Spring 2011

Bioinformatics, Thermodynamics and Kinetics Analysis of an All Alpha Helical Protein with a Gree-Key Topology

Hai Li
Old Dominion University

BIOINFORMATICS, THERMODYNAMICS AND KINETICS ANALYSIS OF AN ALL ALPHA HELICAL PROTEIN WITH A GREEK-KEY TOPOLOGY

by

Hai Li
B.S. July 2001, Central China Normal University, China
M.S. July 2004, Nanjing University, China
M.S. August 2006, Indiana State University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

CHEMISTRY

OLD DOMINION UNIVERSITY
May 2011

Approved by:

Lesley H. Greene (Director)

Patricia Pleban (Member)

Jennifer Poutsma (Member)

Christopher Osgood (Member)

ABSTRACT

BIOINFORMATICS, THERMODYNAMICS AND KINETICS ANALYSIS OF AN ALL ALPHA HELICAL PROTEIN WITH A GREEK-KEY TOPOLOGY

Hai Li
Old Dominion University, 2011
Director: Dr. Lesley H. Greene

Computational and experimental studies focusing on the role of conserved residues for folding and stability is an active and promising area of research. To further expand our understanding we present the results of a bioinformatics analysis of the death domain superfamily. The death domain superfamily fold consists of six α-helices arranged in a Greek-key topology, which is shared by the all β-sheet immunoglobulin and mixed α/β-plait superfamilies. Our sequence and structural studies have identified a group of conserved hydrophobic residues and corresponding long-range interactions, which we propose are important in the formation and stabilization of the hydrophobic core and native topology. Equilibrium unfolding and refolding studies of a model superfamily member, the Fas-associated death domain protein indicate that this process is cooperative, two-state and reversible. Stopped-flow fluorescence studies reveal that the folding is rapid and biphasic with the majority of the hydrophobic core forming in the first phase. Site-directed mutagenesis studies indicate that conserved Trp112, Trp148, Leu115 and Val121 are important to structure, native state stability and folding.

We also present the results of experiments aimed at characterizing the formation of secondary structure. Stopped-flow far-UV CD spectroscopy revealed that the folding process was monophasic and the rate is 23.4 s^{-1}. To gain atomic resolution a combination of quenched-flow methods, hydrogen deuterium exchange (HX) and NMR spectroscopy

was implemented. Twenty-two amide hydrogens in the backbone of the helices and two in the backbone of the loops were monitored and the folding of all six helices was determined to be monophasic with rates between 19 s^{-1} and 22 s^{-1}. These results indicate that the formation of secondary structure is largely cooperative and concomitant with the hydrophobic collapse. Additional insights are gained by calculating the exchange rates of twenty-three residues from equilibrium HX experiments. The majority of protected amide protons are found on helices 2, 4, and 5 which make up core structural elements of the Greek-key topology. These results appear to be the earliest conservation analysis and biophysical characterization conducted on the Fas-associated death domain and folding kinetics using quenched-flow combined with NMR spectroscopy on an all α-helical Greek-key protein.

This thesis is dedicated to my parents, Zhiheng Li and Yiqing Liu

To my wife, Guoxia Zhao

To my daughter, Heidy Li

ACKNOWLEDGMENTS

There are many people who have contributed to the successful completion of this dissertation.

I would first like to express my appreciation to my advisor Dr. Lesley H. Greene for allowing me to conduct my research in her laboratory and teaching me theory and experiments. Her direction, untiring effort, and full support directly led me to succeed in this extraordinary project. I also thank all the group members in Dr. Greene's lab.

I would like to thank my committee members Dr. Patricia Pleban, Dr. Jennifer Poutsma, and Dr. Christopher Osgood for their advice and guidance throughout my years of research, and editing of this manuscript.

I would like to thank my parents, Zhiheng Li and Yiqing Liu, my wife, Guoxia Zhao, and my daughter, Heidy Li for their sacrifice throughout the years.

For the work of conservation analysis, stability, structure, and folding of Fadd-DD, I express our sincere thanks to Professor Paul Driscoll (University College London, UK) for donation of cDNA encoding Fadd-DD; to Pierre Guillemin (Bio-Logic, France) for technical assistance with the experimental studies; to Ruben Igloria and Trisha Ahmed (Old Dominion University, VA) for work on the conservation and network computer programs written in Perl Script; to Alicia Herr and Tammy Subotich for expert assistance in setting up the Greene lab; and to Susan Hatcher for Mass Spectroscopy analysis of the recombinant proteins (COSMIC facility, Old Dominion University).

For the work on the chitinase insertion domain, I express my sincere thanks to Dr. Janet M. Moloney (Old Dominion University) and Jessica Wojtaszek (Old Dominion

University) for critically reading the manuscript. I also thank the reviewers for insightful comments which helped improve the quality of this manuscript.

For the work on the quenched-flow, hydrogen exchange, and NMR, I express my sincere thanks to Dr. Junyan Zhong (COSMIC Facility, Old Dominion University) and John Sonewald (Bio-Logic USA, TN) for technical assistance with the experimental studies. I am grateful to Susan Hatcher for the NMR time.

I would not have been able to obtain my Ph.D. degree without many people's help and contribution.

NOMENCLATURE

β-ME	β-mercaptoethanol
CARD	Caspase recruitment domain
CD	Circular dichroism
CID	Chitinase insertion domain
DD	Death domain
DED	Death effector domain
DTT	Dithiothreitol
Fadd-DD	Fas-associated death domain protein
GH	Glycoside hydrolases
GndHCl	Guanidine hydrochloride
HSQC	Heteronuclear Single Quantum Coherence
HX	Hydrogen deuterium exchange
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MWM	Molecular weight marker
NAG	N-acetylglucosamine
NAGase	β-(1,4)-N-acetyl-glucosaminidase
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance
O.D.	Optical density
PCR	Polymerase chain reaction
PYD	Pyrin domain

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TS	Transition state
WT	Wild-type

TABLE OF CONTENTS

Page

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER I

INTRODUCTION

Proteins consist of a chain of amino acids which are linked together by peptide bonds formed between the amino group of one residue and the carboxyl group of the neighboring residue. There are twenty types of naturally-occurring amino acids. The primary structure of a protein is its amino acid sequence. The secondary structure of a protein is its specific regularly repeating local structures including α-helices, β-strands, and β-turns. The major force forming and stabilizing secondary structure is hydrogen bonding. The tertiary structure of a protein is its overall topology composed of α-helices and β-strands, such as all α-helices, all β-sheets, α/β, and α + β. The forces involved in forming and stabilizing tertiary structure include hydrophobic interactions, ionic interactions, hydrogen bonding and disulfide bonds. The quaternary structure of a protein is in the form of multi-subunits or chains. There are some common topologies in the protein universe, including Greek-key, TIM barrel and jelly roll (May et al., 2010).

X-ray crystallography and NMR spectroscopy are the two primary methods utilized to solve protein structures. The solved structures of a protein are maintained in the Protein Data Bank (PDB). There are two main protein structure classification databases: Structural Classification of Proteins (SCOP) (Murzin et al., 1995) and CATH Protein Structure Classification (Orengo et al., 1997).

Protein folding remains one of the most fundamental and intriguing questions in the field of structural biology (Kang and Kini, 2009). This question can be simply stated

---

This dissertation follows the format of *Molecular Cell*.

as: "how does a linear polypeptide chain of amino acids accurately fold into a distinct native functional conformation?"

Anfinsen's dogma, also known as the thermodynamic hypothesis, states that, at least for small globular proteins, the native structure of a protein is determined only by its amino acid sequence. This discovery led to Christian Anfinsen's winning the Nobel Prize in 1972. Anfinsen's dogma indicates that amino acids interact with each other to produce a well-defined folded three-dimensional structure known as the native state (Anfinsen, 1972).

The Levinthal paradox states that if a protein with 100 amino acids folds by sampling all possible conformations and each side chain has many possible states, it would take an extremely long time to do so. Even if the conformations were sampled at a rate of 10^{-4} nanosecond, it would still take 10^{27} years to fold (Levinthal, 1968; Zwanzig et al., 1992). Since proteins fold much faster than this, Levinthal then proposed that instead of a random conformational sampling, the protein must fold by a specific folding pathway (Levinthal, 1968).

In the folding process, the thermodynamics can be described as:

$$\Delta G_{total} = \Delta H_{chain} + \Delta H_{solvent} - T\Delta S_{chain} - T\Delta S_{solvent}$$

In the equation, ΔG_{total} is the total Gibbs free energy change between the denatured and native states; ΔH_{chain} and $\Delta H_{solvent}$ represent the enthalpy change for the polypeptide chain and solvent, respectively; and ΔS_{chain} and $\Delta S_{solvent}$ represent entropy change for the polypeptide chain and solvent, respectively (Garrett and Grisham, 2005). Van der Waals forces, hydrophobic interaction, hydrogen bonding and electrostatic interactions are the main driving forces for folding. Comparison of all the terms indicates

that the two largest contributions to the stability of a folded protein are $-T\Delta S_{solvent}$ for the hydrophobic side chains ($\Delta S_{solvent}$ is a positive value) and $\Delta H_{solvent}$ (a negative value) for the polar groups, which makes the total free energy a negative value (Garrett and Grisham, 2005). Therefore, the direction of folding from the unfolded state to the native state goes spontaneously. In the folding process, denatured protein goes through the transition state (TS) to the native state and overcomes a free energy barrier, which controls the folding rates. Plaxco et al. (1998) revealed that the intrinsic folding rate decreases with the increase of relative contact order, which reflects the relative importance of local and non-local contacts to the native structure.

It appears that small proteins (100 amino acids or less) can fold rapidly in a two-state (denatured and native states) mechanism with no kinetically detectable intermediates. However, the folding of larger proteins often involves a multi-stage pathway or folding funnel. There are currently four models for protein folding: framework, hydrophobic collapse, nucleation-condensation, and folding funnel models (Nölting and Andert, 2000; Kelly et al., 2005). In the framework model, protein folding starts with the formation of the secondary structure elements independent of tertiary structure, or at least before tertiary structure is formed (Kim and Baldwin, 1982). These elements then assemble into a tightly packed native tertiary structure either by diffusion and collision (Karplus and Weaver, 1994) or by propagation of structure step by step (Wetlaufer, 1973). The folding of acyl-coenzyme A-binding protein which has four α-helices is an example of a sequential framework model (Kragelund et al., 1999). A bioinformatics study of this protein identified twenty-six conserved residues. Among these residues, the mutation of four hydrophobic residues in helix 1 and three

hydrophobic residues in helix 4 significantly decrease the folding rates. The folding process involves regions of helices 1 and 4 in a spatial arrangement that favors the formation of long-range interactions. The backbone hydrogen bonds of helix 1 and 4 form first, and hence the hydrophobic residues from these two helices are brought close together. Afterwards the structure becomes locked and the folding proceeds by the further formation of the other helices. The generation of native-like interactions between two terminal helices (1 and 4) is the rate-limiting step, so the helix formation rate is much faster and a sequential framework model can describe the folding process (Kragelund et al., 1999).

In the hydrophobic collapse model for folding (Dill, 1990), the initial event of the reaction is a relatively uniform collapse of the protein molecule, mainly driven by the hydrophobic interaction. Stable secondary structure starts to grow only in the collapsed state. The folding of cytochrome c, which has four α-helices and a heme group, is presented as a model of hydrophobic collapse (Akiyama et al., 2000). Initially, the unfolded protein condenses into the compact intermediate I, whose helical content is about 20% of the native state. Compaction of intermediate I facilitates the search for stabilizing contacts to create the molten globule-like intermediate II. Thus, the specific contacts between the helices are established in the TS between intermediates I and II. The final process from intermediate II to the native conformation involves the folding of the remaining helices and the coordination of Met80 to the heme. Since initial hydrophobic collapse happens before much of the helix formation in the folding, it is proposed that folding mechanism is better described as the hydrophobic collapse model. Furthermore, the heme may serve as a hydrophobic nucleation core. In the folding of barstar, the

polypeptide rapidly collapses to compact globule with a non-specific hydrophobic core, without involving the formation of concomitant secondary or tertiary structure (Agashe et al., 1995). The rate constants of the fast and slow phases for the secondary structure formation are 7000 s^{-1} and 11 s^{-1}; and those for the tertiary structure formation are 4000 s^{-1} and 4 s^{-1}, respectively. However, Nölting et al. (1997) suggests that barstar also folds through nucleation-condensation model after conducting Φ-value analysis, which will be explained later, indicating that two folding models can co-exist for one protein.

In the nucleation-condensation mechanism (Fersht, 1995; Fersht, 1997; Nölting et al., 1997), early formation of a diffuse folding nucleus catalyzes further folding. The nucleus primarily consists of a few adjacent residues that have some correct secondary structure interactions but is stable only in the presence of further correct tertiary structure interactions. The folding of chymotrypsin inhibitor 2 conforms to a nucleation-condensation model, which may explain the fairly fast folding rate. The nucleus of the protein is part of the only helix and some distant residues with which it interacts (Itzhaki et al., 1995). Only Ala16 on the helix in the hydrophobic core has its full native interaction energy in the TS. The formation of the helix in the TS is stabilized by the interaction between Ala16 and Leu49 and Ile57 that will form the hydrophobic core. The initially diffuse nucleus of relatively low stability becomes increasingly stabilized as further structure grows around it in a hierarchical manner (Nolting, 1999). Measurements of the nucleation site by Φ-value analysis which supports the nucleation-condensation model coincide with predictions by using the folding funnel model. The folding funnel model discussed in the next paragraph concentrates on the rapid decrease of the conformational dispersity during the reaction (Shoemaker et al., 1999). All three models

may be extended to proteins that have intermediates and multiple TSs on their pathways. The final steps of folding usually involve the interlocking of the side chains.

The folding funnel hypothesis is the energy landscape theory of protein folding. Energy landscapes may be considered as energy levels for partially folded intermediate states of the proteins (Onuchic and Wolynes, 2004; Oliveberg and Wolynes, 2005). It assumes that the native state of a protein corresponds to its minimum free energy and entropy. As the protein goes from the unfolded state to the native state, both the conformational energy and entropy value go down and the number of native contacts goes up. In the folding funnel model, the protein has many different starting conformations and eventually converges into a single pathway. The driving force for folding is the sequestration of side chains of hydrophobic residues to the interior of the folded protein. The molten globule state is predicted to be an ensemble of folding intermediates and it corresponds to a species in which hydrophobic collapse has occurred but native contacts have not yet formed. This state may also exist as an intermediate in the three classic models. In addition, it has most of its secondary structure but very little tertiary structure (Kelly and Price, 1997). It belongs to a type of intrinsically disordered protein (Mittag and Forman-Kay, 2007). Molten globule states may exist as an important intermediate in the folding pathway of some proteins, such as α-lactalbumin, calcium-binding lysozyme, apomyoglobin and cytochrome c (Ptitsyn, 1995; Kuwajima and Arai, 2000). Under certain conditions such as low pH, a native protein can be in the molten globule state.

In order to predict predicting folding rates computationally, Plaxco et al. (1998) revealed a significant correlation between the natural logarithm of the intrinsic folding

rate ($\ln k$) and the relative contact order. Contact order is the average sequence distance between residues forming native contacts in the native state divided by the total length of the protein. Their results show that $\ln k$ decreases as the relative contact order increases. Futhermore, although there is no significant relationship between protein length and folding rates, a weak correlation is observed between length and the fraction of solvent-exposed surface area buried in the TS.

The ability to test both empirical and theoretical relationships for predicting protein folding rates has been limited because a wide range of experimental conditions and methods are employed. To overcome these problems, Maxwell et al. (2005) defined a consensus set of experimental conditions such as 25°C at pH 7.0, 50 mM buffer, data analysis methods and data reporting standards. The folding kinetics of thirty two-state proteins or protein domains were further studied under the consensus conditions.

METHODS TO STUDY PROTEIN FOLDING: Φ-VALUE ANALYSIS

Φ-value analysis is an experimental protein engineering method used to study the structure of the folding TS in small domains that fold with a two state transition. Since the folding TS is a transient and partially unstructured state, its structure is difficult to determine by methods such as protein NMR or X-ray crystallography. In Φ-value analysis, the folding kinetics and conformational folding stability of the wild-type (WT) protein are compared with those of one or more point mutants (Fersht et al., 1992; Zarrine-Afsar and Davidson, 2004). This comparison yields a Φ-value that measures the mutated residue's energetic contribution to the TS, and thus the degree of native structure around the mutated residue in the TS.

Usually a high percentage of the residues in a protein are mutated to identify clusters of residues that are well-ordered in the folded TS. Generally the mutations are non-disruptive and conservative and replace the original residue with a smaller one, most commonly alanine (Zarrine-Afsar and Davidson, 2004). Examples of proteins that have been studied by Φ-value analysis include barnase (Dalby et al., 1998), chymotrypsin inhibitor (Jackson et al., 1993; Gromiha and Selvaraj, 2002), SH3 domains (Northey et al., 2002), individual domains of proteins L and G, ubiquitin (Sosnick et al., 2004), Fadd-DD (Steward et al., 2009), BBL (Neuweiler et al., 2009), and chicken brain α-spectrin (Wensley et al., 2009).

The Φ-value is defined as $(\Delta G^{TS\text{-}D}{}_{WT} - \Delta G^{TS\text{-}D}{}_{M})/(\Delta G^{N\text{-}D}{}_{WT} - \Delta G^{N\text{-}D}{}_{M}) = \Delta\Delta G^{TS\text{-}D}/\Delta\Delta G^{N\text{-}D}$ (Figure 1) (Zarrine-Afsar and Davidson, 2004). $\Delta G^{TS\text{-}D}{}_{WT}$ represents the energy difference between the TS and the denatured state for the WT protein, $\Delta G^{TS\text{-}D}{}_{M}$ represents this energy difference for the mutant (M) protein, and $\Delta G^{N\text{-}D}{}_{WT}$ terms the energy difference between the native state and the denatured state for the WT (Fersht and Daggett, 2002; Daggett and Fersht, 2003). Thus, the Φ-value represents the ratio of the energetic destabilization introduced by the mutation to the TS versus that introduced to the native state. The Φ-value ranges from 0 to 1. A Φ-value near 0 suggests that the region surrounding the mutation is relatively unfolded or unstructured in the TS; a value near 1 means that the local structure around the mutation site in the TS closely resembles the structure in the native state. In Figure 1A, the energy of the TS is affected by Ala to Gly by the same energy as in D, so the change in energy of TS relative to that of D, $\Delta\Delta G^{TS\text{-}D}$, is 0. Thus, $\Phi = \Delta\Delta G^{TS\text{-}D}/\Delta\Delta G^{N\text{-}D} = 0$. In Figure 1B, in the TS, $\Delta\Delta G^{TS\text{-}D} = \Delta\Delta G^{N\text{-}D}$, so $\Phi = 1$. $\Delta\Delta G^{TS\text{-}D}$ is calculated from the ratio of rate constants for folding of

WT [$k_{f(WT)}$] and mutants [$k_{f(M)}$] [$\Delta\Delta G^{TS-D} = RTln(k_{f(WT)}/k_{f(M)})$]. The value of $\Delta\Delta G^{N-D}$ is calculated by subtracting the free energy of folding of mutant $\Delta G^{N-D}{}_{M}$ from that of WT protein $\Delta G^{N-D}{}_{WT}$. The free energies of folding are usually measured from urea, GndHCl, or thermal denaturation curves. Calculations of folding rates and equilibrium constants are required for Φ-value analysis (Zarrine-Afsar and Davidson, 2004).

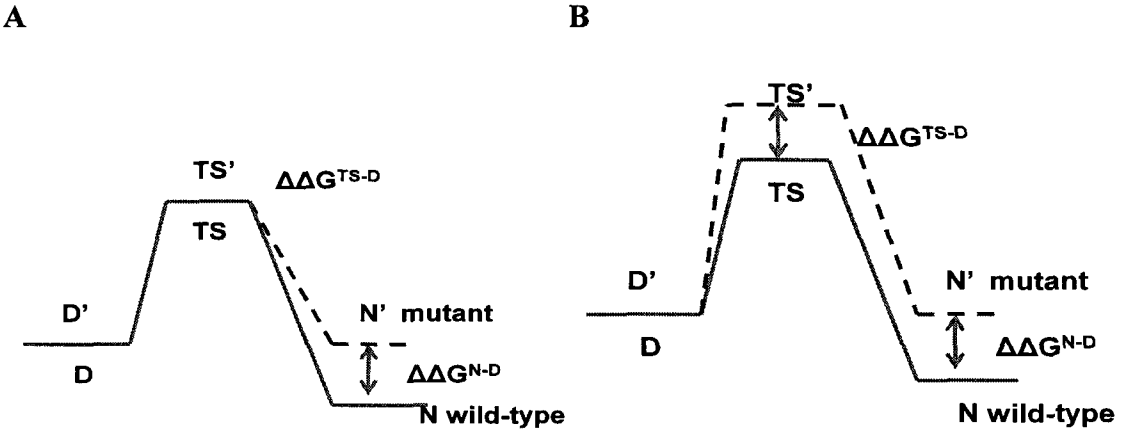


Figure 1. Φ-value analysis (adapted from Zarrine-Afsar and Davidson, 2004)

Schematic profiles are sketched in solid lines for a protein that has an alanine, and in dashed lines for a variant in which the alanine is replaced by glycine.

(A) The TS is at the top of the energy profile. $\Delta\Delta G^{TS-D}$ is 0 and thus, $\Phi = 0$.

(B) In the TS, $\Delta\Delta G^{TS-D} = \Delta\Delta G^{N-D}$, so $\Phi = 1$ (modified from Dagget and Fersht, 2003).

METHODS TO STUDY PROTEIN FOLDING: STOPPED-FLOW FLUORESCENCE

The application of fluorescence in protein chemistry usually employs intrinsic fluorescence of aromatic residues (Lakowicz, 2006). Three aromatic amino acids tryptophan, tyrosine, and phenylalanine are fluorescent; however, they all occupy relative low percentage in proteins. For example, there is generally 1 mol% tryptophan present in proteins (Lakowicz, 2006). The maximal absorption wavelengths for phenylalanine, tyrosine and tryptophan are approximately 260 nm, 280 nm, and 285 nm, respectively; and the maximal emission wavelengths for phenylalanine, tyrosine and tryptophan are about 285 nm, 300 nm and 360 nm, respectively (Lakowicz, 2006). Tryptophan has both a greater molar extinction (absorption) coefficient and fluorescence intensity than the other two aromatic amino acids (Lakowicz, 2006), therefore it is a good probe for monitoring unfolding and refolding. An excitation wavelength of 295 nm is employed in order to specifically detect tryptophan intrinsic fluorescence. Usually, for a protein with a buried tryptophan, the maximum emission wavelength is shorter, and emission intensity is greater in the native state than in the denatured state in the equilibrium study. This is because in the presence of a polar solvent, the energy is transferred to the solvent and the emission is lost (Lakowicz, 2006). Consequently, the fluorescence intensity will increase at the maximum wavelength if such a protein goes from the denatured state to the native state. Furthermore, unfolding and refolding conducted with fluorescence can indicate important thermodynamic values such as the Gibbs free energy of denaturation (Pace et al., 1989). However, tryptophan is very sensitive to its microenvironment. The tryptophan can be quenched in the native state by the neighboring residues such as cysteine, a

disulfide bridge, histidine, aspartic acid, glutamic acid and lysine (Engelborghs, 2001). Moreover, a tryptophan in a protein with multiple tryptophans may behave differently depending on its own microenvironment. In this case, site-directed mutagenesis helps to detect the contribution of each individual tryptophan to the overall fluorescence.

A stopped-flow instrument is a rapid mixing device used to study the kinetics of a reaction in solution and it allows various conditions to be tested (Frieden et al., 1993). Stopped-flow apparatus from Bio-Logic (Claix, France) and Applied Photophysics (Surrey, UK) are commonly used. Different types of spectroscopy such as fluorescence spectroscopy (Lakowicz, 2006), circular dichroism (CD) spectroscopy (Kelly et al., 2005), real time-NMR spectroscopy (Zeeb and Balbach, 2004) and FT-IR spectroscopy (Fabian and Naumann, 2004) can be used in combination with stopped-flow (Frieden et al., 1993). A stopped-flow instrument coupled to either a CD spectrometer or a fluorescence spectrometer is most often used in the field of protein folding to observe rapid reactions (Kuwajima, 1996). In the stopped-flow fluorescence spectroscopy, after one solution containing the unfolded protein and another solution containing the native buffer are mixed, the change of fluorescence intensity using tryptophan or tyrosine as the probe is amplified, revealing the kinetics of the hydrophobic core formation in a protein with such buried residue(s) (Lakowicz, 2006). The kinetics traces can be analyzed with the scientific analysis and graphing program SigmaPlot to calculate the folding rates and residuals using various exponential equations. The best-fit residuals indicate the number of phases in the folding process.

METHODS TO STUDY PROTEIN FOLDING: STOPPED-FLOW CIRCULAR DICHROISM

CD is one of the most widely used types of chiroptical spectroscopy (Woody, 1996) and a very sensitive biophysical tool for detecting the overall structure of the protein and monitoring conformational change (Venyaminov and Yang, 1996). The inherent information content of the far-UV CD spectra (190-250 nm) depends on the difference in absorption of left- and right- handed circularly polarized light at the protein backbone (Bohm et al., 1992). Within these wavelengths the peptide bonds in a regular and folded environment are the chromophore, whereas in fluorescence the side chains of aromatic residues are chromophores. Therefore, the far-UV CD region of proteins can reveal important characteristics of their secondary structure elements including conformations of the α-helix, the β-sheet, the β-turn and random coil (Woody, 1996; Whitmore and Wallace, 2008). The far-UV CD spectra can also examine the types of secondary structure in proteins including all-α, all-β, α + β, α/β and disordered proteins (Venyaminov and Yang, 1996; Kelly et al., 2005). For example, all-α proteins show two strong double minimum at around 208 nm and 222 nm and all-β proteins usually demonstrate a less strong single minimum between 210 and 225 nm (Venyaminov and Yang, 1996). It can be used to study how the protein secondary structure changes as a function of temperature, pH, or the concentration of denaturant such as guanidinium hydrochloride (GndHCl) or urea.

Near-UV CD spectrum (250-320 nm) of proteins provides information on the tertiary structure. The signals obtained in the 250-300 nm region are due to the absorption, dipole orientation and the nature of the surrounding environment of the

phenylalanine, tyrosine, tryptophan and cysteine (or disulfide bond) (Woody and Dunker, 1996). Tryptophan shows a peak near 290 nm with fine structure between 290-305 nm; tyrosine shows a peak between 275-282 nm, with a shoulder at longer wavelengths; phenylalanine shows weaker but sharper signals with fine structure between 255-270 nm (Kelly et al., 2005). Unlike far-UV CD, near-UV CD spectrum cannot be assigned to any one specific three-dimensional structure *a priori*.

CD gives less specific protein structural information than X-ray crystallography and NMR spectroscopy. However, CD spectroscopy is a fast method that does not require large amounts of protein or extensive data processing. Therefore, CD can examine changes in protein structure for a large number of conditions including solvent, temperature, pH and salt concentration. Stopped-flow far- (usually at the wavelength of 222 or 225 nm) and near- UV CD spectroscopy usually at the wavelengths between 275 and 295 nm can measure the rates of secondary structure and the tertiary structure formation, respectively (Kuwajima, 1996; Kelly et al., 2005).

METHODS TO STUDY PROTEIN FOLDING: QUENCHED-FLOW AND HYDROGEN-DEUTERIUM EXCHANGE

A combination of quenched-flow methods, hydrogen deuterium exchange (HX) and NMR spectroscopy can give detailed site specific information about the protein folding process (Englander and Mayne, 1992; Dyson and Wright, 1996). In a typical experiment (Krishna et al., 2004) as shown in Figure 2, the protein is initially unfolded in D_2O and guanidium-HCl (GndHCl) and all the amide hydrogens exchange to deuteriums. Folding is initiated by rapid dilution into a folding buffer in H_2O. The folding buffer is kept at fairly low pH so that no exchange occurs. After select folding times, a brief

hydrogen labeling pulse is applied by mixing with high pH buffer and amide deuteriums that are unprotected exchange to hydrogens but those in stable-formed structures are protected. A third mix into low pH buffer stops the labeling process. The protein sample is then concentrated and exchanged into D_2O buffer for NMR analysis. 2D NMR spectra of the refolded protein are then recorded and the proton occupancy is measured at each refolding time for the observable amide proton resonances (Dyson and Wright, 1996).

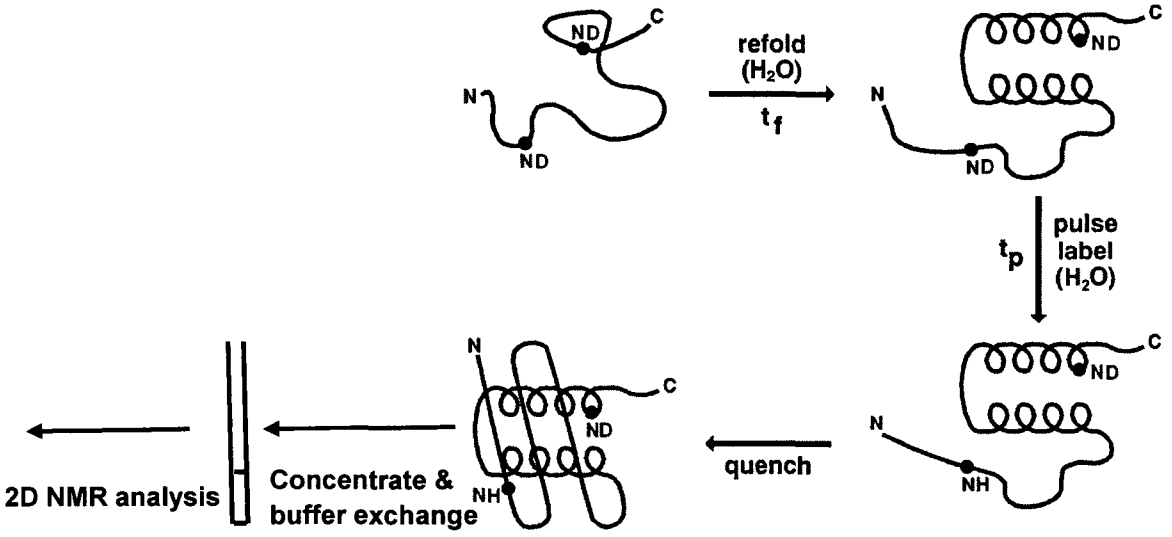


Figure 2. The schematic diagram showing the protein at different stages in a typical quenched-flow, HX and NMR experiment (adapted from Dyson and Wright, 1996) The stages include refolding, pulsing, quenching, concentration and buffer exchange.

Quenched-flow HX studies of the folding of a number of proteins have been reported, such as ribonuclease A (Udgaonkar and Baldwin, 1988), cytochrome c (Roder et al., 1988), T4 lysozyme (Lu and Dahlquist, 1992), ribonuclease T_1 (Mullins et al.,

1993), staphylococcal nuclease (Jacobs and Fox, 1994), immunoglobulin binding domain of streptococcal protein G (Kuszewski et al., 1994), dihydrofolate reductase (Jones and Matthews, 1995), hen lysozyme (Lu et al., 1997), apomyoglobin (Garcia et al., 2000), acyl-CoA binding protein (Teilum et al., 2000), human fibroblast growth factor (Samuel et al., 2001), hisactophilin (Liu et al., 2002), cobrotoxin (Hsieh et al., 2006), onconase (an RNase A homologue from the oocytes of *Rana pipiens*) (Schulenburg et al., 2009) and lysozyme from bacteriophage λ (Di Paolo et al., 2010). The results of several of these studies such as hen lysozyme, human fibroblast growth factor and onconase appear to show that certain parts of the protein may fold earlier than other regions (Lu et al., 1997; Samuel et al., 2001; Schulenburg et al., 2009). However, in other proteins such as phage λ lysozyme, acyl-CoA binding protein and the immunoglobulin binding domain of streptococcal protein G the kinetics of secondary structure formation indicate that they form cooperatively (Kuszewski et al., 1994; Teilum et al., 2000; Di Paolo et al., 2010).

CHAPERONES

The refolding of an unfolded protein *in vitro* is a spontaneous process. The primary driving force for protein folding is the movement of hydrophobic residues away from the aqueous solvent and reduction in solvent accessible area. On the other hand, folding of a nascent protein *in vivo* is a different story (Garrett and Grisham, 2005). The proteins in the cell can be so crowded that new proteins can aggregate due to the interaction between their hydrophobic patches (van den Berg et al., 1999). In order to solve this problem, new proteins are assisted in folding by proteins called molecular chaperones and chaperonin (Martin and Hart, 1997). Chaperones such as heat shock proteins (Hsp) bind to the hydrophobic segments of proteins and release them as each

segment is ready to participate in folding. The main Hsp chaperones are Hsp10, Hsp40 (DnaJ in *Escherichia coli*), Hsp60, Hsp70 (DnaK in *E. coli*) and Hsp90 (HtpG in prokaryotes). In eukaryotes, a protein after release from ribosomes can interact with Hsp70, and then Hsp60 to complete the folding process. GroEL is an example of chaperonin in *E. coli* (Garrett and Grisham, 2005; Kampinga et al., 2009) and GroES is a co-chaperonin. In *E. coli,* nascent proteins fold in the cylindrical cavity of the GroES-GroEL chaperonin complex. The cylindrical cavity protects the protein from the crowded environment of the cytoplasm (Scopes and Truscott, 1998; Garrett and Grisham, 2005). In eukaryotes, prefoldin, which serves as a co-chaperone, binds unfolded chains emerging from the ribosomes and delivers them to the chaperonin, TricC (Garrett and Grisham, 2005).

MISFOLDING CAUSES DISEASE

The misfolding of a protein is often associated with its assembly into fibrillar aggregates, commonly termed amyloid fibrils (Bellotti and Chiti, 2008). Amyloid fibrils are insoluble aggregates, which form when proteins polymerize to form a β-sheet structure (Chiti and Dobson, 2006; Bellotti and Chiti, 2008; Maji et al., 2009). Figure 3 is an illustration of how a native protein forms amyloid fibrils. A partially unfolded protein can form misfolded monomers. The misfolded monomers oligomerize to form a precursor species known as protofibrils which mature into fibrils. Accumulation of amyloid fibrils may lead to various neurodegenerative and other kinds of diseases. Human disease associated with the formation of extracellular amyloid fibrils or intracellular inclusions with amyloid-like characteristics include thirteen

neurodegenerative diseases, thirteen nonneuropathic systemic amyloidoses and fourteen nonneuropathic localized diseases (Chiti and Dobson, 2006).

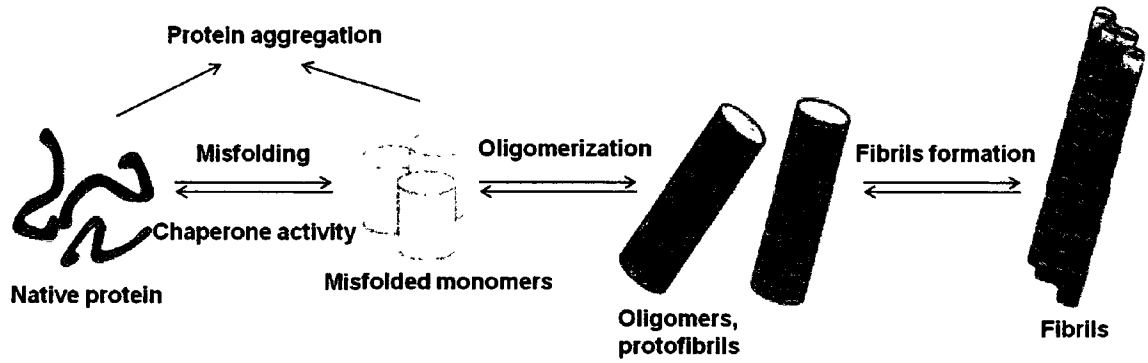


Figure 3. Model for protein misfolding and fibrillization (adapted from Skovronsky et al., 2006)

Soluble native protein is misfolded and associates in the form of oligomers and protofibrils that eventually lead to fibrils.

Neurodegenerative diseases associated with fibrils include Alzheimer's, Huntington's and Parkinson's. Alzheimer's disease is caused by accumulation of β-amyloid (1-42 aa) and tau protein aggregations in the brain (Hashimoto et al., 2003). β-amyloid is a fragment hydrolyzed from a larger integral membrane protein called amyloid precursor protein (Hooper, 2005). One fragment (1-42 aa) aggregates to form ordered insoluble fibrils consisting of β-sheets (Ohnishi and Takano, 2004). Huntington's disease is caused by a tri-nucleotide (CAG) repeat in a gene, making it exceed a normal length (Walker, 2007). The CAG sequence (coding for the glutamine) repeats in Huntingtin protein results in a chain of glutamines (Walker, 2007). A sequence of forty or more

glutamines in the protein causes it to aggregate and form amyloid fibrils (Walker, 2007). Parkinson's disease is thought to be caused by an abnormal accumulation and fibril-formation of α-synuclein in the brain (Galpern and Lang, 2006). However, Volles and Lansbury (2003) propose that and toxicity is caused by the pore-like protofibrils, which lead to membrane permeabilization, rather than the amyloid fibril, fibrillar aggregates, Lewy bodies and the α-synuclein monomer. Bovine spongiform encephalopathy, also known as mad cow disease and humans Creutzfeldt-Jakob disease, is believed to be caused by misfolded prion protein (Prusiner, 1998). PrP^{C} is a normal protein found on the cells membranes. PrP^{Sc}, the infectious isoform of PrP, can change conformation of normal PrP^{C} proteins into the infectious isoform.

Nonneurodegenative diseases include cystic fibrosis and type 2 diabetes. Mutation in the gene encoding cystic fibrosis transmembrane chloride channel causes cystic fibrosis. The most common mutation is ΔF508, a deletion of three nucleotides that results in a loss of phenylalanine508 on the protein. Absence of phenylalanine508 causes the protein to misfold (Cyr, 2005). The folding intermediate remains attached to the chaperone and is not inserted into the membrane (Kopito, 1999). Type 2 diabetes is caused by amylin amyloid fibrils in the pancreas (Jaikaran and Clark, 2001). Haemodialysis-associated amyloidosis is caused by amyloid fibrils of β2 microglobulin, a normal serum protein in the blood (Miyata et al., 1993). The mutation of p53, a key check-point protein in the cell cycle regulation results in its misfolding. The proteasome system removes the misfolded p53. The reduced concentration of p53 fails to arrest the cell cycle in the event of improper DNA replication (Garrett and Grisham, 2005). The above examples show that mutation and/or failure of assistance by chaperones lead to

protein misfolding and fibril formation, resulting in diseases. External factors may also play a role in amyloid fibrils formatin such as metals and other environmental agents (Alexandrescu, 2005).

THE FAS-ASSOCIATED DEATH DOMAIN AND THE DEATH DOMAIN SUPERFAMILY

The death domain superfamily fold consists of six α-helices arranged in a Greek-key topology. This topology is shared with members of two other superfamilies, which are the all β-sheet immunoglobulins and the mixed α/β-plaits (Higman and Greene, 2006). Extensive experimental studies have been conducted on these two superfamilies, but no studies have been conducted on the death domain superfamily members. Thus they are ideal for investigating the determinants of the common topology. The death domain superfamily is comprised of four families: death domain (DD), death effector domain (DED), caspase recruitment domain (CARD) and pyrin domain (PYD). They function in either intracellular signal transduction of apoptosis (DD, DED and CARD), or innate inflammation (PYD) (Park et al., 2007). The Fas-associated death domain (Fadd) consists of an N-terminal DED and a C-terminal DD (Fadd-DD) consisting of one-hundred residues (Berglund et al., 2000; Carrington et al., 2006). Fadd-DD was selected as our model DD protein system because its NMR structure has been solved, the protein is monomeric and the protocols to express and purify the protein are already established (Berglund et al., 2000).

The role of Fadd-DD in the apoptosis is presented in Figure 4. The Fas ligand, an intrinsic membrane protein, is present on the surfaces of cytotoxic T lymphocytes, which remove the virally infected cells (Pollard and Earnshaw, 2004). The Fas ligand associates

with the extracellular binding domain of Fas on the target cell and initiates the apoptotic pathway (Ashkenazi and Dixit, 1998). Intracellular DD of Fas binds with the DD of Fas-associated protein. The DED of Fas-associated protein binds with DED of procaspase 8. Procaspase 8 activates itself proteolytically and activates downstream effector caspases.

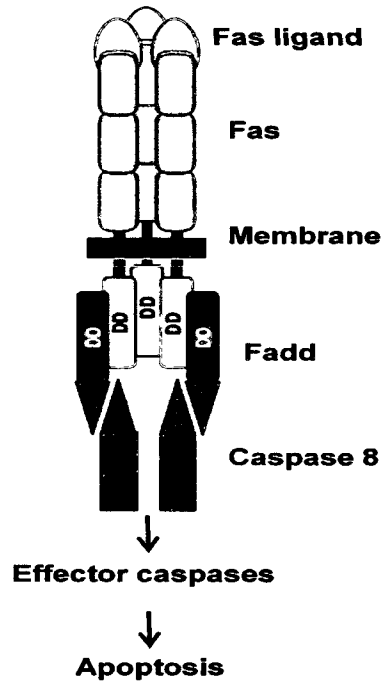


Figure 4. The role of Fadd-DD in the apoptotic signaling (adapted from Ashkenazi and Dixit, 1998)

Fadd: Fas-associated death domain; DD: death domain; DED: death effector domain.

CHAPTER II

MATERIALS AND METHODS

MATERIALS

All chemicals were high quality reagents. Ultra pure 8 M Guanidine-HCl solution (GndHCl) was purchased from Pierce (Rockford, IL), dithiothreitol (DTT), carbenicillin and Isopropyl β-D-1-thiogalactopyranoside (IPTG) from Fisher (Waltham, MA) or VWR (Suwanee, GA), Bis-Tris from Acros Organics (Morris Plains, NJ) and Sephadex G-75 superfine resin from Sigma (St. Louis, MO). All buffers were filtered through either a 0.45 or 0.22 μm filter (Pall Corporation, Ann Arbor, MI). T7 promoter and T7 terminator primers, Pfu DNA polymerase, Phi-X174/Hae III Marker, *E. coli* BL21(DE3), dNTP mix, 2 X PCR mastermix and XL1-blue supercompetent cells were from Stratagene (La Jolla, CA). Qiaquick gel extraction kit and Qiagen plasmid mini Kit were from Qiagen (Valencia, CA). pET-14b vector was from EMD Chemicals (San Diego, CA). *Nde* I, *Xho* I, *Bam*H I and NEB 5-alpha F'I^q competent *E. coli* were from New England Biolabs (Ipswich, MA). NuPAGE Novex 4-12% Bis-Tris Mini Gels were from Invitrogen (Carlsbad, CA). Ni-NTA His bind resin and Ni-NTA Buffer Kit were from Novagen (San Diego, CA).

The ^{15}N labeled-ammonium chloride and 99.0% deuterium oxide were purchased from Cambridge Isotope (Andover, MA). 2-(*N*-morpholino) ethanesulfonic acid (MES), glycine and citric acid were purchased from Fisher. Dipotassium phosphate was purchased from VWR. Deuterated GndHCl, DTT, citric acid and K_2HPO_4 were prepared

by dissolving the chemicals into D_2O and lyophilizing. This procedure was repeated three times. The concentration of GndHCl used in the experiments was determined with an Atago hand-held refractometer (Tokyo, Japan).

Protein purification was conducted on a Perceptive Biosystems BioCAD Sprint (GMI, Ramsey, MN). Mutagenesis was conducted with the Stratagene QuickChange II site-directed mutagenesis kit. DNA sequencing was conducted by Nucleic Acids Research Facility (Virginia Commonwealth University, Richmond, VA). Mass spectroscopy was conducted by the COSMIC facility (Old Dominion University, Norfolk, VA).

METHODS FOR THE BIOINFORMATICS STUDY OF FADD-DD

The death domain superfamily is a large superfamily with different functions, divergent sequences and members with known structures. A bioinformatics analysis to identify conserved residues has not been conducted. This analysis is a necessary prerequisite for structure, stability and folding studies to test the role of conserved residues in structure and folding.

Construction of a Superfamily Sequence Alignment

A multiple sequence alignment is an alignment of three or more protein sequences. Generally the input sequences have an evolutionary relationship by sharing a lineage and descending from a common ancestor. A diverse multiple sequence alignment was created in order to enhance sequence variability and in this way, only the conserved residues for structure and folding could be identified. PSI-BLAST (Altschul et al., 1990) was used to search for divergent (< 25% sequence identity) members of the death domain superfamily. Four proteins served as query sequences when conducting PSI-BLAST

search: human Fadd-DD (PDB code: 1E3Y), human Fadd-DED (PDB code: 1A1W), human Apaf-1-CARD (PDB code: 1CY5) and human ASC-PYD (PDB code: 1UCP). They were selected for the following reasons: the three-dimensional structures are resolved, they are functionally diverse and they represent the four families within the superfamily. The multiple sequence alignment was created using the program Muscle - multiple protein sequence alignment program (Edgar, 2004) in the software Jalview (Clamp et al., 2004). Modifications were made by hand based on the comparison of the side-chain orientation in the four aligned protein structures generated with the online CE-MC-Multiple Protein Structure Alignment Server (Guda et al., 2004). Modifications by hand are necessary to make sure the side chains of the four structures have similar orientation and all sequences without solved structures can be aligned properly because computer algorithms are not perfect. Structures were analyzed using Insight II, version 2005 (Accelrys, San Diego, CA) running on a SUN workstation with Linux.

Calculation of Conservation and Hydropathy

The number of each residue type in each position of the superfamily sequence alignment was calculated using a program written in Perl script. The entropy value was calculated by the following equation:

$$S(i) = \sum_{j=1}^{m} -\{Pj(i)\ln[Pj(i)]\}(i = 1,2,\ldots,20)$$

$Pj(i)$ is the fractional occurrence of amino acid type j at each site i; and m is the number of amino acid types used in the particular analysis (Sander and Schneider, 1991). Since twenty sequences were incorporated, i ranges from 1 to 20. Conservation was calculated by the following equation: $C(i) = 1 - S(i) / \ln(m)$ (Greene et al., 2003). The positions with conservation values greater than 0.45 are considered to be highly

conserved; the positions with conversation values between 0.35 and 0.45 are considered to be moderately conserved; and the positions with conservation values lower than 0.35 are considered to be less conserved. The positions which have more than one gap are considered non-conserved and therefore have a value of zero. Persistent hydropathy is calculated by the following equation: hydropathy = sum of the number of each amino acid * hydrophobicity of that amino acid. The hydrophobicity scale is adapted from Nozaki and Tanford scale (Nozaki and Tanford, 1971). The resultant data were analyzed using SigmaPlot (Ver. 10, SYSTAT Software).

The Long-range Interaction Network within the Fadd-DD

A contact file listing all atom contacts within each structure was generated with the Contact program (Collaborative Computational Project, 1994). Programs written in Perl script were used to calculate the conserved long-range interaction network from the contact files. The focus is on long-range interactions which are defined as pair-wise contacts between residues that are ten or more amino acids apart in the primary structure but within 6 Å in the three-dimensional structure (Greene and Higman, 2003). Interactions are considered between all atoms except hydrogens.

METHODS FOR THE EXPERIMENTAL STUDY OF FADD-DD

The experimental procedures are summarized in Figure 5. The recombinant DNA was obtained after the Fadd-DD gene was amplified and cloned into an expression vector. Fadd-DD protein was expressed and purified and biophysical studies were conducted.

