks. The Nzz spectrum with zero mixing time contains only the folded and unfolded autopeaks. The crosspeaks and unfolded autopeaks have decayed to almost zero in the 0.69 second mixing time spectrum; only the folded resonances are seen at high contours (Figure 2.2). Classification of the peaks was achieved by noting that the folded autopeaks are present in all three spectra, the unfolded autopeaks are in the zero mixing time and summed spectra, and the crosspeaks are only observed in the summed spectrum.

Two types of standard deviations were used in the study. The first is the experimental uncertainty in peak volume, $\sigma$, which was estimated as the mean replication error:

$$\sigma^2 = \frac{1}{2}\left\langle (x-y)^2 \right\rangle . \qquad \text{Eq.A1}$$

The term $\langle (x-y)^2 \rangle$ is the mean deviation squared of peak volumes from replicate spectra. The average is over all peaks, whether auto- or cross-peak, used in the Nzz analysis and is over the four sets of replicate spectra. The second type is specific to each timepoint *t* and to each kind of peak ($I_{FF}(t)$, $I_{FU}(t)$, $I_{UF}(t)$, or $I_{UU}(t)$). These latter estimates ($\sigma_{FF}(t)$, $\sigma_{FU}(t)$, $\sigma_{UF}(t)$, $\sigma_{UU}(t)$) measure the residue-to-residue variability in peak intensity within each spectrum.

The dependence of the peak intensity on mixing time (t) was modeled with the two-state solution of the Bloch-McConnell equations (Farrow et al. 1994b, McConnell 1958). The global kinetic rate constants $\{k_{FU}, k_{UF}\}$ were extracted by simultaneously fitting $\{I_{FF}(t), I_{FU}(t), I_{UF}(t), I_{UU}(t)\}$ data of multiple residues. Due to issues of computer memory and non-linear minimization convergence, it was only possible to simultaneously fit data from a small subset of seven residues. Specifically, a total of 280 data points (4 curves per residue x 10 time points per curve x 7 residues = 280 data points) were fitted to 16 adjustable parameters ($R_{1U}$ and $R_{1F}$ for each of 7 residues plus global $k_{FU}$ and $k_{UF}$). It was critical to the robustness of the fitting that each member of the seven-residue fitting set possessed well-resolved $I_{FF}$ and $I_{UU}$ autopeaks. Since the crosspeak intensities are statistically similar for all residues, the average $I_{FU}(t)$ and $I_{UF}(t)$ were used in the fittings. These intensity data were fitted using Levenberg-Marquardt non-linear least squares minimization (Levenberg 1944, Marquardt 1963) and the uncertainty in intensities were simulated *via* Monte Carlo with 100 repetitions. The value and uncertainty of the adjustable parameters were the mean and standard deviation over the 100 repetitions.

In order to prevent bias in the parameter extraction due to the choice of fitting set members, the data fitting was repeated multiple times, each round using intensities from a set of seven residues randomly selected from the 17 residues that have well-resolved $I_{UU}$ peaks at time zero (Figure 2.4). Allowing the fitting set choice to vary in

10 rounds yields results statistically identical to results obtained using 100 rounds. The average extracted global $k_{FU}$ and $k_{UF}$ were then used as constants to extract $R_{1F}$ and $R_{1U}$ values by separately fitting the intensities of individual residues. All fittings were implemented using Matlab version R2006a.

The quality of fit for each individual residue $\{R_{1F}, R_{1U}\}$ was evaluated with the chi-square probability, Q (Press et al. 1988). Due to the presence of experimental noise, the minimized $\chi^2$ for each fit can vary according to the chi square distribution. The value Q is the probability that, if one re-fit the data, one would find a larger $\chi^2$ simply due to the experimental uncertainty. A model fitting is rejected if $Q<0.001$.

D. Simulation of $R_{1F}$ relaxation rate constants

Anisotropic tumbling alters the relaxation rates of amide groups depending on how they are oriented within the macromolecule (Woessner D.E. 1962). TrAvrPto was modeled as various types of ellipsoids that are axially-symmetric about their unique diffusion axis. The viscosity, 0.0010019 kg/s*m for $H_2O$ at 293.15 K (Weast. 1982), and temperature, 298.15 K, were used in simulations of $R_{1F}$.

$R_{1F}$ values were calculated for each residue with the aid of the structure of TrAvrPto (PDB file 1R5E) (Wulf et al. 2004), the atomic coordinates of which were ensemble-averaged using MOLMOL (Koradi et al. 1996) and then transformed into the inertial reference frame using the in-house program, NORMAdyn (Pawley et al. 2001, Pawley et al. 2002). The principal axis in the inertial frame is aligned with the unique diffusion axis of a prolate ellipsoid and is perpendicular to the unique diffusion axis of an oblate ellipsoid. The spectral density function for completely rigid, axially-symmetric anisotropic proteins (Lipari et al. 1982, Schurr et al. 1994) was used to calculate $R_1$ using the standard approach (Abragam. 1961), assuming $\Delta\sigma$=-163 ppm and $<r_{NH}>$=1.04 Å (Palmer 2001). For the simulation of folded monomer (M) and dimer (D) in fast exchange, the population-weighted $R_{1F}$ was calculated:

$$R_{1F}=p_M*R_{1F}(\text{monomer})+p_D*R_{1F}(\text{dimer}) \quad \text{Eq. A2}$$

where $p_M=1-p_D$.

***Acknowledgements***

We thank Dr. Pete Pascuzzi and Professor Gregory Martin for reagents and helpful discussions, Kyriacos Leptos for construction of the 6His-3CPro expression vector, Kevin Tan for his contributions to the TrAvrPto purification protocol, and Soumya De, Bhargavi Jayaraman, Lea Michel, Valerie Anderson, Beth R. Dawson, and Alex Greenwood for helpful editing comments. This work was supported by NSF grant MCB-0641582 and NIH Molecular Biophysics Training Grant T32 GM08267.

REFERENCES

Abragam, A. 1961. *Principles of nuclear magnetism,* 1st ed. 1-599. Clarendon Press, Oxford.

Akeda, Y. and Galan, J.E. 2005. Chaperone release and unfolding of substrates in type III secretion. *Nature* **437:** 911-915.

Anderson, D.M. and Schneewind, O. 1997. A mRNA signal for the type III secretion of Yop proteins by Yersinia enterocolitica. *Science* **278:** 1140-1143.

Bezsonova, I., Korzhnev, D.M., Prosser, R.S., Forman-Kay, J.D. and Kay, L.E. 2006. Hydration and packing along the folding pathway of SH3 domains by pressure-dependent NMR. *Biochemistry* **45:** 4711-4719.

Blocker, A., Jouihri, N., Larquet, E., Gounon, P., Ebel, F., Parsot, C., Sansonetti, P. and Allaoui, A. 2001. Structure and composition of the Shigella flexneri "needle complex", a part of its type III secreton. *Mol. Microbiol.* **39:** 652-663.

Bogdanove, A.J. and Martin, G.B. 2000. AvrPto-dependent Pto-interacting proteins and AvrPto-interacting proteins in tomato. *Proc. Natl. Acad. Sci. U.S.A.* **97:** 8836-8840.

Brown, I.R., Mansfield, J.W., Taira, S., Roine, E. and Romantschuk, M. 2001. Immunocytochemical localization of HrpA and HrpZ supports a role for the Hrp pilus in the transfer of effector proteins from Pseudomonas syringae pv. tomato across the host plant cell wall. *Mol. Plant-Microbe Interact.* **14:** 394-404.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G.I. and Skelton, N.J. 1996. *Protein NMR spectroscopy: Principles and practice,* 243-299. Academic Press, San Diego, CA.

Chang, J.H., Rathjen, J.P., Bernal, A.J., Staskawicz, B.J. and Michelmore, R.W. 2000. AvrPto enhances growth and necrosis caused by Pseudomonas syringae pv. tomato in tomato lines lacking either Pto or Prf. *Mol. Plant-Microbe Interact.* **13:** 568-571.

Cheng, L.W. and Schneewind, O. 1999. Yersinia enterocolitica type III secretion. On the role of SycE in targeting YopE into HeLa cells. *J. Biol. Chem.* **274:** 22102-22108.

Cordes, F.S., Komoriya, K., Larquet, E., Yang, S., Egelman, E.H., Blocker, A. and Lea, S.M. 2003. Helical structure of the needle of the type III secretion system of Shigella flexneri. *J. Biol. Chem.* **278:** 17103-17107.

Dawson, J.E. and Nicholson, L.K. 2008. Folding kinetics and thermodynamics of Pseudomonas syringae effector protein AvrPto provide insight into translocation via the type III secretion system. *Protein Sci.* **17:** 1109-1119.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6:** 277-293.

DeLano, W.L. 2002. *The PyMOL molecular graphics system,* DeLano Scientific, San Carlos, CA, USA.

Eichelberg, K., Ginocchio, C.C. and Galan, J.E. 1994. Molecular and functional characterization of the Salmonella typhimurium invasion genes invB and invC: homology of InvC to the F0F1 ATPase family of proteins. *J. Bacteriol.* **176:** 4501-4510.

Emerson, S.U., Tokuyasu, K. and Simon, M.I. 1970. Bacterial flagella: polarity of elongation. *Science* **169:** 190-192.

Enninga, J., Mounier, J., Sansonetti, P. and Tran Van Nhieu, G. 2005. Secretion of type III effectors into host cells in real time. *Nat. Methods* **2:** 959-965.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. 1994a. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33:** 5984-6003.

Farrow, N.A., Zhang, O., Forman-Kay, J.D. and Kay, L.E. 1994b. A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR* **4:** 727-734.

Fivaz, M. and van der Goot, F.G. 1999. The tip of a molecular syringe. *Trends Microbiol.* **7:** 341-343.

Frithz-Lindsten, E., Rosqvist, R., Johansson, L. and Forsberg, A. 1995. The chaperone-like protein YerA of Yersinia pseudotuberculosis stabilizes YopE in the cytoplasm but is dispensible for targeting to the secretion loci. *Mol. Microbiol.* **16:** 635-647.

Galan, J.E. and Collmer, A. 1999. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284:** 1322-1328.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D. and Bairoch A. 2005. *Protein Identification and Analysis Tools on the ExPASy Server.* In *The Proteomics Protocols Handbook,* (eds. John M. Walker), 571-607. Humana Press,

Ghosh, P. 2004. Process of protein transport by the type III secretion system. *Microbiol. Mol. Biol. Rev.* **68:** 771-795.

Goddard, T.D. and Kneller, D.G. *Sparky 3,* University of California, San Francisco,

Gorbalenya, A.E. and Koonin, E.V. 1993. Helicases-- Amino-acid-sequence comparisons and structure-function-relationships. *Curr. Opin. Struct. Biol.* **3:** 419-429.

Hauck, P., Thilmony, R. and He, S.Y. 2003. A Pseudomonas syringae type III effector suppresses cell wall-based extracellular defense in susceptible Arabidopsis plants. *Proc. Natl. Acad. Sci. U.S.A.* **100:** 8577-8582.

Jayaraman, B. and Nicholson, L.K. 2007. Thermodynamic dissection of the Ezrin FERM/CERMAD interface. *Biochemistry* **46:** 12174-12189.

Jin, Q., Thilmony, R., Zwiesler-Vollick, J. and He, S.Y. 2003. Type III protein secretion in Pseudomonas syringae. *Microb. Infect.* **5:** 301-310.

Jin, Q., Hu, W., Brown, I., McGhee, G., Hart, P., Jones, A.L. and He, S.Y. 2001. Visualization of secreted Hrp and Avr proteins along the Hrp pilus during type III secretion in Erwinia amylovora and Pseudomonas syringae. *Mol. Microbiol.* **40:** 1129-1139.

Johnson, S., Deane, J.E. and Lea, S.M. 2005. The type III needle and the damage done. *Curr. Opin. Struct. Biol.* **15:** 700-707.

Kenniston, J.A., Baker, T.A., Fernandez, J.M. and Sauer, R.T. 2003. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* **114:** 511-520.

Koradi, R., Billeter, M. and Wüthrich, K. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14:** 51-55.

Lee, C., Schwartz, M.P., Prakash, S., Iwakura, M. and Matouschek, A. 2001. ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Mol. Cell* **7:** 627-637.

Lee, V.T. and Schneewind, O. 2002. Yop fusions to tightly folded protein domains and their effects on Yersinia enterocolitica type III secretion. *J. Bacteriol.* **184:** 3740-3745.

Leigh, J.S. 1971. Relaxation times in systems with chemical exchange: Some exact solutions. *J. Magn. Reson.* **4:** 308-311.

Levenberg, K. 1944. A method for the solution of certain problems in least squares. *Quart. Appl. Math.* **2:** 164-168.

Li, C.M., Brown, I., Mansfield, J., Stevens, C., Boureau, T., Romantschuk, M. and Taira, S. 2002. The Hrp pilus of Pseudomonas syringae elongates from its tip and acts as a conduit for translocation of the effector protein HrpZ. *EMBO J.* **21:** 1909-1915.

Lipari, G. and Szabo, A. 1982. Model-free analysis to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* **104:** 4559-4570.

Luo, Y., Bertero, M.G., Frey, E.A., Pfuetzner, R.A., Wenk, M.R., Creagh, L., Marcus, S.L., Lim, D., Sicheri, F., Kay, C., Haynes, C., Finlay, B.B. and Strynadka, N.C. 2001. Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. *Nat. Struct. Biol.* **8:** 1031-1036.

Lupas, A.N. and Martin, J. 2002. AAA proteins. *Curr. Opin. Struct. Biol.* **12:** 746-753.

Marquardt, D. 1963. An algorithm of least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* **11:** 431-441.

Matouschek, A. 2003. Protein unfolding--an important process in vivo? *Curr. Opin. Struct. Biol.* **13:** 98-109.

McConnell, H.M. 1958. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* **28:** 430-431.

Minamino, T., Imae, Y., Oosawa, F., Kobayashi, Y. and Oosawa, K. 2003. Effect of intracellular pH on rotational speed of bacterial flagellar motors. *J. Bacteriol.* **185:** 1190-1194.

Montelione, G.T. and Wagner, G. 1989. 2D chemical-exchange NMR-spectroscopy by proton-detected heteronuclear correlation. *J. Am. Chem. Soc.* **111:** 3096-3098.

Mulder, F.A.A., Spronk, C.A.E.A., Slijper, M., Kaptein, R. and Boelens, R. 1996. Improved HSQC experiments for the observation of exchange broadened signals. *J. Biomol. NMR* **8:** 223-228.

Nishiguchi, S., Goto, Y. and Takahashi, S. 2007. Solvation and desolvation dynamics in apomyoglobin folding monitored by time-resolved infrared spectroscopy. *J.Mol.Biol.* **373:** 491-502.

Palmer, A.G.I., Kroenke, C.D. and Loria, J.P. 2001.
Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **339:** 204-238.

Palmer, A.G.I. 2001. NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* **30:** 129-155.

Pawley, N.H., Gans, J.D. and Nicholson, L.K. 2002. Factors determining the reliable description of global tumbling parameters in solution NMR. *J. Biomol. NMR* **24:** 215-229.

Pawley, N.H., Wang, C., Koide, S. and Nicholson, L.K. 2001. An improved method for distinguishing between anisotropic tumbling and chemical exchange in analysis of 15N relaxation parameters. *J. Biomol. NMR* **20:** 149-165.

Pedley, K.F. and Martin, G.B. 2003. Molecular basis of Pto-mediated resistance to bacterial speck disease in tomato. *Annu. Rev. Phytopathol.* **41:** 215-243.

Piaggio, M.V., Peirotti, M.B. and Deiber, J.A. 2007. On the application of CZE to the study of protein denaturation. *Electrophoresis* **28:** 2223-2234.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. 1988. *Numerical recipes in C,* 521-522. Cambridge University Press, Cambridge, UK.

Preston, G., Huang, H.C., He, S.Y. and Collmer, A. 1995. The HrpZ proteins of Pseudomonas syringae pvs. syringae, glycinea, and tomato are encoded by an operon containing Yersinia ysc homologs and elicit the hypersensitive response in tomato but not soybean. *Mol. Plant-Microbe Interact.* **8:** 717-732.

Rahme, L.G., Mindrinos, M.N. and Panopoulos, N.J. 1992. Plant and environmental sensory signals control the expression of hrp genes in Pseudomonas syringae pv. phaseolicola. *J. Bacteriol.* **174:** 3499-3507.

Rosqvist, R., Magnusson, K.E. and Wolf-Watz, H. 1994. Target cell contact triggers expression and polarized transfer of Yersinia YopE cytotoxin into mammalian cells. *EMBO J.* **13:** 964-972.

Schlumberger, M.C., Müller, A.J., Ehrbar, K., Winnen, B., Duss, I., Stecher, B. and Hardt, W.D. 2005. Real-time imaging of type III secretion: Salmonella SipA injection into host cells. *Proc. Natl. Acad. Sci. U.S.A.* **102:** 12548-12553.

Schurr, J.M., Babcock, H.P. and Fujimoto, B.S. 1994. A test of the model-free formulas. Effects of anisotropic rotational diffusion and dimerization. *J. Magn. Reson. B* **105:** 211-224.

Scofield, S.R., Tobias, C.M., Rathjen, J.P., Chang, J.H., Lavelle, D.T., Michelmore, R.W. and Staskawicz, B.J. 1996. Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. *Science* **274:** 2063-2065.

Shan, L., Thara, V.K., Martin, G.B., Zhou, J.M. and Tang, X. 2000. The Pseudomonas AvrPto protein is differentially recognized by tomato and tobacco and is localized to the plant plasma membrane. *Plant Cell* **12:** 2323-2338.

Shea, J.E., Onuchic, J.N. and Brooks, C.L.,3rd. 2002. Probing the folding free energy landscape of the Src-SH3 protein domain. *Proc.Natl.Acad.Sci.U.S.A.* **99:** 16064-16068.

Sorg, J.A., Miller, N.C., Marketon, M.M. and Schneewind, O. 2005. Rejection of impassable substrates by Yersinia type III secretion machines. *J. Bacteriol.* **187:** 7090-7102.

Sory, M.P. and Cornelis, G.R. 1994. Translocation of a hybrid YopE-adenylate cyclase from Yersinia enterocolitica into HeLa cells. *Mol. Microbiol.* **14:** 583-594.

Stebbins, C.E. and Galan, J.E. 2001. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* **414:** 77-81.

Tamano, K., Aizawa, S., Katayama, E., Nonaka, T., Imajoh-Ohmi, S., Kuwae, A., Nagai, S. and Sasakawa, C. 2000. Supramolecular structure of the Shigella type III secretion machinery: the needle part is changeable in length and essential for delivery of effectors. *EMBO J.* **19:** 3876-3887.

Tang, X., Xie, M., Kim, Y.J., Zhou, J., Klessig, D.F. and Martin, G.B. 1999. Overexpression of Pto activates defense responses and confers broad resistance. *Plant Cell* **11:** 15-29.

van Dijk, K., Fouts, D.E., Rehm, A.H., Hill, A.R., Collmer, A. and Alfano, J.R. 1999. The Avr (effector) proteins HrmA (HopPsyA) and AvrPto are secreted in culture from Pseudomonas syringae pathovars via the Hrp (type III) protein secretion system in a temperature- and pH-sensitive manner. *J. Bacteriol.* **181:** 4790-4797.

Walker, P.A., Leong, L.E., Ng, P.W., Tan, S.H., Waller, S., Murphy, D. and Porter, A.G. 1994. Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Biotechnology (N.Y)* **12:** 601-605.

Weast, R.C. 1982. *CRC Handbook of Chemistry and Physics,* 63rd ed. CRC Press, Inc, Boca Roton, FL.

Wilharm, G., Lehmann, V., Krauss, K., Lehnert, B., Richter, S., Ruckdeschel, K., Heesemann, J. and Trulzsch, K. 2004. Yersinia enterocolitica type III secretion depends on the proton motive force but not on the flagellar motor components MotA and MotB. *Infect. Immun.* **72:** 4004-4009.

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

Woestyn, S., Allaoui, A., Wattiau, P. and Cornelis, G.R. 1994. YscN, the putative energizer of the Yersinia Yop secretion machinery. *J. Bacteriol.* **176:** 1561-1569.

Wulf, J., Pascuzzi, P.E., Fahmy, A., Martin, G.B. and Nicholson, L.K. 2004. The solution structure of type III effector protein AvrPto reveals conformational and dynamic features important for plant pathogenesis. *Structure* **12:** 1257-1268.

Wulff-Strobel, C.R., Williams, A.W. and Straley, S.C. 2002. LcrQ and SycH function together at the Ysc type III secretion system in Yersinia pestis to impose a hierarchy of secretion. *Mol.Microbiol.* **43:** 411-423.

Xiao, Y., Lu, Y., Heu, S. and Hutcheson, S.W. 1992. Organization and environmental regulation of the Pseudomonas syringae pv. syringae 61 hrp cluster. *J. Bacteriol.* **174:** 1734-1741.

Yonekura, K., Maki-Yonekura, S. and Namba, K. 2003. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424:** 643-650.

Zhang, Z.Y., Clemens, J.C., Schubert, H.L., Stuckey, J.A., Fischer, M.W., Hume, D.M., Saper, M.A. and Dixon, J.E. 1992. Expression, purification, and physicochemical characterization of a recombinant Yersinia protein tyrosine phosphatase. *J. Biol. Chem.* **267:** 23759-23766.

Chapter Three

# Elucidation of a pH-dependent folding switch of the *Pseudomonas syringae* effector protein, AvrPto[1]

## *Introduction*

Although the delivery of bacterial proteins into a host cell is a primary step in pathogen infection by gram-negative bacteria, little is presently known regarding the detailed mechanisms of its delivery process. Gram-negative bacteria, such as *Pseudomonas syringae,* use a morphologically conserved apparatus (Francis et al. 1994, Marlovits et al. 2004) called the Type III Secretion System (TTSS) to inject a suite of bacterial effector proteins, which act as agents of infection, into the host cell's cytoplasm (Galan et al. 1999). Given the critical role of the TTSS in disease, an understanding of the properties of effector proteins that allow efficient transport through the TTSS is mandated.

Secretion through the TTSS is a complicated process dependent on several factors, including the stability and folding kinetics of the effector proteins. The invading effector proteins travel along a hollow needle-like structure of the TTSS, presumably through the inside of the needle (Johnson et al. 2005). Since the inner diameter of the needle is ~2-3 nm (Blocker et al. 2001, Cordes et al. 2003), most proteins would have to at least partially unfold to pass through unimpeded. The requirement for unfolding suggests that there may be stability thresholds above which an effector protein is too stable to be effectively secreted. Effector protein YopE fused to a stable ubiquitin domain cannot be secreted, while a similar YopE fusion protein with a destabilized ubiquitin is secretion-competent (Lee et al. 2002).

[1] Chapter Three is based on a manuscript to be submitted to the *Proceedings of the National Academy of Science of the United States of America.*

The TTSS requires a proton gradient (Wilharm et al. 2004). The *Yersinia enterocolitica* TTSS and the evolutionally-related flagellar apparatus (Emerson et al. 1970, Yonekura et al. 2003) of *E. coli* both require that a proton gradient is maintained across the inner bacterial membrane (Minamino et al. 2003, Wilharm et al. 2004). In light of these requirements of the TTSS (i.e. the proton gradient and the ability of effectors to be unfolded), what properties of an effector protein allow it to be secreted efficiently?

The *P. syringae* pv tomato effector protein, AvrPto, provides an excellent model system with which a better understanding of the requirements for effective secretion can be gained. AvrPto has been well studied in its virulence (disease-causing) and avirulence (disease-preventing) roles in the host cell (Pedley et al. 2003). At least thirty effector proteins including AvrPto are injected into host leaf cells upon infection by *P. syringae* pv tomato (Guttman et al. 2002, Petnicki-Ocwieja et al. 2002), acting to counter cellular defenses, aid bacterial proliferation, and, ultimately, cause bacterial speck disease in tomato and Arabidopsis (Chang et al. 2000, Hauck et al. 2003, Jin et al. 2003, Shan et al. 2000, Whalen et al. 1991). However, some strains of tomato are resistant to this infection, possessing a kinase, Pto, which recognizes the presence of AvrPto in the plant cell cytoplasm and triggers cellular defenses (Scofield et al. 1996, Tang et al. 1999). Both the virulence and avirulence roles require AvrPto to be folded in order to function.

The pH-dependence of AvrPto stability potentially provides a pool of secretion-ready protein for the TTSS that can subsequently refold inside the host cell. The *P. syringae* TTSS and its associated proteins are expressed when the bacteria are in the acidic extracellular space (apoplast) of the host leaves (Grignon et al. 1991, Jin et al. 2003) or when cultured in a minimal *hrp*-inducing medium (Rahme et al. 1992, Xiao et al. 1992, van Dijk et al. 1999). While AvrPto is expressed if the pH of the

external *hrp* medium ($pH_{ext}$) is either pH 6 or 7, it is only secreted efficiently by the TTSS when $pH_{ext}$ is 6 (van Dijk et al. 1999). In analogy to the proton gradient required for the operation of both the *Yersinia enterocolitica* TTSS and the *E. coli* flagellum (Minamino et al. 2003, Wilharm et al. 2004), the pH inside a *P. syringae* bacterium ($pH_{int}$) should be greater than that of the external medium ($pH_{int}>pH_{ext}$). Intriguingly, AvrPto is more stable at pH 7 than at pH 6 (Wulf et al. 2004). If the external medium were at $pH_{ext}$ 7, then, in order to maintain the proton gradient, $pH_{int}$ would be greater than pH 7, a condition that promotes greater stability of AvrPto and should inhibit its secretion. If the pathogen is in a more acidic environment, such as the plant apoplast where the pH is in the range of 5 to 6.5 (Grignon et al. 1991), then $pH_{int}$ would be slightly greater than $pH_{ext}$ to maintain the proton gradient. AvrPto is less stable under acidic conditions, which increases the amount of secretable protein available and, hence, should increase its secretion efficiency. Once AvrPto has passed through the TTSS, the neutral pH of the host cell cytoplasm enhances the stability of the effector, providing the population of folded AvrPto required for its functions within the host cell. This pH-regulated folding provides a potential mechanism for infection, in which AvrPto is efficiently secreted under the acidic conditions that the pathogen experiences inside the plant tissue, yet can refold into its functional folded conformation once inside the neutral plant cytoplasm.

The native conformation of truncated AvrPto (TrAvrPto) has been characterized *via* NMR spectroscopy. The disordered N- and C-terminal tails of AvrPto were removed, leaving TrAvrPto, a three-helix bundle with an Ω loop and a short orthogonal helix (Wulf et al. 2004). Like AvrPto, TrAvrPto is more stable at pH 7 than at pH 6 and discrete populations for the native and non-native conformations are resolved in NMR spectra.

In this study, we applied CD, Trp fluorescence, and NMR spectroscopy to elucidate the mechanism of the pH-regulated stability in TrAvrPto. Each technique probed different aspects of the protein and their use in combination revealed a previously unelucidated intermediate conformation of TrAvrPto. The stability of TrAvrPto changes during two pH titration events with midpoints of stability at pH 3.10±0.07 and at 5.9±0.1. The latter titration occurs in the pH range relevant to effector protein secretion. At this higher pH range, changes in pH titration alter local chemical environments of native TrAvrPto in regions near the seven His sidechains, but leave the core intact, as observed in NMR HSQC spectra. Utilizing the fine resolution of NMR correlation experiments, $pK_a$ values for each His sidechain could be quantified. In native TrAvrPto, most His display $pK_a$ values consistent with their solvent-exposed environment on the protein surface. However, both H54, which is near the partially-positive end of an α-helix, and buried H87 have anomalously depressed $pK_a$ values: $pK_a$(H54)=5.90±0.03 and $pK_a$(H87)=5.2±0.1. These sidechains favor their deprotonated neutral states within native protein and, based on thermodynamic simulations, drive the acid-denaturation of the TrAvrPto. Intriguingly, H87, whose ionization causes by far the greatest change in TrAvrPto stability, is conserved in homologous sequences for both AvrPto and AvrPtoB, further supporting its functional importance. Taken altogether, the stability data, $pK_a$ values, thermodynamic simulations, and the sequence alignment imply a common mechanism in which the stability of not just AvrPto, but a larger subset of effector proteins is predominantly regulated by the protonation state of H87.

***Results***

*Use of multiple techniques in combination reveals an intermediate conformation of TrAvrPto*

Individual spectroscopic techniques yield insights into different aspects of protein folding and the correct interpretation of this information is contingent upon knowledge of what each techniques actually observes. CD and Trp fluorescence spectroscopy are sensitive to the stability of the protein as a whole. CD data are dependent on the proportions of different elements of secondary structure, especially of α-helices (Martin et al. 2008). The fluorescence emission intensity and wavelength for Trp depends on whether it is solvent-exposed or is in the hydrophobic core (Pace 1986). In contrast, NMR spectroscopy resolves individual groups within the protein as separate peaks. In addition to being sensitive to local chemical environment, this technique can depend on the rate of exchange between conformations relative to the timescale of the experiment. The folding of TrAvrPto is on the slow chemical exchange timescale and, so, there are two separate sets of peaks, whose *peak volumes* are proportional to the relative populations of native and non-native protein, respectively. Conformations in fast chemical exchange contribute to the same set of peaks, whose *positions* depend on the relative populations of each conformation. Unfolded ensembles contain many such conformations in fast chemical exchange (Dyson et al. 2002, Mittag et al. 2007). Given the specific sensitivities of each experimental technique, it is plausible that CD, Trp fluorescence, and NMR data could each report on different sub-sets of TrAvrPto conformations. An intermediate conformation with a hydrophobic cluster and intact α-helices, for example, would contribute along with the native conformation to the total "folded" population observed by CD and Trp fluorescence spectroscopy. However, if this intermediate

conformation exchanges rapidly with the unfolded ensemble, it would contribute to the intensity of the unfolded ensemble's peaks.

Indeed, different pH-dependent stability curves are observed for TrAvrPto from CD and Trp fluorescence spectroscopy and from NMR spectroscopy (Figure 3.1A). CD and Trp fluorescence data show simple pH-dependent stability curves. Based on the shape of the mean residue extinction coefficient wavelength profiles, at each pH point CD observes a mixture of α-helices, extended loops, and random coils (Appendix Figure 3.1). The NMR data, in contrast, show more complex acid denaturation curves. The apparent "folded" population observed by NMR is less than that seen by the other two techniques, suggesting the presence of an intermediate conformation that contributes to the CD and fluorescence data in a manner similar to the native conformation. Importantly, this curve shape and smaller population are conserved for all thirteen groups probed by NMR: the backbone amide groups of A47, G48, A61, S64, T76, T91, G92, S94, G95, G99, A112, and G128 and the Trp sidechain indole group, W116 ε1 (Figure 3.1B). The consistency in titration behavior, despite their differences in local chemical environments, is indicative of global changes in conformation. Within the native conformation, some of these groups are involved in the hydrogen-bonds of the α-helices (ex. T76 and A112), others are solvent-exposed (ex. G95), and some are buried (ex. W116 ε1). Based on the specific sensitivities of CD, Trp fluorescence, and NMR spectroscopy and the stability data for TrAvrPto, it is reasonable to propose the existence of an intermediate conformation that, while possessing a buried Trp core and α-helices, is in slow chemical exchange with the native conformation.

The possibility that the TrAvrPto stability is affected by aggregation or some other high concentration effect can be excluded by noting the concentration-independence of wt TrAvrPto $R_1$ data. The concentration of the NMR samples was

much greater than those of the CD and Trp fluorescence samples: 0.5-1.2 mM for NMR and 10-15 μM for CD and fluorescence samples. This raises the question of whether high concentration (e.g. aggregation) affects the measured stability. NMR $R_1$ relaxation data are very sensitive to the mass of the protein. Since aggregation is concentration-dependent, the $R_1$ rate constant for native TrAvrPto should be affected by aggregation, if present. As discussed in Chapter Two, $R_1$ values are concentration-independent at both pH 6 and 7. Most importantly, the observed pH-dependence of these $R_1$ data can be explained by the two-state exchange between the monomeric intermediate and native conformations. NMR samples bearing signs of degradation (*Materials and Methods*) were excluded from the stability analysis, eliminating another potential source of artifacts. Although these arguments yield high confidence the intermediate is present at all concentrations, due to the large concentration difference between the optical and NMR measurements, further efforts to confirm by NMR the presence of the intermediate at micromolar concentrations are currently in progress.

