1-1-2009

Isothermal Titration Calorimetry Studies Of Protein-Mediated Interactions And Preliminary Structural Studies Of Tandem Pdz1-2 Domain Of Psd-95 Protein

Ana Jankovic
Wayne State University

ISOTHERMAL TITRATION CALORIMETRY STUDIES OF PROTEIN-MEDIATED INTERACTIONS AND PRELIMINARY STRUCTURAL STUDIES OF TANDEM PDZ1-2 DOMAIN OF PSD-95 PROTEIN

by

ANA D. JANKOVIC

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2010

MAJOR: CHEMISTRY (Biochemistry)

Approved by:

_____

Advisor Date

_____

_____

_____

_____

DEDICATION

To my mother Ljubica and my brother Marko for all their love and support.

ACKNOWLEDGMENTS

I would like to express gratitude to my advisor, Dr. Mark Spaller for his help, support, and encouragements during the time spent in his research group. I would also like to thank my co-advisor Dr. Louis Romano for all his help during the time of preparing this dissertation. I would like to express my gratitude to Dr. Timothy Stemmler for all his help and useful discussions during the time I have spent in his lab. I would like to extend my gratitude to my committee members Dr. Ladislau Kovari for his insightful suggestions and Dr. Woody Guo for his support.

I had a privilege to work amongst some extraordinary people at Wayne State University. I would like to thank Dr. Christine Chow for providing the necessary help in a very difficult moment. My gratitude extends to all her lab members, but especially Dr. Anne-Cecile Duc, Dr. Mei Li, Dr. Sanjaya Abeysirigunawardena, Dr. Santosh Mahto and Keshab Rijal. I thank Nestor Ocampo for all the help with computer related issues. I thank Sharon Kelly, Melissa Barton, Mary Wood, Erin Scully and Debbie McCreless for their guidance through graduate student's struggles with paperwork. I thank Dr. Brian Shay for his help with MALDI experiments.

Spaller lab members were wonderful group of scientists to work, learn, discuss, and share experiences with. Moreover, I have developed friendships here that will last a lifetime. My special thanks goes to Dr. Dorina Saro (dearest and most dependable friend), Dr. Daljinder Kahlon and Dr. Rachel Parisien (for endless patience for my troubles and mutual love for coffee), Dr Anne-Cecile Duc (thank you for all your kindness and hoping for more "crazy endeavors" still to

come), Dr. Edvin Klosi (for all your help and some opera singing), Dr.Chamila Rupasinghe (for coming to my rescue in most difficult tasks), Dr. Sudhir Sharma (thank you for valuable discussions and also much lighter history and political briefings), Dr. Gomika Udugamasooriya, Dr. Tao Li, Dr. Adnan Memic and Azrael Peredes for making my graduate school experience much more interesting. These projects could not be carried out without the synthetic expertise and major contributions from Dr. Parisien, Dr. Klosi and extraordinary mind of Dr. Rupasinghe. I would like to thank all colleagues from my group and wish them all the best for their future.

I would also like to thank the collaborators I had the wonderful opportunity to work with during my graduate career. Without them the story presented here would not have been complete and I am extremely grateful to them for their contribution and assistance; especially Dr. Krisztina Bencze, Dr. Kalyan Kondapalli, Dr. Madushi Raththagala, Swati Rawat and Jeremy Cook from Dr. Stemmler's lab and Dr. Iulia Kovari, Dr. Philip Martin and Ravikiran Yedidi from Dr. Kovari's lab.

Years spent in graduate school have been very demanding, sometimes exciting, sometimes difficult, but would be impossible without the support of some wonderful people I am proud to have as friends. Without Anne-Cecile Duc and Grigor Georgiev, I believe I would not be able to finish this journey. Katerina Popova, Sandhya Muralidharan, Ying Nualpun Sirinupong, Krisztina Bencze have been so supportive and I wish we will continue our friendship.

My deepest gratitude goes to my Mom and my brother for their love, help, encouragements and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

BIOPHYSICAL METHODS FOR STUDYING PROTEIN-MEDIATED INTERACTIONS

Understanding biological processes in living organisms ultimately means understanding the atomic-level molecular recognition events between the primary biomolecules—proteins, nucleic acids, carbohydrates and lipids—that take place at the cellular level. To these participants can also be added the endogenous array of organic metabolites, inorganic salts and metals, and—last but certainly not least—water, which not only serves as universal host for all the binding interactions among all the aforementioned, but can be an explicit participant as well. The presence of these naturally occurring molecular components, in both static and dynamic roles, results in all the structures and process we associate with cellular function.

Although all the biomolecular classes are involved in generating and maintaining higher-level molecular organization on the cellular level, the formation of macromolecular complexes is predominately the responsibility of *proteins*. To elucidate the function of this diverse group of biomolecules, it is of central importance to reveal their structure. Empirically solving the three-dimensional structure of a protein represents a key step in the determination of its structure-activity relationship (SAR). X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are the methods of choice to accomplish this. To these can be added other methods that aid in characterizing protein structures, circular dichroism (CD) and mass spectrometry (MS). In

combination, all these techniques contribute to assembling a more complete picture of a protein structure, as well as the protein-drug, protein-inhibitor, and protein-protein complexes that it can engage in. Computer-aided visualization and manipulation (e.g., molecular modeling and dynamics), while not strictly empirical techniques, are often invaluable additions to the study of protein structure and behavior.

Despite the enormous amount of knowledge gained from the use of these experimental approaches regarding the structure of a protein (either alone or in a complex), they do not ensure accurate prediction of its function and, by extension, its biological activity. The first step to functional analysis is in fact the understanding of a protein's binding behavior, since the binding of a protein to one or more ligands (broadly defined from small metals and organic molecules all the way to other large proteins) is what will immediately or eventually result in a 'biological activity'. The first rigorous step towards this understanding is to then examine the biomolecular associations themselves, and this at the level of thermodynamic binding parameters. Thermodynamics provides information on the forces that drive biomolecular interactions, providing a detailed energetic foundation to the analysis of the basic behavior of a protein.

In brief, then, a complete understanding of a protein's cellular role means that both structural and thermodynamic properties need to be characterized. In this dissertation, I present research based on both approaches, directed towards two different proteins: the peptide-binding PDZ domains of PSD-95, and iron-binding frataxin. Specifically, the two primary methods utilized in the *in vitro*

characterization of the protein complexes presented here are *isothermal titration calorimetry* (ITC) and *X-ray crystallography*. What follows is a brief overview of the techniques themselves and their theoretical foundations, beginning first with thermodynamic binding parameters and their measurements by ITC, followed by structural analysis as determined by X-ray crystallography.

1.1 Isothermal Titration Calorimetry (ITC) as a Method for Measuring Thermodynamic Parameters.

1.1.1 Basic thermodynamic parameters. There are four basic thermodynamic parameters that govern all biomolecular interactions; these are (1) change in Gibbs free energy (ΔG); (2) change in enthalpy (ΔH); change in entropy (ΔS); and (4) change in heat capacity (ΔC_p).

1.1.1.1 Gibbs free energy change (ΔG). By definition, the Gibbs free energy of a system is the enthalpy of the system (all molecular participants, including solvents) minus the product of the temperature times the entropy of the system (1). Therefore, the change in free energy that occurs during a chemical process (such as a binding interaction or an actual transformation or reaction) is equal to the difference in enthalpy change and the change in entropy at the given temperature (2).

$$G = H - TS \qquad (1)$$

$$\Delta G = \Delta H - T\Delta S \qquad (2)$$

The change in free energy in a chemical process is also related to the association or dissociation equilibrium constants through equation (3).

$$\Delta G = -RT\ln K_a = RT\ln K_d \qquad (3)$$

Thus, by combining equations 2 and 3, it is clear that there is a direct correlation between the key thermodynamic parameters of enthalpy and entropy, and the affinity of a binding interaction as represented by the equilibrium constant.

1.1.1.2 Enthalpy change (ΔH). Enthalpy change of a chemical process can be assessed in a single ITC experiment. By definition, change in enthalpy between products and reactants for a chemical transformation represents heat released or absorbed during the reaction. Exothermic reactions, in which the heat of the system is emitted, have negative ΔH whereas the reactions in which heat is absorbed (i.e. endothermic) have a positive ΔH. Regarding the relations in equation (2), a negative value of ΔH is favorable for the interaction and a positive ΔH is unfavorable. In terms of molecular interactions, changes in enthalpy are most closely linked with actual atomic-level interactions, such as those based on hydrogen bonds, electrostatic (charged, ionic, or dipole) forces, or Van der Waals forces. To a first approximation, the changes in enthalpy are often attributed to the making and breaking of specific interactions, between the binding molecules themselves as well as with (and between) solvent (i.e., buffered water, in both cellular and ITC environments).

1.1.1.3 Entropy change (ΔS). The change in entropy between the products and the reactants for a chemical process is defined as a change of degree of freedom at a molecular level. Theoretically, this can be expressed as the difference in the number of accessible energetic states, before and after a binding interaction. Specifically, this is associated with several processes, such as the net increase or decrease in conformational mobility, as dictated by the changes in rotatable

bonds. Also, the restriction or immobilization (or release) or water molecules is also a major biomolecular process that is linked to changes in entropy. Spontaneous changes occur with an increase in entropy and negative values of ΔS indicate decrease in the number of accessible states during the transformation. In regards to the free energy change given in equation (2), a positive ΔS is favorable for the association and negative ΔS has an opposite effect.

1.1.1.4 Heat capacity change (ΔC_p). The net change in heat capacity between chemical transformation reaction products and reactants, at constant pressure, can be calculated using equations (4) and (5).

$$\Delta C_p = \frac{d(\Delta H)}{dT} \qquad (4)$$

$$\Delta C_p = T\frac{d(\Delta S)}{dT} \qquad (5)$$

In many regards ΔC_p is considered a subtler and more complex parameter than enthalpy and entropy, since it is not as easy to assign to it specific molecular processes. But it is considered by some even more fundamental than the other parameters, since it places limits on what the values of enthalpy and entropy can be at a given temperature, as equations 3 and 4 demonstrate. Nonetheless, there are processes, such as the extent of burial of polar or nonpolar surface area when a ligand binds a protein, that have been correlated to the sign and magnitude of ΔC_p.

1.1.2 ITC as a biophysical technique.

1.1.2.1 General aspects of ITC. Calorimetry, derived from the Latin word *calor*

(heat) and the Greek word *metria* (a measure), is, as the combination implies, a method that quantitatively determines the heat absorbed or released during physicochemical transformations. Calorimetric methods have proven very useful in characterizing biomolecular interactions. However, only recently with technical improvements in instrumentation that have given rise to several highly sensitive titration calorimeters that are increasingly accessible to researchers, has the method became highly valued and widely used. The increased popularity of ITC as a tool for studying biomolecular interactions *in vitro* has been addressed in many review articles lately [1-9]. In order to summarize the wide landscape of applications of the method, Ladbury and coworkers have published yearly surveys on calorimetry from 2002-2005 [10-13].

1.1.2.2 Advantages of ITC. Compared to other methods applied in thermodynamic studies of biomolecular systems, the ITC technique offers some important and unique advantages. The particular aspects involving the interaction of proteins with peptides and metal ions is addressed in Chapter 2 of this dissertation, and the following is a broader survey of ITC as pertaining to five basic attributes of ITC.

1. ITC is the only technique that directly provides the thermodynamic binding constants. van't Hoff analysis of temperature-dependent affinity values from other assay methods has been used to provide ΔH and ΔS parameters (see 1.1.2.4), but there is concern that this does not yield optimal or valid results. In addition, ITC provides those values in the course of a single experiment.
2. ITC relies on the 'universal' signal of heat transfer, which is not restricted to

molecular class or type, and consequently can be used to study almost any type of biomolecular binding process. A detailed list of all the possible applications is given on the official MicroCal website [14]. Further, reliance on a heat signal avoids the need to prepare chemically-modified versions of one or more participants in the process (e.g., as with fluorescently-labeled ligands). All the reagents are used in their 'native' form.

3. Experimental conditions can be easily varied without adversely affecting the binding analysis. Typical perturbations of interest for a particular biological interaction include, but are not limited to, *component concentration*; *temperature*, *ionic strength*, *pH* and *buffer composition*. As part of a systematic and thorough study of a particular protein-ligand system, conditions like the nature and concentration of salts used can be altered as well as mutational changes in the protein itself (site-specific or group mutagenesis). [2, 15]
4. ITC is a true solution method, and unaffected by artifacts that can arise in other assays that involve chemical modification (e.g., affects of fluorescent spectral labels) or when heterogeneous surface immobilization (e.g., as with *surface plasmon resonance* (SPR), such as Biacore) of interacting species is required. This is an important property, since attachment of different reporting groups can alter the binding properties of the system. In addition, unlike the ELISA (enzyme-linked immunosorbent assay) and related methods, the equilibrium state of the interaction is not perturbed since it is not necessary to separate bound from free species.

5. Assuming the appropriate fitting model is used, ITC can also generate the binding stoichiometry (*n*, the number of equivalents of ligand per protein) in a single experiment. This is also a particularly important factor in this dissertation research, as the issue of multivalency is central to the reported PDZ domain and frataxin studies. The reproducibility of ITC measurements regarding molar binding ratios (i.e., *n*) is claimed to be within 1% [16].

1.1.2.3 Limitations of ITC. No experimental method is perfect, and like any other analytical technique, ITC has its own set of disadvantages. But as noted below in the four challenges that can be encountered, most of them can be overcome in one way or another.

1. A large sample requirement still remains the principal limitation of the ITC technique. The actual differences are highly case- and volume-dependent, but ITC may require, for example, hundreds of micrograms of a protein, whereas a fluorescence method or SPR might consume material in the nanogram range [2]. Thus, as currently practiced ITC requires upwards of 10^2-10^3 more material than the most sensitive methods currently available. It is possible, however, to recover the protein sample used, since ITC is a non-destructive method. Also, the next generation of microcalorimeter from MicroCal is reported to use considerable less material (only 200 µL of protein sample) than the current series.
2. A primary requirement for all ITC experiments is that there should be a measurable heat signal. Although noncovalent macromolecular interactions are associated with intrinsically small heat changes (typically 5 to 10

kcal/mol), experimental conditions can be manipulated in order to obtain valid data. However, it is occasionally possible to encounter an entropically-driven high affinity interaction with an insignificant ΔH contribution [17]. For example, a case where ΔG = -8 kcal/mol can be due to TΔS ≈ +8 kcal/mol, and ΔH ≈ 0 kcal/mol (see equation 2). In such a case, it is possible to perform the same experiment using a different buffer system with a higher heat of ionization, or at a higher temperature (if stability of macromolecule is not jeopardized). Such changes can often move the enthalpy change further from zero.

3. ITC experiences a relatively narrow range for the direct determination of binding affinity (K_d), from 10 mM to 0.1 μM. However, a common workaround solution is to conduct a *competitive* titration experiment with a known binding agent of differing affinity. [1, 18, 19]
4. The buffer system might also be a restriction for some proteins that have a requirement for a specific buffer. Protonation and deprotonation reactions result in change of the overall enthalpy, and therefore corrections need to be made [2, 20]. This requires that ITC titrations be performed in buffers with different ionization enthalpies (at least three for reliable data analysis), and if there is any change in the enthalpy values due to proton exchange, these should be adjusted. Otherwise, the observed ΔH values may not be interpretable in terms of only the protein-ligand interactions involved.

1.1.2.4 Indirect estimation of ΔH without ITC: the van't Hoff analysis. In the van't Hoff method, temperature dependence of the equilibrium constant is used

for calculations of ΔH and ΔS; $\ln K_a$ is plotted against 1/T, multiplied by (-R) and used to determine ΔH. This method is potentially valuable since it does not require a calorimeter, and can rely upon much less material, and that of lower purity. This latter aspect alone makes the method attractive for analysis of membrane proteins, such as receptors, which are difficult to access in large quantities in pure form [21]. However, there are significant differences amongst authors as to the accuracy of this approach. There are three major points that van't Hoff omits: (1) the assumption is made that ΔH is constant over the entire experimental temperature range; (2) enthalpy-entropy compensation is not taken into account; and (3) the thermodynamic parameters cannot be determined independently. The matter is not fully resolved. Based on the specific experimental setup, some researchers obtained good agreement in enthalpy values determined by by both ITC and van't Hoff analysis [22, 23], while others that demonstrated significant differences [16, 24].

1.1.3 ITC instrumentation. Not long ago, investigators conducting biocalorimetric analyses did so on custom-made instrumentation specifically built for the purpose in their respective laboratories. Few calorimeters were commercially available with the required sensitivity. The breakthrough period in the field came with the commercial availability of newly developed microcalorimeters (both ITC and DSC (differential scanning calorimetry)) in the beginning of the 1980s. Today, two companies, MicroCal (acquired in 2008 by GE Healthcare) [14] and Calorimetry Sciences Corporation (acquired in 2007 by TA Instruments) [25] dominate the field of ultra-sensitive microcalorimeters

suitable for biomolecular analysis.

The most common type of differential power compensation isothermal titration calorimeters consists of twin coin-shaped cells composed of a highly efficient conducting material (Hastelloy® Alloy C 276 in the case of the VP-ITC, which is the calorimeter used for all measurements reported in this dissertation) surrounded by an adiabatic jacket. [6, 8] A schematic diagram of the VP-ITC is shown in Figure 1.1. Sample cell volumes vary with instrument type; in the case of the VP-ITC the cell volume is exactly 1.4154 mL [26]. Operationally, this requires a (bio)macromolecule sample of approximately 1.8 mL ,even though the working volume of the cell is only 1.4 mL [26]. In addition to the sample cell filled with the macromolecule solution, a reference cell is also present that contains either water or buffer, and plays no direct role in the titration. The titrating agent or ligand is loaded into an injection syringe with a volume of 290 μL.

A thermoelectric device measures the temperature difference between the sample and reference cells, and a second device measures the temperature difference between the cells and the adiabatic jacket. During the experiment, very low constant power is applied to the reference cell and a variable power to the sample cell in order to preserve the temperature difference. The power added to the sample cell to maintain this fixed difference is the experimentally measured quantity related to the heat of reaction [27]. A constant power (< 1mW) is applied before the injections start and represents the baseline signal [6]. The noise level of the VP-ITC instrument is 1 ncal/sec (4 nW) and the baseline constancy is ± 5 ncal/sec (± 20 nW) [26].

In an ITC experiment, the signal actually presents time dependence of the electric power (µcal/sec) necessary to maintain constant the temperature difference between the reaction in the sample cell and the reference cell, after each injection of reactant [8]. The area under each peak is the area associated with the process. Depending on the reaction type, exothermic or endothermic, heat is released or absorbed. For an exothermic reaction, with the temperature increase, feedback power will be deactivated in order to preserve equal temperatures between the cells, resulting in a negative signal. The situation is reversed for the endothermic reaction, where the recorded signal is positive [6].

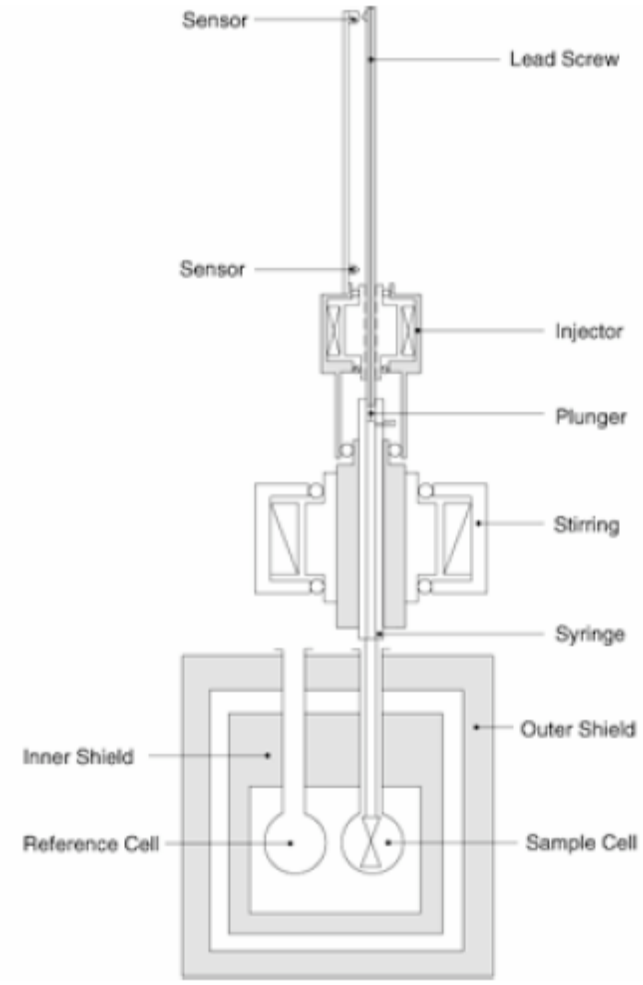

Figure 1.1 A schematic diagram of the ITC instrument. Reproduced from the VP-ITC MicroCalorimeter user's manual [26].

1.1.4 Designing the ITC experiment. There are many different aspects that need to be taken into account before performing an ITC experiment. A critical factor is correct sample preparation, both in terms of purity and concentration determination of both the macromolecule and the ligand. The solutions of the two interacting species must be prepared carefully using the same exact buffer in order to avoid a false heat signal that such solution mismatch can generate. In the experiments described in this dissertation two different techniques were

employed in order to achieve this. For the PDZ domain protein-peptide interactions, both solutions were prepared in the same buffer, and the ligand solution was usually titrated to adjust the pH value to within acceptable limits. A difference of ±0.02 pH units was considered acceptable [1, 2]. For frataxin protein-iron interactions, that procedure was impossible since all sample preparation had to be conducted anaerobically. It was determined in a separate study that dissolving iron salt in the buffer that the protein was dialyzed in did not change the pH significantly.

Accurate concentration estimation of both macromolecule and the ligand is important since it directly affects the calculation of all thermodynamic parameters. Protein concentration can be determined using different assays: UV reading at 280 nm wavelength; Coomassie Plus Assay (modified Bradford assay); and the Advanced Protein Assay, among others. Ligand concentration need also be precisely determined. For both PDZ domain -peptide and frataxin-iron studies, ligand was in the solid form: lyophilized peptide powder or inorganic iron salt, respectively. In both cases, careful weighing of the compound yielded sufficient accuracy for the concentration determination. Ligand concentration affects all the thermodynamic parameters (stoichiometry, K_a, ΔH) obtained from ITC experiments, while the concentration of the macromolecule affects the stoichiometry (n) of *the* reaction [26].

For the PDZ domain-peptide interactions, thorough degassing along with thermo stating was needed. The thermovac is an ITC accessory device used for this purpose, prior to loading protein and peptide samples into the sample cell

and injector [26]. Degassing of the solutions is also important, since the presence of released air bubbles can interfere with the signal that is recorded during the ITC run. In order to accelerate the degassing process stirring the macromolecular sample can also be employed. However, stirring can lead to protein denaturation and precipitation, and therefore it was avoided in all the experiments described in this dissertation. During thermo stating the temperature species is usually kept at few degrees lower than the actual working temperature. This is preferred since the time required for the instrument to heat up is much less than the time required to cool down. For frataxin-iron interactions all these preparations were avoided, due to the anaerobic preparation of samples.

In a typical ITC experiment, the macromolecule solution is added to the sample cell using a long needle glass syringe with a blunt end [6]. Slowly releasing liquid is followed by two or three rapid spurts of at least 2.3 mL solution to ensure the removal of air bubbles from the bottom of the cell. Any excess solution remaining in the reservoir is removed.

ITC experiments are fully automated, and can be designed in several different ways. One key parameter that can be changed easily is the temperature, which is kept fixed throughout the duration of a single experiment. The range of temperatures available with the VP-ITC varies from 2 °C to 80 °C [26]. This is fairly broad, and the very low and high ends assume that the protein remains stable (either avoiding precipitation or denaturation), which is of course case dependent. Other adjustable parameters include injection volume, injection number, spacing between the injections, stirring speed, baseline position, and

feedback mode. At the outset of a study with a new protein, several pilot experiments are performed in which these settings are systematically varied. Once optimal conditions are found, they are employed for all future experiments.

Control ITC experiments are another necessary factor in characterizing the system under study. Usually referred to as *blank experiments* or *blank titrations*, they can be designed in different manners. Titrating ligand solution into the matching buffer in the sample cell is a control experiment with the most pronounced impact. Highly concentrated ligand solution (volume = 290 µL) is diluted several fold when added to the sample cell (volume = 1.41 mL) and can lead to the release or absorption of a considerable amount of heat. This effect must be separated from the heat of the macromolecular interaction by running this control experiment and correcting for the ligand dilution effect.

Another variant of the dilution effect involves that of the *protein*. Buffer from the injector is added to the protein solution in the sample cell. The heat change is usually insignificant, since during the course of the experiment the sample solution is only diluted ca. 10-20%. This type of blank titration is usually only needed when starting a new protein study, or switching to a new buffer system. Finally, the last type of blank titration is buffer into buffer. Assuming care is taken to prepare the protein and ligand in matched buffers to begin with, this is not a required control, as the heat associated is usually negligible under normal circumstances. it may be important, though, if the experiment is performed at more than ±10°C of the environment temperature (since the ligand is in the external syringe at ambient temperature) [26].

Another control experiment that can be useful in investigating errors or artifacts in ITC experiments is *reverse titration*. Here the interacting species switch places: protein is used as a titrating agent, and ligand is placed in the sample cell. The concentration of both components needs to be adjusted accordingly (since the titration typically proceeds to saturation), while keeping all other parameters the same as in the standard “direct titration”. Theoretically, in case of 1:1binding, if the correct model is used for data fitting, the same values for the thermodynamic parameters should be obtained from both experiments.