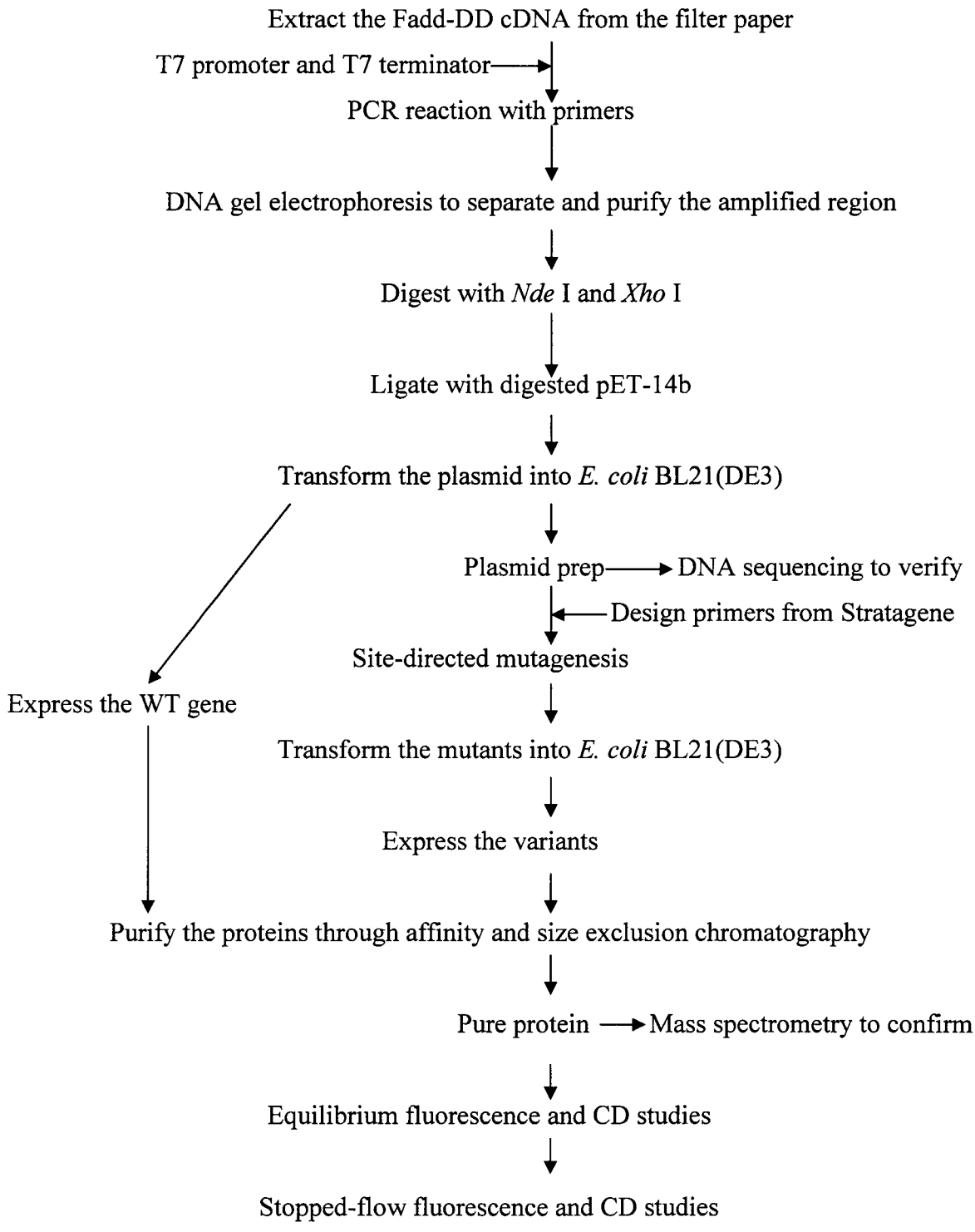


Figure 5. The flow chart showing the experimental procedures

Cloning

Plasmid is a circular double-stranded DNA and separate from the chromosomal DNA. Since a plasmid can replicate independently of the chromosomal DNA, it is commonly used as a vector to multiply and express a particular gene in bacteria. Complementary DNA (cDNA) is the DNA sequence synthesized from a mature mRNA template under the catalysis of the reverse transcriptase and DNA polymerase. The paper containing the plasmid with Fadd-DD as cDNA was kindly supplied by Dr. Paul Driscoll, University of College London (London, UK). The plasmid was contained within two concentric circles. The small inner circle and the big outer circle of paper containing the plasmid was cut and immersed in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). They were preserved at -20°C overnight, followed by incubation at 42°C for fifteen minutes and vortexed to get most of the DNA sample into the solution. Since transformation of the plasmid into the expression cells did not generate the recombinant bacteria probably due to inadequate amount of plasmid, molecular biology methods were used to obtain the cDNA.

The polymerase chain reaction (PCR) is a technique used to amplify a single or a few copies of DNA and generate a large amount of copies of a particular DNA sequence *in vitro*. Forward and reverse primers are DNA oligomers, which serve as starting points for DNA synthesis. PCR was used to amplify the regions between T7 promoter and T7 terminator on the cloning/expression region of pET-14b vector with Eppendorf Mastercycler (Eppendorf, NY) (see Figure A1). The vector harbors the gene encoding β-lactamase which is resistant against ampicillin, therefore *E. coli* transformed with the vector can grow on the media with ampicillin. PCR reactions were set up according to

Table A1 and the temperatures and cycles were set up as Table A2. The T_m of the two primers were calculated according to the equation: $T_m = 64.9°C + 41°C$ X (number of G's and C's in the primer – 16.4)/N (http://www.promega.com/biomath/calc11.htm). Both T7 promoter primer and T7 terminator primer bind to the specific regions of the plasmid. The initial set up stage was increased to 2.5 minutes as Hot-start PCR to reduce non-specific amplification. Pfu DNA polymerase from the hyperthermophilic archaeon *Pyrococcus furiosus* was used to generate fragments which have fewer errors than those generated by Taq DNA polymerase from the thermophilic bacterium *Thermus aquaticus*.

The PCR product was isolated by the agarose gel electrophoresis. After a 1.5% of agarose gel with 50 ng/ml of ethidium bromide was polymerized, it was immersed in 1 X Tris/borate/EDTA (TBE) buffer (pH 8.0). Voltage for electrophoresis was set to 100 V and time was set to 30-40 minutes. Loading dye (0.4% orange G, 0.03% bromophenol blue and 0.03% xylene cyanol FF) was used to track the movement of the samples. Phi-X174/Hae III Marker or ladder (100-1000 bp) was used to calculate the approximate molecular weight of cDNA. After electrophoresis, the DNA in the gel was illuminated with a UV lamp. Part of the gel which contained cDNA was cut and purified with Qiaquick gel extraction kit. cDNA samples were sent to Dr. Wayne Hynes's group (Department of Biological Sciences, ODU, Norfolk, VA) and Nucleic Acid Research Facility (VCU, Richmond, VA) for sequencing.

The cDNA encoding Fadd-DD was cloned into pET-14b plasmid. Two of three restriction sites for cloning (*Nde* I, *Xho* I and *Bam*H I) are commonly used for the insertion of any cDNA clone. Figure A2 shows the sequences of three cloning sites. Three restriction enzymes *Nde* I and *Xho* I, and *Nde* I and *Bam*H I were used to cut the

cDNA and pET-14b plasmid (Table A3). The digestion protocol is presented in Table A4. The addition of bovine serum albumin (BSA) helps reduce the loss of restriction enzymes on tube and pipette tip surfaces and stabilize the enzymes according to the website of New England Biolabs. Sequential digestion was performed in order to get a good yield. The plasmid and cDNA were first cut with *Nde* I, and then cut with *Bam*H I or *Xho* I. Digestion products were run on the agarose gel and then purified with gel extract kit as mentioned previously. One cDNA cut by *Nde* I and *Xho* I and another cDNA cut by *Nde* I and *BamH* I were inserted into the corresponding cut plasmid, resulting in the recombinant plasmid. Inserts were made one fourth dilution to get the correct ratio with the plasmid (Table A5).

Transformation is the process during which a competent bacterial cell takes up an exogenous DNA, particularly a plasmid, from its environment. Transformation was performed according to the protocol of NEB 5-alpha $F'I^q$ competent *E. coli*. 3 µl of recombinant DNA was transformed into 50 µl of competent cell. Cells were heat shocked at 42°C for 30 seconds. 200 µl of reaction solution was added to LB medium containing 200 µg/ml of ampicillin. After drying at room temperature, plates were incubated upside down at 37°C overnight.

Colonies containing the recombinant plasmid grew on the plates because the plasmid conferred ampicillin-resistance. Twelve colonies were picked up and inoculated into 50 ml falcon tubes containing 5 ml LB with 100 µg/ml of ampicillin (six from "*Xho*" and six from "*Bam*"). The culture was incubated at 37°C and shaking at 200 rpm overnight. 3 ml of overnight LB culture was centrifuged at 13,000 rpm for 15 minutes. The resultant pellet was dried for 15 minutes and dissolved in 50 µl of ddH_2O. Plasmid

extraction was conducted with the protocol of Qiagen plasmid mini Kit. To verify if the cDNA is present in the recombinant plasmid, 10 µl of plasmid was cut with 2 µl of *Nde* I and 2 µl of *Xho* I in 0.4 µl of BSA, 4 µl of buffer 4 and 22 µl of water. Mixture was incubated at 37°C for one hour.

Bacterial strains from colony 1 from "*Xho*" and colony 7 from "*Bam*" were inoculated into two flasks and incubated overnight at 37°C. After plasmid extraction, one of the plasmid from "*Xho*" with the highest concentration of 285 ng/µl and one from "*Bam*" with the highest concentration of 214 ng/µl were sent for sequencing. The result showed that the cDNA encoding Fadd-DD was subcloned into pET-14b.

Protein Expression

The small-scale expression of Fadd-DD WT was performed to find out the best experimental conditions. 1 µl of recombinant plasmid was transformed into 100 µl of expression cell *E . coli* BL21(DE3). Pasmid pUC18 (Stratagene) was also transformed into the cell as a control. After growing overnight at 37°C, one single colony was picked up and inoculated into 5 ml of Luria-Bertani (LB) broth in a falcon tube, which was incubated overnight at 37°C. Bacteria containing pUC18 was used as a negative control. 50 µl of culture was transferred into 50 ml of LB broth containing ampicillin (200 µg/ml) and incubated at 37°C. Following the first four hours, the optical density (O.D.) value at 600 nm was measured in the Cary 50 UV-Vis spectrophotometer (Varian, Palo Alto, CA) every hour until the value reached around 0.8. The $O.D._{600}$ reached 0.8 at 7.8 hours. At this point, 100 µl of 100 mg/ml of ampicillin was added to the medium. Additionally, 20 µl of 1 M IPTG as an inducing agent was added to one culture while another was kept free of IPTG.

The $O.D._{600}$ of 1 ml of bacteria culture was measured sixteen hours later after being induced with IPTG. The $O.D._{600}$ of all three cultures (with IPTG, without IPTG and pUC control) were adjusted to be approximately 1.5. 1 ml of culture was centrifuged at 120,000 rpm for 5 minutes to harvest the cell. After the supernatant was decanted, the cell pellet was redissolved in 40 μl of 1 X sample buffer (100 μl of 4 X sample buffer, 300 μl of water and 20 μl of β-ME as a reducing agent). The mixture was heated at 74°C for ten minutes to lyse the cell. The solution was then filtered through syringe several times to shear the sticky DNA. 10 μl more of sample buffer was used to wash out the remaining solution in the syringe.

Protein gel electrophoresis 4 X sample buffer, Coomassie brilliant blue staining buffer and destaining buffer were made according to the protocol from Invitrogen. 15 μl of samples and marker (Invitrogen) were loaded to NuPAGE Novex 4-12% Bis-Tris Mini Gels. Gel electrophoresis was performed at the voltage of 200 V for 45 minutes. The gel was agitated in the staining buffer for three hours on the rotator, and then in the destaining buffer overnight.

The protocol to express the protein was based on that of Berglund et al. (2000) with some variations to accommodate the difference in equipment in our lab such as column sizes. One single colony was inoculated into one 50 ml broth containing carbenicillin (200 μg/ml) in flask and cultivated at 37°C for overnight. 1 ml of bacterial culture was inoculated into each of twelve 500 ml LB media in 2 L baffled flask and cultivated at 37°C. The O.D. value was measured at 600 nm every hour after inoculation. After the O.D. value reached 0.8, carbenicillin and IPTG was added so that the final concentration is 200 μg/ml and 0.4 mM, respectively. The culture was continually

cultivated at 25°C for sixteen more hours and then the media was centrifuged in 500 ml bottles at 7,000 rpm for thirty minutes. The cell pellet was harvested and combined into one bottle and the supernatant was removed. The cell pellet was resuspended into 50 ml Tris buffer (pH 8.0) containing 50 mM Tris base, 300 mM NaCl and 5 mM β-mercaptoethanol (β-ME). The suspension was frozen at -80°C to help lyse the cell. After the cells were thawed, the suspension was sonicated every four seconds per one minute for four hours with amplitude of 0.38 at 20°C with ultrasonic processor (Sonics & Materials, Newton, CT). The cell lysate was spun at 11,000 rpm for one hour and without inclusion bodies which are white sandy percipitants, indicating the protein was produced as a soluable form.

Protein Purification

The established protocols to purify the protein were based on previous publication (Berglund et al., 2000). pET-14b vector can express a protein with an N-terminal His-tag. Since the nickel-nitrilotriacetic acid (Ni-NTA) resin contains Ni^{2+} immobilized on a matrix and binds the His-tag, the resin can be used in an affinity column. An Ni-NTA column (1.5 cm X 30 cm) was loaded with Ni-NTA His bind resin with the column volume of 37 ml. Affinity chromatography was controlled by Perceptive BioSystems BioCAD spring biomolecule purification chromatography systems with double wavelengths of 280 nm and 320 nm. The resin was washed with binding buffer in Ni-NTA Buffer Kit. The lysed cells were loaded onto the Ni-NTA column. After the sample was completely loaded, the column was washed with binding buffer (pH 8.0) containing 50 mM Tris base, 300 mM NaCl and 5 mM β-ME and then washing buffer (pH 8.0) containing 50 mM Tris base, 300 mM NaCl, 5 mM β-ME and 20 mM of imidazole

respectively. After the $O.D._{280}$ became steady, the column was eluted with gradient mixture of washing buffer and elution buffer (pH 8.0) containing 50 mM Tris base, 300 mM NaCl, 5 mM β-ME and 250 mM of imidazole from 100:0 to 0:100. His-tag Fadd-DD began to elute with 150 mM of imidazole with the peak showing up with 200 mM of imidazole. Afterwards, the column was washed with pure elution buffer until $O.D._{280}$ became steady.

Fractions containing His-tag Fadd-DD were collected into Standard Grade RC Membranes: Spectra/Por® 3 with MWCO of 3,500 Da (Spectrum Labs, Rancho Dominguez, CA). In order to remove the imidazole, the protein solution was dialyzed against 4 L buffer (pH 8.4) containing 20 mM Tris-HCl, 150 mM NaCl and 5 mM β-ME overnight at 4°C. The protein solution was concentrated with Vivaspin-20 tubes which have Polyethersulfone Membrane with MWCO of 3,000 Da (Sartorius Stedim Biotech S.A., Aubagne, France). The Vivaspin tubes were centrifuged at 8,000 rpm for 30 minutes in Hettich ROTANTA 460 Tabletop Centrifuge (GMI, Inc., Ramsey, MN).

In order to remove the His-tag, 50 units of thrombin were added to the protein solution and the mixture was incubated at 4°C overnight. SDS-PAGE was run on the protein solution to determine successful thrombin cut. The protein solution was loaded onto Ni-NTA column again in order to separate Fadd-DD and His-tag Fadd-DD. Washing and elution were the same as previously described in the first Ni-NTA column separation. Fractions containing Fadd-DD are collected. Fractions containing His-tag Fadd-DD was also collected for further digestion. Protein concentration was measured at 280 nm in Cary 50 UV-Vis spectrophotometer.

For the gel filtration column, 37 g of Sephadex G-50 or G-75 (Sigma-Aldrich, St Louis, MO) was dissolved in 371 ml of H_2O. The resin was allowed to hydrate and swell overnight. It was then poured into the gel filtration column (5 cm X 70 cm, Uppsala, Sweden) and allowed to settle overnight. The column was equilibrated with phosphate buffer (50 mM phosphate buffer, 150 mM NaCl, 1 mM DTT, pH 6.2). Protein sample was concentrated and loaded to the column through 5 ml of the PEEK Sample Loops (Upchurch Scientific Inc., Oak Harbor, WA). The protein absorbance peak was noticed after approximately 170 ml of buffer elution. Following purification approximately 50 mg of protein per liter of culture was obtained.

The Fadd-DD sample was sent to COSMIC (College of Sciences Major Instrumentation Cluster, ODU) for determination of molecular weight using the instrument of Bruker Apex-Qe, a hybrid Qh-FTICR-MS with an extended high performance actively shielded 12 Tesla magnet. The predicted average molecular mass of Fadd-DD is 11852.32139 Da according to Protein Calculator v3.3 (http://www.scripps.edu/cgi-bin/cdputnam/protcalc3), then $(M+H)^+$ = 11853.32825 u.

Site-directed Mutagenesis

Both forward and reverse primers for seven mutations were designed according to the website provided by Stratagene: http://www.stratagene.com/sdmdesigner/default.aspx (Table A6). The primers were ordered from Invitrogen. The primers were dissolved in autoclaved distilled water and diluted to ideal concentrations of 100 ng/µl.

The PCR reactions were designed according to the QuikChange® Site-Directed Mutagenesis Kit provided by Stratagene. 50 µl PCR solutions consisted of the following components: 5 µl of 10 X reaction buffer (Stratagene), 1 µl of 100 ng/µl plasmid template

encoding Fadd-DD, 1 μl forward primer, 1 μl reverse primer, 1 μl dNTP mix, 41 μl distilled water and 1 μl of 2.5 U/μl *pfuTurbo* DNA polymerase. Alternative, reaction buffer, dNTP mix and *pfuTurbo* DNA polymerase were substituted by 25 μl of 2 X PCR mastermix. 2 μl of 5 ng/μl pWhitescript control plasmid (Stratagene) was used as control. Cycling parameters for the Quikchange site-directed mutagenesis method are as follows: 95°C for 30 sec, 95°C for 30 sec, 55°C for 1 min, 68°C for 4 min (cycles for 16 times). Following temperature cycling, the reaction was cooled to 4°C. 1 μl of the *Dpn* I restriction enzyme (10 U/μl) was added to the mixture, which was incubated in water bath at 37°C for one hour. The enzyme can specifically digest the parental non-mutated supercoiled dsDNA.

The transformation protocol with XL1-blue supercompetent cells was similar to that with *E. coli* BL21(DE3). 50 μl XL1-blue supercompetent cells were mixed with 1 μl *Dpn* I-treated DNA from sample reaction and control. To verify the transformation efficiency, 1 μl of pUC18 control plasmid (0.1 ng/μl) was mixed with the competent cells. After the heat pulse, 0.5 ml of NZY^+ broth which was preheated to 42°C was added to the mixture. 100 μl of 10 mM IPTG (in water) and 100 μl of 2% X-gal (dissolved in dimethylformamide) were spread on the LB agar with 100 μg/ml carbenicillin. For pWhitescript mutagenesis control, 250 μl of the mixture was plated on the LB agar; for pUC18 transformation control, 5 μl was plated; for sample mutagenesis, 250 μl was plated. All the plates were incubated at 37°C overnight. Blue cells of pWhitescript control grew up on the plates the next day. Two colonies from each mutagenesis reaction for Trp112Phe were selected and inoculated into LB broth with 200 μg/ml carbenicillin, which were incubated at 37°C overnight.

The plasmid was extracted according to the protocol of Strataprep Plasmid Miniprep Kit. The DNA solutions were washed with 55 μl of distilled water and the concentrations and purities were checked with Biophotometer with O.D. values at 260 nm and 280 nm. The DNA samples were sent to Nucleic Acid Research Facilities in VCU for sequencing.

Selection and Purification of Variants

The proteins for the experimental study are summarized in Table 1. In order to produce variants for folding while minimizing the effect of significant change to the side chain, two tryptophans were mutated to phenylalanines and Leu115 and Val121 were mutated to alanine. As a control, His160Gly was synthesized since His160 in human Fadd-DD aligns with Gly160 in murine homologue. Moreover, in order to produce variants with greater ellipticity in the near-UV region, two more variants are designed to introduce an exposed or partially exposed tryptophan residue to the protein, since the near-UV CD signal is largely caused by the asymmetric environment of aromatic residues in the protein (Kelly et al., 2005). Leu172 is an exposed residue even though it is hydrophobic by its physico-chemical property. Val158 is a partially exposed residue. To replace these two residues with tryptophans will not cause significant change to their microenvironments.

Table 1. Expression of Fadd-DD and variants used in experimental studies

Protein/variant		Property	Expressed location
WT			Cytoplasm
Trp112	Phe	Conserved	Cytoplasm
Trp148	Phe	Conserved	Inclusion bodies
His160	Gly	Non-conserved	Cytoplasm
Trp112/His160	Phe/Gly	Conserved	Cytoplasm
Leu115	Ala	Conserved	Inclusion bodies
Val121	Ala	Conserved	Inclusion bodies
Leu172	Trp	Conserved (hydrophilic position)	Cytoplasm
Val158	Trp	Non-conserved	Cytoplasm

The Trp112Phe, His160Gly and Trp112Phe/His160Gly recombinant proteins were expressed and purified in the same manner as the WT Fadd-DD. It should be noted that the yields of soluble Trp112Phe protein was substantially decreased in comparison to the WT Fadd-DD and His160Gly variant. The Trp148Phe variant however, was so significantly destabilized that instead of soluble expressed protein, this variant formed inclusion bodies within the bacteria. Purification therefore was initially done under denaturing conditions and refolding *in vitro* yielded large amounts of precipitated protein. Variations to the refolding buffer were tried in order to optimize the system for recovery of some native-like protein and found to be 20 mM Tris–HCl, pH 8.0, 150 mM NaCl and 2 mM DTT at 4°C.

Leu115Ala and Val121Ala variants are very unstable and exist as inclusion bodies, similar to the Trp148Phe variant. These two variants were purified in the similar manner as the Trp148Phe variant (Li et al., 2009). Leu172Trp and Val158Trp are stable

and exist in the cytoplasm, similar to the WT. These two variants were purified in the similar manner as the WT (Li et al., 2009).

Protein Isotope-labeling

Recombinant human Fadd-DD (11.8 kD) uniformly labeled with ^{15}N was purified from a six liter fermentation of *E. coli* BL21(DE3) (Novagen, NJ). The cells were grown at 37°C on M9 minimal medium supplemented with 1 g/l $^{15}NH_4Cl$, 0.2 mg/ml carbenicillin and 1 ml/l poly-Vi-Sol vitamin drops with iron (Mead-Johnson, Evansville, IN). The protein was purified according to the previously described protocols. The yield was approximately 15 mg/liter. The protein was lyophilized in preparation for denaturation and refolding in the quenched-flow studies.

Equilibrium Unfolding and Refolding Monitored by Fluorescence Spectroscopy

Fluorescence emission spectra were obtained with a Cary Eclipse fluorescence spectrophotometer (Varian, Palo Alto, CA). 0.05 mg/ml (4 μM) of WT Fadd-DD and all variants were excited at 295 nm and emission spectra were measured from 310 to 450 nm. Scans were repeated six to eight times at 20°C and averaged. The excitation and emission slits were set at 5 nm and 10 nm, respectively for the WT protein and His160Gly variant; 10 nm for Trp112Phe and Trp112Phe/His160Gly variants; and 20 nm for Trp148Phe variant. The path-length of the fluorescence cuvette was 1 cm. All the samples were in 20 mM Bis-Tris buffer (pH 6.2) with 2 mM DTT to keep the three free cysteines reduced except for Trp148Phe which was in Tris–HCl buffer (pH 8.0) and 2 mM DTT.

For the equilibrium unfolding process, WT Fadd-DD and variants were diluted to 0.05 mg/ml in increasing amounts of GndHCl buffer from 0 M to 6 M. For the refolding

titration, protein was denatured in 6 M GndHCl in buffer for 3 h. The protein was subsequently refolded by dilution to 0.05 mg/ml into decreasing concentrations of GndHCl from 6 M to 0 M. The samples were incubated at room temperature overnight. The fluorescence spectra of the samples were determined with emission monitored from 329 to 371 nm for the WT, Trp112Phe, Trp148Phe and Trp112Phe/His160Gly variants and from 320 to 380 nm for the His160Gly variant. The ratios of emission intensities of 370 over 330 were used for data analysis for all proteins. Fraction unfolded values were calculated using the method of Pace et al. (1989) and the curves were fitted using SigmaPlot.

Circular Dichroism and Equilibrium Unfolding

CD spectra were obtained with the Fast modular polarimeter (MOS-450) (Bio-logic, France). The protein concentrations for far-UV and near-UV CD were 0.2–0.25 mg/ml (16–21 μM) and 0.5–0.6 mg/ml (42–50 μM), respectively. Scans were repeated six to twenty-four times at 20°C and averaged. The slits were both set at 1 mm. The cuvette path-lengths for far-UV and near-UV CD were 1 mm and 1 cm, respectively. All the samples were in 20 mM Bis-Tris buffer (pH 6.2) with 2 mM DTT with the exception of the Trp148Phe variant which was in 20 mM Tris–HCl, pH 8.0, 150 mM NaCl and 2 mM DTT.

For the unfolding titration, WT Fadd-DD, Trp112Phe, His160Gly and Trp112Phe/His160Gly variants were diluted to 0.2 mg/ml in increasing amounts of GndHCl in buffer from 0 M to 6 M. The samples were incubated at room temperature overnight. The far-UV CD spectra of the samples were monitored with excitation from

221 to 223 nm. The ellipticity at 222 nm was used for data analysis. Fraction unfolded values were calculated and the curve was fitted using SigmaPlot.

Stopped-flow Fluorescence Studies

The folding of WT Fadd-DD, Trp112Phe and His160Gly variants were characterized by stopped-flow fluorescence with a SFM-400 (Bio-Logic, France). Native protein was denatured at 0.12 mg/ml to 0.3 mg/ml in 6 M GndHCl, 20 mM Bis-Tris buffer (pH 6.2) and 2 mM DTT at room temperature for 3 h or overnight. Refolding was initiated by five-fold dilution into 20 mM Bis-Tris buffer (pH 6.2) and 2 mM DTT at 20°C. The dead time was 8.6 ms. Changes in fluorescence intensity was monitored in an FC-15 cuvette (1.5 mm of path-length) with excitation at 295 nm and emission from 300–340 nm using a bandpass filter (Semrock, Rochester, N.Y.) selected for optimal signal for the WT Fadd-DD and the His160Gly variant. For the Trp112Phe variant another bandpass filter was selected (362–396 nm) to obtain the optimal signal. The excitation and emission slits were both 1 mm for WT Fadd-DD and 2 mm for the variants. The refolding experiments were repeated four to sixteen times and the averaged traces were fit to a double exponential equation in SigmaPlot for WT Fadd-DD and the His160Gly variant. For the Trp112Phe variant the trace was best fit to a triple exponential equation. The time course for all proteins was monitored up to 10 s.

The unfolding process was initiated with a 1:5 dilution at 20°C. Unfolding of native WT protein at 0.12 mg/ml in 20 mM Bis-Tris buffer (pH 6.2), 2 mM DTT and 1 M GndHCl was mixed with buffer with 7 M GndHCl and monitored by stopped-flow fluorescence spectroscopy. The excitation wavelength and dead time are the same with the refolding and emission was monitored using a bandpass filter (300–340 nm).

Secondary Structure Formation Studied with Stopped-flow Far-UV CD

The kinetics of refolding of Fadd-DD was measured at 20°C on a Bio-Logic MOS 450 stopped-flow instrument using far-UV CD detection. 0.9 mg/ml protein was unfolded in 20 mM MES buffer (pH 6.2), 5 mM DTT and 6 M GndHCl. Refolding experiments were carried out by rapid 1 to 5 dilution of the protein solution in 20 mM MES buffer (pH 6.2) at 20°C, giving the final concentrations of protein and GndHCl of 0.15 mg/ml and 1 M, respectively. The dead time was 9.3 ms. Change in the far-UV CD signal was monitored in an FC-20 cuvette (2 mm of path-length) at 225 nm. The excitation and emission slits were both 2 mm. The refolding experiments were repeated twenty times and the averaged traces were fit to a single exponential equation in SigmaPlot. The denatured baseline was obtained by mixing the denatured protein with denaturing buffer in the stopped-flow system and measuring the CD signal at 225 nm. The native baseline was determined by mixing native protein with refolding buffer in the stopped-flow system and measuring the CD signal at 225 nm. Both baselines were calculated from an average of 20 shots each.

Tertiary Structure Formation Studied with Stopped-flow Near-UV CD

The kinetics of refolding of Fadd-DD was measured at 20°C on a Bio-Logic MOS 450 stopped-flow instrument using near-UV CD detection. 1.5 mg/ml protein was unfolded in 20 mM Bis-Tris buffer (pH 6.2), 5 mM DTT and 5 M GndHCl. Refolding experiments were carried out by rapid 1 to 4 dilution of the protein solution in 20 mM Bis-Tris buffer (pH 6.2) at 20°C, giving the final concentrations of protein and GndHCl of 0.3 mg/ml and 1 M, respectively. The dead time was 5.6 ms. Change in the far-UV CD signal was monitored in an TC-100 cuvette (10 mm of path-length) at 285 nm. The

excitation and emission slits were both 2 mm. The refolding experiments were repeated thirty times.

Quenched-flow Experiments

The preparation of the samples for the HX methods utilized a Bio-Logic SFM-400 and the following delay lines: 17, 90 and 190 (Bio-Logic, France) at 20°C. ^{15}N-Fadd-DD (1 mg/ml) was denatured in 5 M deuterated GndHCl, 10 mM deuterated MES (pD 5.8) and 5 mM deuterated DTT overnight. This pD value was chosen because it was used in the previous folding study with stopped-flow fluorescence and CD spectroscopy. The pD of the deuterated buffer solution was determined by adding 0.4 units to the reading of the pH probe (Primrose, 1993). Refolding was initiated by mixing one volume of the denatured protein solution with four volumes of water-based refolding buffer containing 10 mM MES (pH 6.2) and 5 mM DTT. The concentrations of GndHCl and protein at this point are 1 M and 0.2 mg/ml, respectively. Under this concentration of GndHCl, the protein was shown to be in the native state and the folding of the protein at this concentration is not concentration dependent (Li et al., 2009). After refolding for a specified period of time the protein was mixed with five volumes of water-based pulsing buffer containing 50 mM glycine buffer (pH 9.8) and 5 mM DTT, thus subjected to a high pH pulse step for 5.4 ms. At this moment, the amide deuterium on the protein exchange with hydrogen in the solvent. The first three refolding times (9.9 ms, 28.1 ms, 53.2 ms) were used under the continuous mode and the last five (65 ms, 80 ms, 120 ms, 160 ms and 200 ms) were used under the interrupted mode (Table 2). The protein sample was then allowed to continue refolding at a lower pH followed by mixing with ten volumes of water-based quenching buffer containing 200 mM K_2HPO_4/100 mM citric

acid (pH 3.9) and 5 mM DTT. The final pulse and quench pH after dilution were 9.6 and 4.8, respectively. The average intrinsic exchanging times from amide deuterium to amide hydrogen in H_2O at 20°C are approximately 1.4 s and 0.6 ms at pH 6.2 and pH 9.6, respectively (http://hx2.med.upenn.edu/download.html). Therefore, the hydrogen labeling only occurs when the pulsing buffer was mixing with the protein solution. To prepare the samples for NMR studies the solutions were concentrated using vivaspin 20 concentrators (Sartorius Stedim, Aubagne, France) at 7,500 rpm at 4°C for approximately twelve hours immediately following the quenched flow experiments and stored overnight at 4°C. The next day the buffer of the solution was exchanged into 100 mM deuterated K_2HPO_4/50 mM deuterated citric acid (pD 4.4) and 10 mM deuterated DTT in D_2O for approximately twelve hours at 4°C, in order to exchange the amide hydrogens with no or weak hydrogen bond protection and then immediately used for NMR. A pD of 4.4 was selected because the number of stable amide protons was maximal in comparison to those at pD 7.2, 5.8 and 4.8.

Table 2. The set-up parameters of quenched-flow experiment

Continuous mode

DL1 (µl)	DL2 (µl)		Washing	Collecting	Ageing 1 (ms)	Ageing 2 (ms)
39.4	43.4	Time (ms)	50	50	9.9	5.4
		Syr. 1	40 µl	40 µl		
		Syr. 2	160 µl	160 µl		
		Syr. 3	200 µl	200 µl		
		Syr. 4	400 µl	400 µl		
		Total	800 µl	800 µl		
55.8	43.4	Time (ms)	50	50	13.9	5.4
		Syr. 1	40 µl	40 µl		
		Syr. 2	160 µl	160 µl		
		Syr. 3	200 µl	200 µl		
		Syr. 4	400 µl	400 µl		
		Total	800 µl	800 µl		
112.4	43.4	Time (ms)	85	50	28.1	5.4
		Syr. 1	68 µl	40 µl		
		Syr. 2	272 µl	160 µl		
		Syr. 3	340 µl	200 µl		
		Syr. 4	680 µl	400 µl		
		Total	1360 µl	800 µl		
212.6	43.4	Time (ms)	160	50	53.2	5.4
		Syr. 1	128 µl	40 µl		
		Syr. 2	512 µl	160 µl		
		Syr. 3	640 µl	200 µl		
		Syr. 4	1280 µl	400 µl		
		Total	2560 µl	800 µl		

Interrupted mode

	Phase 1 (µl)	Phase 2 (µl)	Phase 3 (µl)	Phase 4 (µl)
Time (ms)	158.6	T_2*	17.42	35.21
Syr. 1	127.5		14	28.3
Syr. 2	510		56	113.2
Syr. 3			70	141.5
Syr. 4			140	283
Total	637.5		280	566

The ageing time 1 increases with the volume of circles and the ageing time 2 is kept constant. DL: delay line. Washing and phases 1 to 3 are for line-washing and collecting and phase 4 are for sample-collection. The parameters are following the instructions of Bio-Logic.

* $DL_2 = T_2 + 53$. Therefore, T_2 = 12, 27, 67, 107, and 147, respectively for refolding times of 65 ms, 80 ms, 120 ms, 160 ms and 200 ms, respectively.

NMR Studies

The NMR experiments were conducted on a Bruker AVANCE III 400 MHz NMR spectrometer (Bruker-biospin, Billerica, MA) in College of Sciences Major Instrumentation Cluster, Old Dominion University. All NMR spectra (1D and 2D) were collected at 30°C. The acquisition time for 2D NMR is approximately 5.5 hours. In order to obtain residue specific information about the refolding of Fadd-DD from the HX experiments, the backbone peaks of ^{1}H-^{15}N-Heteronuclear Single Quantum Coherence (HSQC) spectra for the native Fadd-DD were obtained and compared with the literature (Berglund et al., 2000). Sensitivity enhanced ^{1}H-^{15}N HSQC spectra were obtained on a 5 mm inverse conventional probe with GARP decoupling in the ^{15}N channel. 1024 complex points were collected along the F2 dimension and 128 transients were accumulated along the F1 dimension. ^{15}N chemical shifts were referenced indirectly using the consensus chemical shift ratio of 0.101329118. All the HSQC spectra were generated in NMRPipe by apodizing the FID's with a cosine window, followed by zero-filling to the next power of 2, Fourier transformation and phasing in both F1 and F2 dimension and visualized with NMRDraw (Delaglio et al., 1995).

Data Analysis of Quenched-flow, HX and NMR

Hydrogen bonds were calculated using the program Contact (CCP4, 1994) and further confirmed with the molecular viewing and analysis program Insight II (Accelrys, CA). The HSQC peak intensities from the quenched-flow studies were normalized to a control experiment. In the control experiment, water based buffers were used in the unfolding and refolding portions of the quenched flow experiment. The protein was then concentrated and buffer exchanged as specified earlier. 1D proton NMR spectra were

used to calibrate the concentration difference among protein samples from different time points in TopSpin software (Bruker) (Nabuurs and Mierlo, 2010). The absolute peak intensity of the well-resolved resonance at -0.37 ppm, which corresponds to a non-exchangeable methyl proton resonance from L119, is a direct measure of the protein concentration in the corresponding NMR sample. The HSQC peak intensities were normalized using the absolute peak intensities of the upfield resonances in the 1D spectra in order to correct the small variation of protein concentration in the same samples. The resultant data were plotted using SigmaPlot (version 10) (Systat Software, Chicago, IL) and kinetic rates determined by applying the single exponential equation to twenty-two peaks monitored during the study were analyzed. The peak intensities were calculated with NMRDraw.

Equilibrium Hydrogen Exchange

In the equilibrium HX study, ^{15}N-Fadd-DD (4 mg/ml) was buffer exchanged into 100 mM deuterated K_2HPO_4/50 mM deuterated citric acid and 50 mM deuterated DTT in 99.0% D_2O (pD 4.4). Each HSQC spectrum was acquired every eight hours over the course of one week. HX rates were determined by fitting peak intensities from HSQC spectra as a function of time using a single exponential equation in SigmaPlot. The intensities of each identified peak are normalized against that in the first HSQC spectrum. The protection against exchange rate is expressed as protection factor, which is the ratio between the sequence specific intrinsic exchange rate for an amide proton k_{int}, and measured exchange rate k_{ex} (Bai et al., 1993; Sasakawa et al., 1999). k_{int} is calculated using an intrinsic exchange rate program (H in D_2O) located at http://hx2.med.upenn.edu/download.html. The protection factors are calculated for the slowly exchanging peaks.

METHODS FOR THE BIOINFORMATICS STUDY OF THE CHITINASE INSERTION DOMAIN (CID)

Construction of a Multiple Sequence and Structure Alignment of the CID

The CID regions within the structures of three proteins: *Bacillus circulans* chitinase A1 (1ITX), *C. immitis* chitinase (1D2K) and human chitotriosidase (1LG1) were used as query sequences in PSI-BLAST to search for distant relatives. They represent chitinases within the kingdoms of Bacteria, Fungi and Animalia, respectively. A plant or archaeal structure was not available at the time however the PSI-BLAST searches did identify plant and archaeal chitinases for inclusion in our study. An initial multiple sequence alignment was made using MUSCLE in Jalview (Clamp et al., 2004; Edgar, 2004). In the searched sequences, some from close relatives have high identities >40% (data not shown). Five sequence relatives from each of the five kingdoms and two from early eukaryotes with sequence identities less than 40% were chosen to make twenty-seven representatives of the CID superfamily (Table A7). The alignment was created in order to enhance sequence variability and in this way, only the key conserved residues for structure, folding and function could be identified. The boundary of the CID in each sequence was identified by aligning with the three model chitinases and the domain was further extracted from each chitinase sequence.

An initial structure alignment containing the CIDs from 1ITX, 1D2K and 1LG1 was generated with the online CE-MC (Guda et al., 2001). The initial sequence alignment was compared with the initial structure alignment and adjusted in Jalview to ensure the sequences with unknown structures were properly aligned with the known structures. Since no structure from plant is available, the secondary structure of tobacco chitinase

CID was predicted by the program of PSIPRED (McGuffin et al., 2000) and the other sequences were aligned with it thereafter. To verify our sequence and structure alignment, structures of eight members of family 18 chitinases (1HKM, 1LJY, 1ITX, 1D2K, 1FFR, 1UR9, 1KFW and 1NWT) were superimposed with CE-MC method (Guda et al., 2001). Two structures, 1D2K and 1ITX, are within the initial alignment with twenty-seven sequences.

A larger multiple sequence alignment of sixty sequences was generated using MUSCLE in Jalview, without being edited according to the three model structures. The large alignment includes the twenty-seven CID sequences from Archaea, Bacteria, Fungi, Plantae and Animalia and thirty-three more sequences from Bacteria, Fungi and Animalia (Table A7) were acquired from searches of the protein database using the PSI-BLAST program. Furthermore, the SAM-T08 program was employed to search for the conserved residues in the CID (http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html) (Karplus, 2009).

Conservation and Hydropathy Analysis

The number of each residue in each position was calculated and analyzed by SigmaPlot 10.0 (SYSTAT Software Inc.). The entropy value and hydropathy were calculated by the same equations previously mentioned. The definitions of $Pj(i)$, m, $C(i)$ are the same as mentioned in the Section of "methods for bioinformatics study of Fadd-DD". The positions with conservation values greater than 0.45 were considered to be highly conserved; the positions with conservation values between 0.35 and 0.45 were considered to be moderately conserved; and those positions with conservation values lower than 0.35 were considered to be less conserved (Greene et al., 2003). The positions

which have more than one gap are considered non-conserved and therefore have a value of zero. The hydrophobicity scale of Nozaki and Tanford was used for our studies (Nozaki and Tanford, 1971).

Select structures from the designated family 18 chitinases in SCOP (http://scop.mrc-lmb.cam.ac.uk/scop/) and CAZy (http://www.cazy.org/fam/GH18.html) were chosen to compare the structure and function of chitinases and chitinase-like proteins (see Table A8). Protein data bank (PDB) files were obtained from SCOP and RCSB (http://www.rcsb.org). All PDB files were visualized and analyzed in either Insight II, version 2005 (Accelrys, CA), Pymol, version 0.99 (DeLano Scientific, CA), or Rasmol, version 2.7. Hydrogen bond calculations and van der Waals radii were determined with Insight II.

Phylogenetic Analysis of the CID

In order to investigate the evolutionary relationship of the CID sequences from different lineages of life, the ClustalW2 program (http://www.ebi.ac.uk/Tools/clustalw2/index.html) was performed with the sixty CID sequences, because the program can produce a multiple sequence alignment of divergent sequences and Cladogram or Phylogram to visualize the evolutionary relationships (Larkin et al., 2007). The phylogenetic tree was constructed using the neighbor-joining algorithm as described by Saitou and Nei (1987). The tree was visualized and drawn with MEGA version 4.0.2 software (Kumar et al., 2008).

CHAPTER III

ANALYSIS OF CONSERVATION IN THE FAS-ASSOCIATED DEATH DOMAIN PROTEIN AND CHITINASE INSERTION DOMAIN

INTRODUCTION

Within the vastness of conformational space the greatest number of possible conformations exists at the onset of folding. The restriction of conformational space and the formation of native-like topology prior to the establishment of the finer details of the three-dimensional structure may in part be based on the early formation of a conserved network of interactions which facilitates rapid and correct folding (Greene et al., 2003; Geierhass et al., 2004; Higman and Greene, 2006). A number of computational and experimental studies provide insight into the link between conserved amino acids and folding (Martinez et al., 1999; Mirny and Shakhnovich, 1999; Kloczkowski and Jernigan, 2002; Ting and Jernigan, 2002; Greene et al., 2003; Gunasekaran et al., 2004; Guo et al., 2004; Zarrine-afsar et al., 2005; Pearce et al., 2007) particularly those experiments which reveal that the conserved residues are preferentially structured in the TS (Kragelund et al., 1999; Hamill et al., 2000; Fowler and Clark, 2001; Heidary and Jennings, 2002; Otzon and Oliveberg, 2002; Hubner et al., 2004; Wilson and Wittung-Stafshede, 2005; Cambell-Valois and Michnick, 2007; Olofsson et al., 2007; Lappalainen et al., 2008). While there are a few studies which argue against conservation of a folding nucleus (Larson et al., 2002; Tseng and Liang, 2004), this avenue of investigation is very promising and new approaches to finding determinants of structure and folding from an evolutionary perspective continue to emerge (Socolich et al., 2005; Marcelino et al.,

2006). We report the results of a computational investigation into the determinants of protein folding and stability by identifying conserved residues in the death domain superfamily and the corresponding long-range interactions. The death domain superfamily fold consists of six α-helices arranged in a Greek-key topology which is shared by the all β-sheet immunoglobulin and mixed α/β-plait superfamilies (Higman and Greene, 2006) (Figure 6). The Greek-key topology in general is one of the most prevalent in nature (Higman and Greene, 2006; Cuff et al., 2009).

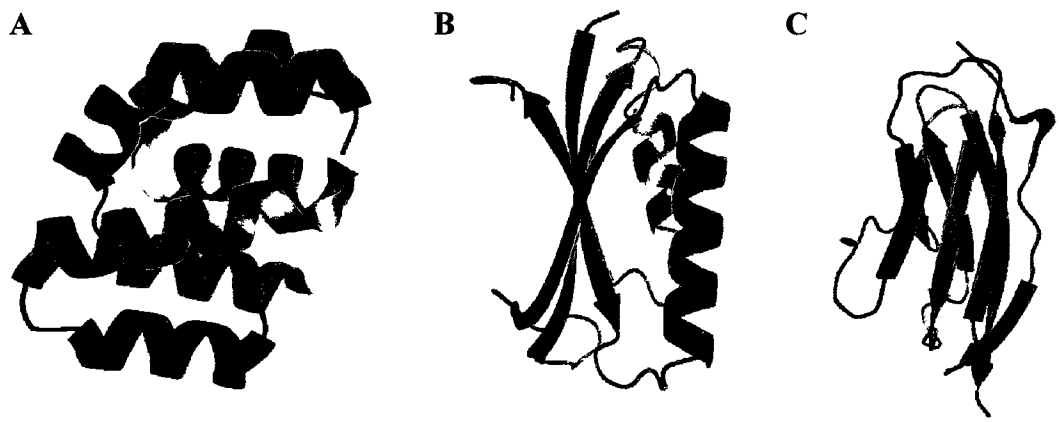


Figure 6. The three superfamilies which share the same Greek-key topology

(A) Apaf-1 CARD (PDB code: 1CY5) represents the death domain superfamily.

(B) Ribosomal S6 (PDB code: 1RIS) represents the α/β plait superfamily.

(C) Titin (PDB code: 1TIT) represents the immunoglobulin superfamily. Ribbon structures created with PyMOL.

Sequence and structural studies show that sixteen conserved hydrophobic residues located within the interior of our model superfamily member, Fadd-DD (see Figure 7),

form a network of long-range interactions, which we propose may play an important and interconnected role in governing protein folding, topology and stability. Additionally, the conservation analysis identified six conserved hydrophilic residues which we propose play a common role in protein–protein interactions between the diverse DDs.

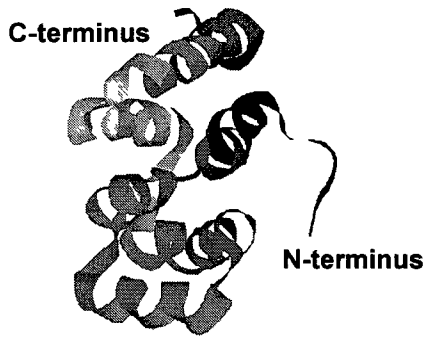


Figure 7. Structural analysis of Fadd-DD (PDB code: 1E3Y)

Ribbon representation of NMR structure of Fadd-DD created by Rasmol (Ver. 2.7).

Chitin and Chitinase

An unrelated side project involved the bioinformatic analysis of conservation in the chitinase insertion domain to further our understanding of the function. This work did however involved the identical approach used to analyze the death domains. Thus in this instance, it is interesting to look at conservation for function versus structure and folding.

Chitin $(C_8H_{13}O_5N)_n$ is a long-chain polymeric polysaccharide of β-glucosamine that forms a hard, semi-transparent material found throughout nature. Chitin is composed of units of N-acetyl-D-glucos-2-amine, which are linked by β-1,4 glycosidic bonds

(Gooday, 1990). Hence, it may also be described as cellulose with one hydroxyl group on each monomer replaced by an acetylamine group. Chitin is the main component of the cell walls of fungi (Gooday, 1990), the shells and radulae of molluscs and of the exoskeletons of arthropods, especially crustaceans and insects (Dahiya et al., 2006).

The breakdown of chitin is catalyzed by chitinases which hydrolyze it to simple sugars. Chitinases can be divided into two major categories: exochitinases and endochitinases (Dahiya et al., 2006; Li, 2006). Exochitinases can be further divided into two subcategories: chitobiosidases, which cleave diacetylchitobiose units from the non-reducing end of the chitin chain, and β-(1,4)-N-acetyl-glucosaminidases (NAGase), which cleave the N-acetylglucosamine (NAG) oligomers, generating NAG monomers. Endochitinases cleave glycosidic linkages randomly at internal sites along the chitin chain, eventually providing a variety of low molecular mass NAG oligomers such as diacetylchitobioses and chitotrioses (Dahiya et al., 2006; Li, 2006).