*TrAvrPto stability altered by two titration events*

The simplest model that explains all of the pH-dependent stability data is a two titration model between the unfolded ensemble (U), the intermediate conformation (I), and the native protein (N).

$$UH_2 \overset{K_{eq1}}{\rightleftharpoons} IH_2 \overset{K_{stab1}}{\rightleftharpoons} IH + H^+ \overset{K_{eq2}}{\rightleftharpoons} NH + H^+ \overset{K_{stab2}}{\rightleftharpoons} N + 2H^+ \qquad \text{Eq. 1}$$

The first titration is between $UH_2$ and IH, where the conformational change (described by $K_{eq1}=p_{IH2}/p_{UH2}$) is explicitly separated from the deprotonation ($K_{stab1}=p_{IH}*[H^+]/p_{IH2}$). The midpoint of stability of the first titration occurs at $pH=pK_{stab1}$. The second titration is between IH and N, in which $K_{eq2}=p_{NH}/p_{IH}$ and $K_{stab2}=p_N*[H^+]/p_{NH}$. For this model, the five fractional populations are

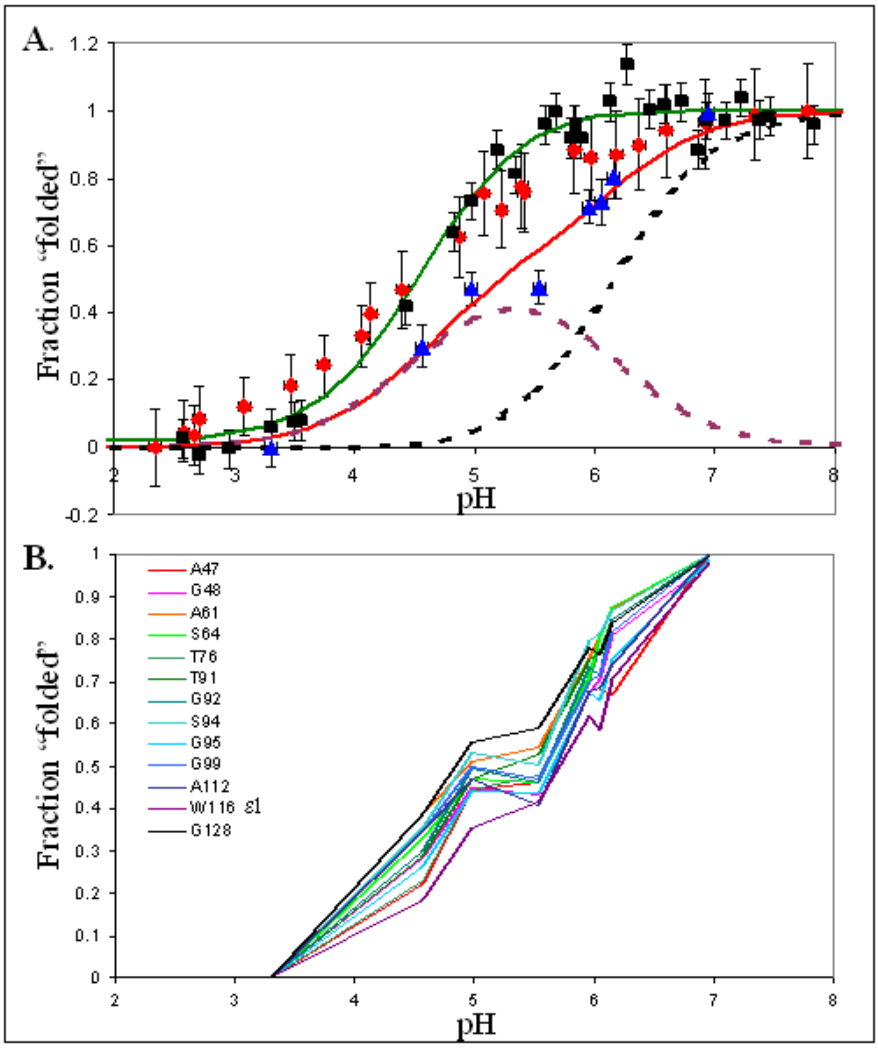



Figure 3.1: **pH-dependent stability of TrAvrPto and fitting to two titration model.** A. The acid denaturation of TrAvrPto was characterized. Stabilities measured with CD (red squares) and Trp fluorescence[2] (black squares) are dependent on the fraction populations of intermediate and native TrAvrPto: $p_{IH2}+p_{IH}+p_{NH}+p_N$. The stabilities measured *via* NMR (blue triangles, the global protein average values shown here) depend on the native populations: $p_{NH}+p_N$. These data can be modeled by a five state, two titration model, fitting the CD and fluorescence data to ($p_{IH2}+p_{IH}+p_{NH}+p_N$) (green line) and the NMR data to ($p_{NH}+p_N$) (red line). The fractional populations $p_{NH}$ (dashed purple line) and $p_N$ (dashed black line) calculated from the model fitting are also shown. B. The shape of NMR stability curve is consistent over all probed residues. The pH-dependent stability observed *via* NMR has complex behavior: there is more than one apparent "inflection point" and the apparent "folded population is less than that observed by CD or Trp fluorescence. These behaviors are conserved in NMR data for 13 residues scattered across TrAvrPto.

[2] Trp fluorescence data taken with the assistance of Aaron Oswald (Ithaca Senior High School).

$$p_{UH2} = 1/D, \quad \text{Eq. 2}$$

$$p_{IH2} = K_{eq1}/D, \quad \text{Eq. 3}$$

$$p_{IH} = (K_{eq1} * 10^{pH-pKstab1})/D, \quad \text{Eq. 4}$$

$$p_{NH} = (K_{eq1} * K_{eq2} * 10^{pH-pKstab1})/D, \quad \text{Eq. 5}$$

$$p_{N} = (K_{eq1} * K_{eq2} * 10^{pH-pKstab1} * 10^{pH-pKstab2})/D, \quad \text{Eq. 6}$$

where $D = 1+K_{eq1}[1+10^{pH-pKstab1}\{1+K_{eq2}(1+10^{pH-pKstab2})\}]$. Eq. 7

Different subsets of conformations are sampled by CD, Trp fluorescence, and NMR. Since the deprotonation of ionizable groups is typically on the fast exchange timescale (Sudmeier et al. 1980), it is reasonable to assume that the "folded" population observed by NMR contains contributions from both the protonated and deprotonated native states, $p_{NH}$ and $p_N$ (Figure 3.1A). As pH is increased, $p_{NH}$ builds up from zero, reaches a maximum, and then decreases back to zero, while $p_N$ increases smoothly from zero to one. The superposition of these two curves ($p_{NH}+p_N$) model the multiple inflection points observed in NMR data. The apparent population of "folded" protein obtained from CD and Trp fluorescence data contain contributions from all states except the unfolded ensemble and, so, is modeled by $(1-p_{UH2}) = (p_{IH2}+p_{IH}+p_{NH}+p_N)$.

The five state model in which two pH titration events alter the stability of TrAvrPto explains data gathered from three different techniques (Figure 3.1A, Table 3.1). During the first titration, the unfolded ensemble is better able to accommodate the doubly-protonated state than the intermediate conformation ($K_{eq1}$=0.016±0.002). The $pK_{stab1}$=3.10±0.07 suggests that the increase in stability during the first, low pH titration is driven by the deprotonation of one or more of the 14 carboxylate groups of TrAvrPto, given that the typical dissociation constant for these groups is $pK_a$=3.1 to 4.35 (Nozaki et al. 1962). The IH state is approximately as stable as the FH ($K_{eq2}$=1.2±0.4) and the midpoint of the second titration is at $pK_{stab2}$=5.9±0.1. Given

that His sidechains typically titrate with a $pK_a$~6.4 (Nozaki et al. 1962), the position of the $pK_{stab2}$ raises the possibility that the pH-dependence of TrAvrPto stability is controlled by the ionization state of one or more of its seven His residues. Since $pH_{ext}$=5-6.5 within the plant apoplast, this second, higher pH titration is of more potential relevance to the pH-dependence of TTSS secretion and will be the focus of the rest of this Chapter.

**Table 3.1: Results of two titration model fitting**

| Group | $K_{eq1}$ | $pK_{stab1}$ | $K_{eq2}$ | $pK_{stab2}$ |
|---|---|---|---|---|
| A47 | 0.017 | 3.07 | 0.94 | 5.9 |
| G48 | 0.016 | 3.05 | 0.96 | 5.8 |
| A61 | 0.016 | 3.16 | 1.6 | 5.9 |
| S64 | 0.017 | 3.10 | 1.1 | 5.8 |
| T76 | 0.019 | 3.13 | 0.95 | 5.8 |
| T91 | 0.019 | 3.19 | 1.2 | 5.8 |
| G92 | 0.017 | 3.11 | 1.2 | 5.9 |
| S94 | 0.016 | 3.18 | 1.6 | 5.9 |
| G95 | 0.016 | 3.05 | 0.93 | 5.9 |
| G99 | 0.014 | 3.05 | 1.3 | 6.0 |
| A112 | 0.013 | 2.96 | 1.1 | 6.0 |
| W116 e1 | 0.018 | 3.03 | 0.63 | 5.8 |
| G128 | 0.015 | 3.22 | 2.1 | 6.1 |
| **Mean±standard deviation** | **0.016±0.002** | **3.10±0.07** | **1.2±0.4** | **5.9±0.1** |

All of the experimental data from CD, Trp fluorescence and NMR spectroscopy were fitted to formulae in which the Hill coefficients of each titration were implicitly constrained to a value of unity (Eq. 2-7, A1). For pH titrations, a Hill coefficient is a measure of how many protons are taken up during denaturation, as well as an estimate of the cooperativity of proton binding. Therefore, the interpretation of fitted Hill coefficient values can be ambiguous. During modeling of the experimental data, any fitting with a chi square probability value of $Q<0.001$ (Press et al. 1988) was rejected. For the TrAvrPto data above, models that allowed the Hill coefficients to

vary could not be rejected (Q>>0.001, data not shown). Nor could models that implicitly constrained the Hill coefficients to unity be rejected. The Hill coefficients are factors of the exponents of 10 and, so, are very insensitive to minor variations in the data. Since the latter models (Eq. 2-7, A1) were the simplest models that explained all of the data, these were the models presented above in this Chapter.

*pH titration affects regions of native TrAvrPto near the His residues*

The separation of native and non-native TrAvrPto backbone amide peaks allows the pH-induced changes to the native conformation to be monitored independently. The deprotonation of individual titratible groups occur on the fast chemical exchange timescale (Sudmeier et al. 1980) and may be reactions distinct from the folding process. Therefore, a native TrAvrPto peak will shift position based on the relative populations of protonated and deprotonated species of the titrating groups. Even groups that do not themselves titrate, such as the backbone amide groups observed in these HSQC spectra, may exhibit these peak shifts if they are in close proximity to a titrating group.

Since the position of a peak depends on its surroundings, changes in the chemical shift, $\Delta\omega = [\Delta\omega_{1H}(\text{ppm})^2 + 0.154^2 * \Delta\omega_{15N}(\text{ppm})^2]^{1/2}$ (Ayed et al. 2001, Seavey et al. 1991), illuminate which regions of native TrAvrPto are affected by pH titration (Figure 3.2). The regions neighboring the seven His residues display the greatest change in local chemical environment due to the pH titration between pH 7.25 and pH 5.54, most likely reflecting changes in local chemical environment due to the change in His protonation state.

The overall structure of the native TrAvrPto conformation is not disrupted by changes in pH. The sidechains of Y69, Y73, and W116 form an aromatic network

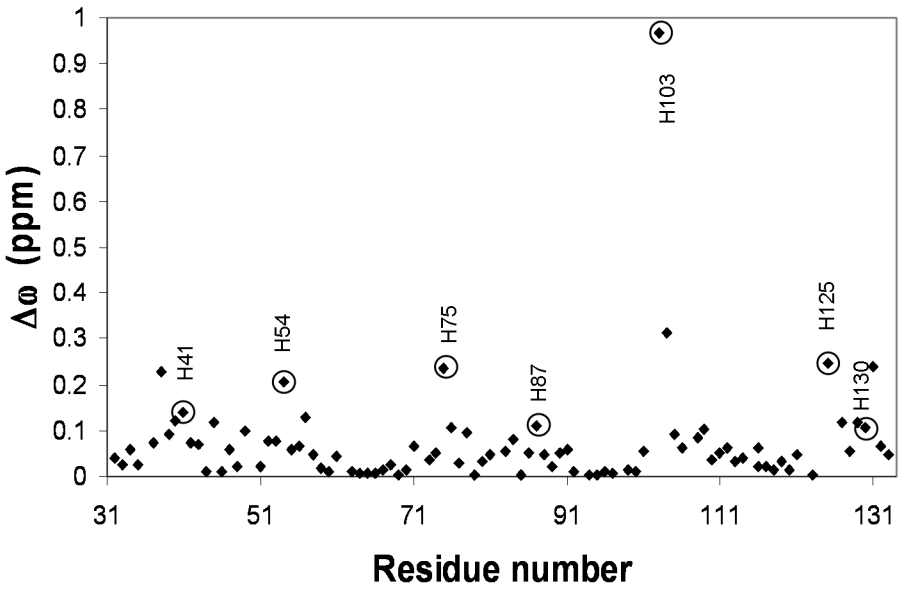



Figure 3.2: **Perturbation of TrAvrPto backbone NH during pH titration.** The greatest changes in chemical shift between pH 5.5 and 7.3 for the native conformation are for residues in near proximity to the seven His sidechains (marked with circles). Therefore, in the pH range expected during effector protein secretion, changes in chemical environment of native TrAvrPto are centered around the His sidechains.

(Burley et al. 1985) within the core of native TrAvrPto with G48 amide packed against the face of the W116 aromatic ring (Figure 3.3A). The G48 amide peak was shifted to lower frequency at all pH values (6.2-6.3 ppm, Appendix Table 3.1), compared with the ~8.3 ppm chemical shifts typical of Gly in random coils (Wishart et al. 2001). This upfield shift is consistent with G48's proximity to the aromatic ring current (Wishart et al. 1991). The peaks of these four NH groups were immobile or moved only slightly (Figure 3.2, Appendix Table 3.1), signifying that the packing structure has changed little. The large downfield (to higher frequency) shifts of the N-terminal residues of all four helices raise the possibility that the native conformation retains at least some helical structure during pH titration. Even at pH 4.6, the lowest pH at which the native population was observed, the amide proton chemical shifts for the first residues of the helices are all shifted downfield (Appendix Table 3.1): 9.25 (S33 of $\alpha_A$), 8.97 ppm (D52 of $\alpha_B$), 8.95 ppm (Q63 of $\alpha_C$), and 9.65 ppm (H103 of $\alpha_D$). While backbone amide proton chemical shifts are not reliable indicators of secondary structure (Wishart et al. 2001), amide protons located at the N-terminal position of helices tend to have anomalous downfield chemical shifts, probably due to the backbone amide groups' alignment with the partially-positive end of the helix dipole (Wishart et al. 1991). Overall, it can be concluded that, during pH titration, the folded population of TrAvrPto retains a hydrophobic core and its helices, based on the NMR and CD data.

*H54 and H87 titrate with depressed $pK_a$ in native TrAvrPto*

The resolution of NMR peaks in HMQC spectra allows the titration of individual His sidechains to be characterized. These HMQC spectra were optimized to probe the local environments of the two $^{15}N$ nuclei in His sidechains, $^{15}N\varepsilon 2$ and $^{15}N\delta 1$ (Bax et al. 1990, Pelton et al. 1993) (*Materials and Methods*), most especially

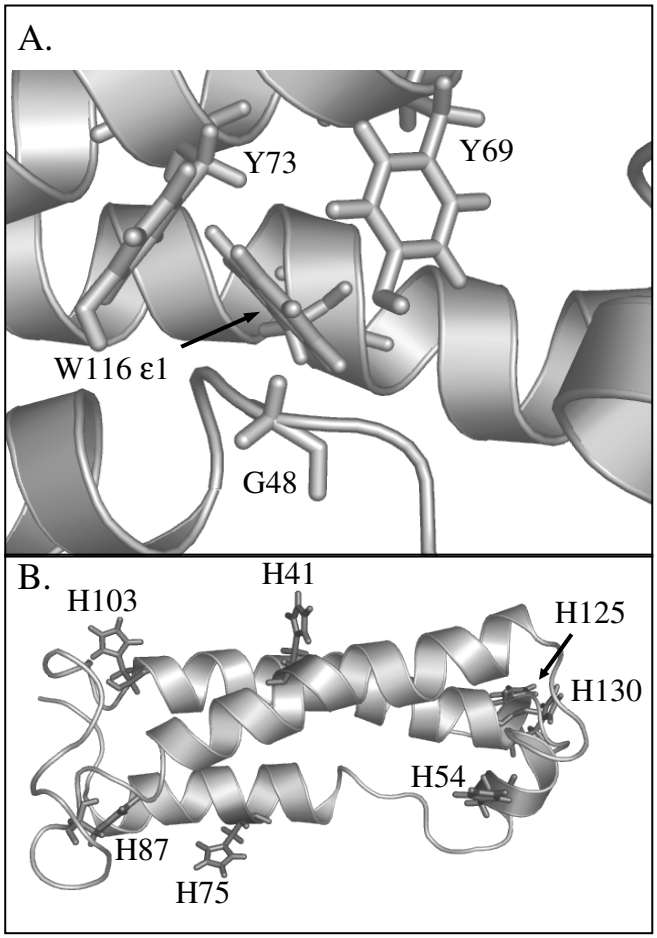



Figure 3.3: **An aromatic network and His of TrAvrPto.** A. Aromatic network within the core of TrAvrPto. B. Distribution of His across TrAvrPto

their charge and protonation state. The pH titration of a His sidechain at physiological pH is between the positively-charged state ($His^+$) where both $^{15}N$ are protonated, and the two neutrally-charged tautomers, one of which is only protonated at $^{15}N\varepsilon2$ ($N\varepsilon2$-H $His^0$) and the other only at $^{15}N\delta1$ ($N\delta1$-H $His^0$). Therefore, even in the absence of protein folding, the HMQC data is affected by two separate reactions: the pH titration between $His^+$ and $His^0$ and the exchange between $N\varepsilon2$-H and $N\delta1$-H $His^0$.

Both pH titration and tautomer exchange were observed for His in native TrAvrPto. All seven His titrate in the pH range 4.9-7.3 in a manner consistent with fast exchange and are predominantly in the more energetically-favored $N\varepsilon2$-H tautomer (Blomberg et al. 1977, Pelton et al. 1993) at pH 7.0 (Figure 3.4A). In the absence of tautomer exchange, as one $^{15}N$ is protonated, its peak should shift to higher $^{15}N$ frequency (downfield), while the other moves upfield. This is precisely what is observed for H130 (Figure 3.4B). Fast and intermediate tautomer exchange, though, is possible and has been observed in the literature (Betz et al. 2004, Geierstanger et al. 1998, Grey et al. 2006, Pelton et al. 1993) and may cause the protonated $^{15}N$ peak to move downfield with increasing pH. The remaining six His of TrAvrPto display this behavior.

If the tautomer exchange is in the intermediate exchange regime, then an increase in $^{15}N$ linewidth with increasing pH is expected for all peaks, since the linebroadening effects of intermediate chemical exchange become more predominant as the total $His^0$ population increases (Blomberg et al. 1977, Geierstanger et al. 1998, Plesniak et al. 1996, Wolff et al. 2002). The linewidths of H87 and H103 broaden in the $^{15}N$ dimension as pH is increased, consistent with intermediate exchange of neutral tautomers (Appendix Figure 3.2). The sidechains H41, H54, H75, and H125 display little to no linewidth broadening, indicating that the tautomerization rates of these sidechains are faster than those of H87 and H103.

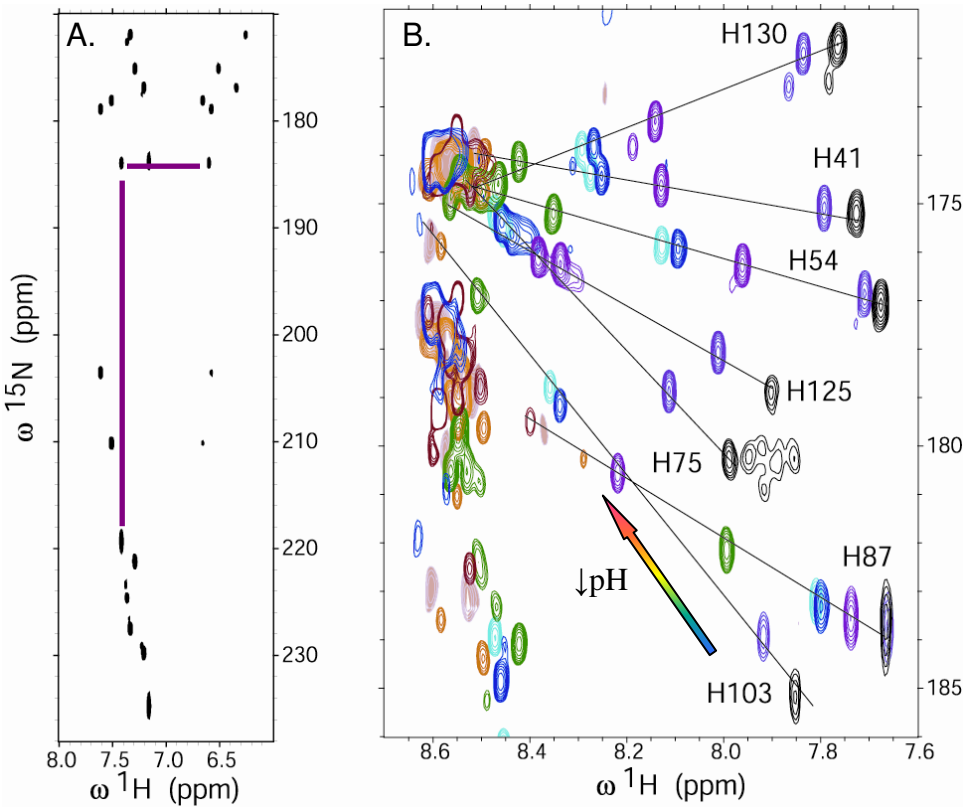



Figure 3.4: **pH titration of His sidechains in native TrAvrPto.** A. At pH 7.0, all seven His exhibit patterns of peaks consistent with Nε2-H $His^0$ tautomers. The peaks for H103 are marked in purple. B. An overlay of all spectra, focusing on the pH titration of the $^{15}$Nε2-$^{1}$Hε1 peaks.

The $pK_a$ values for each His sidechain in native TrAvrPto can be quantified based on the shift in peak positions during pH titration (*Materials and Methods*). While peaks for Nε2 and Nδ1 of each sidechain are resolved, the presence of tautomer exchange means that the fitted $pK_a$ values are the average values for the entire sidechain. Solvent-exposed His are expected to have $pK_a$~6.4. This is the case in native TrAvrPto, except for H54 and H87 (Table 3.2). The depressed $pK_a$ values for these two sidechains shed light on local environments within the protein and, as described in Chapter Four, play critical roles in the pH-dependent stability of TrAvrPto.

Table 3.2: **$pK_a$ values of His in native TrAvrPto**

| Residue | $pK_a$ |
|---|---|
| H41 | 6.29±0.05 |
| H54 | 5.90±0.03 |
| H75 | 6.61±0.08 |
| H87 | 5.2±0.1 |
| H103 | 6.26±0.04 |
| H125 | 6.48±0.02 |
| H130 | 6.25±0.04 |

## *Discussion*

*Experimental data illuminate the native and intermediate conformations of TrAvrPto*

Native TrAvrPto exists in equilibrium with the unfolded ensemble and an intermediate conformation. Even though the relative population of native TrAvrPto decreases with decreasing pH, CD, Trp fluorescence, and NMR data reveal that it retains its compact core and α-helices. The His sidechains, which drive the second titration, are scattered across the protein (Figure 3.3B). H87 is buried in the protein core and, like other His observed in hydrophobic environments (Betz et al. 2004, Geierstanger et al. 1998, Grey et al. 2006, Pelton et al. 1993), has a depressed $pK_a$. The sidechain is a hydrogen-bond donor to the S33 carbonyl. The formation of this

tertiary contact, as well as the sidechain's burial, is most likely the cause of H87's intermediate tautomer exchange. H54 is near the partially-positive N-terminal end of an α-helix and its own reduced $pK_a$ is consistent with previous literature observations (Armstrong et al. 1993). Therefore, since H103 forms a tertiary contact as the hydrogen-bond acceptor to N30 and is the first residue of an α-helix, it would be expected that H103 display both intermediate tautomer exchange and a depressed $pK_a$. While the former is observed, H103's $pK_a$ is more consistent with a solvent-exposed His. The presence of the neighboring, negatively-charged E104 presumably compensates for the energetic cost of placing $H103^+$ at the partially-positive end of a helix dipole. H130 sidechain does not display intermediate or fast tautomerization, perhaps due to a structural constraint, such as a strong hydrogen bond, that slows tautomerization down to the slow exchange timescale and/or restricts $His^0$ to only the Nε2-H tautomer. Indeed, the high pH limit chemical shift $\omega_{His0}$ for Nδ1 is 238 ppm, which, while approximately 10 ppm less than that expected for an unprotonated sidechain $^{15}N$ (Pelton et al. 1993), is consistent with that of His $^{15}N\delta1$ acting as a hydrogen-bond acceptor (Bachovchin 1986). The remaining sidechains, H41, H75, and H125 are solvent-exposed and display typical $pK_a$ values and faster tautomer exchange.

The intermediate conformation of TrAvrPto appears similar to a molten globule. Molten globules are characterized by the presence of secondary structure and the lack of *specific* long-range tertiary contacts (Alexandrescu et al. 1993, Arai et al. 2000). These conformations form hydrophobic cores that are more loosely packed than those of the natively-folded protein. Based on Trp fluorescence data, the TrAvrPto intermediate has a hydrophobic cluster that shields the Trp sidechain from solvent, while CD data indicate that it retains intact α-helices. It must be noted that the hydrophobic clusters within a molten globule, or even within an unfolded

ensemble, are not necessarily those formed within the native structure (Alexandrescu et al. 1993, Evanics et al. 2006, Marsh et al. 2007). For example, W36 sidechain of drkN SH3 is on average more protected from solvent within the unfolded ensemble than within the folded conformation (Evanics et al. 2006). In the case of TrAvrPto, the non-native peak for G48 is not shifted upfield like the G48 peak for the native conformation, implying that this backbone amide group is no longer packed against the W116 sidechain in either the intermediate or the unfolded ensemble.

Three helix bundles like TrAvrPto have been observed to fold at rates up to $10^5$-$10^6$ $s^{-1}$ (Kubelka et al. 2004), so it is not unreasonable that the intermediate conformation and the unfolded ensemble are in fast exchange. During the second titration in the pH-dependent stability curve, the predominant species are intermediate and native TrAvrPto, which are in slow chemical exchange with one another. The $k_{ex}$=2.1 $s^{-1}$ folding exchange rate quantified at pH 6.1 in Chapter Two (Dawson et al. 2008) is between these two conformations. The $R_1$ relaxation rates for the intermediate form are greater than those for the native conformation (referred to as $R_{1U}$ and $R_{1F}$, respectively, in Table 2.1). This difference is consistent with an intermediate that is smaller than the native protein. However, an alternate and more likely interpretation of the difference in $R_1$, in light of all other data, is that the more loosely packed intermediate is more dynamic than the native TrAvrPto.. Additionally, the variation of $R_1$ indicates that the intermediate TrAvrPto is probably anisotropic in shape.

The dynamic behavior of TrAvrPto is complex. Three separate chemical exchange reactions have been observed: tautomer exchange, pH titration, and folding. The questions that remain to be answered are whether the pH titration and folding processes depend on one another and, if so, which titratable groups play the most critical roles in the stability of TrAvrPto?

*The depressed $pK_a$ of H54 and H87 drive the acid-denaturation of TrAvrPto*

In general, the pH-dependence of stability is subject to the relative ability of the different conformations to accommodate each ionization state of a titratable group (Yang et al. 1993). For TrAvrPto during the second titration, these relative abilities are reflected by the $pK_a$ of the group in native protein, $pK_{a,N}$, and the $pK_a$ when the protein is in intermediate form, $pK_{a,I}$. For any titratable group, when $pK_{a,N}<pK_{a,I}$, decreasing the pH will promote destabilization since the protonated group is more energetically favored in the intermediate than in the native conformation. If it is assumed that on average His sidechains in the intermediate form have $pK_a$ typical of solvent-exposed His, $<pK_{a,I}>=6.4$ (Nozaki et al. 1962), then the His with the lowest $pK_{a,N}$, H87, titrates at pH values 1.2 pH units lower than it would in intermediate form. Using the thermodynamic theory of Yang and Honig (Yang et al. 1993), the known $pK_{a,N}$ for the base titration of each His, and $<pK_{a,I}>$, the acid denaturation can be simulated. The midpoint of the simulated curve, $\Delta\Delta G_{IF}(pH)$, occurs at pH 5.92±0.08, which is strikingly comparable to the NMR-derived $<pK_{stab2}>=5.9\pm0.1$. H87 and H54 have the greatest influence on $\Delta\Delta G_{IF}(pH)$, with over half of the energy contributed by buried H87 (Figure 3.5).

The acid denaturation at higher pH can be simulated remarkably well by assuming an average His $pK_a$ for the sidechains in the intermediate molten globule, $<pK_{a,I}=6.4>$, which matches that expected for solvent-exposed His (Nozaki et al. 1962). In native TrAvrPto, H87 is buried in the Ω loop, not in the three-helix bundle (Figure 3.3B). Within this hydrophobic environment in the native protein, the H87 $pK_{a,N}$ is depressed, while in the intermediate, the Ω loop is most likely disordered, exposing the sidechain to solvent. H54 is near the δ+ end of $\alpha_B$, resulting in a depressed $pK_{a,N}$. Collapse of this very short helix into a loop would raise H54's $pK_a$ to a value consistent with solvent-exposure. The remaining five His are on the protein

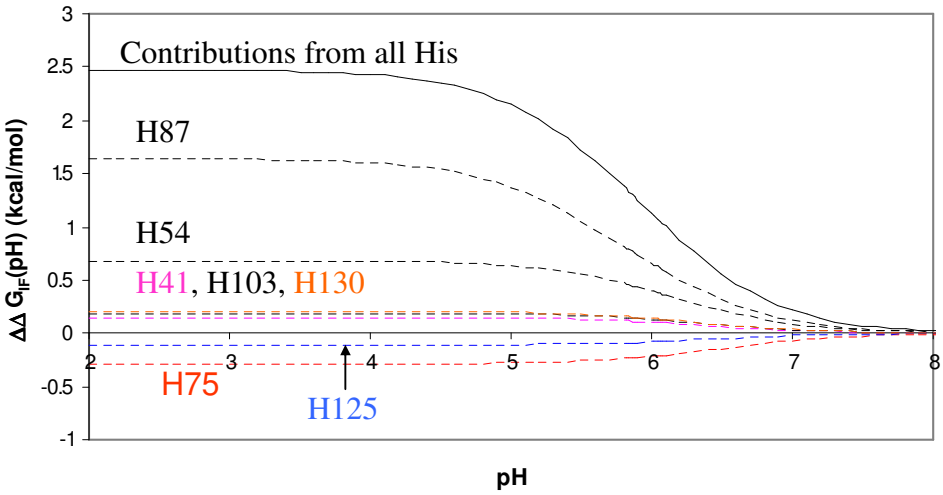



Figure 3.5: **Simulation of acid denaturation due to His ionization.** Using the known $pK_{a,N}$ for the seven His and an assumed $<pK_{a,I}>=6.4$, the observed high pH titration for TrAvrPto destabilization can be simulated (solid black line). The contributions from each individual His are shown in dashed lines, revealing that the high pH titration is the most heavily influenced by H87's protonation.

surface in the native conformation and, as implied by the simulations, are also on the intermediate's surface.

TrAvrPto is not the only protein with both multiple conformations (U,I,N) and acid denaturation that is controlled by the protonation state of a buried His. The villin headpiece has a partially folded intermediate (Grey et al. 2006). The denaturation of native sperm whale apoglobin to its low pH intermediate is driven by ionization of buried H24 (Geierstanger et al. 1998).

The $\{K_{eq1}, pK_{stab1}, K_{eq2}, pK_{stab2}\}$ constants fitted to the NMR data give the relative free energies of the states $\{UH_2, IH_2, IH, NH, N\}$ for TrAvrPto. These data, combined with the theory for pH-dependence of stability, offer a useful guide to predicting the effects of mutating any one of the His, which will be explored in Chapter Four.

*H87 is conserved for AvrPto and AvrPtoB*

The conservation of an amino acid residue across multiple homologous proteins generally implies that that residue has some function within the proteins, whether enzymatic or related to stability. There are three sequence homologues for full-length AvrPto, each with >95% identity; they are very closely related variants of the same protein. AvrPto aligns with an 80-residue segment of AvrPtoB (Xiao et al. 2007). Twelve homologues were found for AvrPtoB, yielding the proteins VirPphA, HolPmaN, AvrPtoB, and their variants. Importantly, all 15 sequences (Table 3.3) are for effector proteins secreted *via* the TTSS. The sequence alignments of the fifteen homologues of AvrPto and AvrPtoB is shown in Figure 3.6. The putative effector protein HolPmaN, aligns with only the first section of the sequences, in a region corresponding to the first three helices of TrAvrPto. It was included only for the sake of completeness and will not be discussed further as it does not include the entire core

```
                      H41          H54                  H75
TrAvrPto 30  NVTSSQLLSVRHQLAESAGLPRDQHEFVSSQAPQSLRNRYNNLYSHT-QR-TLDMADMQ
Q08242   30  NVTSSQLLSVRHQLAESAGLPRDQHEFVSSQAPQSLRNRYNNLYSHT-QR-TLDMADMQ
Q87Y16   30  NVTSSQLLSVRHQLAESAGVPRDQHEFVSNQAPQSLRNRYNNLYSHT-QR-TLDMADMQ
Q4ZLM6   30  NVTSNQLLSVRRQLAESAGLPRDQHEFISSQAPESLRSSYNNLYSHT-QR-TLDFADMQ
Q8RSY1   117 DASAPRRGAVAHANSIVQQLVSEGADI--SHTRNMLRNAMNGDAVAF-SR-VEQNIFRQ
A5AC83   117 DASAPRRGAVAHANSIVQQLVSEGADI--SHTRNMLRNAMNGDAVAF-SR-VEQNIFRQ
Q2QCI9   131 GTSPRHTGAVPHANRIVQQLVDAGADL--AGINTMIDNAMRRHAIALPSR-TVQSILIE
A5AC84   123 ARAPQHAGATAHANSIVQQLVEAGADL--AHTRTMFRNILRGEEIAL-S-LAEQSVLLQ
Q9RBW3   94  -----------QAERIVQELVRGGADL--NNVRTMLRNVMDNNAVAF-SR-VERDILLQ
Q48B61   94  -----------QAERIVQELVRGGADL--NNVRTMLRNVMDNNAVAF-SR-VERDILLQ
Q8RK09   94  -----------QAERIVQELVRGGADL--NNVRTMLRNVMDNNAVAF-SR-VERDILLQ
A2BCU2   94  -----------QAERIVQELVRGGADL--NNVRTMLRNVMDNNAVAF-SR-VERDILLQ
Q4ZMD6   93  AERI-------VAERIVQELVRAGANL--NNVRTMLRNVMDNNAVAF-SR-VEWNILLQ
Q8RK12   94  -----------QAERIVQELVRGGADL--NNVRTMLRNVMDNNAVAF-SR-VERDILLQ
Q8RP04   134 STSPLYTGAVPRANRIVQQLVEAGADL--ANIRTMFRNMLRGEEMIL-SRAEQN-VFLQ
Q8RP02   120 ----QHLGARAHANGIVQQLVDAHADL--AGMLSMIQDGRRG-----------------

             H87            H103                  H125 H130
TrAvrPto     HRYMTGASGINPGMLPHENVDDMRSAITDWSDMREALQHAMGIHADI            133
Q08242       HRYMTGASGINPGMLPHENVDDMRSAITDWSDMREALQHAMGIHADIPPSPERFVATM 144
Q87Y16       HRYMTGASGINPGMLPHENVDDMRSAITDWSDMREALQYAMGIHADIPPSPERFVATM 144
Q4ZLM6       HRFMTGASGINPGMLPRENVDDMRSAISDWSDMREALQHAMGIHADIPPSPERFVTTI 144
Q8RSY1       H---------FPNMPMHGISRDSELAI----ELRGALRRAVHQQAASAP--------- 207
A5AC83       H---------FPNMPMHGISRDSELAI----ELRGALRRAVHQQAASAP--------- 207
Q2QCI9       H---------FPHLLAGELISGSELAT----AFRAALRREVRQQEASAPPRTAARSSV 231
A5AC84       H---------FPNMLATGINRHSELAI----ELRGALRRADSQQAVPAPARTPPRSSA 222
Q9RBW3       H---------FPNMPMTGISSDSVLAN----ELRQRLRQTVRQQ-------------- 168
Q48B61       H---------FPNMPMTGISSDSVLAN----ELRQRLRQTVRQQ-------------- 168
Q8RK09       H---------FPNMPMTGISSDSVLAN----ELRQRLRQTVRQQR------------- 169
A2BCU2       H---------FPNMPMTGISSDSVLAN----ELRQRLRQTVRQQ-------------- 168
Q4ZMD6       H---------FPDMHTNGISSDSVLAN----ELRQTLRQVVHQQRTQRAL-------- 177
Q8RK12       H---------FPNMLMTGISSDSVLAN----ELRQRLRQTVRQQ-------------- 168
Q8RP04       H---------FPDMLPCGIDRNSELAI----ALREALRRADSQQAARAPARTPPRSSV 228
Q8RP02       ------------------------------------------------------------
```

Figure 3.6: **Alignment of AvrPto and AvrPtoB homologues.** Of the seven His in TrAvrPto, H87 is the only one that is conserved in both AvrPto and AvrPtoB homologues.

region of native TrAvrPto. All seven His are absolutely or conservatively conserved in the three AvrPto variants. Several residues are conserved in all 14 sequences, most notably H87. H41, H125, and H130 show only conservative substitutions for polarity and for, in one H125 alignment, aromaticity. However, H87 is the *only* one of the seven His that is absolutely conserved. The conservation of H87 implies that a His is required at that specific position in the protein and, given that H87 has no known catalytic role, the importance of this buried residue is most likely structural. Altogether, given the weight of the stability, titration, structural, and simulation data, it is very probable that the buried H87 acts as a pH switch for TrAvrPto stability and represents a mechanism utilized by a subset of other effector proteins for sensing and responding to changes in pH.