If desired, ITC experiments can also be designed to obtain the enthalpy change (ΔH) alone, or the complete set of binding parameters (n, K_a, ΔH, ΔS). In the first case, introducing ligand in a *single* injection into the protein sample is sufficient. Enthalpy change can be determined at concentrations when the binding partners are fully associated and the saturation is still low, *i.e.* full saturation at partial saturation [28]. In order to obtain all binding parameters, a complete isotherm needs to be recorded. Usually, ligand is added in ca. 3-4 molar excess to ensure complete saturation of the macromolecule at the end of titration. Another important point is the limiting sensitivity of 0.1 µcal/s of the VP-ITC instrument [26]. Therefore, experiment should be designed in such a way that in each injection at least 1-2 µcal/s heat change is produced [9]. The most recent version of the software for VP-ITC has an option for Single Injection Method (SIM) [29] as a separate, innovative ITC technique.

1.1.5 Data analysis. The method of data analysis depends on the system under study. Detailed explanations about the data treatment are given in Appendix A.

ORIGIN software supplied by MicroCal comes preloaded with different models of curve fitting; they include: “One Set of Binding Sites”, “Two Sets of Binding Sites” and “Sequential Binding Sites” models. The latest version of the ORIGIN software offers additional data fitting equations that include “Enzyme/Substrate/Inhibitor Assay”, “Dissociation Model” and “Competitive Binding Model”.

Before peak integration, the heats of binding are normalized as a function of ligand concentration. Additionally, a volume correction is also performed due to the dilution of the macromolecule during each injection (also explained in Appendix A). The area of each peak is then automatically integrated using the routines of the ORIGIN software. Experimentally obtained data is normalized as kilocalories per mole of ligand injected, and plotted against the ratio of the ligand to macromolecule concentration. The derived sigmoidal titration curve is further fitted to yield the final binding isotherm. The final figure obtained characterizes the interaction of the macromolecule and the ligand, and includes the thermodynamic parameters of binding, *i.e.* n, K and ΔH.

The quality of the experimental data fitting procedure is usually judged by the value of the $\chi 2$ (chi squared) parameter [30]. The standard manner of defining the best fit is to choose the parameters so that the sum of the squares of the deviations of the theoretical curve from the experimental points for a range of independent variables is at a minimum. The $\chi 2$ parameter is equal to the sum of the squares of the deviations of the theoretical curve from the experimental points, divided by the number of degrees of freedom.

If a set of experimental data is treated with different models, the most meaningful thermodynamic parameters are obtained from the model that results in the minimum $\chi 2$ value—provided, of course, that they do not violate what is biochemically known about, or reasonable for, the protein under study.

1.1.6 of the ITC *c* value. The critical parameter that determines the shape of the binding isotherm is the unitless constant *c*. The value *c* is defined as product of the binding constant K_a and the total starting macromolecule concentration of the macromolecule in the sample cell, M_{tot}, times the stoichiometry *n*, as shown in the equation (6).

$$c = K_a M_{tot} n \quad (6)$$

Tight binding (i.e., small K_d) as a consequence has a very large *c* value and lower values are signature for weak binding (high K_d). Figure 2.2 depicts shape variants of the binding isotherm as a function of the *c* value.

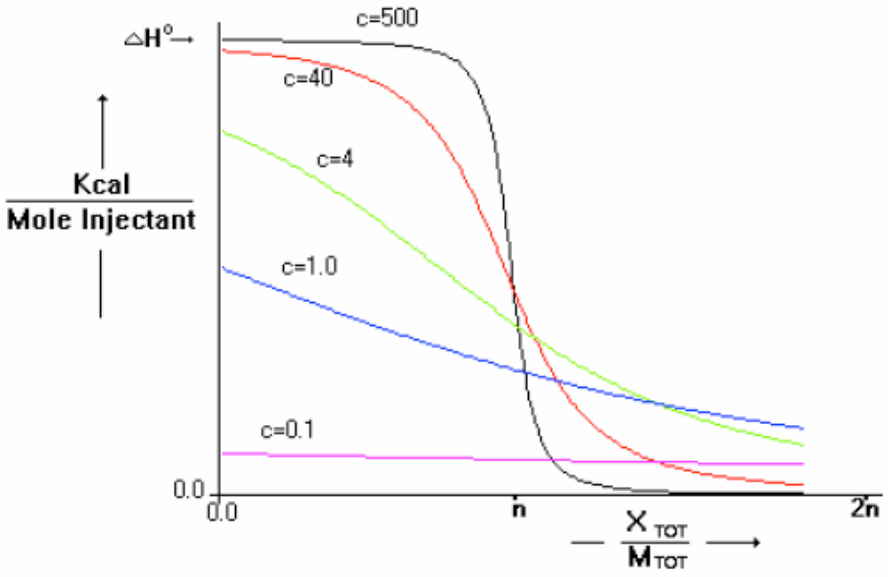

Figure 1.2 Schematic diagram of simulated binding isotherm for different values of *c*. Picture reproduced from VP-ITC user's manual [26].

Close inspection of these isotherms reveals that their shape is reasonably sensitive to the binding constant only when the *c* value is between 1 and 1000 [26]. Often times this range is referred to as the "experimental K window". Curve fitting of the experimental data should allow reliable determination of all thermodynamic parameters (*n*, K_a, ΔH,). Therefore, the binding isotherm should have at least a few data points corresponding to the initial phase of the experiment, additional points for the transition from the unsaturated to the saturated state of the macromolecule, and finally, a few points towards the end of the experiment that demonstrates complete saturation. Each of these parts of the experiment is valuable, and the absence of any makes data interpretation difficult.

Values for binding affinity (i.e., K_d) that can be reliably measured by ITC are typically between 10^{-3} and 10^{-8} M. Very tight binding cannot be determined, since in order to satisfy the *c* value and generate a continuous, fittable curve with all the regions discussed above, the concentration of the macromolecule should be very small, which will result in a very low heat signal. Even with the most recent ITC instrumentation these measurements cannot be accurately made. Similarly, very weak binding cannot be studied. For the good data fitting and an acceptable *c* value, macromolecular concentration would need to be so high that sample aggregation and other effects would preclude the accurate determination of the thermodynamic parameters. However, there is a way to extend the range of such applications by designing appropriate competitive experiments [18, 19]. The most recent version (2004) of the ORIGIN software tutorial [31] includes

complete descriptions of this method, based on the contributions of Sigurskjold [19].

In brief, a competitive displacement experiment can be executed by titrating a very tight binding ligand into a solution of macromolecule already saturated with a weak binder of known affinity [3]. Obtaining thermodynamic parameters is possible only if two ligands have significant differences in binding enthalpy [18, 32-35]. Detailed calculations and fitting routine are provided in Appendix B. Another approach to overcoming this problem in the case of a weak affinity system is to saturate the macromolecule in a single injection of ligand. The obtained enthalpy change is fixed during data fitting in ORIGIN.

1.1.7 Calculating thermodynamic binding parameters in ITC experiments. ITC records the heat signal during a biochemical interaction. The heat released or absorbed during each injection is proportional to the change in concentration of the bound species. The overall heat content during the experiment is related to the total macromolecule concentration, and used to calculate the enthalpy change of binding. Utilizing the volume change in the course of the titration, the software calculates the heat released or absorbed in each injection. Calculated values from the model and experimental values for data fitting are then compared by standard Marquardt methods. Iterations are performed until no further improvement in the data fitting finally gives values for the stoichiometry (*n*), association constant (K_a), and enthalpy (ΔH). From this ΔG is calculated. The change in entropy (ΔS) is obtained using ΔG and ΔH from the standard thermodynamic equation (equation 2). It is in this manner that ITC provides all

the thermodynamic parameters from a single experiment [7]. Detailed calculations are given in the Appendix D.

1.1.8 Thermodynamic data interpretation. Binding affinity is an exponential function of the energy of association, as shown in equation 7.

$$K_a = e - \frac{\Delta G}{RT} \qquad (7)$$

Therefore, ΔG is considered the most important thermodynamic parameter of binding that determines the stability of a biomolecular complex. Free energy is often used to characterize structure-function relationships [9]. In various studies it was noted that, compared to thermodynamic parameters measured by ITC, ΔG has the lowest signal to noise ratio (values close to one another) [7]. Enthalpy-entropy compensation (explained in section 1.1.8) is the reason why free energy change is relatively insensitive to the change in molecular details of interactions [36-38]. Since two different parameter contribute to the value of ΔG, many combinations of ΔH and ΔS can result in the same binding affinity [8].

The enthalpy change measured by ITC is a property of the net change of the whole system. Therefore, the total heat released or absorbed in the cell upon adding the ligand represents the cumulative effect of many components. These include, for example, the enthalpy of association of molecules, the heat due to mixing and the heat of dilution (especially that of the ligand). Due to contributions from nonspecific effects, what is measured is actually the *apparent* enthalpy change [3]. For these reasons it is important to correct the calculated heat and accurately determine the *intrinsic* ΔH of the reaction.

The *enthalpy change* has the highest signal-to-noise ratio in an ITC experiment [7]. This parameter reflects the formation of discrete noncovalent interactions [9]. During the binding of the ligand to the macromolecule, old interactions are broken while new ones are formed. The interactions that are lost include hydrogen bonding and van der Waals interactions between the protein and the solvent (water in most cases) and ligand and the solvent. New bonds formed are protein-ligand interactions that also typically involve hydrogen bonding and van der Waals forces, but can also include hydrophobic interactions and salt bridges, and solvent reorganization near protein surfaces. All these interactions can be favorable or unfavorable, but since ITC measures the net change, the overall effect is usually smaller than the individual specific contributions [17]. Specific interactions, however, can be evaluated through alanine scanning [39] and mutational studies [40], by removing specific amino acid functionality from the binding site of the macromolecule.

The entropy change is calculated indirectly from ΔG and ΔH based on equation (2). Entropy change is less precisely determined than these other two thermodynamic parameters because of error propagation [7]. Several factors contribute to ΔS, but the main one is generally considered the solvation (hydration) effect. Complex formation is often coupled with the release of ordered water from the binding regions of the protein surface and the ligand, which enters into the *bulk water* phase, which is less ordered (higher entropy). The net result for this specific process is a large and positive entropy change upon binding. The factors that affect unfavorably the entropy change of binding are any reductions

of rotational, translational and conformational entropy when free protein and ligand associate. Standard errors within experiments for the ΔH, ΔS and ΔG parameters are typically around 1 kcal/mol [7].

Stoichiometry is calculated from the molar ratio of the interacting species at the equivalence point of the titration curve. In order to obtain an accurate value for *n*, the concentration of the interacting species should be known very precisely [2]. The potential factors that can affect the *n* value are experimental errors in the concentration measurement. For example, if a solid ligand is weighed out in order to prepare the desired concentration solution, any error in weight measurement will be transferred to the error in concentration. If the concentration of the macromolecule (in the sample cell) is lower than what is determined as a result of misfolding, inactivation, or degradation, then the *n* value will be smaller than expected, and *vice versa*. Similarly, the presence of an impurity in the ligand will cause the *n* value to be larger than expected. For this same reason ITC can be applied as a quality control tool for analyzing the protein stability where the stoichiometry is already known [2].

1.1.9 Enthalpy-entropy compensation. As mentioned earlier (section 1.1.8) compensation between ΔH and ΔS is a common phenomenon observed in many biological systems [37, 38]. This is a result of a linear relationship between the two parameters (*i.e.* any increase in favorable enthalpy is often matched by an unfavorable entropy and *vice versa*), resulting in no noticeable improvement in affinity. In such a case, any increase in the bonding that produces a highly favorable enthalpy will be at the expense of freezing rotations and increasing

order, leading to more unfavorable entropy. Enthalpy-entropy compensation has been considered a general consequence of weak intermolecular interactions [41].

1.1.10 Heat capacity change. Since the working temperature range for the VP-ITC is quite wide (2-80 ºC), it is suitable for the experimental calculation of ΔC_p during complex formation. Heat capacity change can be calculated from a plot of ΔH versus temperature (T) using equation (3). The correlation between ΔC_p and the change in surface area buried upon binding of the macromolecular complex, has been very well studied, and forms a link between structural and thermodynamic information.

1.2 X-ray Crystallography as a Method for Structural Determination. Proteins are defined as amino acid heteropolymers, whose exact sequence is defined by their corresponding gene sequence. The amino acid sequence determines its three-dimensional tertiary structure, which dictates its function [42]. Structural information reveals design principles of the molecular foundations of living systems. It not only answers basic mechanistic questions of protein function, but it can also provide insight into evolutionary relationships between cellular components [43].

There are several methods used to determine structural information: X-ray crystallography; electron microscopy (EM); small-angle X-ray scattering (SAXS); and nuclear magnetic resonance (NMR). Among them, X-ray crystallography has historically proven to be the preferred method, judging by the more than 40,000 structures deposited in the official protein structure repository, the Protein Data Bank (PDB; http://www.rcsb.org). Although protein NMR methods have increased

in popularity, X-ray crystallography is still the major source of protein structures since it routinely yields the highest resolution data, and can solve structures that are considerably larger than those solvable by any other current technique. The number of structures in the PDB has increased exponentially over the last few years [44], a trend that will certainly continue with the development of "structural proteomics" as a field of research [45-47].

X–ray crystallography is a powerful technique in resolving the three-dimensional structure of biological macromolecules [48-50]. Details of covalent and noncovalent interactions during protein-ligand complex formation are frequently revealed by this type of analysis [51]. Thus far, it is the only method that can routinely reach atomic resolution, and in addition to providing information to basic questions about protein behavior, it has proven invaluable in the applied biomedical sciences. Data obtained using this method has been instrumental in structure-based design efforts [52, 53] that ultimately led to the discovery and development of many drugs, including HIV protease inhibitors, neuramidase inhibitors, rennin inhibitors [53] and anticancer agents [54], to cite only a few examples.

X-ray crystallography was first used as a method in structural biology in 1934, when Bernal and Crowfoot (later Hodgkin) [55] produced the first diffraction pattern of a protein, *pepsin*. The real breakthrough in the field came with the work of Perutz [56] and Kendrew [57] on the structures of hemoglobin and myoglobin, respectively, in the 1950's. Technological advances brought many improvements to instrumentation, sample preparation, data collection and processing. However,

the methods and mathematical approaches that were developed to solve those first protein structures represent the core of the methodology that is used even today.

In the process of solving a protein structure by X-ray crystallography, a series of steps, each with its own difficulties, is required. In brief, these include (1) crystal formation; (2) data collection; (3) solving the structure. Each of these steps is summarized below.

1.2.1 Crystal formation. The first requirement is a reliable method for producing a good supply of protein crystals. Protein crystals are, unlike mineral or small molecule crystals, mostly made of water (30%-80%), and as a result are extremely fragile due to the relatively weak interactions that hold the lattice together. For determining crystallization conditions and effecting subsequent growth of crystals of suitable quality for structure determination, at least 2 mg of pure protein is needed, in a highly concentrated form of 5–30 mg/mL [43]. Although this amounts is considered high for certain proteins, only a few microliters (1 – 3 µL) of this concentration sample is required for a 'screening set up'; thus, a 1 mL sample can be used to screen for hundreds of different crystallization conditions [58]. In the early days of protein crystallography, experiments were carried out on proteins obtained from natural sources (usually cow and pig organs or blood). Recent advances in recombinant DNA technology [59] and newly developed bacterial expression systems, principally those based on *E. coli* [60, 61], routinely allow overproduction of properly folded target protein

[62]. A wide array of bacterial expression plasmids is nowadays available from commercial vendors (e.g., Novagen and Invitrogen).

However, not only protein *quantity*, but also *quality* is important for crystallization [63]. Protein samples need to be chemically pure; that is, having only one type of macromolecule or complex in the tube. It should be folded into its native conformation and not contain covalently heterogeneous features (such as partial glycosylation or phosphorylation, or mixtures of truncation products). Methods like gel electrophoresis, protein chromatography, and mass spectrometry should be employed in order to provide and confirm a chemically pure sample [63, 64]. The sample must also be conformationally pure, which is usually assessed with size exclusion chromatography or with light scattering measurements [65-67].

Protein crystallization is a phase transition that consists of nucleation and crystal growth [63, 68]. The basic idea is to transfer the concentrated protein of interest into a solution of precipitants that will trigger crystal formation. Precipitants act by promoting the formation of intermolecular interactions by affecting macromolecular hydration, molecular crowding, solubility, hydrophobic interactions, and electrostatics. Unfortunately, the same components that push proteins towards crystallization can also promote the common (and undesired) phenomenon of amorphous precipitation. It is impossible to predict in advance which conditions, if any, will cause a protein or protein complex to crystallize. The list of possible factors is a long one: precipitant type; pH; salt concentration;

detergents; temperature; and the inclusion of cofactors and ligands are the most common variables. The variables are endless.

A prudent route to take involves screening sets of already established 'good' protein crystallization agents. Crystallization screens often focus on a limited set of concentrations of various sizes of polyethylene glycols (PEGs), salts, and alcohols that have been successful for protein crystallization in the past [68-71]. Commercially available crystallization screening kits (e.g., those of Hampton Research, Wizard, and Nextal), especially if the experimental set up is performed utilizing microcrystallization robotics, makes it possible to test thousands of initial conditions. The newest trend that appears to be emerging in the field is to have screenings conducted in a dedicated 'crystallization centers'.

Although there are other techniques available, *vapor diffusion* is the most frequently used method of setting up crystallization experiments. Protein and precipitant are usually mixed in a 1:1 proportion, in volumes from 50 nL to 1 µL, depending on the type of instrumentation used. A solution drop that is formed in such a way is then placed in a sealed chamber that contains a large volume relative to the drop of precipitant, also termed the *mother liquor* solution. The imbalance in vapor pressure slowly draws water from the protein/precipitant drop into the mother liquor, and gradually concentrates the protein solution. Crystals can be formed if the process proceeds through a supersaturated protein solution. Crystal formation time is highly variable. They may form within a few hours (sometimes even only minutes), or, quite commonly, they may require weeks or months, depending on the protein and the particular conditions used.

Vapor-diffusion can be performed in two modes: the *hanging-drop* method and the *sitting-drop* method. In the hanging-drop method, each drop is prepared by mixing the protein sample with the buffer. A small slide with the droplet is inverted over a prepared well with the buffer. The drop is kept on the slide by the surface tension forces. This kind of setup is usually performed in a multiwell plate format, with 4 x 6 wells per plate. The sitting drop method utilizes the same principle, only in this case the drop sits on a microbridge over the precipitant reservoir, in the same well.

Another interesting method used for crystallization of proteins is known as the *microbatch* method. Using this technique a small drop of concentrated protein solution, combined with the crystallization reagent of choice, is pipetted under a layer of oil. All of the reagents involved in the crystallization are present at a specific concentration and no significant concentration of the protein or the reagents occurs in the drop. This method is simple and efficient for screening for crystallization conditions, but not for obtaining good quality crystals.

A new advancement in crystal preparation methods is the use of robotic systems that are able to prepare and evaluate up to 40,000 crystallization experiments per day, using this microbatch technique, or 60,000 micro-vapor diffusion experiments per day. These technological advancements are very important in the new field of high-throughput crystallization and structure determination, which is expected to have a large impact in the drug discovery process.

1.2.2 Data collection. Unfortunately, having crystals form does not necessarily mean that they will be suitable for obtaining a good structure. Crystals can only diffract X-rays if they are well ordered. Since weak interactions hold the lattices together, fragile crystals might, in some cases, have internal disorder and therefore diffract poorly. This is impossible to determine in advance just by visually judging the crystal appearance, The only way to be certain is to actually expose a crystal to the X-ray beam and measure the diffraction pattern.

The question is then why some crystals lead to solvable structures and others do not. A more detailed discussion of the nature of crystals is required. Crystals represent an “orderly three-dimensional array of molecules, held together by noncovalent interactions [55]"; that is, built in space from multiple copies of the same repetitive element. These building blocks are called *unit cells* and are related to each other by simple translations. There are fourteen types of unit cells (cubes, prisms, hexagons).

The unit cell may contain a single copy of the protein or complex of interest, but often it contains many copies, and they are related to each other by a set of rotations and translations defined by crystallographic symmetry. The symmetry of the unit cell is indicated by its *space group*. Theoretically, 230 different space groups exist, but luckily for researchers, only 65 different types of symmetry are actually known. In the unit cell the “largest aggregate of molecules that possesses no symmetry elements, but can be juxtaposed on other identical entities by symmetry operations, is called the *asymmetric unit* [55].” The asymmetric unit and its dimensions (coordinates) are determined in solving the

structure. Often times this can be a biological unit, but it can also be built from symmetrically positioned protein chains such as homodimer or homoteteramer. These symmetry elements are depicted in Figure 1.3.

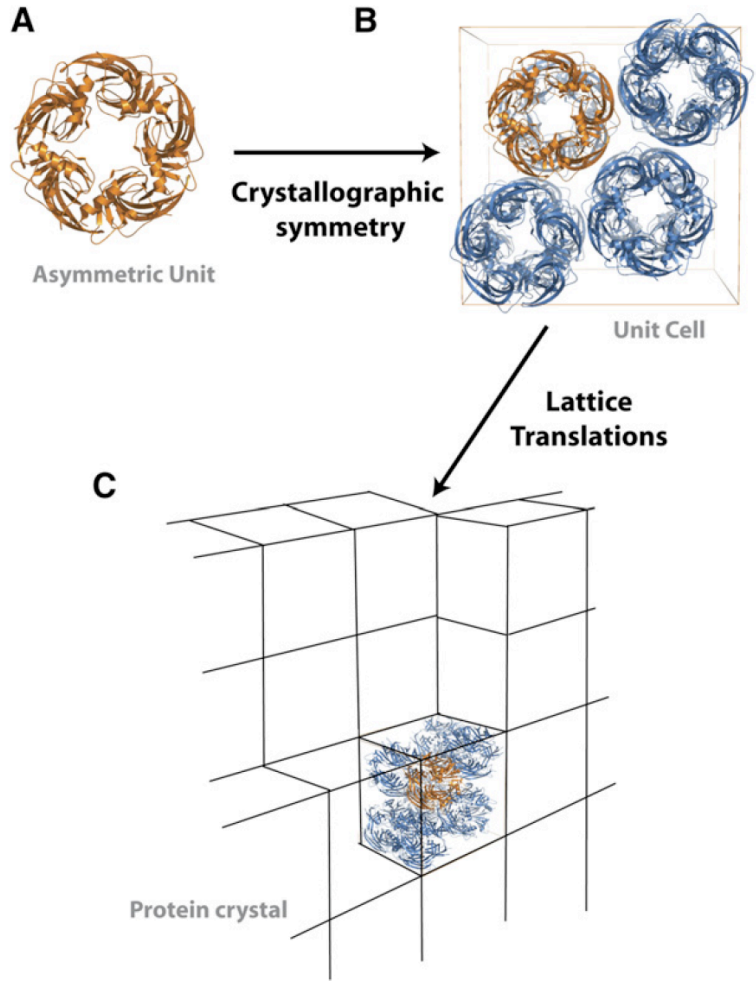

Figure 1.3 Anatomy of a protein crystal. Reproduced from Minor, DL[43] by permission of Elsevier

1.2.3 Data collection. Actual determination of a structure begins with measuring the X-ray diffraction from the crystal. X-rays are electromagnetic waves that have short wavelengths, and the best range for crystallography involves wavelengths between 0.5 Å–1.6 Å. These X-rays are able to penetrate samples of up to 1 mm

thickness. They are also able to discern between features that have magnitudes on the order of angstroms (1-2 Å covalent distance between atoms, and hydrogen bonds at 2.5 Å – 3.5 Å).

X-rays can be produced by exciting a metal (usually copper or molybdenum) with high energy electrons produced by a heated filament, and accelerated by an electric field. X-rays are emitted by excited copper electrons, making a L→ K transition at λ = 1.54 Å. In order to produce monochromatic radiation, usually a nickel filter of 0.015 mm thickness is used.

In an X-ray diffraction experiment the intensity of the coherently scattered X-rays is measured. The diffraction pattern is an array of “reflections” in which the pattern and spacing is set by the unit cell parameters (Figure 1.4).