Chitinases occur in a wide range of organisms including bacteria, fungi, plants, insects and animals. Chitinases from bacteria and fungi are extremely important for maintaining a balance between the large amount of carbon and nitrogen trapped in the biomass as insoluble chitin in nature (Aronson et al., 2003; Li, 2006). Chitinases are needed by fungi to disrupt existing cell walls when normal cells divide (Kuranda and Robbins, 1991) and chitinases from some plants may be essential in inhibition against fungal pathogens (Taira et al., 2002). In insects and crustaceans, chitinases are associated with degradation of old cuticle (Merzendorfer and Zimoch, 2003). Additionally, human chitotriosidase may be important in defence against chitinous pathogens such as *Candida albicans* (Renkema et al., 1998; van Eijk et al., 2005).

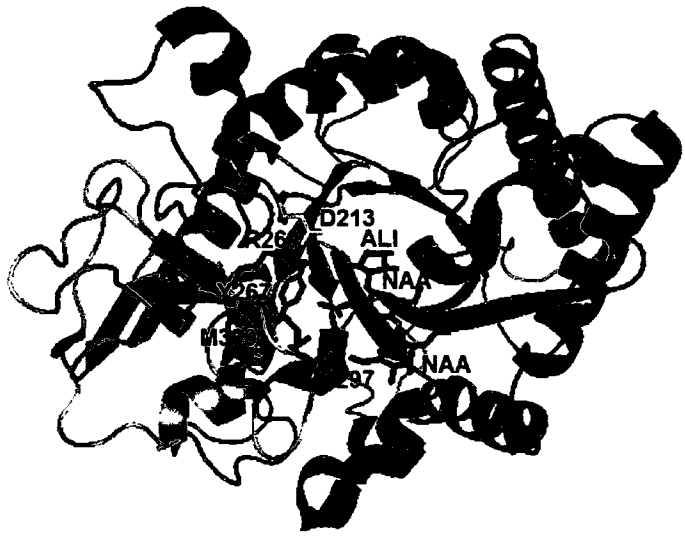
Based on amino acid sequence similarity, chitinases are classified into families 18 and 19 of glycoside hydrolases (GH) (Kawase et al., 2004; Funkhouser and Aronson, 2007). The members of the two different families differ in their amino acid sequences, three-dimensional structures and molecular mechanisms of catalytic reactions (Aronson et al., 2003). Family 18 chitinases have catalytic domains of triosephosphate isomerase (TIM barrel) fold with a conserved DxDxE motif (Vaaje-Kolstad et al., 2004) and catalyze the hydrolytic reaction by a substrate-assisted mechanism (Terwisscha van Scheltinga et al., 1996; van Aalten et al., 2001), whereas family 19 chitinases have high percentage of α-helices and adopt the single displacement catalytic mechanism (Brameld and Goddard, 1998; Hoell et al., 2006). In family 18 chitinases, the leaving group alcohol is protonated by a conserved glutamic acid, the sugar at −1 subsite is distorted into a boat conformation, and an oxazolinium intermediate is stabilized by the sugar N-acetamido group and then hydrolyzed (van Aalten et al., 2001; Songsiriritthigul et al., 2008). In an exochitinase *S. marcescens* chiB, it was proposed that binding of substrate causes the −1 sugar ring to distort to a boat conformation and rotation of Asp142 towards Glu144, thus enabling hydrogen bonding between the acetamido group, Asp142 and Glu144. Later on the oxazolinium ion intermediate was hydrolyzed, leading to protonation of Glu144 and rotation of Asp142, which shares a proton with Asp140 (van Aalten et al., 2001) (Figure A3). Family 18 chitinases are widely distributed in five lineages of life; for example, *Thermococcus kodakarensis* (Fukui et al., 2005) in Archaea, *Serratia marcescens* (*S. marcescens*) (Brurberg et al., 1994) in Bacteria, *Coccidioides immitis* (*C. immitis*) (Hollis et al., 2000; Bortone et al., 2002) in Fungi, tobacco (Melchers et al., 1994) in Plantae, and

the sandfly (Ramalho-Ortigao and Traub-Cseko, 2003) and human (Fusetti et al., 2002) in Animalia.

Family 18 Chitinases

Family 18 chitinases can be classified into three subfamilies A, B and C, in terms of the amino acid sequence similarity (Watanabe et al., 1993). The main structural difference between subfamilies A and B chitinases is that a small $\alpha + \beta$ domain inserts into the TIM barrel catalytic domain in the subfamily A, while this insertion domain is absent in the subfamily B (Suzuki et al., 2002). For example, human chitotriosidase (PDB code: 1HKM), as a family 18 chitinase in the subfamily A, has a TIM domain and a CID, which is a module inserted into the TIM barrel (Figure 8A). In the subfamily A, other additions can occur at N- or C- terminus of the TIM barrel. On the other hand, *S. marcescens* chitinase C (chiC), belonging to the subfamily B, has a catalytic domain, a fibronectin type III-like domain and a chitin-binding domain (Suzuki et al., 2002). Therefore the presence or absence of the insertion domain appears to be subfamily specific (Suzuki et al., 1999). Examples of family 18 chitinases in the subfamily B are only limited to a few bacteria, such as *S. marcescens* and *B. circulans* (Watanabe et al., 1993; Suzuki et al., 1999). Here we mainly discuss family 18 chitinases in the subfamily A.

A



B

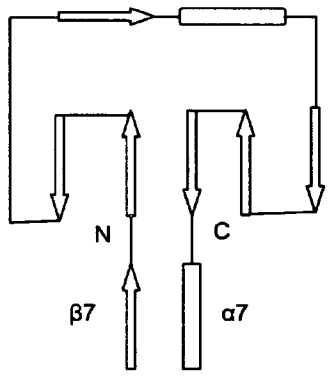


Figure 8. Structural analysis of the CID

(A) Ribbon model of human chitotriosidase (PDB: 1HKM) in complex with the substrate (NAA-NAA-ALI) generated by Pymol, showing the TIM barrel and CID. The helices and strands on the TIM barrel are colored in green and those on the CID are colored in light blue. Some residues (Tyr267, Arg269, Glu297 and Met300) in blue on the CID and Asp213 in yellow on the TIM barrel interact with the substrate in red.

(B) Schematic representation of the CID between β7 and α7 on the TIM barrel, which is composed of two anti-parallel β-strands followed by one β-strand, one short α-helix and

Figure 8 Continued.

lastly three anti-parallel β-strands. The arrows indicate β-strands and the rectangles are α-helices. The lines stand for the loops connecting α-helices or β-strands.

The TIM barrel domain consists of an $(\alpha/\beta)_8$-barrel fold and has been found in many different proteins, most of which are enzymes. The TIM barrel domains share low sequence identity and have a diverse range of functions. The specific enzyme activity is determined by the eight loops at the carboxyl end of the β-strands (Branden and Tooze, 1998). In some TIM barrels, an additional loop from a second domain approaches the active site of the TIM domain and participates in binding and catalysis (Branden and Tooze, 1998; Pestsko and Ringe, 2004).

The CID superfamily has only one family member and is classified as having an FKBP-like fold in the SCOP database (Figure 8B) (Murzin et al., 1995). The CID is composed of five or six anti-parallel β-strands and one α-helix and it inserts between the seventh α-helix and seventh β-strand of the TIM barrel (Srivastava et al., 2006). The CID forms a wall alongside the substrate-binding cleft of TIM barrel of chitinase which increases the depth of the cleft. Thus, it is easy to imagine that the substrate-binding cleft of chitinases from the subfamilies B and C is not as deep as that from the subfamily A (Suzuki et al., 1999). Interestingly, some mammalian glycoproteins with various functions also exhibit the fold of a family 18 chitinase, such as human cartilage glycoprotein-39 (HCgp-39), whose structure consists of a TIM domain and a CID (Fusetti et al., 2003).

In addition to the TIM domain and the CID, some bacterial chitinases in the subfamily A involved in chitin degradation contain one or two additional domains involved in substrate-binding (van Aalten et al., 2000). For example, *S. marcescens* chitinase A (chiA) (PDB code: 1CTN/1FFR) has an additional N-terminal domain (Papanikolau et al., 2001) which belongs to the E-set domain superfamily in SCOP, whereas *S. marcescens* chitinase B (chiB) (PDB code: 1E15/1UR9) has one extra C-terminal domain (Vaaje-Kolstad et al., 2004) which belongs to the carbohydrate-binding domain superfamily. Removal of such domains often results in enzymes that are still active but show extremely impaired binding to substrates (van Aalten et al., 2000; Katouno et al., 2004). For example, mutagenesis studies of two tryptophans on the N-terminal domain of chiA resulted in decreased specific hydrolyzing activity thus showing their importance for the hydrolysis of β-chitin (Uchiyama et al., 2001; Aronson et al., 2003; Ferrandon et al., 2003).

Four Conserved Residues on the CID May Play an Important Role in Chitinase Function

As known previously, the TIM barrel is considered the catalytic domain in family 18 chitinases (Uchiyama et al., 2001; Aronson et al., 2003). Although a number of previous publications showed interactions between a group of residues on the CID and the enzyme substrate and reported the possible functional significance of the CID (van Perrakis et al., 1994; Hollis et al., 2000; Aalten et al., 2001; Papanikolau et al., 2001; Fusetti et al., 2003), the definitive role of the CID in chitinase function has not been completely determined (Fusetti et al., 2002; Fusetti et al., 2003; Songsiriritthigul et al., 2008). For example, the functional contribution of the CID is not clear in the case of *S.*

marcescens chiA (Zees et al., 2009). A previous study showed that by removing the CID from *S. marcescens* chiA, the thermal stability was reduced, the specific activity was decreased, the pH optimum was shifted lower, and the catalytic activity towards long chitin derivatives was lost (Zees et al., 2009). However, none of the residues on the CID have been individually mutated. Hence, the role of the specific residues in binding with substrates remains to be identified.

To identify the specific functional residues on the CID, a multiple sequence and structure alignment of this domain was constructed. The sequence search process revealed that this domain exists in a wide range of organisms. Conservation and hydropathy analysis revealed that four conserved residues, constituting two distinct sequence motifs, interact with the substrate. Furthermore, extensive comparisons among different family 18 chitinases demonstrated that the TIM domains + CID can bind long-chain substrates by providing a deep substrate-binding cleft, while this may not be the case for the enzymes with the TIM domain alone. In general, additional modules fused to a catalytic domain may play a role in substrate specificity by providing a specific binding site or shaping the active site to recognize a substrate with a different shape or size (Todd et al., 2001). We extrapolate that this may be a reason for the insertion of the CID into the TIM barrel. This paper identifies and provides initial computational support for the importance of conserved residues on the CID in chitinase function.

RESULTS

Bioinformatics and Network Studies of the Death Domain Superfamily

A sequence alignment with twenty proteins from the death domain superfamily was constructed as described in the materials and methods (Figure 9; Table A9).

Hydrophobic residues such as leucine and tryptophan are preferentially aligned with the hydrophobic residues, while hydrophilic residues such as lysine and glutamine are preferentially aligned with hydrophilic residues. A conservation analysis was conducted using a modified entropy parameter (Figure 10A) and hydrophobicity measure (Figure 11) for each position in the superfamily alignment. Ten positions were found to have conservation scores greater than 0.45, indicating that they are highly conserved. Among these ten positions, nine are persistently hydrophobic and one persistently hydrophilic. The residues in Fadd-DD that correspond to the hydrophobic positions are: Ala100, Asn107, Leu119, Ser144, Leu145, Trp148, Leu165, Val173 and Val177 (Figure 12A). Thirteen positions had conservation scores between 0.35 and 0.45 and are designated as moderately conserved. Among the thirteen positions of intermediate conservation, seven are persistently hydrophobic, five are persistently hydrophilic and one neutral. The residues in Fadd-DD that correspond to the hydrophobic positions are: Trp112, Leu115, Val121, Ile129, Leu161, Ala164 and Ala174 (Figure 12A). The remaining positions (< 0.35) are designated as less conserved. Therefore, sixteen conserved hydrophobic residues and six conserved hydrophilic residues were identified (Figure 12A–B).

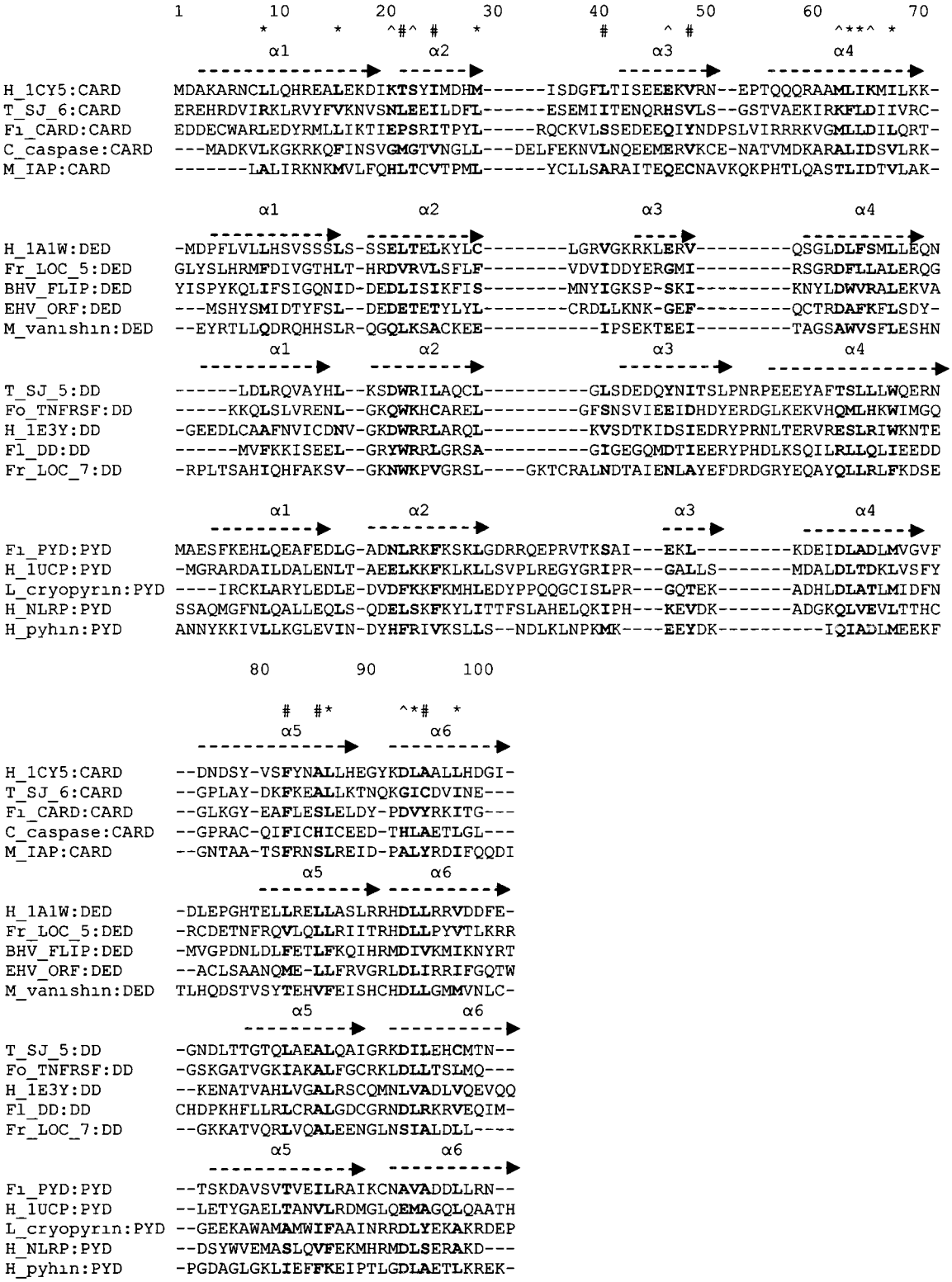
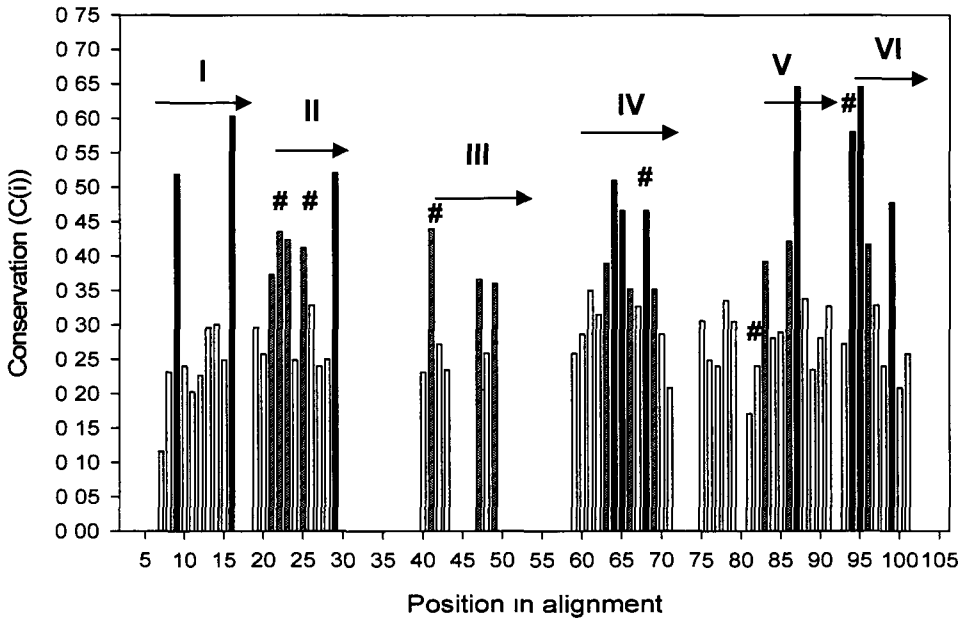


Figure 9. Multiple sequence and structure alignment of twenty proteins from death domain superfamily

Figure 9 Continued.

Hydrophobic positions with high conservation are colored in blue and annotated with "*". Hydrophobic positions with intermediate conservation are colored in green and annotated with "#". Conserved hydrophilic positions are colored in red and annotated with "^". Arrows indicate the sequences that are in the six α-helices. The number scheme represents positions in the alignment.

A



B

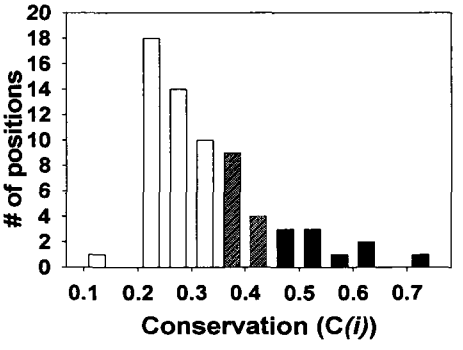


Figure 10. Sequence conservation and hydropathy analyses of death domain superfamily

(A) Black bars represent highly conserved positions ($C(i) \geq 0.45$) (ten positions); gray hashed bars represent moderately conserved sites ($0.35 \leq C(i) < 0.45$) (thirteen positions); white bars represent the less conserved positions ($C(i) < 0.35$). Six helices are annotated with roman numerals on arrows. Residue positions annotated with a "hash mark" represent Trp112, Leu115, Val121, Trp148, His160 and Leu172.

(B) The small diagram shows the distribution of conservation scores ($C(i)$).

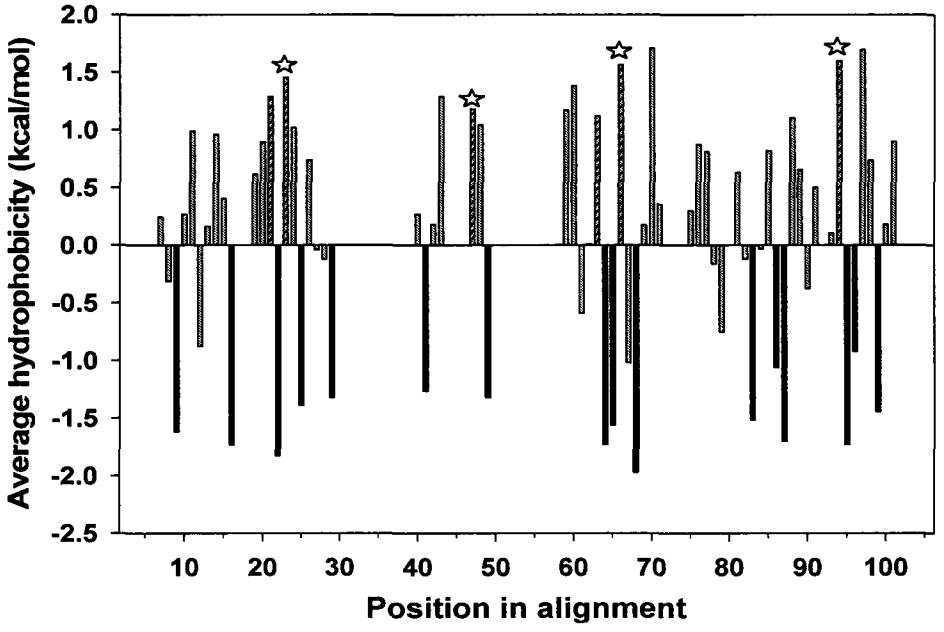
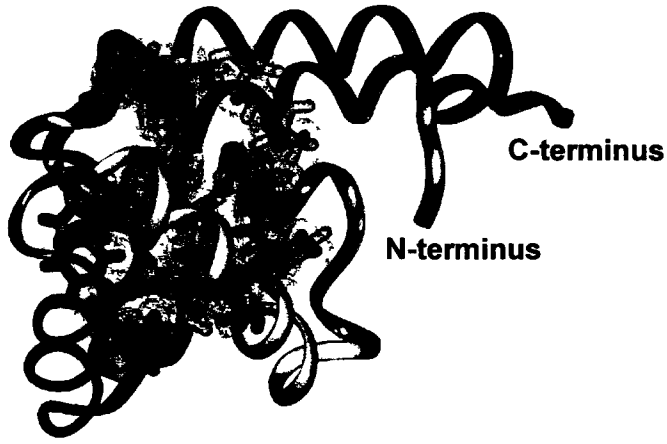


Figure 11. Average hydropathy profile analysis in the death domain superfamily

Conserved hydrophobic residues are represented by black bars ($C(i) \geq 0.35$) (sixteen positions). Conserved hydrophilic residues are represented by gray hashed bars ($C(i) \geq 0.35$) (six positions). Residue positions annotated with a "star" were found to be important in the interaction with Fas-DD in mutagenesis studies.

A



B

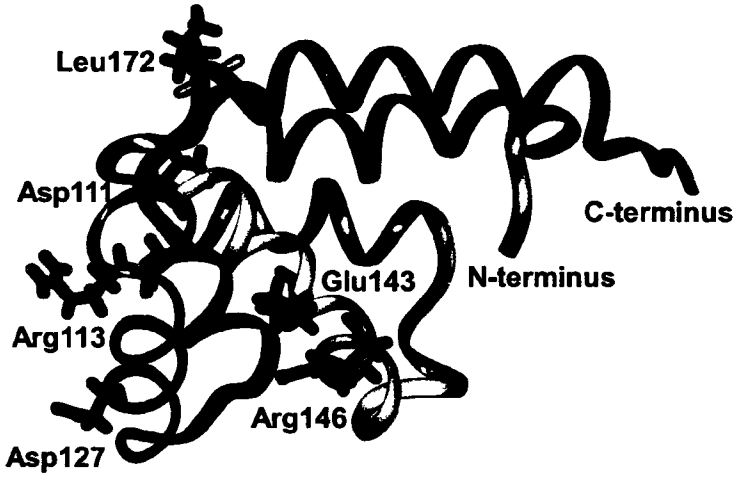


Figure 12. Ribbon diagram showing conserved residues

(A) Ribbon diagram generated by Insight II (Ver. 2005) shows highly conserved (blue) and intermediately conserved (green) hydrophobic residues. The helices are colored 1–6 as follows: purple, blue, green, yellow, orange, red. Van der Waals radii around atoms of the conserved residues show the hydrophobic core.

(B) Ribbon diagram shows conserved hydrophilic residues and numbers.

The long-range interaction network within Fadd-DD is illustrated in Figure 13A. Long-range interactions were denoted by contacts that occur between residues that are ten or more residues apart in the primary structure and within 6 Å in the tertiary structure (Greene and Higman, 2003). There are twenty-two interactions between the sixteen conserved residues. These interactions are all between predominantly hydrophobic residues in the core. If two pairs of residues from two families both form a long-range interaction and these two pairs belong to the same position in the alignment, then this interaction is designated to be conserved. Furthermore, eleven of the interactions in the long-range interaction network in Fadd-DD are conserved in the other three families within the death domain superfamily (Figure 13B).

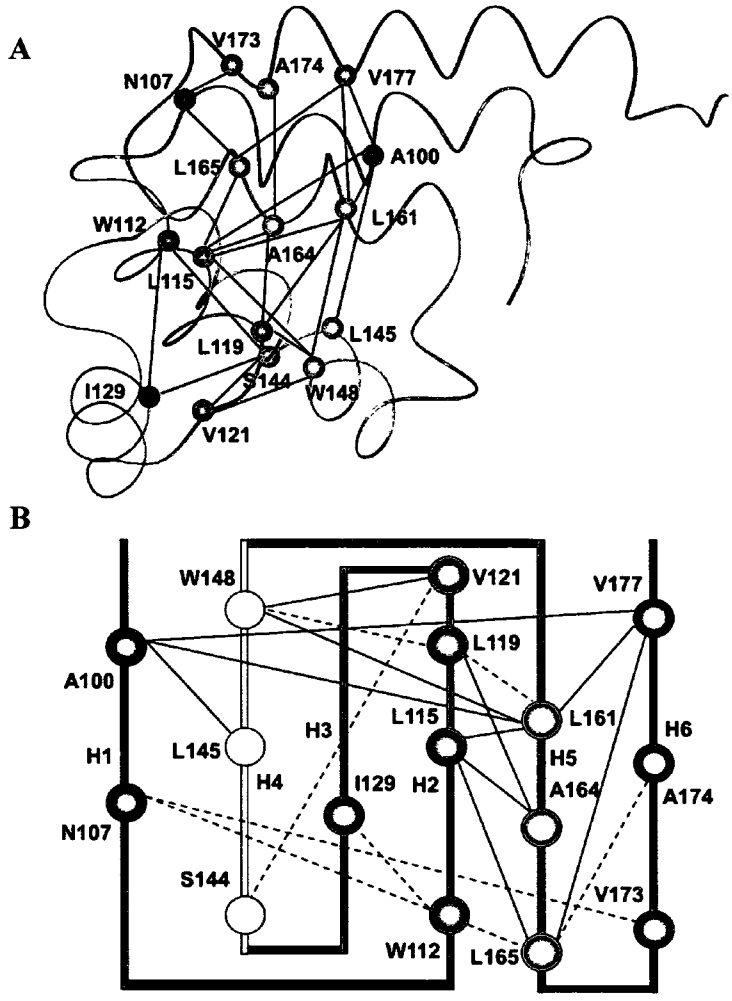


Figure 13. Network of long-range interactions between the conserved residues in Fadd-DD

(A) The circles indicate the locations and identities of the sixteen conserved residues in the Fadd-DD structure. The black lines indicate contacts that are within 6 Å and involve only heavy atoms.

(B) The balls labeled with numbers represent the sixteen conserved residues. Helices 1 to 6 are colored in blue, cyan, green, yellow, orange and red, respectively. The turns are colored in gray. The solid lines are the conserved long-range interactions within the four death domain subfamilies; the dashed lines are the conserved long-range interactions within three of the four subfamilies. The helices are denoted H1 – H6.

Structure-based Sequence Alignment of the CID

The representative family 18 chitinases and chitinase-like proteins from plants, bacteria, fungi and animals with determined three-dimensional structures are listed in Table A8. A multiple sequence alignment of twenty-seven CIDs based on the structures of three model proteins: *B. circulans* chitinase A1 (PDB code: 1ITX), *C. immitis* chitinase (PDB code: 1D2K) and human chitotriosidase (PDB code: 1LG1) was generated by MUSCLE in Jalview (Figure A4). CIDs from organisms in all five kingdoms are aligned, including Archaea, Bacteria, Fungi, Plantae and Animalia (Figure A4). The secondary structure of the CID of tobacco chitinase is quite similar to those of fungal chitinases, and thus the β-strands and α-helix of plant CIDs can be predicted. Smaller alignments can be found in the following references: van Aalten et al. (2000); Varela et al. (2002); Fusetti et al. (2002); Srivastava et al. (2006). Because the CID in chitinases is conserved, we can identify its sequence boundaries. Furthermore, we can predict its structure in family 18 chitinases with absence of solved structures.

Eight chitinase and chitinase-like structures including the three model chitinases and five more structures (PDB codes: 1LJY, 1FFR, 1UR9, 1KFW and 1NWT; explained in Table A8) were superimposed on each other based on the CE-MC method (see Figure 14). Furthermore, a second and larger sequence alignment with sixty CID sequences was generated using MUSCLE (see Figure A5).

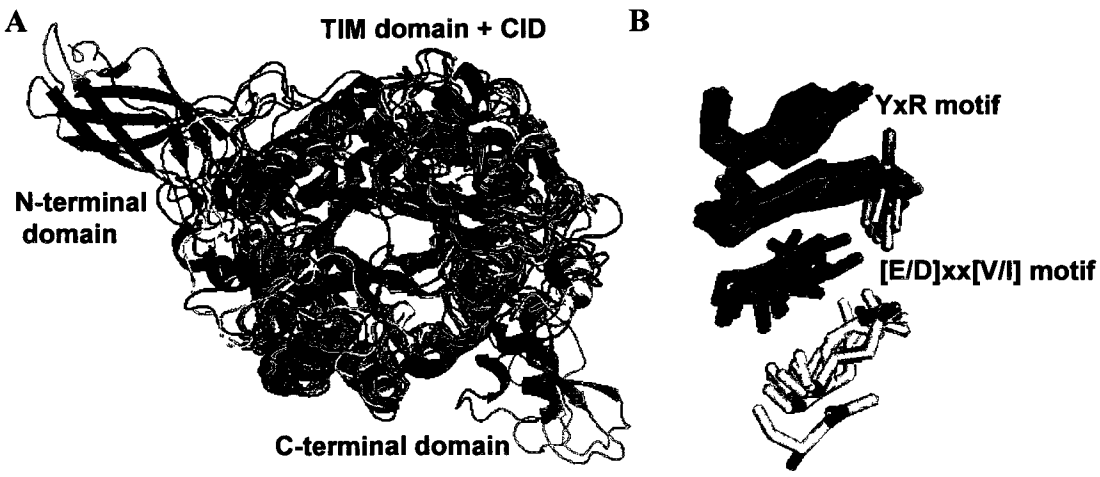


Figure 14. Superimposition of eight family 18 chitinases and chitinase-like structures (PDB codes: 1HKM, 1LJY, 1ITX, 1D2K, 1FFR, 1UR9, 1KFW and 1NWT) and the two conserved motifs on the CIDs

(A) The eight structures 1HKM, 1LJY, 1ITX, 1D2K, 1FFR, 1UR9, 1KFW and 1NWT are coloured in red, orange, yellow, green, blue, cyan, purple and black, respectively. The aligned, blue and cyan parts are TIM domain + CID, N-terminal domain on 1FFR and C-terminal domain on 1UR9, respectively. The structures were superimposed with the CE-MC method.

(B) The two positions in the YxR motif are shown in red and orange, respectively; and the two positions in the [E/D]xx[V/I] motif are shown in yellow and blue, respectively.

Proposed Role of Conserved Residues on the CID

Residues are often conserved in protein families because they either make critical stabilizing interactions or play important functional roles (Schueler-Furman and Baker, 2003). Additionally, residues important for stability are clustered together in the hydrophobic core and functional residues may be close together in protein-ligand binding sites (Schueler-Furman and Baker, 2003). Therefore, an analysis of residue conservation is a reasonable approach in which to identify functionally important sites in the CID.

Positions of highly and moderately conserved residues (Figure 15) and the average hydropathy profile analysis (Figure 16) are shown. Our conservation study indicated that there are nine hydrophobic positions with high conservation and five with moderate conservation; five hydrophilic positions with high conservation and two with moderate conservation; and five neutral positions with high conservation and six with moderate conservation (Figure A4). Among these conserved positions, two hydrophobic and two hydrophilic on the CIDs in chitinases denoted by PDB codes 1LG1, 1D2K and 1ITX are proposed to be important for interactions with the substrate, and five for the formation of the hydrophobic core, as well as the stabilization of the domain (Table 3). Interestingly, these four residues fall into two characteristic motifs, one in the N-terminal region and one in the central region, which are termed the YxR motif and the [E/D]xx[V/I] motif, respectively. These two motifs are also conserved in the larger multiple sequence alignment (see Figure A5) as well as the structural superimpositions (see Figure 14B). It should be noted that the use of SAM-T08 program also identified the two conserved motifs.

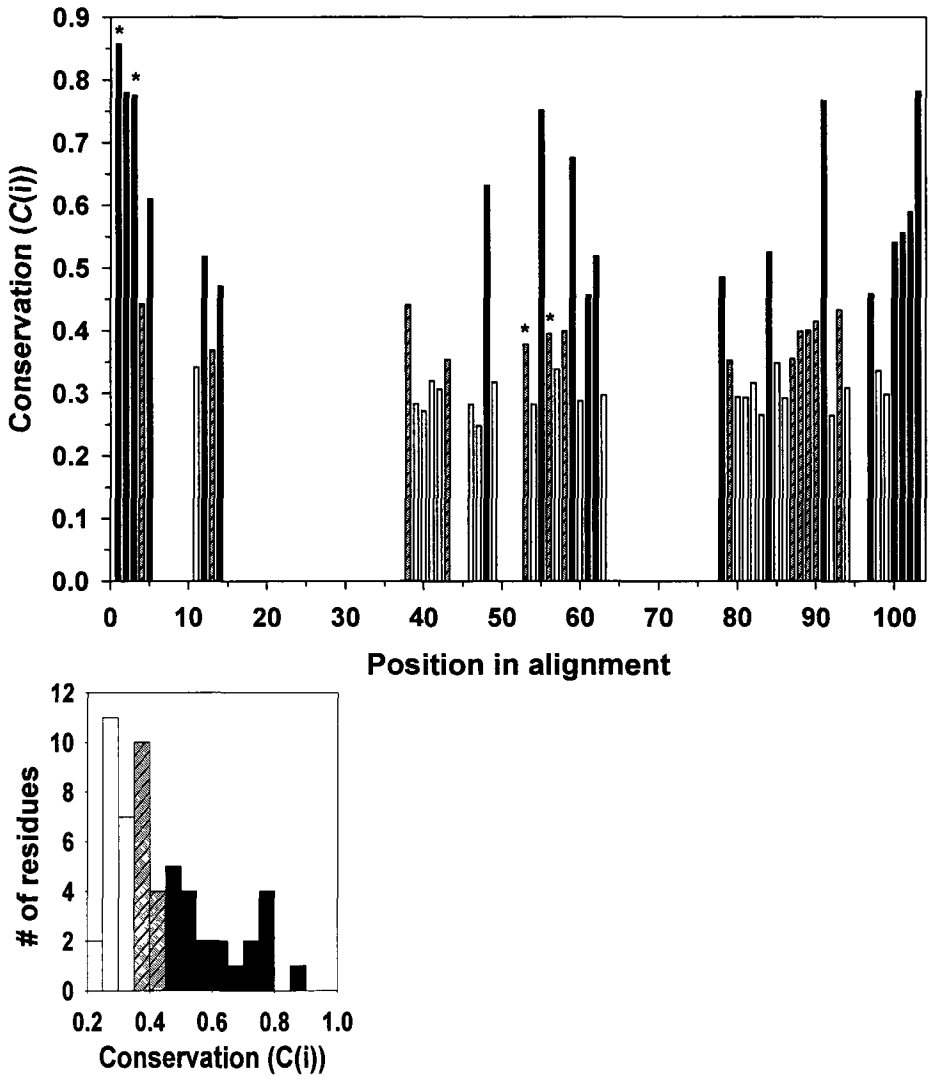


Figure 15. Sequence conservation analysis of the CID (alignment with 27 sequences)

The figure shows the distribution of conservation scores ($C(i)$). Positions with high conservation are represented by black bars ($C(i)\geq 0.45$), positions with moderate conservation by gray hashed bars ($0.35\leq C(i)<0.45$) and positions with less conservation by white bars ($C(i)<0.35$). The conservation values of the positions with more than one gap in the alignment are calculated as zero. The insert below shows the histogram of conservation in terms of the number of positions. Bars annotated with the stars are the conserved residues which may interact with the substrate.

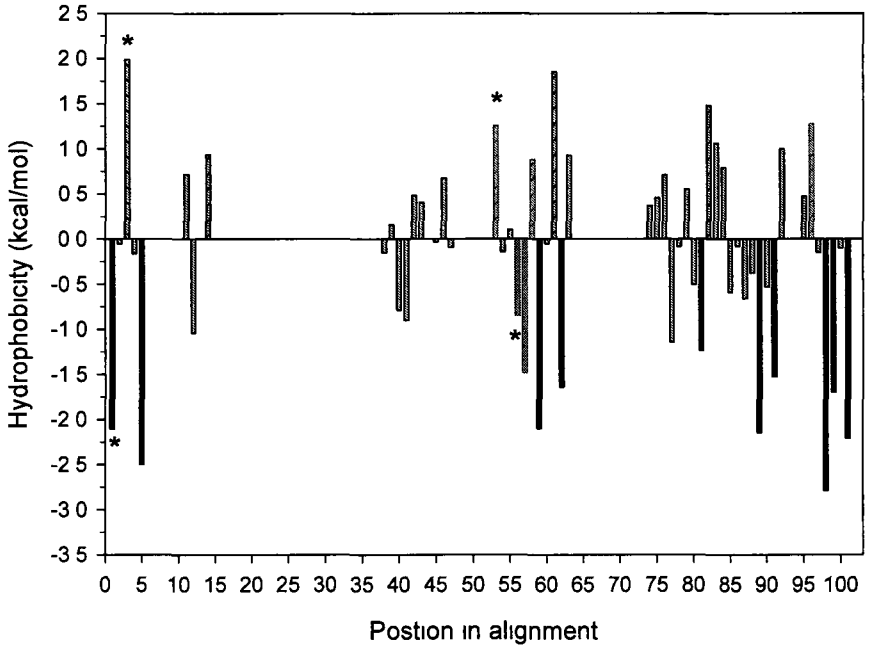


Figure 16. The average hydropathy profile analysis in the CID superfamily

Highly conserved hydrophobic positions are represented by black bars and moderately conserved positions by dark gray bars. Highly conserved hydrophilic positions are represented by black hashed bars and moderately conserved positions by gray hashed bars.

Table 3. Conserved residues on the CIDs in chitinases denoted by PDB codes of 1LG1, 1D2K and 1ITX and their proposed roles

Code	Interaction with the substrate				Formation of the hydrophobic core				
	YxR *motif*		[E/D]xx[V/I] *motif*						
1LG1	Tyr267	Arg269	Glu297	Met300	Tyr303	Val306	Ala312	Val332	Phe334
1D2K	Tyr293	Arg295	Glu316	Val319	Tyr322	Met325	Ala330	Ile352	Tyr354
1ITX	Tyr338	Arg340	Glu366	Ser369	Phe372	Leu375	Tyr385	Ile407	Tyr409
1NWT	Phe261	Arg263	Glu290	Thr293					
1FFR	Tyr444	Arg446	Glu473	Ile476					

1LG1 (1HKM): human chitotriosidase; 1D2K: *C. immitis* chitinase; 1ITX: *B. circulans* chitinase A1.

In the YxR motif, tyrosine and arginine form a pi-cation interaction, which is conserved in all five kingdoms except Plantae. These interactions are also conserved in the other family 18 chitinases which were not included in the alignment. In many family 18 chitinases, a conserved catalytic aspartic acid on the TIM barrel (e.g. Asp213 in human chitotriosidase, Figure 17A), forms an electrostatic interaction with the arginine and hydrogen bonds with both arginine and tyrosine in the motif. The pi-cation interaction, salt bridge and hydrogen bonding are likely to be important to the structural integrity of the active site. *Vibrio harveyi* chitinase A (PDB code: 3B9A) was proposed to catalyze the substrate hydrolysis following the "slide and bend mechanism" as previously described for a long-chain substrate (Songsiriritthigul et al., 2008). First, the sugar chain slides forward towards the reducing end distorting the chain especially in −1 NAG, causing it to bend and take up a transient strained boat conformation (van Aalten et al., 2001). Then the twist of the scissile bond, together with the bending of −1 NAG, makes the glycosidic oxygen accessible to the catalytic residue Glu315 for cleavage

(Songsiriritthigul et al., 2008). This mechanism may also apply to the other family 18 chitinases. In the protein structure 3B9A, Tyr461 and Arg463 in the conserved YxR motif interact with −1 NAG. They also form hydrogen bonds with the conserved catalytic residue Asp392 on the TIM barrel, which interacts with three subsites of $(NAG)_6$ (Songsiriritthigul et al., 2008).

In another exochitinase *S. marcescens* chiA, after the substrate glycosidic bond is protonated, Asp313 which interacts with Asp311 moves to another position where it interacts with the proton donor residue Glu315, forcing the acetamido group of -1 sugar to rotate. Subsequently, the water molecule that forms hydrogen bonds with Tyr390 and the NH of the acetamido group is displaced to a position which allows hydrolysis to complete (Papanikolau et al., 2001). Since the conserved YxR motif on the CID interacts with -1 NAG in *S. marcescens* chiA (Figure 18), it may help cause distortion of substrate, thus facilitating the cleavage of the glycosidic bonds along the long-chain sugar. Moreover, the YxR motif in chiA forms hydrogen bonds and provides a hydrophilic environment for the catalytic residue Asp391 (Figure 18), which is in a nearly symmetrical position with another catalytic residue Glu315 with respect to the plane of the sugar ring (Perrakis et al., 1994). Interestingly, Asp311, Asp313 and Glu315 in chiA's and Asp140, Asp142 and Glu144 in chiB (not discussed here) both belong to the conserved TIM barrel DxDxE motif, indicating that their catalytic mechanisms are very similar.

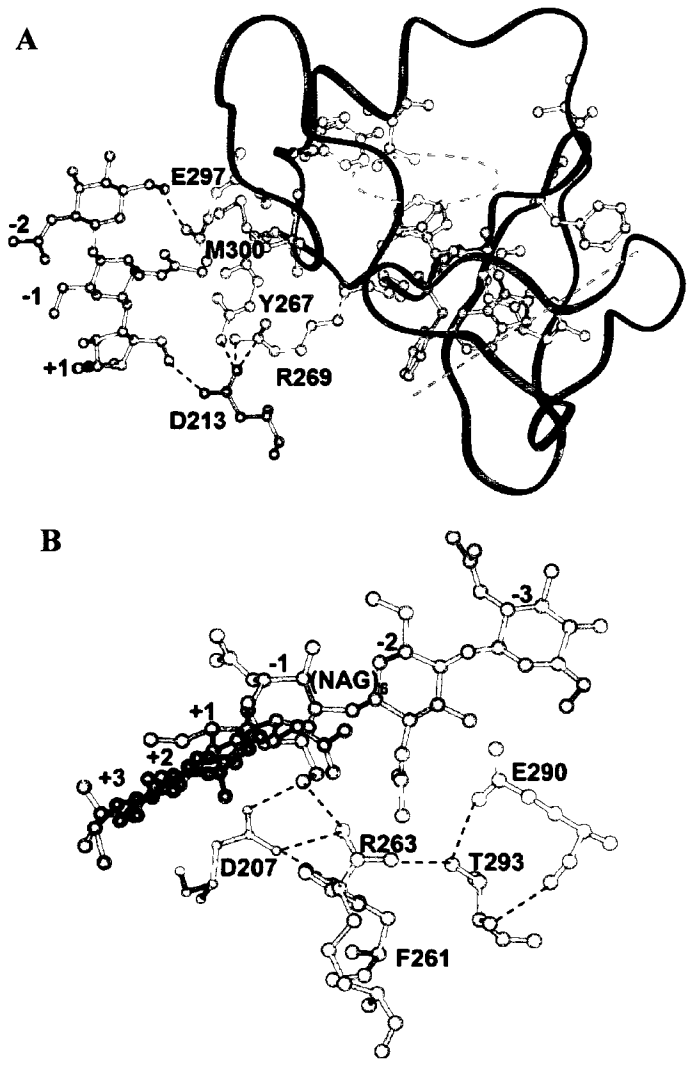


Figure 17. Structures of select family 18 chitinases with their substrates

The TIM barrel residues are coloured in green, the four conserved CID residues are in yellow and gray, and substrates are in red. Hydrogen bonds are indicated as dashed lines. (A) The conserved residues on the CID of human chitotriosidase (1HKM) either interact with the substrate, or presumably form a hydrophobic core (Table 3). The α-carbon backbone of the CID is depicted as a blue ribbon. Glu297 on the CID forms a hydrogen bond directly with the substrate while Tyr267 and Arg269 on the CID have hydrogen bonding interactions indirectly through Asp213 on the TIM domain with +1 subsite of the substrate (Table 3). Tyr267 and Met300 on the CID form hydrophobic interactions

Figure 17 Continued.

with the substrate. Some conserved hydrophobic residues in gray appear to form a hydrophobic core indicated by a dashed pink circle. Other conserved hydrophobic residues face the straight plane indicated by a dashed pink line. They are mostly aromatic and their role is undetermined.

(B) Subsites from +3 to −3 in the structure of HCgp-39 (1NWT) are lined up on the main chitin fragment. On the CID of 1NWT, Arg263 forms a hydrogen bond directly with −1 NAG and indirectly via Asp207 on the TIM domain. Phe261 from the CID forms a hydrophobic interaction with the oligosaccharide (Table 3).

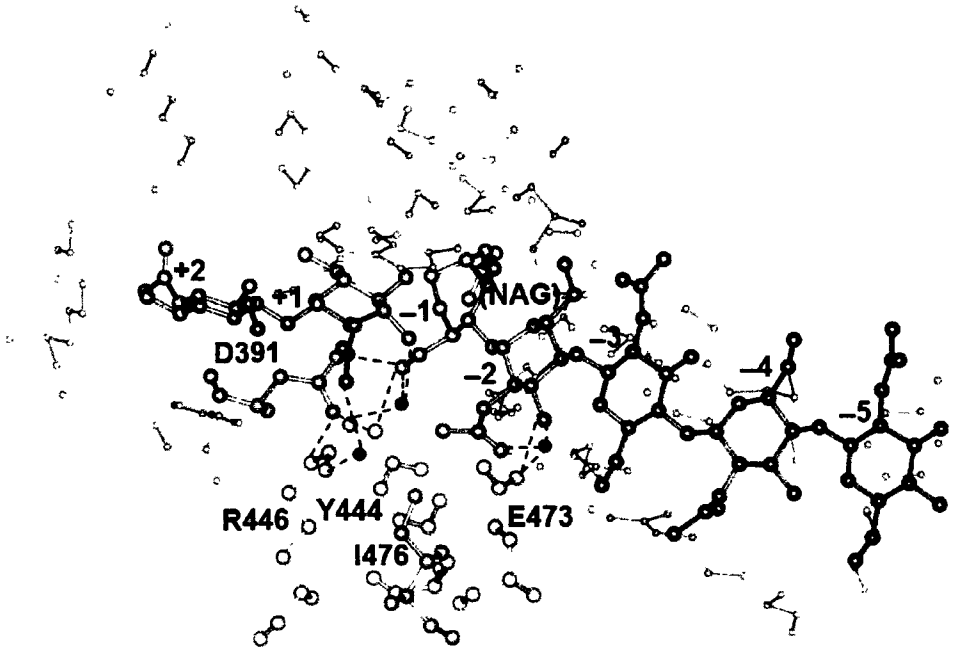


Figure 18. Residues on *S. marcescens* chiA (PDB code: 1FFR) interact with 7-mer of NAG substrate

Residues in yellow on the CID of 1FFR form hydrogen bonds with the substrate (shown as dashed lines), although some interactions are mediated by Asp391 from the TIM barrel and water molecules coloured in blue. Ile476 from the CID forms a hydrophobic interaction with the substrate (Table 3). Additional TIM barrel residues involving in hydrogen bonding and hydrophobic interactions are shown in green and brown, respectively. Structures are visualized and analyzed in Insight II. Structural studies analyzing the interactions between the protein and substrate have been previously conducted by other researchers (Papanikolau et al., 2001; Fusetti et al., 2003; Horn et al., 2006).