Table 3.3: **AvrPto and AvrPtoB homologues**

| Swiss-Prot primary accession number | Common name(s) | Organism |
|---|---|---|
| *AvrPto Homologues* | | |
| Q08242[a] | AvrPto, AvrPto1 | P. syringae pv tomato |
| Q87Y16 | AvrPto, AvrPto1 | P. syringae pv tomato |
| Q4ZLM6 | AvrPto, AvrPto1 | *P. syringae* pv syringae |
| *AvrPtoB Homologues* | | |
| Q8RSY1 | AvrPtoB, HopAB2, HopAB3 | *P. syringae* pv tomato |
| A5AC83 | AvrPtoB homologue | *P. avellanae* |
| Q2QCI9 | AvrPtoB, HopAB2, HopAB3 | *P. syringae* pv tomato |
| A5AC84 | AvrPtoB, HopAB2, HopAB3 | *P. syringae* pv. theae |
| Q9RBW3 | HopAB1, VirPphA | *P.syringae* pv. phaseolicola |
| Q48B61 | HopAB1, VirPphA | *P. syringae* pv phaseolicola (strain 1448A / Race 6) |
| Q8RK09 | HopAB1, VirPphA | *P. syringae* pv. savastanoi |
| A2BCU2 | HopAB1, VirPphA | *P. syringae* pv. phaseolicola |
| Q4ZMD6 | HopAB1, VirPphA | *P. syringae* pv. syringae (strain B728a) |
| Q8RK12 | HopAB1, VirPphA | *P. syringae* pv. glycinea |
| Q8RP04 | HopPmaL, HopAB3 | *P. syringae* pv. maculicola |
| Q8RP02 | HolPmaN | *P. syringae* pv maculicola |

[a] Sequence from which TrAvrPto construct was derived

***Conclusion***

Little is presently known regarding the detailed mechanisms of effector protein secretion and knowledge of the key determinants of the stability of effector proteins, such as AvrPto, is vital for understanding how the protein can be efficiently unfolded and secreted by the TTSS, but still refold into functional form once inside the host cell cytoplasm. Under mildly acidic conditions, mimicking the pH when *P. syringae* is in the host plant tissue, TrAvrPto is less stable, providing a pool of secretion-ready protein. Where the pH is neutral, such as inside the host cell, TrAvrPto is at its most stable, which ensures that most of the injected population is in the functional folded conformation. The observed acid denaturation at high pH is regulated by the ionization of the His sidechains, predominantly by that of the buried H87 sidechain. Furthermore, H87 is conserved for AvrPto and AvrPtoB homologues, implying that the presence of a pH folding switch is a potential mechanism for regulating secretion- and functional-efficiency in a subset of effector proteins.

***Materials and Methods***

A. Sample preparation

$^{15}$N-labeled TrAvrPto samples used for NMR and CD were prepared as described previously (Dawson et al. 2008). Natural abundance TrAvrPto samples used for fluorescence were cultured on LB broth, but otherwise prepared in an identical manner. All samples are in a variant of Sorensen's phosphate buffer: {10 mM $NaH_2PO_4 \bullet H_2O$, 5 mM $KH_2PO_4$, 225 mM NaCl ; ionic strength 240 mM} for the NMR and CD samples and {33 mM $Na_2HPO_4$, 34 mM $KH_2PO_4$, 118 mM NaCl ; ionic strength 251 mM} for the fluorescence samples. The pH of the samples were adjusted using small amounts of NaOH and HCl. All experiments were performed at 25°C.

B. Circular Dichroism spectroscopy

The acid denaturation of 10μM TrAvrPto was quantified from pH 2.3 to 7.8 using circular dichroism spectroscopy (CD) on a Aviv Biomedical Circular Dichroism spectrometer (Model 202-01). The wavelength was sampled every 5 nm between 260 and 200 nm with each sampled point averaged for three seconds. A narrower scan was taken from 225 to 218 nm, sampled every 1 nm with 10 second averaging. The data was converted to $\Delta\varepsilon_{MRW}$(pH), the mean residue CD extinction coefficients (Martin et al. 2008), using TrAvrPto's 11,875.1 Da molecular weight, 105 residues, and 10 μM concentration. The pathlength of the quartz cuvette was 0.1 cm. The protein $\Delta\varepsilon_{MRW}$ was subtracted in reference to the pH 6.9 buffer $\Delta\varepsilon_{MRW}$ profile. The data was fitted to a two state titration model by nonlinear chi-squared minimization:

$$x(pH) = \frac{x^{A} + x^{HA} * 10^{pK-pH}}{1+10^{pK-pH}}. \qquad \text{Eq. A1}$$

For CD data, x is $\Delta\varepsilon_{MRW}$. $x^{A}$ and $x^{HA}$ are $\Delta\varepsilon_{MRW}$ at the high (A) and low (HA) pH limits, respectively. The term pK is the pH at which half of the total population is "folded". The "folded" CD signal is proportional to the sum of the intermediate ($p_{IH2}$+$p_{IH}$) and native ($p_{NH}$+$p_{N}$) TrAvrPto populations:

$$p_{IH2} + p_{IH} + p_{NH} + p_{N} = \frac{x(pH) - x^{HA}}{x^{A} - x^{HA}}. \qquad \text{Eq. A2}$$

The uncertainty in ($p_{IH2}$+$p_{IH}$+$p_{NH}$+$p_{N}$) was derived from the CD signal *via* propagation of error.

C. Trp Fluorescence

The acid denaturation of 15 μM TrAvrPto was characterized from pH 2.6 to 7.8 *via* Trp fluorescence using a Bio-tek Synergy HT 96-well plate reader in time-resolved fluorescence mode. The single, buried Trp of TrAvrPto was excited at 286±5 nm from the monochromator source. The emitted light was collected at 340 nm with a 30 nm full-bandwidth filter. The buffer subtracted fluorescence intensities, x=I,

during denaturation were fitted via nonlinear chi-squared minimization to the two state titration model (Eq. A1). Similarly to the CD data, x(pH)=I(pH) is proportional to the summed intermediate and native populations (Eq. A2). The uncertainty in fractional population ($p_{IH2}+p_{IH}+p_{NH}+p_N$) is 0.05751, estimated as a standard deviation of data points from pH 5.6 to 7.8, where the summed population values were all approximately unity.

D. NMR Spectroscopy

The pH-dependence of TrAvrPto conformation and stability, as well as the $pK_a$ values of the His sidechains, were evaluated using separate $^{15}N$-$^{1}H$ HSQC and HMQC NMR spectra (Mulder et al. 1996) for each pH point investigated: pH 3.3, 4.6, 4.9, 5.0, 5.1, 5.5, 6.0, 6.1, 6.2, 7.0, and 7.3. The concentrations of the samples varied from 0.5 mM (pH 4.6 & 5.0 samples) to 0.6 mM (pH 5.1, 5.5, 6.0, & 7.3) to 1.2 mM (pH 3.3, 4.9, 6.1, 6.2, & 7.0). However, there are no aggregation or dimerization events for TrAvrPto in this concentration range (Dawson et al. 2008).

All NMR data were collected on a Varian Inova 600 MHz spectrometer with a {H,C,N} z-axis gradient probe. The experiments were performed at 25ºC and the data were processed using NMRpipe and NMRDraw (Delaglio et al. 1995).

*1. pH-dependence of TrAvrPto stability*

The pH-dependence of TrAvrPto stability was probed for individual NH groups using separate $^{15}N$-$^{1}H$ HSQC NMR spectrum (Mulder et al. 1996). A five second delay interval ensured that full re-equilibration of magnetization occurred between each repetition of the pulse sequence. The spectra were processed using a shifted sine-bell apodization function and zero-filling. The peak volumes were determined using Sparky (Goddard et al. ), where a Gaussian lineshape was assumed. Since progressive sample degradation contributed to the intensities of the non-native protein peaks, only those spectra with no observed degradation were included in the

stability analysis: pH 4.6, 5.0, 5.5, 6.0, 6.1, 6.2, and 7.0. Partially degraded samples were identified by the presence of new sharp peaks at random coil chemical shifts. Some degradation was observed in the pH 3.3 spectrum. However, the absence of peaks for native TrAvrPto demonstrates that there was no native protein at this pH, an observation that was included in the stability analysis.

TrAvrPto folding occurs on the slow chemical exchange timescale. The volumes of the native protein's peaks are proportional to the total native population, ($p_{NH}$+$p_N$). Both the unfolded ensemble and the intermediate conformation contribute to the non-native TrAvrPto peaks, ($p_{UH2}$+$p_{IH2}$+$p_{IH}$)=1-($p_{NH}$+$p_N$). Using the ratio of native ($I_N$) to non-native ($I_D$) peak volumes,

$$K(pH) = \frac{I_N(pH)}{I_D(pH)} = \frac{p_{NH} + p_N}{p_{UH2} + p_{IH2} + p_{IH}}, \qquad \text{Eq. A3}$$

the total native population can be determined for each pH point:

$$p_{NH} + p_N = \frac{K(pH)}{1 + K(pH)}. \qquad \text{Eq. A4}$$

The experimental uncertainty was estimated as the standard deviation in ($p_{NH}$+$p_N$) across all investigated groups.

2. *Quantifying the $pK_a$ values and peak linewidths of His in native TrAvrPto*

His sidechain $pK_a$ values were quantified *via* a series of modified HMQC spectra (Bax et al. 1990, Pelton et al. 1993) optimized to refocus two-bond magnetization transfers between carbon-bound $^1$H and $^{15}$N nuclei within a histidine (His) sidechain. A J-coupling of 20 Hz was found empirically to yield the optimal spectra, compensating for the relaxation of the spins during the experiment. The desired resonances were recorded using a 10,000 Hz sweepwidth centered at 4.74 ppm in the $^1$H dimension and a 10,000 Hz sweepwidth centered at 220 ppm in the $^{15}$N

dimension. A delay interval of 1 second was used for all spectra. Chemical shift data was collected from spectra apodized as for the HSQC data.

The chemical shift data, $x=\omega(pH)$, were fitted to to Eq. A1. $x^{A}=\omega_{His0}$ and $x^{HA}=\omega_{His+}$, are the chemical shifts for $His^0$ and $His^+$, respectively, and $pK=pK_a$, the pH at which half of the His is protonated ($His^+$). The $^{15}N$ and $^{1}H$ chemical shift data were fitted separately, yielding an estimate of the parameter uncertainty.

$^{15}N$ linewidths (in Hz), unlike the chemical shift and stability data, were obtained from unapodized spectra by fitting the peaks to a Lorentzian lineshape.

3. *Assignment of His sidechain $^{15}N$*

The His sidechain $^{15}N_{\varepsilon 2}$, $^{15}N_{\delta 1}$, $^{1}H_{\varepsilon 1}$, $^{1}H_{\delta 2}$ spins were assigned using a combination of previous assignments (Jennifer Wulf, unpublished data) and the HSQC, HMQC, and $^{15}N$-$^{1}H$ NOESY spectra at pH 6.1. The 3D $^{15}N$-$^{1}H$ NOESY was collected with 1024, 196, and 24 (real+imaginary) points in the $\{^{1}H, ^{1}H, ^{15}N\}$ dimensions. The sweepwidths were 8999.9, 7000, and 1094 Hz, respectively. Due to progressive sample degradation, only the first six $^{15}N$ points were used. The spectrum was apodized, zerofilled in all dimensions and linear predicted in the $^{15}N$ dimension.

E. Two titration model for TrAvrPto pH-dependence of stability

The CD, Trp fluorescence, and NMR data were modeled by a five state equilibrium $\{UH_2, IH_2, IH, NH, N\}$ in which there are two pH titration events that affect the protein stability and the conformational changes are considered separate from the ionizations (Eq. 1-7). The CD and Trp fluorescence data were modeled by $(1-p_{UH2})=(p_{IH2}+p_{IH}+p_{NH}+p_N)$ and the NMR data by $(p_{NH}+p_N)$. NMR data for individual groups were simultaneously fitted along with the CD and Trp fluorescence data to their respective modeling equations for $\{K_{eq1}, pK_{stab1}, K_{eq2}, pK_{stab2}\}$. The fit was optimized *via* nonlinear chi-squared minimization using the experimental uncertainties for their fractional populations. Since each of the thirteen groups probed

by NMR reported on the same global events, the standard deviations over the thirteen fits were used as estimates of the uncertainty in $\{K_{eq1}, pK_{stab1}, K_{eq2}, pK_{stab2}\}$.

F. Alignment of AvrPto and AvrPtoB homologues

Homologous sequences of AvrPto were identified by BLAST (Altschul et al. 1997) and aligned *via* CLUSTAL W. The same process was used for AvrPtoB homologues. These two sets of sequences were then aligned relative to one another using the alignment in the literature (Xiao et al. 2007).

**Acknowledgements**

I thank Dr. Cynthia Kinsland of the Cornell University Protein Facility for providing fluorescence spectrometer and Brian Zoltowski, Dr. Lea Vacca Michel, and Professor Brian Crane for the use and training on the CD spectrometer. I would especially like to thank Dr. Jens Nielsen of University College, Dublin for his comments regarding my Protein Society poster (21st Annual Symposium of the Protein Society, July 2007), which lead me to the previously unknown world of pH denaturation theory. This work was supported by NSF grant MCB-0641582 and NIH Molecular Biophysics Training Grant T32 GM08267.

# REFERENCES


Alexandrescu, A.T., Evans, P.A., Pitkeathly, M., Baum, J. and Dobson, C.M. 1993. Structure and dynamics of the acid-denatured molten globule state of alpha-lactalbumin: a two-dimensional NMR study. *Biochemistry* **32:** 1707-1718.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389-3402.

Arai, M. and Kuwajima, K. 2000. Role of the molten globule state in protein folding. *Adv.Protein Chem.* **53:** 209-282.

Armstrong, K.M. and Baldwin, R.L. 1993. Charged histidine affects alpha-helix stability at all positions in the helix by interacting with the backbone charges. *Proc.Natl.Acad.Sci.U.S.A.* **90:** 11337-11340.

Ayed, A., Mulder, F.A., Yi, G.S., Lu, Y., Kay, L.E. and Arrowsmith, C.H. 2001. Latent and active p53 are identical in conformation. *Nat.Struct.Biol.* **8:** 756-760.

Bachovchin, W.W. 1986. 15N NMR spectroscopy of hydrogen-bonding interactions in the active site of serine proteases: evidence for a moving histidine mechanism. *Biochemistry* **25:** 7751-7759.

Bax, A., Ikura, M., Kay, L.E., Torchia, D.A. and Tschudin, R. 1990. Comparison of different modes of two-dimensional reverse-correlation
NMR for the study of proteins. *J Magn Reson* **86:** 304-318.

Betz, M., Lohr, F., Wienk, H. and Ruterjans, H. 2004. Long-range nature of the interactions between titratable groups in Bacillus agaradhaerens family 11 xylanase: pH titration of B. agaradhaerens xylanase. *Biochemistry* **43:** 5820-5831.

Blocker, A., Jouihri, N., Larquet, E., Gounon, P., Ebel, F., Parsot, C., Sansonetti, P. and Allaoui, A. 2001. Structure and composition of the Shigella flexneri "needle complex", a part of its type III secreton. *Mol. Microbiol.* **39:** 652-663.

Blomberg, F., Maurer, W. and Ruterjans, H. 1977. Nuclear magnetic resonance investigation of 15N-labeled histidine in aqueous solution. *J.Am.Chem.Soc.* **99:** 8149-8159.

Burley, S.K. and Petsko, G.A. 1985. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **229:** 23-28.

Chang, J.H., Rathjen, J.P., Bernal, A.J., Staskawicz, B.J. and Michelmore, R.W. 2000. AvrPto enhances growth and necrosis caused by Pseudomonas syringae pv. tomato in tomato lines lacking either Pto or Prf. *Mol. Plant-Microbe Interact.* **13:** 568-571.

Cordes, F.S., Komoriya, K., Larquet, E., Yang, S., Egelman, E.H., Blocker, A. and Lea, S.M. 2003. Helical structure of the needle of the type III secretion system of Shigella flexneri. *J. Biol. Chem.* **278:** 17103-17107.

Dawson, J.E. and Nicholson, L.K. 2008. Folding kinetics and thermodynamics of Pseudomonas syringae effector protein AvrPto provide insight into translocation via the type III secretion system. *Protein Sci.* **17:** 1109-1119.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6:** 277-293.

Dyson, H.J. and Wright, P.E. 2002. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Adv.Protein Chem.* **62:** 311-340.

Emerson, S.U., Tokuyasu, K. and Simon, M.I. 1970. Bacterial flagella: polarity of elongation. *Science* **169:** 190-192.

Evanics, F., Bezsonova, I., Marsh, J., Kitevski, J.L., Forman-Kay, J.D. and Prosser, R.S. 2006. Tryptophan solvent exposure in folded and unfolded states of an SH3 domain by 19F and 1H NMR. *Biochemistry* **45:** 14120-14128.

Francis, N.R., Sosinsky, G.E., Thomas, D. and DeRosier, D.J. 1994. Isolation, characterization and structure of bacterial flagellar motors containing the switch complex. *J. Mol. Biol.* **235:** 1261-1270.

Galan, J.E. and Collmer, A. 1999. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284:** 1322-1328.

Geierstanger, B., Jamin, M., Volkman, B.F. and Baldwin, R.L. 1998. Protonation behavior of histidine 24 and histidine 119 in forming the pH 4 folding intermediate of apomyoglobin. *Biochemistry* **37:** 4254-4265.

Goddard, T.D. and Kneller, D.G. *Sparky 3,* University of California, San Francisco,

Grey, M.J., Tang, Y., Alexov, E., McKnight, C.J., Raleigh, D.P. and Palmer, A.G.,3rd. 2006. Characterizing a partially folded intermediate of the villin headpiece domain under non-denaturing conditions: contribution of His41 to the pH-dependent stability of the N-terminal subdomain. *J.Mol.Biol.* **355:** 1078-1094.

Grignon, C. and Sentenac, H. 1991. pH and ionic conditions in the apoplast. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42:** 103-128.

Guttman, D.S., Vinatzer, B.A., Sarkar, S.F., Ranall, M.V., Kettler, G. and Greenberg, J.T. 2002. A functional screen for the type III (Hrp) secretome of the plant pathogen Pseudomonas syringae. *Science* **295:** 1722-1726.

Hauck, P., Thilmony, R. and He, S.Y. 2003. A Pseudomonas syringae type III effector suppresses cell wall-based extracellular defense in susceptible Arabidopsis plants. *Proc. Natl. Acad. Sci. U.S.A.* **100:** 8577-8582.

Jin, Q., Thilmony, R., Zwiesler-Vollick, J. and He, S.Y. 2003. Type III protein secretion in Pseudomonas syringae. *Microb. Infect.* **5:** 301-310.

Johnson, S., Deane, J.E. and Lea, S.M. 2005. The type III needle and the damage done. *Curr. Opin. Struct. Biol.* **15:** 700-707.

Kubelka, J., Hofrichter, J. and Eaton, W.A. 2004. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **14:** 76-88.

Lee, V.T. and Schneewind, O. 2002. Yop fusions to tightly folded protein domains and their effects on Yersinia enterocolitica type III secretion. *J. Bacteriol.* **184:** 3740-3745.

Marlovits, T.C., Kubori, T., Sukhan, A., Thomas, D.R., Galan, J.E. and Unger, V.M. 2004. Structural insights into the assembly of the type III secretion needle complex. *Science* **306:** 1040-1042.

Marsh, J.A., Neale, C., Jack, F.E., Choy, W.Y., Lee, A.Y., Crowhurst, K.A. and Forman-Kay, J.D. 2007. Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. *J.Mol.Biol.* **367:** 1494-1510.

Martin, S.R. and Schilstra, M.J. 2008. Circular dichroism and its application to the study of biomolecules. *Methods Cell Biol.* **84:** 263-293.

Minamino, T., Imae, Y., Oosawa, F., Kobayashi, Y. and Oosawa, K. 2003. Effect of intracellular pH on rotational speed of bacterial flagellar motors. *J. Bacteriol.* **185:** 1190-1194.

Mittag, T. and Forman-Kay, J.D. 2007. Atomic-level characterization of disordered protein ensembles. *Curr.Opin.Struct.Biol.* **17:** 3-14.

Mulder, F.A.A., Spronk, C.A.E.A., Slijper, M., Kaptein, R. and Boelens, R. 1996. Improved HSQC experiments for the observation of exchange broadened signals. *J. Biomol. NMR* **8:** 223-228.

Nozaki, Y. and Tanford, C. 1962. Examination of titration behavior. *Advan Protein Chem* **17:** 715-734.

Pace, C.N. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131:** 266-280.

Pedley, K.F. and Martin, G.B. 2003. Molecular basis of Pto-mediated resistance to bacterial speck disease in tomato. *Annu. Rev. Phytopathol.* **41:** 215-243.

Pelton, J.G., Torchia, D.A., Meadow, N.D. and Roseman, S. 1993. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Sci.* **2:** 543-558.

Petnicki-Ocwieja, T., Schneider, D.J., Tam, V.C., Chancey, S.T., Shan, L., Jamir, Y., Schechter, L.M., Janes, M.D., Buell, C.R., Tang, X., Collmer, A. and Alfano, J.R. 2002. Genomewide identification of proteins secreted by the Hrp type III protein secretion system of Pseudomonas syringae pv. tomato DC3000. *Proc.Natl.Acad.Sci.U.S.A.* **99:** 7652-7657.

Plesniak, L.A., Connelly, G.P., Wakarchuk, W.W. and McIntosh, L.P. 1996. Characterization of a buried neutral histidine residue in Bacillus circulans xylanase: NMR assignments, pH titration, and hydrogen exchange. *Protein Sci.* **5:** 2319-2328.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. 1988. *Numerical recipes in C,* 521-522. Cambridge University Press, Cambridge, UK.

Rahme, L.G., Mindrinos, M.N. and Panopoulos, N.J. 1992. Plant and environmental sensory signals control the expression of hrp genes in Pseudomonas syringae pv. phaseolicola. *J. Bacteriol.* **174:** 3499-3507.

Scofield, S.R., Tobias, C.M., Rathjen, J.P., Chang, J.H., Lavelle, D.T., Michelmore, R.W. and Staskawicz, B.J. 1996. Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. *Science* **274:** 2063-2065.

Seavey, B.R., Farr, E.A., Westler, W.M. and Markley, J.L. 1991. A relational database for sequence-specific protein NMR data. *J.Biomol.NMR* **1:** 217-236.

Shan, L., Thara, V.K., Martin, G.B., Zhou, J.M. and Tang, X. 2000. The Pseudomonas AvrPto protein is differentially recognized by tomato and tobacco and is localized to the plant plasma membrane. *Plant Cell* **12:** 2323-2338.

Sudmeier, J.L., Evelhoch, J.L. and Jonsson, N.B.H. 1980. Dependence of NMR Lineshape Analysis upon chemical rates and Mechanisms: Implications for enzyme histidine titrations. *J Magn Reson* **40:** 377-390.

Tang, X., Xie, M., Kim, Y.J., Zhou, J., Klessig, D.F. and Martin, G.B. 1999. Overexpression of Pto activates defense responses and confers broad resistance. *Plant Cell* **11:** 15-29.

van Dijk, K., Fouts, D.E., Rehm, A.H., Hill, A.R., Collmer, A. and Alfano, J.R. 1999. The Avr (effector) proteins HrmA (HopPsyA) and AvrPto are secreted in culture from Pseudomonas syringae pathovars via the Hrp (type III) protein secretion system in a temperature- and pH-sensitive manner. *J. Bacteriol.* **181:** 4790-4797.

Whalen, M.C., Innes, R.W., Bent, A.F. and Staskawicz, B.J. 1991. Identification of Pseudomonas syringae pathogens of Arabidopsis and a bacterial locus determining avirulence on both Arabidopsis and soybean. *Plant Cell* **3:** 49-59.

Wilharm, G., Lehmann, V., Krauss, K., Lehnert, B., Richter, S., Ruckdeschel, K., Heesemann, J. and Trulzsch, K. 2004. Yersinia enterocolitica type III secretion depends on the proton motive force but not on the flagellar motor components MotA and MotB. *Infect. Immun.* **72:** 4004-4009.

Wishart, D.S. and Case, D.A. 2001. Use of chemical shifts in macromolecular structure determination. *Methods Enzymol.* **338:** 3-34.

Wishart, D.S., Sykes, B.D. and Richards, F.M. 1991. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J.Mol.Biol.* **222:** 311-333.

Wolff, N., Deniau, C., Letoffe, S., Simenel, C., Kumar, V., Stojiljkovic, I., Wandersman, C., Delepierre, M. and Lecroisey, A. 2002. Histidine pK(a) shifts and changes of tautomeric states induced by the binding of gallium-protoporphyrin IX in the hemophore HasA(SM). *Protein Sci.* **11:** 757-765.

Wulf, J., Pascuzzi, P.E., Fahmy, A., Martin, G.B. and Nicholson, L.K. 2004. The solution structure of type III effector protein AvrPto reveals conformational and dynamic features important for plant pathogenesis. *Structure* **12:** 1257-1268.

Xiao, F., He, P., Abramovitch, R.B., Dawson, J.E., Nicholson, L.K., Sheen, J. and Martin, G.B. 2007. The N-terminal region of Pseudomonas type III effector AvrPtoB elicits Pto-dependent immunity and has two distinct virulence determinants. *Plant J.* **52:** 595-614.

Xiao, Y., Lu, Y., Heu, S. and Hutcheson, S.W. 1992. Organization and environmental regulation of the Pseudomonas syringae pv. syringae 61 hrp cluster. *J. Bacteriol.* **174:** 1734-1741.

Yang, A.S. and Honig, B. 1993. On the pH dependence of protein stability. *J.Mol.Biol.* **231:** 459-474.

Yonekura, K., Maki-Yonekura, S. and Namba, K. 2003. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424:** 643-650.

Chapter Four

Tuning AvrPto secretion efficiency:

The rational mutagenesis of AvrPto mutants

## *Introduction*

The overarching purpose of this project is to further our understanding of how the Type III secretion system (TTSS) works. Effector proteins, which aid the pathogen in infecting a susceptible host, pass through the TTSS in at least partially unfolded form (Johnson et al. 2005). Some dependence on the effector protein stability and folding rate has been demonstrated: fusion of stable ubiquitin to YopE inhibited secretion (Lee et al. 2002), as did fusion to rapidly-folding DHFR (Sorg et al. 2005). However, this dependence has not been fully quantified. How stable or rapidly-folding may an effector be and still be secreted by the TTSS? The TTSS requires a proton gradient to function (Wilharm et al. 2004) and *P. syringae* has exhibited changes in effector secretion efficiency that depends on the pH of its environment (van Dijk et al. 1999). How then does pH affect effector protein stability and does that in turn affect the protein's ability to be secreted?

In the previous two chapters, the folding kinetics, thermodynamics, and acid denaturation were characterized for an AvrPto construct in which the disordered N- and C-terminal tails had been removed, truncated AvrPto (TrAvrPto). Both protein constructs, AvrPto and TrAvrPto, will be used to probe the kinetic and stability thresholds of the TTSS. As discussed in Chapter Two, the structure of an effector protein may also play a part in secretion, through its putative interaction with the TTSS's ATPase. This complication is removed by using a series of mutants with different stabilities that are based on the same structure, TrAvrPto. Using detailed knowledge of wild-type (wt) TrAvrPto, the series of mutants were designed to

stabilize the protein, remove the pH-dependence of stability, or illuminate further aspects of this complicated protein's folding. The ability of these mutants to be secreted can then be tested using translocation assays developed by the Collmer lab (Schechter et al. 2004), also of Cornell University.

The development of a stabilized mutant is an ambitious goal. *Destabilizing* a protein is simple; almost all amino acid substitutions will reduce protein stability. Indeed, during random mutagenesis screenings of AvrPto, most substitutions reduced the half-life of the protein within the pathogen (Shan et al. 2000). The half-life is a measure of how long it takes for a protein to be degraded within the cell and depends roughly on the protein's stability and folding rate. Making a protein more stable or more rapidly-folding is many orders of magnitude more difficult. No algorithms exist that definitively predict the effects of amino acid substitutions. There are only heuristics, which take advantage of general trends in protein behavior.

Every kind of secondary structural element favors some amino acids over others. For example, α-helices favor residues A, L, M, E, K, R, and Q. Many different mechanisms underlie secondary structure propensities. Within a polypeptide chain, amino acid residues are in close contact with one another. In order to avoid overlap, the amino acids are limited to only a relatively small subset of conformations. In an α-helix, the backbone forms a tight twist with 3.6 amino acids per turn, which limits the available space for the sidechains. Hence, amino acids with bulky aromatic rings that clash with other sidechains or β-branched sidechains that clash with the backbone are disfavored. Electrostatics can also play important roles. Opposite charges in near proximity form stabilizing interactions, while like charges near to one another are energetically unfavorable. The backbone of each amino acid residue is polar with a partially-negative charge (δ–) on the carbonyl and a partially-positive charge (δ+) on the amide group. Due to the geometry of an α-helix, all backbone

amide groups point towards the N-terminal end and all carbonyl toward the C-terminal end, creating an electric dipole across the length of the helix. Therefore, residues with negative charges are favored and those with positive charges are disfavored at the δ+ N-terminus. The reverse holds for the δ- C-terminal end of an α-helix.

Unfolded proteins sample a larger conformational space than folded proteins, which results in a large loss in entropic energy during folding. Proteins recoup this energy loss *via* favorable enthalpic interactions, such as the electrostatic interactions discussed previously. A protein generally maximizes the number of hydrogen-bonds made with other amino acids or solvent. Within an α-helix, each backbone carbonyl hydrogen-bonds with the backbone amide of the residue four amino acids away. However, these periodic bonds cannot be formed at the ends of a helix. Helix capping boxes are formed from sections of polypeptide adjacent to the helix ends that loop back and help satisfy the hydrogen-bonding potential. Packing between hydrophobic residues is another stabilizing interaction. Many helices are packed into the protein with one side facing the hydrophobic core and the other facing the solvent. These amphipathic helices show a periodic pattern of polar residues that interact favorably with the solvent and hydrophobic residues that pack into the nonpolar core.

The amount of entropy lost upon folding can be limited by using residues that are already preordered for the required conformation in the folded protein. Proline's sidechain is fused to its backbone, which dramatically reduces the amount of conformational space it can sample. This small constrained set of conformations matches that required for the center residue of an inverse-γ turn, resulting in only a relatively small loss of entropy upon folding. Placing a flexible Gly at this position would have a large entropic cost of folding.

It has been demonstrated that the stability of a protein can be improved by substituting amino acids for those with increased secondary structural propensities

(Bezsonova et al. 2006, Taddei et al. 2000). There are multiple methods for predicting the propensities of a given protein. The HMMSTR library is an excellent source for secondary structure propensity information and its associated program, I-SITES, uses the library in predicting the likely secondary structure (Bystroff et al. 2000, Bystroff et al. 2002). A similar secondary structure prediction program, Jnet, uses a combination of propensity information and alignment of homologous sequences (Cuff et al. 2000). A nucleation-condensation model, implemented by the program AGADIR, simulates the formation of α-helices (Lacroix et al. 1998, Muñoz et al. 1995a, Muñoz et al. 1995b, Muñoz et al. 1997).