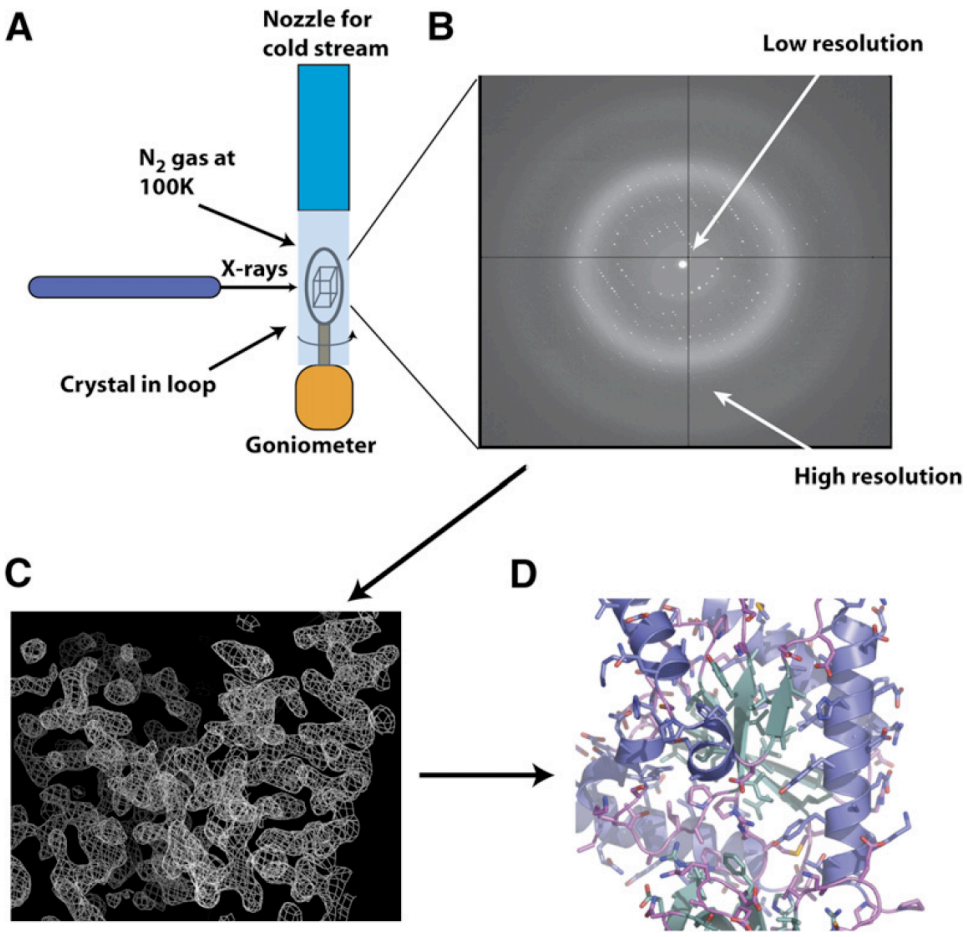

Figure 1.4 Scheme of diffraction experiment. Reproduced from Minor, DL[43] by permission of Elsevier

Each reflection carries information about the entire asymmetric unit. There is an inverse relationship between the position of a reflection in a diffraction pattern and the resolution of the information. Near the center of the diffraction image, close to the position of the incident beam, are where the low-resolution reflections found, while high-resolution data are found at larger scattering angles (Figure 1.4B). High-resolution reflections determine the resolution limits of a given experiment. These limits are a result of the quality of the internal order of the crystal. Once the crystal parameters (i.e., space group and unit cell size) are

determined from the diffraction pattern, the crystallographer knows exactly where all the reflections at any resolution should be.

For data collection either a diffractometer or synchrotron radiation is used. Diffractometers have three main components: X-ray source, goniometer and detectors. The goniometer is able to rotate a crystal by a 40° angle in each of two perpendicular planes during data collection. The steps in diffraction experiments are described in Figure 1.4. Presently, most data collection is done at the synchrotron radiation sources, although for the studies reported in this dissertation, an in-house Rigaku instrument was used. The main advantage of synchrotron radiation is that the intensity of X-rays generated is much higher, so the crystal exposure time can reduced. This is very important since most molecular crystals lose the ability to diffract X-rays within a few seconds of exposure, due to the high energy imparted by X-rays. In order to preserve crystals, they have to be flash frozen before the exposure, usually using liquid nitrogen [72]. As already explained, this is important for crystal lifetime and quality of data [73, 74]. Nearly 90% of all crystallographic data is performed at a temperature of 100 K or lower [73]. During the whole expose time, the crystal is placed in a cryostream of nitrogen gas, which keeps it frozen throughout the experiment.

The X-ray crystal reflections are recorded by detectors. These can be of different types depending on their specific design. *Image plate detectors* and *area detectors* are most commonly used for X-ray crystallography. Image plates are plastic sheets with a coating of small crystals of phosphor material, like

$BaF:Eu^{2+}$. By hitting the detector surface, X-ray photons excite the electrons in the material. Laser is then used to scan the detector surface and record the signal. The pixel size depends mainly on the reading system and varies from 100x100 to 200x200 μm. The image plate detectors have a wide dynamic range and can be used indefinitely.

Area detectors are based on either a gas filled ionization chamber or an image intensifier coupled to a video system. The X-ray photons cause ionization of gas atoms. The disadvantage of this type of detector is the limitation of counting rate due to the build up of charges in the chamber and the processing electronics. The maximum counting rate is about 10^5 Hz. Another disadvantage is the lower sensitivity at shorter wavelengths due to poor absorption, if used in combination with synchrotron radiation.

1.2.4 From diffraction data to electron density. In order to obtain an atomic-scale image from crystallographic data, a computer simulates the action of a lens [55]. Electron density is computed from the list of indexed intensities obtained, which is the final outcome of the X-ray data collection. Each intensity has an index *hkl* assigned corresponding to its position in the reciprocal lattice. This diffraction pattern can be mathematically related, very precisely, to the object using a *Fourier transform* [55]. A Fourier-series description of the reflections is converted into Fourier-series description of electron density. A reflection can be described by a *structure-factor equation*, containing one term for each atom in the unit cell.

Crystallographers use Fourier transforms to convert structure factors to an electron density equation. However, one necessary piece of information is missing. A description of a diffracted wave includes knowing its amplitude, frequency and phase. For each reflection, we can only measure its intensity, I_{hkl} not the complete structure factor F_{hkl}. The amplitude of structure factor F_{hkl} is actually proportional to square root of the measured intensity. Unfortunately, its phase still remains unknown. In an analogy that was mentioned before, in focusing reflected light from the object, a lens maintains all phase relationships among the rays constructing an image accurately. Recorded diffraction intensities do not carry that necessary information. This is often referred to as the *phase problem* of X-ray crystallography. This problem can be overcome, however, and phases can be obtained using different indirect approaches, as explained in the following sections.

1.2.4.1 Heavy atom derivative. This method was developed when solving the first protein structures in the 1960s. It is based on modifying the protein in the crystal with a "heavy atom", such as in the form of a mercury or platinum complex. Because proteins are made of "light atoms" (C, H, O, N, S), the selective addition of a heavy atom or atoms with a large number of electrons (e.g., 80 electrons in Hg) to precise binding sites on the protein causes perturbations to the intensities of the reflections in the diffraction pattern. The key objective is to produce modified crystals in a way that the unit cell is not altered, and possessing the same dimensions and crystal symmetry group. This condition is also known as *isomorphism*. Obtaining phases from only one crystal derivative

is called the *single isomorphous replacement* (SIR) method; if two or more derivatives are required then the method is termed *multiple isomorphous replacement* (MIR).

At the same time, the presence of heavy-atom should result in detectable changes in the reflection intensities. If that is the case, than these atoms can be easily located and the phases can be calculated by different computational programs that are available. Phase estimates are further used for refinement and structure solution. For producing the right crystals two methods are employed: soaking of native crystals with heavy atom derivatives solutions, and cocrystallization of the protein and heavy atoms.

MIR requires a lot of material, since many crystals often need to be screened to yield the appropriate derivatives. Just like the search for crystallization conditions, it cannot be known *a priori* which derivatives will be useful. Often derivatization alters the unit cell parameters in such a way that the isomorphism is not maintained

1.2.4.2 Anomalous Scattering Method. This second method of solving the phase problem takes advantage of having a heavy atom incorporated into the protein. Three very important technical developments make this method possible: variable wavelength of synchrotron radiation, cryocrystallography, and the production of proteins that contain selenomethionine instead of sulfur-containing methionine [55]. For proteins that naturally contain a heavy atom, such as the iron in globin or cytochtrome, the native heavy atom provides the source of anomalous dispersion.

Proteins lacking functional heavy atoms, but which naturally (or through introduced point mutations) encode for methionine, can be expressed in *E. coli* containing exclusively selenomethionine. Selenium atoms serve as heavy atoms in a protein that is essentially identical to the “native” form. Isomorphism does not represent a problem since same protein serves as both native and derivative form. This method requires a tunable X-ray source, generally provided by a synchrotron X-ray beamline. The incorporation of selenium allows a special type of experiment to be done in which data are collected at the wavelength at which the selenium atoms absorb X-rays as well as at wavelengths at which they do not. This sort of experiment is known as *a multiwavelength anomalous dispersion* or *single wavelength anomalous dispersion* (MAD and SAD, respectively) [75-77]. Selenium atoms are located based on differences in the intensities of the reflections, just as similar differences allow the determination of heavy atom positions. The advantage over MIR, which requires multiple crystals, is that in this case the experiment can be conducted with a single crystal. Also, for MIR the treatment of the crystals with the heavy atoms frequently causes serious degradation in the diffraction quality. This issue is completely avoided with the MAD or SAD experiments. The quantity of selenomethionine residues allowed in a single protein or protein complex can vary, but the largest number in a structure solved by MAD was six per protein subunit, in the dimeric enzyme DsbA [55].

1.2.4.3 Molecular replacement: related proteins as phasing models. The third method of solving structures is by *molecular replacement*. This technique, developed by Rossman and Blow [78], takes advantage of the fact that the basic

backbone architecture of many proteins of interest is similar to other proteins or protein domains for which there are already solved, high resolution structures. In this case the known structure of the macromolecule is referred to as the *phasing model.* The model and the molecule of interest should share at least 40% homology. This method is extremely useful in examining a series of mutations to see how they perturb a structure or to see how drugs might bind and interact with a target. In general, the target and the model backbone atoms must be within ~2.0 Å root mean squared deviation (RMSD).

If the model and the new protein are isomorphous, then the phases for the new protein can be calculated directly, after which they are improved during an iterative process. If they are not isomorphous the problem is more complicated, and the position and the orientation of the phasing model in the unit cell are searched [55]. One of the best ways to do this is through calculating the *Patterson map*. The Patterson function $P(u,v,w)$ is a Fourier series used to calculate $\rho(x,y,z)$ from structure factors. The coordinates (u,v,w) locate a point in the Patterson map in the same way that coordinates (x,y,z) locate a point in an electron density map. A Patterson map shows peaks at locations corresponding to vectors between atoms, and is independent of the position of the structure in the unit cell. But if the structure is rotated in the unit cell, the "Patterson map rotates around the origin, changing the position of Patterson atoms in a single Patterson unit cell"[55]. This has important implications for the use of the Patterson map in determining the best orientation of the model in the unit cell of the new protein.

Different programs are available that perform a combination of rotational and translational searches, or both. For the orientation search, which is also known as *rotational search*, the program is looking for large values of the model Patterson function *P*model(*u,v,w*) at locations corresponding to peaks in the Patterson map of the desired protein. The resolution range used for rotational searches is usually within 10–3.5 Å, with 8–4 Å being the most common. For the location search, which is also known as *translational search*, the program is looking for the "correspondence between the expected structure-factor amplitudes from the model in a given trial location and the actual amplitudes derived from the native data on the desired protein."

The parameter used to determine the success of the search procedure is the R-value [55]. For proteins, R-values of 0.3 to 0.4 for the best orientation of the phasing model have proven successful in providing initial estimates of phases. If they are superimposed, then the phases of the model can be used as initial estimates and improved *via* an iterative process.

1.2.5. Solving and refining a structure. Once the phase problem has been solved, than the electron density function ρ(*x,y,z*) can be calculated from the measured intensities. This is only the beginning of an iterative process that attempts to improve the quality of the initial map by different methods, which inlcude solvent flattening, noncrystallographic symmetry averaging, or introducing a partial atomic model (which is building the structure into electron density maps). If the initial electron density maps are good, protein structural features such as α-helices, β-sheets, side chains, and cofactors can be

recognized. The procedure involves matching the covalent structure of the molecule (i.e., the protein sequence) with the density.

How easy or how hard this step is depends on the resolution to which the crystals yield good data. Different protein structural features require different resolutions. The definition of individual atoms requires data at 1.5 Å resolution or better. At ≤ 2.0 Å resolution, the electron density of individual side chains will be sufficiently resolved to define specific conformers. At 3.0 Å resolution or lower, major structural features such as α helices and β sheets are clearly distinguishable, but many side chains may not be resolved. Another problem is that the electron density map may not be of uniform quality, and some parts of the protein may be poorly ordered and not visible at all.

With high-resolution data (≤ 2.0 Å) interpretation is fairly straightforward. At lower resolutions, ambiguities may exist that require multiple rounds of model building and refinement. In both cases the basic procedure of map fitting employs placement of atoms for the interpretable density in the map (i.e., given 3-D coordinates within the asymmetric unit); the positions are then refined against the data, new maps are calculated using the new phase information from the model, and the procedure is repeated. This iterative building and refinement process gradually improves the phases and the quality of the maps such that features that may not have been visible at the outset become visible. In cases where the data are ≤ 2.3 Å, automated building and refinement programs can do much of the work. At lower resolutions, the work still requires building and map interpretation to be done manually. It should be noted that the precision of the

placement of the atoms in a macromolecular structure is typically much greater than the resolution of the data. The positional errors in the core regions (i.e., those that are best defined) range from 0.1–0.2 Å at 1.5 Å resolution to ~0.5 Å at 3.0 Å resolution. Why is this precision better than the diffraction limit? The answer is chemistry. We know the average lengths and angles for all of the types of covalent bonds that hold together a protein. These values are included as restraints in the refinement procedure and ensure that the final structural model makes good chemical sense.

Another important parameter that defines data quality is the R factor. It shows how well a final model fits the experimental data. Fortunately, the data from an X-ray diffraction experiment are redundant to some degree. Thus, 5%–10% of the data can be actually excluded from the entire refinement procedure and this set used (known as the R_{free} set) as an unbiased metric of how correct the structure refinement is. During the structure refinement, the crystallographer compares how well the structural model predicts the R_{free} dataset.

It is now routine to report the R and R_{free} values for a crystal structure. These two numbers serve as one metric for how correctly things have been done. R and R_{free} should be similar (R_{free} is always higher by ~2%–6% for well-refined structures). During the refinement process, the value of R may decrease, but R_{free} will stay the same (or increase). If the R values are already low, this is one way for crystallographers to know when to stop refining. Large differences between R and R_{free} indicate that some portion of the model is incorrect and that something needs to be corrected before the refinement can be completed.

Examination of stereochemistry and Ramachandran analysis provide two other good measures of the quality of a structure. In good structures, one should see RMSD for bond lengths of < 0.02 Å and bond angles of < 2°; curiously, lower values than these do not necessary indicate higher quality structures but may reflect the use of too-tight constraints during refinement. Ramachandran analysis examines which parts of conformational space the protein is in.

Although the final refined model represents the average positions of the atoms, the structures do contain information about the mobility of different parts. One indicator is a parameter known as the *B-factor*. Each atom in the structure has a B-factor value that describes the average mean displacement from the position seen in the structure. Low B-factors show areas of low mobility and high B-factors show regions of higher mobility. This simple interpretation can break down in low-resolution structures (> 2.7 Å), as these generally have higher average B-factors that also reflect the general disorder in the lattice.

1.2.4 Finale and future: integrated studies. The methods outlined in this chapter, which have been utilized in the research presented in this dissertation, offer a powerful way to gain valuable molecular insight into the fine anatomy that underlies how macromolecules look and behave. Besides the explanatory power that such approaches offer for understanding the native structure and function itself, they are often the first step in protein design, allowing the engineering of variant proteins with novel function, and they also provide guidance for drug discovery and development, as commonly employed in “structure-based design” (or “rational design”) methodologies. Although each method offers complimentary

information, integrated studies that use various methods to get at different levels of structural questions offer a great advantage for exploring large macromolecular complexes in different functional states.

CHAPTER 2

PDZ PROTEIN INTERACTION DOMAINS

The central theme of the thesis project is multivalency of biological systems and the core study was done on the dual PDZ domain of PSD-95 protein. Special focus was on utilizing ITC technique for analyzing affinity of protein–protein interactions. This lead us to design of potent molecular probes to target these interactions. Ligand design and subsequent in vivo studies are also described in this chapter.

Different cellular processes are regulated by proteins through their various interactions with other polypeptides, small molecules, phospholipids or nucleic acids [79]. Many of the associated regulatory proteins are constructed of different *protein interaction domains*, or PIDs [80]. In structural biology or biochemistry the definition of a 'domain' is somewhat broad, although usually they are described as relatively independent (structurally and/or functionally) structures of 35-150 amino acids that retain their fold and binding characteristics even when they are expressed separately. A striking but not unusual structural feature is that the N- and C- termini are often located very close, on the opposite side of the ligand-binding site. This arrangement enables the assembly of PIDs into different host proteins while keeping their binding affinity very similar. Utilizing such "building blocks" introduces diversity in complex proteins with a limited number of components.

Many protein domains are known to bind to different sequences in their target molecules. Another interesting feature is that different interaction domains are present in the same host molecule (single or multiple copies) thus forming a complex protein that can mediate multiple protein-protein or protein-ligand interactions. Based on their sequence homology or their ligand-binding properties, domains are organized into different groups; for example, SH2 (Src Homology 2) [81], SH3 (Src Homology 3) [82], WW (two conserved tryptophan residues) [83], PTB (Phosphotyrosine binding domain) [84], PDZ (PSD-95, Discs-large, ZO-1), LIM (Lin11, Isl-1, Mec-3) domain, *etc [80]*.

2.1. The PDZ Domain. PDZ domains are members of the same PID family of proteins, and likewise are relatively small, independently folded 'modules' that occur in longer polypeptides [85-87]. As with the other members of the PID family, PDZ domains are largely responsible for promoting transient non-covalent protein-protein interactions. These occur primarily within the intracellular space of a cell, and represent a number of different signaling pathways that affect all manner of cellular functions, from growth to death.

The PDZ domains themselves are among the youngest of the identified PIDs. They were discovered on the basis of sequence repeats first found in three different proteins: postsynaptic density 95 (PSD-95); Disc Large (DLG) and zonula occludens 1 **(**ZO-1) [88]. Initially these regions of polypeptide were referred to as *GLGF domains [85]*, because of the highly conserved Gly-Leu-Gly-Phe motif found within the amino-terminal sequence, or *DHR domain* for discs large homology region. Later, the acronym “PDZ” rose to dominance (from **P**SD-

95, **D**lg and **Z**O-1, the initial letters of the first three proteins in which they were identified), and was widely adopted by the scientific community. Since the sequences of several genomes have been determined, it has been shown that PDZ domains occur in hundreds of proteins, and are among the most abundant protein domains in multicellular eukaryotic genomes, *Caerohabitis elegans* (89), *Drosophila* (128) and *Homo sapiens* (320).

2.2 PDZ Domain Specificity and Nomenclature. PDZ domains consist of approximately 100 amino acid residues and bind in a sequence-specific manner to the carboxyl terminus of their protein-binding partners [85, 86]. While almost exclusively recognizing peptides with free carboxyl termini, some PDZ domains can also interact with internal peptide sequences, and in at least one case, with a non-C terminal cyclic peptide [89]. PDZ domains are predominantly found in cytoplasmic proteins, in which they occur in one or multiple copies. These protein domains are widespread in organisms, ranging from simple bacteria to complex eukaryotes, including humans, indicating that “they are either hyper adaptable or that they are essential or fundamental to cellular processes” [90]. Their high conservation level throughout the species indicates that their molecular function has been retained since the beginning of cellular life. These modular protein domains show no catalytic activity and preferentially bind to C-terminal sequences of the target protein, like cytoplasmic tails of transmembrane receptors and ion channels [91, 92]. However, it has been observed that PDZs can bind in non-canonical fashion as well to internal sequences, cyclic peptides or even lipid molecules [89, 93]. As stated before, PDZ domains are found in

multiple copies in complex proteins, often in combination with other domains like SH3, WW, LIM [94], *etc.* A recent search in the SMART (Simple Modular Architecture Research Tool) database found 20,072 in 12,617 proteins. PDZ proteins can be classified into three principal families based on their modular organization (Figure 2.1).

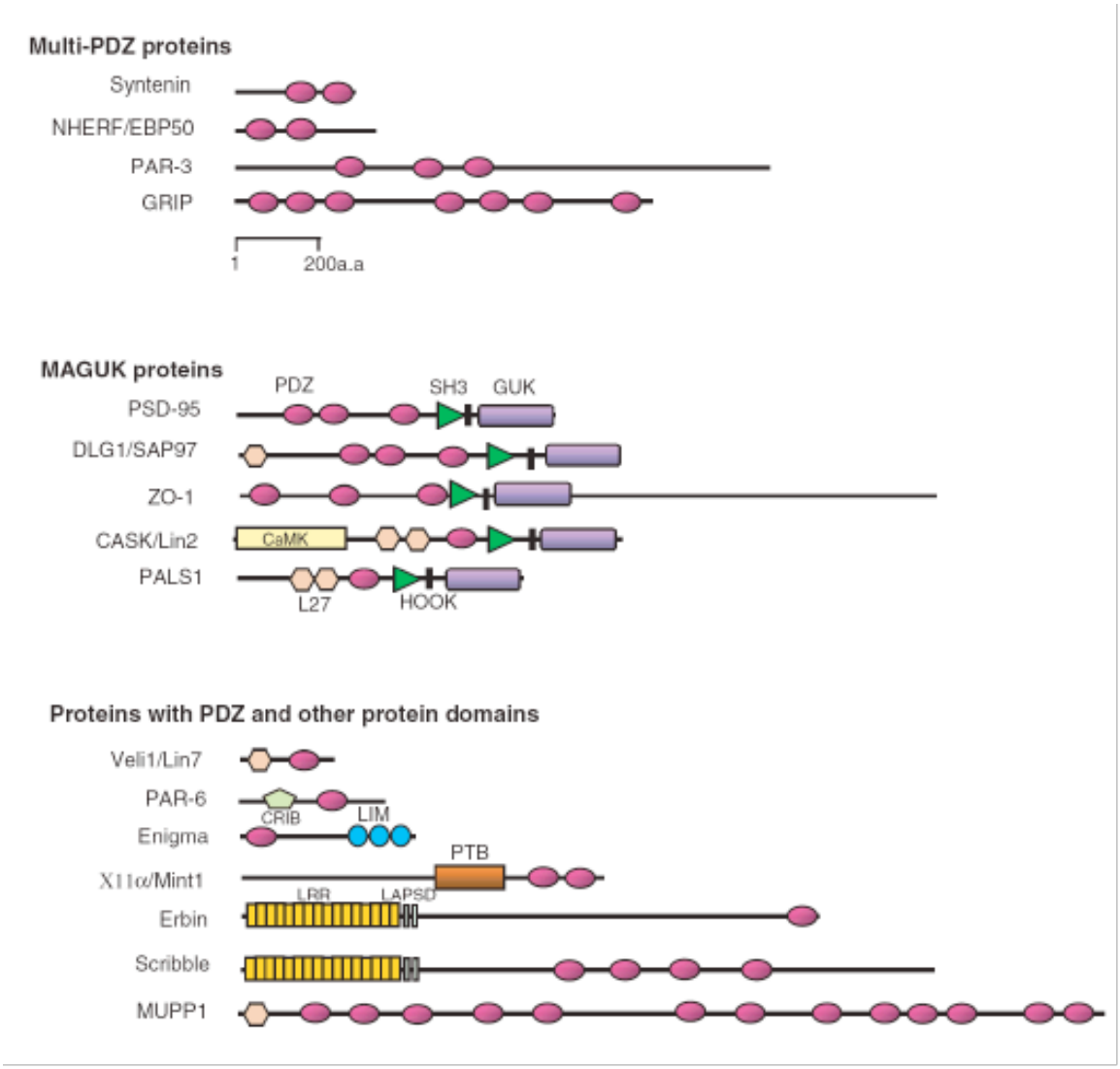

Figure 2.1 Examples of PDZ domain classification according to modular organization**.** Reprinted by permission [94]

The first family contains proteins consisting entirely of PDZ domains, though their number may vary from two to more than ten in certain proteins. The

second family is the MAGUK (membrane-associated guanylate kinase) proteins. These are a well studied family of PDZ containing proteins, particularly the mammalian MAGUK proteins that include PSD-95/SAP-90, PSD-93/chapsyn-110, SAP-97/hdlg, and SAP-102 [94]. They are found at synapses and behave as protein-protein interaction modules that target and cluster ion channels, receptors, and cell adhesion proteins. MAGUKs all share a common domain organization, consisting of (one or three) PDZ domains at the N-terminus, followed by an SH3 domain, and at the C-terminus a guanylate kinase-like (GK) domain. PSD-95, the focal protein of the PDZ domain research presented in this dissertation, belongs to this family [95]. Overall it is probably one of the best studied PDZ domain-containing proteins to date, if not *the* best studied. The third family encompasses proteins that, in addition to PDZ domains, contain other domains, such as LIM, L27, C2, PH, WW, DEP and LRR [96].

2.3 The Significance of PDZ domains

PDZ domains have been found to perform a wide variety of functions on different cell and tissue types, and many of those occurrences are associated with disease states. Thus, the importance of these domains spans from their basic endogenous functions all the way to disorders that are associated with protein-protein interactions that involve binding (or disruption thereof) of PDZ domains. A brief overview is provided here that exemplifies the range and significance of these domains, especially as it applies to human health.

2.3.1 PDZ domains as scaffolding proteins. A well documented property of PDZ domain-bearing proteins is their ability to serve as scaffolding elements for

the assembly of macromolecular complexes at specific subcellular locations, particularly at the cell surface [97, 98]. PDZ domains have a specific fold and interact with their binding partner proteins through the latter's variable C-terminal region. Along with their often being accompanied in their host protein by one or more different PIDs, such as SH2, SH3, WW and PTB, PDZ domains are usually present in multiple copies. This enables them to assemble and anchor large multimeric protein complexes. A primary role of such protein assemblies is to maintain signal transduction specificity and efficiency. Sheng *et al*. have described some of the best known systems where PDZ domains are involved [96]. A representative example of a PDZ domain protein acting as such a scaffold in the cell is INAD (inactivation-no-afterpotential D), which contains five domains that interact in the phototransduction network [99].