In the substrate-binding site in CID of human chitinase (1HKM), Tyr267 and Arg269 both form hydrogen bonding indirectly by Asp213 with +1 site, and Glu297 directly with −2 site; and Met300 forms a hydrophobic interaction with substrate (Figure 17A) (Rao et al., 2003). These amino acids, together with neighboring residues from the TIM domain, may constitute part of the substrate-binding site of the chitinase. Some of the clustered hydrophobic residues (Tyr303, Val306, Ala312, Val332 and Phe334) form a hydrophobic core indicated by the dashed pink circle (Figure 17A). The roles of the other aromatic residues (Phe271, Tyr324, Phe326 and Trp331) are not exactly known. Interestingly, they face a straight plane indicated by the dashed pink line (Figure 17A).

The [E/D]xx[V/I] motif also appears to form contacts with substrate (Table 3). In human cartilage glycoprotein-39 (HCgp-39) (PDB code: 1NWT), six sugar-binding subsites in the carbohydrate-binding groove across the C-terminal ends of the β-strands of the barrel were identified from −3 to +3 from the non-reducing end (Figure 17B). The CID also plays a role in sugar-binding because a complex hydrogen bonding network involving conserved residues Arg263, Glu290 and Thr293 on the CID interacts with −1 NAG and Phe261 forms a hydrophobic interaction (Figure 17B) (Fusetti et al., 2003).

The other highly conserved neutral positions contain mostly alanine, glycine, or proline; the latter two frequently occur in the structure of β-turns (Creighton, 1993) and may be conserved for structural reasons. CID has a large percentage of aromatic residues (e.g. 21% in 1ITX). With the exception of some residues which interact with sugar, many of them exist in the hydrophobic core, which may be important for folding and stability. Aromatic residues have been found to play an important role in stabilizing proteins and

peptides (Subramaniam et al., 2001; Palermo et al., 2008). Therefore, the combination of the CID with TIM barrel may increase the thermal stability of the whole enzyme.

Comparison of GH 18 Proteins with the CID and Those without the CID

Both the NAGase from *Elizabethkingia meningoseptica* (PDB code: 1EOM) and the NAGase from *Streptomyces plicatus* (PDB code: 1EDT) are composed of one TIM domain. They break down the glycosidic bond of $(NAG)_2$ to NAG, therefore, they do not have complete chitinolytic activities. In the crystal structure of 1EOM in complex with biantennary octa-saccharide, only the reducing end NAG and two mannoses of the tri-mannose core are in direct contact with the protein (Waddling et al., 2000), while the other sugars extend away from the protein (data not shown). 1EDT hydrolyzes the central β1→4 bond of the diacetylchitobiose core, NAG-(β1-4)-NAG, of asparagine linked oligosaccharides. Unlike the chitinases, the enzyme acts on branched oligosaccharides and has specificities for distinct forms of asparagine-linked oligosaccharides (Rao et al., 1995; Rao et al., 1999).

While only four out of eight units of the substrate interact directly with some residues on 1EOM (Table 4) (Waddling et al., 2000), proteins with the TIM and CID domains have a broad network of contacts including hydrophobic interactions and hydrogen bonding with the substrate. This can be seen, for example, in the analysis of the structure of *S. marcescens* chiA (Figure 18, Table 4B) (Papanikolau et al., 2001).

Table 4. Comparison of substrate-binding residues between chitinases with the CID and without the CID

PDB Code	Sugar ring number	Protein residue			
		Residues on the TIM barrel		Conserved residues on the CID	
		Hydrogen bonding	Hydrophobic interaction	Hydrogen bonding	Hydrophobic interaction
1EOM (A)	632	D126, E128, Q211, Y213, Y272	F39		
	633	D18, R20, E245, Y272,			
	634	K42, N85, D87, H129	F39		
	641	R20, E245			
	635, 636, 642, 643				
1FFR (B)	+2	K369, D391	W275, F396, Y418		
	+1	E315, D391	W275, F316, M388	**R446**	
	-1	D313, E315, D391	Y163, W275, A362, M388, W539	**Y444, R446**	
	-2	E540	W275, W539	**E473**	**I476**
	-3	T276	W167	**E473**	
	-4	R172			
	-5		Y170		

(A) Interactions between select residues on *F. meningosepticum* NAGase (1EOM) and bound polysaccharide.

(B) Interactions between some residues on *S. marcescens* chiA (1FFR) and bound substrate $(NAG)_7$. The data are adapted from Waddling et al. (2000) and Papanikolau et al. (2001). Conserved residues on the CID from our conservation analysis are in bold.

Sun et al. specified that the CID of mouse lectin Ym1 (PDB code: 1E9L) was not involved in the saccharide-binding (2001). Furthermore, they were unable to assign any definitive function for this domain. However, the results of our study indicate that at least four conserved residues on the CID of many chitinases were found to have either hydrogen bonding or hydrophobic interaction with the substrate of more than three units of NAG. While 1E9L was not included in the original structural alignment, a close homologue 1NWT was studied and suggests that the authors may have seen saccharide-binding by the CID if a longer substrate was used.

In CAZy database (Cantarel et al., 2009), *S. marcescens* chiA and chiB, *B. circulans* chitinase A1 and *Aspergillus fumigatus* chitinase B (PDB code: 1W9P) are "bacterial-type" exochitinases with a deep or even a tunnel-shaped substrate-binding cleft, formed by the TIM barrel and CID (van Aalten et al., 2001; Watanabe et al., 2003; Sikorski et al., 2006; Horn et al., 2006). *S. marcescens* chiC (Horn et al., 2006; Sikorski et al., 2006) and "plant-type" chitinases such as hevamine from Para rubber tree (*Hevea brasiliensis*) (PDB code: 1HVQ) (Terwisscha van Scheltinga et al., 1996), ScCTS1 from *Saccharomyces cerevisiae* (PDB code: 2UY2) (Hurtado-Guerrero and van Aalten, 2007), PPL2 from *Parkia platycephala* seeds (PDB code: 2GSJ) (Cavada et al., 2006) and a hyperthermophilic chitinase from *Pyrococcus furiosus* (PDB code: 2DSK) (Nakamura et al., 2006) are endochitinases with a shallow substrate-binding cleft since they lack the CID. Therefore, overall it appears that the CID may enhance the exo-type activity by forming a deep substrate-binding cleft on the top of the TIM barrel (Horn et al., 2006; Sikorski et al., 2006; Zees et al., 2009).

Structures of TIM domain alone, TIM domain + CID, and TIM domain + CID + N- (or C-) terminal domain align very well with their respective counterparts (data not shown). Interactions between residues and substrates are shown in Table 4 for 1EOM (TIM domain alone) and 1FFR (TIM domain + CID). It appears as if more sugar residues interact with amino acid residues when the CID is included in the TIM domain. Therefore, the CID may facilitate stronger association with the substrate, particularly with increasing substrate length. By removing the CID from *S. marcescens* chiA, a processive exochitinase (Uchiyama et al., 2001; Sikorski et al., 2006), the truncated active enzyme appeared to have a shallower tunnel in the catalytic domain than that of the intact enzyme (Zees et al., 2009) and it resembled the catalytic domain of *S. marcescens* chiC, which acts as a non-processive endochitinase (Sikorski et al., 2006). Therefore, the CID of chiA enhances not only the exo-*N*,*N*'-diacetyl-glucosaminidase activity, but also the processivity during the degradation of the polysaccharide chains (Zees et al., 2009).

Phylogenetic Analysis of the CID and Evolutionary Scheme of Family 18 Chitinases (Subfamilies A and B)

The ubiquitous TIM barrel fold is adopted by seven enzyme superfamilies, one of which is the TIM barrel GH (Todd et al., 2001). The evolutionary relationships between different enzymes with TIM barrel are well studied (Todd et al., 2001; Nagano et al., 2001; Nagano et al., 2002). Gene duplication, gene fusion and incremental mutations are three mechanisms by which new functions are created in proteins (Todd et al., 1999; Todd et al., 2001). Molecular phylogenetic analyses of mammalian GH 18 chitinase and chitinase-like members suggest that active chitinases result from an early gene

duplication event, which is followed by mutations, leading to chitinase-like proteins, such as chito-lectins (Bussink et al., 2007). Comprehensive genomic studies of animal GH 18 proteins showed a similar result (Funkhouser and Aronson, 2007). Another phylogenetic analysis of catalytic domain sequences from various organisms showed that sequences of animal, fungi and bacteria belong to different lineage; however, chitinase genes from lepidopteran insects and baculoviruses originated from bacteria and were maintained through evolution since they transferred laterally (Daimona et al., 2005).

Since the CID sequences are present in all of sixty archaeal, bacterial and eukaryal genomes in this study, it is possible that the CIDs were present in the Last Universal Common Ancestor (LUCA) (Ranea et al., 2006). However, no evolutionary study has been conducted on the CID by itself. To establish the phylogenetic relationships between the CIDs from different organisms, a preliminary phylogenetic tree was constructed based on the sixty sequences from five kingdoms (Archaea, Bacteria, Fungi, Plantae and Animalia) (Figure A6). Overall, the CID sequences grouped into five major clusters, each representing one kingdom as to be expected. In the cluster of Animalia, members from early eukaryotes and early Animalia branch out earlier than those from vertebrates and mammals.

DISCUSSION

Bioinformatics and Network Analysis of Fadd-DD

Our computational analysis shows that sixteen residues with intermediate to high levels of conservation and dominant hydrophobic character form a long-range interaction network in Fadd-DD (Figure 10-11 and Figure 13). These residues are postulated to be important for folding, topology and stability. Furthermore, the residues constituting the

network may act by organizing the hydrophobic collapse during the folding process. In the network, Ala100, Leu115, Trp148 and Leu161 have more interactions than the other twelve residues suggesting that these four interconnected residues may be particularly important (Figure 13). Given the scarcity of tryptophan in proteins in general (e.g., in most of CARD proteins and PYD proteins), it is unusual that some DD proteins have two tryptophans (some DED proteins have just one). Since the two conserved tryptophans were taken as probes in the stopped-flow fluorescence studies, it will be interesting to see if the more conserved Trp148 becomes structured on an earlier timescale, and thus is key to the first phase.

Our structural studies also identified another potentially important residue, Phe101. It is hydrophobic and has an orientation similar to Ala100 but is only 60% buried according to the program Naccess (www.bioinf.manchester.ac.uk/naccess). It does however form long-range interactions with Trp148 and Leu161. While Phe101 does not occupy a conserved position in our analysis it is adjacent to Ala100 and may, in Fadd-DD, play an important structural role in concert with this residue. This suggests that while conserved residues and interactions may be used by all members within a superfamily for folding, there may also be some residues specific for individual proteins. Interestingly, there are eleven conserved interactions within the four structures studied and these may be important discriminators of the all α-helical Greek-key topology.

Conserved Residues for Structure and Stability versus Function in Fadd-DD

It is well established that amino acids are conserved in proteins because they are important for stability and function (Schueler-Furman and Baker, 2003). With respect to stability, mutational experiments indicate that hydrophobic core residues make

substantial contributions (Shortle, 1992). Therefore, mutation of any of these sixteen hydrophobic residues in Fadd-DD to any hydrophilic amino acid may affect the structure and stability of Fadd-DD. This in turn can disrupt the functional association with other DDs in the formation of the apoptotic complex. Berglund et al. (2000) proposed that mutation of Val121Asn in human Fadd-DD would destabilize the protein fold and thus indirectly affect function. Two mutations towards conserved residues in mouse Fadd-DD (Val121Asn and Leu119Asn) were also implicated in a functional role due to reduced or abolished binding affinity with Fas-DD (Jeong et al., 1999). We speculate that the observed effect of Val121 on function may be explained because it forms the long-range interaction in the hydrophobic core and mutation probably causes the disruption of the structure and stability in mouse and human Fadd-DD. This is supported by the fact that in mouse Fadd-DD Val121Asn significantly reduced the secondary structure measured by far-UV CD (Bang et al., 2000). Additionally, previous mutation of Ile225Asn in mouse Fas-DD which aligns with Val121 in human Fadd-DD, disrupted its binding and caused autoimmune lymphoproliferative (*lpr*) syndromes (Watanabe et al., 1992; Imtiyaz et al., 2005). In mouse Fas-DD, the mutation may destabilize the structure and thus indirectly affect function.

The conservation and hydropathy analysis also identified six hydrophilic residues with intermediate to high levels of conservation (Figures 10-11, 12B). Mutational experiments indicate that some conserved residues in protein–protein interfaces contribute significantly to binding between proteins (DeLano et al., 2002). Fourteen sites on mouse or human Fadd-DD were experimentally mutated by other researchers and found to be important for the interaction or binding with Fas-DD, which triggers Fas-

mediated apoptosis (Table 5) (Jeong et al., 1999; Bang et al., 2000; Imtiyaz et al., 2005; Hill et al., 2004; Sandu et al., 2005). Of these fourteen experimentally determined positions, four are shown on our hydropathy plot to be conserved and persistently hydrophilic (Figure 12). Our analysis also indicates that there are two additional conserved hydrophilic residues (Glu143 and Asp111) which have not been experimentally studied (Table 5). Mutagenesis of these two residues may reveal their importance in function and further highlight the value of these conservation studies.

Table 5. Sites on Fadd-DD which are important for the interaction with Fas-DD to cause Fas-induced apoptosis

Residue	D106	K110	R113	R114	R117	Q118	D123
Mutation	A	A	E, A	A	E, A	A	R, A
Residue	K125	D127	R142	R146	R166	L172	D175
Mutation	A	A	E, A	A	E	E	R

Underlined residues may be explained by the conservation and hydropathy analysis.

It could also be argued that the conserved hydrophilic residues on the surface may play a role in protein folding and stability as in the case of Lys5 and Lys7 of the cold shock protein CspB (Garcia-Mira and Schmid, 2006). Additional support for considering the role of charged residues for folding comes from studies of de novo libraries which successfully generate stable protein folds based on polar–nonpolar patterns (Go et al.,

2008). With regard to stability, hydrophilic residues involved in ionic interactions (salt bridges) can often be crucial determinants as typified by studies of proteins from thermophilic organisms (Sadeghi et al., 2005). Mutagenesis of these residues should address this possible role. However, in support of their proposed role in function, the conserved Asp127 in the mouse Fadd-DD was changed to alanine by Bang et al. (2000) and resulted in loss of interaction with Fas-DD. The change also did not perturb the structure judging from the far-UV CD signal and the $^{1}H–^{15}N$ HSQC spectra (Bang et al., 2000).

Bioinformatics Analysis of the CIDs

In the study conducted by Nagano et al., family 18 GH were divided into two functional groups; F4 includes chitinases and F5 includes both hevamine and NAGase (Nagano et al., 2001). A proposed evolution of the structure and function of family 18 chitinases and chitinase-like proteins in the subfamilies A and B can be potentially described as follows. Due to divergent evolution, a TIM domain line may initially have evolved as hevamine, xylanase inhibitor protein, or seed storage protein (e.g. Concanavalin B) in some higher plants, as well as NAGase in some bacteria. While hevamine has lysozyme/endochitinase function (Terwisscha van Scheltinga et al., 1996; Terwisscha van Scheltinga et al., 1994), xylanase inhibitor protein (Payan et al., 2003) and seed storage protein (Hennig et al., 1995) do not have known chitinolytic activity. One possible evolutionary scheme suggests that a TIM barrel evolved to a more potent family 18 chitinase in two routes: 1) with the incorporation of the CID to form a subfamily A chitinase and 2) with the other domains (e.g. chitin-binding domain) to form a subfamily B chitinase. In the first route, this double-domain chitinase evolved in

archaea, bacteria, fungi, plants and animals, as well as the triple-domain chitinase with the fusion of N- or C- terminal domain in *S. marcescens*. Subsequently, the double-domain chitinase gene was mutated to have novel functions in animals (Bussink et al., 2007).

CHAPTER IV

IMPORTANCE OF CONSERVED HYDROPHOBIC RESIDUES IN STRUCTURE AND STABILITY OF FADD-DD

INTRODUCTION

The mutation of conserved amino acids on some proteins led to decreased stability and structural change (Jäger et al., 2009; Xiao et al., 2009). In the section on bioinformatics, we propose that nine significantly conserved and seven moderately conserved hydrophobic residues form a network that is important for the structure and stability of Fadd-DD. In order to test this hypothesis, site-directed mutagenesis was employed to produce the protein variants. Equilibrium fluorescence and CD experiments were performed and the structure and stability were compared between the WT and variants.

To probe the role of the conserved residues hypothesized to be important in structure and native state stability, six protein variants were analyzed: conserved Trp112 to phenylalanine, conserved Trp148 to phenylalanine, non-conserved His160 to glycine, a double variant Trp112Phe/His160Gly, conserved Leu115 to alanine and conserved Val121 to alanine. Trp148 interacts with His160 in the native state via van der Waals forces and through a tertiary hydrogen bond. The significantly reduced side chain of His160 to glycine and loss of the hydrogen bond had no negative effects on native state stability. However, the small reduction in side chain of Trp148 to phenylalanine, Leu115

and Val121 to alanine had significant destabilizing effects on the structure of Fadd-DD. The change of Trp112 to phenylalanine was also destabilizing though to a lesser degree.

RESULTS

Cloning, Expression and Purification of WT Fadd-DD

Pure PCR products for WT Fadd-DD were obtained (Figure 19). The length is about 590 bp. Bands on the 3rd and 4th lanes were combined to get more concentrated DNA. The purified cDNA was eluted with ddH_2O (pH 8.2). The DNA concentration was measured at 260 nm in Eppendorf Biophotometer (Eppendorf, NY). The concentrations of the samples "circle" and "PCR product" were 54 and 156 μg/ml, respectively. The ratio of O.D. values at 260 nm/280 nm was also measured to check the purity of the DNA product. Their ratios of approximately 1.8 proved their purity.

Sequencing results show that the DNA sample contains Fadd-DD cDNA (Figure 20). Its site is between *Nde* I and *Xho* I and it is 303 bases long which encode 100 amino acids (TGA as stop codon). BLAST search in NCBI shows that it has 99% identity with the human Fadd-DD gene. Two nucleotides were different from the original gene sequences noted in Figure 20. These "mismatches" encode the same amino acids. Gel electrophoresis shows the insert in the plasmid, which was cut by both *Nde* I and *Xho* I (Figure 21). The results show that cDNA was successfully inserted into the plasmid (Figure 20-21).

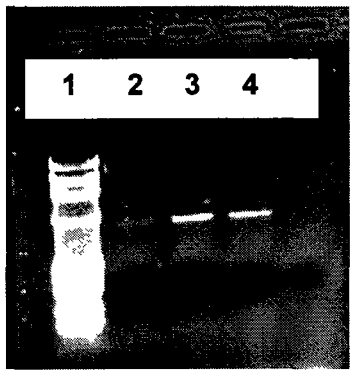


Figure 19. Agarose gel electrophoresis showing pure PCR product

Bright bands on the 3rd and 4th lanes were PCR product of cDNA in "small circle". Molecular weight marker (MWM) was in the 1st lane.

TATTTCCCTCTAGAAAAATTTTGTTTACTTTAAAAGGAGATATACCATGGGCA

GCAGCCATCATCATCATCATCACGGCAGCGGCCTGGTGCCGCGCGGCAGCCA

His_6·Tag thrombin digestion site

TATGGGGGAAGAAGACCTG**TGC**GCAGCATTTAACGTC**ATC**TGTGATAATGTG

Nde I start Cys Ile

GGGAAAGATTGGAGAAGGCTGGCTCGTCAGCTCAAAGTCTCAGACACCAAG

ATCGACAGCATCGAGGACAGATACCCCCGCAACCTGACAGAGCGTGTGCGG

GAGTCACTGAGAATCTGGAAGAACACAGAGAAGGAGAACGCAACAGTGGCC

CACCTGGTGGGGGCTCTCAGGTCCTGCCAGATGAACCTGGTGGCTGACCTGG

TACAAGAGGTTCAGCAGGCCCGTGACCTCCAGAACAGGAGTGGGGCCTGA**C**

Stop

TCGAGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTG

Xho I *Bam*H I

CCACCGCTGAGCAATAACTAGCAATAA

Figure 20. Sequencing result of the cDNA sample for WT Fadd-DD

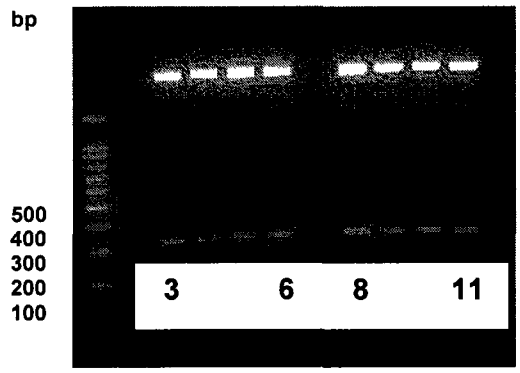


Figure 21. Gel electrophoresis of inserts after cutting the recombinant plasmid with *Nde* I and *Xho* I or *BamH* I

Faint bands in each lane show insert at 300 kb (lanes from 3 to 6 were cut by *Nde* I and *Xho* I, while lanes from 8 to 11 were cut by *Nde* I and *BamH* I). MWM is in the 1st lane.

Fadd-DD (100 aa) attached with His-tag and thrombin digestion site (21 aa) has a molecular weight of roughly 13.7 kDa. The protein was successfully expressed by *E. coli* BL21(DE3) (Figure 22), since 13 kD bands were observed for the bacteria induced by IPTG, but not for the negative control. Moreover, the bacteria induced by IPTG produced more Fadd-DD than those without IPTG. However, this is a leaky system, because the *E. coli* RNA polymerase makes a small amount of T7 RNA polymerase without induction. High-level expression of cloned genes in *E. coli* exploits a phage T7 promoter whose activity depends on a regulatable transcription unit supplying the specific T7 RNA polymerase (Mertens et al., 1995). However, T7 RNA polymerase in some pET vectors may bind upstream of the *lac* UV5 promoter and read through it (Gerstein, 1992). Therefore, the target protein is expressed before induction with IPTG causing leaky

expression (Spehr et al., 2000). This phenomenon may be due to a small amount of lactose in the medium (Trudy et al., 1998).

After purification through the first Ni-NTA column, the fractions containing His-tag Fadd-DD were pooled. After the protein was digested with thrombin and loaded onto the SDS-PAGE, the bands on the gels showed one smaller band at 11 kDa and one bigger band at 13 kDa, indicating that part of the His-tag Fadd-DD was cut (Figure 23). After purification through the second Ni-NTA column, the fractions containing Fadd-DD were collected. Protein gel electrophoresis showed a pure protein sample (Figure 24). After the gel filtration column, pure protein samples were obtained.

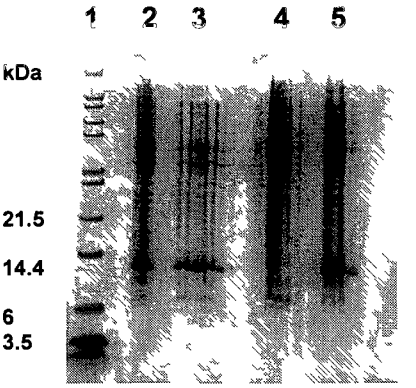


Figure 22. Protein analyses on SDS-PAGE

The 1st lane is the MWM. The 2nd lane is bacteria induced with IPTG; the 3rd lane is bacteria without IPTG; the 4th lane is bacteria with pUC18 control; the 5th lane is bacteria in 6 L culture.

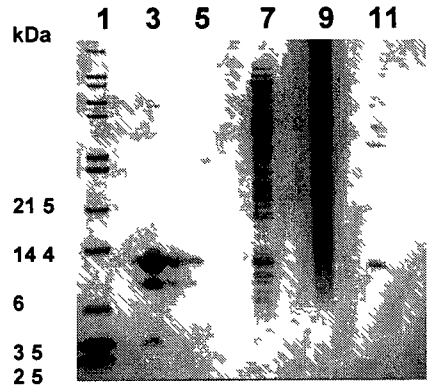


Figure 23. His-tag Fadd-DD in the SDS-PAGE

Protein sample on SDS-PAGE after thrombin digestion. MWM is in the 1st lane. The 3rd lane is the mixture of His-tag Fadd-DD and Fadd-DD.

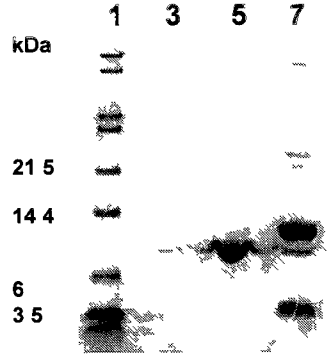


Figure 24. Fadd-DD in the fractions and SDS-PAGE

Protein samples on SDS-PAGE after the second Ni-NTA column. The 5th lane is Fadd-DD. The 7th lane indicates His-tag Fadd-DD and His-tag.

Mass spectrometry was performed on the purified Fadd-DD in COSMIC. The observed average molecular mass of $(M+H)^+$: 11853.217401 u was obtained from the in-

house deconvolution program attached to the instrument (Figure 25). The predicted isotopically averaged molecular weight is 11853.3213 according to http://www.scripps.edu/cgi-bin/cdputnam/protcalc3. Therefore, the difference of -9.4 ppm indicates that the protein is the correct one. DNA sequencing, SDS-PAGE gel electrophoresis and mass spectroscopy data confirmed that the cDNA was cloned and the expressed recombinant protein is WT Fadd-DD. The recombinant Fadd-DD is composed of 104 amino acids, which includes 4 residues left from the thrombin digestion sites and 100 residues of the original Fadd-DD. Following purification approximately 50 mg of protein per liter of culture was obtained.

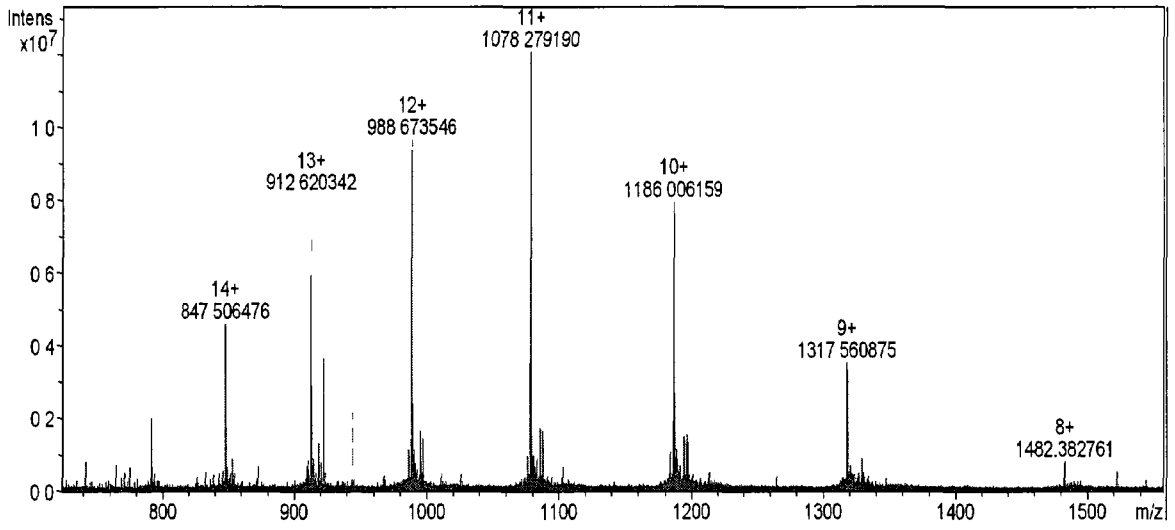


Figure 25. Mass spectrum analysis of Fadd-DD

The peaks indicate the different charge states of the protein.

The Trp112Phe, Trp148Phe, His160Gly, Trp112Phe/His160Gly, Leu115Ala, Val121Ala, Leu172Trp and Val158Trp mutants were synthesized with the Stratagene Quick Change Kit and confirmed by DNA sequencing.

Intrinsic Fluorescence Spectra and Equilibrium Refolding and Unfolding

The overall character of the fluorescence emission spectra was assessed by exciting the protein at 295 nm and observing emission intensities from 310 to 450 nm in native buffer and buffer with GndHCl (Figure 26A). Native WT Fadd-DD had a fluorescence emission maximum of 337 nm and the denatured protein exhibited a fluorescence emission maximum of 360 nm. There is little difference in maximum intensities between the native and denatured states considering that Fadd-DD has two buried tryptophans (Figure 26A). This led us to carefully assess the environments of the tryptophans in the solution structure. We identified that the side chain of Trp148 is within close proximity to His160 and forms a hydrogen bond with its side chain (Figure 27). Therefore, it is quite probable that the fluorescence intensity is quenched in the native state by the Trp–His interaction (Lakowicz, 1999). A quenching affect is also noted for the human γD crystallin domains (Kosinski-Collins and King; 2003; Kosinski-Collins et al., 2004). Therefore, in accordance with the analysis procedure established by Kosinski-Collins et al. (2004), all equilibrium unfolding/refolding data for Fadd-DD were analyzed using the ratio of fluorescence emission intensities at 370 nm over 330 nm.

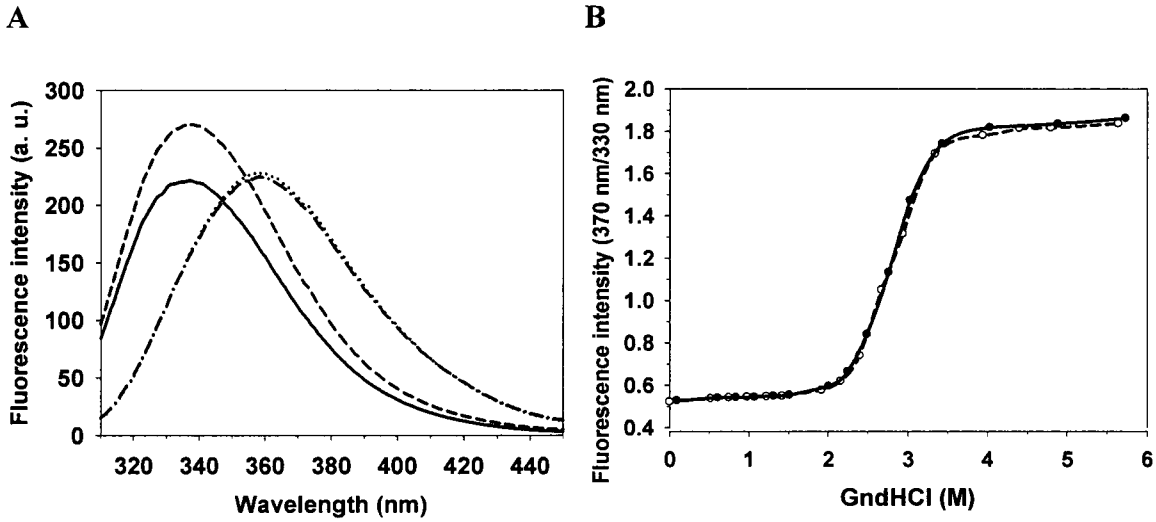


Figure 26. Fluorescence spectra of native and denatured states of Fadd-DD

(A) The fluorescence of 0.05 mg/ml of WT Fadd-DD (4 μM) in 20 mM Bis-Tris buffer (pH 6.2, 2 mM DTT) with different concentrations of GndHCl was measured at 20°C with excitation at 295 nm and emission from 310 nm to 450 nm. The solid line, dashed line, dash-dot line and dotted line were protein in 0 M, 1 M, 5 M and 6 M of GndHCl, respectively.

(B) The ratio of fluorescence emission intensities at 370 nm/330 nm was used to concurrently monitor changes in the unfolding and native maxima. The equilibrium unfolding (open circle on the dashed line) and refolding (closed circle on the solid line) curves show that the folding is cooperative, two-state and reversible.

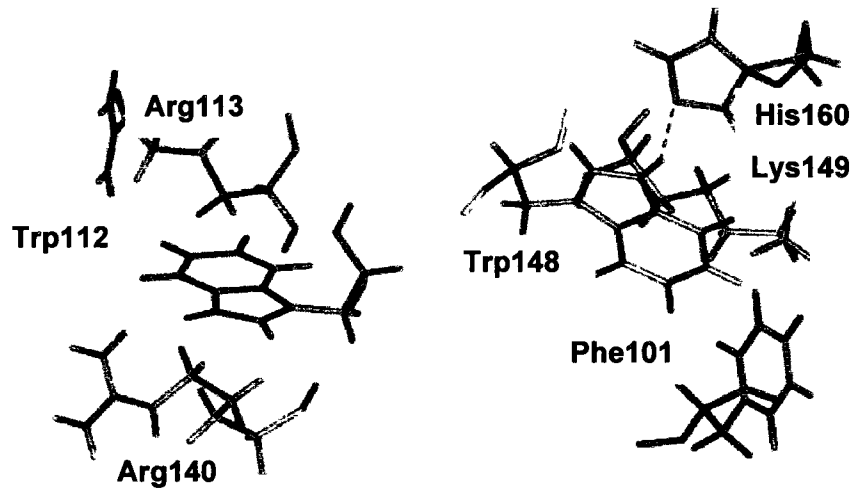


Figure 27. Side chain environments around the two buried tryptophans

Trp112 is flanked by Arg113 and Arg140 and Trp148 is surrounded by Phe101, Lys149 and His160. The calculated hydrogen bond between the side chains of His160 and Trp148 is shown as a dotted line.

The equilibrium refolding transition of WT Fadd-DD appears to be cooperative and reversible since it resembles the unfolding curve (Figure 26B). The overall character of the fluorescence emission spectra was assessed by exciting Trp112Phe at 295 nm and observing emission intensities from 310 to 450 nm in native buffer and buffer with GndHCl (Figure 28A). The conformational free energy, ΔG^{D-N}, indicates how much more stable the native conformation of the protein is than the unfolded conformation (Pace, 1986) and the cooperativity index, m, is the slope of the curve ΔG^{D-N} over the molar GndHCl concentration (Dalal and Pio, 2006). ΔG^{D-N} and m determined from fits of the data to the two-state model are 6.60 ± 0.03 kcal/mol and − 2.41 ± 0.02 kcal/mol/M, respectively (Figure 28B). The ΔG^{D-N} and m, determined for the variants are listed in Table 6 (Figure 28B).

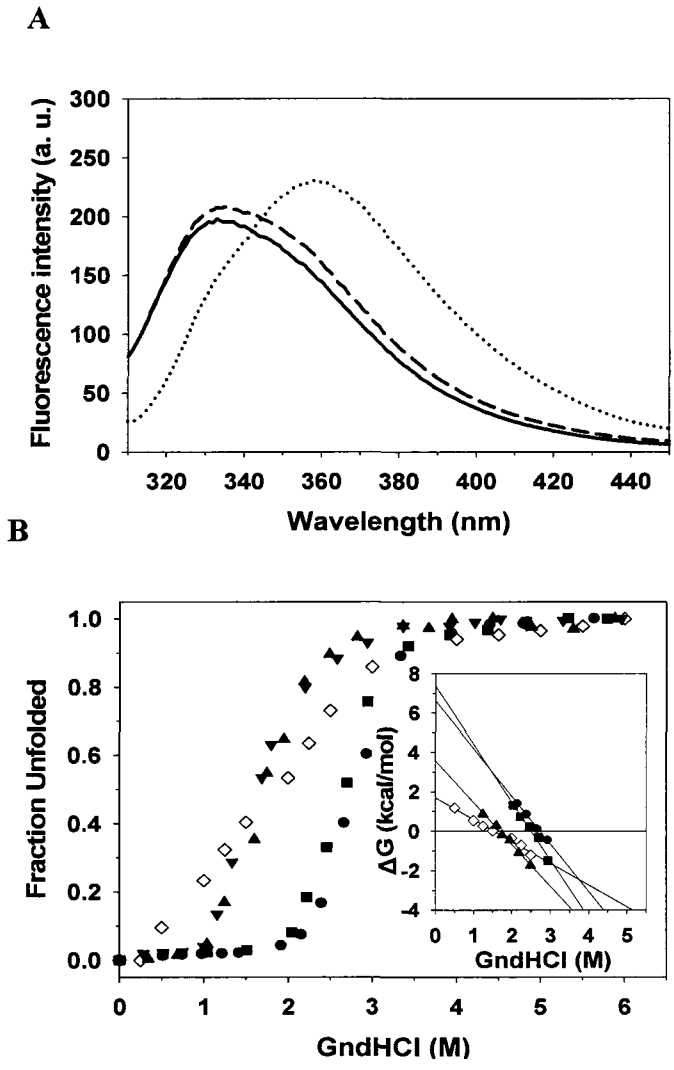


Figure 28. Fluorescence spectra of native and denatured states of Fadd-DD variants

(A) The fluorescence of 0.05 mg/ml of Trp112Phe in 20 mM Bis-Tris buffer with different concentrations of GndHCl was measured. The line types are the same as those in Figure 26A.

(B) Normalized fraction unfolded of the WT and variants as a function of GndHCl concentration at 20 °C. The fraction unfolded was measured as the ratio of fluorescence intensities 370 nm/330 nm. The proteins are annotated as follows: WT (●); Trp112Phe (▲); Trp148Phe (◊); His160Gly (▪); Trp112Phe/His160Gly (▼). The inset shows ΔG_U as a function of GndHCl concentration. The data are graphed using SigmaPlot (Ver. 10).

Table 6. The ΔG^{D-N} and m value of the WT and variants

Protein/variant	ΔG^{D-N} (kcal/mol)	m value (kcal/mol/M)
WT	6.60 ± 0.03	-2.41 ± 0.02
Trp112Phe	3.57 ± 0.06	-2.12 ± 0.04
Trp148Phe	1.70 ± 0.07	-1.11 ± 0.04
His160Gly	7.35 ± 0.15	-2.93 ± 0.07
Trp112Phe/His160Gly	3.42 ± 0.06	-2.15 ± 0.04
Leu115Ala	1.81 ± 0.15	-1.39 ± 0.10
Val121Ala	2.94 ± 0.46	-2.28 ± 0.31
Leu172Trp	6.54 ± 0.62	-2.38 ± 0.23
Val158Trp	6.53 ± 1.10	-2.60 ± 0.43

Leu172Trp and Val158Trp Variants

The fluorescence spectra and fraction unfolded of Leu172Trp are both similar to those of the WT (Figure 29). The fluorescence spectra and fraction unfolded of Val158Trp are both similar to those of the WT (Figure 30). The ΔG^{D-N} and m values of these two variants are shown in Table 6.

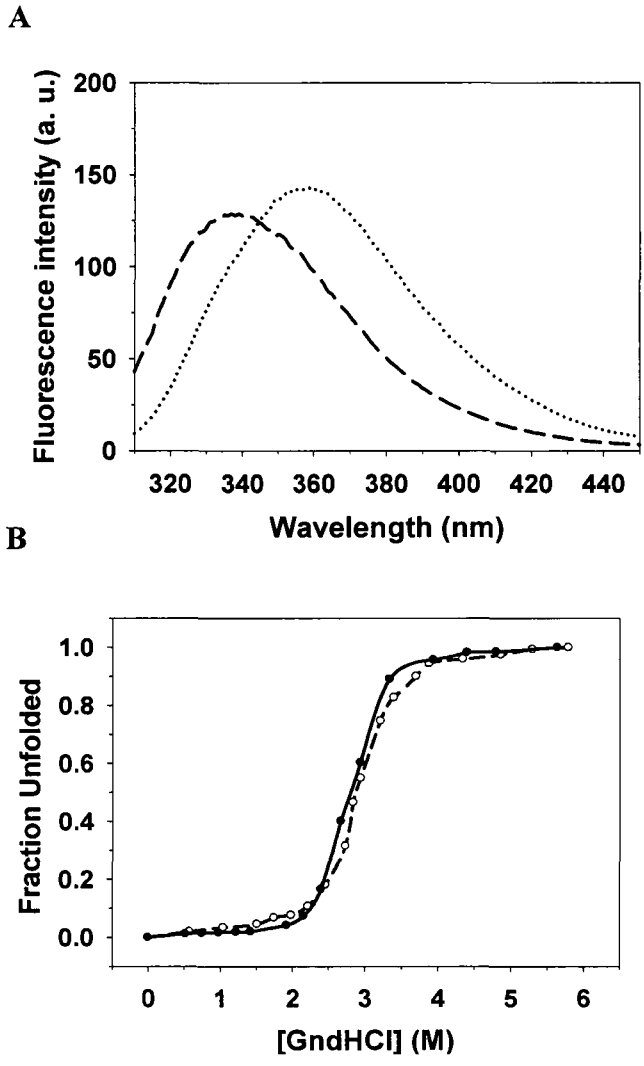


Figure 29. Fluorescence spectra of native and denatured states of Leu172Trp

(A) The fluorescence of 0.05 mg/ml of Leu172Trp in 20 mM Bis-Tris buffer (pH 6.2, 2 mM DTT) with different concentrations of GndHCl was measured at 20°C with excitation at 295 nm and emission from 310 nm to 450 nm. The dashed line and dotted line were protein in 0 M and 5 M of GndHCl, respectively.

(B) The ratio of fluorescence emission intensities at 370 nm/330 nm was used to concurrently monitor changes in the unfolding and native maxima. The open circle on the dashed line and closed circle on the solid line are equilibrium unfolding of Leu172Trp and WT.

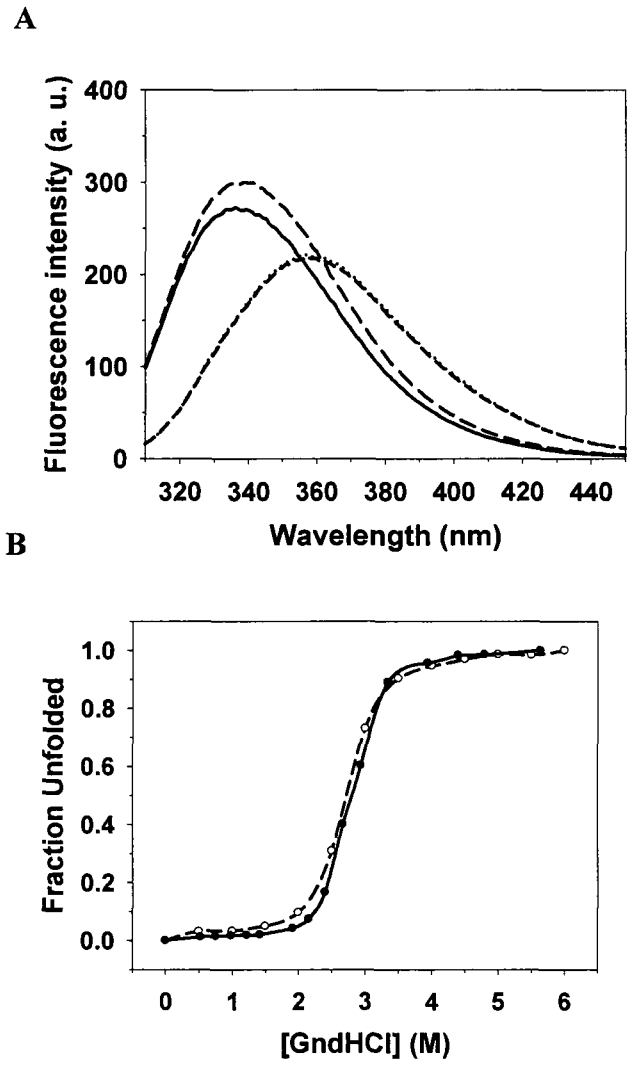


Figure 30. Fluorescence spectra of native and denatured states of Val158Trp

(A) The fluorescence of 0.05 mg/ml of Val158Trp in 20 mM Bis-Tris buffer (pH 6.2, 2 mM DTT) with different concentrations of GndHCl was measured at 20°C with excitation at 295 nm and emission from 310 nm to 450 nm. The solid line, dashed line, dash-dot line and dotted line were protein in 0 M, 1 M, 5 M and 6 M of GndHCl, respectively.

(B) The ratio of fluorescence emission intensities at 370 nm/330 nm was used to concurrently monitor changes in the unfolding and native maxima. The open circle on the dashed line and closed circle on the solid line are unfolding of Val158Trp and WT.

Leu115Ala and Val121Ala Variants

Equilibrium fluorescence was conducted on the two variants, showing faction unfolding as the function of the GndHCl concentration (Figure 31A) and free energy (Figure 31B). Leu115 and Val121 were shown in our previous study to be moderately conserved (Li et al., 2009). This is not surprising because in the study conducted by Steward et al. (2009), the mutation of Leu115 to alanine was too unstable to obtain enough proteins for study and so Leu115 was converted to methionine. It can therefore be inferred that Val121Ala is also unstable since it misfolds, similar to Leu115Ala variant. The ΔG^{D-N} and m value of these two variants are shown in Table 6.

The Effect of pHs on Protein Fluorescence

To test the effect of different pHs on the conformation of Fadd-DD, the fluorescence experiment was conducted with the protein in different buffers with pH values ranging from 3.9 to 9.8. It appears that the maximum fluorescence intensity increases as pH decreases, while the fluorescence maximum wavelength is only marginally affected (Figure 32). This phenomenon may due to the fact that the hydrophobic core is more buried in lower pH than in higher pH or because of fluorescence quenching effect since fluorescence of one of the tryptophans may be quenched in the native state.

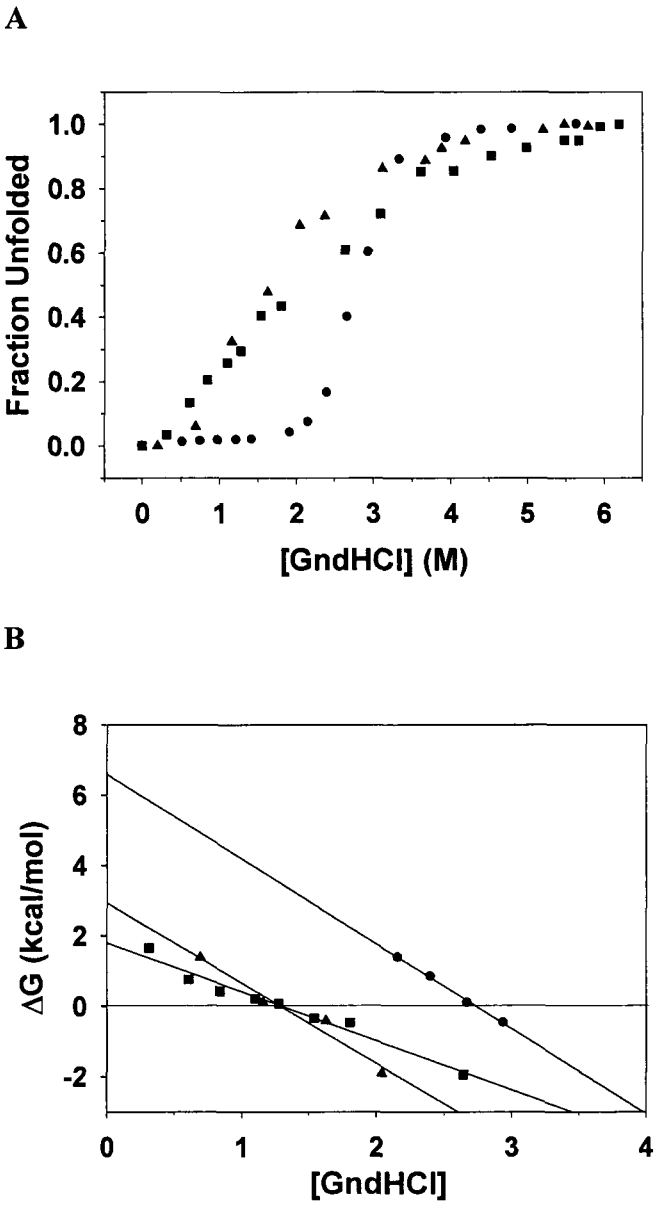


Figure 31. The fraction unfolded and free energy of WT, Leu115Ala and Val121Ala

(A) Normalized fraction unfolded of the WT and variants as a function of GndHCl concentration at 20°C. The fraction unfolded was measured as the ratio of fluorescence intensities 370 nm/330 nm. The proteins are annotated as follows: WT (●); Val121Ala (▲); Leu115Ala (▪).