While the testing of the AvrPto mutants by NMR and the translocation assay is still in the preliminary stages, the design work is complete for a number of promising mutants. The identification of these potential candidates drew together information from myriad sources: secondary structural prediction and simulation techniques, thermodynamics theories on pH-dependent protein stability, literature review, and detailed knowledge of how the wild-type protein folds. Some proposed mutants may potentially increase the overall stability of the protein (ex. G95P, Q124L), while others may diminish the pH-dependence of stability (ex. H87Y, H87Q, H54Q). Mutants E104Q and E55Q, while unlikely to increase the protein stability, should yield considerable insight into the low pH titration of TrAvrPto. The overall stability of the effector protein, the rate of its folding, and the tuning of its pH-dependence of stability are all hypothesized to affect its ability to be secreted by the TTSS. Indeed, during translocation assays, both the modestly-stable AvrPto and unstable H54P-AvrPto demonstrated the ability to be transported into the host cell.

***Use of secondary structural prediction methods in mutation design***

The first strategy used for mutation design hinged on secondary structure prediction techniques. This method is useful as a way of identifying mutations that may increase the secondary structure propensities. To avoid disrupting the core, only solvent-exposed residues were changed. Proposed substitutions were tested using AGADIR to see if it increased the predicted α-helix propensity and, if applicable, the capping box propensity. This strategy was also a fruitful aid for aligning the AvrPto and AvrPtoB homologues, as discussed later in this chapter and published elsewhere (Xiao et al 2007).

In order to establish a baseline, both AGADIR and Jnet were used to predict the secondary structure of wt TrAvrPto. AGADIR can predict propensities for a range of solution temperatures, ionic strengths, and pH values. For TrAvrPto and all of the proposed mutants, T=298 K (room temperature), the ionic strength of AvrPto buffer is I=0.24 M, and, since the goal is to improve the overall stability of the protein, pH 6.1 was used. Of the four helices of TrAvrPto (Figure 4.1), AGADIR and Jnet correctly predicted $\alpha_A$, the beginning and end of $\alpha_C$, and parts of $\alpha_D$ (Figure 4.2). There is a sharp decline in predicted α-helix propensity in the middle of helix $\alpha_C$ and a relatively high propensity is predicted in the beginning of the omega loop, where part of the long loop forms a structure similar to a single turn of an α-helix. Helix $\alpha_B$ is only 5 or 6 residues long depending on the NMR structure, has barely any predicted propensity by AGADIR, and is not predicted by Jnet. This deficiency makes it difficult to design $\alpha_B$ mutants using propensities alone. The identification of $\alpha_B$'s H54 as a possible mutagenesis target required an understanding of the pH-dependence of TrAvrPto stability, as discussed in a later section. Secondary structural methods were used to design mutants in TrAvrPto helices $\alpha_A$, $\alpha_C$, and $\alpha_D$ and the Ω loop. The predictions presented in this section are for TrAvrPto mutants with single substitutions. However,

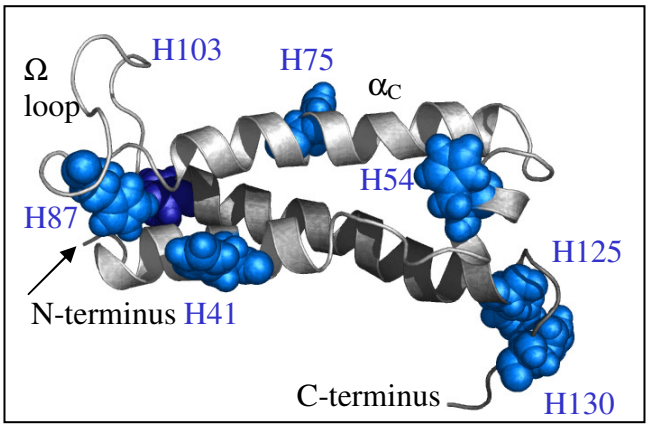



Figure 4.1: **α-helices and histidine sidechains of TrAvrPto.** TrAvrPto is a three-helix bundle ($\alpha_A$, $\alpha_C$, and $\alpha_D$) with a short orthogonal helix ($\alpha_B$) and an Ω loop. The seven His are scattered across the structure. Image rendered with PyMOL (DeLano. 2002) using NMR structure (PDB file 1R5E) (Wulf et al. 2004), ensemble-averaged *via* MOLMOL (Koradi et al. 1996).

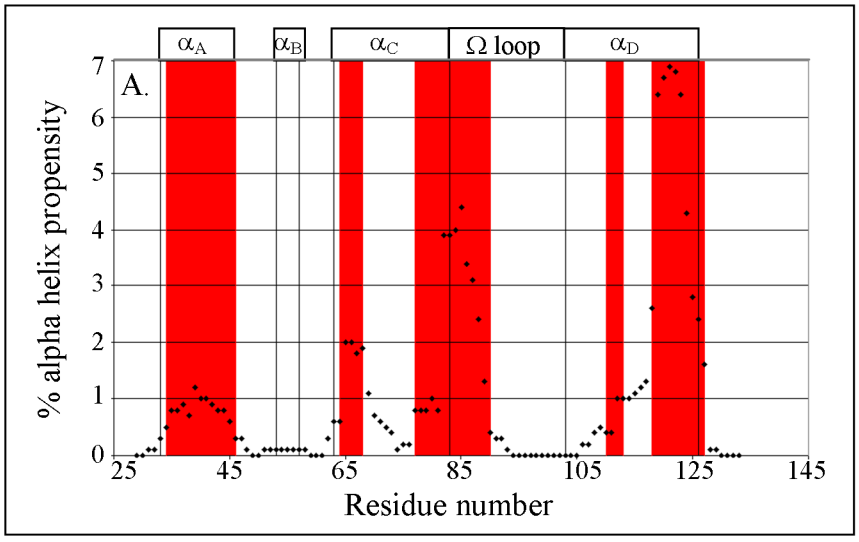



Figure 4.2: **Predicted secondary structure of AvrPto and AvrPtoB(121-200).** Actual and predicted secondary structure of TrAvrPto. The positions of the α-helices were predicted using AGADIR (black diamonds) and Jnet (red shading). The actual positions of the helices are marked by boxes on top.

this strategy may be extended for mutants with multiple substitutions.

*Proposed mutants to improve α-helix, capping box, and turn propensities*

The helix closest to the C-terminus of TrAvrPto, $\alpha_D$ (Figure 4.1), has the greatest predicted α-helix propensity (Figure 4.2). However, the secondary structural propensities may still be improved. Residue Q124 is involved in a Schellman-like C-terminal capping box at the end of $\alpha_D$. Capping boxes are structural motifs at the N- or C-termini of α-helices that stabilize the helix by satisfying the hydrogen bond potentials and, in some cases, by countering the partial charges at either end of the helix dipole (Aurora et al. 1998).

Figure 4.3: **Residues involved in a Schellman C-capping box.**

| | **Helix $\alpha_D$** | | | | C-terminal tail | |
|---|---|---|---|---|---|---|
| | **Q124** | **H125** | **A126** | **M127** | G128 | I129 |
| C-cap index | C3 | C2 | C1 | Ccap | C′ | C′′ |

In a Schellman motif, residue C' (Figure 4.3) is generally Gly. Residues that counter the δ- charge at the C-terminal end of the helix tend to be favored at positions C2, C1, and Ccap. Residues C3 and C" should be hydrophobic, packing together to stabilize the motif. In TrAvrPto helix $\alpha_D$, C3 is Q124, a polar residue. Substituting Q124 with Leu, Ala, or Met, the most preferred residues at this position (Aurora et al. 1998), should increase the C-cap propensities, which in turn should raise the α-helix propensities for this helix. Using AGADIR to predict propensities, it was found that all three proposed mutants (Q124L, Q124A, and Q124M) did indeed increase the C-cap and α-helix propensities (Figure 4.4A,B) with Q124L having the greatest effect, followed by Q124A. Residue T114 is in the middle of $\alpha_D$. The program I-SITES predicted that T114 should occupy a β-strand. Substituting Lys for Thr at this position

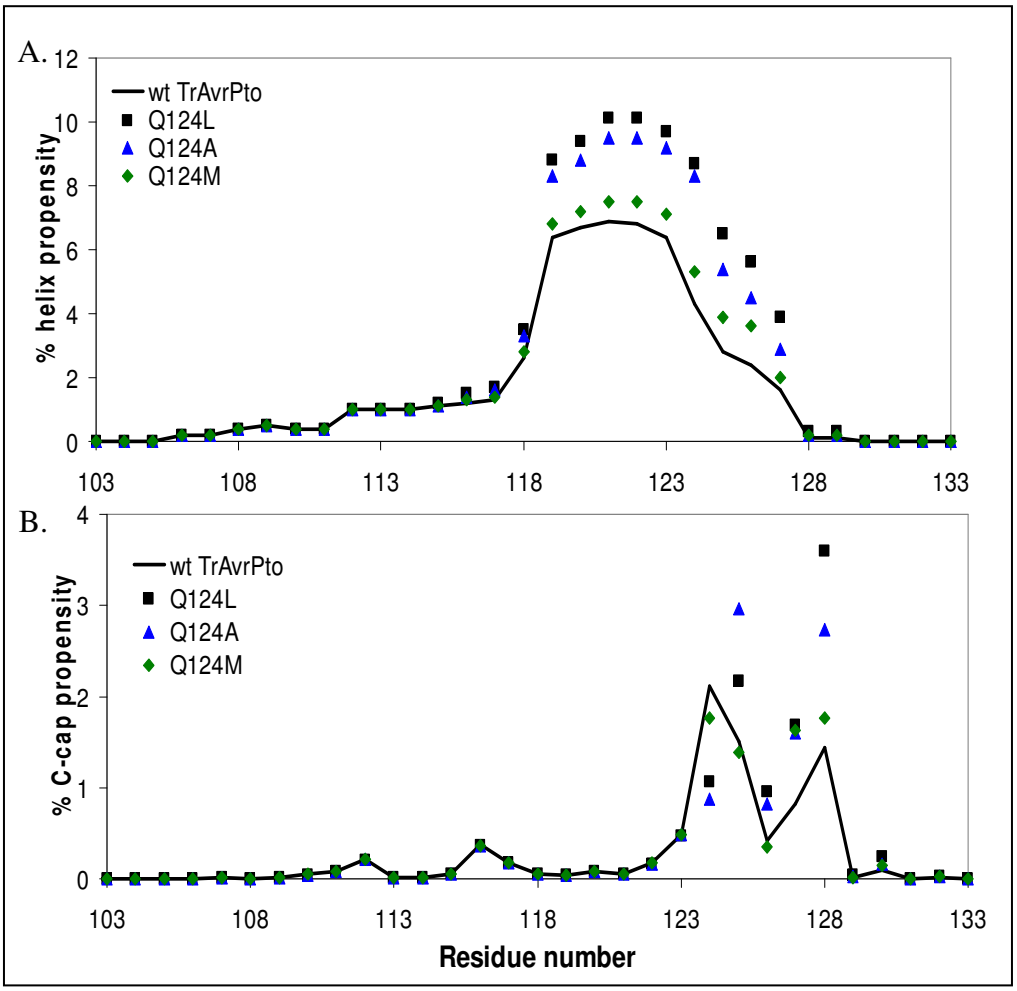



Figure 4.4: **Change in secondary structural propensities in $\alpha_D$ and C-terminus due to proposed Q124 mutants.** Residues 103-133 are shown. A. α-helix propensities for Q124 mutants (Q124L, Q124A, Q124M). B. C-cap propensities for Q124 mutants.

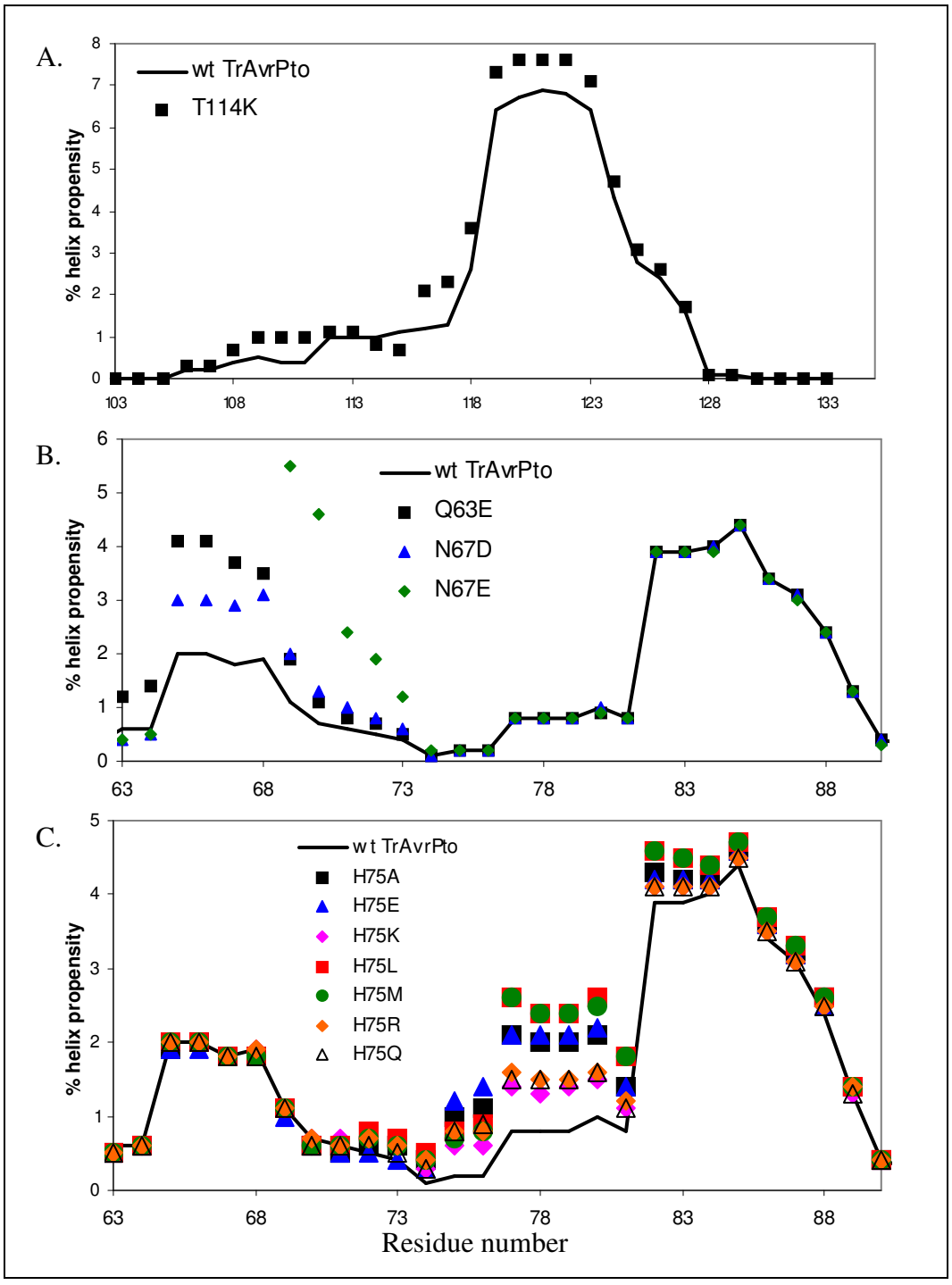



Figure 4.5: **Change in secondary structural propensities for proposed $\alpha_C$ and $\alpha_D$ mutants.** A. α-helix propensities for $\alpha_D$ mutant T114K. B. α-helix propensities for $\alpha_C$ mutants (Q63E, N67D, N67E). C. α-helix propensities for H75 mutants of $\alpha_C$ (H75A, H75L, H74M, H75E, H75R, H75Q).

resulted in an increase in predicted α-helix propensities for a T114K mutant (Figure 4.5A).

G95 is at the apex of the Ω loop and is the center residue of an inverse-γ turn. The backbones of the adjacent residues S94 and I96 are hydrogen-bonded, placing tight constraints on the conformation of G95. AGADIR focuses on α-helices and their associated capping boxes. The design of mutants within the loops must rely on known loop and turn propensities from other literature sources. Gly is the most flexible amino acid found in proteins due to the small size of its sidechain, a proton. In general, the greater the flexibility, the greater the loss of entropy that occurs upon folding. Indeed, Gly is strongly disfavored at the center position of an inverse-γ turn (Guruprasad et al. 2000). Pro is constrained to the geometry required for this position (torsion angle $\phi \sim -60^{\circ}$), resulting in reduced entropy loss. A G95P mutant takes advantage of the high propensity for Pro at this position (Guruprasad et al. 2000) and is predicted to be more stable due to the reduced entropy loss. G95P TrAvrPto might also fold more rapidly the wt TrAvrPto, since Pro would have to sample fewer conformations before folding. If G95 sampling is the event that limits the rate of folding, then a G95P mutation could potentially increase the folding rate to the intermediate or fast chemical exchange timescale in NMR spectra.

There is a low predicted α-helical propensity for the center region of $\alpha_C$ (Figure 4.2), which may contribute to the low overall stability of TrAvrPto. Q63 is at the δ+ end of $\alpha_C$. A Q63E mutant should balance this partial charge and did increase predicted propensity (Figure 4.5B). N67 is between positively-charged R66 and R68. Mutants N67D and especially N67E dramatically increase the predicted propensity (Figure 4.5B), no doubt by countering the neighboring charges. Substitution of H75

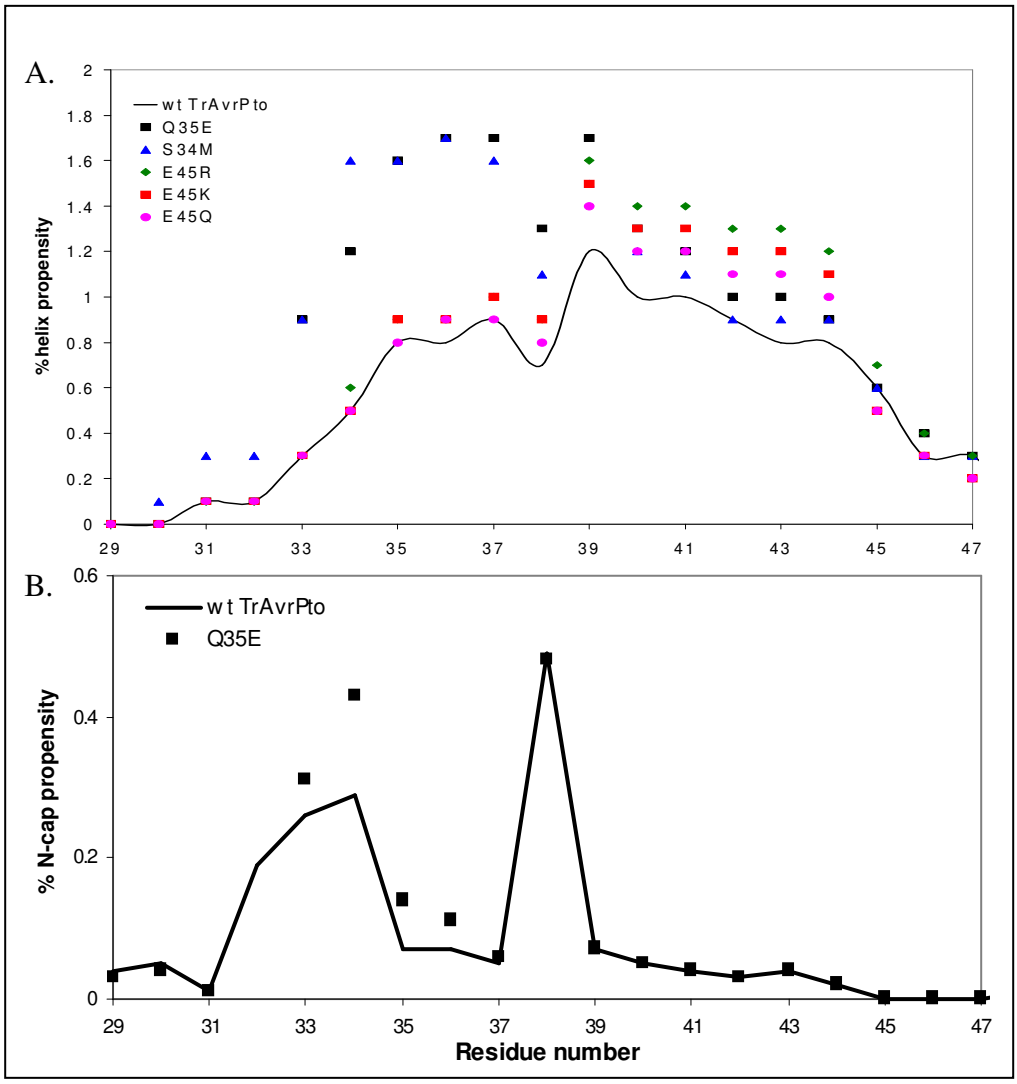



Figure 4.6: **Change in secondary structural propensities for proposed $\alpha_A$ mutants.** A. α-helix propensities for $\alpha_A$ mutants (Q35E, S34M, E45R, E45K, E45Q). B. N-cap propensities for Q35E mutant.

for A, L, M, E, R, K, or Q (all residues favored by α-helices) significantly improved the α-helical character of a region in which little to no propensity had been previously predicted (Figure 4.5C). The H75L and H75M substitutions had the greatest predicted change in propensity. Care should be taken, however, with substitutions such as H75L and H75M, which replace a polar group with a hydrophobic group at a solvent-exposed position. Additionally, as discussed in Chapter Three, the His residues are important to the acid denaturation of TrAvrPto, not just to the overall stability. The effects of removing H75 upon the pH dependence of stability will be explored in the next section.

Within $\alpha_A$, the α-helix nearest to the N-terminus of TrAvrPto, substitutions are proposed for residues at three positions: S34, Q35, and E45. S34 is the second residue of $\alpha_A$. According to the HMMSTR library, Met is much more favored than Ser at this position within an amphipathic helix and, so, increased the predicted α-helix propensity (Figure 4.6A). Both T32, which immediately precedes $\alpha_A$, and Q35 have high propensities for occupying a N-terminal capping box (Aurora et al. 1998). A Q35E substitution further increased the predicted N-cap and α-helix propensities (Figure 4.6 A,B). E45 is at the δ- end of $\alpha_A$, where a positively-charged substituent should be more favored. Both E45R and E45K increase the α-helix propensity (Figure 4.6A). Simply removing the negative charge from this position with an E45Q mutant also increased this propensity to a more minor extent.

*Alignment of AvrPto and AvrPtoB homologues with aid of secondary structure prediction*

Standard methods of identifying homologous proteins rely on sequence alignments. Functionally important residues are conserved during evolution and, so, related proteins have statistically significant numbers of residues that are identical or

chemically similar between proteins. The length and composition of loops are typically less conserved than the rest of the proteins. Alignment programs like BLAST (Altschul et al. 1997) search through the available database of sequences, allowing for the possibility of gaps in parts of the alignment to reflect potential differences in structure. These kinds of programs only use amino acid sequences. Since they do not reference any known structure, they cannot always correctly predict the locations of loops.

AvrPtoB is a 553-residue, multi-domain protein secreted by *P. syringae* pv tomato (Kim et al. 2002). Like AvrPto, it interacts with host kinase Pto within resistant tomato strains. The sets of Pto residues required for interaction with AvrPto or AvrPtoB are very similar, which implies that there may be some structural similarity between these two effector proteins (Bernal et al. 2005, Wu et al. 2004). Residues 121-200 of AvrPtoB constitute a minimal region for the AvrPtoB-Pto interaction (Xiao et al. 2007). Both AGADIR and Jnet predict that AvrPtoB(121-200) contains four α-helices with a pattern of helix propensities that is similar to that of AvrPto under the same solution conditions (Figures 4.2, 4.7).

The locations of loops in TrAvrPto are known from the NMR structures, where the longest is the Ω loop. Multiple motifs are present within the Ω loop, including an inverse-γ turn, as well as a sub-structure similar to a turn of an α-helix at the beginning of the loop. These sub-structures would not necessarily be conserved between proteins.

Homology searches using BLAST uncovered three homologous sequences for AvrPto and 12 for AvrPtoB (Table 3.3), but did not predict homology between these two proteins. The manual alignment between these two sets of sequences was based on the secondary structure prediction and the structure of TrAvrPto, maximizing the

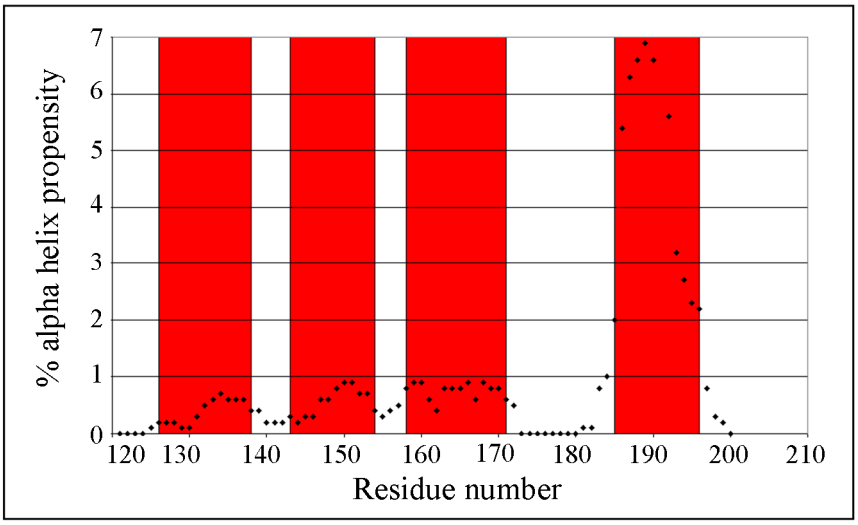



Figure 4.7: **Predicted secondary structure for AvrPtoB(120-200).** The secondary structure for the 80-residue fragment of AvrPtoB(121-200) was predicted using AGADIR and J-net.

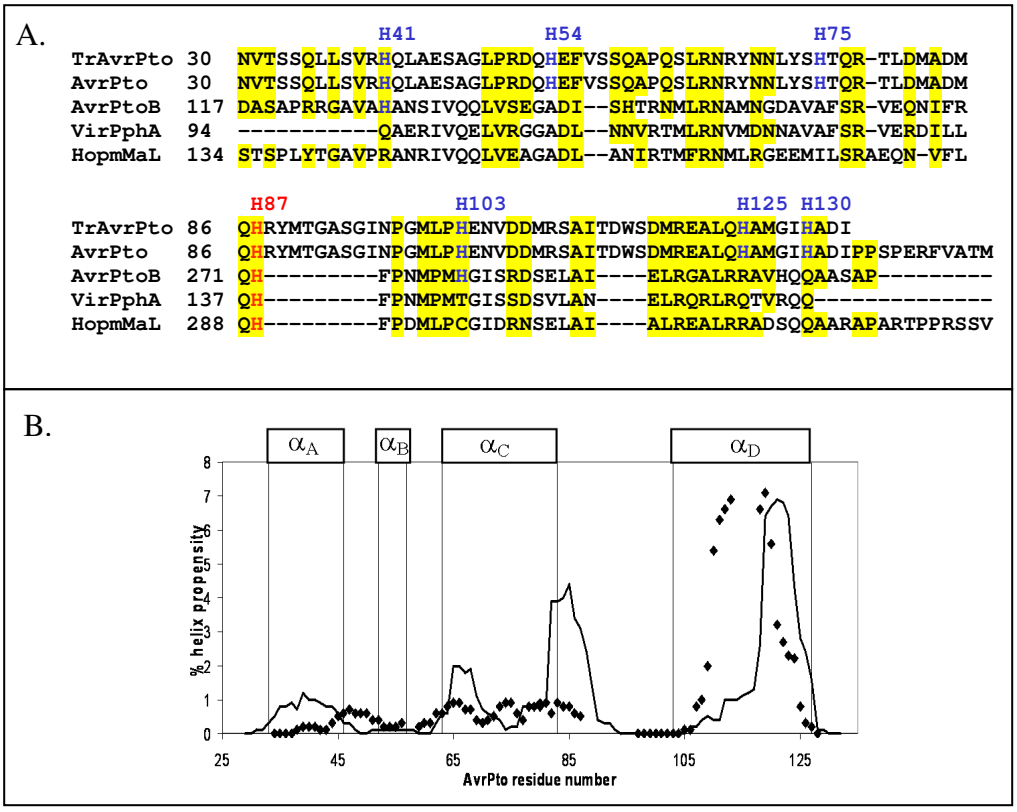



Figure 4.8: **Alignment of AvrPto and AvrPtoB(121-200).** A. Sequence alignment of AvrPto and AvrPtoB homologues. Only non-redundant sequences are shown. Identical and chemically conserved residues are highlighted in yellow. B. α-helix propensities for aligned AvrPto (black line) and AvrPtoB(121-200) (black diamonds) sequences. The four helices of TrAvrPto are indicated in boxes above the graph.

number of aligned residues that are identical or chemically similar and minimizing the number and lengths of the gaps required for the alignment. The best alignment (Figure 4.8A) places the largest gap within the Ω loop (residues 84-102 of AvrPto) (Xiao et al. 2007), meaning that the loop critical for AvrPto's interaction with Pto (Chang et al. 2001, Shan et al. 2000, Wulf et al. 2004) is shorter, but present in AvrPtoB. There is a gap in the AvrPtoB sequences corresponding to one entire turn of $\alpha_D$ (residues T114 to S117). Mutation of T114 to T114K is predicted to improve the helix propensity of $\alpha_D$ in AvrPto (Figure 4.5A). Therefore, it is interesting that AvrPtoB, which does not have this residue, also has a greater predicted propensity in this region (Figure 4.8B). The predicted secondary structure and propensities for AvrPto and AvrPtoB(121-200) are the most similar for the C-terminal half of the protein, where significant α-helix propensities were predicted for $\alpha_C$ and $\alpha_D$. AGADIR predicts a dip in helix propensity in the middle of $\alpha_C$ and Jnet (Figures 4.2, 4.7) predicts two helices in this region in both cases. The alignment of $\alpha_A$ is offset towards the N-terminus in AvrPto and no $\alpha_B$ is predicted for AvrPtoB(121-200).

***Alteration of the pH-dependence of stability***

The folding of TrAvrPto is fascinating and complex, especially for a small 105-residue protein. The native conformation exists in equilibrium with both the unfolded ensemble and an intermediate conformation (Chapter Three). The transition between these conformations is controlled by two pH-dependent stability titrations. The first, at low pH, is between the unfolded ensemble and the intermediate conformation and is possibly driven by the deprotonation of a carboxylate group. The second, higher pH titration drives the folding from the intermediate to the native conformation. Remarkably, it was possible to simulate the second titration using the known individual His $pK_a$ values for the native conformation and an assumed average

$pK_a$ for those in the intermediate form. Each His contributed, either positively or negatively, to the ionization energy. However, H87, which contributes over half of this energy, is conserved for AvrPto and AvrPtoB homologues (Figure 4.8A), highlighting its importance in the folding process.

A full treatment of the thermodynamic theory behind pH-dependent stability was beyond the scope of Chapter Three. However, in order to predict how the mutation of His residues would alter the pH-dependence, a more thorough discussion of the theory is warranted.

The pH-dependence of stability depends on the relative ability of each conformation involved in a titration to accommodate the different ionization states (Yang et al. 1993). These relative abilities are reflected by the $pK_a$ values for titratable groups in each conformation. The high pH titration of TrAvrPto is controlled by His $pK_a$ values in the native, $pK_{a,N}$, and intermediate, $pK_{a,I}$, forms. TrAvrPto has four possible states that contribute to its stability: the native conformation with protonated His ($His^+$), native with singly deprotonated His ($His^0$), intermediate with $His^+$, and intermediate with $His^0$. The ionization energy associated with each titration can be separated into pH-independent and pH-dependent components (Yang et al. 1993). For the high pH titration,

$$\Delta G^{high}(pH) = \Delta G_{indep} + \Delta\Delta G^{high}{}_{ion}(pH). \qquad \text{Eq. 1}$$

$\Delta G_{indep}$ is the pH-independent component, containing contributions from any interactions that do not depend on the pH of solution. $\Delta\Delta G^{high}{}_{ion}(pH)$ measures the relative abilities of native and intermediate TrAvrPto to accommodate the ionizations of the protein's seven His sidechains:

$$\Delta\Delta G^{high}{}_{ion}(pH) = \Delta G_{ion}(Native) - \Delta G_{ion}(Intermediate). \qquad \text{Eq. 2}$$

At physiological pH, each His is in one of two ionization states: neutral ($His^0$) or positively-charged ($His^+$). Since there are N=7 His in TrAvrPto, there are a total of

$2^N=2^7=128$ possible ionization states for the protein. For example, in one possible state, H54 is protonated and the other six His are neutral. For the native conformation, the ionization energy for each state *n* depends on the pH of the solution and the $pK_{a,N}^{i}$ values of each His, *i*:

$$\Delta G^n(Native) = \sum_{i=1}^{N=7}\left[\delta_n(i) * \gamma(i) * \ln(10) * RT * \left(pH - pK_{a,N}^{i}\right) + \sum_{1\le j<i}^{N=7}\delta_n(i) * \delta_n(j) * \Delta G^{ij}\right]. \quad \text{Eq. 3}$$

The first term within the brackets of Eq. 3 describes the ionization energy for each of the *i*=1,…,7 His, independent of any interaction with other titratible groups. The second term characterizes any possible interaction between the ionizing His during the titration and was assumed to be negligible during the simulations presented in this chapter. This assumption is supported by the highly localized chemical shift changes with pH titration that are limited to regions surrounding individual His residues in native TrAvrPto (Chapter Three). Moreover, as discussed below, the simplified form of Eq. 3 fully explained the pH-dependent stability during the high pH titration using His sidechain titration data. The vector $\delta_n$ specifies whether each His is charged or neutral in state *n*. When His *i* is neutral, $\delta_n(i)=0$, and when it is charged, $\delta_n(i)=+1$. Therefore, for the state in which H54 is the only charged group, $\delta_n=(\delta_n(H41), \delta_n(H54), \delta_n(H75), \delta_n(H87), \delta_n(H103), \delta_n(H125), \delta_n(H130))=(0,1,0,0,0,0,0)$. The term $\gamma$ distinguishes between acid ($\gamma=-1$) and base ($\gamma=+1$) titration reactions. Since only the His base titration reactions are considered here, $\gamma=+1$ for all titrating groups. These ionization energies (Eq. 3) were used to calculate the Boltzmann distribution of ionization states within the native protein:

$$Z_{native} = \sum_{n=1}^{2^N}\exp\left\{-\Delta G^n(Native)/RT\right\}. \quad \text{Eq. 4}$$

Finally, the free energy of ionization for the native conformation (left term in Eq. 2) is:

$$\Delta G_{ion}(Native) = -RT\ln(Z_{native}). \quad \text{Eq. 5}$$

If the N=7 TrAvrPto His are assumed to not interact during their titrations, Eq. 3 simplifies to

$$\Delta G^{n}(Native) = \sum_{i=1}^{N=7}\left[\delta_{n}(i) * \ln(10) * RT * \left(pH - pK^{i}_{a,N}\right)\right]. \qquad \text{Eq. 6}$$

Under this condition, Eq. 5 is dramatically simplified (see *Supplementary Material*) to the sum of the ionization energies for each individual His.

$$\begin{aligned}\Delta G_{ion}(Native) &= \sum_{i=1}^{N=7}\Delta G^{i}_{ion}(Native) \\ &= -RT\sum_{i=1}^{N=7}\ln\left[1+\exp\left\{-\ln(10) * \left(pH - pK^{i}_{a,N}\right)\right\}\right]\end{aligned} \qquad \text{Eq. 7}$$

Similarly, the free energy of ionization for intermediate TrAvrPto depends on the pH and $pK_a$ values of His within the intermediate, $pK_{a,I}$.

$$\begin{aligned}\Delta G_{ion}(Intermediate) &= \sum_{i=1}^{N=7}\Delta G^{i}_{ion}(Intermediate) \\ &= -RT\sum_{i=1}^{N=7}\ln\left[1+\exp\left\{-\ln(10) * \left(pH - pK^{i}_{a,I}\right)\right\}\right]\end{aligned} \qquad \text{Eq. 8}$$

$\Delta G_{ion}$(Native) and $\Delta G_{ion}$(Intermediate), therefore, depend upon any environmental conditions that affect the $pK_a$ values.