2.3.2 PDZ domains and their roles in cancer development. Numerous PDZ domain-containing proteins have been shown recently to play a role in cancer progression, and thus emerge as potential drug targets [100]. Syntenin, a structurally well characterized protein that contains two PDZ domains, represents one such example [100]. Syntenin is overexpressed in many metastatic breast and gastric cell lines. Since syntenin is localized in the area of cell-cell contacts that act as cell adhesion sites, it might be involved in tumor cell invasion. In support of this mechanism, forced *in vitro* overexpression in MCF7 (non-metastatic breast cancer cell line) or poorly metastatic AZ-521 gastric cancer cells increases the invasion and migration properties of those cell lines [100]. As shown in work by Boukerche *et* al [101], altering syntenin expression leads to

modifications in phenotype and alters the metastatic capabilities of melanoma cells.

Another PDZ domain-bearing protein, PCD1 (pancreatic cancer derived), is found to be highly overexpressed in different types of tumor tissues, such as colon, breast, liver, pancreas, stomach and prostate [102]. Rarely is PCD1 expressed in normal tissues. Further studies have shown that the *PCD1* gene may play a role in colorectal cancer development from adenomas [103]. The same group has shown that PCD1 is implicated in breast cancer progression and nodal metastasis [104]. Therefore, this protein is a potential candidate for diagnosis and treatment of these types of cancer.

A*ctivated in prostate cancer* (AIPC) is another PDZ domain protein containing six such domains, and is highly expressed in prostate tumor tissues The initial immunohistochemical studies have linked accumulation of this protein with the promotion of prostate tumorigenesis [105].

Finally, another cancer-associated protein is MAGI-3, a member of the MAGI family of PDZ domain-containing proteins that interact with proteins from viruses like human T-cell leukemia virus type 1 (HTLV-1) and high-risk human papilloma-virus (HPV) types. MAGIs can be viewed as a distinct subgroup of the MAGUK family. They have three PDZ domains, but instead of an SH3 domain there is a WW domain and guanylate kinase (GK) is found at the N-terminus. MAGI-3 has been shown to interact through its PDZ domain with Tax1 viral protein in HTLV-1 infected cells. This binding event alters the cellular localization of these proteins and might have implications for their biological activity, leading

to continuous infection of HTVL-1 in T-cells. Therefore, MAGI-3 is a very interesting target for understanding the mechanism of these dangerous infections, and ultimately for the design of inhibitors that could control them.

2.3.3 PDZ domain proteins in cystic fibrosis. Cystic fibrosis (CF) is the most common lethal genetic disease in Caucasians [106]. The CF conductance regulator (CFTR) is a low conductance chloride channel that mediates cAMP dependent chloride secretion [107]. Changes in CFTR function are associated with CF, secretory diarrhea and pancreatitis. Mutation of the gene encoding CFTR reduces its chloride transport capacity and/or its level of cell surface expression. Another disease caused by the overstimulation of CFTR activity through bacterial enterotoxins is secretory diarrhea, the second largest cause of infant mortality in developing countries [106].

The mechanism of action is mediated through the C-terminal tail of CFTR that directly binds to the PDZ domain proteins NHERF (Na^+/H^+ exchanger regulatory factor) [108] and EBP50 (ezrin-binding protein) [109]. It has been shown that binding of tandem PDZ domains to the C-terminal of CFTRs has a modulatory effect on the channel activity, revealing a new function of PDZ domains, which might not be limited only to this complex. It has been reported that regulation of CFTR with the PDZ domains of NHERF involves phosphorylation of one PDZ domain. This mechanism is under control of protein kinase C (PKC) [110]. PKC phosphorylates PDZ2 of NHERF, which triggers a conformational change that results in loss of affinity for CFTR and inability to activate the channel [110].

2.3.4 PDZ domain involvement in neurodegenerative diseases. Early studies of PDZ domains were especially interested in a single such repeat in CASK protein (calcium/calmoudlin-dependent serin protein kinase, a mammalian homologue of Lin2) [111]. Through its PDZ domain, CASK is shown to interact with parkin, a protein involved in progression of Parkinson's disease [111, 112]. Parkin is an E3 ligase responsible for the ubiquitination of protein substrates as a tag or marker, a critical step in the protein degradation pathway. It is believed that CASK acts as a scaffold enabling cellular trafficking and localization of parkin. However, it remains unclear what the consequence is when this interaction is lost. Designing ligands to selectively disrupt the interaction would be one possible way to study the associated complex functional properties, and subsequently might lead to developing therapeutics for this neurodegenerative disorder.

Another very interesting example is GRIP1 (glutamate receptor-interacting protein 1), which is shown to have a direct functional link with the Fraser syndrome protein Fras1 [113]. GRIP1 contains seven PDZ domains, and through them it can form a diverse network of protein-protein interactions. This adaptor protein has been implicated in the regulation of neuronal synaptic function, but its physiological roles have not been defined *in vivo*. Thus far studies in mouse model have shown that loss of GRIP1 leads to epidermal blistering, renal agenesis and defects in the brain, a condition similar to human Fraser syndrome [113, 114].

A novel protein, human like X11-like (hX11L), contains a phosphotyrosine interaction (PI) domain and two PDZ domains [115]. Through the PI domain, hX11-L interacts with a sequence containing the same motif found in the cytoplasmic domain of Alzheimer's (amyloid precursor protein APP). Further analysis of binding to APP of different hX11-L constructs indicates that PDZ domains at the carboxy-terminal region are essential for modulating APP degradation [115]. These observations should contribute in the current understanding of the molecular mechanism of APP in Alzheimer's disease.

After reviewing all these therapeutic areas that have been summarized, it is evident that the PDZ domain is a critical player in a variety of protein binding interactions. What unites these incidents is a molecular recognition event between the partner protein and the PDZ domain itself. Focusing attention on this, what is revealed is that there is both a conformity and diversity in the peptide sequences that different PDZ domains bind, and which gives rise to how these structures are classified.

2.4 Classification of PDZ domains. PDZ domains predominantly bind to C-terminal sequences of their binding partners [116]. The mode of recognition represents the basis for their derived nomenclature, describing residue position away from the carboxy end, and which was adopted for their binding targets. C-terminal residue is referred to as the "0" position with the subsequent residues labeled "-1", "-2", "-3", etc. Classification by Songyang *et al*. showed that PDZ domains can be classified based on the preference for certain amino acid residues at key positions "0" and "-2" in their sequence (Figure 2.2). *Class I* PDZ

domains prefer peptides with small hydrophobic residues at the “0” position and a β-hydroxyl group (Ser or Thr) at the “-2” position, while *class II* PDZ domains prefer hydrophobic or aromatic residues at this position. Most of the PDZs fall into these two classes, but later on novel domains were discovered that have different preferences and that led to establishing *class III* and *class IV* categories as well (Table 2.1).

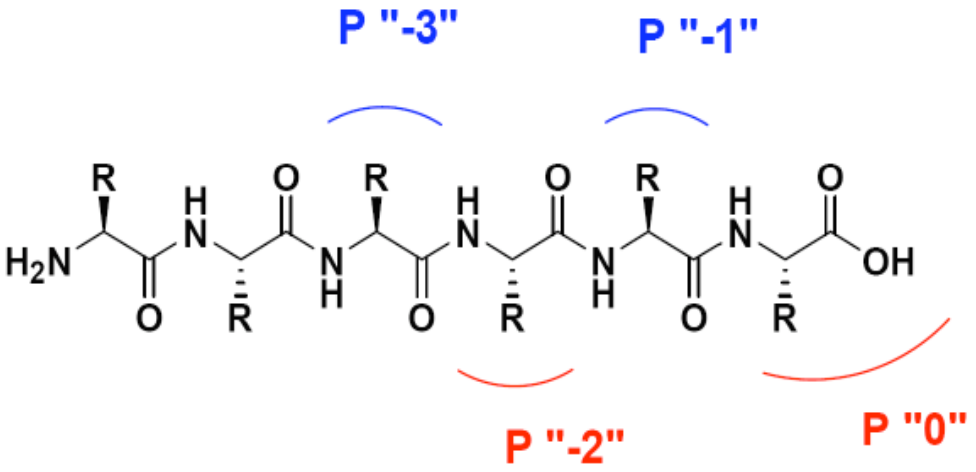

Figure 2.2 Schematic diagram of peptide orientation in the PDZ binding site. P“0” and P”-2” point toward PDZ binding pocket while P”-1” and P”-3” are solvent exposed.

Class III PDZ domains bind to their partner ligands with Cys at position “0” and any amino acid residue at the “-2” position. According to the latest classification *class IV* PDZs recognize Val at the “0” position and a negatively-charged residue (Asp or Glu) at “-2”. However, there are quite a few PDZ domains that cannot be listed in any of these classes, since they show unique binding patters.

A different study has taken into account the nature of amino acids in two critical positions. First would be the position that immediately follows the second

β-strand (βB5) and another would be the first position in the second alpha helix (αB1) of the binding pocket [117]. This method was successful for classification for the limited number of PDZ domains, however not for all, since the criterion based on nature of amino acids at these two positions was insufficient to classify all PDZs.

	Position **"-3"**	Position **"-2"**	Position **"-1"**	Position "**0**"
Class I	X	Ser/Thr	X	Val/Leu/Ile
Class II	X	Phe/Tyr	X	Phe/Tyr
Class III	X	X	X	Cys
Class IV	X	Asp/Glu	X	Val

Table 2.1 Classification of PDZ domains based on the preference in amino acid sequence of their binding partners; X stands for multiple possible standard amino acids.

The idea of Bezprozvanny's work [117] can be utilized by expanding the number of residues in the binding pocket of the domain that are taken into consideration. Thus proposed classification of PDZ domains by Bezprozvanny and Maximov was listed in SMART website (yielding 25 groups overall) [118].

Insight into structure of both class I and II PDZ domains reveals that they have a very similar fold of six β-sheets and two α-helices despite their differences in amino acid sequence. Protein domain structure is the property that is conserved, not its sequence [90].

2.4.1 Class I PDZ domains. The majority of studies on PDZ domains was done on class I PDZs. Doyle et *al*. [88] pioneered the work by solving the crystal structure of PDZ3 of PSD-95 bound with the peptide derived from C-terminal of CRIPT (cysteine rich interactor of PDZ domain).

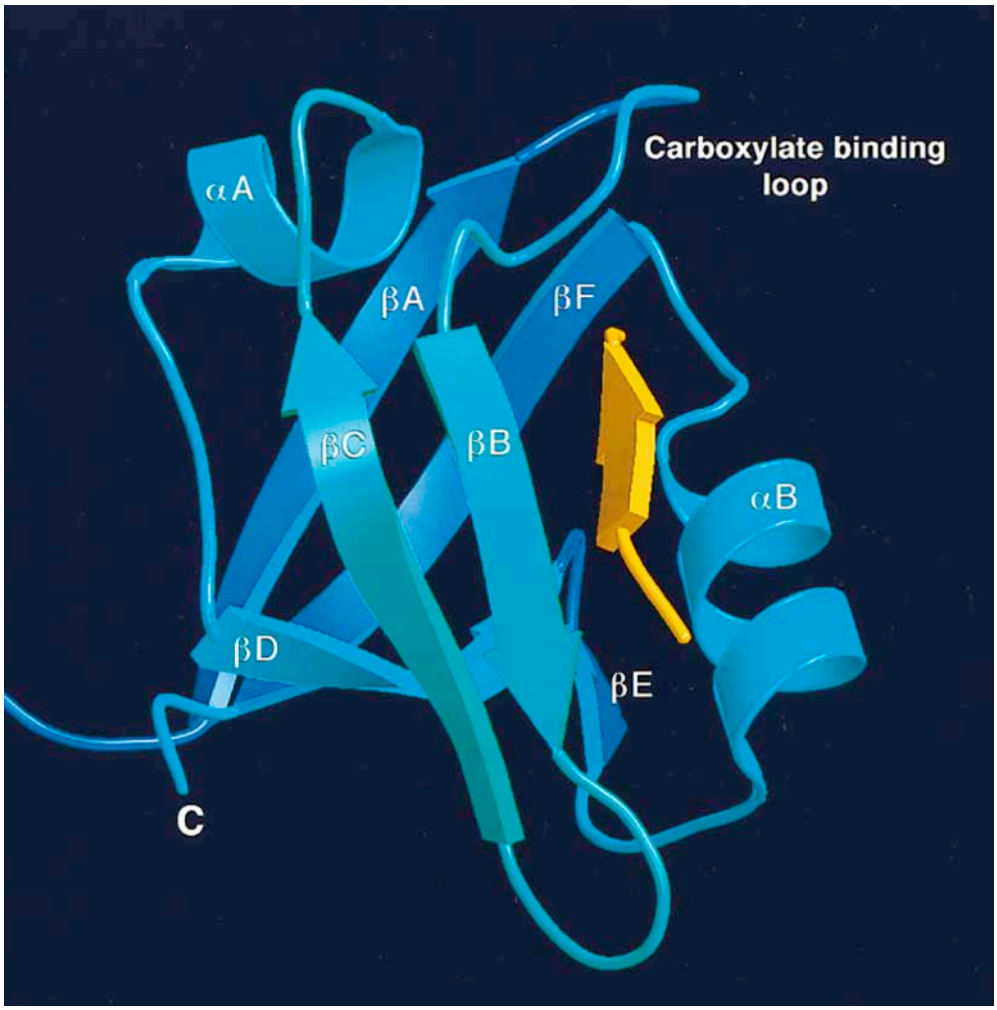

Figure 2.3 Structure of the PDZ3 domain (blue) from PSD-95 complexed with peptide TKNYKQTSV (golden) derived from C-terminal of CRIPT. Only four C-terminal residues (QTSV) were resolved in the crystal structure. The picture is reprinted from Doyle *et al.* with the permission from Elsevier.

Published crystal structure of PDZ3 (Figure 2.2) is comprised of six β strands (βA- βF) and two α-helices (αA and αB), which fold in an overall six-stranded β sandwich.

Four residues of the ligand's C-terminal region have been shown to interact with the peptide binding groove between the βB strand and αB helix as an antiparallel β strand extending one of the β sheet. The C-terminal hydrophobic residue valine at the "0" position of the peptide binds to the hydrophobic pocket created in the PDZ domain by several hydrophobic residues, which are Leu 323, Phe 325, Ile 327, and Leu 379. Since this pocket is the only one present on the binding surface of the domain, authors hypothesize that selection of biding parthner is done based on the volume of this pocket. The tight contact of the side chain of valine with the hydrophobic pocket explains the preference for the sequence having a hydrophobic residue at the "0" position [88]. Another critical interaction is the hydrogen-bonding interaction of the hydroxyl group of the "-2" position Thr with N3 of His 372 (first residue of the αB helix). The side chain of the "-1" Ser residue is shown pointing away from the interacting surface.

The general feature of PDZ domains to bind to C-terminal sequences is supported by the stabilization of carboxylate through the series of H-bonds to the so called carboxylate binding loop formed by the conserved Gly-Leu-Gly-Phe (GLGF) and an Arg 318 through a highly ordered water molecule.

Superposition of crystal structures of the peptide-free and peptide-bound PDZ3 domain of PSD-95 showed very little conformation change leading to peptide binding event

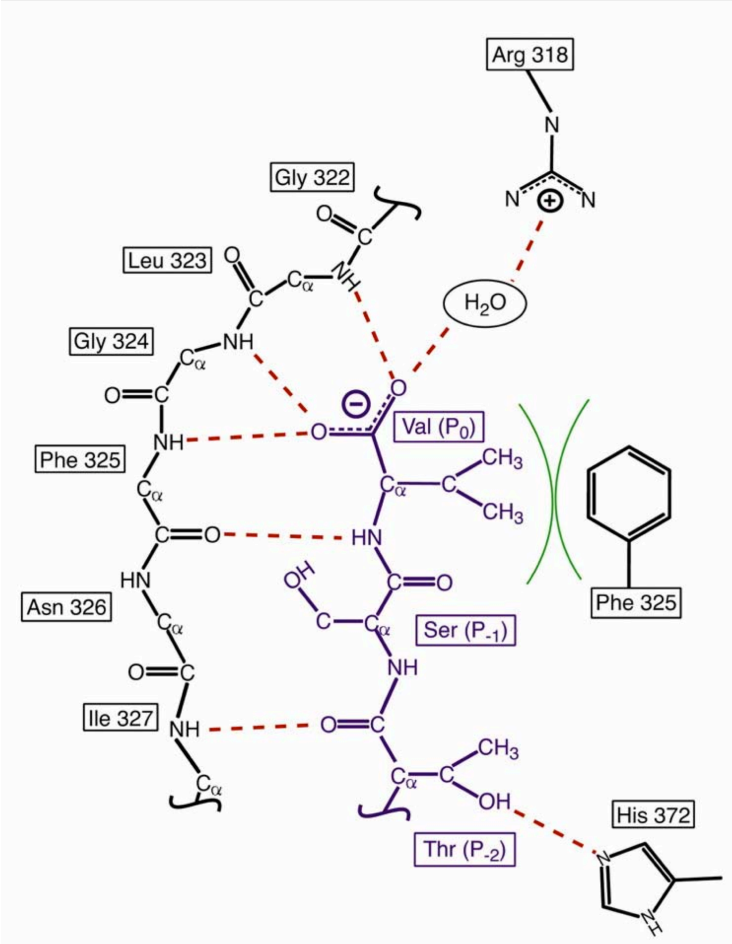

Figure 2.4 Mechanism of peptide recognition by PDZ domain. Reprinted figure from Harris *et al*.[119]

2.4.2 Class II PDZ domains The class II recognition mechanism was first studied on crystal structure of hCASK/LIN-2 PDZ domain [120]. The structure of hCASK/Lin-2 PDZ domain is consistent with the typical PDZ fold formed by six β–strands and two α-helices shown also for class I PDZ domains, even though these classes only share a low percentage (20%) of sequence identity. An unexpected feature of this structure is that the peptide binding pocket of this domain is occupied by the C-terminal sequence of a neighboring non-

crystallographically related molecule. Peptide binding site consists of two hydrophobic pockets for residues at positions “0” and “-2” which is consistent with the class II binding sequence: X-ϕ-X-ϕ (where X is any amino acid and ϕ is a hydrophobic residue). Binding pocket can accommodate smaller and larger hydrophobic groups.

After closely examining the binding pocket, it is noted that “-2” position is formed by residues Met 507, Val 549 (at position αB1) and Leu 552. Among these, Val 549 is highly conserved in the class II PDZ domains. This residue is replaced by histidine in class I PDZ domains and forms an important H-bond with the Ser/Thr residue at “-2” position of their binding partner. Another interesting feature of this domain is that the carboxylate binding loop is structurally very similar with that of class I, although sequences are very different (Pro-Met-Gly-Ile in the hCASK instead of Gly-Leu_Gly-Phe). Despite these sequence differences, both class I and II PDZ domains are structurally similar and form H-bonding through an ordered water molecule.; yet, they bind to completely different target sequences.

2.4.3 Non-C-terminal PDZ domain-peptide binding Non-canonical binding of PDZs to cyclic peptides was first discovered through the phage dislay library containing peptides cyclized via an intramolecular disulfide bond [89]. Later reports have shown that PDZ domains can bind to internal sequences in other PDZ domains [121, 122]. Both these complexes are characterized by interactions of PDZ domains with “pre-organized β-finger motifs” in their partners. This β-finger motif is formed by an extension of the canonical PDZ sequence that

contains around 30 amino acids. The β-finger motif binds into the traditional peptide binding site of the PDZ domain and its first strand mimics a C-terminal peptide binding. The terminal carboxylate of a peptide ligand is replaced by the sharp β turn at the tip of the nNOS β-finger. The loss of the important interaction of the C-terminal carboxylate with the GLGF motif in the PDZ domain is most probably compensated by the tertiary interactions between these two PDZ domains. Evidence that supports this hypothesis is the calculation of the surface area buried upon binding of the β-finger motif of nNOS PDZ domain to the PDZ domain of syntrophin, which comes to 800 $Å^2$. This value is twice as much as the burial of the surface area for binding of a C-terminal peptide ligand. This type of interaction might be typical for other PDZ domains than PDZ of nNOS if they fulfill the requirements for the β-finger motif in their structure.

2.5 Ligand recognition by PDZ domain. As previously stated positions “0” [116, 123] and “-2” [88, 124] are the critical residues for the recognition of the peptide sequence by PDZ domains. Evidence is emerging that other residues play role in the specificity of the peptide recognition [116]. Structural data of PDZ domains complexed with peptides [88, 120, 121] reveals the key components for the molecular interaction event; (1) stabilization of C-terminal carboxylate group, (2) binding of the “0” residue at a hydrophobic pocket, (3) hydrogen bonding of peptide backbone with the side chain of different residues in the PDZ domain, and (4) binding of the “-2” residue with the conserved His residue (the first residue in the αB helix) in PDZ domain.

PDZ domains have a unique ability to stabilize the C-terminal carboxylate

group. They do so due to the presence of carboxylate-binding loop containing the conserved motif R/K-X-X-X-G-L-G-F. The peptide C-terminal carboxylate is coordinated by a network of hydrogen bonds with main chain amide group of this loop. The carboxylate group forms hydrogen bonds with conserved Arg or Lys mediated through an ordered water molecule. A mutational study in PSD-95 and syntrophin PDZ domains has revealed the importance of the identity of conserved but interchangeable Arg or Lys in peptide recognition [125]. The result also indicates that the identity of the residues at this position in the carboxylate-binding loop is critical for the interaction of one PDZ domain with another PDZ domain. PDZ domains with Lys at this position (PDZ2 of PSD-95, PDZ domains of both α- and β-syntrophins) dimerize with the PDZ domain of nNOS (neuronal nitric oxide synthase) while PDZ with Arg at this position (PDZ1 and 3 of PSD-95, PDZ of γ1-syntrophin) restricts the dimerization with PDZ of nNOS [125].

N-methyl amide scanning of the peptide backbone shows that hydrogen bonding plays an important role in the PDZ ligand interaction [126]. The order of affinity of the N-methyl amide peptides suggests that amide hydrogens at the “0” and “-2” positions are more important than at other positions, consistent with the hydrogen bonding inferred from the structural information of peptide-bound PDZ3 of PSD-95 [88]. Replacing the carboxylate group with neutral alcohol and hydroxamide group shows that electrostatic interaction of carboxylate group andconserved Lys (in MAGI-3 PDZ2) is more important than preserving the size of the “0” residue.[68] This study also points out that the interaction between the “-2” Ser/Thr hydroxyl group and the His side chain (from αB1 of the PDZ domain)

is strictly regulated by the approximate size of binding pocket as peptide side chain modifications at “-2” position. Such modifications include inversion of chiral center, branching, and elongation; all lead to abolition of binding.

Requirement of a free terminal carboxylate group for the ligand led to the belief that electrostatic interaction of the C-terminus with the domain plays an important role in recognition. But examining the salt dependence of both terminal and internal ligands and the effect of mutation of a conserved basic residue (thought to be important in electrostatic recognition) in syntrophin PDZ showed that direct electrostatic interaction with the terminal peptide do not play a significant part in binding specificity [127]. Binding analysis of a series of chemically modified C-terminal peptides with syntrophin PDZ demonstrates that steric features and hydrogen bonding complementarity are vital for recognition of the free carboxy group.

One of the major projects in our laboratory is based on the design of macrocyclic ligands for PDZ domain using structure-based information [128, 129]. Cyclic peptides show higher stability to proteolytic cleavage in *in vivo* studies, along with significant improvement in entropic contribution to the binding affinity[130]. Another approach utilized a peptide library containing nonstandard amino residues at position “0” was used to improve the ligand affinity by optimizing the size of hydrophobic group at this position to best fit the binding cavity [131].

2.6 Affinity of PDZ domain-mediated interactions. Binding affinities for the PDZ-mediated interactions have been reported in studies that used both solution-

based methods (fluorescence [105, 132-135] and ITC [128, 129, 136, 137]) and solid-phase methods (surface plasmon resonance [116, 138, 139] and ELISA based assay [140]). Predominantly, solution methods are gaining lower values of affinity values (in the micromolar range) as opposed to the values from solid phase methods (in the nanomolar range). Such findings are attributed to solid phase methods possessing artifacts that lead to overestimation of binding affinity[141, 142]. PDZ domains are involved in transient, reversible interactions with other macromolecular targets, these low micromolar interactions are consistent with their regulatory function [143]. However, it is worth noting that actual *in vivo* values may be affected by local concentration in the subcellular compartments or the cellular enviroment (molecular crowding phenomenon).

2.7 Regulation of PDZ domain interactions. Regulation of protein-protein interaction involving modular domains and their binding partners is often critical for cell signaling [80]. Several mechanisms have been proposed for the regulation of PDZ domain interactions. One is phosphorylation and dephosphorylation of Ser and Thr at the "-2" position in *class I* PDZ domains [144]. Specific phosphorylation at the "-2" position of β2 adrenergic receptor with GRK-5 (G-protein coupled receptor kinase 5) abolishes its interaction with PDZ of NHERF (Na^+/H^+ exchanger regulatory factor) [145]. Similarly, serine phosphorylation at the "-2" position in the inward rectifier K^+ channel Kir2.3 by protein kinase A (PKA) disrupts Kir2.3 binding to the PDZ2 domain of PSD-95 [146] In another example, phosphorylation of the "-3" residue of the AMPA

receptor subunit GluR2 C-terminus by another kinase PKC prevents GluR2 binding to the GRIP PDZ domain [143].