(B) ΔG_U as a function of GndHCl concentration.

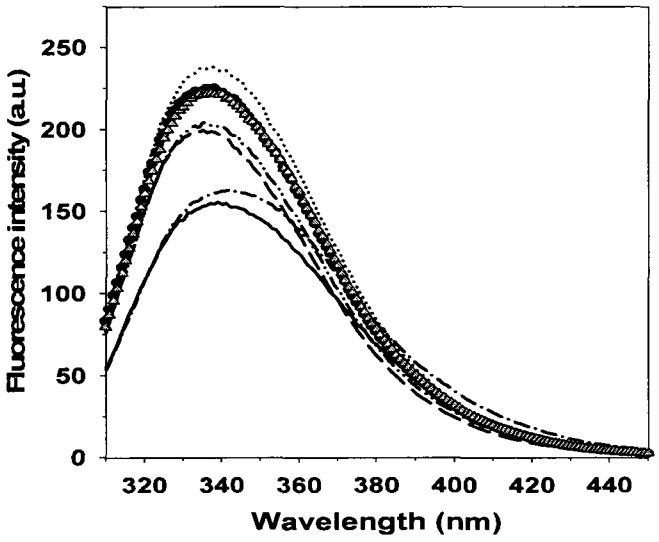


Figure 32. Fluorescence spectra of 0.05 mg/ml Fadd-DD in buffers with different pHs

The parameters of the fluorescence experiment are the same as previous. The buffers used are the same as those used in the quenched-flow experiments. The lines indicate pH as follows: 3.9: dotted black; 4.2: gray triangle; 4.8: black circle; 5.2: dashed-dotted-dotted black; 6.2: dashed black; 9.6: dashed-dotted black; 9.8: solid black.

CD Spectra and Equilibrium Unfolding

The overall character of the far-UV and near-UV CD spectra was assessed by observing ellipticities from 190 to 250 nm and from 250 to 320 nm in native buffer and buffer with GndHCl, respectively (Figure 33A–B). For far-UV CD, the native WT Fadd-DD had one maximum around 195 nm and two minima at 207 nm and 221 nm indicating Fadd-DD is a typical α-helical bundle protein. The denatured protein, as expected, lost this characteristic. For near-UV CD, the native WT Fadd-DD had two maxima around

260 nm and 295nm, and one minimum around 285 nm. Because there are only two exposed aromatic residues, Tyr133 and Phe101, it is reasonable that the near-UV CD ellipticity is low. The depth at around 285 nm is probably due to the asymmetrical arrangement of Tyr133 (Kelly et al., 2005). The equilibrium unfolding transition monitored by far-UV CD at 222 nm shows a two-state transition (Figure 33C).

Trp112Phe and Trp112Phe/His160Gly variants showed diminished CD signal and decreased stability (Figure 33A–C). Interestingly, the Trp148Phe variant had significant effects on the structure and stability of the protein. Equilibrium unfolding occurred rapidly and there was a 4.9 kcal/mol decrease in the free energy of stability (Figure 26B; Table 6). The far-UV CD signal was greatly diminished and the near-UV CD signal indicated loss of tertiary structure (Figure 33A–B). The His160Gly variant showed both a similar equilibrium CD spectra and two-state unfolding transition as the WT Fadd-DD (Figure 33A–C).

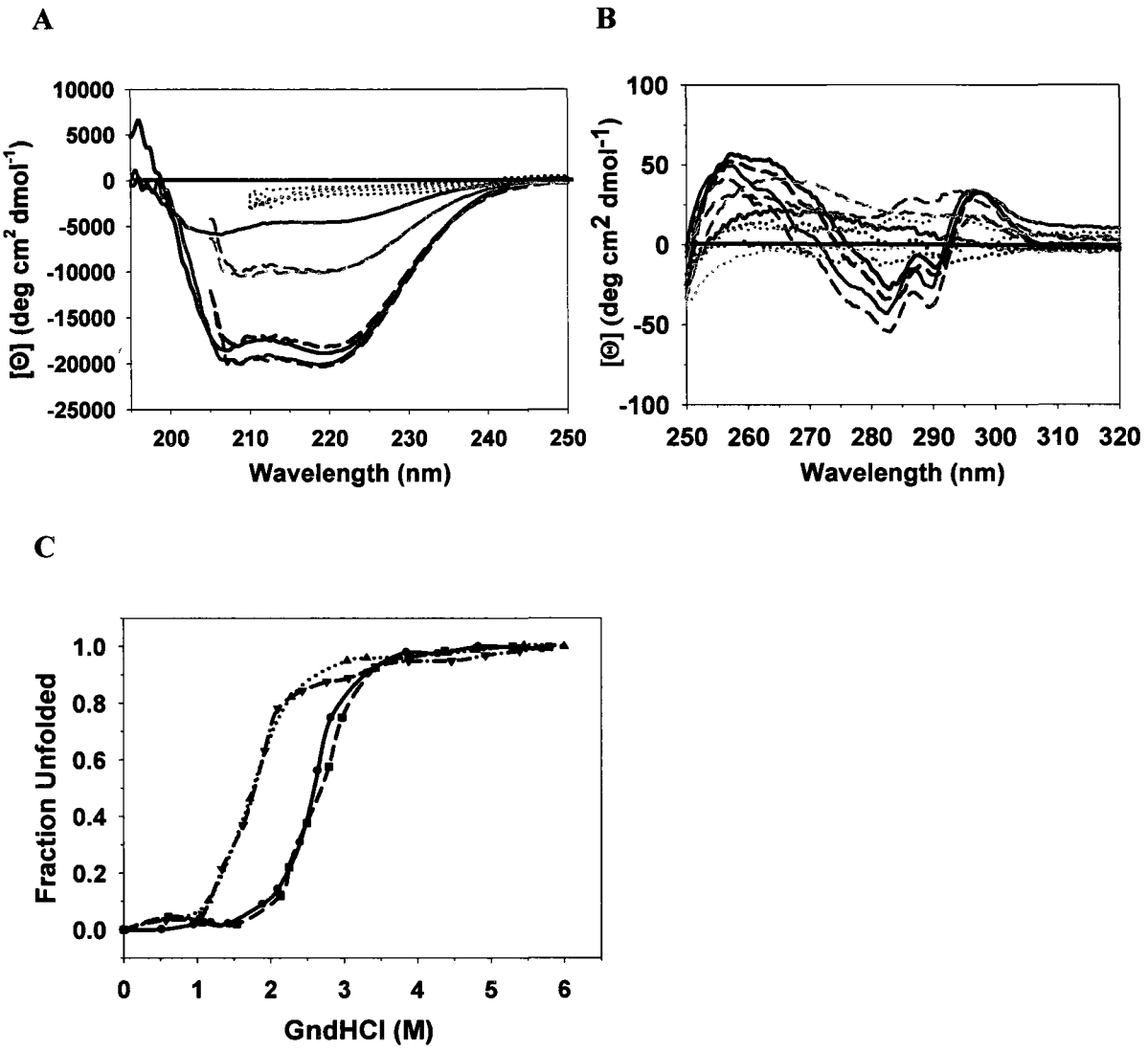


Figure 33. Circular dichroism spectra of native and denatured states of Fadd-DD

(A) The far-UV CD of 0.2 mg/ml of Fadd-DD (16 μM) in 20 mM Bis-Tris buffer (pH 6.2, 2 mM DTT) with different concentrations of GndHCl was measured at 20°C with the excitation from 190 nm to 250 nm. The solid line, dashed line and dotted line in black represent WT protein in 0 M, 1 M and 6 M of GndHCl, respectively. The dashed and dotted lines in cyan represent Trp112Phe in 1 M and 6 M GndHCl, respectively. The solid line in blue represents Trp148Phe in the native buffer. The solid line, dashed line and dotted line in green represent His160Gly in 0 M, 1 M and 6 M of GndHCl,

Figure 33 Continued.

respectively. The dashed and dotted lines in pink represent Trp112Phe/His160Gly in 1 M and 6 M GndHCl, respectively.

(B) The near-UV CD of 0.6 mg/ml of Fadd-DD (50 μM) in 20 mM Bis-Tris buffer with different concentrations of GndHCl was measured at 20 °C with excitation from 250 nm to 320 nm. The line types and colors are the same as those in Figure 33A.

(C) The equilibrium unfolding curves for the WT Fadd-DD and select variants appear to be a two-state transition. The proteins are annotated as follows: WT (solid-line with circles); Trp112Phe (dotted-line with triangle up); His160Gly (dashed-line with squares); Trp112Phe/His160Gly (dash-dot-line with triangle down). The data are graphed using SigmaPlot (Ver. 10).

Leu172Trp and Val158Trp Variants

The far- and near- UV CD spectra of Leu172Trp are shown in Figure 34. The far-UV CD spectrum of Leu172Trp is similar to the WT (Figure 34A). The near-UV CD spectrum of Leu172Trp is different from the WT (Figure 34B) and the ellipticity is smaller than that of the WT. The far- and near- UV CD spectra of Val158Trp are shown in Figure 35. The far-UV CD spectrum of Val158Trp is similar to the WT (Figure 35A). The ellipticity from 260 to 285 nm in the near-UV CD spectrum of Val158Trp is greater than that of the WT (Figure 35B), due to the introduction of the extra tryptophan.

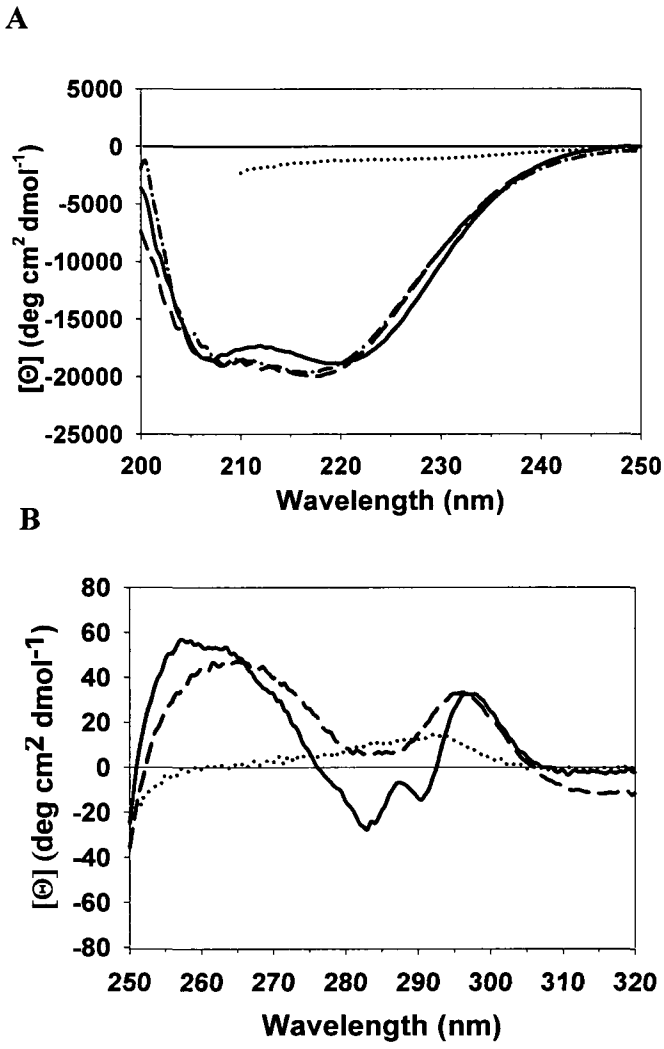


Figure 34. Circular dichroism spectra of WT Fadd-DD and Leu172Trp

(A) The far-UV CD of 0.2 mg/ml of Fadd-DD in 20 mM Bis-Tris buffer (pH 6.2, 2 mM DTT) measured at 20 °C with the excitation from 200 nm to 250 nm. The solid line represents WT protein in the buffer with 0 M GndHCl. Dashed line, dashed-dotted and dotted lines indicate Leu172Trp in the buffer with 0 M, 1 M and 5 M GndHCl, respectively.

(B) The near-UV CD of 0.5 mg/ml of WT Fadd-DD and Leu172Trp in 20 mM Bis-Tris buffer with different concentrations of GndHCl was measured at 20°C with excitation from 250 nm to 320 nm.

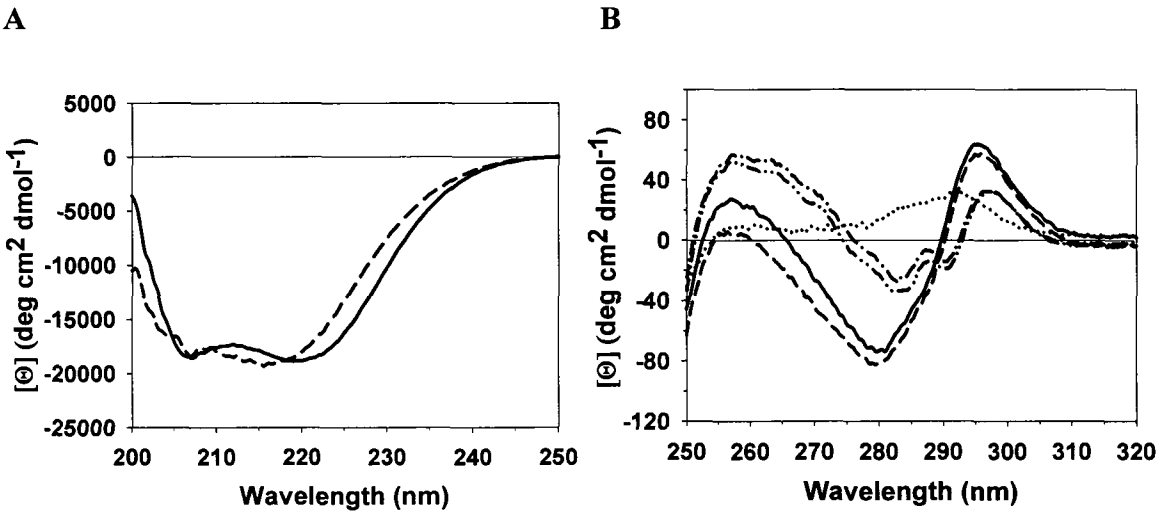


Figure 35. Circular dichroism spectra of WT Fadd-DD and Val158Trp

(A) The far-UV CD of 0.2 mg/ml of Fadd-DD in 20 mM Bis-Tris buffer (pH 6.2, 2 mM DTT) and Val158Trp was measured at 20 °C with the excitation from 200 nm to 250 nm. The solid line and dashed line represent WT protein and Val158Trp in native buffer, respectively.

(B) The near-UV CD of 0.5 mg/ml of WT Fadd-DD and Val158Trp in 20 mM Bis-Tris buffer with different concentrations of GndHCl was measured at 20°C with excitation from 250 nm to 320 nm. The solid line, dashed line and dotted line represent Val158Trp in the buffer with 0 M, 1 M and 6 M of GndHCl, respectively. The dashed-dotted and dashed-dotted-dotted lines represent WT protein in the buffer with 0 M and 1M GndHCl, respectively.

DISCUSSION

The equilibrium unfolding and refolding results clearly indicate that the unfolding transition of WT Fadd-DD is well-described by a two-state equilibrium folding model

and is reversible (Figure 26B, 33C). Interestingly, Fadd-DD exhibits fluorescence quenching. There are only a few proteins which demonstrate native-state fluorescence quenching effects. The most notable example is barnase. The emission intensity of wild-type barnase increases dramatically as the pH is increased from 7.0 to 8.5 (Loewenthal et al., 1991), which can be explained by the loss of the quenching effect of His18. This suggests that Trp94 is quenched by the protonated form of histidine (Lakowicz, 1999). Investigation of the NMR structure of barnase revealed that His18 and Trp94 are very close in the three-dimensional structure. As for Fadd-DD, His160 forms a hydrogen bond with Trp148, which was thought to be the most likely cause of the intrinsic fluorescence quenching. Mutagenesis of the histidine residue however, did not eliminate the fluorescence quenching effect. Other amino acids are known to quench tryptophan fluorescence, such as arginine, lysine and phenylalanine (Chen and Barkley, 1998; Clark et al., 1996; Hennecke et al., 1997). Therefore, the likely candidates for the source of fluorescence quenching are Trp112 by one or both of the flanking arginines and/or Trp148 by Phe101 or Lys149 (Figure 27).

Biophysical characterization of the four Fadd-DD variants revealed that changes to the conserved Trp112 and Trp148 had adverse effects on the stability. The degree of the effect for the Trp148Phe substitution was, however, surprising. Within the superfamily this site is highly conserved, but there are instances where other superfamily members have phenylalanine or even isoleucine in place of the tryptophan (Figure 9-11). The large effect of a Trp mutation is however, not without precedent. The Trp109Phe substitution in the cellular retinoic acid-binding protein also resulted in significant destabilization of the native protein, substantial aggregation and little yield (Clark et al.,

1996). The substitution of Trp133 to valine in *E. coli* dihydrofolate reductase resulted in a partially unfolded protein and a 5 kcal/mol decrease in free energy (Ohmae et al., 2001). In lysozyme, mutation of Trp62 to glycine led to misfolding due to the significant loss of long-range interactions (Zhou et al., 2007). In β2-microglobulin, mutation of Trp95 to glycine destabilized the native state and greatly diminished secondary and tertiary structures (Esposito et al., 2008).

Conversely, the His160Gly had no adverse affects on structure and stability. In fact, the stability marginally increased. The substitution for glycine was made on the basis that a naturally occurring glycine substitution occurs at this location in the mouse sequence. Thus, the mutagenesis and biophysical study on a pair of interacting residues where one is conserved and one is not provides insight into the role of the conserved residues play in structure and stability. The effect of mutating Trp112 to phenylalanine was pronounced. The change to this conserved residue produces a protein which is destabilized. The Trp112Phe variant in conjunction with the other four variants (Trp148Phe, His160Gly, Leu115Ala and Val121Ala) thus provides initial support for the hypothesis that the conserved residues are important for stability and the structure of Fadd-DD. Further, these results indicate that conserved residues themselves can have varying roles of importance in stability and structure. To summarize, Trp112Phe, Trp148Phe, Trp112Phe/His160Gly, Leu115Ala and Val121Ala variants decrease the stability. His160Gly, Leu172Trp and Val158Trp variants do not change the stability significantly.

A research group led by Dr. Jane Clarke (University of Cambridge, England) (Steward et al., 2009) showed that the variants of Trp112Ala, Leu115Met, Leu119Met,

Ile129Ala, Leu145Met, Trp148Phe, Leu161Ala, Leu165Ala, Val173Ala and Val177Ala had decreased equilibrium stability compared to the WT. These ten residues are conserved according to our bioinformatics study (Li et al., 2009). These results further support our hypothesis about the nature of conservation in Fadd-DD.

Fadd-DD is the only member of death domain superfamily to be studied by site-directed mutagenesis. So the effects on stability in Fadd-DD cannot be compared at this time to other DDs. See future work where I propose studying pyrin protein from this superfamily. The PYD has been studied thermodynamically (Dalal and Pio, 2006). Its stability is 1.75 kcal/mol, lower than Fadd-DD.

CHAPTER V

IMPORTANCE OF CONSERVED HYDROPHOBIC RESIDUES IN FOLDING OF FADD-DD

INTRODUCTION

Attention has been paid to the link between protein sequence conservation and folding (Greene et al., 2003). In the section of bioinformatics, we propose that nine significantly conserved and seven moderately conserved hydrophobic residues form a network, which is important for the folding of Fadd-DD. In order to test this hypothesis, site-directed mutagenesis was employed to produce the protein variants. Experimental studies were conducted to characterize the folding behavior of Fadd-DD using four biophysical techniques; which include stopped-flow fluorescence spectroscopy and CD. The folding kinetics were then compared between the WT and variants.

To probe the role of the conserved residues hypothesized to be important in folding, four protein variants were analyzed: Trp112 to phenylalanine, Trp148 to phenylalanine, His160 to glycine and Trp112Phe/His160Gly. Previous results showed that the significantly reduced the side chain of His160 to glycine had no negative effects on native state stability; the small reduction in side chain of Trp148 to phenylalanine had significant destabilizing effects on the structure; and the change of Trp112 to phenylalanine was also destabilizing though to a lesser degree.

Folding kinetics studies of three CARD proteins in the death domain superfamily have been described previously and display complex behavior consisting of multiple

phases, parallel pathways and kinetically trapped species (Chen and Clark, 2004; Chen and Clark, 2006; Milam et al., 2007). In contrast, the folding behavior of WT Fadd-DD is rapid, biphasic and the majority of the hydrophobic core is formed in the first phase.

RESULTS

Folding Kinetics of WT Fadd-DD, Trp112Phe and His160Gly Variants

Denatured protein was refolded in a final concentration of 1 M GndHCl to achieve the maximum change in fluorescence between the native and unfolded states. We established that Fadd-DD is native at this concentration of denaturant, because the ratio of fluorescence intensities at 370 nm over 330 nm and the ellipticities at both near- and far-UV CD are similar for Fadd-DD in 0 M and 1 M GndHCl (Figure 26 and Figure 33). The time course of refolding of WT Fadd-DD was examined in single-mixing stopped-flow experiments by monitoring changes in fluorescence emission (Figure 36A). The residuals of the fit to a biphasic relaxation equation are shown in Figure 36A. The relaxation time of the first phase in the hydrophobic core formation is 44.0 ms and that of the second is 123.0 ms (Table 7). The gap between the denatured state baseline and the start of the refolding trace indicates that there is a burst phase within the dead time of the experiment. Analysis of the amplitudes indicates that 28%, 55% and 17% of the folding occurs in the burst phase, the first phase and the second phase, respectively. The folding rates of different concentrations of Fadd-DD (ranging from 0.01 to 0.2 mg/ml) were similar to those of 0.05 mg/ml, indicating that the folding rates are not concentration dependent (see Figure A7 and Table A10).

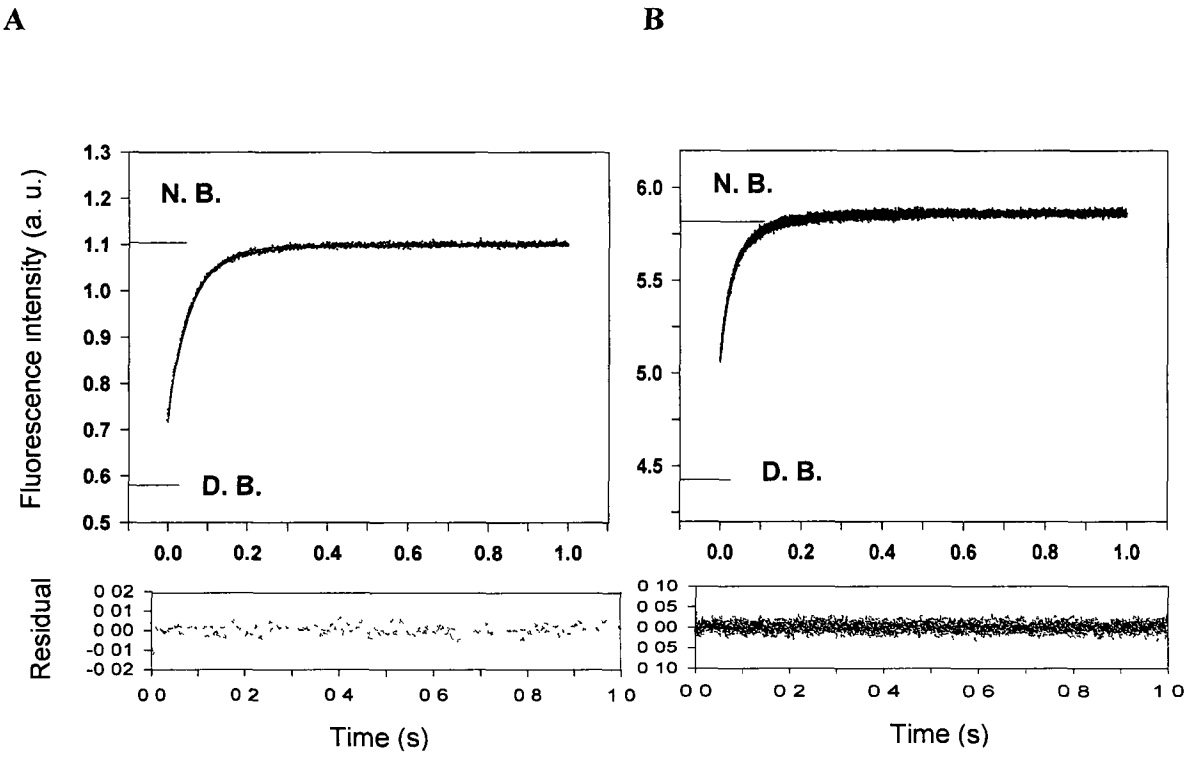


Figure 36. Stopped-flow fluorescence studies of Fadd-DD WT and His160Gly

The folding process was initiated with a 1:5 dilution at 20°C. The native state and denatured state baselines are denoted by N. B. and D. B., respectively.

(A) Refolding of denatured WT protein at 0.05 mg/ml in 20 mM Bis-Tris buffer (pH 6.2) and 2 mM DTT was monitored by stopped-flow fluorescence spectroscopy. The excitation wavelength was 295 nm and emission was monitored using a bandpass filter (300–340 nm) therefore the emission increases (see Figure 26A). The dead time was 8.6 ms. The residuals of the fit to a double exponential equation are shown below the time course.

(B) Refolding of denatured His160Gly at 0.05 mg/ml in Bis-Tris buffer was monitored by stopped-flow fluorescence spectroscopy. The emission was monitored with the same bandpass as the WT protein. The residuals based on the fit to a double exponential equation are also shown.

Table 7. Refolding kinetics of WT Fadd-DD, His160Gly and Trp112Phe and unfolding of WT monitored by intrinsic tryptophan fluorescence

Protein	A_1	k_1 (s^{-1})	A_2	k_2 (s^{-1})	A_3	k_3 (s^{-1})
WT refolding	0.29 ± 0.01	22.73 ± 0.57	0.09 ± 0.01	8.13 ± 0.47	–	–
His160Gly	0.58 ± 0.01	34.05 ± 0.33	0.20 ± 0.01	8.51 ± 0.15	–	–
Trp112Phe	0.81 ± 0.01	11.97 ± 0.15	0.26 ± 0.01	4.13 ± 0.15	0.08 ± 0.00	0.33 ± 0.01
WT unfolding	0.28 ± 0.00	6.02 ± 0.08	0.04 ± 0.00	1.39 ± 0.21	-	-

Refolding and unfolding are performed in 1 M GndHCl (pH 6.2) and 6 M GndHCl (pH 6.2), respectively.

The time course of refolding for the His160Gly variant was analyzed by stopped-flow fluorescence spectroscopy (Figure 36B). The timescales were relatively similar indicating that the substitution had only a small effect on folding kinetics (Table 7). In fact, the slight increase in stability appears to correspond to the marginal increase in folding rates. Due to the inability to acquire sufficient amounts of the Trp148Phe variant and its partially unfolded character, folding kinetic studies could not be conducted. The Trp112Phe variant was also characterized by stopped-flow fluorescence and the folding analysis revealed three phases (Figure 37). The relaxation timescales are as follows: 84 ms, 242 ms and 3 s (Table 7). This change from a double to a triple exponential equation is based on the significant improvement in the fit and the distributions of the residuals (Figure 37).

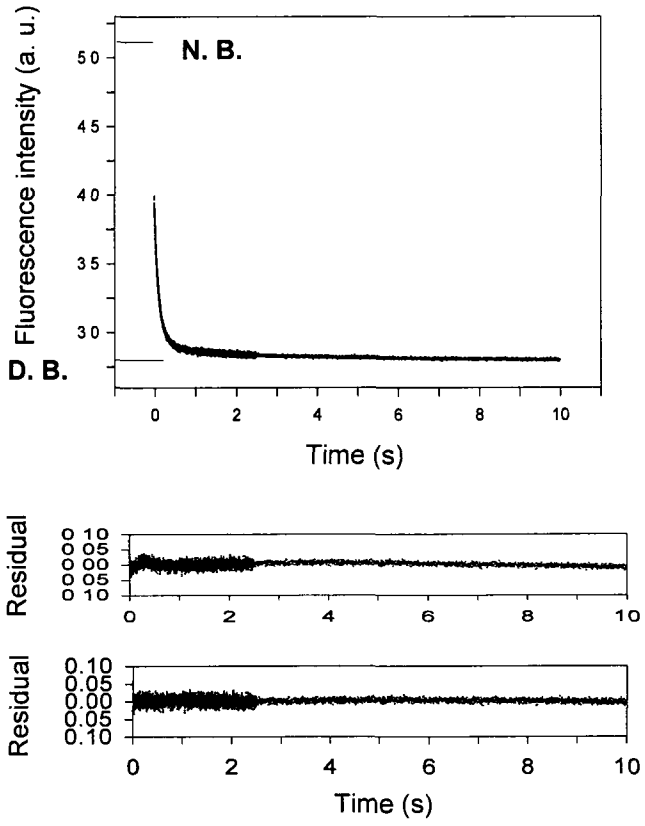


Figure 37. Stopped-flow fluorescence studies of Trp112Phe

Refolding of denatured Trp112Phe at 0.12 mg/ml in Bis-Tris buffer was monitored by stopped-flow fluorescence spectroscopy. The emission was monitored by a bandpass filter (362–396 nm) therefore the intensity change goes down (see Figure 28A). The residuals of the fits are shown for both a double and triple exponential equations are shown.

The time course of unfolding of WT Fadd-DD was examined in single-mixing stopped-flow experiments by monitoring changes in fluorescence emission (Figure 38). The residuals of the fit to a biphasic relaxation equation are also shown. The relaxation

time of the first phase in the hydrophobic core formation is 166 ms and that of the second is 719 ms (Table 7).

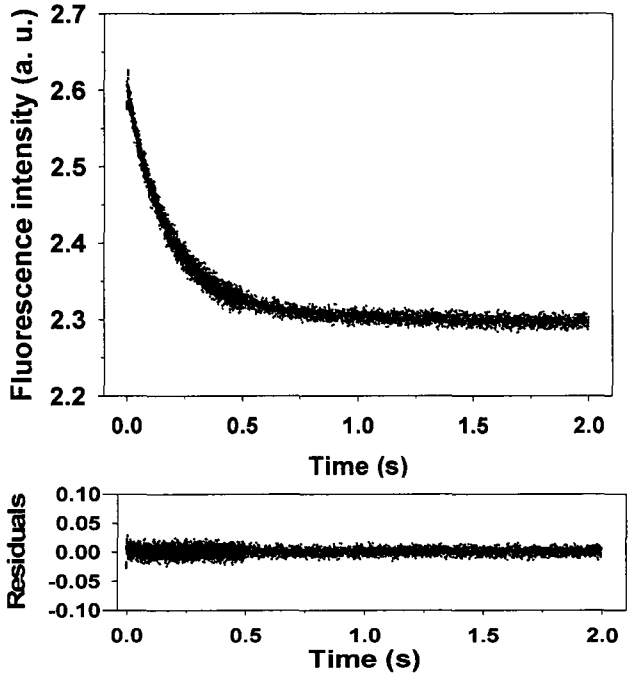


Figure 38. Stopped-flow fluorescence unfolding studies of Fadd-DD in 6 M GndHCl The unfolding process was initiated with a 1:5 dilution at 20°C. Unfolding of native WT protein at 0.12 mg/ml in 20 mM Bis-Tris buffer (pH 6.2), 2 mM DTT and 1 M GndHCl was mixed with buffer with 7 M GndHCl and monitored by stopped-flow fluorescence spectroscopy. The excitation wavelength was 295 nm and emission was monitored using a bandpass filter (300–340 nm) therefore the emission change goes down. The dead time was 8.6 ms. The residuals of the fit to a double exponential equation are shown below the time course.

Stopped-flow Far-UV CD Spectroscopy

The folding kinetics of Fadd-DD at a final GndHCl concentration of 1 M, monitored by far-UV CD reveals the rate of the secondary structure formation (Figure 39). One single phase with a rate of 23.4 ± 0.4 s^{-1} and amplitude of 16.8 ± 0.1 millidegrees was calculated.

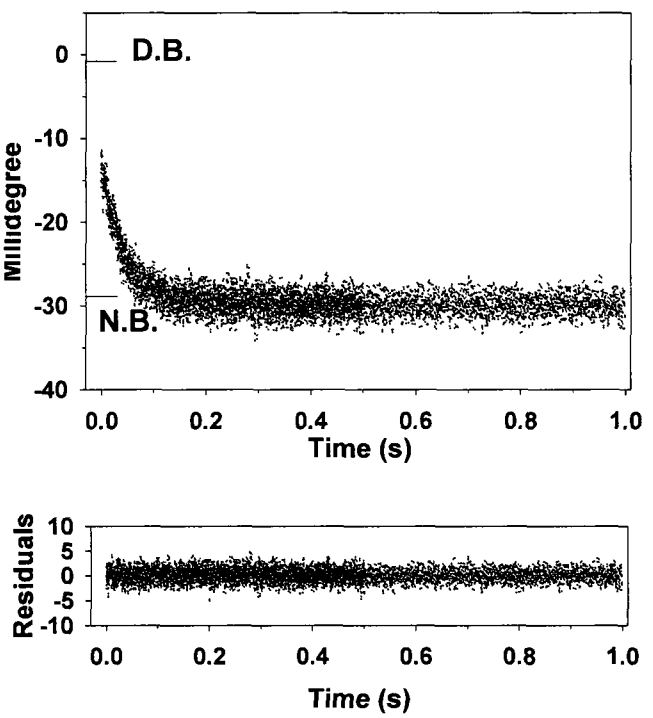


Figure 39. Stopped-flow far-UV CD study of Fadd-DD

The folding process was initiated with a 1:5 dilution at 20 °C. Refolding of denatured WT protein at 0.15 mg/ml in 20 mM MES buffer (pH 6.2) and 5 mM DTT was monitored by stopped-flow far-UV CD spectroscopy at 225 nm. The dead time was 9.3 ms. The residuals are shown below the time course. D.B. and N.B. stand for the denatured baseline and native baseline, respectively.

Stopped-flow Near-UV CD Spectroscopy

The stopped-flow near-UV CD experiment conducted on Val158Trp does not demonstrate any noticeable change of CD signal in folding (Figure 40). This is probably due to the inadequte difference of ellipticity between the native and denatured states, which is approximately 80 deg*cm^2*dmol^{-1} at 280 nm (Figure 35B).

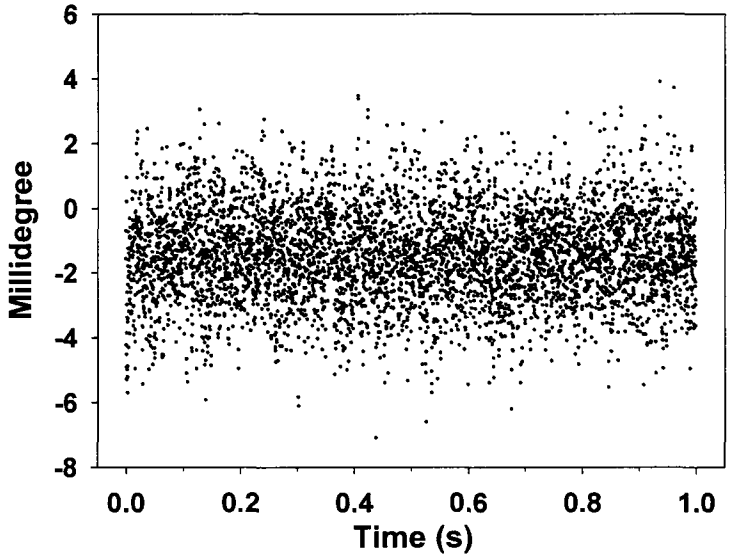


Figure 40. Stopped-flow near-UV CD study of 0.3 mg/ml Val158Trp

The folding process was initiated with a 1:5 dilution at 20 °C. Refolding of denatured Val158Trp variant at 1.5 mg/ml in 5 M GndHCl, 20 mM Bis-Tris buffer (pH 6.2) and 2 mM DTT was monitored at 285 nm by stopped-flow near-UV CD spectroscopy with TC-100 cuvette. The path-length is 10 mm. The traces shown are an average of thirty shots.

By comparison, the folding of hen lysozyme was implemented at 289 nm with the final concentration of 1 mg/ml (Radford et al., 1992) and the difference of of ellipticity between the native and the denatured states is about 120 deg*cm^2*$dmol^{-1}$ at 289 nm (Sasahara et al., 2000). Construction of one more tryptophan to the Val158Trp is needed in order to create more near-UV CD signal for folding study.

DISCUSSION

The folding behavior of WT Fadd-DD is both rapid and biphasic. Approximately 80% consolidation of the hydrophobic core appears to occur in the intermediate state. The absence of a slow phase suggests proline isomerization does not exist in the folding process. Because Fadd-DD contains only one tyrosine on the surface, the difference in near-UV CD ellipticity between the native and denatured states is relatively small. Therefore, the timescale of the tertiary structure formation could not be obtained. Mutation of one exposed residue to tryptophan could increase the difference in the near-UV CD spectra. Fadd-DD uniquely has a hydrogen bond between the side chain of His160 and the highly conserved Trp148 in the core. Therefore, Trp148 is not only an ideal probe of the hydrophobic collapse but also to monitor the structuring of the core by techniques such as quenched-flow hydrogen–deuterium exchange.

Within the death domain superfamily the folding behavior of three members of the CARD family have been previously studied. These are procaspase-1, RICK-CARD and Apaf-1-CARD. All three show complex kinetics. The refolding of RICK-CARD contains multiple phases as well as kinetically trapped species, which are unrelated to proline isomerization (Chen and Clark, 2003; Chen and Clark, 2006). Apaf-1-CARD, like RICK-CARD, appears to fold via parallel paths due to two unfolded conformations

(Milam et al., 2007); while procaspase-1, like RICK-CARD, also folds through kinetically trapped species (Chen and Clark, 2004). In comparison, the folding pathway of Fadd-DD appears to be more straightforward and is thus a good model in which to investigate the determinants of the all-α-helical Greek-key topology.

Biophysical characterization of four Fadd-DD variants revealed that changes to the conserved Trp112 had adverse effects on folding while the Trp148 variant could not even be studied due to its low stability. The degree of the effect for the Trp148Phe substitution is not without precedence. In antichymotrypsin, the Trp194Phe mutation lowered the kinetic barrier to misfolding (Pearce et al., 2007).

Conversely, the His160Gly had no adverse affects on structure, stability or folding. In fact, the refolding rates marginally increased. Thus the mutagenesis and biophysical study on a pair of interacting residues where one is conserved provides initial insight into the role of the conserved residues in folding. The affect of mutating Trp112 to phenylalanine was pronounced. The change to this conserved residue produces a protein which folds significantly more slowly. In addition to a doubling in the folding timescales of the first two phases, a third phase now appears and is on the second not millisecond timescale. Analysis of mutated conserved residues in other proteins such as apo-azurin also show changes to the folding rate in proteins (Engman et al., 2004). The Trp112Phe variant in conjunction with the other two variants (Trp148Phe and His160Gly) thus provides initial support for the hypothesis that the conserved residues are important for folding of Fadd-DD. Furthermore, these results indicate that conserved residues themselves can have varying roles of importance at different stages of folding.

To summarize, the folding rate of Trp112Phe is slower than the WT. Trp148Phe, Leu115Ala and Val121Ala variants cause misfolding. There are two phases for the hydrophobic core formation and only one phase for the secondary structure formation. The majority of hydrophobic core formation occurs at similar timescale with the secondary structure formation, indicating the hydrophobic core and secondary structure form concomitantly. The second minor phase is probably due to the complete exclusion of the two tryptophans from the solvent at the later stage of folding.

A research group led by Dr. Jane Clarke (Steward et al., 2009) showed that the folding rates of Trp112Ala, Leu115Met, Leu119Met, Leu145Met, Leu161Ala, Leu165Ala, Val173Ala and Val177Ala in 2 M urea were slower than that of the WT. These eight residues are conserved according to our bioinformatics study (Li et al., 2009). These studies provide support for our hypothesis about the nature of conservation in Fadd-DD.

Fadd-DD is the only member of death domain superfamily to be studied by site-directed mutagenesis. So the effects on folding in Fadd-DD cannot be compared at this time to other DDs. See future work where I propose studying pyrin protein from this superfamily.

CHAPTER VI

FOLDING OF FADD-DD MONITORED BY QUENCHED-FLOW, HYDROGEN-DEUTERIUM EXCHANGE AND NMR SPECTROSCOPY

INTRODUCTION

Fadd-DD is composed of six α-helices which is shown in Figure 41A. The folding pathway of Fadd-DD is more straightforward than the folding of CARDs and is thus a good model with which to investigate the determinants of the all α-helical Greek-key topology. We are interested in determining if the helices which make up the canonical Greek-key structure (helices 1, 2, 4, 5) form on a faster folding time-scale than the other helices (3 and 6) (Higman and Greene, 2006; Steward et al., 2009). Stopped-flow fluorescence studies by our group indicate that folding is biphasic (k_1 = 22.73 $s^{-1}$and k_2 = 8.13 s^{-1}) with the majority of the folding occurring in the first phase (amplitude = 76%) (Li et al., 2009). In our condition, the refolding buffer is 20 mM Bis-Tris buffer (pH 6.2) and temperature is 20°C. In folding studies conducted by Steward et al. (2009) one rapid phase with a rate of 40 s^{-1} was detected using different experimental conditions, including the refolding buffer of 50 mM phosphate buffer (pH 7.0) and 150 mM NaCl and temperature is 25°C. In our present experiments using stopped-flow far-UV CD spectroscopy, the folding of the secondary structure of Fadd-DD was determined to be monophasic with a rate similar to that of the majority of hydrophobic core formation. In a more detailed study using a combination of quenched-flow HX and NMR spectroscopy the folding of twenty-two amide hydrogens in the backbone of helices and two amide

hydrogens in the backbone of loops are monitored. The results indicate that the folding of all six helices is monophasic and the formation of hydrogen bonded secondary structure is fundamentally cooperative. In addition, the equilibrium HX of Fadd-DD was also performed and the exchange rates of twenty-three residues are calculated. Most of the amide protons that are slowest to exchange are in the core region. This experiment is the first time to the best of our knowledge that folding kinetics monitored by hydrogen bond formation was conducted on an all α-helical protein with a Greek-key topology.

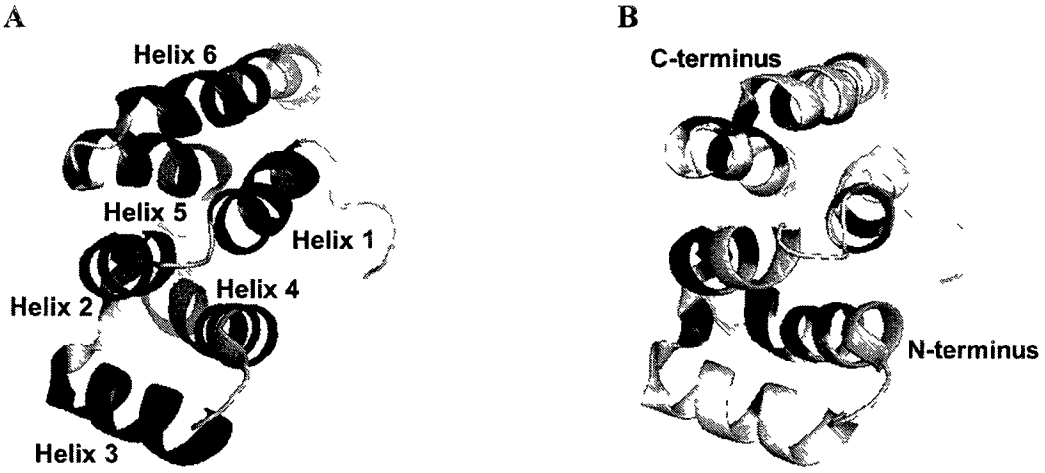


Figure 41. Structure of Fadd-DD showing stable backbone amides

(A) Ribbon diagram of Fadd-DD drawn in PyMOL v0.99 (DeLano Scientific). The six helices are annotated.

(B) Location of the twenty-three backbone amides which persist for over twenty-four hours and thus have the greatest protection in Fadd-DD (F101, V103, C105, L115, R117, Q118, L119, V121, I129, V141, S144, L145, I147, W148, K149, V162, G163, A164, R166, M170, A174, V177, Q182) are shown in black in the context of the secondary and tertiary structure using PyMOL.

RESULTS

Test of Protein State in Different pHs

The fluorescence was used to monitor the state of the protein in different pHs in a sequence of quenched-flow experiments. The fluorescence spectra of the sample starting with the denatured protein resembles that of the native protein of all the pHs tested (Figure 42), indicating the protein maintains its native state after refolding.

^{1}H-NMR Spectra

The ^{1}H-NMR spectrum of 4 mg/ml Fadd-DD in 100% D_2O, 200 mM K_2HPO_4/100 mM citric acid and 5 mM DTT (pD 4.4) at 30°C are shown in Figure 43. The peak height of the methyl proton of Leu119 was used to calibrate the protein concentration in the samples.

Comparing the Exchange Rates from Equilibrium HX Studies

The native HSQC spectrum of Fadd-DD in K_2HPO_4/citric acid buffer (pH 4.8) and 10% D_2O is shown in Figure 44 and is similar to the previously published HSQC in potassium phosphate buffer (pH 6.2) (Berglund et al., 2000).