The pH-dependence of the high pH titration, $\Delta\Delta G^{high}{}_{ion}$(pH), can therefore be simplified to the sum of the ionization energies of each individual His.

$$\begin{aligned}\Delta\Delta G^{high}{}_{ion}(pH) &= \Delta G_{ion}(Native) - \Delta G_{ion}(Intermediate) \\ &= \left(\sum_{i=1}^{N=7}\Delta G^{i}_{ion}(Native)\right) - \left(\sum_{i=1}^{N=7}\Delta G^{i}_{ion}(Intermediate)\right) \\ &= \sum_{i=1}^{N=7}\left[\Delta G^{i}_{ion}(Native) - \Delta G^{i}_{ion}(Intermediate)\right] \\ &= \sum_{i=1}^{N=7}\Delta\Delta G^{high,i}{}_{ion}(pH)\end{aligned} \qquad \text{Eq. 9}$$

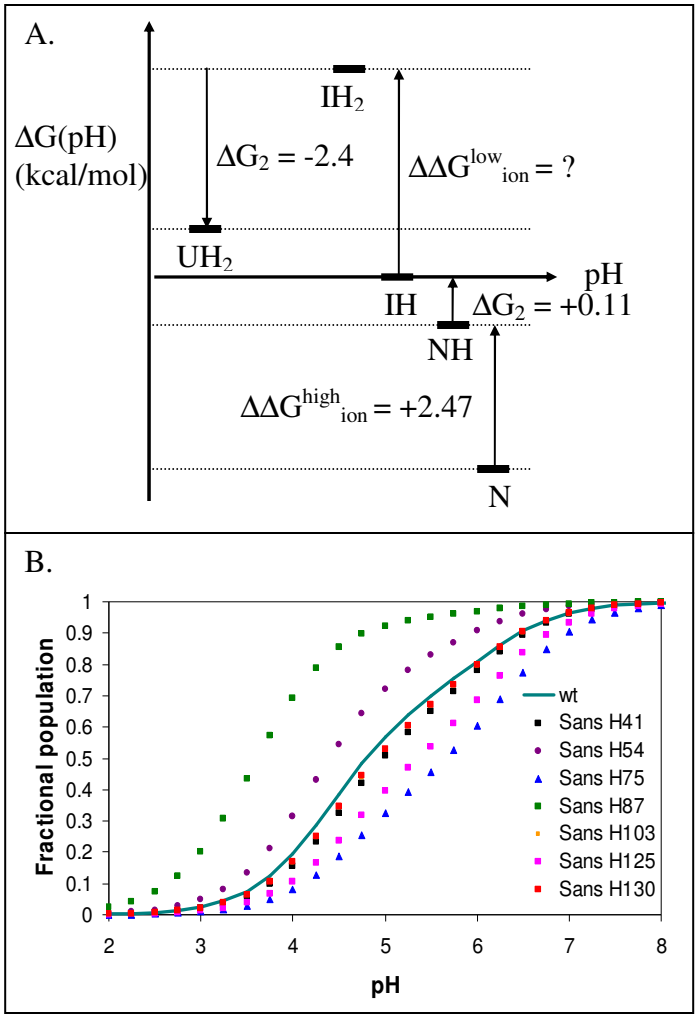



Figure 4.9: **Predicting the effects of removal of His on pH-dependence of stability.** **A.** The approximate free energies between each of the five states $UH_2$, $IH_2$, IH, NH, and N for wild-type TrAvrPto. It is assumed that a His substitution would affect only the high pH titration, i.e. only the free energy of NH is affected. B. The fractional ($p_{NH}+p_N$) populations fitting the NMR data (Chapter Three) and the predicted ($p_{NH}+p_N$) curves for mutants, each substituting one of the seven His for another residue.

The $\Delta\Delta G^{high,i}_{ion}(pH)$ for each His and the total $\Delta\Delta G^{high}_{ion}(pH)$ ionization curves were calculated assuming an average $<pK_{a,I}>=6.4$ for His in the intermediate conformation (Chapter Three). Most of the destabilization energy was due to the protonation of H54 and H87 (Figure 3.5). The other five His contributed only minor changes. Since H54 and H87 are on opposite ends of TrAvrPto (Figure 4.1) and since all other His contribute little to the high pH titration, the assumption that the titrating His do not interact and, therefore, that the second term of Eq. 3 could be neglected, was reasonable.

TrAvrPto stability can be modeled by a five-state, two-titration model (Chapter Three):

$$UH_2 \overset{K_{eq1}}{\rightleftharpoons} IH_2 \overset{K_{stab1}}{\rightleftharpoons} IH + H^+ \overset{K_{eq2}}{\rightleftharpoons} NH + H+ \overset{K_{stab2}}{\rightleftharpoons} N + 2H^+ \qquad \text{Eq.10}$$

The low pH titration is between the doubly-protonated unfolded ensemble ($UH_2$) and the protonated intermediate form (IH), which involves both a change in conformation ($K_{eq1}=p_{IH2}/p_{UH2}$) and a deprotonation ($K_{stab1}=p_{IH}*[H^+]/p_{IH2}$) event that are treated explicitly as separate steps. The midpoint of stability for this titration occurs at $pH=pK_{stab1}$. The high pH titration is between IH and N, where $K_{eq2}=p_{NH}/p_{IH}$ and $K_{stab2}=p_N*[H^+]/p_{NH}$. For wt TrAvrPto (Table 3.2),

$$K_{eq1}=0.016\pm0.002$$

$$pK_{stab1}=3.10\pm0.07$$

$$K_{eq2}=1.2\pm0.4$$

$$pK_{stab2}=5.9\pm0.1.$$

These values can be used to estimate the relative energies of the conformations. Since the conformational change and protonation are considered separately, the total change of free energy during the low pH titration between IH and $UH_2$ is the sum of $\Delta\Delta G^{low}_{ion}$ and $\Delta G_1$. $\Delta\Delta G^{low}_{ion}$ is the total change of free energy

during the $IH \rightarrow IH_2$ protonation and $\Delta G_1$ is the difference in free energy during the conformation exchange between $IH_2$ and $UH_2$ (Figure 4.9A).

$$\Delta G_1 = -RT*\ln(p_{UH2}/p_{IH2}) = +RT*\ln(K_{eq1}) = -2.4 \text{ kcal/mol} \quad \text{Eq. 11}$$

Similarly, for the high pH titration,

$$\Delta\Delta G_{high} = \Delta\Delta G^{high}_{ion} + \Delta G_2 = \Delta\Delta G^{high}_{ion} + RT\ln(K_{eq2})$$
$$= \Delta\Delta G^{high}_{ion} + 0.11 \text{ kcal/mol}, \quad \text{Eq. 12}$$

where $\Delta\Delta G^{high}_{ion}$ is the total change of free energy during the $N \rightarrow NH$ protonation and $\Delta G_2$ is the free energy of the conformational exchange.

Simulating the effects of His protonation using Eq. 1-9, the free energy is changed by $\Delta\Delta G^{high}_{ion}$=+2.47 kcal/mol (Figure 3.5). Therefore, the total change in energy due to the high pH titration is $\Delta\Delta G_{high}$=+2.47 kcal/mol + 0.11 kcal/mol =+2.6 kcal/mol.

How would the pH-dependence of TrAvrPto stability be affected by substituting a His with a non-titrating group? Assume that such a mutation would not change the low pH titration, leaving the energies for $UH_2$, $IH_2$, and IH unchanged (Figure 4.9A). The energy for N, the native conformation in which all His are deprotonated, is also assumed not to be affected by the mutation. Therefore, only the energy of NH is changed and the overall stability of the native conformation N is unchanged under these conditions. $\Delta G_1$ and $\Delta\Delta G^{low}_{ion}$ are invariant upon His mutation and the free energy difference between N and $UH_2$ is also constant. Both $\Delta\Delta G^{high}_{ion}$ and $\Delta G_2$ can be altered by a His substitution. However, since the energies of $UH_2$, $IH_2$, IH, and N are constant, the total energy for the high pH titration $\Delta\Delta G_{high}=\Delta\Delta G^{high}_{ion}+\Delta G_2$=+2.6 kcal/mol must also be constant. The new ionization energy curve can be simulated for the high pH titration in the same manner as for wt TrAvrPto using Eq. 1-9, with the mutated His residue's contribution excluded from the calculation. The new equilibrium between NH and IH is characterized by

$$\Delta G_2^{new} = +2.6 \text{ kcal/mol} - \Delta\Delta G^{high}_{ion}(new) \quad \text{Eq. 13}$$

and $$K_{eq2}^{new} = \exp\{+\Delta G_2^{new}/RT\}. \quad \text{Eq. 14}$$

Table 4.1: **Effects of mutating individual His on the pH-dependence of stability**

| Altered residue | $\Delta G_2^{new}$ (kcal/mol) | $K_{eq2}^{new}$ | $pK_{stab2}^{new}$ | $\Delta\Delta G^{high}_{ion}$(new) (kcal/mol) |
|---|---|---|---|---|
| H41 | +0.28 | 1.61 | 5.8-5.98 | +2.32 |
| H54 | +0.81 | 3.94 | 5.8-5.9 | +1.79 |
| H75 | -0.16 | 0.77 | 5.9-6.06 | +2.76 |
| H87 | +1.77 | 19.8 | 6.02-6.2 | +0.83 |
| H103 | +0.32 | 1.72 | 5.8-5.98 | +2.28 |
| H125 | +0.02 | 1.04 | 5.86-6.02 | +2.58 |
| H130 | +0.34 | 1.76 | 5.8-5.98 | +2.26 |

H54 and H87 have the lowest $pK_{a,N}$ (Table 3.2), so it is unsurprising that mutating these residues should have the greatest effect on the high pH titration, which is centered in this region of pH (Table 4.1). The intensity of the "folded" peaks in NMR spectra are proportional to the sum of the deprotonated and protonated native populations, ($p_{NH}$+$p_N$). The relative populations can be calculated in terms of $K_{eq1}$, $pK_{stab1}$, $K_{eq2}^{new}$, and $pK_{stab2}^{new}$ using equations 2-7 in Chapter Three. Replacement of H87 with an amino acid that does not titrate in this pH range should remove most of the pH-dependence between pH 5-8 (Figure 4.9B). Substituting H54 for another residue should also reduce the pH-dependence to a lesser extent. All other His mutants are predicted to shift the pH-dependence of stability slightly towards higher pH.

The predicted stability curves are a valuable resource in designing His mutants, though they should be used with some caution. These calculations were made assuming that neither the overall stability of native TrAvrPto nor the low pH titration was affected by the mutations. However, as demonstrated by the series of H75 mutants designed in the previous section, the overall stability can be very sensitive to the identity of the substituting residue. Mutating H75 may dramatically increase the

pH-dependence, but the H75L and H75M mutants might also stabilize $\alpha_C$ and the protein enough to inhibit secretion *via* the TTSS. H54 is near the δ+ end of $\alpha_B$, is next to negatively-charged E55, and its imidazole sidechain may form a hydrogen bond with the T76 hydroxyl group. This eliminates both the charged and the nonpolar residues as possible substituents. Of the residues most favored by α-helices (A, L, M, E, R, K, Q), only Q remains a possibility. A H54Q should reduce the pH-dependence of stability and, based on AGADIR predictions (Figure 4.10), should increase the overall stability of TrAvrPto, both changes predicted to reduce the protein's secretion efficiency. The most promising mutagenesis target is H87. It is the largest contributor to the ionization energy and is absolutely conserved within a subset of effector proteins. It is part of the Ω loop, so AGADIR propensities are of limited use in design. Since the H87 sidechain forms a hydrogen-bond to the backbone carbonyl group of S33, any proposed substituent should potentially maintain this tertiary contact. H87Y preserves the aromatic-like character of His, while H87Q retains a more similar shape. If the hydrogen bond is found not to be critical, then a H87F mutant may be fruitful.

The remaining residues (H41, H103, H125, H130) increase the pH-dependence of stability and are less promising mutagenesis candidates. H125 and H130 are spatially close to both each other and the δ- end of $\alpha_D$ within a C-capping box, so it would be difficult to predict the outcome of a mutation to either of these residues. H103 is at the δ+ end of $\alpha_D$ and next to E104, so, again, prediction is difficult.

Still, E104 is a tantalizing candidate as the residue controlling the low pH titration. Its midpoint of stability, $pK_{stab1}$=3.1, is consistent with a titration controlled by the protonation of a carboxylate, i.e. $E104^- + H^+ \rightleftharpoons E104^0$. If this is the case, then within the doubly-protonated unfolded ensemble ($UH_2$), all titratable groups are protonated with His charged ($His^+$) and E104 neutral ($E104^0$). The intermediate form

would have all $His^+$ and $E104^-$ and the native conformation, $His^0$ and $E104^-$. The effects of mutating E104 are difficult to predict since less is known about the low pH titration. Additionally, the local environment of E104 is such that mutating this residue could either destabilize or stabilize the protein. At intermediate pH values, $E104^-$ would compensate for the placement of $H103^+$ at the δ+ end of $\alpha_D$. An E104Q mutation would remove this charge compensation, resulting in the destabilization of $\alpha_D$ and the rest of the protein, and could eliminate the intermediate conformation. Both of these results should shift the acid denaturation to higher pH, even if the overall stability is unchanged. Conversely, being at the δ+ end of $\alpha_D$ without the presence of $E104^-$ could depress H103's $pK_{a,N}$. $H103^0$ is a hydrogen-bond acceptor for N30, which stabilizes the protein at $pH>pK_{a,N}(H103)$. These last two effects should shift the acid denaturation to lower pH values. If E104 indeed controls the low pH titration, some combination of all four effects will likely occur. While the E104Q mutant is unlikely to produce a more stable protein, it may illuminate TrAvrPto folding behavior at low pH values.

An alternate, but less promising, candidate to explore the low pH titration is E55Q. E55 is adjacent to H54, which is near to the δ+ end of $\alpha_B$. However, since H54 *has* a depressed $pK_{a,N}$, the presence of E55 might not be critical to protein stability.

***Literature review and the dangers of scientific parallax***

H54P AvrPto was hypothesized to be translocation-incompetent (Shan et al. 2000). When injected into *N. benthamiana* by *P. syringae via* the TTSS, the mutant failed to trigger the hypersensitive response (HR), a cellular defense that is observed on a macroscopic level as localized death of leaf tissue. *P. syringae* was able to express and maintain H54P AvrPto within its cytoplasm. Additionally, H54P

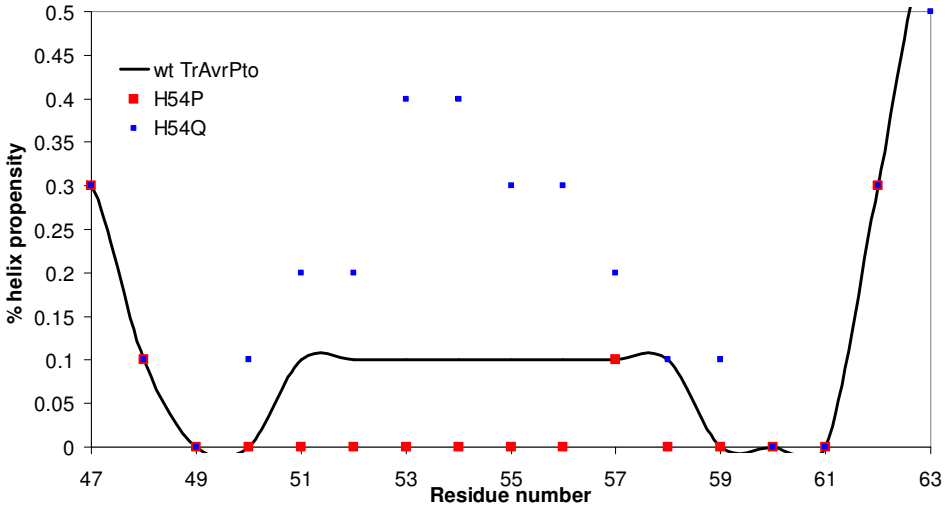



Figure 4.10: **H54 mutants.** The H54P mutation removes all predicted α-helix propensity in $\alpha_B$. H54Q improves the propensity.

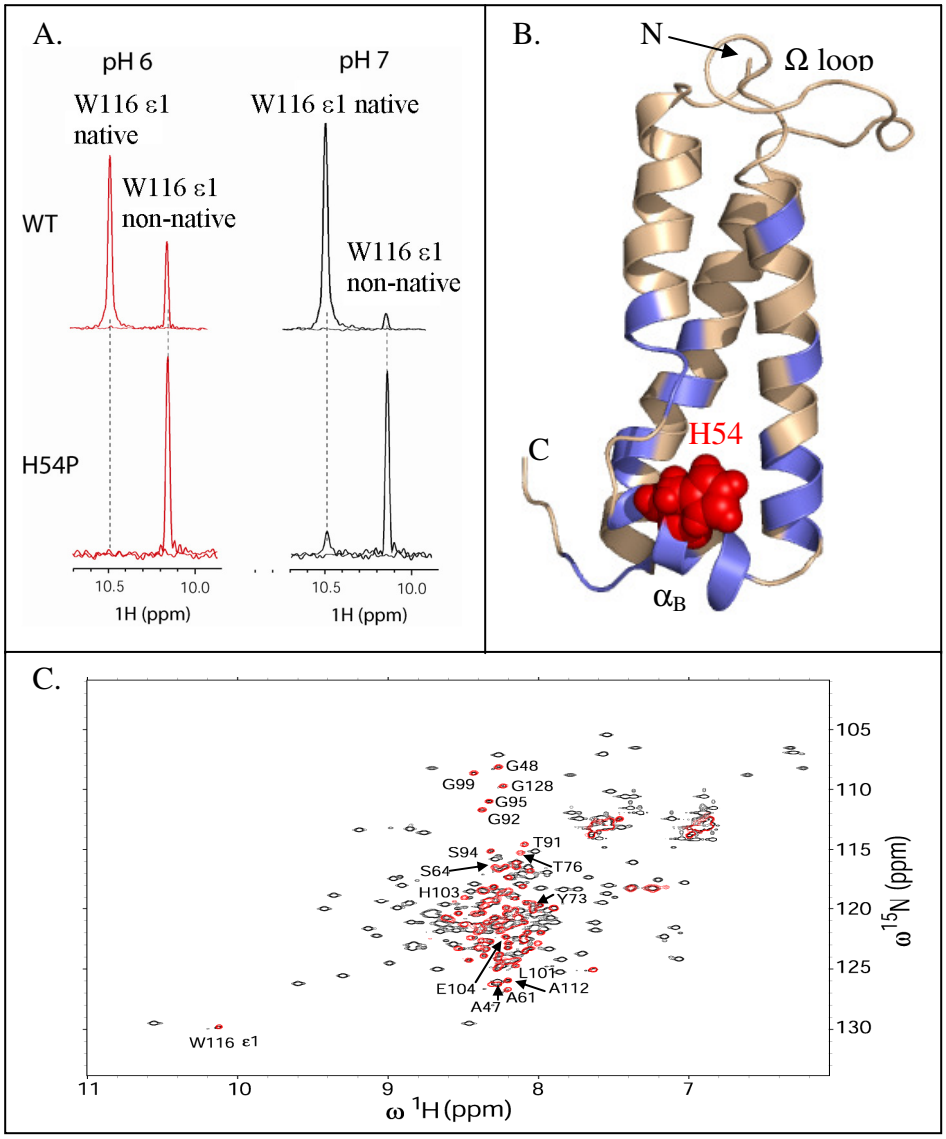



Figure 4.11: **The effects of the H54P mutation on TrAvrPto.** A. H54P TrAvrPto is less stable than wt TrAvrPto, but retains pH-dependence of stability. For both proteins, the one-dimensional lineshapes of the W116 ε1 peaks were extracted from $^{15}$N-$^{1}$H HSQC spectra using NMRDraw (Delaglio et al. 1995) and were overlaid using Adobe Illustrator. B. The changes due to H54P substitution cluster near $\alpha_B$. Mutant residues whose backbone amide peaks moved more than a linewidth (~20 Hz) from wt shifts are mapped onto the TrAvrPto structure in blue. H54 is shown in red. Image rendered with PyMOL (DeLano. 2002) using NMR structure (PDB file 1R5E) (Wulf et al. 2004), ensemble-averaged *via* MOLMOL (Koradi et al. 1996). C. At pH 6, wt TrAvrPto has both native and non-native HSQC spectral peaks (black). Only the non-native peaks are observed in a pH 6 H54P TrAvrPto HSQC (red). The seventeen best-resolved non-native peaks overlay in both spectra.

interacted with Pto during yeast 2-hybrid screenings. It should have been able to elicit HR. Indeed, when H54P AvrPto enters the host cell through another injection method, transient expression *via A. tumefaciens*, the mutant does elicit HR. Therefore, it was reasonably hypothesized that H54P AvrPto is unable to be translocated by the TTSS.

The secondary structure predictions and the hydrogen-bonding patterns argue against the H54P mutant being too stable to be secreted[1]. According to AGADIR predictions, the H54P substitution eliminated what little α-helix propensity there was for $\alpha_B$ (Figure 4.10). In wt TrAvrPto, the H54 backbone amide group is hydrogen-bonded to the R51 carbonyl group and the imidazole sidechain may be hydrogen-bonded to the T76's hydroxyl group. A H54P mutant would lose these stabilizing interactions, since Pro has neither a backbone amide group nor a sidechain that can form a hydrogen-bond.

H54P TrAvrPto was found *via* NMR to be less stable than wt TrAvrPto (Figure 4.11A). Wildtype TrAvrPto is mostly in its native form at pH 7. 0 and is only 16% non-native at pH 6.1 (Dawson et al. 2008). The mutant has a minor native population at pH 7, but none at pH 6. Therefore, the H54P mutant is dramatically destabilized, yet still displays acid denaturation.

The regions affected by the H54P substitution can be illuminated by comparing the backbone amide chemical shifts of wt and H54P TrAvrPto in HSQC spectra at pH 7. The native H54P residues whose backbone amide peaks either moved from their wild-type chemical shifts or disappeared are shown mapped onto the TrAvrPto structure in Figure 4.11B. Most of the changes are near $\alpha_B$, leaving the Ω loop unperturbed. The mutation probably destabilized $\alpha_B$, leaving a very long loop from

[1] Expression and purification of H54P TrAvrPto was aided by Sam Schueler (Cornell University) and Aaron Oswald (Ithaca Senior High School).

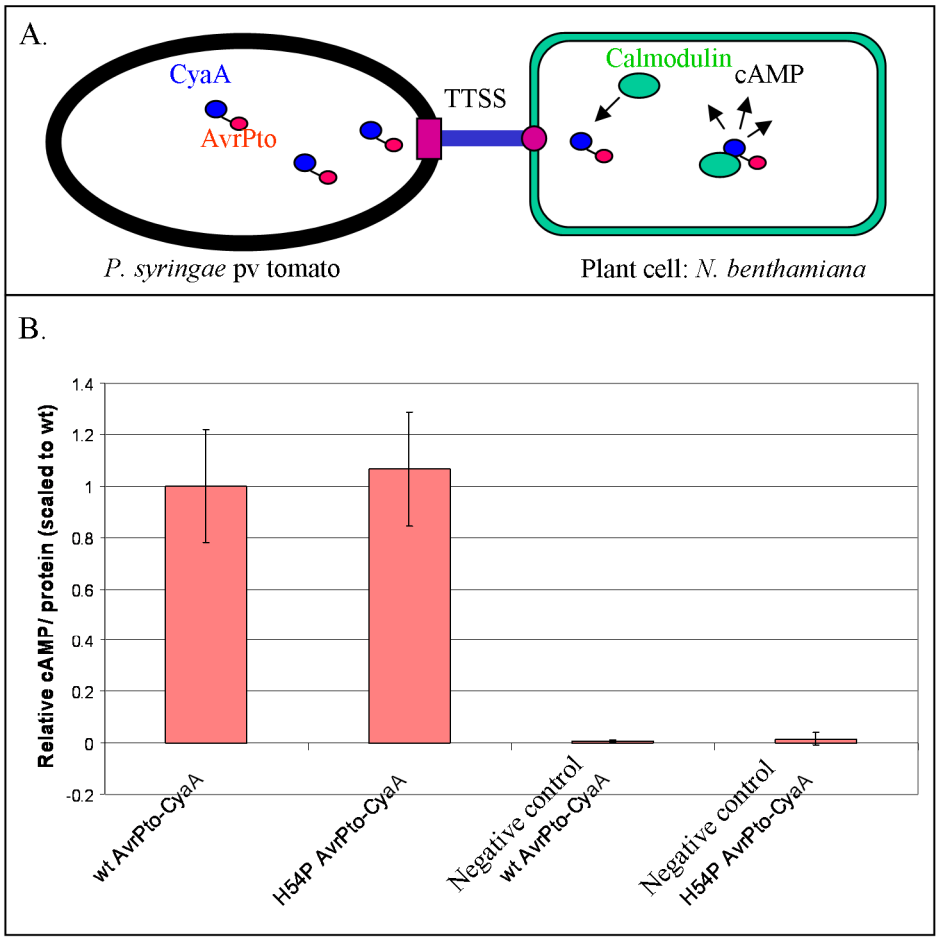



Figure 4.12: **Assay for effector protein translocation**. A. Diagram for the translocation assay. The effector protein is fused to CyaA, an adenyl cyclase, which once inside the host plant cell interacts with calmodulin and produces cAMP. B. Both wt and H54P AvrPto-CyaA are translocated in significant amounts. The nmol cAMP/ µg protein measured by the assay were scaled relative to the wt AvrPto-CyaA measurement.[2]

[2] The translocation assay data was taken by Kristina Govorovska (Cornell University)

A47 to D62. Without packing interactions with $\alpha_B$, the core in turn may have been destabilized. The chemical environment surrounding H87 was not altered, which likely explains the retention of the pH-dependence of stability. While there was no native population observed for H54P TrAvrPto at pH 6, the non-native conformations appear to be unperturbed by the mutation (Figure 4.11C). The sets of non-native peaks for wt and H54P TrAvrPto overlay each other for the seventeen best-resolved non-native peaks (A47, G48, A61, S64, Y73, T76, T91, G92, S94, G95, G99, L101, H103, E104, A112, W116 ε1, and G128), indicating that the local chemical environments of the non-native conformations are not altered by the H54P substitution. This indicates that the H54P mutation has trapped the intermediate form and provides an opportunity for future detailed biophysical studies of the intermediate.

In contrast to what was reported in the literature (Shan et al 2000), studies in the Nicholson laboratory have shown that H54P AvrPto can be translocated. These studies were carried out in collaboration with rotation student Kristina Govorovska and the Collmer laboratory (Cornell University) and utilized a translocation assay based on the adenyl cyclase enzyme, CyaA (Schechter et al. 2004). This assay is a very sensitive way of detecting small amounts of effector protein that have been translocated from the pathogen (*P. syringae* pv tomato) through the TTSS into the host cell's cytoplasm (*N. benthamiana,* domestic tobacco leaf cells). The first 400 amino acids of CyaA form a domain that converts ATP to cAMP only when in the presence of calmodulin and is used as a reporter for translocation (Ladant et al. 1999). CyaA can be fused to the C-terminus of AvrPto or any of the AvrPto mutants using DNA recombination (*Experimental Methods*). *P. syringae* cells containing the fusion protein are inoculated into tobacco leaves. The pathogen translocates the fusion proteins into the host cell *via* the TTSS (Figure 4.12A). Once inside the host, the fusion protein encounters calmodulin, a protein that interacts with CyaA and is only

found in eukaryotic cells. This interaction triggers a cAMP cascade. The presence of elevated cAMP, as measured seven hours after inoculation, indicates that the TTSS can translocate both AvrPto-CyaA and H54P AvrPto-CyaA in significant amounts (Figure 4.12B).

As was the case for the stability data measured *via* CD, Trp fluorescence, and NMR spectroscopy (Chapter Three), similar assays may yield different results. This kind of scientific parallax was encountered when comparing the results of the translocation assay used in this research, which depends only on the fusion protein's ability to be translocated, and that used in the literature (Shan et al. 2000), which required that the effector be translocated *and* cause a macroscopic HR response. Native H54P AvrPto still has a folded Ω loop, which is critical for interactions with Pto, as observed in the yeast 2-hybrid screenings. However, only a small percentage of H54P AvrPto is in its native functional form, even at pH 7.0. The most plausible explanation for the discrepancy between of the two translocation assays' results is that there was not enough functional H54P AvrPto delivered into the host cell by the TTSS to elicit an HR observable on the macroscopic scale. Then, when H54P AvrPto was overexpressed during transient expression, the amount of functional protein was raised to high enough levels that HR could be observed.

***Summary of the rationally designed AvrPto mutants***

A number of promising mutants have been designed with the purpose of improving our understanding of the thermodynamics and kinetics of TrAvrPto folding and how these properties relate to secretion competency *via* the TTSS. The design relied on a number of sources, from secondary structure propensities and prediction programs to thermodynamic theory to knowledge of the wild-type protein's behavior. The most interesting candidates are summarized in Table 4.2. While these mutants

should be efficacious in theory, the consequences of each mutation can only be determined by experimental testing, as demonstrated for the case of H54P AvrPto.

Table 4.2: **The most promising rationally designed TrAvrPto mutants**

| Element | Proposed mutants |
|---|---|
| *Mutants designed to improve the overall TrAvrPto stability*[a] | |
| $\alpha_A$ | Q35E, S34M > E45R > E45K > E45Q |
| Beginning of $\alpha_C$ | N67E > N67D > Q63E |
| Middle of $\alpha_C$ | H75L, H54M > H75A, H75E |
| Ω loop | G95P |
| $\alpha_D$ | Q124L > Q124A > T114K, Q124M |
| *Mutants designed to reduce the pH-dependence of stability* | |
| $\alpha_B$ | H54Q |
| Ω loop | H87Y, H87Q, H87F |
| *Mutants designed to explore the low pH titration* | |
| $\alpha_D$ | E104Q, E55Q |

[a] *Listed in decreasing order of improved predicted secondary structural propensity*

***Experimental Methods***

A. Site-directed mutagenesis

The H54P substitution was made using Stratagene's QuikChange® Mutagenesis kit and two primers, forward and reverse, which contain the codons for the new residue (Table 4.3). The kit employs a technique that is a variation on PCR, producing amplified amounts of plasmid containing the desired mutant. For the TrAvrPto mutants, the substitution was done on a 5254 basepair pQE9 plasmid containing an N-terminal 6-His tag, a 3C-Pro cut site, and residues 29-133 of AvrPto (the original plasmid was a generous gift From Pete Pascuzzi and Greg Martin, Cornell University). The 2601 basepair pENTR vector containing full-length AvrPto (vector pCPP5154) was mutated to produce H54P AvrPto for the translocation assay. The protocol for site-directed mutagenesis followed that given with the kit. Afterwards, the plasmids were sequenced to ensure that the correct mutations had been made (Applied

Biosystems ABI 3700 automated DNA sequencer, Cornell Bioresource Center). The pENTR vectors were sequenced using M13F and M13R primers supplied by the Cornell Bioresource Center. The TrAvrPto constructs were transformed *via* heat shock into DH5α *E. coli* for DNA storage and M15 *E. coli* for protein expression. Heat shock was also used to transform the pENTR AvrPto-CyaA constructs into DH5α cells.