Another mechanism proposed is allosteric regulation, as seen in the case of the Par6 PDZ domain. The binding of Cdc42 to CRIB (Cdc42/Rac interactive binding) domain of Par6 induces conformational changes in the Par6 PDZ domain, leading to enhanced affinity for the ligand [147, 148]. A NMR binding analysis of the hPTP1E (human protein tyrosine phosphatase 1E) PDZ2 domain with RA- GEF2 (Ras-associated guanine nucleotide exchange factor 2) peptide indicates a significant change in PDZ2 dynamics upon ligand binding, and the distal surface in PDZ2 are shown dynamically linked to the ligand binding site [149].

2.8 Materials and Methods.

This section describes all the materials and experimental procedures employed in conducting the research presented in this dissertation.

2.8.1 Materials. Difco laboratories Bacto® Lennox LB Broth and granulated agar were used to prepare the LB/amp50 plates and LB/amp50 Media. Corning® 100 mm/tissue culture dishes were used for the LB/amp50 plates. The ampicillin sodium salt was acquired from Fisher Biotech and a 20 mg/mL stock solution was prepared by dissolving ampicillin sodium salt (2%) in water. The ampicillin stock solution was filtered with a Star LB™ 0.22 μm filter. Corning® 0.22 μm polyethersulfone filters were used to sterilize the LB/amp50 media. The incubators used for bacterial culturing were the Fisher Isotemp® 200 series and the Lab-line® Incubator-Shaker incubators. Handling and plate streaking with glycerol stock, as well as transfer of the liquid cultures, were performed within a Laconco purifier class II biosafety cabinet. The 10 mL and 20 mL LB broth/amp50 cultures were grown in Falcon® Blue Max™ 50 mL conical tubes.

The electrophoresis chamber used for the agarose gels was the Small Horizontal Gel System FB MSU-1 by Fisher Biotech and the gels were visualized on the Transilluminator FB TIV-88 UV Box by Fisher Biotech. The E-C Apparatus Corporation EC 105 electrophoresis power source was utilized. Ethidium bromide was obtained from LKB Instruments Incorporated. The QIAquick PCR Purification Kit and QIAprep Spin Miniprep kit were acquired from Qiagen. Invitrogen® DNA Dipsticks™ were used for quantitation of DNA samples. Eppendorf Centrifuge 5810 R and Centrifuge 5417 C were utilized for centrifugation steps.

DNA primers were prepared by Sigma-Genosys and Invitrogen. The VentR® DNA Polymerase, T4 DNA ligase, and Buffer B were purchased from New England Biolabs. The Wizard® *Plus* Maxipreps DNA Purification System, PCR Nucleotide Mix, Buffer D, Buffer H, acetylated BSA, 1 kb DNA ladder, and the *Bam* HI, and *Sal* I restriction endonucleases were acquired from Promega Corporation. Ampliwax PCR Gem 100 was obtained from Perkin Elmer. The PCR thermocycler used was the Ericomp Inc. Twinblock™ System Easycycler Series. VWR Scientific Products Standard Heatblock and Model 1162 water bath were used. The *E. coli* cells were Epicurian Coli BL21-Gold competent cells by Stratagene. The BL21(DE3)RIL/pRK793 cells were obtained from Science Reagents.

All biochemical reagents were purchased from Sigma, Fischer Scientific, and VWR. Molecular Weight Standard Mixture was obtained from Sigma. Coomassie® Plus-200 Protein Assay Reagent and TPCK Trypsin were obtained from Pierce. Advanced Protein Assay Reagent was obtained from Cytoskeleton. Glutathione agarose beads were purchased from Sigma. DTT and IPTG were obtained from Inalco Spa. SnakeSkin™ Pleated Dialysis Tubing and Slide-A-Lyzer™ Dialysis Cassettes were obtained from Pierce. Amicon™ Ultra-15 Centrifugal Filter Devices were obtained from Millipore. Mini-Protean® III Electrophoresis Cell from BioRad was used for PAGE analysis. HiTrap™ QHP column was obtained from Pharmacia Biotech. The liquid samples were loaded into the chromatographic columns (both for isolation and purification) with the help of Dynamix® peristaltic pump. OLIS-CD spectropolarimeter was used for

CD measurements. Beckman Coulter™ DU® 530 UV/VIS spectrophotometer was used to read the absorbance. Corning PH meter 430 and VWR symPHony electrode were used for pH measurements. Hamilton Gastight Syringes from Hamilton were used for ITC experiments. ThermoVac and VP-ITC Microcalorimeter from MicroCal™ Incorporated were used for ITC experiments. Wizard® *plus* maxiprep DNA purification kit (Promega) was used for the purification of plasmid DNA.

2.8.1.1 Preparation of solid and liquid media. The agar plates were prepared by dissolving Lenox LB (Difco) broth (2%), agar (1.5%) and deionized water. The mixture was autoclaved for 30 min at 121-124 ºC. Once the temperature of the media reached 50 ºC, ampicillin stock solution (20 mg/mL) was added to obtain a final ampicillin concentration of 50µg/mL. Ampicillin stock solution was prepare by dissolving ampicillin sodium salt (2%) in deionized water followed by filtering through Star LB™ 0.22 µm filter to remove any potential contaminants.

The agar solution was then pipetted into Corning® 100 mm tissue culture dishes and allowed to solidify at room temperature. Once solidified plates were inverted, labeled and stored at 4 ºC.

TB (Terrific Broth)/amp50 media was prepared by dissolving Bacto Terrific Broth (4.76%), glycerol (0.4%) in deionized water. To this solution 20 mg/mL ampicillin stock was added to yield a final ampicillin concentration of 50 µg/mL. Finally, media was sterilized by filtering through Corning® 0.22 µm polyethersulfone filters and stored in autoclaved bottles at room temperature.

2.8.2 Methods

2.8.2.1 Expression of fusion protein. A glycerol stock solution of *E .coli* BLR-Gold cells, containing the gene that encodes for the desired GST fusion, was used to streak LB/amp50 plates and incubated for 12 h at 37 ºC. An isolated colony of moderate size was picked out and used to inoculate 10 mL of TB/amp50 broth in 50 mL Falcon tube, and left to grow with shaking for approximately 8 h. This culture was then placed in an autoclaved 1 L flask containing 90 mL of fresh TB/amp50 nutrient media, and incubated in the same manner for 2 h. The culture was then split in half, and 450 mL of the fresh TB/amp50 media was added to each. These cultures were incubated in 2x2 L flasks, with shaking at 37 ºC for approximately 3 h, until the optical density at 600 nm (OD_{600}) value reached 1.2. At this point the expression of fusion protein was induced by IPTG (isopropyl-1-thio-β-D-galalctoside), used at a final concentration 0.8 mM. The incubation was allowed to continue at 30 ºC with shaking for an additional 5 h. The culture was transferred into 250 mL centrifuge tubes and spun down at 3220xg for 20 min at 4 ºC. The supernatant was decanted and the harvested bacterial pellets were stored at -80ºC.

2.8.2.3 Bacterial cell resuspension and lysis. The bacterial cell pellets were resuspended in ice-cold PBS buffer (Phosphate Buffered Saline) containing 1% Triton-X and 10mM β-mercaptoethanol. The Branson 250 Sonifier (Branson Ultrasonics Corporation) was used to lyse the cells by sonication. During sonication (3.2 mm probe, instrument set at 50% duty cycle and output control 5) that lasted for 55 s, cells were kept on ice. The sample was then centrifuged at

10,000x*g* for 20 min at 4 ºC to pellet the cell debris. The supernatant was decanted into a 50 mL Falcon tube. A clear cell lysate was obtained, and kept at 4 ºC.

2.8.2.4 Glutathione agarose beads (GST-affinity) column preparation. Approximately 0.75 mL of glutathione agarose beads (Sigma) were placed in a 50 mL Falcon tube and allowed to swell in 10 volumes of 1xPBS ice-cold buffer for 1 h. The beads were than centrifuged at 10,000x*g* for 10 min, the supernatant discarded, and 5 mL of 1xPBS added to the bead slurry. Pre-swelled beads were then transferred to a glass column (20 mm diameter, 150 mm length). The glutathione bead column was then washed with ten times the volume of 1xPBS and stored in 10 mL of the same buffer (4 ºC).

2.8.2.5 Isolation of fusion protein. Prepared glutathione-bound agarose beads column was washed with 10 volumes of ice-cold PBS buffer. Cell lysate was loaded onto the column at a rate of 5 mL/min, using Dynamix Peristaltic pump, followed by repeated washing with 10 volumes of PBS. The fusion protein was eluted with 50mM Tris-HCl, 5 mM reduced glutathione, pH 8.0. This extraction procedure was repeated till all the fusion protein was isolated from the cell lysate (determined by Coomassie plus-200 protein assay and SDS-PAGE analysis).

2.8.2.6 SDS-PAGE analysis. In order to determine the effectiveness of the fusion protein extraction, protein samples were analyzed by 20% SDS-PAGE. 20 μL aliquots were combined with 10 μL of sample buffer (12.5% Tris-HCl (pH 6.8), 10% glycerol, 10%(w/w) sodium dodecylsulfate (SDS), 5% β-mercaptoethanol, 15%(w/v) bromophenol blue and 47.5% deionized water). Molecular weight

marker (MW) was prepared by taking 20 μL of molecular weight standard (Sigma) and 5 μL of sample buffer. All samples were placed in a boiling water bath for 3 min (except for the MW; this period was 2 min). The samples were loaded on the gel, and run in the mini Protean III Electrophoresis Cell (BioRad) for approximately 50 min. at 200 V. The gel was then stained in Coomassie blue stain (50% methanol, 0.05% Coomassie brilliant blue, 10% acetic acid, 40% deionized water) for 15 min. The gel was then left overnight in destaining solution (5% acetic acid, 0.5% methanol, 85% deionized water). The gel was scanned using a CanoScan D1250 U2F scanner.

2.8.2.7 Determination of fusion protein concentration. The Coomassie plus-200 protein assay was used to determine GST fusion protein concentration. The spectrophotometer was auto-zeroed and measurements taken with deionized water as reference. According to the protocol provided (Pierce), a 50 μL aliquot of each fusion protein sample was added to 1.5 mL of Coomassie Plus-200 protein reagent in a 2.0 mL microtube. A blank was prepared by mixing 50 μL of elution buffer with reagent. Contents of the tubes were gently mixed by inverting, and the absorbance measured at 595 nm. The results obtained were compared against a standard curve prepared by performing the Coomassie plus 200 Protein assay on diluted samples of 200 50 μg/mg Bovine Serum Albumin (BSA) stock.

2.8.2.8 Trypsin cleavage of fusion protein. Different trypsin to protein ratios were used for different fusion protein samples. For GST-PDZ1 and GST-PDZ2, the ratio used was 1:1400(w/w), and for GST-PDZ1-2 it was 1:7000 (w/w) of

trypsin (TPKC-trypsin, Pierce) to fusion protein. Trypsin working buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, pH 7.4) was added, if required, to bring the final trypsin concentration to 1 μg/mL (from 1 mg/mL trypsin stock solution). The cleavage reactions were allowed to proceed for 2 h and 2.5 h for PDZ1 and PDZ1-2, respectively.

In order to stop the trypsin cleavage reaction, immobilized Soybean Trypsin Inhibitor on beaded agarose (Pierce) was added. The inhibitor was made into a 50:50 v/v slurry according to the Pierce protocol. One milligram of the trypsin inhibitor in a slurry can stop 4.5 mg of trypsin. Twice the theoretical amount was added. The contents of the tube were mixed by inverting for 2 min at room temperature, and then left for 5 min on ice. The procedure was repeated two more times. The cleaved product was then filtered through a 0.45 μm nylon Bio Express syringe filter to remove the immobilized inhibitor. The cleaved sample was placed in Snake Skin 3500 MWCO dialysis bag (Pierce) and dialyzed into 1 L of 20 mM MES, 10 mM NaCl, 1 mM DTT, pH 6.0. Thr dialysis buffer was changed twice.

2.8.2.9 GST removal. A GST affinity column was used to remove GST from dialyzed cleavage mixture. GST was bound to the glutathione agarose beads, while PDZ proteins were collected in the flow-through. The column was pre-washed with 10 column volumes of ice cold PBS, and trypsin cleavage samples were loaded at the rate of 5mL/min, collecting the flow in 50 mL Falcon tubes on ice. The column was washed and the bound GST eluted with 50 mM Tris-HCl, 5 mM reduced glutathione pH 8.0. The GST removal procedure was repeated

twice. The efficiency of GST removal after a second passage through the column was visualized by SDS-PAGE. The samples were further dialyzed in MES buffer (20 mM MES, 10 mM NaCl, 1 mM DTT, pH 6.0) for PDZ1 and MES buffer (20 mM MES, 10 mM NaCl, 1 mM DTT, pH 5.9) for PDZ1-2.

2.8.2.10 Purification of protein with ion exchange chromatography. Dialyzed protein samples were purified by ion exchange chromatography. The Hi-trap Q column (Amersham Pharmacia) was pre-washed with 5 column volumes of start buffer (20 mM MES, 10 mM NaCl, 1 mM DTT), followed by 5 volumes of regeneration buffer (20 mM MES, 1 M NaCl, 1 mM DTT), and then again 5 column volumes of start buffer. Previously dialyzed protein samples were then loaded into the Q column at a rate of 5 mL/min, using the Dynamax peristaltic pump. The column was washed with 5 column volumes of start buffer, and the bound protein was then eluted with 20 mM MES, 10 mM NaCl, 1 mM DTT, pH 6.0 or pH 5.9 (depending on the protein of interest) with varying ionic strength buffer. The elutions were conducted with an overall volume of 3 column volumes (15mL) of 25, 35, 45, 50, 55, 60, 65, 70, 75, and 80 mM NaCl for PDZ1, and 25, 40, 50, 70, 80, 90, 100, 110, 120, 140, 150, 200, and 250 mM NaCl for PDZ1-2. The column was then regenerated by washing with 5 column volumes of regeneration buffer, and 10 column volumes of start buffer. All the elutions were collected in 50 mL Falcon tubes on ice. 20 μL aliquots of all the elution fractions were analyzed by 20% SDS-PAGE. The elution fractions that contained protein of interest were dialyzed into final MES buffer (20 mM MES, 10 mM NaCl, pH 6.0) for PDZ1 and PDZ2 and (10 mM sodium phosphate, pH 6.0) for PDZ1-2.

2.8.3 Concentrating sample and determining protein concentration. Dialyzed protein samples (dialysis buffer changed three times) were concentrated using Millipore Ultrafree-15 Centrifugal Filter Device MWCO 5000. The filter was rinsed with dialysis buffer, used for the sample for 10 min at 2000x*g*, at 4 °C. The samples were added in 50 mL portions, and then concentrated by centrifuging at the same speed until a satisfactory concentration for ITC experiments was obtained.

A final protein concentration was defined by UV measurements at λ = 280 nm. Extinction coefficients for individual PDZ domains were obtained from calculations based on the protein sequence. Two quartz cuvettes (path length of 1 cm) were used for the measurements. Sample dialysis buffer was used as a reference in the measurements and to auto-zero the instrument.

2.8.3.1 Advanced Protein Assay. The absorbance was measured at 590 nm for this assay. Advanced Protein Assay Reagent (ADV01) was used as a reference and to auto-zero the instrument. The samples were gently mixed, transferred to disposable cuvettes, and the absorbance was measured with a UV-VIS spectrophotometer. Each 15 μL of protein sample was added to 1.5 mL ADV01, and the absorbance was read and concentration calculated based on correlation 1.0 OD_{590} = 30 μg/mL of reagent, per cm of path length.

2.8.3.2 Coomassie plus-200 protein assay. For the final protein sample, the Coomassie plus-200 Protein Assay was used. The same protocol described in 2.3.2.7 was employed, with the only difference being that the dialysis buffer was used instead of reduced glutathione solution.

2.8.4 Characterization of PDZ1, PDZ2 and PDZ1-2 proteins of PSD-95.

2.8.4.1 ESI mass spectroscopy. PDZ1, PDZ2 and PDZ1-2 tandem protein samples, whose purity were assessed by SDS-PAGE, were placed in 0.5 mL Slide-A-Lyzer dialysis cassettes (3500 MWCO) and dialyzed in 0.1 % trifluroacetic acid (TFA) in deionized water, volume of buffer 600 mL. The dialysis buffer was changed twice. The sample was taken out of the dialysis cassette and submitted for electrospray ionization mass spectroscopy (ESI-MS) analysis (Central Instrumentation Facility, WSU).

2.8.4.2 Plasmid DNA preparation for sequencing. Glycerol stocks of *E. coli* bearing the individual GST fusion were used to streak a single LB/amp50 plate, which was then incubated at 37 °C overnight. One colony was chosen to inoculate 10 mL of TB/amp50 media in a 50 mL falcon tube incubated with shaking for 10 h at 37 °C. Only 2 mL of the prepared culture was centrifuged at 20,800x*g* for 5 min to pellet the cells. The supernatant was removed, and a Qiagen Spin Miniprep Kit was used to extract the plasmid.

2.8.4.3. Quantitation of GST-PDZ plasmids and DNA sequencing. The DNA DipStick Kit was used to determine DNA concentration in the sample. Plasmid preparation sample dilutions (1:1; 1:10; and 1:100) were spotted on the DNA DipStick membrane and developed accordingly. Based on comparison of the color intensities to the standard chart on the DNA DipStick Quick Reference Card, the concentration was estimated to be >500 ng/μL, sufficient for analysis. The result showed the DNA concentration to be 100 ng/mL or more. 20 μL of the

plasmid DNA sample and 10 μL of PGEX5 primer (1 μM) were sent to the sequencing facility in the Biological Sciences Department of WSU.

2.8.5 Subcloning PDZ4 of MUPP-1.

2.8.5.1 Isolation of donor plasmid. The template DNA for MUPP-1 PDZ4 was isolated from cells containing plasmid pcDNA-MUPP1. The gene that encodes for the whole rat MUPP-1 protein was constructed into plasmid DNA of the host cells. Glycerol stock of *E. coli* was used to streak a single LB/amp50 plate, which was then incubated at 37 °C overnight. One colony was chosen to inoculate 10 mL of TB/amp50 media in a 50 mL falcon tube, incubated with shaking for 10 h at 37 °C. Only 2 mL of the prepared culture was centrifuged at 20,800x*g* for 5 min to pellet the cells. The supernatant was removed, and a Qiagen Spin Miniprep Kit was used to extract the plasmid**.**

2.8.5.2 Quantification of the MUPP-1 donor plasmid. The DNA DipStick Kit was used to determine the DNA concentration in the sample. Plasmid preparation sample dilutions (1:1; 1:10; and 1:100) were spotted on DNA DipStick membrane and developed accordingly. Based on comparison of the color intensities to the standard chart on the DNA DipStick Quick Reference Card, the concentration was estimated to be >500 ng/μL.

2.8.5.3 Polymerase chain reaction of the MUPP-1 plasmid.

Primers were designed to hybridize to sequences flanking the PDZ4 encoding region for amplification. The lyophilized forward (4.6 nmol) and reverse primer (4.8 nmol) (obtained from Invitrogen) were resuspended in sterile TE buffer (pH 8.0) to make a 100 μM stock. From these stocks, working solutions (20 μM) were

prepared by dilution. Two separate reaction mixtures were prepared and Hot Start PCR was used.

Solution 1	Solution 2
Sterile deionized water (63 μL)	Sterile deionized water (6 μL)
Forward primer (5 μL)	10 x Thermo buffer (1 μL)
Reverse primer (5 μL)	Vent Polymerase (1 μL)
10 x Thermo buffer (9 μL)	Plasmid DNA (2 μL)
nucleotide bases (8 μL)	
Total volume (90 μL)	Total volume (10 μL)

Table 2.2 PCR reaction conditions

Each component of the Solution1 was added to a PCR microtube (50 μL) and an Ampliwax PCR Gem 100 wax bead was placed on the top of the liquid. The wax was melted by placing the PCR microtube in a 80 °C heat block for 5 min, and then it was transferred on ice to allow the wax layer to solidify. On the top of the layer, Solution2 was carefully pipetted, and the tube was then placed in the Thermocycler. A second reaction was prepared exactly as the previous one, but with an increased concentration of $MgSO_4$ (from 2 μM to 4 μM), while keeping the overall volume constant.

The thermocycler was programmed for 3 cycles. Cycle 1-5 min at 94 °C for 1 repetition, cycle 2-10 seconds at 94 °C, 30 seconds at 45 °C, and 1 min at 72 °C for 25 repetitions, cycle 3 was set for 10 min at 72 °C for one repetition.

2.8.5.4 Agarose gel analysis of PCR product. A 1% agarose gel was prepared by mixing 0.5 g agarose powder with 50 mL 1xTBE buffer (prepared by dilution from 5 mL of 10xTBE buffer: 0.9M Tris base, 0.9M Boric acid, 0.025M EDTA (pH

8.0) and 45 mL of deionized water). The mixture was heated in a water bath for approximately 20 min, until the agarose was completely dissolved (solution became clear). After cooling for a few minutes, 2.5 μL of ethidium bromide (10 mg/mL) was added to yield a final concentration of 0.5 μg/mL, and then the mixture was poured into a gel casting frame, with a comb inserted in it. After the gel solidified (15-20 min), the comb was removed and the gel placed into the electrophoresis apparatus and covered in 300 mL of 1xTBE buffer with 15 μL same ethidium bromide stock solution, final concentration being 0.5 μg/mL. Samles were prepared as follows using 10 X Loading buffer-(20% w/v) Ficoll 400, 0.1M Na_2EDTA (pH 8), 1.0% (w/v) SDS, 0.25% (w/v) Bromophenol blue, 0.25% (w/v) Xylene cyanol and 1 X TBE buffer.

	sample	1X TBE	10X Loading buffer
LANE1	Promega 1kB DNA ladder	8 μL	1 μL
LANE2	PCR reaction (2 μM $MgSO_4$)	8 μL	1 μL
LANE3	PCR reaction (4 μM $MgSO_4$)	8 μL	1 μL

Table 2.3 Gel analysis of PCR products

The samples (10 μL) of each were loaded into the wells and the gel was run at 82 V for 1 h. Lane 2 showed a band at approximately 300 bp, as expected. The increased concentration of $MgSO_4$ didn't result in any band at lane 3. The calculated size of the PCR product was 303 bp, which corresponds to the band visualized under the UV light.

2.8.5.5 Purification and quantification of the PCR product sample. Qiagen PCR Purification Kit was used to purify the PCR product. The DNA DipStick™ Kit

was used to quantitate the purified sample. PCR product samples that were undiluted, (1:5), (1:10) diluted were spotted on a DipStick™ membrane and developed accordingly. Concentration estimated to be10 ng/µL.

2.8.5.6 Double restriction enzyme digest of PCR product sample. The PDZ4 PCR product was cleaved using *Bam* HI and *Sal* I restriction enzymes to form complementary ends for ligation into the GST plasmid. Double digest reaction set up was as follows: purified PCR product (20 µL), 10 X Buffer D (4 µL), acetylated BSA (0.4 µL), sterile deionized water (13.6 µL), *Bam* HI (1 µL) and *Sal* I. The reaction was mixed in a microtube and incubated for 2 hours at 37 °C.

2.8.5.7 Preparatory agarose gel of the restriction digest (PCR of PDZ4). A 1% agarose gel was prepared as described previously (2.1.3) except different comb with larger wells was used to accommodate the increased sample volume. After the gel was solidified, comb was removed and the gel placed into the electrophoresis apparatus and covered with 300 mL of 1 X TBE buffer, Ethidium bromide final conc. 0.5 µg/mL. To the overall volume of 40 µL of restriction digest 5 µL of 10 X Loading buffer and 5 µL of 1 X TBE were added. The sample was run on the gel along with molecular weight marker (2 µL 1 kb DNA ladder, 2 µL of 10 X Loading buffer, 16 µL 1 X TBE buffer) at 82 V for approximately 1 hour. Gel was then taken out from the chamber and placed on the UV illumination box. The band containing PDZ4 DNA was cut out and weighted.

2.8.5.8 Preparation of PDZ4 DNA for ligation. The Geneclean II kit Bio 101 was used with the TBE modifier to dissolve the gel slice and remove contaminants.

2.8.5.9 Quantification of the purified, digested PDZ4 DNA sample. Once again DNA DipStick™ Kit was used to determine DNA concentration in the sample. Purified, digested samples that were undiluted and (1:10) were spotted on DNA DipStick™ membrane and developed according to the protocol. When the membrane dried, color intensities were compared to the standard chart. Concentration was estimated to be 100 ng/µL.

2.8.5.10 Plasmid prep of GST/PDZ3 vector. A plate LB/amp50 was streaked with glycerol stock of *E. coli* cells which contained GST/PDZ3 vector and the plate was incubated overnight. Single, isolated colony was picked out and used to inoculate 10 mL of TB/amp50 media, followed by incubation at 37 °C with shaking for 8 h. Then 500 µL of the culture was used to inoculate 500mL of fresh TB/amp50, and the growth was continued overnight (37 °C, shaking).

The culture was spun down in 250 mL centrifuge tube at 3220 x g for 20 min, to pellet the cells. Supernatant was discarded and Wizard® Plus Maxiprep purification system from Promega was used to extract the plasmid from cells.

2.8.5.11 Restriction digest of the GST/PDZ3 plasmid. GST/PDZ3 plasmid has also unique restriction sites *Bam* HI and *Sal* I (as previously 1.5). Therefore, double restriction digest was performed on GST/PDZ3 to remove PDZ3 insert and create complementary ends for the ligation of the PDZ4 (MUPP-1) into GST plasmid. Double restriction digest reaction was set up in the following order: sterile deionized water (5 µL) 10 X Buffer D (10 µL), acetylated BSA (1 µL),), at this point everything was mixed thoroughly by pipetting, and then enzymes were added *Bam* HI (2 µL) and *Sal* I (2 µL). 1.5mL microcentrifuge tube containing

reaction mixture was incubated at 37 °C, for 2 hours, and then it was immediately analyzed by agarose gel.