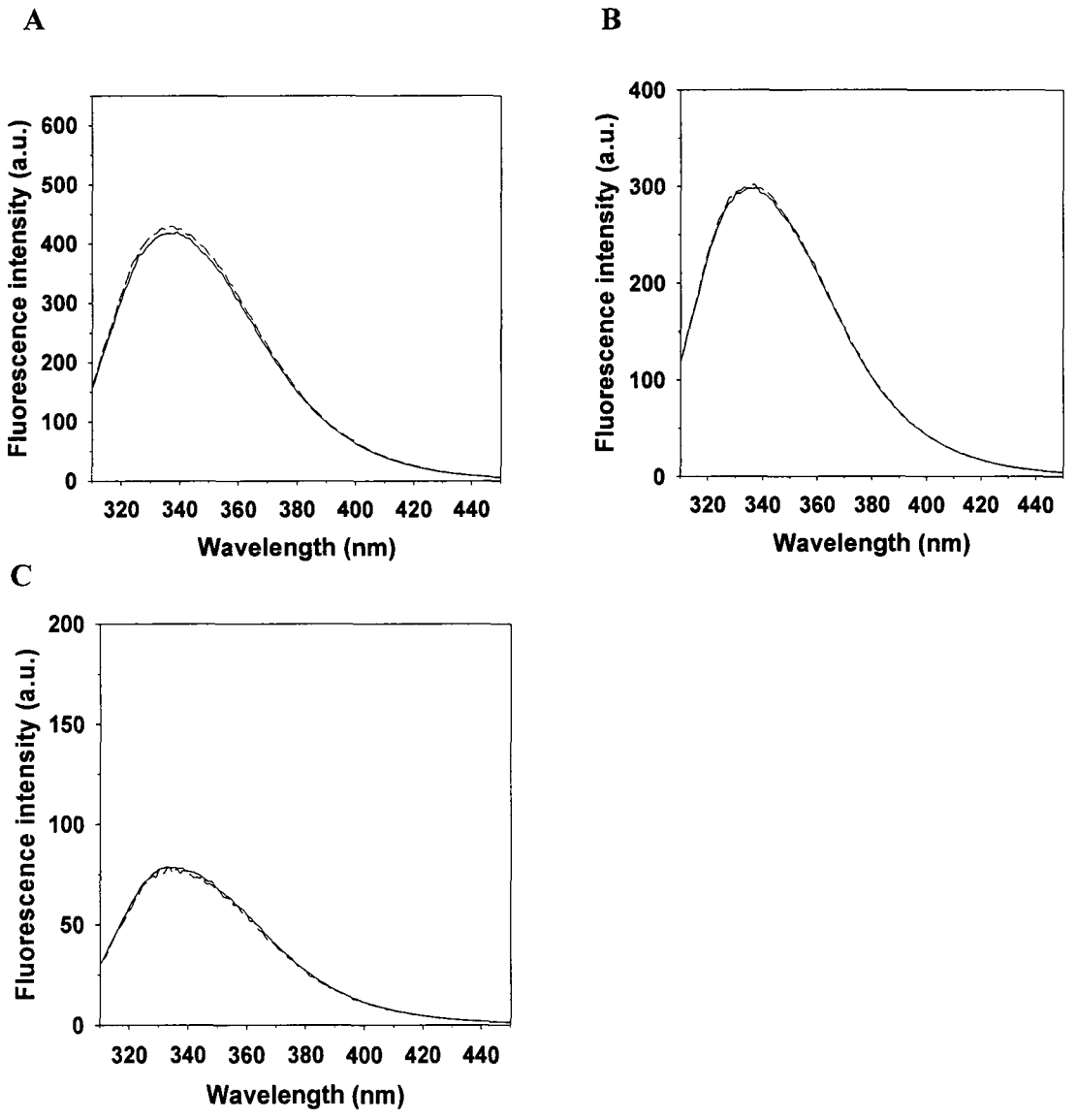


Figure 42. The fluorescence spectra of Fadd-DD in different buffers in the quenched-flow process

The solid line and dashed line indicate the fluorescence spectra of the sample starting with the denatured protein and that starting with the native protein, respectively. The parameters of the fluorescence experiment are the same as previous. Panels A, B and C show 0.2 mg/ml protein in refolding buffer (pH 6.2), 0.1 mg/ml protein in pulsing buffer (pH 9.8) and 0.05 mg/ml protein in quenching buffer (pH 4.8), respectively.

The HX rate is dependent on the conditions such as pH and temperature (Englander and Mayne, 1992). The rates of HX (k_{ex}) of slowly exchanged amide protons of native Fadd-DD were measured at pD 4.4 and 30°C by recording HSQC spectra every eight hours for one week. Exchanges of three amide protons were observed at this pH but were too fast for the rates to be measured. Except for some overlapping peaks in the spectra, a large number of identifiable backbone amide protons are exchanged during the process of buffer exchange. Thus, the exchange rates of twenty-three slowly exchanging amide protons which are defined as those that persist for twenty-four hours or more were calculated. The range is between $3.7x10^{-3}$ min^{-1} and $6.7x10^{-6}$ min^{-1}. Normalized intensities of the eight peaks with the slowest exchange rates are shown as a function of time in Figure 45A. Protected amides occur in all six helices (Figure 45B). However, the number of protected amides did not distribute evenly among the six helices; for example, only one amide proton I129 in helix 3 was protected. The HSQC spectra are shown in Figure 47. Figure 48 shows the hydrogen bonding pattern in Fadd-DD. Most assigned peaks are not visible after one week, except for three, W148, V162 and V177, which have the slowest exchange rates.

V177 has both the slowest exchange rate and the highest protection factor (Figure 45). The very stable and local hydrogen bond between W148 and S144 as well as between S144 and R140 suggests that these regions of helix 4 are very stable. Thus, it appears as if helix 4 is the most stable. The locations of the backbone amides which are stable for more than twenty-four hours is shown in Figure 41B and are distributed among the six helices, although most are located on helices 2, 4 and 5 (Figure 45B, 46). Among

the twenty-three stable amides, ten residues are conserved according to the bioinformatics study, indicating a moderate correlation between their stability and conservation.

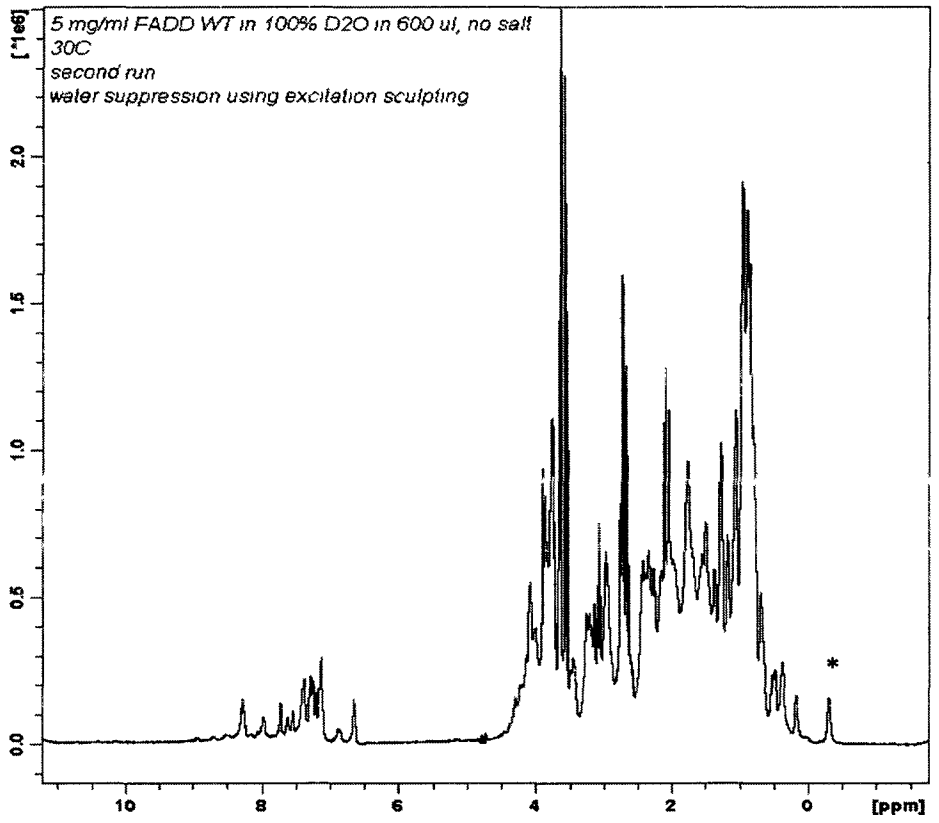


Figure 43. The ^{1}H-NMR spectrum of 4 mg/ml Fadd-DD in 100% D_2O, 200 mM phosphate/100 mM citric acid and 5 mM DTT (pD 4.4) at 30°C

The upfield peak marked with an asterisk shows the methyl proton of Leu119.

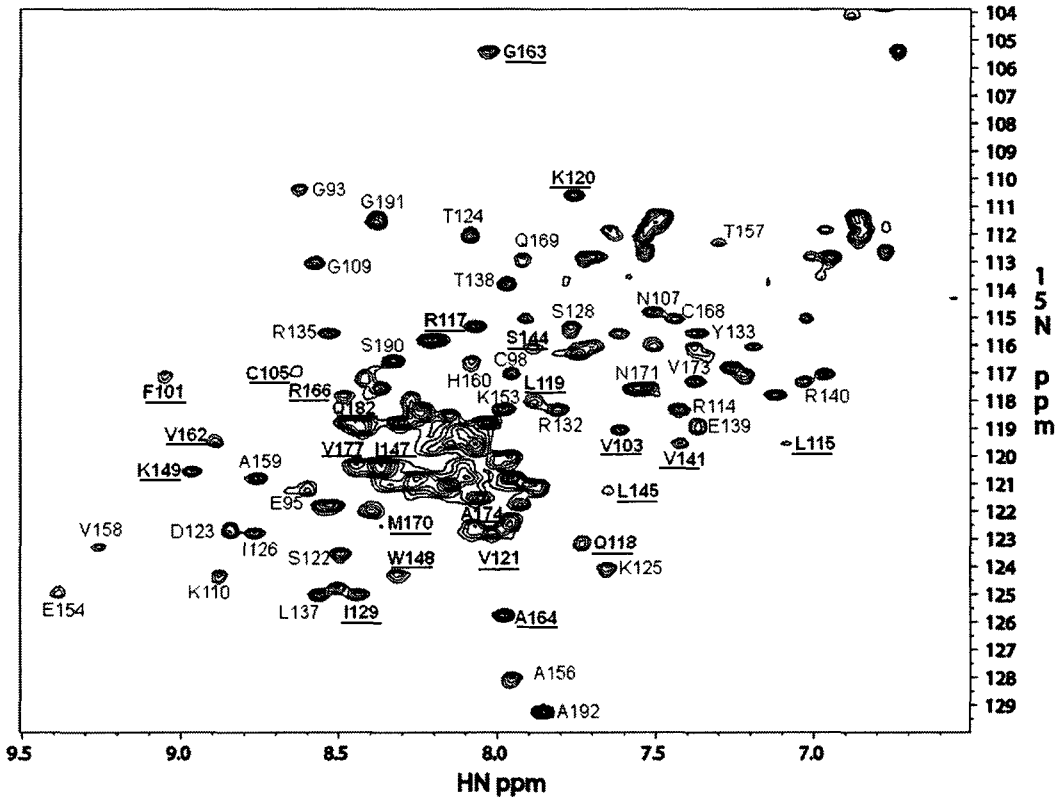


Figure 44. The HSQC spectrum of Fadd-DD in 50 mM citric acid/100 mM K_2HPO_4, 5 mM DTT (pH 4.8) and 10% D_2O at 30°C

The twenty-four stable amide protons in our kinetic study are indicated in bold and underlined.

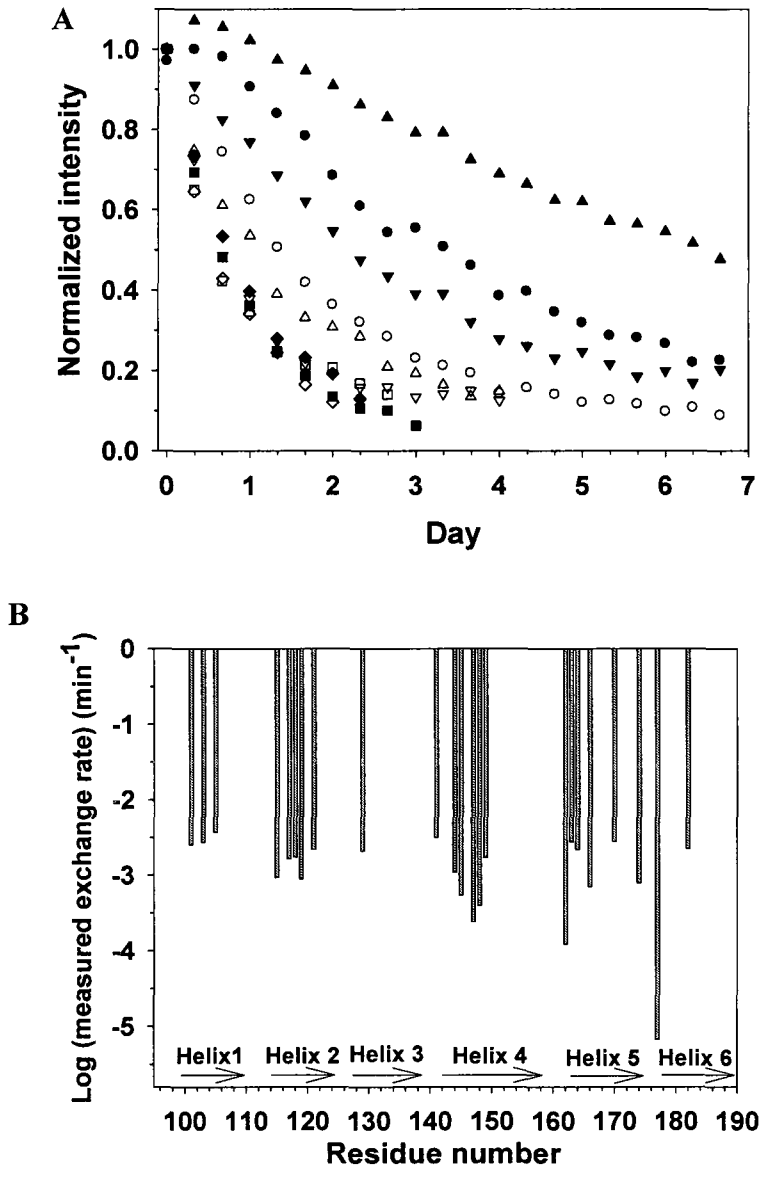


Figure 45. Equilibrium amide hydrogen exchange in Fadd-DD

(A) The HX of the ten slowest exchanging peaks are plotted (L115: white diamond, L119: white triangle down, S144: white square, L145: white triangle up, I147: black triangle down, W148: white circle, V162: black circle, R166: black diamond, A174: black square, V177 black triangle up).

Figure 45 Continued.

(B) The histogram shows the distribution of experimentally calculated exchange rates for the native protein at pD 4.4 and 30°C versus residue number. Regions of secondary structure are indicated schematically.

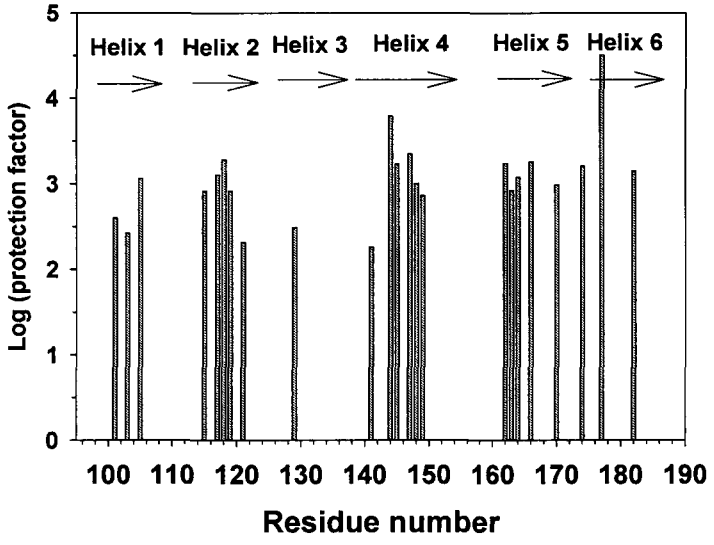


Figure 46. Distribution of protection factors for Fadd-DD

The histogram shows the distribution of protection factors for the native protein at pD 4.4 and 30°C versus residue number. Regions of secondary structure are indicated schematically.

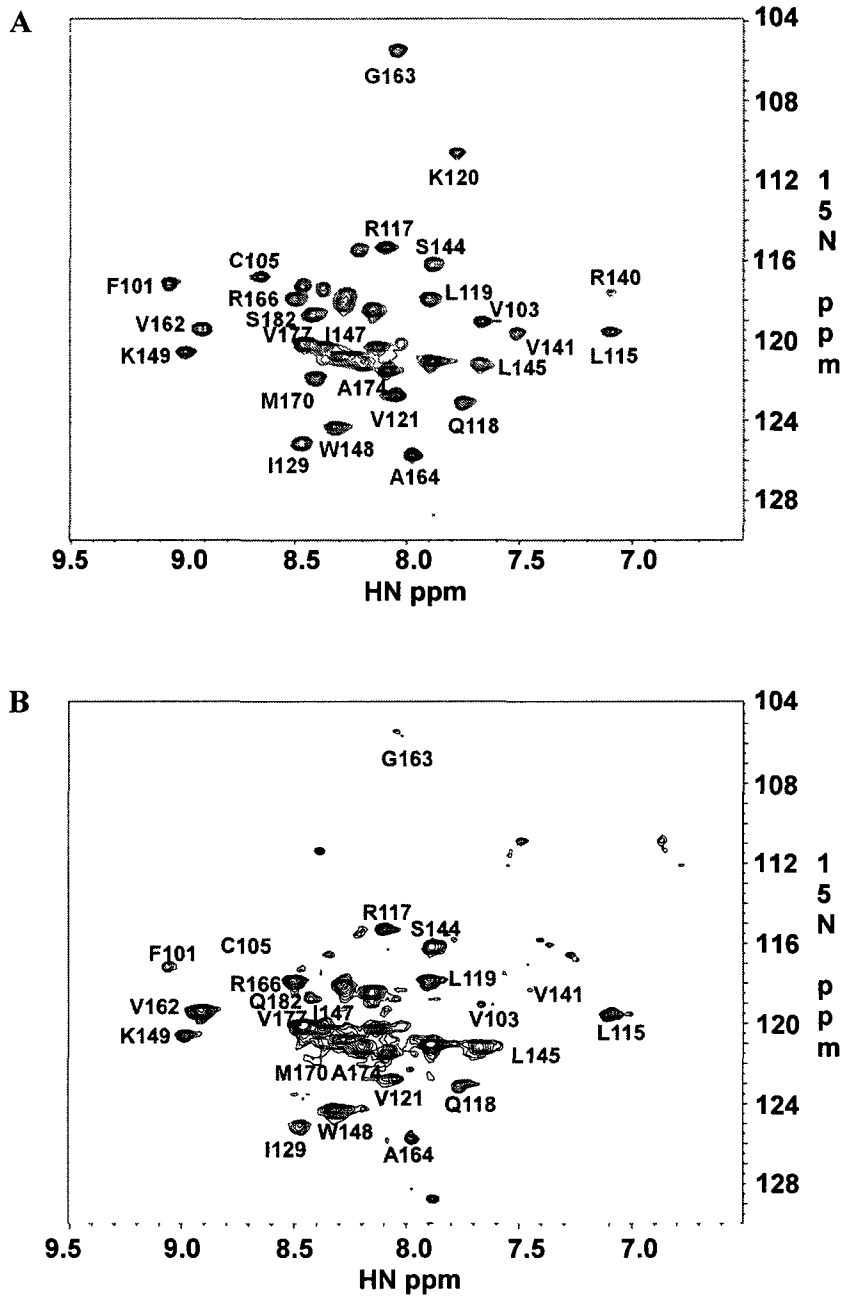
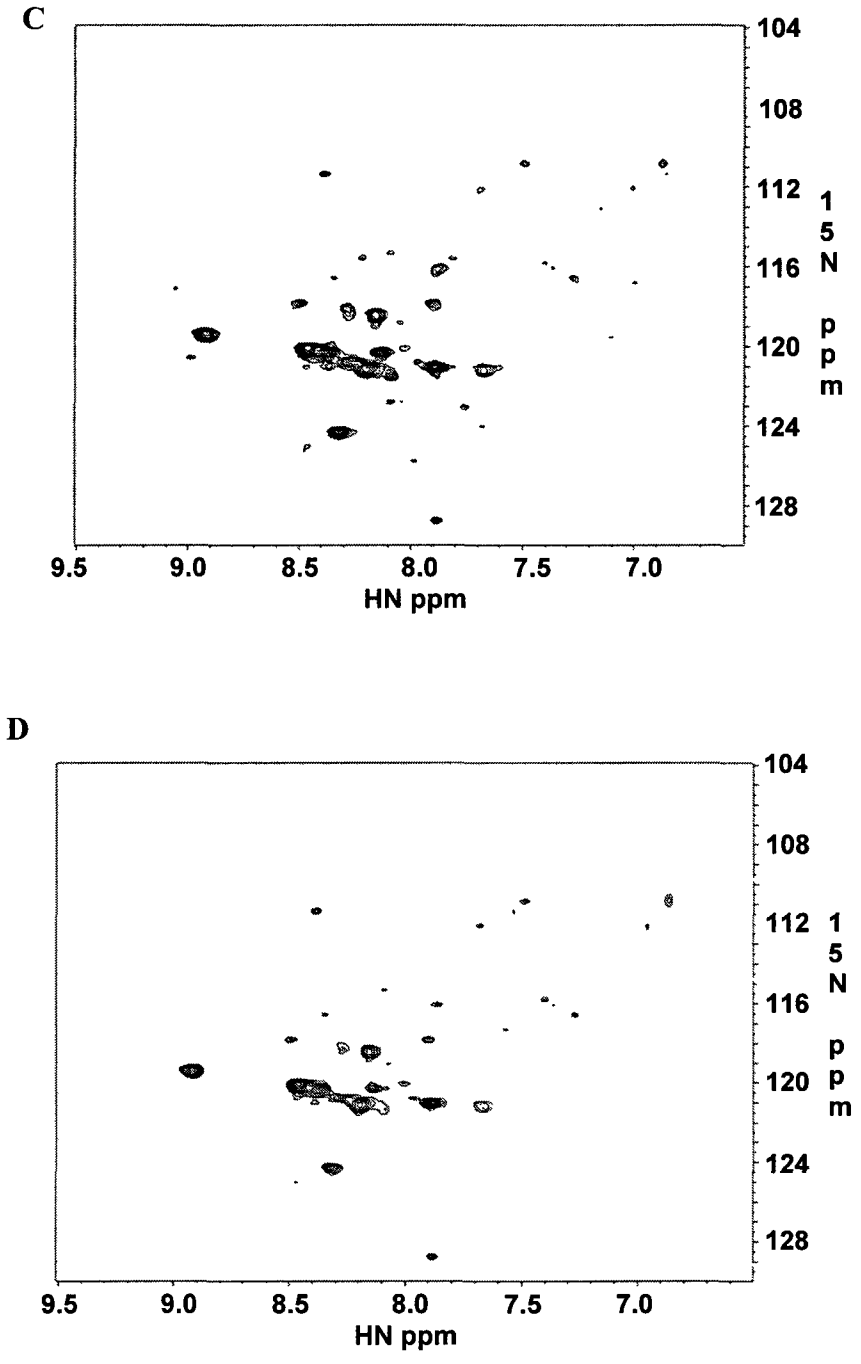


Figure 47. The HSQC spectra of 4 mg/ml Fadd-DD after exchange into D_2O buffer containing 50 mM citric acid/100 mM K_2HPO_4 (pD 4.4)

Panels A, B, C and D are day 1, 2, 3 and 4, respectively. The contour level of the HSQC spectrum of day 1 is 6 x 10^6. In order to see all the peaks, the contour level of the spectrum of day 2 is reduced to 1.5 x 10^6, which is the same as those of day 3 and 4.

Figure 47 Continued.



Hydrogen Exchange and Quenched-flow Coupled with NMR Spectroscopy

Following HX in D_2O buffer, a total of twenty-four well-resolved and assignable amide protons were identified to be stable. There is one more stable amide because the time for acquiring NMR spectra for the quenched-flow study is about six hours, which is shorter than that for the equilibrium study. The amide protons of N107 and R140 have very high exchange rates and the peaks disappear after sixteen hours of exchange. They are therefore not included in the quenched-flow kinetic studies, leaving twenty-two residues to be analyzed. A total of twenty-two amide protons could be monitored in the six helices.

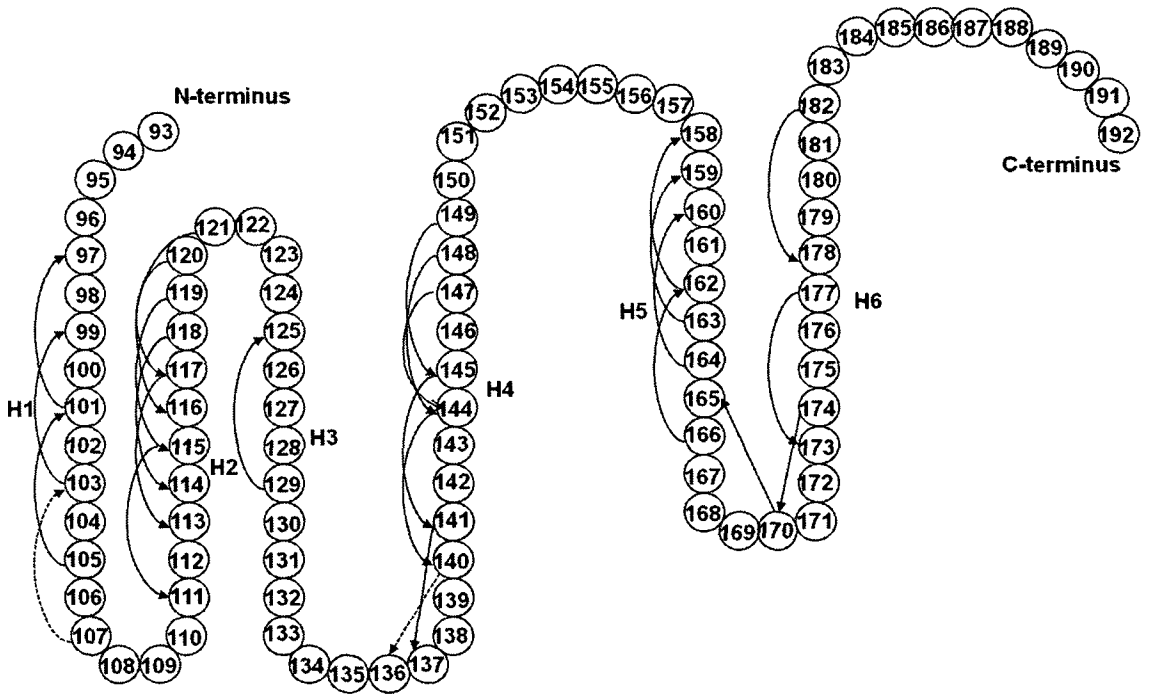


Figure 48. The hydrogen bonding pattern of Fadd-DD

The curves and arrows indicate the hydrogen bonding and the direction. The dashed lines show the two unstable hydrogen bonds of N107 and R140.

The ratio of volume among the protein solution, refolding buffer and pulse buffer is 1:4:5. The content of proton and deuterium after mixing of the three buffers are 90% and 10%, respectively. Therefore, the maximum proton occupancy of the quenched-flow study is 90%. In our study, the maximal proton occupancy is about 90%.

After the quenched-flow experiments were performed with eight different refolding times, the structural details of the folding reaction were obtained by analyzing the samples by 2D ^{1}H-^{15}N HSQC measurements. Figure 49 displays the protection time courses obtained for individual amides, grouped according to their distribution in different helices. HSQC spectra are also shown in Figure 50. In addition, two amide protons from two residues (V121 and M170) located in loop structures could also be monitored with folding rates of $20.72 \pm 7.03\ s^{-1}$ and $21.34 \pm 6.74\ s^{-1}$, respectively (Figure 51). All of the kinetics were found to be monophasic and a single exponential function could be used to fit the data satisfactorily (Figure 49). The resulting curves are all virtually identical, with rates between $19\ s^{-1}$ and $22\ s^{-1}$ (Table 8). The average rate of hydrogen bond formation is $20.85 \pm 1.66\ s^{-1}$ (Table 8). The results indicate that all detectable amides acquire protection from exchange concomitantly, with simple mono-exponential time courses. No intermediate or partially protected species are observable in these experiments and folding appears to be highly cooperative.

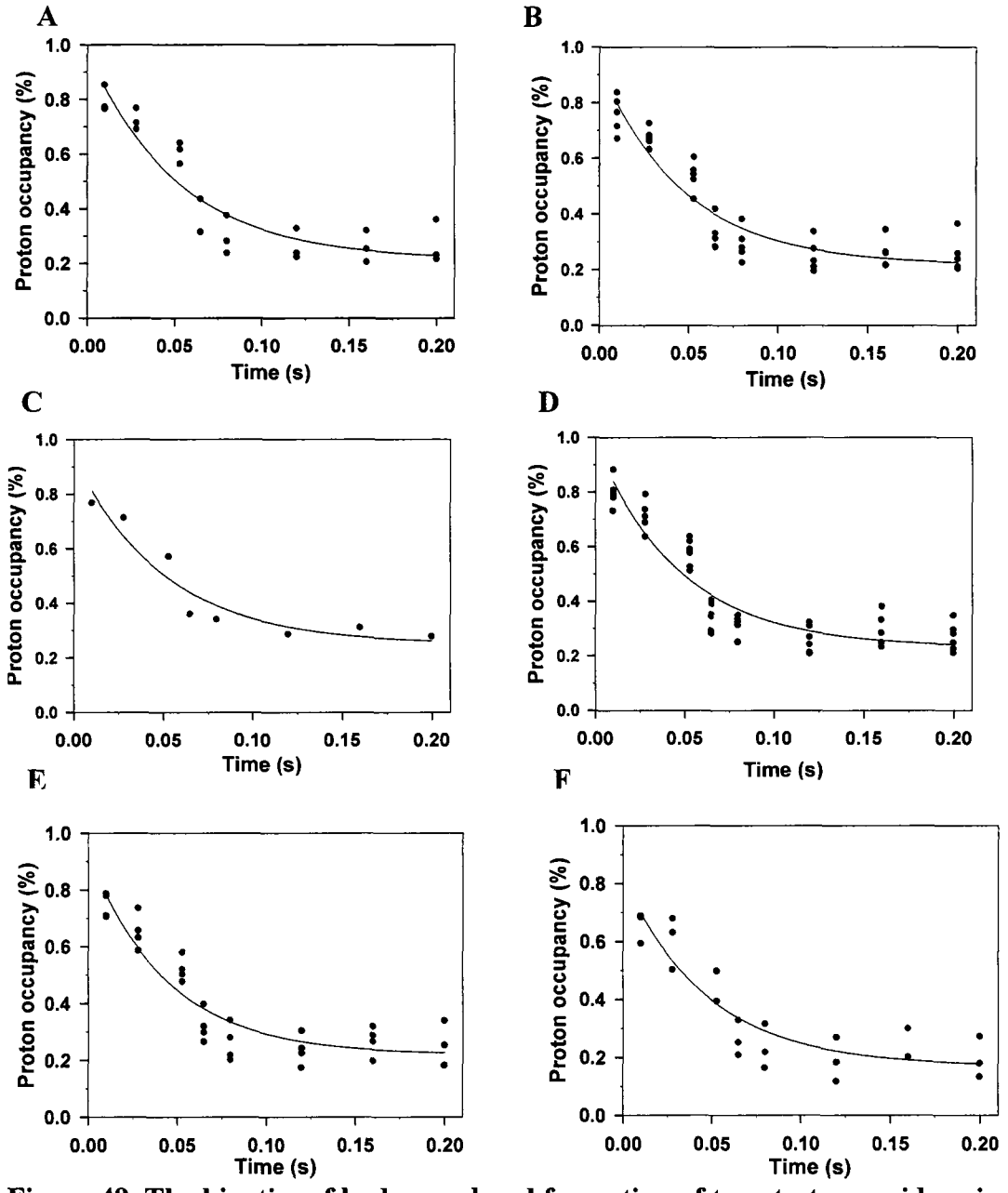


Figure 49. The kinetics of hydrogen bond formation of twenty-two residues in the six helices of Fadd-DD

Panels A, B, C, D, E and F show the proton occupancies of backbone amides from helix 1 (F101, V103 and C105), 2 (L115, R117, Q118, L119 and K120), 3 (I129), 4 (V141, S144, L145, I147, W148 and K149), 5 (V162, G163, A164 and R166) and 6 (A174, V177 and Q182), respectively. This data was fit to a single exponential equation (black curve in plots) to calculate the folding rate using SigmaPlot.

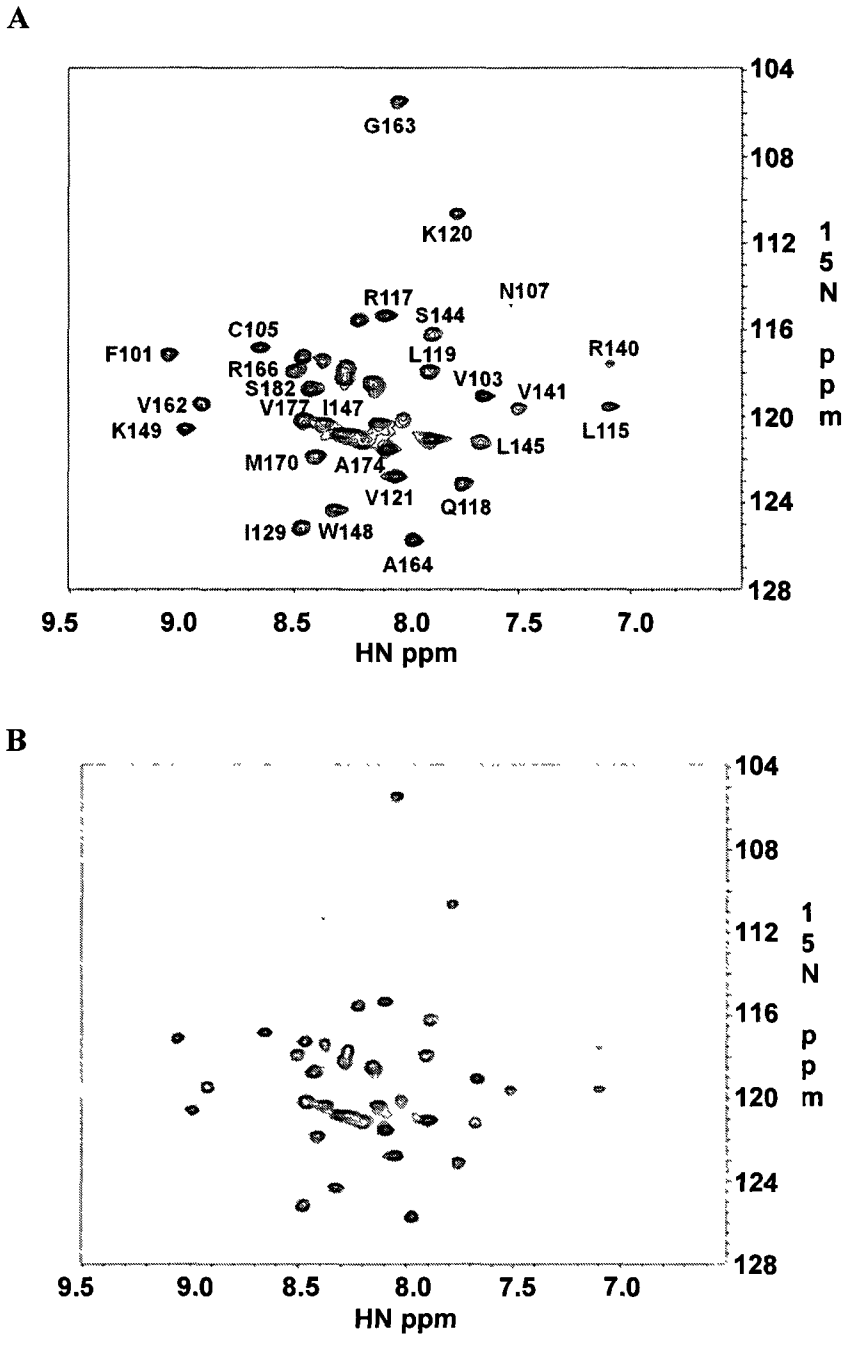
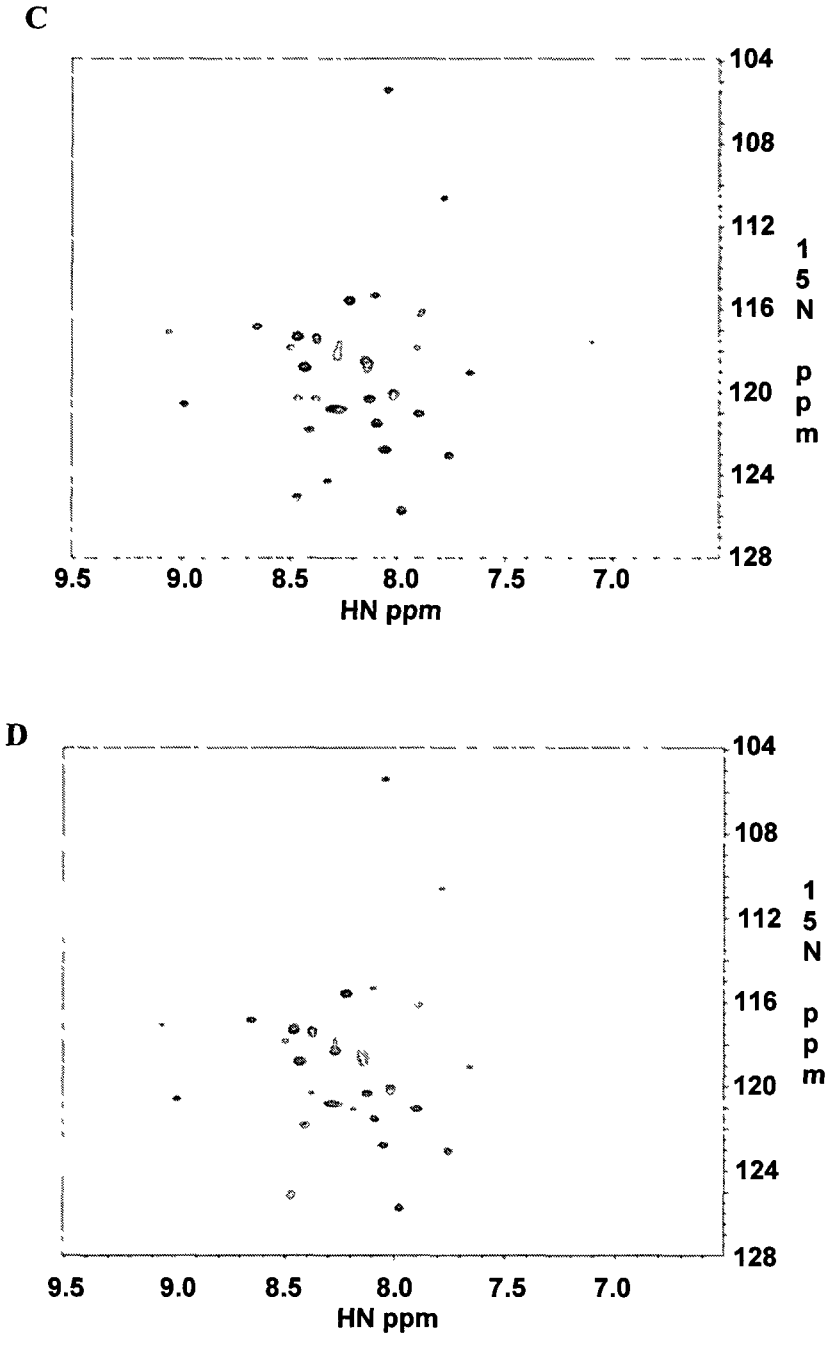


Figure 50. The HSQC spectra of 4 mg/ml Fadd-DD in D_2O after quenched-flow folding and buffer exchange into D_2O buffer (50 mM citric acid/100 mM K_2HPO_4, pD 4.4)

Select refolding times are shown in the following panels: (A) 9.9 ms, (B) 53.2 ms, (C) 80 ms and (D) 200 ms. The hydrogen bond formation of twenty-two residues are followed.

Figure 50 Continued.



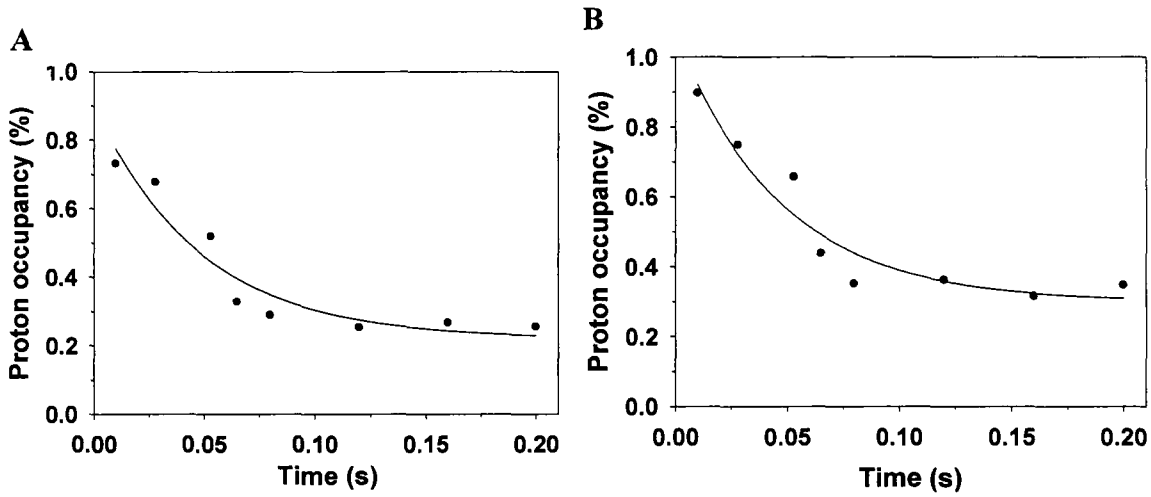


Figure 51. The kinetics of hydrogen bond formation of two loop residues (V121 and M170)

Panels A and B show the proton occupancy of V121 and M170, respectively.

Table 8. The folding rates and the formation of secondary structure and hydrogen bonds for each helix of Fadd-DD

	Rate (s^{-1})	**Amplitude**
Secondary structure formation (Stopped-flow far-UV CD)	23.4 ± 0.4	16.8 ± 0.1
Average Hydrogen bond formation (Quenched-flow/NMR spectroscopy)	20.9 ± 1.7	0.72 ± 0.0
Helix 1	19.1 ± 4.0	0.8 ± 0.1
Helix 2	21.0 ± 3.4	0.7 ± 0.0
Helix 3	19.7 ± 7.0	0.7 ± 0.1
Helix 4	21.0 ± 2.8	0.7 ± 0.0
Helix 5	22.5 ± 4.2	0.7 ± 0.1
Helix 6	20.9 ± 5.4	0.7 ± 0.1

DISCUSSION

In the equilibrium HX study conducted by Jeong et al. with mouse Fadd-DD, thirty-five residues spanning the six helices were shown to have the slowest exchanging protons in an equilibrium HX study (Jeong et al., 1999). Invariably, residues involved in the hydrophobic core formation have the slowest exchanging protons. Among these thirty-five residues, nineteen residues are strongly protected in the human Fadd-DD as well (Figure 41B). The backbone amides with hydrogen bonds in the loops of both mouse and human Fadd-DD, V121 and L170 (M170 in human homologue), are also protected. More backbone amides from mouse Fadd-DD were found to have slower exchange rates than human Fadd-DD. It may be explained that in our study, the protein goes through approximately twelve hours of buffer exchange from the water based buffer to D_2O based buffer. Since there are three free cysteines in Fadd-DD, the protein cannot be lyophilized. Therefore, some of the backbone residues with weak hydrogen bonds may be exchanged during the process whereas the mouse Fadd-DD was lyophilized and directly dissolved in D_2O. The pH also differed and Fadd-DD pH 4.0 was used for the mouse Fadd-DD study.

For human Fadd-DD, only one slowly exchanging amide group was found in helix 3 in comparison to the other helices. This observation is in good agreement with the discovery that helix 3 is more mobile than the other helices (Berglund et al., 2000). The observation that helix 3 has higher relative accessibility compared to the other helices shows this helix is clearly more exposed. In mouse Fadd-DD helix 3 also has the most internal flexibility of all helical elements according to the profile of amide solvent exchange rates (Jeong et al., 1999).

There appears to be a limited number of equilibrium HX studies conducted on proteins with the Greek-key topology. One study is on ribosomal S6 (Haglund et al., 2009), which belongs to the α/β-plait superfamily, two are members of all-β immunoglobulin superfamily, Llama antibody fragment (Perez et al., 2001) and cell-surface receptor protein CD2.D1 (Parker et al., 1997) and another is on the human ephrin-B2 ectodomain, which also possesses the Greek-key topology (Ran et al., 2008). The amide protons in ribosomal S6 located in β1, α1 and β3 have greater protection factors (Haglund et al., 2009). They correspond with helix 1, 2 and 4 in Fadd-DD, respectively (Higman and Greene, 2006). Of these three helices, 2 and 4 have significant protection in Fadd-DD. Similar to our result, the residues on β2 in ribosomal S6 which corresponds to helix 3 in Fadd-DD has the lowest protection factor of all secondary elements (Haglund et al., 2009). The llama antibody fragment appears to have more complex secondary elements consisting of eleven β-strands. It appears that strands 4 and 9, which correspond to helices 2 and 5 in Fadd-DD, have higher protection factors (Perez et al., 2001; Higman and Greene, 2006). Again, helix 2 as well as helix 5 in Fadd-DD are well protected. In the cell surface receptor protein, CD2.D1, protected amides are located throughout the various β-stands as well as in loops (Parker et al., 1997). Interestingly, the three highest protection factors are in β-strand C and the loop between β-strands D and E (Parker et al., 1997). These correspond to helix 3 and the turn between helices 4 and 5 in Fadd-DD (Higman and Greene, 2006). All of these studies indicate that proteins with Greek-key topology have similar stable cores despite different secondary structures.

There is good agreement between folding rates measured by stopped-flow far-UV CD and quenched-flow indicating that they all monitor the same cooperative transition. It

is also similar to the first and dominant phase of folding monitored by stopped-flow fluorescence spectroscopy (22.73 s^{-1}) (Li et al., 2009). While there is only one phase indicated by stopped-flow far-UV CD and quenched-flow and two phases for the study of stopped-flow fluorescence it is not an unprecedented result. There are examples of several proteins which also show similar kinetic behavior. For example, the study conducted on bacteriophage λ lysozyme showed that there are two phases for intrinsic fluorescence and far-UV CD, but only one phase for both HX and NMR as well as HX and mass spectrometry (Di Paolo et al., 2010). Additionally, for human fibroblast growth factor, two phases were detected for stopped-flow fluorescence spectroscopy, one phase for stopped-flow far-UV CD and multiple rates for HX and NMR (Samuel et al., 2001).

In the quenched-flow experiment conducted on phage λ lysozyme, it was suspected that two factors account for the 15-20% lack of protection: 1), some back exchange of deuterium for hydrogen occurred following quenching and before the final buffer exchange into the D_2O buffer; 2), sample preparation required several hours due to the large amount of protein needed for NMR (Di Paolo et al., 2010). We think that these factors also apply to our experiment, which could explain the fact that proton occupancy of all amides does not go to approximately zero. It should be noted that in our study the data were plotted with peak intensity because the standard errors are less than those plotted with peak volume although the time scales of hydrogen bond formation are similar.

Our quenched-flow experiments indicate that all six helices fold concomitantly. The folding behavior of individual amide protons with HX quenched-flow and NMR spectroscopy have been conducted on well over ten proteins. In some proteins, it is

shown that select amide protons fold earlier than the others. For example, in β-lactoglobulin, it was found that during the folding process, the intermediate contains a hydrogen bonded structure in the core of βF, βG, βH and the major α-helix (Kuwata et al., 2001). In hen lysozyme, the fast phase and slow phase of the α-domain both have greater folding rates than those of the β-domain (Lu et al., 1997). Furthermore, in human fibroblast growth factor, among the various β-strands, strands I, IV, IX and X form in similar time scales and appear to provide the basic β-trefoil framework (Samuel et al., 2001). On the other hand, there are cases of other proteins where the formation of hydrogen bonds appears to occur on a similar time scale. For example, in the immunoglobulin binding domain of streptococcal protein G, the twenty-six slowly-exchanging backbone amides fold simultaneously with a single rate of 133 s^{-1} (Kuszewski et al., 1994). In the acyl-CoA binding protein, the measured proton occupancies at different later folding times could be fitted to a single exponential decay with an average rate constant of 20 s^{-1} (Teilum et al., 2000). Pulsed HX study coupled with mass spectrometry showed that the N/C-terminal regions of cytochrome c fold cooperatively on the same timescale (Yang and Smith, 1997). Also in bacteriophage λ lysozyme, all of the kinetics were monophasic and the resulting curves are all virtually identical (Di Paolo et al., 2010). In our study, the folding of Fadd-DD is similar to that of the immunoglobulin binding domain, acyl-CoA binding protein, N/C-terminal regions of cytochrome c and bacteriophage λ lysozyme, indicating significant cooperative hydrogen bond formation in the secondary structure of the protein. It is also concomitant with the hydrophobic collapse. Overall, the studies presented here with Fadd-DD provide insight

into the formation of secondary structure for the all α-helical Greek-key proteins and potentially the all-β and mixed α/β Greek-key proteins more generally.