Table 4.3: **Forward and reverse primer for mutagenesis and sequencing**

| | Sequence | Direction |
|---|---|---|
| *Site-directed Mutagenesis* [a,b] | | |
| H54P | 5′-GTCTACCAAGAGATCAGCCTGAATTTGTTAGTAGCC-3′ | Forward |
| | 5′-GGCTACTAACAAATTCAGGCTGATCTCTTGGTAGAC-3′ | Reverse |
| G95P [c] | 5′-GGTACATGACGGGAGCCTCA**CCA**ATCAATCCGGGAATGCTG-3′ | Forward |
| | 5′-CAGCATTCCCGGATTGAT**TGG**TGAGGCTCCCGTCATGTACC-3′ | Reverse |
| H87Y | 5′-GGCGGACATGCAGTATAGGTACATGACGGGAGCG-3′ | Forward |
| | 5′-CGCTCCCGTCATGTACCTATACTGCATGTCCGCC-3′ | Reverse |
| H54Q | 5′-GTCTACCAAGAGATCAGCAGGAATTTGTTAGTAGCCAAG-3′ | Forward |
| | 5′-CTTGGCTACTAACAAATTCCTGCTGATCTCTTGGTAGAC-3′ | Reverse |
| Q124A | 5′-CATGCGCGAAGCTCTGGCGCACGCAATGGGTATC-3′ | Forward |
| | 5′-GATACCCATTGCGTGCGCCAGAGCTTCGCGCATG-3′ | Reverse |
| H87Q | 5′-GGCGGACATGCAGCAGAGGTACATGACGGG-3′ | Forward |
| | 5′-CCCGTCATGTACCTCTGCTGCATGTCCGCC-3′ | Reverse |
| H87F | 5′-GGCGGACATGCAGTTTAGGTACATGACGGG-3′ | Forward |
| | 5′-CCCGTCATGTACCTAAACTGCATGTCCGCC-3′ | Reverse |
| H41Q | 5′-CTGAGCGTCAGACAGCAACTTGCGGAGTCTG-3′ | Forward |
| | 5′-CAGACTCCGCAAGTTGCTGTCTGACGCTCAG-3′ | Reverse |
| *DNA Sequencing* | | |
| pQE9 [d] | 5′-CGGATAACAATTTCACACAG-3′ | Forward |
| | 5′-GTTCTGAGGTCATTACTGG-3′ | Reverse |
| pVLT35 [e] | 5′-CTGCGGGGATGCC-3′ | Reverse |

[a] The primers and mutants are based on the wt TrAvrPto nucleotide sequence (UniProtKB/TrEMBL entry Q08242)

[b] The codon for the mutated residue is marked in red.

[c] Nucleotides mutated to prevent possible formation of secondary structure in primer (shown in **green**).

[d] The forward and reverse primers basepair with the Type III/IV and Reverse Sequencing sections of the plasmid, respectively (Anonymous 2003)

[e] Primer basepairs with the beginning of *cyaA* (Brian Kvitko, personal communication).

The H54P, G95P, H87Y, and H54Q mutagenesis primers and the pQE9 and pVLT35 sequencing primers have been constructed (Invitrogen Custom Primers) and are listed in Table 4.3. While the Q124A, H87Q, H87F, and H41Q forward and reverse primers have been designed, they have not yet been synthesized.

Some caution was required with the G95P primers. The primers make a coincidental match with another section of the pQE9 vector, an unavoidable matching of up to 17 nucleotides that include the codon for the new Pro. After performing the G95P mutagenesis, plasmids were transformed into DH5α, plated on selective 100 μg/ml ampicillin LB plates and grown overnight at 37ºC. Taking one colony for each test, several 10 ml LB tester cultures were grown with ampicillin overnight. Glycerol stocks were made, then DNA was extracted (QIAprep® spin miniprep kit, QIAGEN) from the rest of the culture to be sequenced. Some sequenced plasmids had the correct gene; others contained the wild-type gene. A plate was streaked from a glycerol stock with the correct sequence and the test was repeated to check that the colony contained only plasmids with the correct sequence.

A tendency for pQE9 vectors to lose the ability to correctly express 6His-labeled protein was observed. *E. coli* (M15 and B21 strains) bearing these vectors still retain their ampicillin resistance, which is encoded in another section of the pQE9 plasmid. However, either no protein is over-expressed during induction or a protein with greater molecular weight than TrAvrPto (>12 kDa) was observed. This deficiency was observed on different occasions for 6His-TrAvrPto (both wt and mutant constructs) and 6His-AvrPtoB(121-200) and tends to occur after the glycerol stocks have been exposed to multiple freeze-thaw cycles. While the root cause of this problem remains unknown, it is hypothesized that mutations in the plasmid accumulate over time and result in frameshifts that alter the gene for expressed protein

and/or disrupt the stop codon. Any mutations disrupt the ampicillin resistance would be selected against during protein expression.

B. NMR Spectroscopy

$^{15}$N-labeled NMR samples were prepared as described elsewhere (Dawson et al. 2008). As before, the NMR data was collected on a Varian Inova 600 MHz spectrometer at 25ºC and were processed with NMRpipe and NMRDraw (Delaglio et al. 1995). The wt TrAvrPto HSQC spectra at pH 6.1 and 7.0 used in Figure 4.11A are the same spectra included in the characterization of TrAvrPto's pH-dependence in Chapter Three. $^{15}$N-$^{1}$H HSQC spectra (Mulder et al. 1996) were collected at pH 6 and 7 for 0.2 mM H54P TrAvrPto using a delay interval of one second.

C. Construction of AvrPto-CyaA vectors

The vectors for the AvrPto-CyaA fusion proteins were constructed using the recombination of two plasmids. The entry vector (pCCP5154 pENTR SD/D TOPO) contains *avrPto* and kanamycin resistance. The destination vector (pCPP3234 pVLT35) contains genes for CcdB, CyaA, and spectinomycin resistance. These vectors were a generous gift from Hanh Nguyen and Greg Martin, Cornell University, Department of Plant Pathology and Plant-Microbe Biology). The recombination was performed using the Invitrogen Gateway Technologies kits and its recommended protocol. After recombination, there are two new plasmids: one with AvrPto-CyaA and one with CcdB. Heatshock-transformed DH5α cells were plated on selective plates with 30 μg/ml kanamycin and 30 μg/ml spectinomycin. Since the protein CcdB is fatal to *E. coli*, only colonies with AvrPto-CyaA grow. The H54P AvrPto-CyaA fusion protein was constructed in the same manner. The DNA of both plasmids was extracted and then transformed into *P. syringae via* electroporation. Both the wt and H54P AvrPto-CyaA constructs were confirmed by DNA sequencing.

D. Translocation Assay Protocol

Both *P. syringae* strains used in the translocation assay contain the full complement of *hrp* and effector protein genes required to elicit HR or cause disease in addition to the AvrPto-CyaA construct.

**Samples:**

1. DC3000 with AvrPto-CyaA

*Strain*: *P. syringae* pv tomato DC3000 (resistant to ampicillin and rifampicin)

*Vector*: AvrPto-CyaA in pCCP3234 pVLT35(CyaA) plasmid (spectinomycin resistant). The plasmid contains a *lac* promoter, allowing protein expression to be induced with IPTG.

2. 5112 with AvrPto-CyaA (Negative control)

*Strain*: *P. syringae* pv tomato DC3000 CUCPB5112 (resistant to ampicillin, kanamycin, and rifampicin). Its TTSS has a defective outer membrane protein and will secrete into the periplasmic space between the two bacterial membranes.

*Vector*: AvrPto-CyaA in pCCP3234 pVLT35 (CyaA) plasmid (spectinomycin resistant).

3. DC3000 with H54P AvrPto Cya

*Strain*: *P. syringae* pv tomato DC3000 (resistant to ampicillin and rifampicin)

*Vector*: H54P AvrPto-CyaA in pCCP3234 pVLT35(CyaA) plasmid (spectinomycin resistant). H54P AvrPto is a destabilized mutant of wt TrAvrPto

4. 5112 with H54P AvrPto Cya (Negative control)

*Strain*: *P. syringae* pv tomato DC3000 CUCPB5112 (resistant to ampicillin, kanamycin, and rifampicin).

*Vector*: AvrPto-CyaA in pCCP3234 pVLT35(CyaA) plasmid (spectinomycin resistant).

**Protocol:** All steps done at room temperature. For each sample…

I. Sample preparation

1. From an isolated colony, make a primary streak onto King's Broth (KB) (King et al. 1954) plates and grow overnight.
2. Add 200 µl (KB) and spread the primary streak and grow overnight into a lawn.
3. Scrape cells from the lawn with a sterile loop and suspend in 5 ml MES buffer (5mM MES, pH 6.5 + 100uM IPTG) to $O.D._{600} = 0.3$.
4. Dilute 1/10 with the MES buffer

II. Wild or domestic tobacco inoculation

The tobacco plants (generous gifts from the Collmer lab) were grown under green-house conditions and transferred into the lab at least one day before inoculation.

For each sample, inoculate leaves on two different plants. Keep all four samples on the same leaf.

1. Prick leaves with a dissecting needle and then, while covering one end of the hole with one finger, use a tuberculin (needleless) syringe to inoculate. Make certain that liquid goes inside the leaf. Dab the leaf surface with a KimWipe to remove any residual liquid.
2. Mark where inoculation occurred with a wide tip Sharpie
3. Allow 7 hours for inoculation
4. With a 1 cm diameter cork borer, harvest two disks from each inoculation site.
5. Store at -80°C overnight in 1.5 ml microcentrifuge tube.

III. Adenylate cyclase assays

*Leaf preparation*

1. Freeze leaf disks in liquid nitrogen, grind to a powder with Kontes microfuge tube pestle
2. Suspend in 300 µl 0.1M HCl
3. Agitate for 5 mins
4. Vortex to pellet leaf material
5. Dilute 1/100 or 1/1000 (carefully avoid picking up any leaf material during the dilutions

*Adenylate cyclase assay*

The concentrationof cAMP produced in response to the translocation of AvrPto-CyaA fusion proteins into the host cell was quantified using an immunocompetition assay (Direct Cyclic AMP Enzyme Immunoassay kit, Assay Designs) and its recommended protocol. Four leaf samples were analyzed for wt and H54P AvrPto-CyaA.

*Bradford assay*

The quantification of protein concentration was run in parallel with the cAMP assay. In a 96 well plate, set up the following standards in duplicate:

| 100ug/ml BSA(µl) / dH20 (µl) | [final conc] |
|---|---|
| 0/200 | [0.0 µg/ml] |
| 2/198 | [0.8 µg/ml] |
| 4/196 | [1.6 µg/ml] |
| 8/192 | [3.2 µg/ml] |
| 12/188 | [4.8 µg/ml] |
| 16/184 | [6.4 µg/ml] |
| 20/180 | [8.0 µg/ml] |

Prepare duplicates of samples: 5 µl of 1X supernatant 195µl of dH20.

Add 50 µl of 5X Bradford's reagent to each well.

Mix wells and let the plate sit for 5 minutes (pop any bubbles).

Read absorbance of Bradford assay samples at 595 nm.

***Supplementary Material***

*Simplifying the partition function and ionization energy equations*

At physiological pH, a His sidechain can occupy one of two states: neutral ($His^0$) or positively charged ($His^+$). For a protein with *N* His sidechains, there are $2^N$ possible ionization states. The energy of each state $n=1,\ldots,2^N$ is

$$\Delta G^n = \sum_{i=1}^{N}\left[\delta_n(i) * \gamma(i) * \ln(10) * (pH - pK_a{}^i) + \sum_{1\le j<i}^{N}\delta_n(i) * \delta_n(j) * \Delta G^{ij}\right], \quad \text{Eq. A1}$$

where the first term is the energy for the state *n* independent of any other His and the second term is the interaction energy between His (Yang et al 1993). In the simulations presented in this chapter, the interaction energies were assumed to be negligible. The vector $\delta_n$ is of length *N*. Each term of the vector, $\delta_n(i)$, specifies the ionization state of His *i* as either neutral ($\delta_n(i)=0$) or charged ($\delta_n(i)=+1$). The ionization of His were modeled as base titration reactions ($His^0 + H^+ \rightleftharpoons His^+$). Therefore, $\gamma(i)=+1$ for all His. Under the simulation conditions presented here, Eq. A1 simplifies to

$$\Delta G^n = \sum_{i=1}^{N}\delta_n(i) * \ln(10) * (pH - pK_a{}^i). \quad \text{Eq. A2}$$

The Boltzmann distribution and the ionization energy depend on the energies of all $2^N$ states:

$$Z = \sum_{n=1}^{2^N}\exp\{-\Delta G^n / RT\} \quad \text{Eq. A3}$$

$$\Delta G_{ion} = -RT\ln(Z). \quad \text{Eq. A4}$$

Can Eq. A4 be simplified? Consider a protein with $N$=2 His: His1 and His2. In this system, there are $2^N = 4$ possible ionization states.

State $n$=1: $His1^0$, $His2^0$ (both His are neutral)

$\delta^1 = (0,0)$

$\Delta G^1 = 0 + 0$ (see Eq. A2)

State $n$=2: $His1^0$, $His2^+$

$\delta^2 = (0,1)$

$\Delta G^2 = 0 + \ln(10)*RT(pH-pK_a^{His2})$

State $n$=3: $His1^+$, $His2^0$

$\delta^3 = (1,0)$

$\Delta G^3 = \ln(10)*RT(pH-pK_a^{His1}) + 0$

State $n$=4: $His1^+$, $His2^+$

$\delta^4 = (1,1)$

$\Delta G^4 = \ln(10)*RT(pH-pK_a^{His1}) + \ln(10)*RT(pH-pK_a^{His2})$

Using Eq. A3, the partition function is found.

$$Z = 1 + \exp\{-\ln(10)*(pH-pK_a^{His2})\} + \exp\{-\ln(10)*(pH-pK_a^{His1})\} + \ldots$$
$$\ldots + \exp\{-\ln(10)*(pH-pK_a^{His1}+pH-pK_a^{His2})\} \quad \text{Eq. A5}$$

The fourth term of Eq. A5 can be factored into the product of two exponents.

$$Z = 1 + \exp\{-\ln(10)*(pH-pK_a^{His2})\} + \exp\{-\ln(10)*(pH-pK_a^{His1})\} + \ldots$$
$$\ldots + \exp\{-\ln(10)*(pH-pK_a^{His1})\}* \exp\{-\ln(10)*(pH-pK_a^{His2})\} \quad \text{Eq. A6}$$

Equation A6 is equivalent to

$$Z = [1 + \exp\{-\ln(10)*(pH-pK_a^{His1})\}]* [1 + \exp\{-\ln(10)*(pH-pK_a^{His2})\}]. \quad \text{Eq. A7}$$

Since the natural log of the partition function is taken to find the ionization energy, $\Delta G_{ion}$ (Eq. A4), the contributions of each His to the energy can be separated into individual terms.

$$\Delta G_{ion} = -RT\ln[1 + \exp\{-\ln(10)*(pH-pK_a^{His1})\}] - \ldots$$
$$\ldots -RT\ln[1 + \exp\{-\ln(10)*(pH-pK_a^{His2})\}]. \qquad \text{Eq. A8}$$

In general, when the interaction energy (Eq. A1) is neglected, the ionization energy equation (Eq. A4) is simplified to the sum of the contributions from each of the *N* titratible groups.

***Acknowledgements***

This section of my thesis project relied heavily on the generous donations of time and material from many people. Pete Pascuzzi, Hahn Nguyen, and Professor Gregory Martin provided both the TrAvrPto and CyaA fusion protein construction vectors. Brian Kvitko, Joanne Morello, and Professor Alan Collmer gave their expertise, materials, and patience. Sam Schueler and Aaron Oswald assisted in protein purification and Kristina Govorovska collected the translocation assay data. To each these people, I offer my gratitude. This work was supported by NSF grant MCB-0641582 and NIH Molecular Biophysics Training Grant T32 GM08267.

# REFERENCES


Anonymous 2003. *The QIAexpressionist: A handbook for high-level expression and purification of 6xHis-tagged proteins.* In 5th ed. (eds. Anonymous ), 118. QIAGEN,

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.

Aurora, R. and Rose, G.D. 1998. Helix capping. *Protein Sci.* **7:** 21-38.

Bernal, A.J., Pan, Q., Pollack, J., Rose, L., Kozik, A., Willits, N., Luo, Y., Guittet, M., Kochetkova, E. and Michelmore, R.W. 2005. Functional analysis of the plant disease resistance gene Pto using DNA shuffling. *J.Biol.Chem.* **280:** 23073-23083.

Bezsonova, I., Korzhnev, D.M., Prosser, R.S., Forman-Kay, J.D. and Kay, L.E. 2006. Hydration and packing along the folding pathway of SH3 domains by pressure-dependent NMR. *Biochemistry* **45:** 4711-4719.

Bystroff, C. and Shao, Y. 2002. Fully automated ab initio protein structure prediction using I-SITES, HMMSTR and ROSETTA. *Bioinformatics* **18 Suppl 1:** S54-61.

Bystroff, C., Thorsson, V. and Baker, D. 2000. HMMSTR: a hidden Markov model for local sequence-structure correlations in proteins. *J.Mol.Biol.* **301:** 173-190.

Chang, J.H., Tobias, C.M., Staskawicz, B.J. and Michelmore, R.W. 2001. Functional studies of the bacterial avirulence protein AvrPto by mutational analysis. *Mol. Plant-Microbe Interact.* **14:** 451-459.

Cuff, J.A. and Barton, G.J. 2000. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins: Struct. Funct. Bioinform.* **40:** 502-511.

Dawson, J.E. and Nicholson, L.K. 2008. Folding kinetics and thermodynamics of Pseudomonas syringae effector protein AvrPto provide insight into translocation via the type III secretion system. *Protein Sci.* **17:** 1109-1119.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6:** 277-293.

DeLano, W.L. 2002. *The PyMOL molecular graphics system,* DeLano Scientific, San Carlos, CA, USA.

Guruprasad, K. and Rajkumar, S. 2000. Beta-and gamma-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials. *J. Biosci. (Bangalore)* **25:** 143-156.

Johnson, S., Deane, J.E. and Lea, S.M. 2005. The type III needle and the damage done. *Curr. Opin. Struct. Biol.* **15:** 700-707.

Kim, Y.J., Lin, N.C. and Martin, G.B. 2002. Two distinct Pseudomonas effector proteins interact with the Pto kinase and activate plant immunity. *Cell* **109:** 589-598.

King, E.O., Ward, M.K. and Ranay, D.E. 1954. Two simple media for the demonstration of pyocyanin and fluorescin. *J.Lab.Clin.Med.* **44:** 301-307.

Koradi, R., Billeter, M. and Wüthrich, K. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14:** 51-55.

Lacroix, E., Viguera, A.R. and Serrano, L. 1998. Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284:** 173-191.

Ladant, D. and Ullmann, A. 1999. Bordatella pertussis adenylate cyclase: a toxin with multiple talents. *Trends Microbiol.* **7:** 172-176.

Lee, V.T. and Schneewind, O. 2002. Yop fusions to tightly folded protein domains and their effects on Yersinia enterocolitica type III secretion. *J. Bacteriol.* **184:** 3740-3745.

Mulder, F.A.A., Spronk, C.A.E.A., Slijper, M., Kaptein, R. and Boelens, R. 1996. Improved HSQC experiments for the observation of exchange broadened signals. *J. Biomol. NMR* **8:** 223-228.

Muñoz, V. and Serrano, L. 1997. Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* **41:** 495-509.

Muñoz, V. and Serrano, L. 1995a. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *J. Mol. Biol.* **245:** 297-308.

Muñoz, V. and Serrano, L. 1995b. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* **245:** 275-296.

Schechter, L.M., Roberts, K.A., Jamir, Y., Alfano, J.R. and Collmer, A. 2004. Pseudomonas syringae type III secretion system targeting signals and novel effectors studied with a Cya translocation reporter. *J.Bacteriol.* **186:** 543-555.

Shan, L., He, P., Zhou, J.M. and Tang, X. 2000. A cluster of mutations disrupt the avirulence but not the virulence function of AvrPto. *Mol. Plant-Microbe Interact.* **13:** 592-598.

Sorg, J.A., Miller, N.C., Marketon, M.M. and Schneewind, O. 2005. Rejection of impassable substrates by Yersinia type III secretion machines. *J. Bacteriol.* **187:** 7090-7102.

Taddei, N., Chiti, F., Fiaschi, T., Bucciantini, M., Capanni, C., Stefani, M., Serrano, L., Dobson, C. and Ramponi, G. 2000. Stabilisation of alpha-helices by site-directed mutagenesis reveals the importance of secondary structure in the transition state for acylphosphatase folding. *J. Mol. Biol.* **300:** 633-647.

van Dijk, K., Fouts, D.E., Rehm, A.H., Hill, A.R., Collmer, A. and Alfano, J.R. 1999. The Avr (effector) proteins HrmA (HopPsyA) and AvrPto are secreted in culture from Pseudomonas syringae pathovars via the Hrp (type III) protein secretion system in a temperature- and pH-sensitive manner. *J. Bacteriol.* **181:** 4790-4797.

Wilharm, G., Lehmann, V., Krauss, K., Lehnert, B., Richter, S., Ruckdeschel, K., Heesemann, J. and Trulzsch, K. 2004. Yersinia enterocolitica type III secretion depends on the proton motive force but not on the flagellar motor components MotA and MotB. *Infect. Immun.* **72:** 4004-4009.

Wu, A.J., Andriotis, V.M., Durrant, M.C. and Rathjen, J.P. 2004. A patch of surface-exposed residues mediates negative regulation of immune signaling by tomato Pto kinase. *Plant Cell* **16:** 2809-2821.

Wulf, J., Pascuzzi, P.E., Fahmy, A., Martin, G.B. and Nicholson, L.K. 2004b. The solution structure of type III effector protein AvrPto reveals conformational and dynamic features important for plant pathogenesis. *Structure* **12:** 1257-1268.

Xiao, F., He, P., Abramovitch, R.B., Dawson, J.E., Nicholson, L.K., Sheen, J. and Martin, G.B. 2007. The N-terminal region of Pseudomonas type III effector AvrPtoB elicits Pto-dependent immunity and has two distinct virulence determinants. *Plant J.* **52:** 595-614.

Yang, A.S. and Honig, B. 1993. On the pH dependence of protein stability. *J.Mol.Biol.* **231:** 459-474.

Chapter Five

Biological Relevance and Future Plans

***Biological Relevance***

Secretion *via* the TTSS is a complex process relying on the tight control of genetic expression pathways with multiple feedback regulations (Jin et al. 2003) , the pH of the cell's cytoplasm and its outside environment (Minamino et al. 2003, Wilharm et al. 2004), and the folding kinetics and thermodynamics of the secreted effector proteins themselves (Lee et al. 2002, Sorg et al. 2005). Expression of *P. syringae* genes associated with Type III secretion only occurs within the plant leaf tissue (Jin et al. 2003) or in *hrp*-derepressing minimal medium (Rahme et al. 1992, Xiao et al. 1992, van Dijk et al. 1999). The bacteria proliferates in the extracellular space (apoplast) within the tissue, where the pH is mildly acidic, $pH_{ext}$~5-6.5 (Grignon et al. 1991) (Figure 5.1). The internal pH of the bacterial cytoplasm should be slightly greater than this $pH_{ext}$ in order to maintain the required proton gradient for the TTSS (Wilharm et al. 2004). The effector proteins pass through the TTSS in partially unfolded form into the more basic host cell cytoplasm, in a process called translocation (Johnson et al. 2005). In minimal medium during *in vitro* secretion assays, the TTSS secretes the effector protein AvrPto into the outside medium preferentially at pH 6 (van Dijk et al. 1999). Both TrAvrPto (Chapter Three) and full length AvrPto (Wulf et al. 2004) are destabilized under acidic conditions. This acid denaturation should increase the amount of secretion-ready AvrPto available within the pathogen cytoplasm during secretion. Once translocated into the neutral host cytoplasm, AvrPto can then refold independently into its functional form.

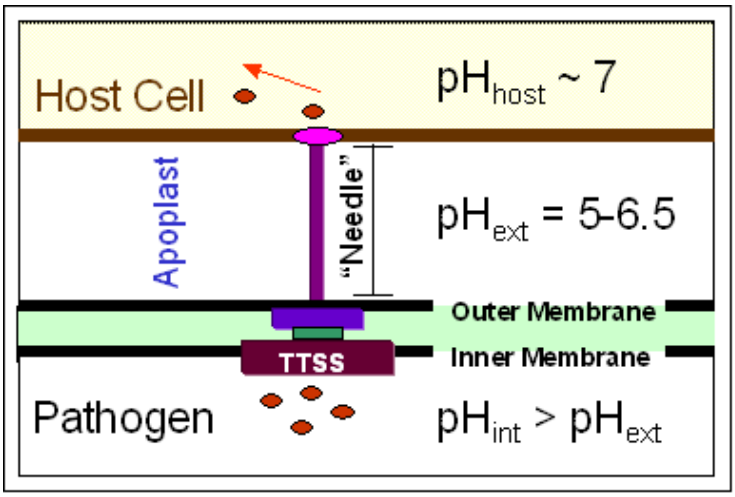



Figure 5.1: **pH and Type III secretion.** When infecting plant leaves, *P. syringae* resides in the mildly acidic ($pH_{ext}$=5-6.5) apoplast (extracellular space). To keep the proton gradient required for TTSS function, when the outside environment is acidic, the pathogen cytoplasm's $pH_{int}$ also becomes more acidic. Effector protein AvrPto is destabilized under these conditions of the bacterial cytoplasm, allowing it to unfold and pass through the TTSS into the host cytoplasm where it refolds at neutral pH.

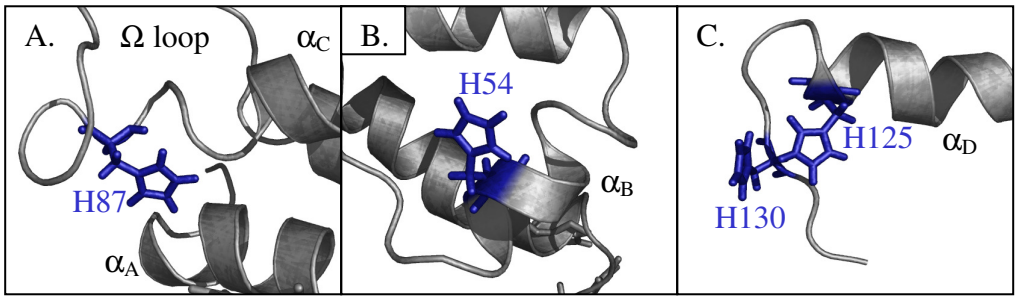



Figure 5.2: Local sub-structures within **native forms of TrAvrPto.** A. Burial of H87 in native TrAvrPto Ω loop. B. H54 near the δ+ end of $\alpha_B$ in native TrAvrPto. C. H125, H130, and the C-capping box at the end of $\alpha_D$.

The slow folding rate also influences the available pool of secretable protein by extending the lifetime of the non-native protein. The folding process in the pH range relevant to secretion (pH 5-7) is primarily between intermediate and native forms of TrAvrPto. The intermediate conformation has a relatively long lifetime in this pH range: $\tau=1/k_{UF}=0.6$ s at pH 6.1 (Chapter Two).

Some information about the intermediate of TrAvrPto can be gleaned from the experimental data (Chapter Three). It has α-helices and a hydrophobic core, like a molten globule. The acid denaturation at higher pH can be simulated remarkably well by assuming an average His $pK_a$ for the sidechains in the intermediate, $<pK_{a,I}=6.4>$, which matches that expected for solvent-exposed His (Nozaki et al. 1962). In native TrAvrPto, H87 is buried in the Ω loop, not in the three-helix bundle (Figure 5.2A). H87 is in a hydrophobic environment, depressing its $pK_{a,N}$ in the native protein, while in the intermediate form, the loops are most likely disordered, exposing the sidechain to solvent. Fascinatingly, this sidechain, which drives the acid denaturation and is conserved for AvrPto and AvrPtoB homologues, is also located in the Ω loop critical for interaction with Pto by both AvrPto and AvrPtoB (Bernal et al. 2005, Wu et al. 2004, Wulf et al. 2004, Xiao et al. 2007). H54 is near the δ+ end of $\alpha_B$ (Figure 5.2B), resulting in a depressed $pK_{a,N}$. Collapse of this very short helix into a loop should raise H54's $pK_{a,N}$. The remaining five His are on the native protein surface; therefore, the removal of loop sub-structures, such as the $\alpha_D$ capping box (Figure 5.2C), should result in all His being solvent-exposed.

The presence of the intermediate conformation is enigmatic. Is it involved in tuning the pH-dependence of stability and secretion efficiency? Removal of this intermediate should shift the pH-dependence towards higher values, making it more sensitive to changes in environmental pH conditions. Could the presence of the intermediate simply be a consequence of protein chemistry? The villin headpiece and

sperm whale apoglobin both display pH-dependent stability that is heavily influenced by a buried His and both have an intermediate conformation (Geierstanger et al. 1998, Grey et al. 2006). Could the presence of the partially folded intermediate keep the protein from being degraded within the pathogen? Lon protease of *P. syringae* acts both to negatively-regulate the expression levels of the TTSS and associated genes and to degrade effector proteins, including AvrPto, after translation (Bretz et al. 2002, Losada et al. 2005). Some effector proteins—HopPsyB, HopPsyV1, and HopPtoM – have secretion chaperones within the pathogen cytoplasm to protect them from Lon protease (Losada et al. 2005). AvrPto does not appear to have a secretion chaperone. The intermediate conformation of AvrPto, however, might conceal its degradation sites even under acidic conditions. The intermediate appears to be a molten globule. This type of conformation, while lacking in specific tertiary contacts, is typically more compactly-packed than a random coil (Alexandrescu et al. 1993, Arai et al. 2000). The intermediate has a destabilized core, which may be relevant for its ability to be trafficked through the TTSS. The dimensions of the TTSS lumen would permit passage of a string of intact alpha helices. This suggests that only partial unfolding would be necessary, requiring a smaller energy expenditure, as observed here for TrAvrPto.

### *Future Plans*

#### *Mutagenesis*

A series of promising mutants were designed with the purpose of improving the overall stability of AvrPto or changing the tuning of its pH-dependence of stability (Chapter Four). However, the only way to determine the effects of mutation is by experimental measurements. NMR HSQC spectra may be used as a quick test for the presence of the native conformation, which has a characteristic pattern of peaks in

these spectra (Wulf et al. 2002). NMR HSQC, HMQC, and Nzz spectra are available to quantitate on a per-residue basis the changes in local chemical environment, pH-dependent stability, His $pK_a$, and folding kinetics due to the mutation. The presence of the intermediate was revealed for wt TrAvrPto by comparing the pH-dependent stability data from CD, Trp fluorescence, and NMR spectroscopy, so any mutant that alters this conformation should also change the observations from at least one of these techniques. Once the folding and pH-dependence of the mutant has been tested, mutant AvrPto-CyaA constructs will be used in conjunction with the translocation assay to test the kinetic and thermodynamics thresholds of the TTSS.

As an alternative to site-directed mutagenesis, AvrPto homologues Q87Y16 and Q4ZLM6 (TrEMBL accession number; Table 3.3) could be used. These sequences are extremely (>95%) similar, but do exhibit variations, most notably the H41R and H103R substitutions in Q4ZLM6 and H125Y in Q87Y16. AvrPtoB(121-200) and its homologues are other potential probes of translocation thresholds.

*Exploring the low pH titration*

The low pH titration (Chapter Three) is an unexplored avenue of TrAvrPto folding. Neither the ionization energy for the intermediate form, nor the identities of the titratable groups that control this transition are known (Chapter Four). The elucidation of the low pH titration is potentially more challenging than the characterization of the high pH titration. Exchange between the unfolded ensemble (U) and the intermediate (I) occurs on the fast chemical exchange timescale, so both conformations contribute to the same set of peaks. The intermediate and native (N) forms are in slow exchange, resulting in a second set of peaks at higher pH that complicate the spectra further. There are several model systems that may allow the

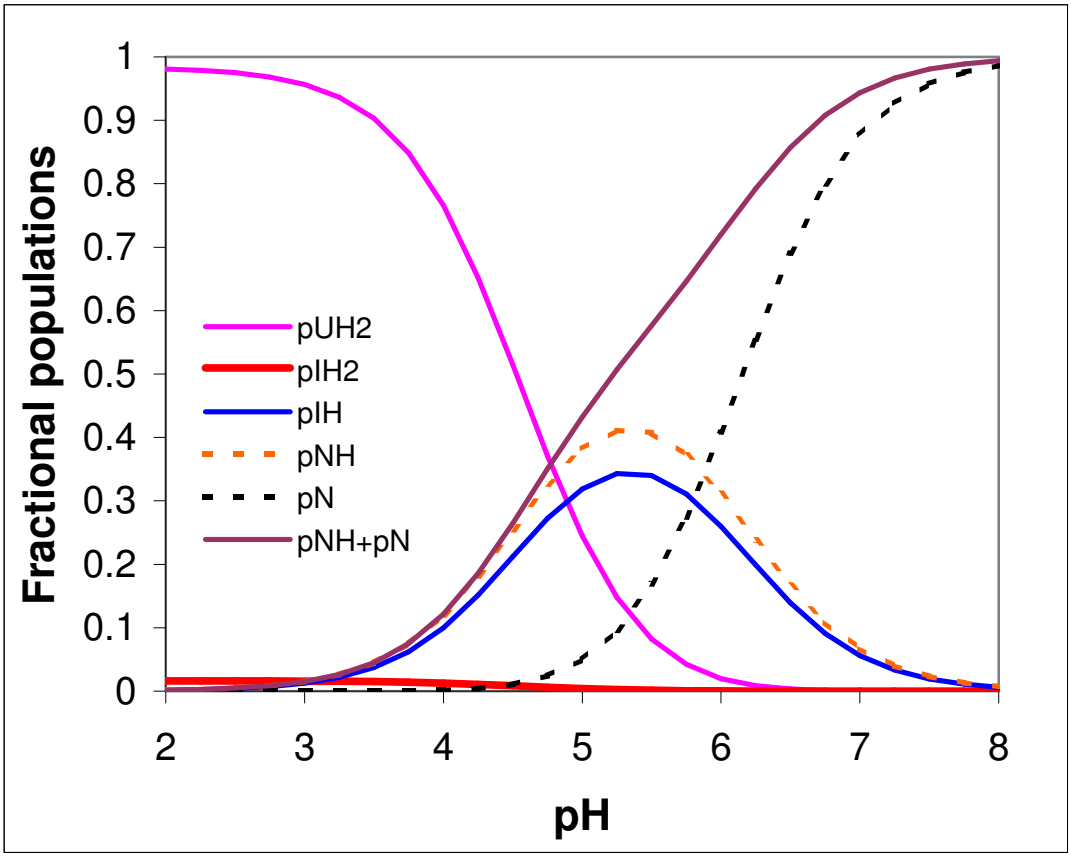



Figure 5.3: **The dependence of fractional TrAvrPto populations upon pH.** The fractional population curves generated from the fitting of CD, Trp fluorescence, and NMR HSQC data to the two-titration, five-state model (Chapter Three Eq.1-7, Figure 3.1, Table 3.1). The unfolded ensemble population (solid pink line) is minor at pH 6, but is the predominant species at pH<3. The intermediate form, $p_{IH2}$ (solid red line) and $p_{IH}$ (solid blue line), is not predicted to be the dominant species under any pH condition.

low pH titration to be characterized: wt TrAvrPto at low pH, H54P TrAvrPto, and the E104Q and E55Q mutants.