2.8.5.12 Preparatory agarose gel of the restriction digest product-GST vector. Again 1% agarose gel was prepared with large wells as previously described (section 2.1.6) and to the restriction digest sample 10 x Loading buffer (5 µL) and 1 x TBE (5 µL) were added. The sample was run on the gel along with a molecular weight marker (2 µL of 1kb DNA ladder, 2 µL 10 x Loading buffer, 16µL 1 x TBE buffer) at 82 V for 55 min. The gel was placed on the UV illumination box, under minimum UV exposure and least possible time to minimize DNA damage. Using sharp razor agarose piece (digested plasmid) was cut out and placed into 2 x 1.5 mL microcentrifuge tubes.

2.8.5.13 Preparation of plasmid for ligation and quantification of the purified plasmid. For the extraction and purification of DNA sample, again QIAquick Gel Extraction Kit was used (according to the protocol). The DNA DipStick™ Kit was used to determine the concentration of the DNA in the sample. Purified, digested plasmid samples that were undiluted, 1:1, 1:5 diluted were spotted on the DipStick™ membrane and developed. After comparing the color intensities on the membrane to the standard chart, results showed DNA concentration of 100 ng/µL.

2.8.5.14 Ligation reaction of the GST-plasmid with the PDZ4 DNA using FAST-LINK DNA Ligation Kit. The formula shown below was used to determine the amounts of PDZ4 DNA and GST plasmid DNA for the ligation reaction.

$$\text{ng insert} = \frac{(\text{ng vector}) \text{ x } (\text{kb size of insert})}{\text{kb size of vector}} \left[\frac{\text{moles insert}}{\text{moles vector}} \right]$$

According to the Fast-Link protocol, insert to vector ratios (2:1), (5:1) and (10:1) are most likely to give positive results, that's why three different ligation reactions were prepared at the time.

Fast-Link DNA Ligation Kit from Epicentre® was used containing Fast-Link DNA Ligase, 10 X Fast-Link ligation buffer and ATP. Based on protocol, final ligation reactions were set up according to Table 2.4.

component	(2:1)	(5:1)	(10:1)
10 x Fast-Link ligation buffer	1.5 µL	1.5 µL	1.5 µL
10mM ATP	1.5 µL	1.5 µL	1.5 µL
Vector DNA	2.0 µL	2.0 µL	2.0 µL
Insert DNA	0.4 µL	1.0 µL	2.0 µL
Sterile deionized water	8.6 µL	8.0 µL	7.0 µL
Fast-Link DNA ligase	1.0 µL	1.0 µL	1.0 µL
Total volume	15.0 µL	15.0 µL	15.0 µL

Table 2.4 Fast-Link ligation conditions

Actual incubation time at room temperature was 20 min instead of the suggested 5 min, since length of incubation can be extended with no deleterious effects. All three different ligation reactions were then transferred into heat block at 70 °C for 15 min to inactivate the Fast-Link DNA ligase, a very important step, since failure to inactivate ligase may decrease transformation efficiencies.

2.8.5.15 Chemical transformation of* E. coli *cells. The Epicurian Coli BL-21 Gold competent cells were transformed with DNA plasmid from ligation reactions

previously described. The protocol followed was the one provided in the Instruction Manual for the transformation of BL-21 Gold competent cells, from Stratagene®. After thawing the competent cells on ice, three aliquots 3 x 100 µL were placed in 3x15 mL pre-chilled falcon tubes. The whole amount of each ligation reaction mixture (15 µL) was added to the falcon tubes accordingly. The tubes were then incubated on ice for 30 min. The transformation reactions were then heat-pulsed in a 42 ° C water bath for 20 seconds. Since duration of the heat pulse is critical for optimal transformation efficiencies, an optimum was observed in the 20-25 s range. Reactions were then incubated on ice again for 2 min. 0.9 mL of SOC media (stored at -80 ° C and allowed to thaw on ice prior to use) were added to all three samples, and then they were incubated at 37 °C, with shaking for 1 h 15 min. Transformed cells were concentrated by centrifuging at 10,000x*g* for 10 min. 750 µL of media was removed, and the cells were resuspended in 250 µL of the media that was left. Aliquots (3 x 15 µL) of each transformed cells were spread evenly onto LB/amp50 plates using sterile spreaders. The plates were placed upside-down in the incubator for 30-45 min, until the media was dried and then flipped and incubated for 37 °C overnight.

2.8.6 Isothermal titration calorimetry (ITC) experiments. Calorimetry experiments were performed with a VP-ITC titration calorimeter (Microcal, Inc., Northampton, MA). The protein samples were extensively dialyzed for at least 24 hours in the experimental buffer prior to experiments. The peptides used in the study were dissolved into the same dialysis buffer as that of the protein samples in each experiment. The pH of both solutions was measured at room temperature

using the pH-meter. The instrument was calibrated using three standard solutions of pH 4.0, 7.0 and 10.0 before measurements. The pH of peptide sample was corrected to perfectly match that of the protein sample. Differences of less than 0.02 pH units were considered a perfect match. The standard buffer for the calorimetry experiments was 10 mM sodium phosphate, pH 7.0 in case of PDZ1-2. Different buffer conditions were also tried for PDZ1-2 (20 mM MOPS, 10 mM NaCl, pH 7 or 20 mM HEPES, 10 mM NaCl, pH 7). For the different proteins tested in this study various buffer conditions were used 20 mM MES, 10 mM NaCl, pH 6.0 for PDZ1, PDZ2 and PDZ3 titrations. The standard experimental temperature was 25 °C. Solutions were degassed for 15-45 minutes under vacuum, without stirring, at a temperature a few degrees less than that of the experiment using the ThermoVac instrument.

For a typical titration, 65−150 μM samples of the PDZ protein were placed in the 1.4 mL reaction cell while 1.0−4.0 mM peptides were loaded into the 290 μL injection syringe. The c value, defined by Wiseman *et. al.* to be the product of the association constant for the reaction, K, and the macromolecule concentration, M, varied from 1 to 500. Each titration experiment consisted of a first (1 or 2 μL) injection (which was not used in data fitting) followed by 5 or 10 μL injections performed at 3 minutes intervals. The stirring speed used was 270 rpm and the reference power 10 μcal/sec. Heats of dilution were measured in blank titrations by injecting the peptides into the buffer used in the particular experiment. These results were used to correct the binding heats. This correction could sometimes be performed by taking the average of the last 5 microinjections

at the end of titration. Also, control experiments were performed by injecting the buffer into the buffer and the PDZ protein into the buffer.

2.8.6.1 Analysis of the ITC Data The data from each titration experiment were collected by the ORIGIN software (version 5.0, MicroCal Inc.). Thermodynamic parameters were determined by non-linear least-squares fitting for One Set of Sites, Two Set of Sites and Sequential Binding Sites model (all being standard model analysis in the Origin software used).

2.9 Results.

2.9.1 Subcloning of PDZ4 domain MUPP-1 into pGEX-2T vector. Primers for the PCR reaction were designed according to the known sequence of MUPP-1 in the pcDNA plasmid. This plasmid was used as a template for the polymerase chain reaction. Primers were designed to hybridize the sequences flanking the PDZ4 encoding region, in order for it to be isolated and amplified. The PDZ4 product from PCR was then ligated into the GST/PDZ4-containing plasmid in which the PDZ3 domain of PSD-95 was removed by restriction digest. *Bam* HI and *Sal* I restriction sites in the PDZ3 sequence were utilized for the removal of the PDZ3 domain from the plasmid. Recognition sites for these two enzymes therefore must be engineered into the PCR primers, so complementary ends could be created for ligation of the PDZ4 (MUPP-1) fragment into the donor plasmid.

When designing PCR primers, two other issues were taken into account. First, before the actual restriction, endonuclease recognition site series of short oligonucleotides were added in order to increase cleavage efficiency. Second,

primer lengths were increased by 9 extra oligonucleotides on both the forward and reverse PCR primers, outside the actual gene for PDZ4 (MUPP1), in order to ensure the accuracy of the obtained sequence. The DNA sequences for both forward and reverse primer are shown below. Primers were obtained from Invitrogen.

PDZ4FOR 5'- CGC GGA TCC ATG GGA ATT AAC TAT GAA ATA G -3'

PDZ4REV 5'- A CGC GTC GAC TCA GGT GGG TGG CAC AG – 3'

The PCR was run using the "hot start" protocol, in which the primers and the nucleotide bases were separated from the polymerase and the template by a thin wax layer. The actual composition of the reactions and the set up of the thermocycler are described in detail in the experimental section (section 2.2.5). Both the PCR product (calculated length 303 bp) and the GST vector (approximately 5 kb length), were digested using restriction enzymes. *Bam* HI and *Sal* I cut the sequence in such way that it leaves 'sticky ends' and facilitates insertion of the desired fragment.

The Fast-Link DNA Ligation Kit (Epicentre) was used as described in detail in the experimental section (2.2.5.14). After chemical transformation of competent cells, 15 µL aliquots of each of transformed cells were spread evenly onto LB/amp50 plates and left for overnight growth. Only a single colony was spotted on a plate that had (5:1) ratio of ligated transformants. The colony was used to inoculate 10 mL of TB/amp50 media. After incubation and subsequent extraction of the plasmid, it was sent for DNA sequencing. Results confirmed that it had the right insert (PDZ4), and that the subcloning was successful.

2.9.1.2 GST-PDZ4 expression and purification. This was a rather challenging task of expressing and characterizing PDZ domain of MUPP-1 (Multi PDZ-domain Protein), a structurally uncharacterized protein consisting of 13 PDZ domains, with no obvious catalytic activity, for which no binding studies had been performed. The protein domain chosen (PDZ4 of MUPP-1) needed to be produced in higher quantities, and to do so the well-established glutathione-*S*-transferase (GST) Gene Fusion System (Amersham Pharmacia) was selected. The GST tag from *Schistosoma japonicum* has been successfully used to express a significant number of proteins [150]. It has the advantage of relatively easy isolation from the other proteins expressed in the cells by using its glutathione substrate in the affinity column. The high affinity of GST for glutathione is important in the purification of this protein (and those fused to it) from the mixture of proteins expressed in the bacterial cells.

Expression of PDZ4 in the *E. coli* cells in the form of a GST fusion showed that it was mostly present in the insoluble inclusion bodies of the bacterial isolate. In an attempt to generate soluble protein, chemical lysis of the cells containing GST-PDZ4 fusion protein (using B-PER® bacterial protein extraction reagent from Pierce) was employed. The approach was intended as a milder way to break the bacterial cell membranes, in order to avoid any 'precipitating' processes and to access a higher concentration of the protein in the soluble form. Unfortunately, no soluble protein was harvested. However, the results are not surprising; in a comparative study done in the lab for a known protein expressed in high yield when sonication was used to rupture the cells, B-PER®

bacterial protein extraction gave poor results (Dorina Saro Ph.D. dissertation, WSU). A possible explanation lies in the composition of the reagent that contains a mild, nonionic detergent in 20 mM Tris-HCl (pH 7.5) buffer.

A refolding procedure was also applied to the GST-PDZ4 from the inclusion bodies, but was not successful in isolating the protein in a properly folded form. The reason for this is likely related to the fact that the mixture of reduced and oxidized glutathione (which is considered a redox shuffling agent for disulfide bond formation [151]) is also a substrate for the GST tag of the fusion protein, so it cannot produce the correctly folded protein.

2.9.2 Expression and purification of recombinant PDZ1-2 of PSD-95

Although considerable effort was spent on expressing and characterizing individual PDZ domains (constructs PDZ1 and PDZ2), the main focus of this research study is the tandem PDZ1-2 domain. To the best of our knowledge this dual domain was never expressed in this fusion vector. Previous attempts made in the group resulted in the fragmented polypeptide, not a fully functional protein.

The *E. coli* bacterial cells that were chemically transformed with pGEX-2T vector encoding the GST-PDZ (PDZ1, PDZ2 or PDZ1-2) fusion protein were used for overexpression of this protein. The vector is constructed in such a way that the gene that encodes for the GST enzyme (26 kDa protein) is fused to the N-terminal of PDZ protein. There is an engineered thrombin cleavage site in between the GST and the PDZ sequence that is used for proteolytic separation of these proteins. pGEX-2T vector contains an ampicillin resistance gene that allows the growth of the transformed bacterial cells in ampicillin containing

media. The protein expression from a pGEX plasmid is under the control of the *tac* promoter, which is induced using the lactose analog *isopropyl* β-*D-thiogalactoside* (IPTG).

The expression and purification of PSD-95 PDZ proteins (PDZ1, PDZ2 and PDZ1-2 tandem) were successful and provided high purity samples as determined by SDS-PAGE analysis. However, in terms of final protein amounts and stability, considerable differences were observed. In particular, in the case of PDZ2 by itself, the required concentration of the protein for binding studies by ITC couldn't be achieved.

The plasmid PGEX-2T, used for the expression of GST-fusions, was engineered with the thrombin cleavage site LVPRGS on the C-terminal of GST. This serine protease cleaves after Arg or Lys residues in the sequence P4-P3-Pro-Arg ↓ P1'-P2' where P3 and P4 are hydrophobic amino acids and P1' and P2' are nonacidic amino acids. In the experiment, however, trypsin (a less specific protease) was used for cleavage reaction. Assuming that trypsin would have cleaved the fusion after any arginine or lysine residue, different results for mass spectra analysis of individual PDZ domains were expected.

2.9.2.1. Expression and characterization of individual PDZ1 and PDZ2 and PDZ1-2 dual domain. For PDZ1 the yield of the fusion protein was approximately 120 mg/mL of bacterial culture, as determined by Coomassie® plus-200 Protein Assay (Bradford assay). After cleavage and further purification as described in detail in the experimental part, approximately 15 mg of protein

was isolated (value obtained based on three assays UV measurements (λ = 280nm), Advanced Protein Assay and Coomassie plus-200 Protein Assay).

The determined molecular weight for trypsin-cleaved PDZ1 by ESI-MS was 10,067±1.55 Da, which was 109 units less than calculated value of 10,176 Da based on thrombin cleavage (95 amino acid residues).

PDZ2 fusion protein yields for GST-PDZ2 were slightly lower than for GST-PDZ1, approximately 100 mg of protein per liter of bacterial culture (Coomassie® plus-200 Protein Assay). However, final purification provided a very low yield of pure PDZ2 (high purity sample as confirmed by SDS-PAGE) of only approximately 4.5 mg. Even an attempted concentrating of the sample still gave a very low concentration. When such samples were utilized for ITC binding studies against several of the high affinity peptides, the obtained data could not be processed. Saturation of the protein was reached within a few microinjections of a highly concentrated ligand. Therefore, the obtained results could not be interpreted. Further studies should focus on stability of the PDZ2 under the conditions tested.

The determined molecular weight for trypsin cleaved PDZ2 by ESI-MS was 10,111.98±2.71 Da, which was the same as the calculated value based on thrombin cleavage (97 aa).

	amino acid residues	MW (calculated)	MW (mass. spec)	Yield fusion (mg/L culture)	Final protein (mg/L culture)
PDZ1	95	10,176	10,067	120	15
PDZ2	97	10,111.96	10,111.98	100	4.5

Table 2.5 Protein yield and mass for PDZ1 and PDZ2.

Expression and purification of the PDZ1-2 dual domain actually provided the highest quantities when compared to results from single PDZ domain expression attempts. As determined by the Bradford assay, approximate yields were 200 mg from 1 L of cell culture.

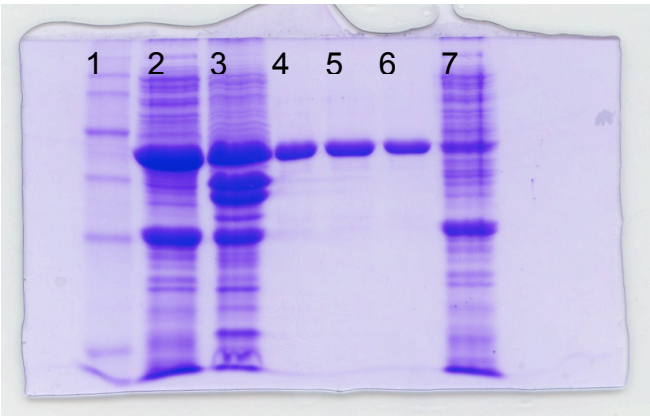

Figure 2.5 Separation steps for isolation of fusion protein, approximately 46 kDa from soluble cell lysate. Lanes: **1**, molecular weight marker; **2,** soluble cell lysate; **3,** cell pellet after lysis; **4-6,** fusion protein extractions I, II, III, respectively; **7,** last load.

A major concern of the PDZ1-2 dual domain, with regard to the five amino acid linker region between PDZ1 and PDZ2, was that the attempted cleavage with trypsin (protease of choice) would result in 'over-cleaving' and separate the two domains. This was based on the previous results obtained in the laboratory. However, optimization of the cleavage reaction finally resulted in a protocol

utilizing 1:7000 trypsin to protein ratio, and a reaction time of 2.5 h at room temperature.

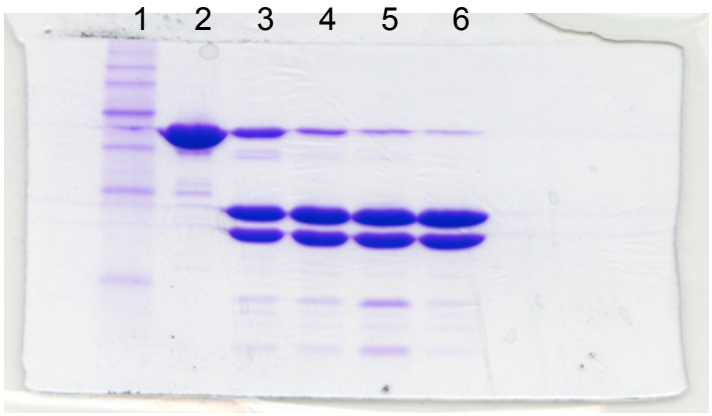

Figure 2.6 Showing trypsin cleavage, reaction for 2,5h, (1:7000) trypsin to protein ratio used. Lanes: **1**, molecular weight marker; **2,** fusion protein; **3-6,** reaction aliquots taken out after 1 h,1.5 h, 2 h, 2.5 h, respectively.

In the Figure 2.6 shown above are the results from trypsin cleavage reaction (1:7000), determined after series of analytical scale cleavages with varying enzyme to protein ratios and incubation times.

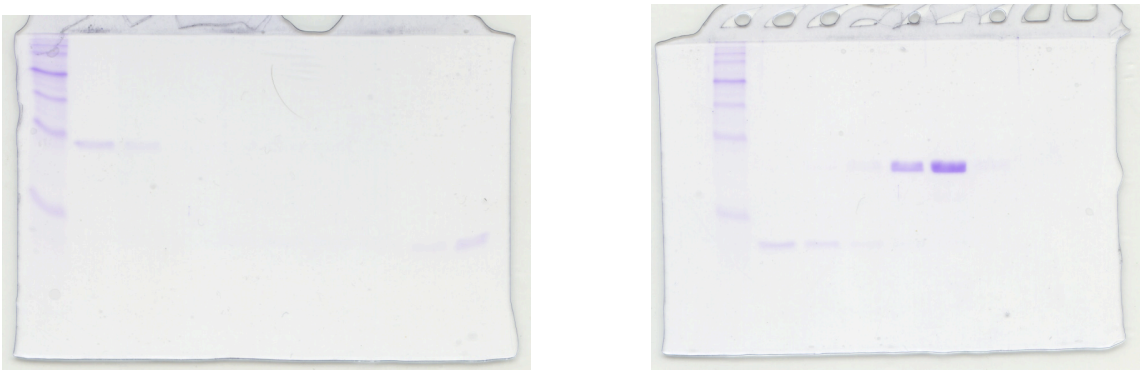

Figure 2.7 Final purification of PDZ1-2. MW- Molecular weight markers. Final PDZ1-2 purified samples in lanes marked f1 and f2, which corresponds to ion exchange fractions of 100mM and 150 mM of NaCl.

Final separation is shown in Figure 2.7. Gels in this figure contain the elutions from the ion exchange column. The GST portion, approximately 27 kDa, began eluting with the start buffer, at an ionic strength of 10 mM. The PDZ1-2 dual domain eluted in later fractions at higher ionic strengths of 100 and 150 mM NaCl. The isolated PDZ1-2 domain is shown in lane labeled f1 and f2.

2.9.3. Energetics of binding of separate PDZ1 and PDZ2 with peptides derived from their natural ligands. The initial design of peptides mimicking natural binding partners for the first two PDZ domains, was mainly focused on already reported results for protein-protein interactions of PSD-95.

The first in hand was PDZ1 and peptides based on binding sequences of best known binders were tested by ITC in 20 mM MES, 10 mM NaCl, pH 6 buffer. Figure 2.8 shows a typical isotherm (top panel) of titration of a peptide into a solution of PSD-95 PDZ1. The binding reactions are exothermic, and integrated heats of reaction for the data in Figure 2.8 (top panel) are shown as a binding curve in Figure 2.8 (bottom panel).

A non-linear least-square fit of the binding curve gives the stoichiometry (n), equilibrium dissociation constant (K_d), and the change in enthalpy (ΔH) for binding interaction. These parameters (ΔH and K_d) were used to calculate the free energy of binding, ΔG, and change in entropy, ΔS, using equations (1) and (2). The solid line shown in the Figure 2.8 (bottom panel) is the best fit for one of the reaction performed at 25 °C. Three different peptide sequences corresponding to C-terminus of natural binding partners were tested against the PDZ1.The ITC results are summarized in the Table 2.6 and Table 2.7.

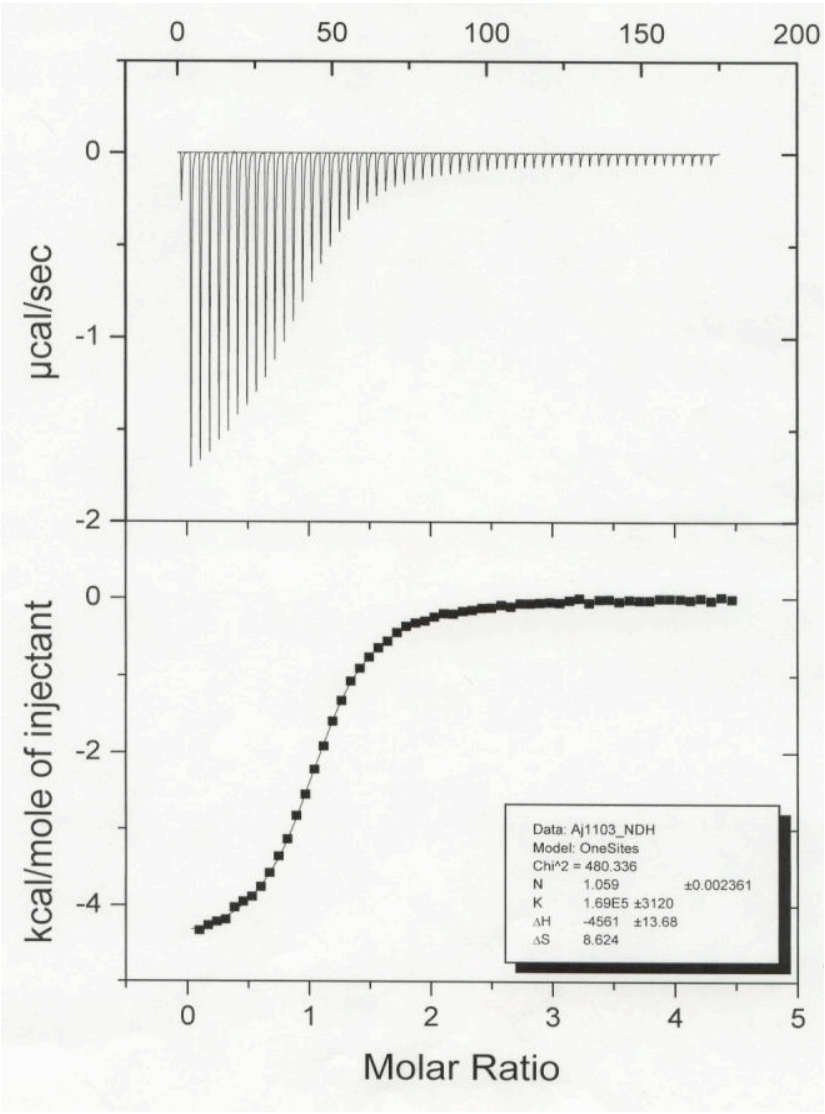

Figure 2.8 A representative thermogram of binding of PDZ1 with the AKAVETDV peptide in MES buffer pH 6 at 25 °C.

Among the natural binders of PDZ1 of PSD- 95, the first series of peptides studied by ITC were synthesized based on C-terminal sequence of K^+ channels of the Shaker type Kv1 family. The first two peptides, comprising 8 amino acid residues, were designed to probe the molecular recognition of binding. The C-terminal ends with Thr-Asp-Val; having Thr at -2 and Val at position 0 is consistent with the Class I PDZ domain binding preferences. The last two

ligands, an octapeptide and a hexapeptide, are derived from the NMDA receptor NR2b subunit. This is one of the most well studied interactions of PSD-95 and is critical for a number of neuronal functions, such as synaptic transmission and synaptic plasticity. The best binder was proven to be that possessing the Kv1.4 sequence, with a K_d of 5.93 μM, and which was slightly enthropically driven with ΔH of -4.42 kcal/mol. The potassium ion channel Kv1.3-derived sequence exhibited almost two times less binding affinity for PDZ1, with a K_d of 13.9 μM, with favorable enthalpy and similar contributing entropy.