A model for the folding of Fadd-DD can be constructed by taking into account the results of HX studies and those from stopped-flow fluorescence and φ-value analysis that allows the identification of key interactions in the TS (Steward et al., 2009) (Figure 52). Steward et al. (2009) generated over twenty variants and after the effect of the mutations was examined, select residues on helices 1, 2, 4 and 5 were found to form native interactions in the TS. In our study, all six helices are formed in a cooperative manner. Therefore, while all the helices are forming concomitant with the hydrophobic collapse based on our stopped-flow fluorescence studies, helices 3 and 6 associate through tertiary interactions with canonical core of Greek-key at a later stage of folding.

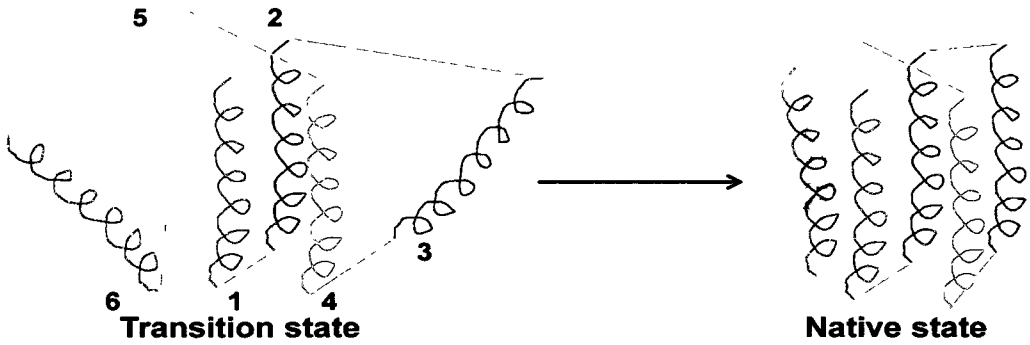


Figure 52. The proposed folding model of Fadd-DD

Helices 1, 2, 3, 4, 5 and 6 are colored in purple, blue, green, yellow, orange and red, respectively.

CHAPTER VII

CONCLUSIONS

ANALYSIS OF CONSERVATION, STRUCTURE, STABILITY AND FOLDING OF WT FADD-DD

Fadd-DD is a six-helical bundle protein with Greek-key topology and belongs to the death domain superfamily. This topology is shared by two other superfamilies, mixed α/β-plait and all-β immunoglobulin. Fadd-DD functions in the signal transduction pathway of apoptosis. A multiple sequence and structure alignment of the death domain superfamily consisting of all four families was constructed and nine significantly conserved hydrophobic and seven moderately conserved hydrophobic positions were identified.

Fadd-DD WT was expressed in *E. coli* BL21(DE3) and purified after using column chromatography. The WT protein is quite stable with ΔG^{D-N} value of 6.60 ± 0.03 kcal/mol. The folding of WT protein is rapid and straightforward. The hydrophobic core formation studied by stopped-flow fluorescence spectroscopy has two phases, one major fast phase with the rate of 22.73 ± 0.57 s^{-1} and one minor slow phase with the rate of 8.13 ± 0.47 s^{-1}. The secondary structure formation studied by stopped-flow far-UV CD spectroscopy has one phase with the rate of 23.4 ± 0.4 s^{-1}.

STRUCTURE, STABILITY AND FOLDING OF FADD-DD VARIANTS

Eight mutants were synthesized by site-directed mutagenesis. Trp112, Leu115 and Val121 are moderately conserved and Trp148 is significantly conserved. His160 and

Val158 are not conserved. Leu172 belongs to a conserved position where most of the residues among the homologues are hydrophilic. The variants were expressed in *E. coli* BL21(DE3) and purified through a series of column chromatography. The $\Delta G^{D\text{-}N}$ values of five variants Trp112Phe, Trp148Phe, Leu115Ala, Val121Ala and Trp112Phe/His160Gly are lower than that of the WT, indicating the replacement of conserved residues with Phe or Ala reduced the stability of the protein. On the other hand, three variants His160Gly, Val158Trp and Leu172Trp are similar to that of the WT. The secondary and tertiary contents of Trp112Phe, Trp148Phe and Trp112Phe/His160Gly significantly decreased compared to the WT, while the secondary and tertiary structures of His160Gly, Val158Trp and Leu172Trp closely resembled the WT. Furthermore, the folding rates of Trp112Phe and Trp112Phe/His160Gly are significantly slower than that of the WT. Instead of two phases, Trp112Phe has one more third long phase. Overall, mutation of the conserved hydrophobic residues had substantial effects on the protein stability, structure and folding.

QUENCHED-FLOW, HX AND NMR STUDIES OF FADD-DD

Quenched-flow, HX and NMR studies give the site specific information of individual hydrogen bond formation at atomic resolution. After deuterated unfolded protein is mixed with refolding buffer, the solution is then mixed with pulsing buffer with high pH. After a certain refolding time, the protected backbone amide deteriums will remain and the unprotected backbone amide deuteriums exchange with protons in the solution. The solution is then mixed with quenching buffer with low pH to stop the labeling process. The final protein is concentrated and buffer exchanged into deuterium buffer. The amide peak intensities are analyzed with 2D NMR spectroscopy (^{1}H,^{15}N-

HSQC). Hydrogen bond formation of twenty-four stable backbone amides are followed. It appears that all the six helices fold on a similar timescale with rates between 19 and 21 s^{-1}, which is similar to the secondary structure formation detected by stopped-flow far-UV CD spectroscopy. Furthermore, exchange rates and protection factors of twenty-three residues were studied by equilibrium HX, and these residues are in the core region.

CHITINASE INSERTION DOMAIN

Four conserved amino acids identified in this study are proposed to be essential for binding with the substrate and they form two distinguishable sequence motifs. The CID may be inserted into the TIM domain to facilitate orienting and binding to longer (e.g.>3) saccharide substrates. Because of the wide distribution in diverse organisms and the high conservation of the CID, we can identify the sequence and predict the structure of this domain in family 18 chitinases in the subfamily A. An evolutionary scheme is presented which places the emergence of the CID in the context of chitinase function; with the addition of the CID leading to an evolutionary shift of the protein from a non-chitinolytic protein, or a NAGase, to a subfamily A or B family 18 chitinase. We also identify a group of conserved hydrophobic residues in the core which we propose are important for folding and structural stability. To test the hypothesis about the role of the CID, a myriad of experimental and computational techniques such as molecular modeling, *in vitro* and *in silico* binding studies coupled to site-directed mutagenesis, enzymatic assays, and crystallization of the holo-protein can be carried out.

FUTURE WORK

Characterize the Tertiary Structure Formation

The timescale of tertiary structure formation of Fadd-DD cannot be obtained with the WT and Val158Trp variant. Another variant will be synthesized to obtain more ellipticity in the near-UV region and the near-UV CD folding will be conducted to study the tertiary structure formation. The two promising locations, Val162 and Val180, will be mutated to tryptophan in order to increase the ellipticity in the near-UV CD spectra.

Solve the Fluorescence Quenching Problem

Equilibrium fluorescence study of the native state of Fadd-DD reveals that the fluorescence intensity at 1 M GndHCl is greater than that at 0 M GndHCl. This result indicates that the fluorescence of the WT is quenched in the native state. One tryptophan (112 or 148) may be quenched by one or more of its surrounding amino acids. Since the Trp112Phe variant and His160Gly also demonstrate the quenching phenomenon, Trp148 is more than likely to be quenched by some other neighboring residues. A variant will be synthesized in order to solve the fluorescence quenching problem. The two promising locations, Phe101 and Lys149 will be mutated to Leu and Asn, respectively.

Show the Early Folding Intermediates of Trp112Phe Variant

The study of quenched-flow, HX and NMR on the WT demonstrates that all helices fold fast and on the same timescale. There is no folding intermediate in the process. The stopped-flow fluorescence study on Trp112Phe indicates that this variant folds more slowly than the WT with one more phase. In order to show if any intermediate exists in the folding process of the Trp112Phe variant, the quenched-flow, HX and NMR

experiments will be conducted to test if all helices still fold cooperatively or helix 2 will fold more slowly than the others.

Study the Effect of Conserved Hydrophobic Residues on Folding and Stability of Another Death Domain Superfamily Member, Pyrin (PDB code: 1UCP)

Buried Tyr60 will be mutated to Trp as the fluorescence probe (Figure 53). L27, equivalent to L119 and L56, equivalent to W148 will be mutated to A. And equilibrium and stopped-flow fluorescence will be performed to compare stability and folding rates between the WT and variants as well as calculate Phi-values.

Figure 53. Pyrin protein (PDB code: 1UCP)

REFERENCES

Agashe, V.R., Shastry, M.C., and Udgaonkar, J.B. (1995). Initial hydrophobic collapse in the folding of barstar. Nature *377*, 754–757.

Akiyama, S., Takahashi, S., Ishimori, K., and Morishima, I. (2000). Stepwise formation of alpha-helices during cytochrome c folding. Nat. Struct. Biol. *7*, 514–520.

Alexandrescu, A.T. (2005). Amyloid accomplices and enforcers. Protein Sci. *14*, 1–12.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Anfinsen, C.B. (1972). The formation and stabilization of protein structure. Biochem. J. *128*, 737–749.

Aronson, N.N. Jr., Halloran, B.A., Alexyev, M.F., Amable, L., Madura, J.D., Pasupulati, L., Worth, C., and van Roey, P. (2003). Family 18 chitinase-oligosaccharide substrate interaction: subsite preference and anomer selectivity of *Serratia marcescens* chitinase A. Biochem. J. *376*, 87–95.

Ashkenazi, A., and Dixit, V.M. (1998). Death receptors: signaling and modulation. Science *281*, 1305–1308.

Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. (1993). Primary structure effects on peptide group hydrogen exchange. Proteins *17*, 75–86.

Bang, S., Jeong, E.J., Kim, I.K., Jung, Y.K., and Kim, K.S. (2000). Fas- and tumor necrosis factor-mediated apoptosis uses the same binding surface of FADD to trigger signal transduction: a typical model for convergent signal transduction. J. Biol. Chem. *275*, 36217–36222.

Bellotti, V., and Chiti, F. (2008). Amyloidogenesis in its biological environment: challenging a fundamental issue in protein misfolding diseases. Curr. Opin. Struct. Biol. *18*, 771–779.

Berglund, H., Olerenshaw, D., Sankar, A., Federwisch, M., McDonald, N.Q., and Driscoll, P.C. (2000). The three-dimensional solution structure and dynamic properties of the human FADD death domain. J. Mol. Biol. *302*, 171–188.

Böhm, G., Muhr, R., and Jaenicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. Protein Eng. *5*, 191–195.

Bortone, K., Monzingo, A.F., Ernst, S., and Robertus, J.D. (2002). The structure of an allosamidin complex with the *Coccidioides immitis* chitinase defines a role for a second acid residue in substrate-assisted mechanism. J. Mol. Biol. *320*, 293–302.

Brameld, K.A., and Goddard, W.A. III (1998). The role of enzyme distortion in the single displacement mechanism of family 19 chitinases. Proc. Natl. Acad. Sci. USA *95*, 4276–4281.

Branden, C., and Tooze, J. (1999). Alpha/beta structures. In Introduction to Protein Structure (2nd ed.), (New York, NY: Garland Publishing), pp. 53–55.

Brurberg, M.B., Eijsink, V.G., and Nes, I.F. (1994). Characterization of a chitinase gene (chiA) from *Serratia marcescens* BJL200 and one-step purification of the gene product. FEMS Microbiol. Lett. *124*, 399–404.

Bussink, A.P., Speijer, D., Aerts, J.M., and Boot, R.G. (2007). Evolution of mammalian chitinase(-like) members of family 18 glycosyl hydrolases. Genetics *177*, 959–970.

Campbell-Valois, F.X., and Michnick, S.W. (2007). The transition state of the ras binding domain of Raf is structurally polarized based on Phi-values but is energetically diffuse. J. Mol. Biol. *365*, 1559–1577.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucl. Acids Res. *37*, D233–238.

Carrington, P.E., Sandu, C., Wei, Y., Hill, J.M., Morisawa, G., Huang, T., Gavathiotis, E., Wei, Y., and Werner, M.H. (2006). The structure of FADD and its mode of interaction with procaspase-8. Mol. Cell *22*, 599–610.

Cavada, B.S., Moreno, F.B., da Rocha, B.A., de Azevedo, W.F. Jr., Castellón, R.E., Goersch, G.V., Nagano, C.S., de Souza, E.P., Nascimento, K.S., Radis-Baptista, G., Delatorre, P., Leroy, Y., Toyama, M.H., Pinto, V.P., Sampaio, A.H., Barettino, D., Debray, H., Calvete, J.J., and Sanz, L. (2006) cDNA cloning and 1.75 A crystal structure determination of PPL2, an endochitinase and N-acetylglucosamine-binding hemagglutinin from *Parkia platycephala* seeds. FEBS J. *273*, 3962–3974.

Chen, Y., and Barkley, M.D. (1998). Toward understanding tryptophan fluorescence in proteins. Biochemistry *37*, 9976–9982.

Chen, Y.R., and Clark, A.C. (2003). Equilibrium and kinetic folding of an alpha-helical Greek key protein domain: caspase recruitment domain (CARD) of RICK. Biochemistry *42*, 6120–6310.

Chen, Y.R., and Clark, A.C. (2004). Kinetic traps in the folding/unfolding of procaspase-1 CARD domain. Protein Sci. *13*, 2196–2206.

Chen, Y.R., and Clark, A.C. (2006). Substitutions of prolines examine their role in kinetic trap formation of the caspase recruitment domain (CARD) of RICK. Protein Sci. *15*, 395–409.

Chiti, F., and Dobson, C.M. (2006). Protein misfolding, functional amyloid, and human disease. Annu. Rev. Biochem.*75*, 333–366.

Clamp, M., Cuff, J., Searle, S.M., and Barton, G.J. (2004). The Jalview Java Alignment Editor. Bioinformatics *20*, 426–427.

Clark, P.L., Liu, Z.P., Zhang, J., and Gierasch, L.M. (1996). Intrinsic tryptophans of CRABPI as probes of structure and folding. Protein Sci. *5*, 1108–1117.

Collaborative Computational Project, No. 4 (1994). Acta Crystallogr. Sect. D *50*, 760–763.

Creighton, T.E. (1993). The folded conformations of globular proteins. In Proteins: Structures and Molecular Properties (2nd ed.), (New York, NY: W. H. Freeman and Company), pp. 255–257.

Cuff, A., Redfern, O.C., Greene, L., Sillitoe, I., Lewis, T., Dibley, M., Reid, A., Pearl, F., Dallman, T., Todd, A., Garratt, R., Thornton, J., and Orengo, C. (2009). The CATH hierarchy revisited-structural divergence in domain superfamilies and the continuity of fold space. Structure *17*, 1051–1062.

Cyr, D.M. (2005). Arrest of CFTRDeltaF508 folding. Nat. Struct. Mol. Biol. *12*, 2–3.

Daggett, V., and Fersht, A.R. (2003). Is there a unifying mechanism for protein folding? Trends Biochem. Sci. *28*, 18–25.

Dahiya, N., Tewari, R., and Hoondal, G.S. (2006). Biotechnological aspects of chitinolytic enzymes: a review. Appl. Microbiol. Biotechnol. *71*, 773–782.

Daimona, T., Katsumaa, S., Iwanagaa, M., Kang, W.K., and Shimada, T. (2005). The BmChi-h gene, a bacterial-type chitinase gene of *Bombyx mori*, encodes a functional exochitinase that plays a role in the chitin degradation during the molting process. Insect Biochem. Mol. Biol. *35*, 1112–1123.

Dalal, K., and Pio, F. (2006). Thermodynamics and stability of the PAAD/DAPIN/PYRIN domain of IFI-16. FEBS Lett. *580*, 3083–3090.

Dalby, P.A., Oliveberg, M., and Fersht, A.R. (1998). Folding intermediates of wild-type and mutants of barnase. I. Use of phi-value analysis and m-values to probe the cooperative nature of the folding pre-equilibrium. J. Mol. Biol. *276*, 625–646.

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) Unraveling hot spots in binding interfaces: progress and challenges. Curr. Opin. Struct. Biol. *12*, 14–20.

Di Paolo, A., Balbeur, D., De Pauw, E., Redfield, C., and Matagne, A. (2010). Rapid collapse into a molten globule is followed by simple two-state kinetics in the folding of lysozyme from bacteriophage λ. Biochemistry *49*, 8646–8657.

Dill, K.A. (1990). Dominant forces in protein folding. Biochemistry *29*, 7133-7155.

Dyson, H.J., and Wright, P.E. (1996). Insights into protein folding from NMR. Annu. Rev. Phys. Chem. *47*, 369–395.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. *32*, 1792–1797.

Engelborghs, Y. (2001). The analysis of time resolved protein fluorescence in multi-tryptophan proteins. Spectrochim. Acta A Mol. Biomol. Spectrosc. *57*, 2255–2270.

Englander, S.W., and Mayne, L. (1992). Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. Annu. Rev. Biophys. Biomol. Struct. *21*, 243–265.

Engman, K.C., Sandberg, A., Leckner, J., and Karlsson, B.G. (2004). Probing the influence on folding behavior of structurally conserved core residues in *P. aeruginosa* apo-azurin. Protein Sci. *13*, 2706–2715.

Esposito, G., Ricagno, S., Corazza, A., Rennella, E., Gumral, D., Mimmi, M.C., Betto, E., Pucillo, C.E.M., Fogolari, F., Viglino, P., Raimondi, S., Giorgetti, S., Bolognesi, B., Merlini, G., Stoppini, M., Bolognesi, M., and Bellotti, V. (2008). The controlling roles of Trp60 and Trp95 in β2-microglobulin function, folding and amyloid aggregation properties. J. Mol. Biol. *378*, 887–897.

Fabian, H., and Naumann, D. (2004). Methods to study protein folding by stopped-flow FT-IR. Methods *34*, 28–40.

Ferrandon, S., Sterzenbach, T., Mersha, F.B., and Xu, M.Q. (2003). A single surface tryptophan in the chitin-binding domain from *Bacillus circulans* chitinase A1 plays a pivotal role in binding chitin and can be modified to create an elutable affinity tag. Biochim. Biophys. Acta *1621*, 31–40.

Fersht, A.R. (1995). Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. Proc. Natl. Acad. Sci. USA *92*, 10869–10873.

Fersht, A.R. (1997). Nucleation mechanism in protein folding. Curr. Opin. Struct. Biol. *7*, 3–9.

Fersht, A.R., and Daggett, V. (2002). Protein folding and unfolding at atomic resolution. Cell *108*, 573–582.

Fersht, A.R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J. Mol. Biol. *224*, 771–782.

Fowler, S.B., and Clark, J. (2001). Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state. Structure *9*, 355–366.

Frieden, C., Hoeltzli, S.D., and Ropson, I.J. (1993). NMR and protein folding: equilibrium and stopped-flow studies. Protein Sci. *2*, 2007–2014.

Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S., and Imanaka, T. (2005). Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. Genome Res. *15*, 352–363.

Funkhouser, J.D., and Aronson, N.N. Jr. (2007). Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. BMC Evol. Biol. *7*, 96–112.

Fusetti, F., Pijning, T., Kalk, K.H., Bos, E., and Dijkstra, B.W. (2003). Crystal structure and carbohydrate-binding properties of the human cartilage glycoprotein-39. J. Biol. Chem. *278*, 37753–37760.

Fusetti, F., von Moeller, H., Houston, D., Rozeboom, H.J., Dijkstra, B.W., Boot, R.G., Aerts, J.M., and van Aalten, D.M. (2002). Structure of human chitotriosidase. Implications for specific inhibitor design and function of mammalian chitinase-like lectins. J. Biol. Chem. *277*, 25537–25544.

Galpern, W.R., and Lang, A.E. (2006). Interface between tauopathies and synucleinopathies: a tale of two proteins. Ann. Neurol. *59*, 449–458.

Garcia, C., Nishimura, C., Cavagnero, S., Dyson, H.J., and Wright, P.E. (2000). Changes in the apomyoglobin folding pathway caused by mutation of the distal histidine residue. Biochemistry *39*, 11227–11237.

Garcia-Mira, M.M., and Schmid, F.X. (2006). Key role of coulombic interactions for the folding transition state of the cold shock protein. J. Mol. Biol. *364*, 458–468.

Garrett, R.H., and Grisham, C.M. (2005). Protein techniques. In Biochemistry (3rd ed.), (Belmont, CA: Thomson Learning), pp. 185-192.

Geierhaas, C.D., Paci, E., Vendruscolo, M., and Clarke, J. (2004). Comparison of the transition states for folding of two Ig-like proteins from different superfamilies. J. Mol. Biol. *343*, 1111–1123.

Gerstein, M. (1992). A resolution-Sensitive procedure for comparing protein surfaces and its application to the comparison of antigen-combining sites. Acta Cryst. A *48*, 271–276.

Go, A., Kim, S., Baum, J., and Hecht, M.H. (2008). Structure and dynamics of *de novo* proteins from a designed superfamily of 4-helix bundles. Protein Sci. *17*, 821–832.

Gooday, G.W. (1990). The ecology of chitin degradation. In Advances in Microbial Ecology, K.C. Marshall ed. (New York, NY: Plenum Press), pp. 387–430.

Greene, L.H., Hamada, D., Eyles, S.J., and Brew, K. (2003). Conserved signature proposed for folding in the lipocalin superfamily. FEBS Lett. *553*, 39–44.

Greene, L.H., and Higman, V.A. (2003). Uncovering network systems within protein structures. J. Mol. Biol. *334*, 781–791.

Greene, L.H., and Higman, V.A. (2005). Conserved networks and the determinants of protein topology. FEBS J. *272* (Suppl. 1), 87–87.

Gromiha, M.M., and Selvaraj, S. (2002). Important amino acid properties for determining the transition state structures of two-state protein mutants. FEBS Lett. *526*, 129–134.

Grossman, T.H., Kawasaki, E.S., Punreddy, S.R., and Osburne, M.S. (1998). Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. Gene *209*, 95–103.

Guda, C., Lu, S., Sheeff, E.D., Bourne, P.E., and Shindyalov, I.N. (2004). CE-MC: a multiple protein structure alignment server. Nucl. Acids Res. *32*, 100–103.

Guda, C., Scheeff, E.D., Bourne, P.E., and Shindyalov, I.N. (2001). A new algorithm for the alignment of multiple protein structures using Monte Carlo optimization. Pac. Symp. Biocomput. 275–286.

Gunasekaran, K., Hagler, A.T., and Gierasch, L.M. (2004). Sequence and structural analysis of cellular retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions. Proteins *54*, 179–194.

Guo, Y., Yu, X., Rihani, K., Wang, Q.Y., and Rong, L. (2004). The role of a conserved acidic residue in calcium-dependent protein folding for a low density lipoprotein (LDL)-

A module: implications in structure and function for the LDL receptor superfamily. J. Biol. Chem. *279*, 16629–16637.

Haglund, E., Lind, J., Oman, T., Ohman, A., Mäler, L., and Oliveberg, M. (2009). The HD-exchange motions of ribosomal protein S6 are insensitive to reversal of the protein-folding pathway. Proc. Natl. Acad. Sci. USA *106*, 21619–21624.

Hamill, S.J., Steward, A., and Clarke, J. (2000). The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. J. Mol. Biol. *297*, 165–178.

Hashimoto, M., Rockenstein, E., Crews, L., and Masliah, E. (2003). Role of protein aggregation in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's diseases. Neuromolecular. Med. *4*, 21–36.

Heidary, D.K., and Jennings, P.A. (2002). Three topologically equivalent core residues affect the transition state ensemble in a protein folding reaction. J. Mol. Biol. *316*, 789–798.

Hennecke, J., Sillen, A., Huber-Wunderlich, M., Engelborghs, Y., and Glockshuber, R. (1997). Quenching of tryptophan fluorescence by the active-site disulfide bridge in the DsbA protein from *Escherichia coli*. Biochemistry *36*, 6391–6400.

Hennig, M., Jansonius, J.N., Terwisscha van Scheltinga, A.C., Dijkstra, B.W., and Schlesier, B. (1995). Crystal structure of concanavalin B at 1.65 Å resolution. An "inactivated" chitinase from seeds of *Canavalia ensiformis*. J. Mol. Biol. *254*, 237–246.

Higman, V.A., and Greene, L.H. (2006). Elucidation of conserved long-range interaction networks in proteins and their significance in determining protein topology. Physica A *368*, 595–606.

Hill, J.M., Morisawa, G., Kim, T., Huang, T., Wei, Y., Wei, Y., and Werner, M.H. (2004). Identification of an expanded binding surface on the FADD death domain responsible for interaction with CD95/Fas. J. Biol. Chem. *279*, 1474–1481.

Hoell, I.A., Dalhus, B., Heggset, E.B., Aspmo, S.I., and Eijsink, V.G.H. (2006). Crystal structure and enzymatic properties of a bacterial family 19 chitinase reveal differences from plant enzymes. FEBS J. *273*, 4889–4900.

Hollis, T., Monzingo, A.F., Bortone, K., Ernst, S., Cox, R., and Robertus, J.D. (2000). The X-ray structure of a chitinase from the pathogenic fungus *Coccidioides immitis*. Protein Sci. *9*, 544–551.

Hooper, N.M. (2005). Roles of proteolysis and lipid rafts in the processing of the amyloid precursor protein and prion protein. Biochem. Soc. Trans. *33*(Pt 2), 335–338.

Horn, S.J., Sørbotten, A., Synstad, B., Sikorski, P., Sørlie, M., Varum, K.M., and Eijsink, V.G. (2006). Endo/exo mechanism and processivity of family 18 chitinases produced by *Serratia marcescens*. FEBS J. *273*, 491–503.

Hsieh, H.C., Kumar, T.K., Sivaraman, T., and Yu, C. (2006). Refolding of a small all beta-sheet protein proceeds with accumulation of kinetic intermediates. Arch. Biochem. Biophys. *447*,147–154.

Hubner, I.A., Oliveberg, M., and Shakhnovich, E.I. (2004). Simulation, experiment and evolution: understanding nucleation in protein S6 folding. Proc. Natl. Acad. Sci. USA *101*, 8354–8359.

Hurtado-Guerrero, R., and van Aalten, D.M.F. (2007). Structure of *Saccharomyces cerevisiae* chitinase 1 and screening-based discovery of potent inhibitors. Chem. Biol. *14*, 589–599.

Imtiyaz, H.Z., Zhang, Y., and Zhang, J. (2005) Structural requirements for signal-induced target binding of FADD determined by functional reconstitution of FADD deficiency. J. Biol. Chem. *280*, 31360–31367.

Itzhaki, L.S., Otzen, D.E., and Fersht, A.R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. J. Mol. Biol. *254*, 260–288.

Jackson, S.E., Masry, N., and Fersht, A.R. (1993). Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: A critical test of the protein engineering method of analysis. Biochemistry *32*, 11270–11278.

Jacobs, M.D., and Fox, R.O. (1994). Staphylococcal nuclease folding intermediate characterized by hydrogen exchange and NMR spectroscopy. Proc. Natl. Acad. Sci. USA *91*, 449–453.

Jäger, M., Dendle, M., and Kelly, J.W. (2009). Sequence determinants of thermodynamic stability in a WW domain--an all-beta-sheet protein. Protein Sci. *18*, 1806–1813.

Jaikaran, E.T., and Clark, A. (2001). Islet amyloid and type 2 diabetes: from molecular misfolding to islet pathophysiology. Biochim. Biophys. Acta *1537*, 179–203.

Jeong, E.J., Bang, S., Lee, T.H., Park, Y.I., Sim, W.S., and Kim, K.S. (1999). The solution structure of FADD death domain. Structural basis of death domain interactions of Fas and FADD. J. Biol. Chem. *274*, 16337–16342.

Jones, B.E., and Matthews, C.R. (1995). Early intermediates in the folding of dihydrofolate reductase from *Escherichia coli* detected by hydrogen exchange and NMR. Protein Sci. *4*, 167–177.

Kampinga, H.H., Hageman, J., Vos, M.J., Kubota, H., Tanguay, R.M., and Bruford, E.A. (2009). Guidelines for the nomenclature of the human heat shock proteins. Cell Stress Chaperones *14*, 105–111.

Kang, T.S., and Kini, R.M. (2009). Structural determinants of protein folding. Cell. Mol. Life Sci. *66*, 2341–2361.

Karplus, K. (2009). SAM-T08: HMM-based protein structure prediction. Nucl. Acids Res. *37* (Suppl. 2), 492–497.

Karplus, M., and Weaver, D.L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. Protein Sci. *3*, 650–668.

Katouno, F., Taguchi, M., Sakurai, K., Uchiyama, T., Nikaidou, N., Nonaka, T., Sugiyama, J., and Watanabe, T. (2004). Importance of exposed aromatic residues in chitinase B from *Serratia marcescens* 2170 for crystalline chitin hydrolysis. J. Biochem. *136*, 163–168.

Kawase, T., Saito, A., Sato, T., Kanai, R., Fujii, T., Nikaidou, N, Miyashita, K., and Watanabe, T. (2004). Distribution and phylogenetic analysis of family 19 chitinases in Actinobacteria. Appl. Environ. Microbiol. *70*, 1135–1144.

Kelly, S.M., Jess, T.J., and Price, N.C. (2005). How to study proteins by circular dichroism. Biochim. Biophys. Acta *1751*, 119–139.

Kelly, S.M., and Price, N.C. (1997). The application of circular dichroism to studies of protein folding and unfolding. Biochim. Biophys. Acta *1338*, 161–185.

Kim, P.S., and Baldwin, R.L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu. Rev. Biochem. *51*, 459-489.

Kloczkowski, A., and Jernigan, R.L.J. (2002). Loop folds in proteins and evolutionary conservation of folding nuclei. Biomol. Struct. Dyn. *20*, 323–325.

Kopito, R.R. (1999). Biosynthesis and degradation of CFTR. Physiol. Rev. *79* (Suppl. 1), 167–173.

Kosinski-Collins, M.S., Flaugh, S.L., and King, J. (2004). Probing folding and fluorescence quenching in human γD crystallin Greek key domains using triple tryptophan mutant proteins. Protein Sci. *13*, 2223–2235.

Kosinski-Collins, M.S., and King, J. (2003). In vitro unfolding, refolding, and polymerization of human γD crystallin, a protein involved in cataract formation. Protein Sci. *12*, 480–490.

Kragelund, B.B., Osmark, P., Neergaard, T.B., Schiødt, J., Kristiansen, K., Knudsen, J., and Poulsen, F.M. (1999). The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. Nat. Struct. Biol. *6*, 594–601.

Krishna, M.M., Hoang, L., Lin, Y., and Englander, S.W. (2004). Hydrogen exchange methods to study protein folding. Methods *34*, 51–64.

Kumar, S., Dudley, J., Nei, M., and Tamura, K. (2008). MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. *9*, 299–306.

Kuranda, M.J., and Robbins, P.W. (1991). Chitinase is required for cell separation during growth of *Saccharomyces cerevisiae*. J. Biol. Chem. *266*, 19758–19767.

Kuszewski, J., Clore, G.M., and Gronenborn, A.M. (1994). Fast folding of a prototypic polypeptide: the immunoglobulin binding domain of streptococcal protein G. Protein Sci. *3*, 1945–1952.

Kuwajima, K. (1996). Stopped-flow circular dichroism. In Circular Dichroism and the Conformational Analysis of Biomolecules, G.D. Fasman, ed. (New York, NY: Plenum Press), pp. 159–178.

Kuwajima, K., and Arai, M. (2000). The molten globule state: the physical picture and biological significance. In Mechanisms of Protein Folding, R.H. Pain, ed. (New York, NY: Oxford University Press), pp. 138–163.

Kuwata, K., Shastry, R., Cheng, H., Hoshino, M., Batt, C.A., Goto, Y., and Roder, H. (2001). Structural and kinetic characterization of early folding events in beta-lactoglobulin. Nat. Struct. Biol. *8*, 151–155.

Lakowicz, J.R. (1999). Principles of Fluorescence Spectroscopy (2nd ed.), (New York, NY: Kluwer Academic), pp. 238.

Lakowicz, J.R., (2006). Principles of Fluorescence Spectroscopy (3rd ed.), (New York, NY: Springer), pp. 529–574, 954.

Lappalainen, I., Hurley, M.G., and Clarke, J. (2008). Plasticity within the obligatory folding nucleus of an immunoglobulin-like domain. J. Mol. Biol. *375*, 547–549.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007). ClustalW and ClustalX version 2. Bioinformatics *23*, 2947–2948.

Larson, S.M., Ruczinski, I., Davidson, A.R., Baker, D., and Plaxco, K.W. (2002). Residues participating in the protein folding nucleus do not exhibit preferential evolutionary conservation. J. Mol. Biol. *316*, 225–233.

Levinthal, C. (1968). Are there pathways for protein folding? J. Chim. Phys. Physico-Chim. Biol. *65,* 44–45.

Li, D.C. (2006). Review of fungal chitinase. Mycopathol. *161,* 345–360.

Li, H., Wojtaszek, J. L., and Greene, L.H. (2009). Analysis of conservation in the Fas-associated death domain protein and the importance of conserved tryptophans in structure, stability and folding. Biochim. Biophys. Acta *1794*, 583–593.

Liu, C., Gaspar, J.A., Wong, H.J., and Meiering, E.M. (2002). Conserved and nonconserved features of the folding pathway of hisactophilin, a beta-trefoil protein. Protein Sci. *11,* 669–679.

Lu, H., Buck, M., Radford, S.E., and Dobson, C.M. (1997). Acceleration of the folding of hen lysozyme by trifluoroethanol. J. Mol. Biol. *265*, 112–117.

Lu, J., and Dahlquist, F.W. (1992). Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. Biochemistry *31*, 4749–4756.

Loewenthal, R., Sancho, J., and Fersht, A.R. (1991). Fluorescence spectrum of barnase: contributions of three tryptophan residues and a histidine-related pH dependence. Biochemistry *30*, 6775–6779.

Maji, S.K., Wang, L., Greenwald, J., and Riek, R. (2009). Structure-activity relationship of amyloid fibrils. FEBS Lett. *583*, 2610–2617.

Marcelino, A.M., Smock, R.G., and Gierasch, L.M. (2006). Evolutionary coupling of structural and functional sequence information in the intracellular lipid-binding protein family. Proteins *63*, 373–384.

Martin, J., and Hart, F.U. (1997). The effect of macromolecular crowding on chaperonin-mediated protein folding. Proc. Natl. Acad. Sci. USA *94*, 1107–1112.

Martinez, J.C., and Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. Nat. Struct. Biol. *6*, 1010–1016.

Maxwell, K.L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M.A., Brown, A.G., Friel, C.T., Hedberg, L., Horng, J.C., Bona, D., Miller, E.J., Vallée-Bélisle, A., Main, E.R., Bemporad, F., Qiu, L., Teilum, K., Vu, N.D., Edwards, A.M., Ruczinski, I., Poulsen, F.M., Kragelund, B.B., Michnick, S.W., Chiti, F., Bai, Y., Hagen, S.J., Serrano, L.,

Oliveberg, M., Raleigh, D.P., Wittung-Stafshede, P., Radford, S.E., Jackson, S.E., Sosnick, T.R., Marqusee, S., Davidson, A.R., and Plaxco, K.W. (2005). Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. Protein Sci. *14*, 602–616.

May, P., Kreuchwig, A., Steinke, T., and Koch, I. (2010). PTGL: a database for secondary structure-based protein topologies. Nucl. Acids Res. *38*, D326–330.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404–405.

Melchers, L.S., Apotheker-de Groot, M., van der Knaap, J.A., Ponstein, A.S., Sela-Buurlage, M.B., Bol, J.F., Cornelissen, B.J., van den Elzen, P.J., and Linthorst, H.J. (1994). A new class of tobacco chitinases homologous to bacterial exo-chitinases displays antifungal activity. Plant J. *5*, 469–480.

Mertens, N., Remaut, E., and Fiers, W. (1995). Tight transcriptional control mechanism ensures stable high-level expression from T7 promoter-based expression plasmids. Biotechnology (NY) *13*, 175–179.

Merzendorfer, H., and Zimoch, L. (2003). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. J. Exp. Biol. *206*, 4393–4412.

Milam, S.L., Nicely, N.I., Feeney, B., Mattos, C., and Clark, A.C. (2007). Rapid folding and unfolding of Apaf-1 CARD. J. Mol. Biol. *369*, 290–304.

Mirny, L.A., and Shakhnovich, E.I. (1999). Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. J. Mol. Biol. *291*, 177–196.

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

Miyata, T., Oda, O., Inagi, R., Iida, Y., Araki, N., Yamada, N., Horiuchi, S., Taniguchi, N., Maeda, K., and Kinoshita, T. (1993). beta 2-Microglobulin modified with advanced glycation end products is a major component of hemodialysis-associated amyloidosis. J. Clin. Invest. *92*, 1243–1252.

Mullins, L.S., Pace, C.N., and Raushel, F.M. (1993). Investigation of ribonuclease T1 folding intermediates by hydrogen-deuterium amide exchange-two-dimensional NMR spectroscopy. Biochemistry *32*, 6152–6156.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536–540.

Nabuurs, S.M., and van Mierlo, C.P. (2010). Interrupted hydrogen/deuterium exchange reveals the stable core of the remarkably helical molten globule of alpha-beta parallel protein flavodoxin. J. Biol. Chem. *285*, 4165–4172.

Nagano, N., Orengo, C.A., and Thornton, J.M. (2002). One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J. Mol. Biol. *321*, 741–765.

Nagano, N., Porter, C.T., and Thornton, J.M. (2001). The $(\beta\alpha)_8$ glycosidases: sequence and structure analyses suggest distant evolutionary relationships. Protein Eng. *14*, 845–855.

Nakamura, T., Mine, S., Hagihara, Y., Ishikawa, K., and Uegaki, K. (2006). Structure of the catalytic domain of the hyperthermophilic chitinase from *Pyrococcus furiosus*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. *63*(Pt 1), 7–11.

Neuweiler, H., Sharpe, T.D., Rutherford, T.J., Johnson, C.M., Allen, M.D., Ferguson, N., and Fersht, A.R. (2009). The folding mechanism of BBL: Plasticity of transition-state structure observed within an ultrafast folding protein family. J. Mol. Biol. *390*, 1060–1073.

Nölting, B. (1999). Analysis of the folding pathway of chymotrypsin inhibitor by correlation of phi-values with inter-residue contacts. J. Theor. Biol. *197*, 113–121.

Nölting, B., and Andert, K. (2000). Mechanism of protein folding. Proteins *41*, 288–298.

Nölting, B., Golbik, R., Neira, J.L., Soler-Gonzalez, A.S., Schreiber, G., and Fersht, A.R. (1997). The folding pathway of a protein at high resolution from microseconds to seconds. Proc. Natl. Acad. Sci. USA *94*, 826–830.

Northey, J.G., Maxwell, K.L., and Davidson, A.R. (2002). Protein folding kinetics beyond the phi value: using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state. J. Mol. Biol. *320*, 389–402.

Nozaki, Y., and Tanford, C. (1971). The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions: establishment of a hydrophobicity scale. J. Biol. Chem. *246*, 2211–2217.

Ohmae, E., Sasaki, Y., and Gekko, K. (2001). Effects of five-tryptophan mutations on structure, stability and function of *Escherichia coli* dihydrofolate reductase. J. Biochem. *130*, 439–447.

Ohnishi, S., and Takano, K. (2004). Amyloid fibrils from the viewpoint of protein folding. Cell Mol. Life Sci. *61*, 511–524.

Oliveberg, M., and Wolynes, P.G. (2005). The experimental survey of protein-folding energy landscapes. Q. Rev. Biophys. *38*, 245–288.

Olofsson, M., Hansson, S., Hedberg, L., Logan, D.T., and Oliveberg, M. (2007). Folding of S6 structures with divergent amino acid composition: pathway flexibility within partly overlapping foldons. J. Mol. Biol. *365*, 237–248.

Onuchic, J.N., and Wolynes, P.G. (2004). Theory of protein folding. Curr. Opin. Struct. Biol. *14,* 70–75.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. (1997). CATH - a hierarchic classification of protein domain structures. Structure *5*, 1093–1108.

Otzon, D.E., and Oliveberg, M. (2002). Conformational plasticity in folding of the split beta-alpha-beta protein S6: evidence for burst-phase disruption of the native state. J. Mol. Biol. *317*, 613–627.

Pace, C.N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. In Methods in Enzymology, C.H.W. Hirs, and S.N.Timasheff, eds. (New York, NY: Academic Press), pp. 266–280.

Pace, C.N., Shirley, B.A., and Thomson, J.A. (1989). Measuring the conformational stability of a protein. In Protein Structure: A Practical Approach. (Oxford, U.K.: Oxford University Press), pp. 311–330.

Palermo, N.Y., Csontos, J., Murphy, R.F., and Lovas, S. (2008). The role of aromatic residues in stabilizing the secondary and tertiary structure of avian pancreatic polypeptide. Int. J. Quantum Chem. *108*, 814–819.

Papanikolau, Y., Prag, G., Tavlas, G., Vorgias, C.E., Oppenheim, A.B., and Petratos, K. (2001). High resolution structural analyses of mutant chitinase A complexes with substrates provide new insight into the mechanism of catalysis. Biochemistry *40,* 11338–11343.

Park, H.H., Lo, Y.C., Lin, S.C., Wang, L., Yang, J.K., and Wu, H. (2007). The death domain superfamily in intracellular signaling of apoptosis and inflammation. Annu. Rev. Immunol. *25*, 561–586.

Parker, M.J., Dempsey, C.E., Lorch, M., and Clarke, A.R. (1997). Acquisition of native beta-strand topology during the rapid collapse phase of protein folding. Biochemistry *36*, 13396–13405.

Payan, F., Flatman, R., Porciero, S., Williamson, G., Juge, N., and Roussel, A. (2003). Structural analysis of xylanase inhibitor protein I (XIP-I), a proteinaceous xylanase inhibitor from wheat (*Triticum aestivum*, var. Soisson). Biochem. J. *372*, 399–405.

Pearce, M.C., Cabrita, L.D., Ellisdon, A.M., and Bottomley, S.P. (2007). The loss of tryptophan 194 in antichymotrypsin lowers the kinetic barrier to misfolding. FEBS J. *274*, 3622–3632.

Pérez, J.M., Renisio, J.G., Prompers, J.J., van Platerink, C.J., Cambillau, C., Darbon, H., and Frenken, L.G. (2001). Thermal unfolding of a llama antibody fragment: a two-state reversible process. Biochemistry *40*, 74–83.

Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A.B., Chet, I., Wilson, K.S., and Vorgias, C.E. (1994). Crystal structure of a bacterial chitinase at 2.3 Å resolution. Structure *2*, 1169–1180.

Pestsko, G., and Ringe, D. (2004). From sequence to structure. In Protein Structure and Function. (London, England: New Science Press), pp. 33–35.

Plaxco, K.W., Simons, K.T., and Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. J. Mol. Biol. *277*, 985–994.

Pollard, T.D., and Earnshaw, W.C. (2004). Programmed cell death. In Cell Biology (1st ed.), (Philadelphia, PA: Saunders), pp. 779–780.

Primrose, W.U. (1993). Samples preparation. In NMR of Macromolecules: A Practical Approach (1st ed.), G.C.K. Roberts, ed. (New York, NY: Oxford University Press), pp. 7–34.

Prusiner, S.B. (1998). Prions. Proc. Natl. Acad. Sci. USA *95*, 13363–13383.

Ptitsyn, O.B. (1995). Structures of folding intermediates. Curr. Opin. Struct. Biol. *5*, 74–78.

Radford, S.E., Dobson, C.M., and Evans, P.A. (1992). The folding of hen lysozyme involves partially structured intermediates and multiple pathways. Nature *358*, 302–307.

Ramalho-Ortigao, J.M., and Traub-Cseko, Y.M. (2003). Molecular characterization of Llchit1, a midgut chitinase cDNA from the leishmaniasis vector *Lutzomyia longipalpis*. Insect Biochem. Mol. Biol. *33*, 279–287.

Ran, X., Qin, H., Liu, J., Fan, J.S., Shi, J., and Song, J. (2008). NMR structure and dynamics of human ephrin-B2 ectodomain: the functionally critical C-D and G-H loops are highly dynamic in solution. Proteins *72*, 1019–1029.

Ranea, J.A.G., Sillero, A., Thornton, J.M., and Orengo, C.A. (2006). Protein superfamily evolution and the Last Universal Common Ancestor (LUCA). J. Mol. Evol. *63*, 513–525.

Rao, F.V., Houston, D.R., Boot, R.G., Aerts, J.M., Sakuda, S., and van Aalten, D.M. (2003). Crystal structures of allosamidin derivatives in complex with human macrophage chitinase. J. Biol. Chem. *278*, 20110–20116.

Rao, P.N., Reddy, K.S., and Bhuyan, A.K. (2009). Amyloid fibrillation of human Apaf-1 CARD. Biochemistry *48*, 7656–7664.

Rao, V., Cui, T., Guan, C., and van Roey, P. (1999). Mutations of endo-beta-N-acetylgluco-saminidase H active site residue Asp130 and Glu132: activities and conformations. Protein Sci. *8*, 2338–2346.

Rao, V., Guan, C., and van Roey, P. (1995). Crystal structure of endo-beta-N-acetylgluco-saminidase H at 1.9 Å resolution: active-site geometry and substrate recognition. Structure *3*, 449–457.

Renkema, G.H., Boot, R.G., Au, F.L., Donker-Koopman, W.E., Strijland, A., Muijsers, A.O., and Hrebicek, M, and Aerts, J.M. (1998). Chitotriosidase, a chitinase, and the 39-kDa human cartilage glycoprotein, a chitin-binding lectin, are homologues of family 18 glycosyl hydrolases secreted by human macrophages. Eur. J. Biochem. *251*, 504–509.

Roder, H., Elöve, G.A., and Englander, S.W. (1988). Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. Nature *335*, 700–704.

Sadeghi, M., Naderi-Manesh, H., Zarrabi, M., and Ranjbar, B. (2005). Effective factors in the thermostability of thermophilic proteins. Biophys. Chem. *119*, 256–270.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425.

Samuel, D., Kumar, T.K., Balamurugan, K., Lin, W.Y., Chin, D.H., and Yu, C. (2001). Structural events during the refolding of an all beta-sheet protein. J. Biol. Chem. *276*, 4134–4141.

Sander, C., and Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins *9*, 56–68.

Sandu, C., Gavathiotis, E., Huang, T., Wegorzewska, I., and Werner, M. (2005). A mechanism for death receptor discrimination by death adaptors. J. Biol. Chem. *280*, 31974–31980.

Sasahara, K., Demura, M., and Nitta, K. (2000). Partially unfolded equilibrium state of hen lysozyme studied by circular dichroism spectroscopy. Biochemistry *39*, 6475–6482.

Sasakawa, H., Tamura, A., Fujimaki, S., Taguchi, S., and Akasaka, K. (1999). Secondary structures and structural fluctuation in a dimeric protein, *Streptomyces subtilisin* inhibitor. J. Biochem. *126*, 859–865.

Schueler-Furman, O., and Baker, D. (2003). Conserved residue clustering and protein structure prediction. Proteins *52*, 225–235.

Schulenburg, C., Löw, C., Weininger, U., Mrestani-Klaus, C., Hofmann, H., Balbach, J., Ulbrich-Hofmann, R., and Arnold, U. (2009). The folding pathway of onconase is directed by a conserved intermediate. Biochemistry *48*, 8449–8457.