1. wt TrAvrPto at low pH

Investigating wt TrAvrPto at low pH is the most complicated strategy for characterizing the intermediate conformation. The pH-dependence of stability, as measured by CD, Trp fluorescence, and NMR spectroscopy (Chapters Three and Four), can be modeled using a five state equilibrium: $UH_2$, $IH_2$, IH, NH, and N (Figure 5.3). The unfolded ensemble ($p_{UH2}$) constitutes the majority of the total population at pH<3. Unfortunately, there is no pH predicted to isolate the intermediate ($p_{IH2}$ or $p_{IH}$); it is always in equilibrium with significant amounts of the unfolded or native populations.

Exchange between conformations during NMR relaxation experiments affects the apparent relaxation rate constants. This effect was observed in $T_1$ experimental data for TrAvrPto at pH 6.1 (Chapter Two). The effective relaxation rate constants, $R_1$(eff), extracted from this data contained contributions from the unfolding rate. Exchange also affects $T_2$ experiments, such as CPMG (Hansen et al. 2008). Here, since the relaxation rate constants $R_2$ control the width of the peaks, exchange alters the peaks' lineshapes. CPMG experiments are sensitive to the exchange of even very minor conformational populations (1-2%), allowing the kinetic rates, populations, and chemical shifts to be determined. Therefore, the kinetics can be characterized from the exchange rates, the thermodynamics from the extracted populations, and structural information can be gleaned from the chemical shifts. CPMG experiments are available to probe $^{15}N$, $^{1}H_N$, $^{13}C_\alpha$, $^{13}CO$, and sidechain methyl and methylene groups. The presence of the native state complicates the proposed analysis adding another conformational exchange and ionization event.

2. Removing the native conformation: H54P TrAvrPto characterization

The system would be simplified by using a mutation that removes the native conformation. The native conformation is dramatically destabilized in the H54P mutant, eliminating its peaks from NMR spectra at lower pH. Even at neutral pH, the peaks are weak. Since the H54P substitution does not appear to perturb the intermediate conformation (Chapter Four), this mutant is a good model system for the low pH titration.

The removal of the native population would not only simplify any CPMG analysis, it may simplify the spectra enough that the structure of the intermediate can be determined. The population of unfolded ensemble ($p_{UH2}$) is predicted to be nearly zero at pH 6 for wt TrAvrPto (Figure 5.3) and there is no observed native conformation at this pH for H54P TrAvrPto (Figure 4.11A). Therefore, the NMR peaks should reflect the chemical environments of the intermediate conformation. Only 22 of the non-native peaks have been assigned for wt TrAvrPto at pH 6.1 (Appendix Table 2.1). Standard NMR experiments that transfer magnetization between covalently linked spins (ex. TOCSY) and those that transfer through space (ex. NOESY) will be needed to assign the rest of the peaks and may yield enough information to determine the structure of the intermediate conformation (Cavanagh et al. 1996).

3. Removing the intermediate conformation

The groups controlling the low pH titration are as yet unidentified. However, based on structural arguments, E104 is the best candidate for this pH switch, followed by E55. If E104Q or E55Q substitutions remove the intermediate, then the pH-dependence of stability data observed by CD, Trp fluorescence, and NMR spectroscopy should be different from that of wt TrAvrPto. For wt TrAvrPto, CD and fluorescence data does not distinguish between the intermediate and native forms,

while the distinct NMR "folded" peaks are proportional to only the native population. This is responsible for the discrepancy between the two sets of data. If either mutation removes the intermediate, then the data from CD, Trp fluorescence, and NMR spectroscopy should be identical and should no longer be described by the two-titration, five-state model in Chapter Three.

### *Closing Remarks*

TrAvrPto is a fascinating protein. It is destabilized under non-denaturing conditions, folds extremely slowly for a three-helix bundle protein, and is a small one-domain protein with a folding intermediate. Based on general trends in protein behavior, it could not have been predicted that it would behave in this manner. However, TrAvrPto is under a different set of evolutionary pressures than most proteins: it must be destabilized relatively simply in order to pass through the TTSS. The bulk of this research focused on elucidating the folding of wild-type TrAvrPto. The mutagenesis, translocation, and microbiological aspects of the project will be passed on to others, who will further tie the *in vitro* work to the biological realm.

# REFERENCES


Alexandrescu, A.T., Evans, P.A., Pitkeathly, M., Baum, J. and Dobson, C.M. 1993. Structure and dynamics of the acid-denatured molten globule state of alpha-lactalbumin: a two-dimensional NMR study. *Biochemistry* **32:** 1707-1718.

Arai, M. and Kuwajima, K. 2000. Role of the molten globule state in protein folding. *Adv.Protein Chem.* **53:** 209-282.

Bernal, A.J., Pan, Q., Pollack, J., Rose, L., Kozik, A., Willits, N., Luo, Y., Guittet, M., Kochetkova, E. and Michelmore, R.W. 2005. Functional analysis of the plant disease resistance gene Pto using DNA shuffling. *J.Biol.Chem.* **280:** 23073-23083.

Bretz, J., Losada, L., Lisboa, K. and Hutcheson, S.W. 2002. Lon protease functions as a negative regulator of type III protein secretion in Pseudomonas syringae. *Mol.Microbiol.* **45:** 397-409.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G.I. and Skelton, N.J. 1996. *Protein NMR spectroscopy: Principles and practice,* 243-299. Academic Press, San Diego, CA.

Geierstanger, B., Jamin, M., Volkman, B.F. and Baldwin, R.L. 1998. Protonation behavior of histidine 24 and histidine 119 in forming the pH 4 folding intermediate of apomyoglobin. *Biochemistry* **37:** 4254-4265.

Grey, M.J., Tang, Y., Alexov, E., McKnight, C.J., Raleigh, D.P. and Palmer, A.G.,3rd. 2006. Characterizing a partially folded intermediate of the villin headpiece domain under non-denaturing conditions: contribution of His41 to the pH-dependent stability of the N-terminal subdomain. *J.Mol.Biol.* **355:** 1078-1094.

Grignon, C. and Sentenac, H. 1991. pH and ionic conditions in the apoplast. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42:** 103-128.

Hansen, D.F., Vallurupalli, P. and Kay, L.E. 2008. Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J.Biomol.NMR*

Jin, Q., Thilmony, R., Zwiesler-Vollick, J. and He, S.Y. 2003. Type III protein secretion in Pseudomonas syringae. *Microb. Infect.* **5:** 301-310.

Johnson, S., Deane, J.E. and Lea, S.M. 2005. The type III needle and the damage done. *Curr. Opin. Struct. Biol.* **15:** 700-707.

Lee, V.T. and Schneewind, O. 2002. Yop fusions to tightly folded protein domains and their effects on Yersinia enterocolitica type III secretion. *J. Bacteriol.* **184:** 3740-3745.

Losada, L.C. and Hutcheson, S.W. 2005. Type III secretion chaperones of Pseudomonas syringae protect effectors from Lon-associated degradation. *Mol.Microbiol.* **55:** 941-953.

Minamino, T., Imae, Y., Oosawa, F., Kobayashi, Y. and Oosawa, K. 2003. Effect of intracellular pH on rotational speed of bacterial flagellar motors. *J. Bacteriol.* **185:** 1190-1194.

Nozaki, Y. and Tanford, C. 1962. Examination of titration behavior. *Advan Protein Chem* **17:** 715-734.

Rahme, L.G., Mindrinos, M.N. and Panopoulos, N.J. 1992. Plant and environmental sensory signals control the expression of hrp genes in Pseudomonas syringae pv. phaseolicola. *J. Bacteriol.* **174:** 3499-3507.

Sorg, J.A., Miller, N.C., Marketon, M.M. and Schneewind, O. 2005. Rejection of impassable substrates by Yersinia type III secretion machines. *J. Bacteriol.* **187:** 7090-7102.

van Dijk, K., Fouts, D.E., Rehm, A.H., Hill, A.R., Collmer, A. and Alfano, J.R. 1999. The Avr (effector) proteins HrmA (HopPsyA) and AvrPto are secreted in culture from Pseudomonas syringae pathovars via the Hrp (type III) protein secretion system in a temperature- and pH-sensitive manner. *J. Bacteriol.* **181:** 4790-4797.

Wilharm, G., Lehmann, V., Krauss, K., Lehnert, B., Richter, S., Ruckdeschel, K., Heesemann, J. and Trulzsch, K. 2004. Yersinia enterocolitica type III secretion depends on the proton motive force but not on the flagellar motor components MotA and MotB. *Infect. Immun.* **72:** 4004-4009.

Wu, A.J., Andriotis, V.M., Durrant, M.C. and Rathjen, J.P. 2004. A patch of surface-exposed residues mediates negative regulation of immune signaling by tomato Pto kinase. *Plant Cell* **16:** 2809-2821.

Wulf, J., Pascuzzi, P.E., Martin, G.B. and Nicholson, L.K. 2002. 1H, 15N and 13C chemical shift assignments of the structured core of the pseudomonas effector protein AvrPto. *J. Biomol. NMR* **23:** 247-248.

Wulf, J., Pascuzzi, P.E., Fahmy, A., Martin, G.B. and Nicholson, L.K. 2004. The solution structure of type III effector protein AvrPto reveals conformational and dynamic features important for plant pathogenesis. *Structure* **12:** 1257-1268.

Xiao, F., He, P., Abramovitch, R.B., Dawson, J.E., Nicholson, L.K., Sheen, J. and Martin, G.B. 2007. The N-terminal region of Pseudomonas type III effector AvrPtoB elicits Pto-dependent immunity and has two distinct virulence determinants. *Plant J.* **52:** 595-614.

Xiao, Y., Lu, Y., Heu, S. and Hutcheson, S.W. 1992. Organization and environmental regulation of the Pseudomonas syringae pv. syringae 61 hrp cluster. *J. Bacteriol.* **174:** 1734-1741.

APPENDIX

Appendix Table 2.1**: Peak assignments of Nzz auto- and cross-peaks**

| | Peak assignment (ppm) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Residue | FF $\omega_{1H}$ [a] | FF $\omega_{15N}$ [b] | FU $\omega_{1H}$ | FU $\omega_{15N}$ | UF $\omega_{1H}$ | UF $\omega_{15N}$ | UU $\omega_{1H}$ | UU $\omega_{15N}$ |
| Q35 | 7.854 | 125.46 | 8.318 | 125.46 | 7.854 | 122.88 | - | - |
| L37 | 8.152 | 118.89 | 7.909 | 118.91 | - | - | 7.909 | 124.91 |
| H41 | 8.441 | 118.38 | - | - | 8.44 | 119.70 | - | - |
| E45 | 8.03 | 116.61 | 8.366 | 116.57 | - | - | - | - |
| A47 | 7.169 | 122.44 | 8.266 | 122.48 | 7.169 | 126.44 | 8.27 | 126.42 |
| G48 | 6.352 | 107.13 | - | - | 6.354 | 108.25 | 8.262 | 108.26 |
| D52 | 8.966 | 117.76 | - | - | 8.965 | 121.44 | - | - |
| H54 | 8.717 | 120.32 | - | - | 8.716 | 122.69 | - | - |
| E55 | 8.866 | 121.38 | - | - | 8.866 | 121.96 | - | - |
| S58 | 7.415 | 113.34 | 8.369 | 113.32 | 7.414 | 120.10 | - | - |
| S59 | 7.576 | 112.49 | 8.352 | 112.48 | - | - | - | - |
| A61 | 7.09 | 121.76 | - | - | 7.088 | 126.91 | 8.213 | 126.93 |
| Q63 | 8.984 | 124.72 | 8.611 | 124.72 | 8.984 | 120.97 | | |
| S64 | 8.846 | 113.42 | 8.286 | 113.42 | 8.846 | 116.63 | 8.285 | 116.70 |
| N67 | 8.905 | 118.15 | - | - | 8.905 | 119.35 | | |
| [c]L72 | 8.178 | 124.31 | 8.066 | 124.35 | 8.176 | 122.14 | 8.069 | 122.14 |
| Y73 | 9.602 | 126.31 | 8.033 | 126.32 | 9.604 | 119.65 | 8.032 | 119.71 |
| T76 | 8.842 | 119.51 | 8.106 | 119.53 | 8.842 | 115.37 | 8.112 | 115.47 |
| R78 | 7.497 | 118.82 | - | - | 7.496 | 122.85 | - | - |
| L80 | 8.669 | 125.14 | 8.402 | 125.11 | - | - | 8.401 | 122.13 |
| Q86 | 8.661 | 122.01 | - | - | 8.658 | 119.39 | - | - |
| H87 | 9.117 | 121.61 | - | - | 9.117 | 118.42 | - | - |
| M90 | 9.408 | 120.10 | - | - | 9.406 | 121.76 | - | - |
| T91 | 7.514 | 105.56 | 8.087 | 105.57 | 7.512 | 114.74 | 8.087 | 114.76 |
| G92 | 7.513 | 110.33 | 8.367 | 110.34 | 7.511 | 111.85 | 8.372 | 111.86 |
| S94 | 7.413 | 110.75 | 8.312 | 110.75 | 7.413 | 115.27 | 8.314 | 115.30 |
| G95 | 8.752 | 113.78 | 8.322 | 113.78 | 8.754 | 111.14 | 8.324 | 111.16 |
| N97 | 8.463 | 129.68 | 8.34 | 129.67 | - | - | - | - |
| G99 | 8.261 | 107.29 | 8.426 | 107.29 | 8.26 | 108.78 | 8.426 | 108.80 |
| L101 | 9.277 | 125.71 | 8.138 | 125.70 | 9.279 | 124.90 | 8.14 | 124.92 |
| H103 | 9.475 | 113.68 | 8.52 | 113.66 | 9.477 | 119.21 | 8.523 | 119.22 |
| E104 | 7.691 | 121.63 | - | - | 7.685 | 122.62 | 8.468 | 122.65 |
| N105 | 7.37 | 116.39 | 8.624 | 116.40 | 7.369 | 120.95 | - | - |
| M109 | 8.675 | 120.25 | - | - | 8.672 | 120.92 | - | - |
| R110 | 9.039 | 119.44 | 8.213 | 119.43 | 9.039 | 120.08 | 8.207 | 120.12 |
| A112 | 7.849 | 124.43 | - | - | 7.849 | 126.08 | 8.204 | 126.11 |
| W116 ε1 | 10.534 | 129.60 | 10.118 | 129.59 | 10.533 | 129.94 | 10.121 | 129.96 |
| R120 | 8.623 | 119.86 | 8.176 | 119.90 | - | - | 8.18 | 121.28 |
| E121 | 7.924 | 120.27 | 8.274 | 120.26 | - | - | - | - |
| G128 | 7.548 | 107.13 | - | - | 7.548 | 109.87 | 8.228 | 109.87 |
| I129 | 7.582 | 119.80 | - | - | 7.583 | 120.15 | - | - |

[a] The uncertainty in $\omega_{1H}$ is $\sigma_{iH}$~±0.008 ppm

[b] The uncertainty in $\omega_{15N}$ is $\sigma_{15N}$~±0.06 ppm

[c] The UU peak of L72 was assigned. However, since the residue's FF peak is overlapped at all timepoints, it was not used in the global kinetic rate fitting.

Appendix Table 2.2: **$R_{1F}$ data from Nzz, $T_1$, and simulation experiments**[a]

| Amide group | 0.63mM, pH6.1 $R_{1F}^{Nzz}$ [b] | 1.2 mM, pH7.0 $R_{1F}$(eff) [c] | 0.92 mM, pH 6.8 $R_{1F}$(eff) [d] | $R_{1F}$(sim)[e] | 0.34 mM, pH6.1 $R_{1F}$(eff) [f] | 0.87 mM, pH6.1 $R_{1F}$(eff) [g] | $R_{1F}$(eff) calculated from Nzz[h] |
|---|---|---|---|---|---|---|---|
| Q35 | 1.11±0.06 | 1.23±0.01 | 1.33±0.08 | 1.18 | 1.61±0.11 | 1.95±0.17 | 1.45±0.07 |
| L37 | 1.17±0.06 | 1.20±0.01 | 1.32±0.11 | 1.24 | 1.63±0.10 | 1.65±0.09 | 1.50±0.07 |
| H41 | 1.10±0.05 | 1.25±0.01 | 1.33±0.13 | 1.31 | 1.60±0.09 | 1.72±0.10 | 1.43±0.06 |
| E45 | 1.16±0.06 | 1.23±0.01 | 1.31±0.16 | 1.26 | 1.50±0.06 | 1.56±0.09 | 1.49±0.07 |
| A47 | 1.11±0.06 | 1.20±0.02 | 1.26±0.07 | 1.17 | 1.52±0.05 | 1.60±0.08 | 1.44±0.07 |
| G48 | 1.18±0.06 | 1.32±0.01 | 1.39±0.17 | 1.47 | 1.54±0.07 | 1.65±0.05 | 1.51±0.07 |
| D52 | 1.47±0.08 | 1.50±0.04 | 1.58±0.10 | 1.51 | 2.05±0.15 | 2.11±0.15 | 1.81±0.09 |
| H54 | 1.34±0.05 | 1.51±0.03 | 1.63±0.11 | 1.50 | 1.76±0.09 | - | 1.67±0.07 |
| E55 | 1.55±0.10 | 1.56±0.02 | 1.70±0.05 | 1.49 | 2.04±0.08 | 2.00±0.07 | 1.88±0.11 |
| S58 | 1.52±0.06 | 1.54±0.01 | 1.64±0.28 | 1.47 | 2.08±0.18 | 2.24±0.12 | 1.85±0.07 |
| A61 | 1.29±0.05 | 1.30±0.01 | 1.37±0.08 | 1.45 | 1.69±0.10 | 1.76±0.10 | 1.62±0.07 |
| Q63 | 1.32±0.08 | - | - | 1.27 | - | - | 1.65±0.09 |
| S64 | 1.25±0.07 | - | - | 1.14 | - | - | 1.58±0.08 |
| N67 | 1.17±0.06 | 1.20±0.02 | 1.31±0.09 | 1.20 | 1.58±0.07 | 1.73±0.08 | 1.50±0.07 |
| Y73 | 1.14±0.07 | 1.21±0.02 | 1.31±0.10 | 1.22 | 1.47±0.07 | 1.62±0.11 | 1.47±0.08 |
| T76 | 1.22±0.08 | 1.21±0.02 | 1.27±0.13 | 1.18 | 1.62±0.06 | 1.61±0.10 | 1.55±0.09 |
| R78 | 1.28±0.09 | 1.23±0.02 | 1.36±0.11 | 1.28 | 1.61±0.06 | 1.64±0.09 | 1.61±0.10 |
| L80 | 1.09±0.06 | 1.25±0.01 | 1.36±0.10 | 1.18 | 1.50±0.28 | 1.66±0.06 | 1.43±0.07 |
| Q86 | 1.26±0.07 | 1.15±0.02 | 1.27±0.10 | 1.23 | 1.56±0.06 | 1.58±0.11 | 1.59±0.08 |
| H87 | 1.18±0.07 | 1.20±0.02 | 1.33±0.23 | 1.14 | 1.60±0.09 | 1.77±0.11 | 1.51±0.08 |
| M90 | 1.20±0.07 | 1.13±0.01 | 1.28±0.07 | 1.51 | 1.64±0.09 | 1.74±0.08 | 1.53±0.08 |
| T91 | 1.34±0.07 | 1.26±0.01 | 1.35±0.11 | 1.47 | 2.07±0.17 | - | 1.68±0.08 |
| G92 | 1.40±0.06 | 1.41±0.01 | 1.54±0.10 | 1.19 | 1.89±0.17 | - | 1.73±0.07 |
| S94 | 1.36±0.05 | 1.41±0.01 | 1.50±0.08 | 1.22 | 1.94±0.16 | - | 1.69±0.06 |
| G95 | 1.48±0.07 | 1.27±0.03 | 1.32±0.09 | 1.51 | - | - | 1.82±0.08 |
| L101 | 1.39±0.06 | 1.42±0.02 | 1.52±0.07 | 1.35 | 1.79±0.05 | 1.74±0.03 | 1.72±0.08 |
| E104 | 1.44±0.07 | 1.26±0.03 | 1.40±0.15 | 1.17 | 1.56±0.06 | 1.64±0.05 | 1.78±0.08 |
| N105 | 1.13±0.07 | 1.12±0.02 | 1.23±0.08 | 1.16 | 1.51±0.06 | 1.51±0.04 | 1.47±0.08 |
| M109 | 1.21±0.06 | 1.15±0.02 | 1.24±0.08 | 1.17 | 1.54±0.07 | - | 1.55±0.07 |
| A112 | 1.12±0.04 | 1.17±0.01 | 1.20±0.07 | 1.16 | 1.37±0.06 | 1.47±0.06 | 1.46±0.06 |
| W116 ε1 | 1.17±0.06 | 1.14±0.01 | 1.22±0.12 | 1.14 | 1.40±0.07 | 1.42±0.07 | 1.51±0.07 |
| R120 | 1.12±0.06 | 1.18±0.01 | 1.24±0.08 | 1.19 | 1.43±0.07 | 1.49±0.06 | 1.45±0.08 |
| E121 | 1.25±0.05 | 1.18±0.01 | 1.24±0.10 | 1.17 | 1.42±0.05 | 1.46±0.06 | 1.58±0.06 |
| G128 | 1.35±0.06 | 1.38±0.02 | 1.43±0.10 | 1.27 | 1.66±0.06 | 1.68±0.08 | 1.69±0.07 |
| I129 | 1.34±0.05 | 1.43±0.02 | 1.47±0.11 | 1.31 | 1.74±0.08 | 1.84±0.06 | 1.67±0.07 |
| **$\langle R_{1F}\rangle\pm\sigma$[i]** | **1.26±0.06** | **1.28±0.02** | **1.37±0.11** | **1.28** | **1.65±0.09** | **1.70±0.08** | **1.60±0.08** |

[a] $R_{1F}$ values are in units of $s^{-1}$.
[b] $R_{1F}^{Nzz}$ values are from the 2-parameter Nzz fitting using global kinetic rate constants, $k_{FU}$ and $k_{UF}$.
[c] $R_{1F}$(eff,1.2mM) data overlays $R_{1F}^{Nzz}$ with an average difference of $\Delta\langle R_{1F}\rangle = \langle R_{1F}^{Nzz}\rangle - \langle R_{1F}(\text{eff,1.2mM})\rangle = -0.02 \pm 0.07\ s^{-1}$.
[d] $R_{1F}$(eff,0.92mM) data overlays $R_{1F}$(eff,1.2mM) with an average difference of $+0.09 \pm 0.11\ s^{-1}$.
[e] $R_{1F}$(sim) simulated as a 6.1nm by 3.1nm prolate monomer using NH bond orientations from TrAvrPro structure. Data overlays $R_{1F}^{Nzz}$ with an average difference of $-0.02 \pm 0.06 s^{-1}$.
[f] $R_{1F}$(eff,0.34mM) and $R_{1F}$(eff,1.2mM) are offset by $+0.37 \pm 0.09\ s^{-1}$ on average.
[g] $R_{1F}$(eff,0.87mM) overlays $R_{1F}$(eff,0.34mM) with an average difference of $-0.04 \pm 0.13\ s^{-1}$.
[h] Calculated using $R_{1F}(\text{eff}) = R_{1F} + k_{FU}$ and Nzz-derived parameters. Data overlays 0.34 mM $R_{1F}$(eff) with an average difference of $-0.06 \pm 0.12\ s^{-1}$.
[i] $\langle R_{1F}\rangle\pm\sigma$ is the mean and standard deviation over all residues.

Appendix Table 2.3**: Simulating $R_{1F}$ as an equilibrium between prolate monomer and oblate dimer in fast chemical exchange**

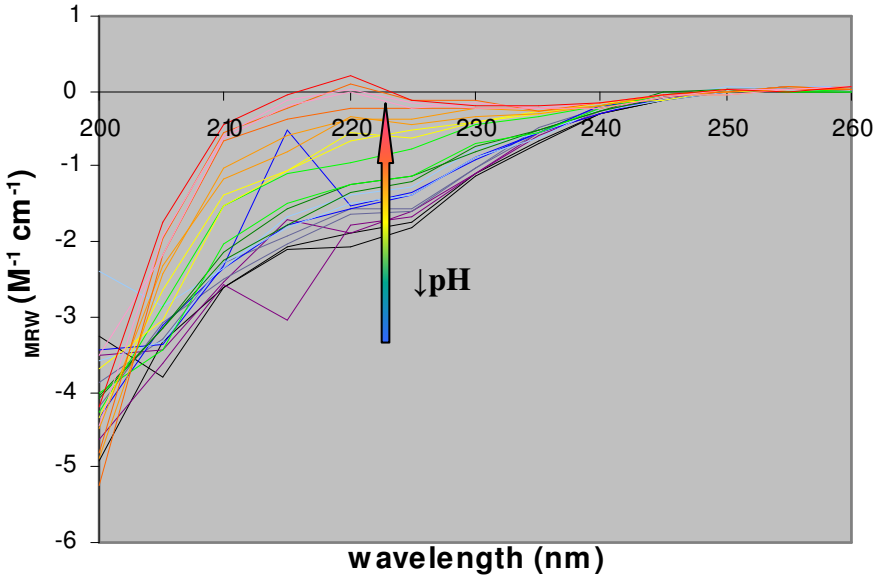
| Amide group | $p_M$=0.7 $R_{1F}$ simulated ($s^{-1}$) [a] | $p_M$=0.79 $R_{1F}$ simulated ($s^{-1}$) [b] | $p_M$=0.65 $R_{1F}$ simulated ($s^{-1}$) [c] |
|---|---|---|---|
| Q35 | 1.25 | 1.35 | 1.19 |
| L37 | 1.30 | 1.41 | 1.23 |
| H41 | 1.37 | 1.49 | 1.30 |
| E45 | 1.32 | 1.43 | 1.25 |
| A47 | 1.23 | 1.33 | 1.17 |
| G48 | 1.51 | 1.65 | 1.43 |
| D52 | 1.54 | 1.68 | 1.46 |
| H54 | 1.54 | 1.68 | 1.46 |
| E55 | 1.53 | 1.67 | 1.45 |
| S58 | 1.51 | 1.65 | 1.43 |
| A61 | 1.50 | 1.63 | 1.42 |
| Q63 | 1.33 | 1.45 | 1.27 |
| S64 | 1.21 | 1.31 | 1.15 |
| N67 | 1.26 | 1.37 | 1.20 |
| Y73 | 1.28 | 1.39 | 1.22 |
| T76 | 1.24 | 1.35 | 1.18 |
| R78 | 1.34 | 1.46 | 1.28 |
| L80 | 1.24 | 1.35 | 1.19 |
| Q86 | 1.30 | 1.41 | 1.23 |
| H87 | 1.20 | 1.30 | 1.15 |
| M90 | 1.55 | 1.69 | 1.47 |
| T91 | 1.52 | 1.65 | 1.44 |
| G92 | 1.25 | 1.36 | 1.19 |
| S94 | 1.28 | 1.39 | 1.22 |
| G95 | 1.55 | 1.69 | 1.47 |
| L101 | 1.41 | 1.53 | 1.34 |
| E104 | 1.23 | 1.34 | 1.18 |
| N105 | 1.22 | 1.32 | 1.16 |
| M109 | 1.23 | 1.34 | 1.18 |
| A112 | 1.22 | 1.33 | 1.17 |
| W116 ε1 | 1.20 | 1.30 | 1.15 |
| R120 | 1.24 | 1.34 | 1.18 |
| E121 | 1.33 | 1.45 | 1.27 |
| G128 | 1.37 | 1.49 | 1.30 |
| I129 | 1.47 | 1.60 | 1.40 |
| **$\langle R_{1F} \rangle$** [d] | **1.34** | **1.46** | **1.28** |

[a] Fitted to the $R_{1F}$ extracted from 0.63 mM, pH 6.1 Nzz data. A monomer population $p_M$=0.7 corresponds to $K_d$=2 mM. The model used a 5.6 by 2.6 nm prolate monomer and a 2.6 by 7.7 nm oblate dimer.

[b] Generated using the monomer-dimer model and $p_M$=0.79. The $p_M$ was estimated using $K_d$ =2 mM and a total TrAvrPto concentration of 0.34 mM.

[c] Generated using the monomer-dimer model and $p_M$=0.65. The $p_M$ was estimated using $K_d$ =2 mM and a total TrAvrPto concentration of 0.87 mM. The simulations for $p_M$ = 0.79 and 0.65 are offset by +0.18 $s^{-1}$ on average.

[d] $\langle R_{1F} \rangle$ is the average over all residues.




Appendix Figure 3.1: **Acid denaturation of TrAvrPto observed by CD spectroscopy.** The CD data displays a mixture of α-helical and random coil behavior at each pH value, becoming increasingly random coil-like as pH is decreased.