Association with the octameric C-terminal NR2b subunit peptide was weaker than expected, with K_d = 11.2 μM, and the interaction is primarily driven by a favorable change in entropy, $T\Delta S$ = 3.88 kca/mol. Surprisingly, the hexapeptide had diminished signal, and no thermodynamic parameters could be calculated.

Peptide	K_d (μM)	ΔH (kcal/mol)	-TΔS (kcal/mol)	ΔG (kcal/mol)	n	c*
AKAVETDV	5.93 ± 0.02	-4.42 ± 0.14	-2.71 ± 0.14	-7.13 ± 0.01	1.09 ± 0.03	17
IKKIFTDV	13.9 ± 0.5	-3.51 ± 0.12	-3.11 ± 0.14	-6.62 ± 0.02	1.25 ± 0.01	7
LSSIESDV	11.2 ± 0.7	-2.87 ± 0.04	-3.88 ± 0.01	-6.75 ± 0.04	0.84 ± 0.13	9
SIESDV	no detectable binding					

Table 2.6 Thermodynamic parameters for peptides based on potassium channel proteins (Kv1.x) Kv1.4 and Kv1.3 and on NMDA receptor (NR2a subunit).

Peptide	K_d (μM)	ΔH (kcal/mol)	-TΔS (kcal/mol)	ΔG (kcal/mol)	n	c*
KKKETEV	10.9	-3.02	-3.75	-6.77	1.63	9

Table 2.7 Thermodynamic parameters for peptides based on consensus sequence for PDZ3.

Of all the three protein domains, PDZ2 was proven to be the least stable. The quantities of isolated pure protein were too low, and insufficient for ITC studies. Protein was dialyzed into 20 mM MES, 10 mM NaCl, pH 6, conditions that were observed to stabilize the PDZ1. In numerous attempts, during ITC titrations protein sample would precipitate in the cell. A huge shift in baseline could be observed, and it was impossible to integrate the data. Randomly, it was possible to collect a whole data set and thermogram of a single experiment was shown below (Figure 2.9).

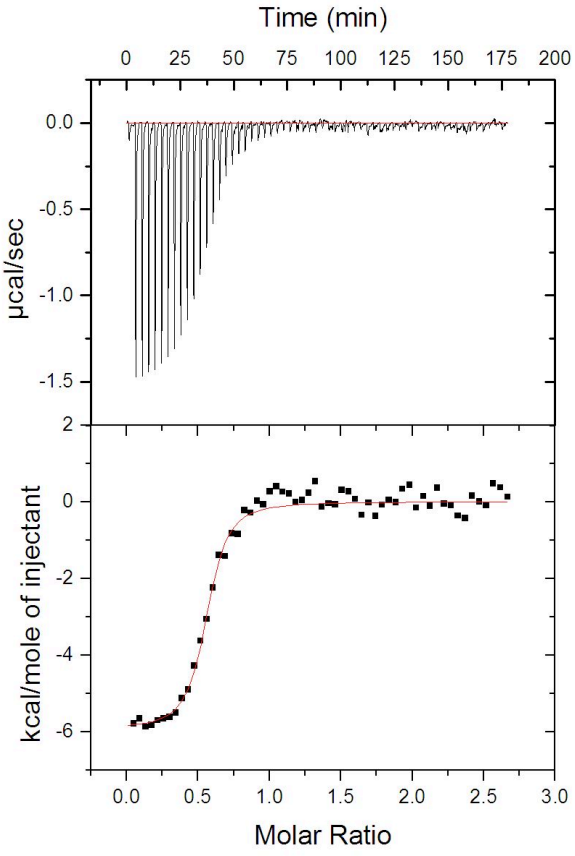

Figure 2.9 Thermogram of PDZ2 titration with AKAVETDV peptide.

When treated with a Single Binding Site model from Origin software this data set gave an n value of 0.5. Normally this would point to a binding stoichiometry wherein two peptides associate to each equivalent of protein. This in itself would be unlikely based on what is presumed about the binding behavior of individual PDDZ domains. In this case, however, there are experimental concerns such that this value is more likely attributable to an inadequate protein concentration estimate, or is indicative of the protein sample not being entirely in its native, folded state.

PDZ2 was not pursued in further experiments; this was due to such poor yields at the end of the long purification, which could not be improved through optimization efforts, and because of its poor stability (protein precipitates after 2 days at 4 °C).

2.9.4. Energetics of binding of the PDZ1-2 dual domain.

PDZ1-2 data was collected in 10 mM sodium-phosphate buffer, pH 7 at 25 °C. For data processing all three models provided by Origin software were used (One Set of Sites, Two Set of Sites and Sequential Binding Sites). The best model fitting, as judged by the $\chi 2$ value, was achieved using the Single Site binding model. This provided a stoichiometry of n = 1 (Table 2.8). This was not expected, based on the presence of two binding sites in the PDZ1-2 dual domain. Consequently, the thermodynamic data was hard to interpret from a biochemical perspective, and further studies were focus on obtaining the structure of tandem PDZ1-2 of PSD-95.

Peptide	**Kd** (μM)	**ΔH** (kcal/mol)	**-TΔS** (kcal/mol)	**ΔG** (kcal/mol)	**n**	**c***
AKAVETDV	3.86 ± 0.08	-4.25 ± 0.33	- 3.13 ± 0.31	-7.38 ± 0.01	1.12 ± 0.01	14
KAVETDV	4.81 ± 1.93	-4.41 ± 0.09	-2.88 ± 0.18	-7.29 ± 0.27	1.03 ± 0.05	21
AVETDV	2.81 ± 0.21	-4.60 ± 1.08	-2.57 ± 1.12	-7.56± 0.04	1.11 ± 0.14	41

Table 2.8 Tables thermodynamic parameters for peptides based on natural binding sequences of Kv1.4.

2.9.5 Binding of PDZ1-2 to modified cyclic peptide ligands.

CN2180 is one of several modified peptides designed to bind to PDZ3 of PSD-95, part of a larger program to develop novel cellular probes to investigate the *in vivo* roles of various PDZ domain-mediated interactions. CN2180 is a nontraditional form of cyclic peptide in which the ring bridges two side chains residues via a β-alanine linkage (Figure 2.10), and the sequence based on that of the CRIPT protein (KNYKQTSV), a known binder of PDZ3. Based on a simpler cyclic peptide structure that itself had been shown to exhibit good binding to PDZ3 [128], CN2180 contains two additional N-terminal residues and a myristic acid moiety to promote membrane permeability (Figure 2.10).

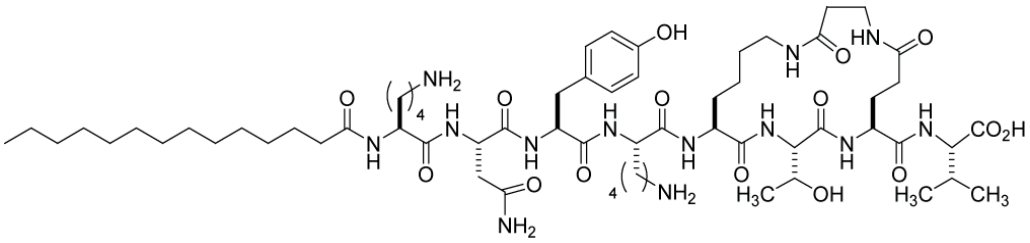

Figure 2.10 Chemical structure of the *N*-myristylated cyclic peptide CN2180.

Although CN2180 was originally designed to bind PDZ3, screening (with a N-biotinylated version of CN2180) against a membrane-bound array of 96 different PDZ domain constructs demonstrated that significant binding also occurred with PDZ1-2 (Dr. C. Rupasinghe, Ph.D. dissertation). This screen was based on the binding of ligand to solid-phase immobilized protein, and thus might

not have represented a true solution interaction. With the prospects of identifying a new binding ligand for PDZ1-2, as well as the newly discovered neuroprotective activities that CN2180 has exhibited (unpublished results, DJ Goebel, CN Rupasinghe and MR Spaller), there were compelling reasons to conduct a calorimeteric binding study of the system.

Experimentally, these studies proved more onerous than the earlier ones utilizing linear peptide. The previous ITC data for the PDZ1-2 dual domain was collected using final dialysis buffer conditions of sodium phosphate (10 mM, pH 7), which was proven to stabilize the protein. Numerous attempts were made to dissolve CN2180 under these conditions. ITC experiments were carried out varying the concentration of peptide, even at very low values (0.5 mM), only to observe diminished signal and precipitate at the end of the titration. Different buffer compositions were also tried (10 mM MOPS; 10 mM HEPES) at pH 7, and for all of them CN2180 was insoluble. In an attempt to match the conditions of previously obtained data for PDZ3 of PSD-95 binding to the same ligand (where K_d = 8μM), following purification PDZ1-2 sample was dialyzed against the same final buffer (20 mM, 10 mM NaCl, pH 6). These final conditions yielded titration curves suitable for subsequent analysis (Figure 2.11).

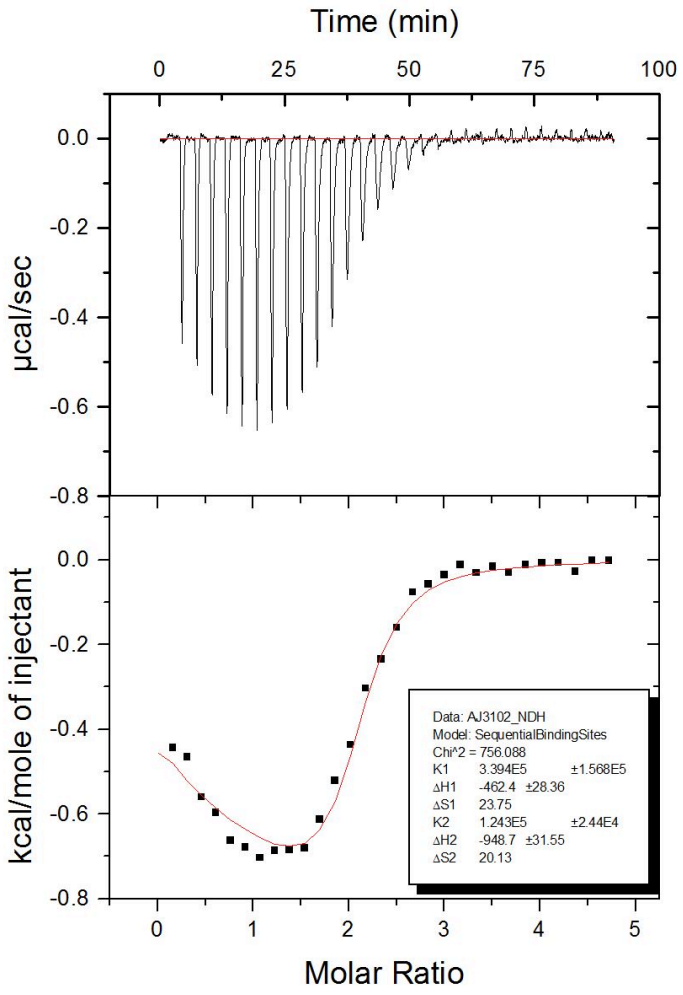

Figure 2.11 ITC thermogram and integrated data with fitted curve for PDZ1-2 titration with CN2180.

Calorimetric binding analysis was performed in the same manner with the previously described work. Data was treated with all three models: One Set of Sites, Two Set of Sites and Sequential Binding Sites. After applying the Single Binding Model, it became clear that the stoichiometric *n* value is 2, meaning that both binding sites on the tandem PDZ1-2 are fully occupied. However, fitting of the data improved significantly if Sequential Binding Sites model was implemented, as judged by χ2 values. Therefore, it is clear that the Sequential

binding Sites Model gives the best description of the experimental data (Table 2.8).

Thermodynamics of the first binding site				Thermodynamics of the second binding site			
Kd (μM)	ΔH (kcal/mol)	-TΔS (kcal/mol)	ΔG (kcal/mol)	Kd (μM)	ΔH (kcal/mol)	-TΔS (kcal/mol)	ΔG (kcal/mol)
6.09 ±3.13	-0.37 ±0.09	-6.83 ±0.25	-7.20 ±0.34	9.00 ±0.95	-1.16 ±0.21	-5.73 ±0.27	-6.89 ±0.48

Table 2.8 Thermodynamics for CN2180 binding to PDZ1-2.

The data reflect the outcomes of three independent measurements. It is apparent that CN2180 populates both binding sites in PDZ1-2 tandem with the fairly good binding affinity for each domain (K_{d1} = 6 μM and K_{d2} = 9 μM). Although two different values are obtained, the model fitting does not allow a specific assignment to be made as to which is that of PDZ1 or PDZ2. These values are comparable to the previously obtained result for PDZ3 (K_{d1} = 8 μM), and thus make this compound a reasonably effective inhibitor of the PSD-95 interaction with the NMDA receptor (NMDAR).

2.9.5.1 CN2180 and PDZ1-2. A key organizing protein in the postsynaptic density of the excitatory synapses is PSD-95. Especially important is the role it plays in coupling NMDAR to intracellular proteins and signaling enzymes [152, 153]. It has been known that through its second PDZ domain (PDZ2), PSD-95 binds the COOH-terminus, tSXV motif of NMDAR NR2 subunits as well as neuronal nitric oxide synthase (nNOS) [91, 153] This binding couples NMDAR activity to the production of nitric oxide (NO), a signaling molecule that mediates

NMDAR-dependent excitotoxicity [154]. Previous studies have shown that NMDAR activity is not suppressed by genetically disrupting PSD-95 *in vivo* [155] or by suppressing its expression *in vitro* [156]. Complete PSD-95 deletion dissociates NMDAR activity from production of NO and suppresses excitotoxicity [156].

NMDARs mediate ischemic brain injury [157], but targeting them has been proven deleterious in humans and animals [158-160] Targeting PSD-95, which is downstream of the NMDAR and would not be expected to impact other NMDA-induced pathways, therefore represents an alternative therapeutic approach for diseases that involve excitotoxicity. Genetic mutation or reducing expression levels of PSD-95 is therapeutically impractical; therefore an attractive alternative approach would be perturbing its interaction with NMDARs in order to suppress excitotoxicity and ischemic brain damage. This might be achieved by intracellular introduction of ligands that bind to PDZ2 of PSD-95. In this case, CN2180 has proven to be a strong lead compound, based on its potential inhibiting of all three PDZ domains of PSD-95. Because CN2180 is myristylated, it is more bioavailable than a non-modified peptide, and its diffusion and uptake capacity by neurons in intact tissues should be enhanced.

In a follow up study, CN2180 was tagged with fluorescein (FITC) to yield CN3205, and administered via intravitreal injection at the maximal effective dose to monitor its rate of diffusion of the retinal ganglion cells (unpublished results, CN Rupasinghe, DJ Goebel, MR Spaller). This is a common approach since NMDA-induced neuronal cell death has been extensively studied in the retina.

Results showed that as early as 1 hr following injection, uniform fluorescent staining was noted in the inner half of the retina, and that by 2 hrs, intensely stained vacuoles could be seen within select cells in the ganglion cell layer (GCL). Although further testing needs to be done in order to determine to what extent these cells express PSD-95, the fact that diffusion and localization of the modified peptide into putative NMDA-receptive cells occurs within a 2 hour time frame indicates that this preincubation period is sufficient for testing the neuroprotective properties of these compounds against NMDA-receptor mediated cell death.

With the impressive biological properties of CN2180 now documented, the calorimetric results serve as an absolutely essential step towards fully characterizing this compound. The solution data make it clear that CN2180 is capable of binding both PDZ1 and PDZ2, and in so doing provide evidence for a mechanism by which the observed activity takes place. Although other possible cellular binding sites cannot be absolutely ruled yet, the combination of the ITC data, in conjunction with the absence of binding to other non-PSD-95 PDZ domains on the 96 domain array, strongly suggest that PSD-95 is indeed the *in vivo* target.

2.9.5.2 In vivo retinal cell death model. NMDA-induced neuronal cell death has been extensively studied in the retina [161, 162]. Our laboratory conducted studies in collaboration with Dr. Goebel, WSU. Relying on previously published data [161] it was proven that following a moderate insult (20 nmol NMDA injected intravitreally) only 70-85% of the retinal ganglion cells and a significant number of

amacrine cells are lost. Extensive research in the past few years in the Goebel laboratory improved this *in vivo* NMDA-toxicity model and also led to development of several quantitative assays to monitor biochemical, physical and energy dependent/independent pathways that are believed to be linked to the cell death pathway [163]. This includes measuring PARP-1 activation (believed to be linked to the process of necrosis), loss of plasma membrane selectivity (a process directly linked to necrosis), DNA cleavage using TUNEL labeling, endogenous ATP pools as well as ganglion cell survival. This system offers several advantages to *in vitro* systems, with the most significant of them being that we can monitor the cell death process in an intact, living animal.

In their previous work, Goebel *et al* have established a baseline for all of the above assays using NMDA-treated and control eyes; they have been able to assess putative neuroprotective compounds using this model and have recently reported that inhibition of PARP-1 activity, using the potent inhibitor PJ-34, abolished NMDA-induced PARP-1 activation in the retina, blocked NMDA-induced loss of endogenous ATP and provided some neuroprotection to retinal ganglion cells. However, this blockade did not attenuate NMDA-induced loss of membrane selectivity or prevent TUNEL labeling [163].

A different set of experiments has shown that blockade of CaMKII-α_B completely abolishes NMDA-induced TUNEL labeling and did provide significant neuroprotection of retinal ganglion cells; however, this treatment had no effect in blocking PARP-1 activation nor loss of membrane selectivity of neurons in the ganglion cell and Inner nuclear layers of the retina (Goebel *et al*, manuscript in

preparation). Collectively, these studies clearly indicate that the pathway leading to cell death is complex and that it can no longer be confined to the classification of being necrotic or apoptotic.

All the results presented in the following (chapter 2.3.6.2) represent the *in vivo* work done in Dr. Goebel's laboratory. Only finalized data representations were outlined as a strong indication of CN2180 biological potential.

2.9.5.2.1 ***CN2180 and CN2097 reduce NMDA-induced poly(ADP-ribosyl)ation (PAR) in a dose-dependent manner***. Two compounds were tested in this study; myristylated peptide CN2180 (detailed description given in paragraph 2.3.6. of this dissertation) and its poly-Arg derivative (CN2097), designed to enhance its diffusion and uptake capacity by neurons in intact tissues. These two compounds were individually administered into the vitreal chamber of the left eye of male Spague-Dawley rats at designated doses 2 hours prior to the retina being exposed to a second injection containing an excitotoxic dose (20 nmol) of the glutamate agonist N-methyl-D-aspartate (NMDA) in the presence of the corresponding dose of the peptide (CN2180 or CN2097). The companion eye (right eye of each animal) served as either the positive or negative control.

Four hours following NMDA insult, the animals were decapitated (to eliminate any potential anesthetic influence on the results of the assay) and retinas were individually homogenized and processed for the detection of poly(ADP-ribosy)lation (an indicator of enhanced PARP-1 activity) as previously described [163].

Results show that both CN2080 and CN2097 reduce NMDA-induced poly(ADP-ribosy)lation (PAR) of large molecular weight proteins. Dose-response data shows that although both compounds did lower PAR-formation, the poly-Arg derived peptide (CN2097) was roughly 10 times more effective than the myristylated compound (CN2180). Interestingly, a myristylated randomized-cyclic peptide (of the same size; CN3500) administered at the optimal effective dose determined for CN2180 and CN2097, showed no significant decline in NMDA-induced PAR-formation.

The obtained data is a clear indication that the decline in NMDA-induced PAR-formation is KNYKKTEV peptide-dependent, and that competitive binding to the PDZ domain of PSD-95 (or related proteins) containing this domain, aids in diminishing NMDA-induced PARP-1 activation. This is consistent with the understanding that PSD-95 is linked to nNOS (neuronal nitric oxide synthetase), which generates nitrous oxide (NO) within the neuron. It has been hypothesized that NO hyperstimulates PARP-1 activation in response to NO-free radical induced DNA-strand breaks [164] and that this response leads to the loss of plasma membrane permeability properties (i.e. necrosis) [163, 165].

2.9.5.2.2 ***CN2180 and CN2097 attenuate NMDA-induced loss of plasma membrane selectivity.*** Same method as described in the previous chapter was employed to assess NMDA-induced loss of plasma membrane selectivity (LPMS) following NMDA insult in the presence of CN2180 or CN2097 and compared with NMDA-treatment alone or with retinas treated with the non-specific peptide CN3500 followed by NMDA insult as a control. LPMS is visualized by the ability

of ethidium bromide to pass through the affected neuron's plasma membrane and stain the cell. This process has been shown to be an NMDA-dependent process, as retinas treated with buffer alone, or those that were untreated are not responsive to neuronal ethidium bromide staining [163].

Retinas pre-exposed to CN2180 or CN2097 subsequently were subjected to NMDA insult. These samples showed a significant reduction in the density of ethidium bromide-stained cells in the ganglion cell layer (GCL), as compared with retinas treated with NMDA alone. This decrease in ethidium bromide staining density was shown to be dose dependent, and peptide-specific as the myristylated randomized cyclic peptide (CN3500), given at the same dose, had no effect in blocking NMDA-induced LPMS. Consistent with the PAR data above, the poly-Arg peptide (CN2097) provided almost complete blockade of ethidium bromide staining in the GCL and inner nuclear layers, whereas the myristylated peptide given at its maximum dose only partially attenuated LPMS.

Presented data support the hypothesis that NMDA-induced cell death is mediated by the NMDA-receptor's linkage to PSD-95 and shows, for the first time, that disconnection of the NMDA-receptor from PSD-95, through competitive binding of the peptide-ligand(s) with the PDZ-domain of PSD-95 (or similar protein), renders the NMDA-receptive neurons insensitive to NMDA-induced LPMS (necrosis).

2.9.5.2.3 ***CN3205-FITC localizes to select subpopulations of retinal neurons that reside in the ganglion cell layer (GCL) and inner nuclear layer of the retina.*** Another cellular probe was developed by tagging the myristylated peptide

CN2180 with fluorescein (FITC: CN3205) and administered via intravitreal injection at the maximal effective dose (6 nmol/eye) to monitor its rate of diffusion into the retina. Results showed 1 hr following injection and after 4 hrs, numerous cells showed preferential accumulation of the tagged peptide and were intensely stained in both the GCL and inner nuclear layer (INL) of the retina (Fig 2.12). Interestingly, this intense staining of the cells (in both layers) appears to be selective, as there were neighboring cells in both layers that did not fluoresce above tissue background levels (Fig 2.12). Furthermore, the staining and density patterns of the cells that accumulated CN3205 showed strong similarities to those neurons that are susceptible to NMDA-induced LPMS.

In the Figure 2.12 panel (a) 4 hrs following intravitreal injection of 12 nmol of CN3205 shows specific uptake into cells located in the ganglion cell layer and by 6 hrs (b) the staining density of labeled cells appears to parallel the staining of cells that are susceptible to NMDA-induced excitotoxcity as detected by ethidium bromide. (c) Detection of CN3205 in ganglion cells was noted as early as 2 hrs following treatment and shows that the compound is localized to the cytoplasm. Panel (d) depicts the ability of the peptide to penetrate the retina radial sectioning of the retina shows that there is extensive staining in cells located in the ganglion cell layer (GCL) as well as in cells located in the inner portion of the inner nuclear layer (INL) (arrows).

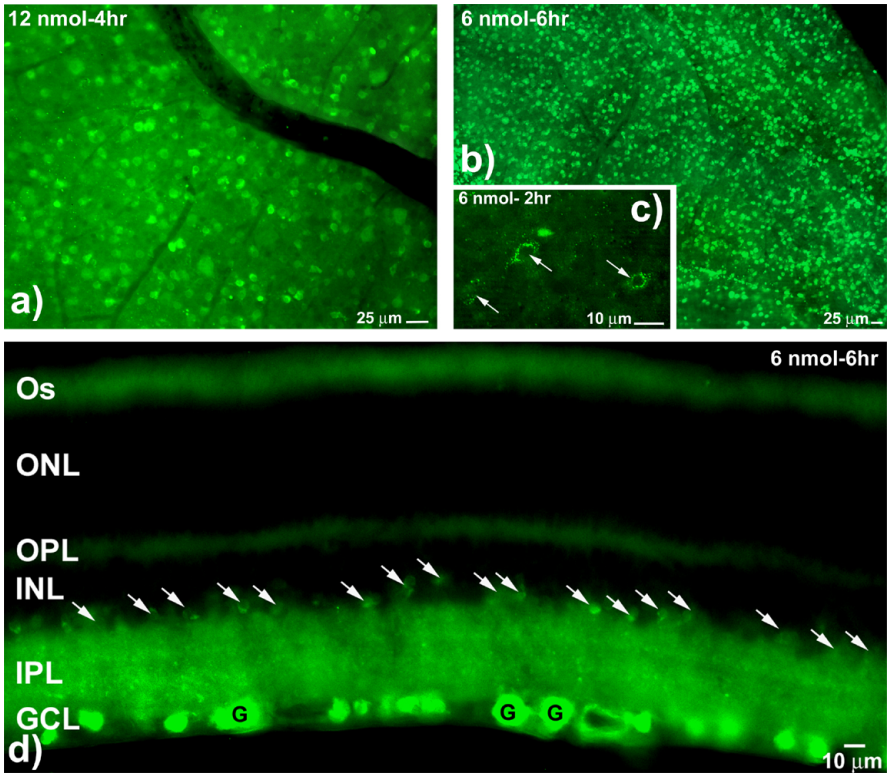

Figure 2.12. Uptake of FITC-labeled KNYKKTEV peptide (CN3205) in the retina

Although these findings are preliminary and will require further investigation to determine whether the stained cells are NMDA-receptive, and more importantly express PSD-95, the fact that diffusion and localization of the modified peptide into putative NMDA-receptive cells occurs within a 2 hour time frame indicates that this preincubation period is sufficient for testing the neuroprotective properties of these compounds against NMDA-receptor mediated cell death.