Scopes, R.K., and Truscott, K. (1998). Chaperonins from Thermoanaerobacter species. In Methods in Enzymology, G.H. Lorimer and T.O. Baldwin, eds. (New York, NY: Academic Press), pp. 161–171.

Shoemaker, B.A., Wang, J., and Wolynes, P.G. (1999). Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble. J. Mol. Biol. *287*, 675–694.

Shortle, D. (1992). Mutational studies of protein structures and their stabilities. Q. Rev. Biophys. *25*, 205–250.

Sikorski, P., Sørbotten, A., Horn, S.J., Eijsink, V.G., and Vårum, K.M. (2006). *Serratia marcescens* chitinases with tunnel-shaped substrate-binding grooves show endo activity and different degrees of processivity during enzymatic hydrolysis of chitosan. Biochemistry *45*, 9566–9574.

Skovronsky, D.M., Lee, V.M., and Trojanowski, J.Q. (2006). Neurodegenerative diseases: new concepts of pathogenesis and their therapeutic implications. Annu. Rev. Pathol. *1*, 151-170.

Socolich, M., Lockless, S.W., Russ, W.P., Gardner, K.H., and Ranganathan, R. (2005). Evolutionary information for specifying a protein fold. Nature *437*, 512–518.

Songsiriritthigul, C., Pantoom, S., Aguda, A.H., Robinson, R.C., and Suginta, W. (2008). Crystal structures of *Vibrio harveyi* chitinase A complexed with chitooligo-saccharides: Implications for the catalytic mechanism. J. Struct. Biol. *162*, 491–499.

Sosnick, T.R., Dothager, R.S., and Krantz, B.A. (2004). Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. Proc. Natl. Acad. Sci. USA *101*, 17377–17382.

Spehr, V., Frahm, D., and Meyer, T.F. (2000). Improvement of the T7 expression system by the use of T7 lysozyme. Gene *257*, 259–267.

Srivastava, D.B., Ethayathulla, A.S., Kumar, J., Singh, N., Sharma, S., Das, U., Srinivasan, A., and Singh, T.P. (2006). Crystal structure of a secretory signalling Glycol-protein from sheep at 2.0 Å resolution. J. Struct. Biol. *156*, 505–516.

Steward, A., McDowell, G. S., and Clarke, J. (2009). Topology is the principal determinant in the folding of a complex all-alpha Greek key death domain from human FADD. J. Mol. Biol. *389*, 425–437.

Subramaniam, V., Jovin, T.M., and Rivera-Pomar, R.V. (2001). Aromatic amino acids are critical for stability of the bicoid homeodomain. J. Biol. Chem. *276*, 21506–21511.

Suginta, W., Vongsuwan, A., Songsiriritthigul, C., Prinz, H., Estibeiro, P., Duncan, R.R., Svasti, J., and Fothergill-Gilmore, L.A. (2004). An endochitinase A from *Vibrio carchariae*: cloning, expression, mass and sequence analyses, and chitin hydrolysis. Arch. Biochem. Biophys. *424*, 171–180.

Sun, Y.J., Chang, N.C., Hung, S.I., Chang, A.C., Chou, C.C., and Hsiao, C.D. (2001). The crystal structure of a novel mammalian lectin, Ym1, suggests a saccharide binding site. J. Biol. Chem. *276*, 17507–17514.

Suzuki, K., Sugawara, N., Suzuki, M., Uchiyama, T., Katouno, F., Nikaidou, N., and Watanabe, T. (2002). Chitinases A, B, and C1 of *Serratia marcescens* 2170 produced by recombinant *Escherichia coli*: enzymatic properties and synergism on chitin degradation. Biosci. Biotechnol. Biochem. *66*, 1075–1083.

Suzuki, K., Taiyoji, M., Sugawara, N., Nikaidou, N., Henrissat, B., and Watanabe, T. (1999). The third chitinase gene (chiC) of *Serratia marcescens* 2170 and the relationship of its product to other bacterial chitinases. Biochem. J. *343*, 587–596.

Taira, T., Ohnuma, T., Yamagami, T., Aso, Y., Ishiguro, M., and Ishihara, M. (2002). Antifungal activity of rye (*Secale cereale*) seed chitinases: the different binding manner of class I and class II chitinases to the fungal cell walls. Biosci. Biotechnol. Biochem. *66*, 970–977.

Teilum, K., Kragelund, B.B., Knudsen, J., and Poulsen, F.M. (2000). Formation of hydrogen bonds precedes the rate-limiting formation of persistent structure in the folding of ACBP. J. Mol. Biol. *301*, 1307–1314.

Terwisscha van Scheltinga, A.C., Hennig, M., and Dijkstra, B.W. (1996). The 1.8 Å resolution structure of hevamine, a plant chitinase/lysozyme, and analysis of the conserved sequence and structure motifs of glycosyl hydrolase family 18. J. Mol. Biol. *262*, 243–257.

Terwisscha van Scheltinga, A.C., Kalk, K.H., Beintema, J.J., and Dijkstra, B.W. (1994). Crystal structures of hevamine, a plant defence protein with chitinase and lysozyme activity, and its complex with an inhibitor. Structure *2*, 1181–1189.

Ting, K.L., and Jernigan, R.L. (2002). Identifying a folding nucleus for the lysozyme/alpha-lactalbumin family from sequence conservation clusters. J. Mol. Evol. *54*, 425–436.

Todd, A.E., Orengo, C.A., and Thornton, J.M. (1999). Evolution of protein function, from a structural perspective. Curr. Opin. Chem. Biol. *3*, 548–556.

Todd, A.E., Orengo, C.A., and Thornton, J.M. (2001). Evolution of function in protein superfamilies, from a structural perspective. J. Mol. Biol. *307*, 1113–1143.

Tseng, Y.Y., and Liang, J. (2004). Are residues in a protein folding nucleus evolutionarily conserved? J. Mol. Biol. *335*, 869–880.

Uchiyama, T., Katouno, F., Nikaidou, N., Nonaka, T., Sugiyama, J., and Watanabe, T. (2001). Roles of the exposed aromatic residues in crystalline chitin hydrolysis by chitinase A from *Serratia marcescens* 2170. J. Biol. Chem. *276*, 41343–41349.

Udgaonkar, J.B., and Baldwin, R.L. (1988). NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. Nature *335*, 694–699.

Vaaje-Kolstad, G., Vasella, A., Peter, M.G., Netter, C., Houston, D.R., Westereng, B., Synstad, B., Eijsink, V.G., and van Aalten, D.M. (2004). Interactions of a family 18 chitinase with the designed inhibitor Hm508 and its degradation product, chitobiono-δ-lactone. J. Biol. Chem. *279*, 3612–3619.

Van Aalten, D.M., Komander, D., Synstad, B., Gåseidnes, S., Peter, M.G., and Eijsink, V. G. (2001). Structural insights into the catalytic mechanism of a family 18 exo-chitinase. Proc. Natl. Acad. Sci. USA *98*, 8979–8984.

Van Aalten, D.M., Synstad, B., Brurberg, M.B., Hough, E., Riise, B.W., Eijsink, V.G., and Wierenga, R.K. (2000). Structure of a two-domain chitotriosidase from *Serratia marcescens* at 1.9 Å resolution. Proc. Natl. Acad. Sci. USA *97*, 5842–5847.

Van den Berg, B., Ellis, R.J., and Dobson, C.M. (1999). Effects of macromolecular crowding on protein folding and aggregation. EMBO J. *18*, 6927–6933.

Van Eijk, M., van Roomen, C.P.A.A., Renkema, G.H., Bussink, A.P., Andrews, L., Blommaart, E.F., Sugar, A., Verhoeven, A.J., Boot, R.G., and Aerts, J.M. (2005). Characterization of human phagocyte-derived chitotriosidase, a component of innate immunity. Int. Immunol. *17*, 1505–1512.

Varela, P.F., Llera, A.S., Mariuzza, R.A., and Tormo, J. (2002). Crystal structure of imaginal disc growth factor-2. A member of a new family of growth-promoting glycol-proteins from *Drosophila melanogaster*. J. Biol. Chem. *277*, 13229–13236.

Venyaminov, S.Y., and Yang, J.T. (1996). Determination of protein secondary structure. In Circular Dichroism and the Conformational Analysis of Biomolecules, G.D. Fasman, ed. (New York, NY: Plenum Press), pp. 69–104.

Volles, M.J., and Lansbury, P.T. Jr. (2003). Zeroing in on the pathogenic form of alpha-synuclein and its mechanism of neurotoxicity in Parkinson's disease. Biochemistry *42*, 7871–7878.

Waddling, C.A., Plummer, T.H. Jr., Tarentino, A.L., and van Roey, P. (2000). Structural basis for the substrate specificity of endo-beta-N-acetylglucosaminidase F(3). Biochemistry *39*, 7878–7885.

Walker, F.O. (2007). Huntington's disease. Lancet *369*, 218–228.

Watanabe, F.R., Brannan, C.I., Copeland, N.G., Jenkins, N.A., and Nagata, S. (1992). Lymphoproliferation disorder in mice explained by defects in Fas antigen that mediates apoptosis. Nature *356*, 314–317.

Watanabe, T., Ariga, Y., Sato, U., Toratani, T., Hashimoto, M., Nikaidou, N, Kezuka, Y., Nonaka, T., and Sugiyama, J. (2003). Aromatic residues within the substrate-binding cleft of *Bacillus circulans* chitinase A1 are essential for hydrolysis of crystalline chitin. Biochem. J. *376*, 237–244.

Watanabe, T., Kobori, K., Miyashita, K., Fujii, T., Sakai, H., Uchida, M., and Tanaka, H. (1993). Identification of glutamic acid 204 and aspartic acid 200 in chitinase A1 of *Bacillus circulans* WL-12 as essential residues for chitinase activity. J. Biol. Chem. *268*, 18567–18572.

Wensley, B.G., Gärtner, M., Choo, W.X., Batey, S., and Clarke, J. (2009). Different members of a simple three-helix bundle protein family have very different folding rate constants and fold by different mechanisms. J. Mol. Biol. *390*, 1074–1085.

Wetlaufer, D.B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. Proc. Natl. Acad. Sci. USA *70*, 697–701.

Whitmore, L., and Wallace, B.A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers *89*, 392–400.

Wilson, C.J., and Wittung-Stafshede, P. (2005). Role of structural determinants in folding of the sandwich-like protein *Pseudomonas aeruginosa* azurin. Proc. Nat. Acad. Sci. USA *102*, 3984–3987.

Woody, R.W. (1996). Theory of circular dichroism of proteins. In Circular Dichroism and the Conformational Analysis of Biomolecules, G.D. Fasman, ed. (New York, NY: Plenum Press), pp. 25–61.

Woody, R.W., and Dunker, A.K. (1996). Aromatic and cysteine side-chain circular dichroism in proteins. In Circular Dichroism and the Conformational Analysis of Biomolecules, G.D. Fasman, ed. (New York, NY: Plenum Press), pp. 109–144.

Xiao, S., Bi, Y., Shan, B., and Raleigh, D.P. (2009). Analysis of core packing in a cooperatively folded miniature protein: the ultrafast folding villin headpiece helical subdomain. Biochemistry *48*, 4607–4616.

Yang, H., and Smith, D.L. (1997). Kinetics of cytochrome c folding examined by hydrogen exchange and mass spectrometry. Biochemistry *36*, 14992–14999.

Zarrine-Afsar, A., and Davidson, A.R. (2004). The analysis of protein folding kinetic data produced in the protein engineering experiments. Methods *34*, 41–50.

Zarrine-Afsar, A., Larson, S.M., and Davidson, A.R. (2005). The family feud: do proteins with similar structures fold via the same pathway? Curr. Opin. Struct. Biol. *15*, 42–49.

Zeeb, M., and Balbach, J. (2004). Protein folding studied by real-time NMR spectroscopy. Methods *34*, 65–74.

Zees, A.C., Pyrpassopoulos, S., and Vorgias, C.E. (2009). Insights into the role of the (alpha+beta) insertion in the TIM-barrel catalytic domain, regarding the stability and the enzymatic activity of Chitinase A from *Serratia marcescens*. Biochim. Biophys. Acta *1794*, 23–31.

Zhou, R., Eleftheriou, M., Royyuru, A.K., and Berne, B.J. (2007). Destruction of long-range interactions by a single mutation in lysozyme. Proc. Natl. Acad. Sci. USA *104*, 5824–5829.

Zwanzig, R., Szabo, A., and Bagchi, B. (1992). Levinthal's paradox. Proc. Natl. Acad. Sci. USA *89*, 20–22.

APPENDIX I

METHODS FOR CLONING, EXPRESSION, AND PURIFICATION OF FADD-DD

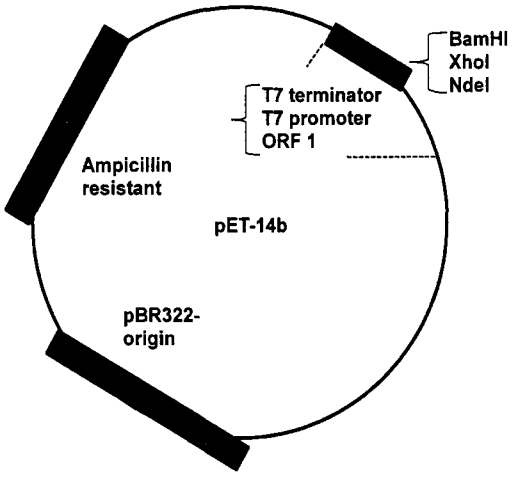


Figure A1. The diagram of pET-14b map showing the cloning site (adapted from http://www.biovisualtech.com/bvplasmid/pET-14b.htm)

Table A1. Components (in µl) of the PCR reactions to amplify the cDNA

Component	Rxn 1	Rxn 2	Rxn 3	Control
Sterile ddH_2O	36	35	33	37
10 X *Pfu* reaction buffer	5	5	5	5
Deoxyribonucleotide (dNTPs) (12.5 mM)	0.8	0.8	0.8	0.8
DNA template	1	2	4	1
Promoter primer (30 ng/µl)	3.1	3.1	3.1	3.1
Terminator primer (30 ng/µl)	3.1	3.1	3.1	3.1
Pfu hotstart polymerase (2.5 U/µl)	1	1	1	0
Total volume	50	50	50	50

Table A2. Temperature and cycles of the PCR reaction to amplify the cDNA

Segment	Number of cycles	Temperature	Duration	Period
1	1	95°C	2.5 minutes	Initial set up
2	30	95°C	30 seconds	Denaturation
		52°C	30 seconds	Annealing
		72°C	1 minute	Elongation
3	1	72°C	10 minutes	Final elongation
4	1	4°C	Until use	Hold

Bgl II **T7 promoter**

AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAG

AAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCAT

Met Gly Ser Ser His His His His His

TCS *Nde* I *Xho* I *BamH* I

CACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTCGAGGATCCGGCTGCTAACAAAGC

His Ser Ser Gly Leu Val Pro Arg Gly Ser His Met Leu Glu Asp Pro Ala Ala Asn Lys Ala

CCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGG

Arg Lys Glu Ala Glu Leu Ala Ala Ala Thr Ala Glu Gln End **T7 terminator**

CCTCTAAACGGGTCTTGAGGGGT TTTTTG

Figure A2. The cloning/expression region of pET-14b showing T7 promoter, His-tag sequence, thrombin cleavage site (TCS), three cloning sites (*Nde* I, *Xho* I and *Bam*H I), and T7 terminator

Table A3. Restriction enzymes used to digest the plasmid and insert

Restriction enzyme	Recognition site	NEB buffer	Bovine serum albumin
*Bam*H I	5'...GGATCC...3'	3	Y
Nde I	5'...CATATG...3'	4	N
Xho I	5'...CTCGAG...3'	2	Y

Table A4. Digestion mixture (concentration is empirically determined)

	Plasmid (µl)	cDNA (µl)
DNA	52	31
Buffer	6	4
Enzyme	2	2
Water	0	3
Total	60	40

Table A5. Ligation mixture (in µl) (concentration is empirically determined)

	Xho I	Xho I	BamH I	BamH I	Xho I	BamH I	Xho I	BamH I
Ratio	1:3	1:6	1:3	1:6	-	-	-	-
plasmid	4	4	4	4	4	4	4	4
Buffer	1	1	1	1	1	1	1	1
cDNA	2	4	1.5	3	0	0	0	0
ATP	1	1	1	1	1	1	1	1
Water	1.5	0	2	0.5	3.5	3.5	4	4
Ligase	0.5	0.5	0.5	0.5	0.5	0.5	0	0
Total	10	10.5	10	10	10	10	10	10

"Xho I" means the cDNA and plasmid were digested by both *Nde* I and *Xho* I; and "BamH I" means the cDNA and plasmid were digested by both *Nde* I and *BamH* I.

Table A6. The sequences of primers for the mutagenesis

Primer Name	Primer Sequence (5' to 3')
Trp112Phe	5'-atctgtgataatgtggggaaagatttcagaaggctggctc-3'
Trp112Phe_antisense	5'-gagccagccttctgaaatctttccccacattatcacagat-3'
Leu115Ala	5'-ggaaagattggagaagggcggctcgtcagctcaaag-3'
Leu115Ala_antisense	5'-ctttgagctgacgagccgcccttctccaatctttcc-3'
Val121Ala	5'-gctcgtcagctcaaagcctcagacaccaagatc-3'
Val121Ala_antisense	5'-gatcttggtgtctgaggctttgagctgacgagc-3'
Trp148Phe	5'-cgggagtcactgagaatcttcaagaacacagagaaggaga-3'
Trp148Phe_antisense	5'-tctccttctctgtgttcttgaagattctcagtgactcccg-3'
Val158Trp	5'-acagagaaggagaacgcaacatgggcccacctgg-3'
Val158Trp_antisense	5'-ccaggtgggcccatgttgcgttctccttctctgt-3'
His160Gly	5'-cgcaacagtggccggcctggtgggggct-3'
His160Gly_antisense	5'-agcccccaccaggccggccactgttgcg-3'
Leu172Trp	5'-gtcctgccagatgaactgggtggctgacctggta-3'
Leu172Trp_antisense	5'-taccaggtcagccacccagttcatctggcaggac-3'

APPENDIX II

BIOINFORMATICS STUDY OF FADD-DD AND CID

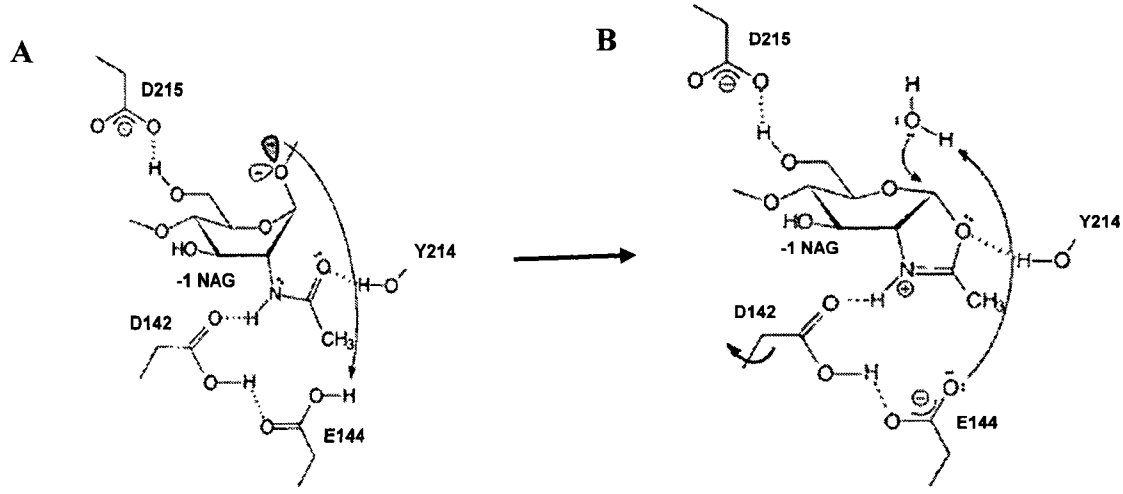


Figure A3. Proposed catalytic mechanism showing Asp142 and Glu144 in *S. marcescens* chitinase B, during two stages of catalysis (adapted from van Aalten et al., 2001)

(A) Binding of substrate -1 NAG residue causes distortion of the pyranose ring to a boat or skewed boat conformation and rotation of Asp142 toward Glu144, enabling hydrogen bond formation between the acetamido group, Asp142, and Glu144.

(B) Hydrolysis of the oxazolinium ion intermediate results in protonation of Glu144 and rotation of Asp142 to its original position.

Table A7. List of the sequence names, species name, and GI numbers of thirty-three more CID sequences that are included in the phylogenetic tree and the larger multiple sequence alignment

Sequence name	Species name	GI number
B_LysobacterE	*Lysobacter enzymogenes*	50429005
B_Stenotrophomonas	*Stenotrophomonas sp.* SKA14	254524159
B_HerpetosiphonA	*Herpetosiphon aurantiacus* ATCC 23779	159899269
B_CellulomonasU	*Cellulomonas uda*	17865808
B_DoohwaniellaC	*Doohwaniella chitinasigens*	6649589
B_ChromobacteriumV	*Chromobacterium violaceum* ATCC 12472	34499695
B_StreptosporangiumR	*Streptosporangium roseum* DSM 43021	229855168
B_KribbellaF	*Kribbella flavida* DSM 17836	227377795
B_StackebrandtiaN	*Stackebrandtia nassauensis* DSM 44728	229863927
B_SaccharopolysporaE	*Saccharopolyspora erythraea* NRRL 2338	134102989
B_NocardiopsisD	*Nocardiopsis dassonvillei* subsp. dassonvillei DSM 4311	229205033
B_CatenulisporaA	*Catenulispora acidiphila DSM 44928*	256395265
B_Synechococcus	*Synechococcus sp.* RCC307	148242001
F_PenicilliumM	*Penicillium marneffei* ATCC 18224	212534216
F_GrifolaU	*Grifola umbellata*	28436151
F_HypocreaS	*Hypocrea seppoi*	220701877
F_BionectriaO	*Bionectria ochroleuca*	88696577
F_PyrenophoraT	*Pyrenophora tritici-repentis* Pt-1C-BFP	189188560
F_RhizopusM	*Rhizopus microsporus* var. oligosporus	1565203
F_NeosartoryaF	*Neosartorya fischeri* NRRL 181	119470878
EE_EntamoebaH	*Entamoeba histolytica* HM-1:IMSS	67472835
EA_CaenorhabditisE	*Caenorhabditis elegans*	17551250
EA_DrosophilaM	*Drosophila melanogaster*	45550474
EA_CionaI	*Ciona intestinalis*	167830427
EA_StrongylocentrotusP	*Strongylocentrotus purpuratus*	115608306
V_XenopusT	*Xenopus (Silurana) tropicalis*	4262194
V_GallusG	*Gallus gallus*	45383307
V_DanioR	*Danio rerio*	41055329
M_MusM	*Mus musculus*	12597291
M_BosT	*Bos taurus*	27807261
M_EquusC	*Equus caballus*	219689080
M_RattusN	*Rattus norvegicus*	119120779
M_CapraH	*Capra hircus*	66361429
Ar_ThermococcusK	*T. kodakarensis* KOD1	57641700
Ar_HalogeometricumB	*Halogeometricum borinquense* DSM 11551	227882613
Ar_HalomicrobiumM	*Halomicrobium mukohataei* DSM 12286	257388962

Table A7 Continued.

Sequence name	Species name	GI number
F_CoccidioidesI	*C. Immitis*	1D2K
F_AspergillusF	*A. fumigatus*	70985392
F_TrichodermaA	*Trichoderma atroviride*	71143448
F_CandidaA	*C. albicans* SC5314	68466729
F_SaccharomycesC	*S. cerevisiae*	6320579
B_BacillusC	*B. circulans*	1ITX
B_StreptomycesT	*Streptomyces thermoviolaceus*	436784
B_ClostridiumP	*Clostridium paraputrificum*	2696017
B_HahellaC	*Hahella chejuensis* KCTC 2396	83644516
B_SerratiaM	*S. marcescens*	1E15
M_HomoS	*Homo sapiens*	1LG1
EA_PenaeusM	*Penaeus monodon*	5114426
EA_AcanthocheilonemaV	*Acanthocheilonema viteae*	804649
EA_LutzomyiaL	*Lutzomyia longipalpis*	28863959
EA_DermatophagoidesP	*Dermatophagoides pteronyssinus*	78128018
EE_HydractiniaE	*Hydractinia echinata*	46016169
EE_DictyosteliumD	*Dictyostelium discoideum* AX4	66818433
P_NicotianaT	*Nicotiana tabacum*	899342
P_RobiniaP	*Robinia pseudoacacia*	119721188
P_MomordicaC	*Momordica charantia*	20269861
P_OryzaS	*Oryza sativa*	115485441
P_ArabidopsisT	*Arabidopsis thaliana*	22328814

Table A8. List of twenty-one structures of family 18 chitinases and chitinase-like proteins from plants, bacteria, fungi, and animals

PDB Code	Species	Name/ Kingdom	With CID?	Other domain	Ligand	Function
1HVQ	*H. brasiliensis*	Para rubber tree	No		3 NAG	Hevamine A (endochitinase/lysozyme)
1TA3	*Triticum aestivum*	Wheat	No		NAG, EDO (scatter)	Xylanase inhibitor protein
1CNV	*Canavalia ensiformis*	Jack bean	No			Concanavalin B, seed storage protein
1NAR	*Vicia narbonensis*	Purple broad vetch	No			Narbonin, seed storage protein
1EOM	*E. meningoseptica*	Bacterium	No		3 NAG, 3 MAN, 2 GAL	NAGase
1EDT	*S. plicatus*	Bacterium	No			NAGase
2EBN	*E. meningoseptica*	Bacterium	No			NAGase
1ITX	*B. circulans*	Bacterium	Yes			Chitinase A1
1FFR/1CTN	*S. marcescens*	Bacterium	Yes	N-terminal	7 NAG	Chitinase A
1UR9/1E15	*S. marcescens*	Bacterium	Yes	C-terminal	NAG, GDL, PHJ	Chitinase B
1KFW	*Arthrobacter* sp.	Bacterium	Yes			Psychrophilic chitinase B
3B9A	*V. harveyi*	Bacterium	Yes	N-terminal	6 NAG	Chitinase A
1D2K	*C. immitis*	Fungus	Yes			Chitinase
1WNO/1W9P	*A. fumigatus*	Fungus	Yes		NAG, NDG	Chitinase B
1LG1/1HKM	*H. sapiens*	Human	Yes		2 NAA, 1 ALI	Chitotriosidase
3FXY	*H. sapiens*	Human	Yes			Acidic Mammalian Chitinase
1E9L	*Mus musculus*	Mouse	Yes		NAG	Mammalian lectin
1NWT	*H. sapiens*	Human	Yes		8 NAG	Cartilage gp39
1JND	*Drosophila melanogaster*	Fruit fly	Yes		NAG, MAN	Disc growth factor-2
2DPE	*Ovis aries*	Sheep	Yes		2 NAG, 3 MAN	Signal processing protein
1LJY	*Capra hircus*	Goat	Yes		NAG	Mammary gland protein

Structures in bold are described and compared in the text.

Table A9. The twenty proteins that are utilized to make sequence and structure alignment of the death domain superfamily

Family	Sequence number	GI number	Name of organism	Name
CARD	1	6137576	Human	**H_1CY5**
	2	76156998	Trematode	T_SJ_6
	3	21326823	Carp fish	Fi_CARD
	4	57163901	Cat	C_caspase
	5	6680696	Mouse	M_IAP
DED	6	4139474	Human	**H_1A1W**
	7	56540946	Frog	Fr_LOC_5
	8	13095651	Bovine herpesvirus	BHV_FLIP
	9	9628077	Equid herpesvirus	EHV_ORF
	10	40389470	Mouse	M_vanishin
DD	11	76155247	Trematode	T_SJ_5
	12	45383358	Junglefowl	Fo_TNFRSF
	13	11513952	Human	**H_1E3Y**
	14	10732781	Fruit fly	Fl_DD
	15	80476543	Frog	Fr_LOC_7
PYD	16	18203023	Zebra fish	Fi_PYD
	17	38493024	Human	**H_1UCP**
	18	37791117	Dwarf Lemur	L_cryopyrin
	19	24660214	Human	H_NLRP
	20	18088559	Human	H_pyhin

Each family has one known structure indicated in bold and four unknown structures. Sequence are chosen from diversified organisms and have identity percentage less than 25% in order to find those residues conserved for structure and folding.

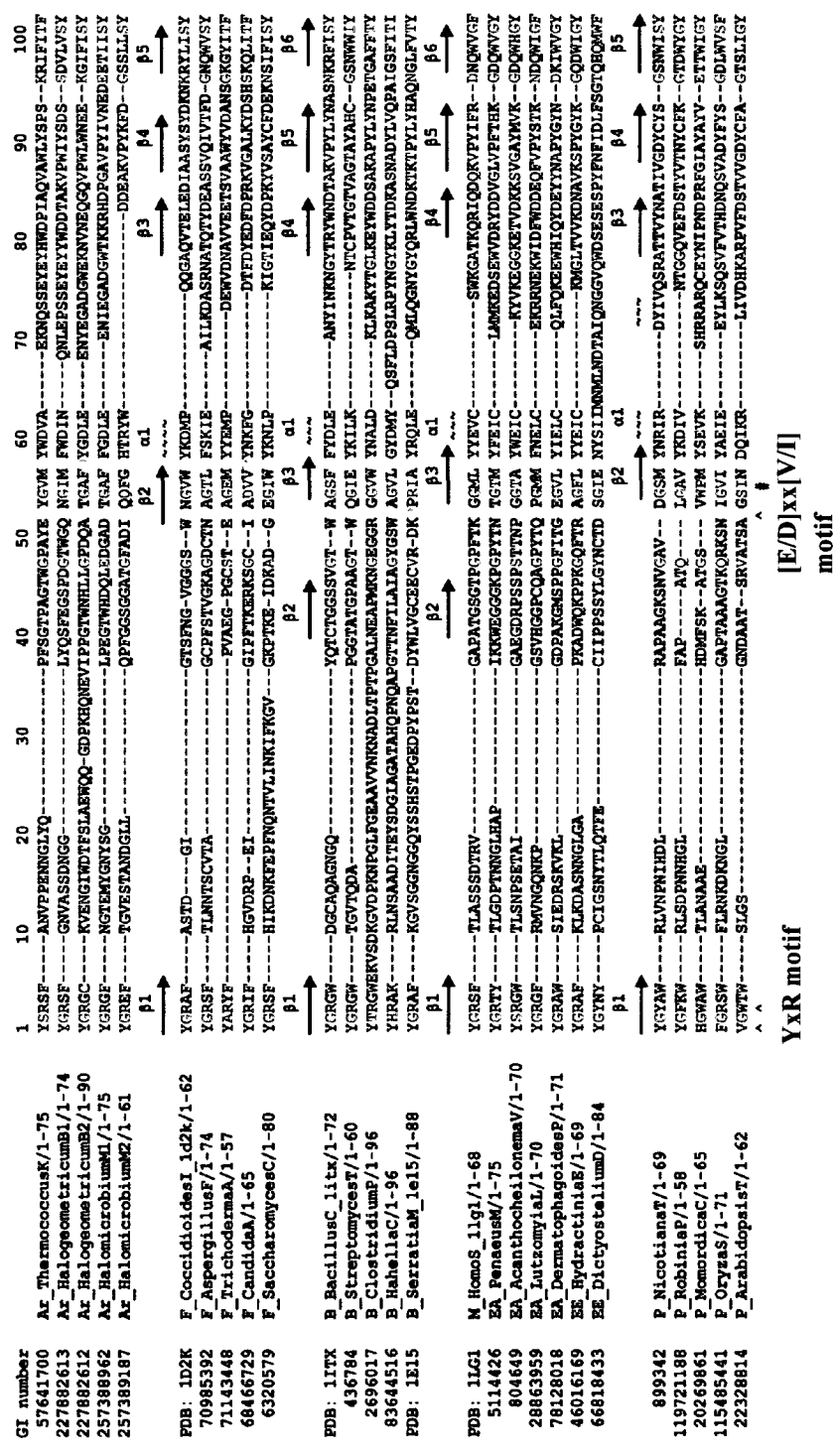


Figure A4. Structure-based multiple sequence alignment of the CID

Hydrophobic positions with high conservation ($C(i)\geq 0.45$) are coloured in blue and positions with moderate conservation ($0.35\leq C(i)<0.45$) are coloured in light blue.

Figure A4 Continued.

Hydrophilic positions with high conservation are coloured in red and positions with moderate conservation are coloured in pink. Neutral positions with high conservation containing mostly glycine, alanine, or proline are coloured in orange, while positions with moderate conservation are not highlighted. "~" and "→" indicate the sequences in α-helices and β-strands, respectively. The secondary structure of tobacco chitinase CID was predicted by the program of PSIPRED. "^" and "#" represent the positions which form hydrogen bonding and the hydrophobic interaction with the substrate, respectively. The full genus name and the first letter of species name are shown for each organism in the figure. If two sequences are from one species, a number is added after the species name. All the sequences were obtained from the protein database at the NCBI. Abbreviations: Ar, Archaea; B, Bacteria; F, Fungi; P, Plantae; EE, early eukaryotes; EA, early Animalia; M, mammal.

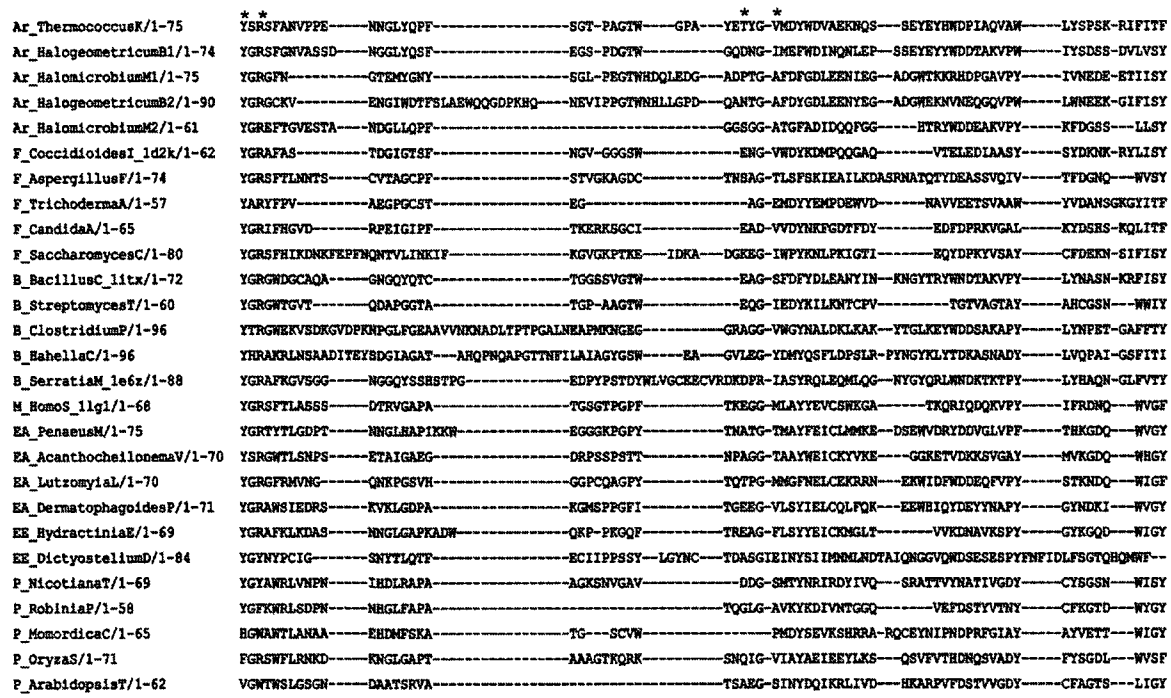


Figure A5. The larger multiple sequence alignment of sixty CID sequences

The alignment was generated by MUSCLE in Jalview.

Figure A5 Continued.

It is not edited according to the three model structures. The two conserved motifs YxR and [E/D]xx[V/I] are highlighted in the frames and the four conserved positions are labeled with the asterisks. The species names and GI numbers refer to Table 7.

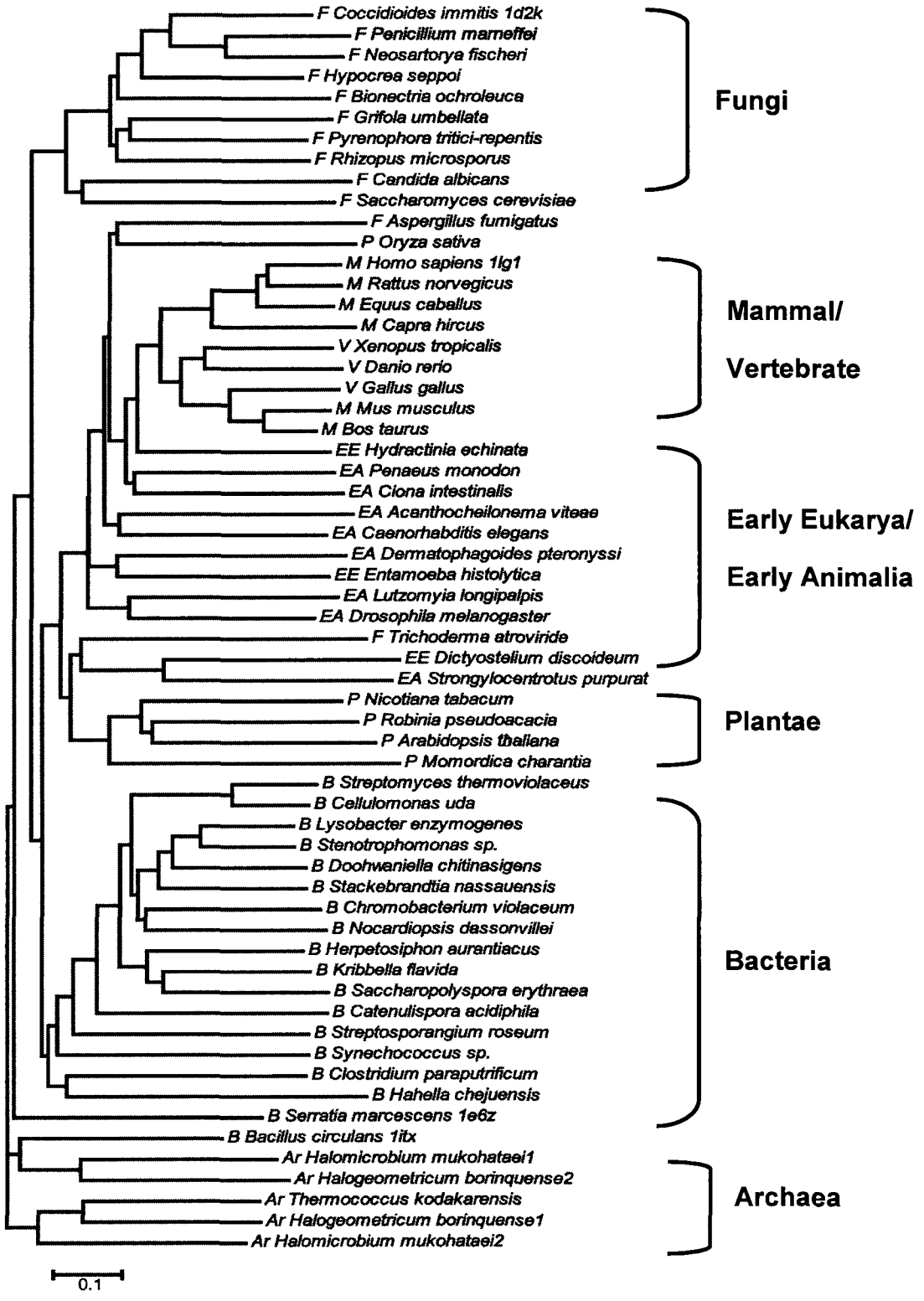


Figure A6. Phylogenetic analysis of the CID sequences from different lineages of organisms

Figure A6 Continued.

The phylogenetic tree was constructed by the neighbour-joining method based on the CID sequences: five from Archaea, eighteen from Bacteria, twelve from Fungi, five from Plantae, three from early eukaryotes, eight from early Animalia, and nine from vertebrates (V) including six from mammals. The sequence names, corresponding GI numbers, and abbreviations are listed in Figure 11 and Table 7. All the sequences were obtained from the protein database at the NCBI.

APPENDIX III

STOPPED-FLOW FLUORESCENCE OF FADD-DD WITH DIFFERENT CONCENTRATIONS

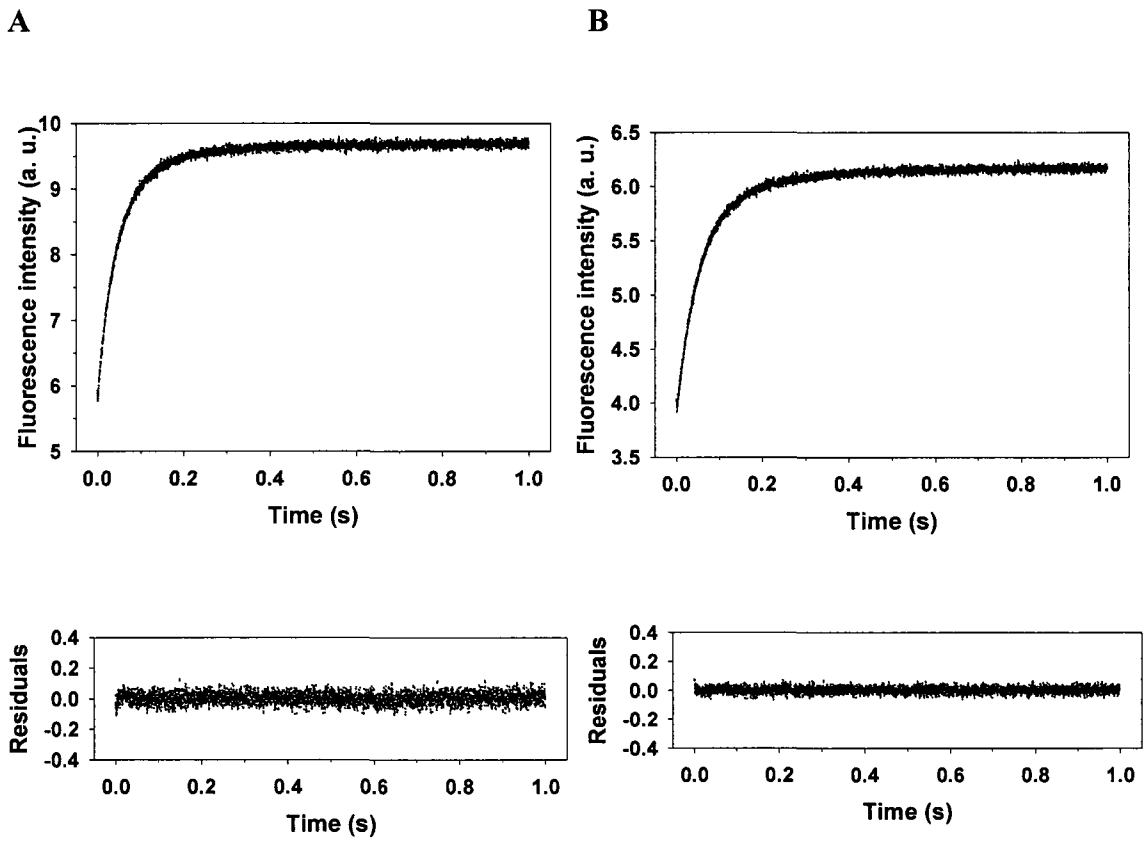
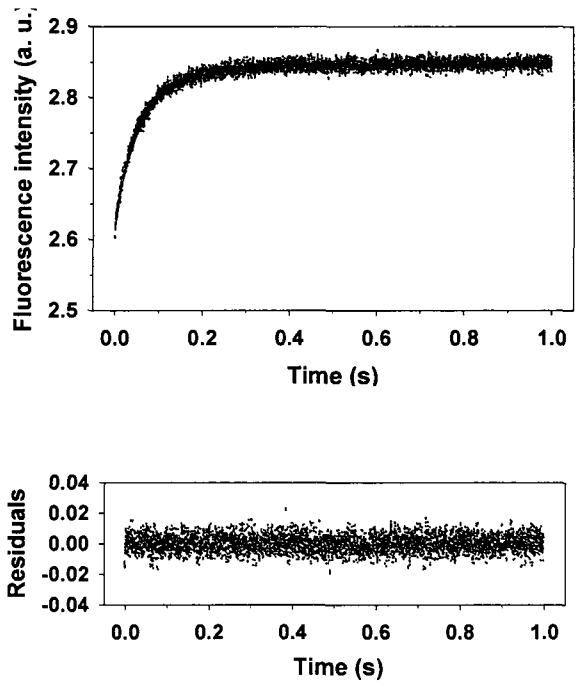


Figure A7. Stopped-flow fluorescence studies of Fadd-DD with different concentrations

Figure A7 Continued.

C



Refolding of denatured WT protein with three different concentrations: (A) 0.2 mg/ml; (B) 0.1 mg/ml; (C) 0.01 mg/ml in refolding buffer was monitored by stopped-flow fluorescence spectroscopy. The conditions are same with previously mentioned.

Table A10. The folding rates and amplitude of hydrophobic core formation of Fadd-DD with the concentration ranging from 0.01 mg/ml to 0.2 mg/ml

Concentration	A_1	k_1	A_2	k_2
0.2 mg/ml	3.226 ± 0.024	23.577 ± 0.183	0.682 ± 0.025	6.479 ± 0.160
0.1 mg/ml	1.839 ± 0.013	20.033 ± 0.150	0.438 ± 0.014	5.512 ± 0.125
0.01 mg/ml	0.161 ± 0.006	24.386 ± 0.780	0.071 ± 0.006	7.980 ± 0.391

VITA

Department of Chemistry and Biochemistry, Old Dominion University, 4541 Hampton Boulevard, Norfolk, VA, 23529, USA

EDUCATION

• 2006 – present (GPA: 3.7/4.0): Ph. D. in Biochemistry with a specialization in protein biochemistry (expected, May, 2011), Old Dominion University, Norfolk, VA, USA.

• 2006 (GPA: 3.6/4.0): M.S. in Microbiology, Indiana State University, Terre Haute, IN, USA.

• 2004 (GPA: 3.4/4.0): M.S. in Biology, Nanjing University, Nanjing, China.

• 2001 (GPA: 3.2/4.0): B.S. in Biochemistry, Central China Normal University, Wuhan, China.

PUBLICATIONS

1) Hai Li, Jessica L. Wojtaszek, Lesley H. Greene. Analysis of conservation in the Fas-associated death domain protein and the importance of conserved tryptophans in structure, stability and folding. *Biochimica Biophysica Acta* 2009, 1794(4): 583-593.

2) Hai Li, Lesley H. Greene. Sequence and structural analysis of the chitinase insertion domain reveals two conserved motifs involved in chitin-binding. *PLoS One* 2010, 5(1): e8654.

3) Li Shen, Yong-Hao Ye, Xiao-Ting Wang, Hai-Liang Zhu, Chen Xu, Yong-Cun Song, Hai Li, Ren-Xiang Tan. Structure and total synthesis of aspernigerin: a novel cytotoxic endophyte metabolite. *Chemistry* 2006, 12(16): 4393-4396.

4) Yong-Chun Song, Hai Li, Yong-Hao Ye, Cheng-Ying Shan, Yu-Ming Yang, Ren-Xiang Tan. Endophytic naphthopyrone metabolites are co-inhibitors of xanthine oxidase, SW1116 cell and some microbial growths. *FEMS Microbiology Letters* 2004, 241(1): 67-72.

CONFERENCE ABSTRACT

Hai Li, Meghan Woughter, Jessica Wojtaszek, Lesley Greene. The importance of conserved hydrophobic core residues on the structure, stability and folding of Fas-associated death domain protein. 23rd Symposium of the Protein Society, Boston, MA, US, July 24-29, 2009.