Appendix Table 3.1: **Native TrAvrPto backbone amide and Trp indole group chemical shifts[a] during pH-titration**

| pH 4.6 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ |
| T32 | 8.00 | 117.3 | V57 | 8.58 | 120.9 | A83 | 8.23 | 124.8 | R110 | 8.96 | 119.3 |
| S33 | 9.25 | 118.6 | S58 | 7.37 | 113.2 | D84 | - | - | S111 | 8.09 | 116.4 |
| S34 | 8.21 | 116.7 | S59 | 7.53 | 112.4 | M85 | - | - | A112 | 7.82 | 124.3 |
| Q35 | 7.85 | 125.4 | Q60 | - | - | Q86 | 8.68 | 121.7 | I113 | 8.56 | 121.4 |
| L37 | - | - | A61 | 7.06 | 121.6 | H87 | 9.05 | 120.2 | T114 | 7.77 | 118.4 |
| S38 | 8.11 | 116.2 | Q63 | 8.95 | 124.6 | R88 | 8.39 | 118.5 | W116 ε1 | 10.48 | 129.5 |
| V39 | 7.95 | 123.4 | S64 | 8.81 | 113.3 | Y89 | 8.02 | 121.6 | W116 | 8.80 | 123.2 |
| R40 | - | - | L65 | 7.01 | 124.2 | M90 | 9.27 | 119.5 | S117 | 8.10 | 118.0 |
| H41 | 8.39 | 118.0 | R66 | - | - | T91 | 7.38 | 105.3 | D118 | - | - |
| Q42 | - | - | N67 | 8.88 | 118.0 | G92 | 7.50 | 110.1 | M119 | 8.53 | 123.0 |
| L43 | 8.80 | 123.2 | R68 | 7.60 | 121.2 | S94 | 7.38 | 110.6 | R120 | - | - |
| A44 | - | - | Y69 | - | - | G95 | 8.72 | 113.6 | E121 | - | - |
| E45 | 7.97 | 115.9 | N70 | 8.90 | 120.1 | I96 | 7.99 | 121.7 | L123 | 8.69 | 120.4 |
| S46 | 8.00 | 117.3 | N71 | - | - | N97 | 8.42 | 129.4 | H125 | - | - |
| A47 | 7.15 | 122.5 | Y73 | 9.55 | 126.1 | G99 | 8.24 | 107.1 | M127 | 7.88 | 117.0 |
| G48 | 6.22 | 106.9 | S74 | 8.27 | 115.7 | M100 | 7.47 | 117.9 | G128 | 7.52 | 107.0 |
| L49 | 7.87 | 121.3 | H75 | - | - | L101 | 9.08 | 125.3 | I129 | 7.55 | 119.6 |
| R51 | 8.80 | 123.2 | T76 | 8.80 | 119.2 | H103 | 9.65 | 114.0 | H130 | 8.46 | 123.0 |
| D52 | 8.97 | 117.9 | Q77 | 7.94 | 121.1 | E104 | 7.75 | 121.1 | A131 | 8.27 | 126.1 |
| Q53 | 7.71 | 123.7 | R78 | 7.46 | 118.7 | N105 | 7.37 | 116.5 | D132 | 8.43 | 120.5 |
| H54 | 8.72 | 119.6 | T79 | - | - | V106 | 7.98 | 118.5 | I133 | 7.63 | 124.8 |
| E55 | 8.83 | 121.2 | L80 | 8.63 | 124.9 | D108 | 8.25 | 121.4 | | | |
| F56 | 7.92 | 120.9 | D81 | - | - | M109 | 8.62 | 120.1 | | | |
| pH 4.9 | | | | | | | | | | | |
| Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ |
| T32 | 8.01 | 117.3 | V57 | 8.58 | 120.9 | A83 | 8.22 | 124.7 | R110 | 8.97 | 119.3 |
| S33 | 9.27 | 118.6 | S58 | 7.37 | 113.2 | D84 | - | - | S111 | 8.08 | 116.3 |
| S34 | 8.17 | 116.8 | S59 | 7.54 | 112.3 | M85 | - | - | A112 | 7.82 | 124.3 |
| Q35 | 7.84 | 125.3 | Q60 | - | - | Q86 | 8.67 | 121.7 | I113 | 8.56 | 121.3 |
| L37 | - | - | A61 | 7.06 | 121.6 | H87 | 9.05 | 120.4 | T114 | 7.77 | 118.4 |
| S38 | 8.05 | 116.2 | Q63 | 8.96 | 124.6 | R88 | 8.39 | 118.5 | W116 ε1 | 10.48 | 129.4 |
| V39 | 7.95 | 123.3 | S64 | 8.81 | 113.3 | Y89 | 8.00 | 121.6 | W116 | 8.80 | 123.1 |
| R40 | - | - | L65 | 7.02 | 124.1 | M90 | 9.28 | 119.6 | S117 | 8.10 | 118.0 |
| H41 | 8.39 | 118.0 | R66 | - | - | T91 | 7.39 | 105.3 | D118 | - | - |
| Q42 | - | - | N67 | 8.87 | 118.0 | G92 | 7.50 | 110.1 | M119 | 8.53 | 123.0 |
| L43 | 8.80 | 123.1 | R68 | 7.59 | 121.2 | S94 | 7.37 | 110.6 | R120 | - | - |
| A44 | - | - | Y69 | - | - | G95 | 8.72 | 113.6 | E121 | - | - |
| E45 | 7.98 | 116.0 | N70 | 8.90 | 120.0 | I96 | 7.99 | 121.6 | L123 | 8.69 | 120.4 |
| S46 | 8.01 | 117.3 | N71 | - | - | N97 | 8.42 | 129.4 | H125 | - | - |
| A47 | 7.15 | 122.5 | Y73 | 9.55 | 126.1 | G99 | 8.24 | 107.0 | M127 | 7.87 | 117.0 |
| G48 | 6.24 | 106.9 | S74 | 8.26 | 115.7 | M100 | 7.48 | 117.9 | G128 | 7.52 | 107.0 |
| L49 | 7.87 | 121.2 | H75 | - | - | L101 | 9.11 | 125.4 | I129 | 7.56 | 119.7 |
| R51 | 8.80 | 123.1 | T76 | 8.80 | 119.2 | H103 | 9.67 | 113.9 | H130 | 8.46 | 123.1 |
| D52 | 8.96 | 117.8 | Q77 | 7.95 | 121.2 | E104 | 7.74 | 121.1 | A131 | 8.27 | 126.1 |
| Q53 | 7.72 | 123.8 | R78 | 7.46 | 118.7 | N105 | 7.36 | 116.5 | D132 | 8.42 | 120.5 |
| H54 | 8.73 | 119.6 | T79 | - | - | V106 | 7.98 | 118.5 | I133 | 7.62 | 124.8 |
| E55 | 8.83 | 121.2 | L80 | 8.62 | 124.9 | D108 | 8.24 | 121.4 | | | |
| F56 | 7.93 | 120.9 | D81 | - | - | M109 | 8.61 | 120.1 | | | |

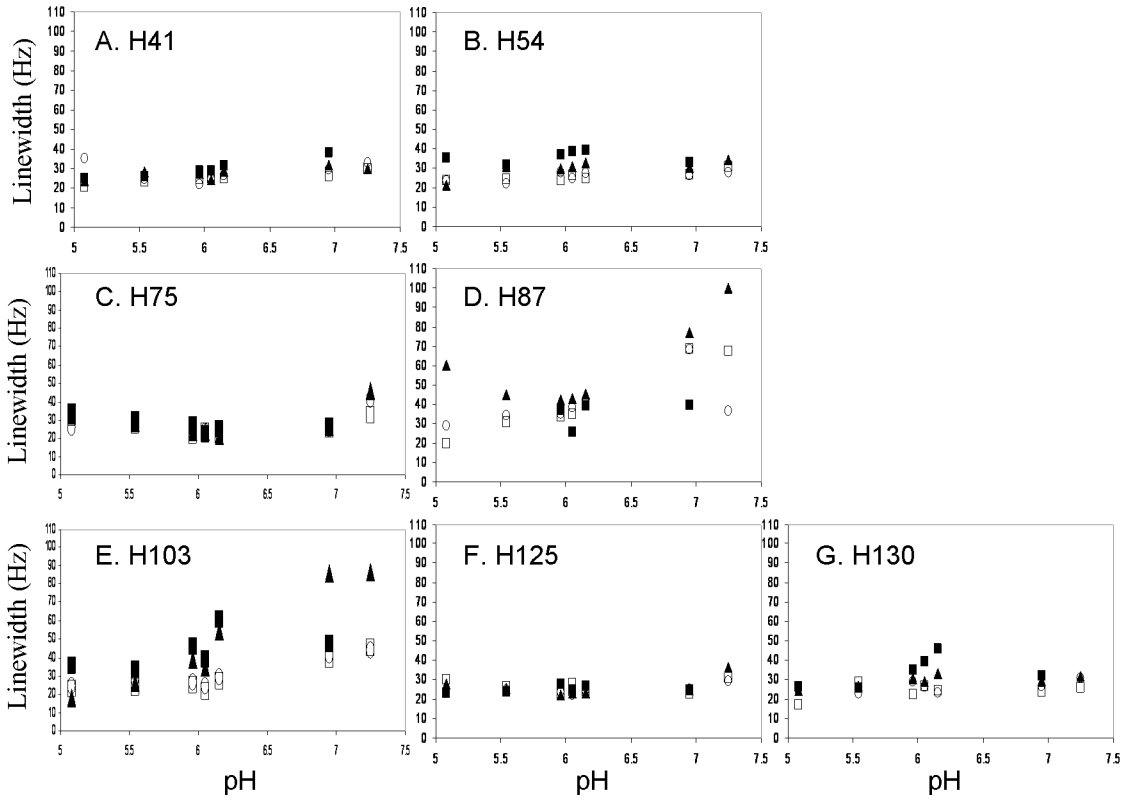
[a] Chemical shifts are measure in parts-per-million (ppm)

| pH 5.0 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ |
| T32 | 8.01 | 117.3 | V57 | 8.58 | 120.9 | A83 | 8.22 | 124.8 | R110 | 8.97 | 119.3 |
| S33 | 9.27 | 118.7 | S58 | 7.36 | 113.2 | D84 | - | - | S111 | 8.09 | 116.4 |
| S34 | 8.14 | 116.8 | S59 | 7.53 | 112.4 | M85 | - | - | A112 | 7.82 | 124.3 |
| Q35 | 7.84 | 125.4 | Q60 | - | - | Q86 | 8.66 | 121.7 | I113 | 8.56 | 121.3 |
| L37 | - | - | A61 | 7.05 | 121.6 | H87 | 9.06 | 120.5 | T114 | 7.77 | 118.4 |
| S38 | 8.04 | 116.1 | Q63 | 8.95 | 124.6 | R88 | 8.38 | 118.4 | W116 ε1 | 10.48 | 129.4 |
| V39 | 7.94 | 123.4 | S64 | 8.80 | 113.3 | Y89 | 8.00 | 121.6 | W116 | 8.80 | 123.1 |
| R40 | 8.16 | 122.9 | L65 | 7.01 | 124.2 | M90 | 9.30 | 119.6 | S117 | 8.10 | 118.0 |
| H41 | 8.39 | 118.0 | R66 | - | - | T91 | 7.39 | 105.3 | D118 | - | - |
| Q42 | - | - | N67 | 8.87 | 118.0 | G92 | 7.49 | 110.1 | M119 | 8.53 | 123.1 |
| L43 | 8.80 | 123.1 | R68 | 7.59 | 121.2 | S94 | 7.37 | 110.6 | R120 | - | - |
| A44 | - | - | Y69 | - | - | G95 | 8.71 | 113.6 | E121 | - | - |
| E45 | 7.97 | 116.2 | N70 | 8.90 | 120.1 | I96 | 7.99 | 121.7 | L123 | 8.69 | 120.4 |
| S46 | 8.01 | 117.3 | N71 | - | - | N97 | 8.43 | 129.4 | H125 | - | - |
| A47 | 7.15 | 122.5 | Y73 | 9.55 | 126.1 | G99 | 8.23 | 107.1 | M127 | 7.87 | 117.0 |
| G48 | 6.25 | 107.0 | S74 | 8.26 | 115.7 | M100 | 7.48 | 117.9 | G128 | 7.52 | 107.0 |
| L49 | 7.86 | 121.4 | H75 | 8.52 | 120.9 | L101 | 9.14 | 125.4 | I129 | 7.55 | 119.7 |
| R51 | 8.80 | 123.1 | T76 | 8.80 | 119.2 | H103 | 9.65 | 113.8 | H130 | 8.45 | 123.1 |
| D52 | 8.95 | 117.8 | Q77 | 7.94 | 121.1 | E104 | 7.73 | 121.1 | A131 | 8.27 | 126.1 |
| Q53 | 7.72 | 123.8 | R78 | 7.45 | 118.7 | N105 | 7.35 | 116.5 | D132 | 8.42 | 120.6 |
| H54 | 8.72 | 119.6 | T79 | - | - | V106 | 7.97 | 118.4 | I133 | 7.62 | 124.9 |
| E55 | 8.83 | 121.3 | L80 | 8.63 | 124.9 | D108 | 8.23 | 121.4 | | | |
| F56 | 7.92 | 120.9 | D81 | - | - | M109 | 8.62 | 120.1 | | | |
| pH 5.1 | | | | | | | | | | | |
| Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ |
| T32 | 8.00 | 117.3 | V57 | 8.58 | 120.9 | A83 | 8.22 | 124.8 | R110 | 8.98 | 119.3 |
| S33 | 9.27 | 118.7 | S58 | 7.36 | 113.2 | D84 | - | - | S111 | 8.09 | 116.4 |
| S34 | 8.14 | 116.8 | S59 | 7.54 | 112.4 | M85 | - | - | A112 | 7.82 | 124.3 |
| Q35 | 7.84 | 125.3 | Q60 | - | - | Q86 | 8.66 | 121.7 | I113 | 8.57 | 121.3 |
| L37 | - | - | A61 | 7.06 | 121.6 | H87 | 9.06 | 120.6 | T114 | 7.77 | 118.4 |
| S38 | 8.04 | 116.1 | Q63 | 8.95 | 124.6 | R88 | 8.38 | 118.4 | W116 ε1 | 10.48 | 129.4 |
| V39 | 7.94 | 123.4 | S64 | 8.81 | 113.3 | Y89 | 8.00 | 121.6 | W116 | 8.80 | 123.1 |
| R40 | 8.16 | 123.0 | L65 | 7.01 | 124.2 | M90 | 9.30 | 119.7 | S117 | 8.10 | 118.0 |
| H41 | 8.39 | 118.0 | R66 | - | - | T91 | 7.41 | 105.3 | D118 | - | - |
| Q42 | - | - | N67 | 8.87 | 118.0 | G92 | 7.49 | 110.2 | M119 | 8.53 | 123.1 |
| L43 | 8.80 | 123.1 | R68 | 7.59 | 121.2 | S94 | 7.37 | 110.6 | R120 | - | - |
| A44 | - | - | Y69 | - | - | G95 | 8.71 | 113.6 | E121 | - | - |
| E45 | 7.97 | 116.2 | N70 | 8.90 | 120.1 | I96 | 7.99 | 121.7 | L123 | 8.69 | 120.4 |
| S46 | 8.00 | 117.3 | N71 | - | - | N97 | 8.42 | 129.4 | H125 | - | - |
| A47 | 7.15 | 122.4 | Y73 | 9.56 | 126.1 | G99 | 8.23 | 107.1 | M127 | 7.86 | 117.0 |
| G48 | 6.28 | 107.0 | S74 | 8.26 | 115.7 | M100 | 7.48 | 117.9 | G128 | 7.51 | 107.0 |
| L49 | 7.86 | 121.4 | H75 | 8.51 | 120.9 | L101 | 9.15 | 125.5 | I129 | 7.55 | 119.7 |
| R51 | 8.80 | 123.1 | T76 | 8.80 | 119.2 | H103 | 9.65 | 113.8 | H130 | 8.45 | 123.2 |
| D52 | 8.95 | 117.8 | Q77 | 7.94 | 121.1 | E104 | 7.72 | 121.2 | A131 | 8.27 | 126.2 |
| Q53 | 7.72 | 123.8 | R78 | 7.45 | 118.7 | N105 | 7.35 | 116.5 | D132 | 8.41 | 120.6 |
| H54 | 8.72 | 119.7 | T79 | - | - | V106 | 7.97 | 118.4 | I133 | 7.61 | 124.9 |
| E55 | 8.84 | 121.3 | L80 | 8.62 | 124.9 | D108 | 8.23 | 121.4 | | | |
| F56 | 7.92 | 120.9 | D81 | - | - | M109 | 8.62 | 120.1 | | | |

| pH 5.5 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N |
| T32 | 8.01 | 117.3 | V57 | 8.57 | 120.9 | A83 | 8.22 | 124.7 | R110 | 9.00 | 119.3 |
| S33 | 9.30 | 118.8 | S58 | 7.37 | 113.2 | D84 | 8.46 | 121.3 | S111 | 8.08 | 116.2 |
| S34 | 8.10 | 116.9 | S59 | 7.54 | 112.3 | M85 | 8.09 | 120.9 | A112 | 7.82 | 124.3 |
| Q35 | 7.82 | 125.3 | Q60 | 7.93 | 120.4 | Q86 | 8.64 | 121.8 | I113 | 8.58 | 121.3 |
| L37 | 8.10 | 118.6 | A61 | 7.06 | 121.6 | H87 | 9.07 | 121.2 | T114 | 7.77 | 118.3 |
| S38 | 8.02 | 115.8 | Q63 | 8.95 | 124.6 | R88 | 8.39 | 118.1 | W116 ε1 | 10.5 | 129.4 |
| V39 | 7.92 | 123.4 | S64 | 8.81 | 113.3 | Y89 | 7.99 | 121.7 | W116 | 8.79 | 123.1 |
| R40 | 8.13 | 123.5 | L65 | 7.02 | 124.2 | M90 | 9.35 | 119.9 | S117 | 8.09 | 118.0 |
| H41 | 8.39 | 118.1 | R66 | 7.85 | 121.1 | T91 | 7.46 | 105.4 | D118 | 8.30 | 122.9 |
| Q42 | 8.00 | 120.4 | N67 | 8.87 | 118.0 | G92 | 7.48 | 110.2 | M119 | 8.52 | 123.2 |
| L43 | 8.79 | 123.1 | R68 | 7.58 | 121.2 | S94 | 7.38 | 110.6 | R120 | 8.58 | 119.7 |
| A44 | 8.37 | 121.5 | Y69 | 8.42 | 121.5 | G95 | 8.72 | 113.6 | E121 | 7.86 | 120.0 |
| E45 | 7.98 | 116.3 | N70 | 8.90 | 120.0 | I96 | 7.99 | 121.7 | L123 | 8.70 | 120.5 |
| S46 | 8.01 | 117.3 | N71 | 8.36 | 119.9 | N97 | 8.42 | 129.5 | H125 | 8.18 | 116.8 |
| A47 | 7.14 | 122.4 | Y73 | 9.56 | 126.1 | G99 | 8.23 | 107.1 | M127 | 7.87 | 117.0 |
| G48 | 6.29 | 107.0 | S74 | 8.26 | 115.7 | M100 | 7.48 | 118.0 | G128 | 7.51 | 107.0 |
| L49 | 7.88 | 121.6 | H75 | 8.51 | 121.2 | L101 | 9.22 | 125.5 | I129 | 7.55 | 119.7 |
| R51 | 8.79 | 123.1 | T76 | 8.80 | 119.3 | H103 | 9.55 | 113.6 | H130 | 8.43 | 123.2 |
| D52 | 8.94 | 117.7 | Q77 | 7.94 | 121.1 | E104 | 7.69 | 121.3 | A131 | 8.28 | 126.2 |
| Q53 | 7.71 | 123.8 | R78 | 7.45 | 118.7 | N105 | 7.34 | 116.3 | D132 | 8.40 | 120.6 |
| H54 | 8.70 | 120.0 | T79 | 8.19 | 118.4 | V106 | 7.96 | 118.4 | I133 | 7.61 | 125.1 |
| E55 | 8.83 | 121.3 | L80 | 8.63 | 125.0 | D108 | 8.21 | 121.5 | | | |
| F56 | 7.91 | 120.8 | D81 | 8.09 | 120.3 | M109 | 8.63 | 120.1 | | | |
| pH 6.0 | | | | | | | | | | | |
| Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N |
| T32 | 8.01 | 117.3 | V57 | 8.59 | 121.1 | A83 | 8.21 | 124.7 | R110 | 9.00 | 119.3 |
| S33 | 9.31 | 118.8 | S58 | 7.39 | 113.2 | D84 | 8.49 | 121.4 | S111 | 8.09 | 116.1 |
| S34 | 8.07 | 117.0 | S59 | 7.54 | 112.4 | M85 | 8.10 | 120.8 | A112 | 7.81 | 124.3 |
| Q35 | 7.81 | 125.3 | Q60 | 7.94 | 120.4 | Q86 | 8.62 | 121.9 | I113 | 8.59 | 121.3 |
| L37 | 8.11 | 118.8 | A61 | 7.05 | 121.6 | H87 | 9.08 | 121.5 | T114 | 7.78 | 118.4 |
| S38 | 7.99 | 115.4 | Q63 | 8.95 | 124.6 | R88 | 8.38 | 117.9 | W116 ε1 | 10.50 | 129.5 |
| V39 | 7.90 | 123.4 | S64 | 8.81 | 113.3 | Y89 | 7.99 | 121.7 | W116 | 8.79 | 123.1 |
| R40 | 8.11 | 123.6 | L65 | 7.02 | 124.2 | M90 | 9.37 | 120.0 | S117 | 8.09 | 117.9 |
| H41 | 8.41 | 118.3 | R66 | 7.85 | 121.1 | T91 | 7.49 | 105.5 | D118 | 8.30 | 122.9 |
| Q42 | 7.99 | 120.5 | N67 | 8.86 | 118.0 | G92 | 7.48 | 110.2 | M119 | 8.51 | 123.0 |
| L43 | 8.79 | 123.1 | R68 | 7.57 | 121.2 | S94 | 7.38 | 110.6 | R120 | 8.58 | 119.7 |
| A44 | 8.37 | 121.5 | Y69 | 8.42 | 121.4 | G95 | 8.71 | 113.6 | E121 | 7.89 | 120.1 |
| E45 | 8.00 | 116.5 | N70 | 8.90 | 120.0 | I96 | 7.99 | 121.7 | L123 | 8.69 | 120.4 |
| S46 | 8.01 | 117.3 | N71 | 8.34 | 119.9 | N97 | 8.42 | 129.5 | H125 | 8.16 | 117.1 |
| A47 | 7.13 | 122.3 | Y73 | 9.56 | 126.2 | G99 | 8.22 | 107.2 | M127 | 7.88 | 116.9 |
| G48 | 6.29 | 107.0 | S74 | 8.25 | 115.8 | M100 | 7.49 | 118.0 | G128 | 7.52 | 107.0 |
| L49 | 7.90 | 121.7 | H75 | 8.49 | 121.4 | L101 | 9.24 | 125.6 | I129 | 7.54 | 119.6 |
| R51 | 8.79 | 123.1 | T76 | 8.81 | 119.4 | H103 | 9.33 | 113.5 | H130 | 8.39 | 123.3 |
| D52 | 8.92 | 117.5 | Q77 | 7.93 | 121.1 | E104 | 7.63 | 121.6 | A131 | 8.25 | 126.2 |
| Q53 | 7.68 | 123.8 | R78 | 7.47 | 118.7 | N105 | 7.33 | 116.2 | D132 | 8.38 | 120.6 |
| H54 | 8.66 | 120.3 | T79 | 8.24 | 118.6 | V106 | 7.95 | 118.3 | I133 | 7.60 | 125.1 |
| E55 | 8.82 | 121.2 | L80 | 8.63 | 125.0 | D108 | 8.22 | 121.5 | | | |
| F56 | 7.88 | 120.8 | D81 | 8.08 | 120.2 | M109 | 8.65 | 120.2 | | | |

| pH 6.1 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N |
| T32 | 8.00 | 117.2 | V57 | 8.60 | 121.1 | A83 | 8.20 | 124.7 | R110 | 8.99 | 119.3 |
| S33 | 9.31 | 118.8 | S58 | 7.39 | 113.2 | D84 | 8.49 | 121.4 | S111 | 8.09 | 116.1 |
| S34 | 8.06 | 117.0 | S59 | 7.54 | 112.3 | M85 | 8.09 | 120.8 | A112 | 7.80 | 124.2 |
| Q35 | 7.81 | 125.3 | Q60 | 7.93 | 120.4 | Q86 | 8.61 | 121.9 | I113 | 8.58 | 121.3 |
| L37 | 8.11 | 118.8 | A61 | 7.04 | 121.6 | H87 | 9.08 | 121.6 | T114 | 7.78 | 118.3 |
| S38 | 7.99 | 115.4 | Q63 | 8.94 | 124.5 | R88 | 8.37 | 117.9 | W116 ε1 | 10.50 | 129.5 |
| V39 | 7.90 | 123.4 | S64 | 8.80 | 113.3 | Y89 | 7.97 | 121.5 | W116 | 8.78 | 123.0 |
| R40 | 8.10 | 123.6 | L65 | 7.01 | 124.2 | M90 | 9.37 | 120.0 | S117 | 8.07 | 117.9 |
| H41 | 8.40 | 118.4 | R66 | 7.84 | 121.1 | T91 | 7.49 | 105.4 | D118 | 8.31 | 122.9 |
| Q42 | 7.98 | 120.5 | N67 | 8.86 | 118.0 | G92 | 7.47 | 110.2 | M119 | 8.51 | 122.9 |
| L43 | 8.79 | 123.2 | R68 | 7.58 | 121.2 | S94 | 7.37 | 110.6 | R120 | 8.57 | 119.7 |
| A44 | 8.37 | 121.5 | Y69 | 8.41 | 121.4 | G95 | 8.71 | 113.6 | E121 | 7.89 | 120.2 |
| E45 | 8.00 | 116.4 | N70 | 8.89 | 120.0 | I96 | 7.98 | 121.7 | L123 | 8.69 | 120.4 |
| S46 | 8.00 | 117.2 | N71 | 8.33 | 119.8 | N97 | 8.41 | 129.5 | H125 | 8.14 | 117.1 |
| A47 | 7.12 | 122.4 | Y73 | 9.55 | 126.2 | G99 | 8.22 | 107.1 | M127 | 7.88 | 116.9 |
| G48 | 6.24 | 106.9 | S74 | 8.24 | 115.8 | M100 | 7.48 | 118.0 | G128 | 7.52 | 107.0 |
| L49 | 7.90 | 121.6 | H75 | 8.50 | 121.5 | L101 | 9.24 | 125.6 | I129 | 7.53 | 119.6 |
| R51 | 8.78 | 123.0 | T76 | 8.80 | 119.4 | H103 | 9.31 | 113.5 | H130 | 8.38 | 123.3 |
| D52 | 8.92 | 117.5 | Q77 | 7.92 | 121.1 | E104 | 7.62 | 121.6 | A131 | 8.24 | 126.1 |
| Q53 | 7.67 | 123.7 | R78 | 7.48 | 118.7 | N105 | 7.33 | 116.2 | D132 | 8.37 | 120.5 |
| H54 | 8.66 | 120.4 | T79 | 8.24 | 118.5 | V106 | 7.94 | 118.3 | I133 | 7.59 | 125.1 |
| E55 | 8.81 | 121.2 | L80 | 8.62 | 125.0 | D108 | 8.22 | 121.5 | | | |
| F56 | 7.88 | 120.8 | D81 | 8.07 | 120.2 | M109 | 8.64 | 120.2 | | | |
| pH 6.2 | | | | | | | | | | | |
| Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N | Group | $\omega^1$H | $\omega^{15}$N |
| T32 | 8.00 | 117.2 | V57 | 8.61 | 121.2 | A83 | 8.19 | 124.7 | R110 | 9.00 | 119.4 |
| S33 | 9.31 | 118.8 | S58 | 7.40 | 113.2 | D84 | 8.50 | 121.3 | S111 | 8.10 | 116.1 |
| S34 | 8.06 | 117.0 | S59 | 7.53 | 112.3 | M85 | 8.09 | 120.8 | A112 | 7.79 | 124.2 |
| Q35 | 7.81 | 125.3 | Q60 | 7.94 | 120.4 | Q86 | 8.60 | 121.9 | I113 | 8.59 | 121.3 |
| L37 | 8.11 | 118.8 | A61 | 7.04 | 121.6 | H87 | 9.08 | 121.7 | T114 | 7.79 | 118.3 |
| S38 | 7.97 | 115.2 | Q63 | 8.94 | 124.5 | R88 | 8.38 | 117.9 | W116 ε1 | 10.51 | 129.5 |
| V39 | 7.88 | 123.4 | S64 | 8.80 | 113.3 | Y89 | 7.97 | 121.5 | W116 | 8.78 | 123.0 |
| R40 | 8.09 | 123.7 | L65 | 7.02 | 124.2 | M90 | 9.38 | 120.0 | S117 | 8.08 | 117.9 |
| H41 | 8.40 | 118.5 | R66 | 7.84 | 121.1 | T91 | 7.50 | 105.4 | D118 | 8.30 | 122.9 |
| Q42 | 7.98 | 120.5 | N67 | 8.86 | 118.0 | G92 | 7.47 | 110.2 | M119 | 8.51 | 122.9 |
| L43 | 8.79 | 123.3 | R68 | 7.57 | 121.2 | S94 | 7.37 | 110.6 | R120 | 8.57 | 119.7 |
| A44 | 8.37 | 121.5 | Y69 | 8.41 | 121.4 | G95 | 8.71 | 113.6 | E121 | 7.89 | 120.2 |
| E45 | 8.02 | 116.5 | N70 | 8.89 | 119.9 | I96 | 7.98 | 121.7 | L123 | 8.69 | 120.4 |
| S46 | 8.00 | 117.2 | N71 | 8.32 | 119.8 | N97 | 8.41 | 129.5 | H125 | 8.14 | 117.3 |
| A47 | 7.11 | 122.3 | Y73 | 9.55 | 126.2 | G99 | 8.22 | 107.1 | M127 | 7.89 | 116.9 |
| G48 | 6.25 | 106.9 | S74 | 8.24 | 115.8 | M100 | 7.48 | 118.0 | G128 | 7.52 | 107.1 |
| L49 | 7.91 | 121.7 | H75 | 8.50 | 121.8 | L101 | 9.25 | 125.6 | I129 | 7.52 | 119.5 |
| R51 | 8.78 | 123.0 | T76 | 8.81 | 119.6 | H103 | 9.14 | 113.4 | H130 | 8.34 | 123.3 |
| D52 | 8.92 | 117.4 | Q77 | 7.92 | 121.1 | E104 | 7.57 | 121.8 | A131 | 8.22 | 126.0 |
| Q53 | 7.66 | 123.7 | R78 | 7.49 | 118.7 | N105 | 7.32 | 116.1 | D132 | 8.36 | 120.5 |
| H54 | 8.64 | 120.6 | T79 | 8.23 | 118.5 | V106 | 7.93 | 118.2 | I133 | 7.59 | 125.1 |
| E55 | 8.80 | 121.1 | L80 | 8.62 | 125.0 | D108 | 8.23 | 121.5 | | | |
| F56 | 7.87 | 120.8 | D81 | 8.07 | 120.2 | M109 | 8.65 | 120.3 | | | |

| pH 7.0 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ |
| T32 | 8.01 | 117.1 | V57 | 8.64 | 121.4 | A83 | 8.17 | 124.8 | R110 | 9.00 | 119.5 |
| S33 | 9.32 | 118.8 | S58 | 7.41 | 113.2 | D84 | 8.53 | 121.3 | S111 | 8.11 | 116.0 |
| S34 | 8.04 | 117.1 | S59 | 7.52 | 112.4 | M85 | 8.08 | 120.8 | A112 | 7.76 | 124.1 |
| Q35 | 7.79 | 125.3 | Q60 | 7.93 | 120.4 | Q86 | 8.59 | 122.0 | I113 | 8.60 | 121.4 |
| L37 | 8.09 | 119.0 | A61 | 7.02 | 121.5 | H87 | 9.10 | 121.8 | T114 | 7.80 | 118.4 |
| S38 | 7.93 | 114.5 | Q63 | 8.95 | 124.5 | R88 | 8.37 | 117.8 | W116 ε1 | 10.53 | 129.5 |
| V39 | 7.83 | 123.3 | S64 | 8.81 | 113.3 | Y89 | 7.96 | 121.5 | W116 | 8.77 | 123.0 |
| R40 | 8.03 | 123.9 | L65 | 7.02 | 124.2 | M90 | 9.38 | 120.0 | S117 | 8.08 | 117.9 |
| H41 | 8.39 | 118.9 | R66 | 7.83 | 121.1 | T91 | 7.50 | 105.5 | D118 | 8.28 | 122.9 |
| Q42 | 7.95 | 120.7 | N67 | 8.85 | 117.9 | G92 | 7.46 | 110.2 | M119 | 8.52 | 122.9 |
| L43 | 8.79 | 123.5 | R68 | 7.54 | 121.1 | S94 | 7.37 | 110.6 | R120 | 8.58 | 119.6 |
| A44 | 8.37 | 121.5 | Y69 | 8.41 | 121.4 | G95 | 8.71 | 113.6 | E121 | 7.89 | 120.2 |
| E45 | 8.06 | 116.7 | N70 | 8.89 | 119.9 | I96 | 7.98 | 121.7 | L123 | 8.68 | 120.4 |
| S46 | 8.01 | 117.4 | N71 | 8.30 | 119.9 | N97 | 8.41 | 129.5 | H125 | 8.09 | 118.0 |
| A47 | 7.08 | 122.2 | Y73 | 9.55 | 126.3 | G99 | 8.20 | 107.1 | M127 | 7.93 | 117.0 |
| G48 | 6.30 | 106.9 | S74 | 8.22 | 115.9 | M100 | 7.48 | 118.0 | G128 | 7.53 | 107.2 |
| L49 | 7.92 | 122.0 | H75 | 8.47 | 122.5 | L101 | 9.27 | 125.6 | I129 | 7.49 | 119.2 |
| R51 | 8.77 | 123.0 | T76 | 8.82 | 119.8 | H103 | 8.69 | 113.1 | H130 | 8.33 | 123.1 |
| D52 | 8.90 | 117.2 | Q77 | 7.91 | 121.2 | E104 | 7.44 | 122.3 | A131 | 8.12 | 125.7 |
| Q53 | 7.63 | 123.7 | R78 | 7.52 | 118.7 | N105 | 7.30 | 115.8 | D132 | 8.33 | 120.5 |
| H54 | 8.59 | 121.0 | T79 | 8.19 | 118.4 | V106 | 7.91 | 118.0 | I133 | 7.57 | 125.2 |
| E55 | 8.78 | 121.0 | L80 | 8.64 | 125.1 | D108 | 8.27 | 121.5 | | | |
| F56 | 7.84 | 120.7 | D81 | 8.06 | 120.1 | M109 | 8.67 | 120.5 | | | |
| pH 7.3 | | | | | | | | | | | |
| Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ | Group | $\omega^1H$ | $\omega^{15}N$ |
| T32 | 8.01 | 117.1 | V57 | 8.66 | 121.4 | A83 | 8.17 | 124.8 | R110 | 9.01 | 119.5 |
| S33 | 9.33 | 118.8 | S58 | 7.42 | 113.2 | D84 | 8.54 | 121.3 | S111 | 8.13 | 116.0 |
| S34 | 8.05 | 117.1 | S59 | 7.52 | 112.4 | M85 | 8.10 | 120.8 | A112 | 7.77 | 124.1 |
| Q35 | 7.80 | 125.3 | Q60 | 7.94 | 120.5 | Q86 | 8.59 | 122.0 | I113 | 8.61 | 121.4 |
| L37 | 8.11 | 119.1 | A61 | 7.02 | 121.5 | H87 | 9.11 | 121.8 | T114 | 7.81 | 118.4 |
| S38 | 7.93 | 114.4 | Q63 | 8.94 | 124.6 | R88 | 8.38 | 117.8 | W116 ε1 | 10.55 | 129.5 |
| V39 | 7.83 | 123.3 | S64 | 8.81 | 113.3 | Y89 | 7.97 | 121.6 | W116 | 8.77 | 123.1 |
| R40 | 8.03 | 124.0 | L65 | 7.02 | 124.2 | M90 | 9.39 | 120.1 | S117 | 8.07 | 117.9 |
| H41 | 8.40 | 119.0 | R66 | 7.84 | 121.1 | T91 | 7.52 | 105.5 | D118 | 8.29 | 122.9 |
| Q42 | 7.96 | 120.8 | N67 | 8.86 | 118.0 | G92 | 7.47 | 110.2 | M119 | 8.53 | 123.0 |
| L43 | 8.81 | 123.5 | R68 | 7.56 | 121.1 | S94 | 7.38 | 110.6 | R120 | 8.58 | 119.6 |
| A44 | 8.38 | 121.5 | Y69 | 8.42 | 121.5 | G95 | 8.72 | 113.6 | E121 | 7.89 | 120.2 |
| E45 | 8.08 | 116.8 | N70 | 8.90 | 119.9 | I96 | 7.99 | 121.7 | L123 | 8.69 | 120.5 |
| S46 | 8.02 | 117.4 | N71 | 8.29 | 119.9 | N97 | 8.42 | 129.5 | H125 | 8.09 | 118.3 |
| A47 | 7.09 | 122.2 | Y73 | 9.56 | 126.4 | G99 | 8.22 | 107.2 | M127 | 7.99 | 117.1 |
| G48 | 6.28 | 106.9 | S74 | 8.23 | 116.0 | M100 | 7.49 | 118.0 | G128 | 7.55 | 107.2 |
| L49 | 7.95 | 122.0 | H75 | 8.49 | 122.7 | L101 | 9.27 | 125.6 | I129 | 7.47 | 119.1 |
| R51 | 8.77 | 123.1 | T76 | 8.84 | 120.0 | H103 | 8.59 | 113.1 | H130 | 8.33 | 123.1 |
| D52 | 8.91 | 117.2 | Q77 | 7.91 | 121.2 | E104 | 7.43 | 122.4 | A131 | 8.07 | 125.5 |
| Q53 | 7.63 | 123.7 | R78 | 7.55 | 118.7 | N105 | 7.31 | 115.8 | D132 | 8.34 | 120.5 |
| H54 | 8.60 | 121.1 | T79 | 8.19 | 118.4 | V106 | 7.93 | 118.0 | I133 | 7.56 | 125.2 |
| E55 | 8.78 | 121.1 | L80 | 8.65 | 125.1 | D108 | 8.30 | 121.5 | | | |
| F56 | 7.84 | 120.7 | D81 | 8.07 | 120.0 | M109 | 8.69 | 120.6 | | | |




Appendix Figure 3.2: **Lorentzian $^{15}$N linewidths of His sidechains**. The $^{15}$N linewidths of H87 and H103 broaden with increasing pH, consistent with His tautomer exchange on the intermediate timescale. The other His display tautomer exchange on either the slow exchange (H130) or the fast exchange timescales.