2.9.5.2.4 ***CN2097 attenuates NMDA-induced loss of retinal ganglion cells***. Based on the data presented, there is a strong indication that disruption of the PDZ binding disconnects NMDA-receptor activation from initiating the cell death signal for the short term (up to 4 hours following insult). However, it was still not

clear whether this short-term protection was transient or if it does provide long-term neuroprotection. To assess this, a series of animals were treated in such a way that the left eye received the optimal dose (3 nmol) of CN2097 followed by NMDA treatment 2 hours later and the right eye received NMDA alone or left untreated. Two weeks following insult each animal was surgically administered the retrograde tracer fluorogold. Retinas were harvested and photographed. Fluorogold (FG)-labeled cells densities were then determined at each eccentricity and then compared with the densities obtained from the companion eye (for full methodology [163]).

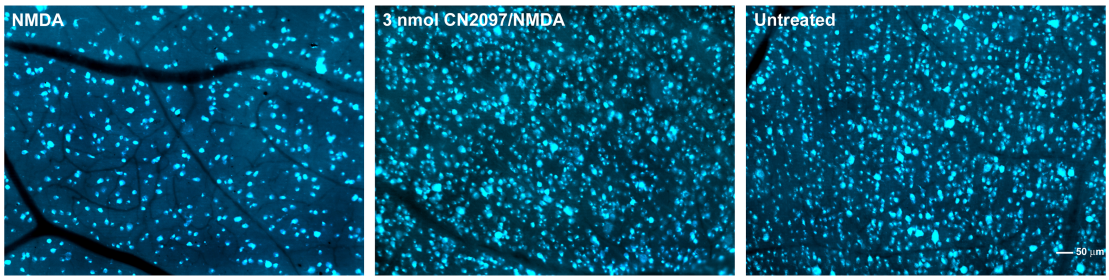

Figure 2.13. Representative fluorogold labeling of retinal ganglion cells sampled from the same region of the retina at 1-2 mm from the optic nerve head are shown.

First panel (Fig. 2.13) shows the NMDA treated retina (20 nmol) and a dramatic reduction in ganglion cell density of surviving ganglion cells is visible. In the second panel retinas were pretreated with 3 nmol CN2097 for 2 hrs followed by 20 nmol NMDA w/ 3 nmol CN2097 (3nmol CN2097/NMDA) and show a dramatic increase in ganglion cell density which was essentially indistinguishable from the regional density of the untreated retina (Untreated labeled) third panel on Fig 2.13.

Preliminary results from 2 retinal pairings show that CN2097 provides significant neuro-protection against NMDA-induced ganglion cell loss (Fig. 2.13 and 2.14). In comparing cell densities from the CN2097/NMDA-treated retinas with those obtained from retinas that were untreated little differences at each eccentricity were observed between the two treatment conditions (Fig. 2.13 and 2.14).

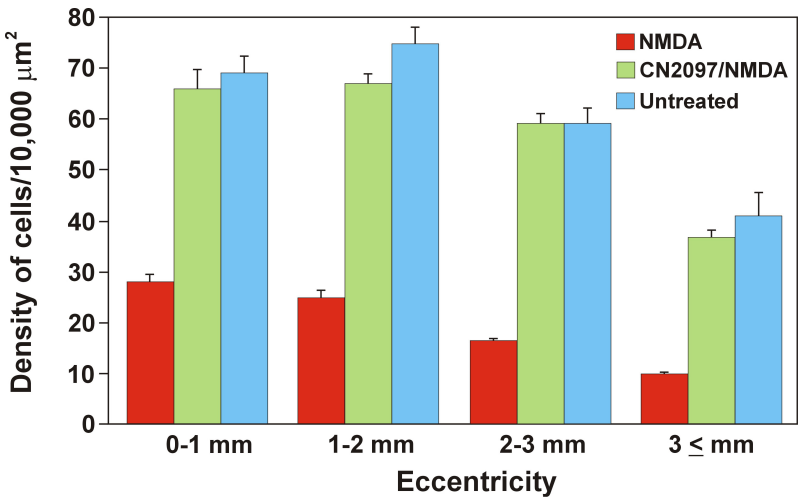

Figure 2.14 Effect on CN2097 pre and post insult was induced (NMDA)

Figure 2.14 represents retinal pairings (n=2), where one eye was pretreated with 3 nmol CN2097 and 2 hrs later treated with 20 nmol /NMDA were compared with companion retinas treated with NMDA-alone show that CN2097 offers almost full protection against NMDA-induced ganglion cell loss when compared with the densities found in untreated retinas (n=4).

Although the full analysis of these experiments are still forthcoming, indications from the two retina pairs analyzed here indicate that the competitive

binding of CN2097 to the PDZ-binding domain does offer almost complete protection against NMDA-induced ganglion cell loss and indicates that the interaction of the NMDA receptor subunits to a PDZ containing protein (PSD-95) is the initiator of the cell death signal that mediates both necrosis and apoptotic-like pathways.

2.10 Discussion

The primary focus of this dissertation is exploring the nature of PDZ interactions, with the special interest in domain organization and multiplicity. As stated before, although modular in design, PDZ domains are often linked in groups within a multi-PDZ protein. Our work has confirmed previously reported data that the first two PDZ domains in the PSD-95 of the MAGUKs subfamily are immediately adjacent to each other and form a protease-resistant unit. The functional significance of such linkages is still uncertain, although it has been studied extensively lately.

2.10.1 PDZ1 and PDZ2 domain binding characteristics Our first aim was examining the postulated ability of the combined PDZ1-2 of PSD-95 to show significantly different binding specificity for C-terminal sequences than did the individual PDZ1 and PDZ2 domains [166]. We possessed both domains (rat ortholog sequence) expressed in *E. coli* and purified. PDZ1 and PDZ2 belong to Class I based on their preferred binding partner, and we recognized that there may be similar recognition properties based on this fact alone when testing the binding of our peptide ligands.

To evaluate binding to the short peptide ligands, we employed isothermal titration calorimetry (ITC). ITC has been successfully used to characterize the binding interactions in the case of protein domains like Src homology 2 (SH2) and Src homology 3 (SH3). ITC presents great advantages over other methods (see Chapter I), as it characterizes the molecular interactions with the native, unmodified form of macromolecules in solution. It allows the measurement of thermodynamic parameters; namely, the change in enthalpy (ΔH), binding constant (Kd), and stoichiometry (*n*) from a single experiment. The change in Gibbs free energy (ΔG) and change in entropy (ΔS) can be calculated using equations (1) and (2) respectively. ITC is a method of choice extensively used for binding studies of PDZ3 of PSD-95 and published in our lab, Saro et al.[136, 167]

Lacking structural information for our PDZ1/PDZ2 system, which is abundant for PDZ3, we have relied on yeast two-hybrid results published earlier and well-described interactions with the NMDA receptor. A series of short peptide ligands, corresponding to the C-termini of endogenous binding partner proteins, were designed and tested against each individual domain. PDZ1 showed characteristic binding parameters, similar to PDZ3 reported for its natural binding partners [167]. Among the natural binders of PDZ1 of PSD-95, the first series of peptides studied by ITC were synthesized based on C-terminal sequence of K^+ channels of the Shaker type Kv.1 family. The first two peptides (8 amino acid residues) were designed to probe the molecular recognition of binding. The C-terminal ends with Thr-Asp-Val ; and having Thr at -2 and Val at position 0 is consistent with the Class I PDZ preferences. The last two peptides (8-mer and 6-

mer) are derived from the NMDA receptor NR2b subunit. This is one of the most well-studied interactions of PSD-95, and is critical for a number of neuronal functions such as synaptic transmission and synaptic plasticity. The best binder proved to be that derived from the Kv1.4 sequence, with K_d of 5.93 μM, and with a slightly enthropically driven of -4.42 kcal/mol. Potassium ion channel Kv1.3 had almost two times reduced binding preference for PDZ1, K_d of 13.89 μM, with favorable enthalpy and similar contributing entropy (Table 2.6 and 2.7).

Interaction with the octapeptide derived from the C-terminus of the NR2b subunit was weaker than expected at K_d of 11.2 μM, and the interaction is primarily entropic at 3.88 kcal/mol. Surprisingly, the hexapeptide sequence exhibited a diminished signal, and no thermodynamic parameters could be calculated. Results for all the natural PDZ1 binders show the changes in the enthalpy (ΔH) and in the entropy (ΔS) for the binding of the peptides to the PDZ3 protein were favorable, reflected in an enhanced, *i.e.*, negative change, in Gibbs free energy (ΔG) for the interaction. These findings are consistent with binding affinities for PSD-95, and comparable to the extensively studied PDZ3 [168] interactions with its natural binding partners.

PDZ2, on the other hand, was unstable during titration and would precipitate in the cell. After extensive trials, only a few complete datasets were collected due to these problems. These showed a binding stoichiometry of 0.5, which clearly denotes a problem with the samples; we hypothesized that the protein is poorly folded and all binding sites are not fully functional (Fig. 2.9).

Since our primary focus was on PDZ1-2 we have discussed individual domain data merely in a context of control experiments.

2.10.2. PDZ1-2 dual domain binding characteristics. The PDZ1-2 dual domain had expressed well and demonstrated an increased stability as compared to the single domains. Binding thermodynamics of the peptide ligands based on the potassium ion channel Kv1.4 sequence provided some interesting functional insight.

Sequence truncation of peptide binders (from 8 to 6 residues) did not diminish binding affinity, which would indicate that the maximum contribution of binding is found in the residues of the hexapeptide. Results (Table 2.8) were comparable to data for the single PDZ1 domain. The dual domain showed an increased affinity of K_d = 3.86 μM for the AKAVETDV peptide, as opposed to K_d = 5.93 μM for PDZ1. Changes in ΔH and in ΔS for the association of the peptides with the domain had roughly similar values, although binding entropy was very unfavorable (TΔS = -3.13 kcal/mol) indicating a substantial loss of degrees of freedom upon binding (Fig. 2.15).

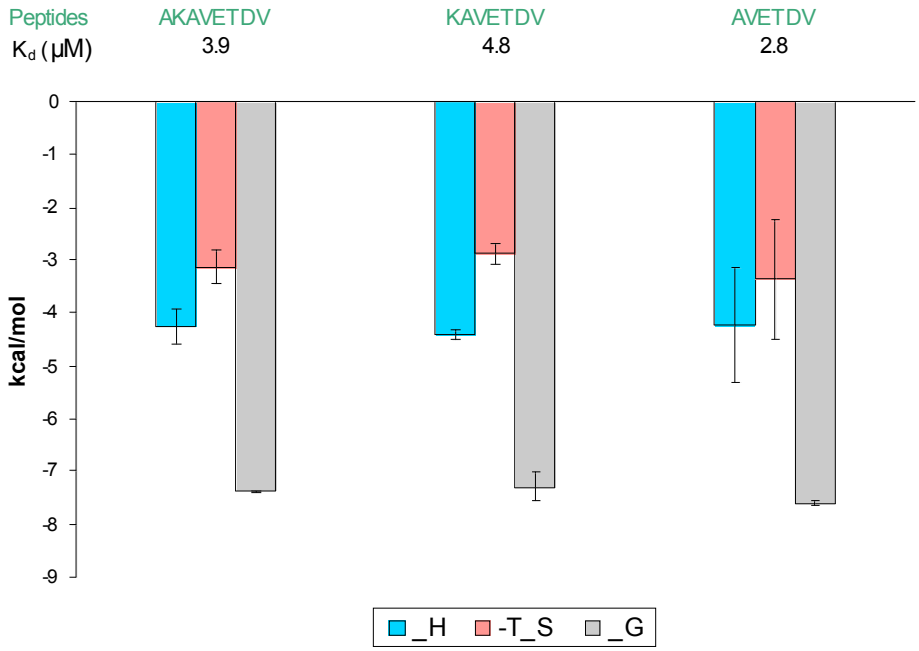

Figure 2.15. Binding energetic of PDZ1-2 domain to its natural derived peptides.

When integrating the ITC data, each of three models were applied (the One Set of Sites (OSS), the Sequential Binding Sites (SBS), and the Two Sets of Sites (TSS) models, each as implemented in the Origin program). In addition to assessing whether the assumptions of each model made biochemical sense, the chi-squared (χ^2) value was used as the primary criterion for the data fit. OSS yielded the best fitting of the binding isotherm, which appeared contradictory at first, since there are two binding sites in the domain. Taking into account all our data presented, we can only conclude that these two binding sites are not equally exposed or properly folded. Our results are in good agreement with published work by Cierpicki et al.[169] investigating the dual PDZ domain of syntenin. Their

crystallographic data showed that peptide binds to a single binding pocket. No electron density was observed for the lower affinity domain. Their NMR titration, however, showed that peptides bind to both PDZ1 and PDZ2, although the magnitudes of the associated chemical shift changes differ significantly. The data are consistent with fast exchange kinetics and relatively weak affinities [169].

At this point, it is worth mentioning that although ITC represents a very powerful technique with all its numerous advantages (outlined in Chapter 1) it is often hard to deconvolute the heat signal. Therefore, any interaction beyond the simplified 1:1 binding stoichiometry is challenging, especially since the relevant literature is very scarce. Work presented on dual domain binding, such as presented in this dissertation, is unique in the field. At the same time it was clear that without the proper structural information on PDZ1-2 tandem of PSD-95, it would be extremely difficult to progress with binding studies. Our efforts toward solving the structure of the dual domain are discussed in detail in Chapter 3 of this dissertation.

2.10.3. PDZ1-2 dual domain binding to modified cyclic peptide ligand (CN2180). In an ongoing effort to develop new cellular probes for *in vivo* investigations on the nature of protein-protein interactions, our lab has chosen PSD-95 as a target of interest. This is due to the extensive biological information available, and especially due to its important roles in the functioning of the post-synaptic density of neuronal cells. Initially designed to bind PDZ3, modified ligand CN2180 was developed and synthesized by Chamila Rupasinghe in the Spaller lab. CN2180 is a cyclic peptide based upon KNYKQTSV, in which the

ring bridges two side chains residues via β-alanine linkage (Figure 2.7). An elaboration of a simpler cyclic peptide structure that itself had been shown to exhibit good binding to PDZ3 [128], CN2180 contains two additional N-terminal residues and a myristic acid moiety to promote membrane permeability. However, screening (with a N-biotinylated version of CN2180) against a membrane-bound array of 96 different PDZ domain constructs demonstrated that comparable or enhanced binding occurred with the PDZ1-2 dual domain of PSD-95.

In light of the exquisite results that we were observing for the biological activity of CN2180 and its derivatives (in collaboration with Dr. Goebel, WSU), in parallel we have tested *in vitro* binding to the dual domain PDZ1-2 by ITC. After applying the OSS curve fitting, it became clear that the stoichiometric *n* value is 2, meaning that both binding sites on the tandem PDZ1-2 are fully occupied. However, fitting of the data improved significantly if SBS model was implemented (details Appendix C), as judged by χ2 values. Therefore, it is clear that the Sequential binding Sites Model gives the best description of the experimental data.

CN2180 populates both binding sites in the PDZ1-2 tandem, with fairly good binding affinity for each domain (K_{d1} = 6 μM and K_{d2} = 9 μM). Although two different values are obtained, the model fitting does not differentiate between the binding sites. These values are comparable to the previously obtained result for PDZ3 (K_{d1} = 8 μM), and thus make this compound a reasonably effective inhibitor of the PSD-95 interaction with the NMDA receptor (NMDAR).

***2.10.3.1. In vivo* validation of CN2180 activity.** Thermodynamic validation of CN2180 binding to PDZ1-2 represents a strong support of the *in vivo* tests carried out in rat retinal model. CN2180 peptide and its polyArg counterpart (CN2097) showed strong activity in the NMDA-induced stressed neurons. Decline in NMDA-induced PAR-formation is KNYKKTEV peptide-dependent, and that competitive binding to the PDZ domain of PSD-95 (or related proteins) containing this domain, aids in diminishing NMDA-induced PARP-1 activation. This is consistent with the understanding that PSD-95 is linked to nNOS (neuronal nitric oxide synthetase), which produces NO within the neuron. Also, results shown indicate that NMDA-induced cell death is mediated by the NMDA-receptor's linkage to PSD-95 and connect this finding, for the first time, that disconnection of the NMDA-receptor from PSD-95, through competitive binding of the peptide-ligand(s) with the PDZ-domain of PSD-95 (or similar protein), renders the NMDA-receptive neurons insensitive to NMDA-induced necrosis.

More extensive analysis of these experiments is still forthcoming, although indications from the two retina pairs analyzed here clearly point out that the competitive binding of CN2097 to the PDZ domain-binding domain does offer almost complete protection against NMDA-induced ganglion cell loss and indicates that the interaction of the NMDA receptor subunits to a PDZ containing protein (PSD-95) is the initiator of the cell death signal that mediates both necrosis and apoptotic-like pathways.

2.11 Conclusions

The central purpose of this dissertation research was to investigate multiplicity in PDZ domain binding, and our system of choice was a biologically critical and abundant protein in the cytoplasm of neuronal cells, PSD-95. Since the broader goal of our laboratory was exploring the nature of PDZ interactions, both monovalent and multivalent, we had independently expressed three separate constructs: PDZ1, PDZ2 and PDZ1-2. A series of peptide ligands derived from the C-terminus of natural binding partners were synthesized and tested against all three proteins. In the case of PDZ1, association of these peptide ligands is driven by dominant favorable enthalpic contributions to the binding free energy. The interactions are usually entropically favorable. The reported assoiciation constant are in good agreement with the previously published fluorescence data [170].

PDZ2 did not express well in recombinant form from bacteria, and it was likely misfolded, and therefore the resulting ITC binding data with peptide ligands was inconclusive. Future efforts in this area should perhaps focus on using the full-length (uncut) GST-PDZ fusion for binding studies. Although GST fusion proteins can dimerize, the alternative would be the preparation of other types of fusion protein, such as with maltose-binding protein (MPB) or NusA and use these for the binding studies.

The dual domain PDZ1-2 showed strong binding preferences toward the same ligands, although a stoichiometry of n=1 clearly points out that only one binding pocket equivalent is occupied. In the case of a very strong binding to

modified peptide CN2180, the stoichiometry of n=2 indicated favorable orientation and occupancy of both binding sites of the domains. We should not rule out, however, the possibility of CN2180 binding to a remote site on the surface of tandem domain, although there is no biochemical evidence in the literature relating to PDZ domains that would suggest this. A lack of structural information for the intact dual domain in complex with ligands encouraged our efforts to crystallize and solve PDZ1-2 tandem structure, further discussed in detail (Chapter 3). CN2180 showed very promising biological activity and efforts are ongoing in this direction.

CHAPTER 3

PRELIMINARY STRUCTURAL STUDIES OF THE TANDEM PDZ1-2 DOMAIN PROTEIN OF PSD-95

PSD-95 is one of the key proteins in the postsynaptic density. Although, its individual domains have been structurally characterized, PDB deposition for dual PDZ1-2 is only modeled. Also, ITC data was very difficult to interpret without true structural insight. In this Chapter we report for the first time successful crystallization of dual PDZ1-2 of PSD-95. This study was carried out in collaboration with Dr. Ladislau Kovari, WSU. The work is featured in a manuscript that is in preparation.

3.1. PDZ domain protein crystallography

X-ray crystallography represents a very powerful method used to determine the three dimensional structures of the major biomacromolecules, namely, proteins, DNA and RNA and their complexes. The information provided by a crystal structure is fundamental in explaining the biological function of macromolecules. Obtaining detailed pictures of large molecules at the molecular and atomic level was practically impossible before employing the technique of single-crystal X-ray crystallography [55, 171, 172]. This method utilizes diffraction of X-rays from many identical molecules in an ordered array of a crystal. It is important to point out that proteins mostly retain their fold, and therefore their function, in the crystal form, therefore obtaining high resolution structures helps in our understanding of proteins' biological functions in terms of their binding

properties and thermodynamics. A detailed explanation of the technique is provided in Chapter 1 of this dissertation.

3.1.1. PDZ domain structures. There have been a considerable number of PDZ domain structures, currently on the order of 200, deposited in the Protein Data Bank (PDB). This includes those solved by either X-ray crystallography or NMR; at least 120 of them were obtained by the former technique. The vast majority of these are structures of just the PDZ domain substructure subcloned from their larger parent protein, which indicates that this class of domain is relatively stable under laboratory conditions and well suited for structural studies. It also indicates that even though they might possess a similar three-dimensional fold, there exist many differences in each particular PDZ domain in both the free or complexed forms as revealed by these structures. Most importantly, these structures have contributed tremendously to our understanding of the biological role of these relatively small protein modules.

3.1.1.1 The general structural form of the PDZ domain. The numerous entries in the PDB databank illustrate vast interest in these PDZ proteins, but also underlie the importance of understanding protein domain organization in general. Most importantly, these studies helped in formulating the rules of peptide (protein binding partner) recognition, which is essential for their biological function. One of the first studies that revealed the molecular recognition events taking place in the PDZ mediated complex formation involved solving the X-ray crystallographic structure of PDZ3 of PSD-95 complexed with a nonapeptide ligand derived from the PDZ3 binding partner, CRIPT protein [88].

The crystal structure (Figure 3.1) of PDZ3 comprises six β strands (βA-βF) and two α-helices (αA and αB), which fold in an overall six-stranded β sandwich. Four residues of the ligand's C-terminal region have been shown to interact with the peptide-binding groove between the βB strand and αB helix as an antiparallel β strand extending one of the β sheet. The C-terminal hydrophobic residue valine at the "0" position of the peptide binds to the hydrophobic pocket created in the PDZ domain by several hydrophobic residues, which are Leu323, Phe325, Ile327, and Leu379 [88]. The tight contact of the side chain of valine with the hydrophobic pocket explains the preference for the sequence having a hydrophobic residue at the "0" position. Another critical interaction is the hydrogen-bonding interaction of the hydroxyl group of the "-2" threonine with N3 of His372 (first residue of the αB helix). The side chain of the "-1" Ser residue is shown pointing away from the interacting surface.

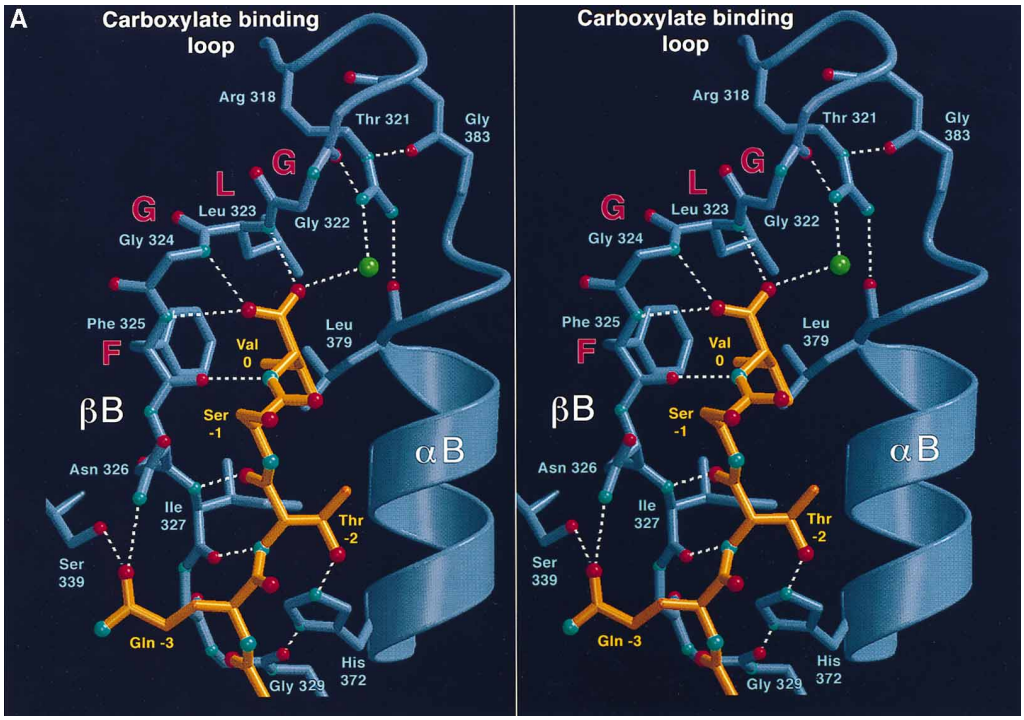

Figure 3.1 Stereo view of the peptide binding site demonstating protein – peptide interactions [88].

The peptide (yellow) inserts between the β strand and β helix and forms an antiparallel sheet with βB. The connecting loop between αB and βB is involved in binding the peptide C-terminus and therefore is designated the carboxylate-binding loop.˚

More recently, crystal structure of the same protein at higher resolution was solved by Dorina Saro in Spaller lab. Two different structures of PDZ3 complexed with hexapeptides KKETWV and KKETPV were deposited, with PDB codes 1TP5 and 1TP3, respectively. The method employed in both cases was that of molecular replacement (details in Chapter 1). Figure 3.2 depicts the domain in association with KKETWV (shown in dark blue) and its position between β strand and β helix.

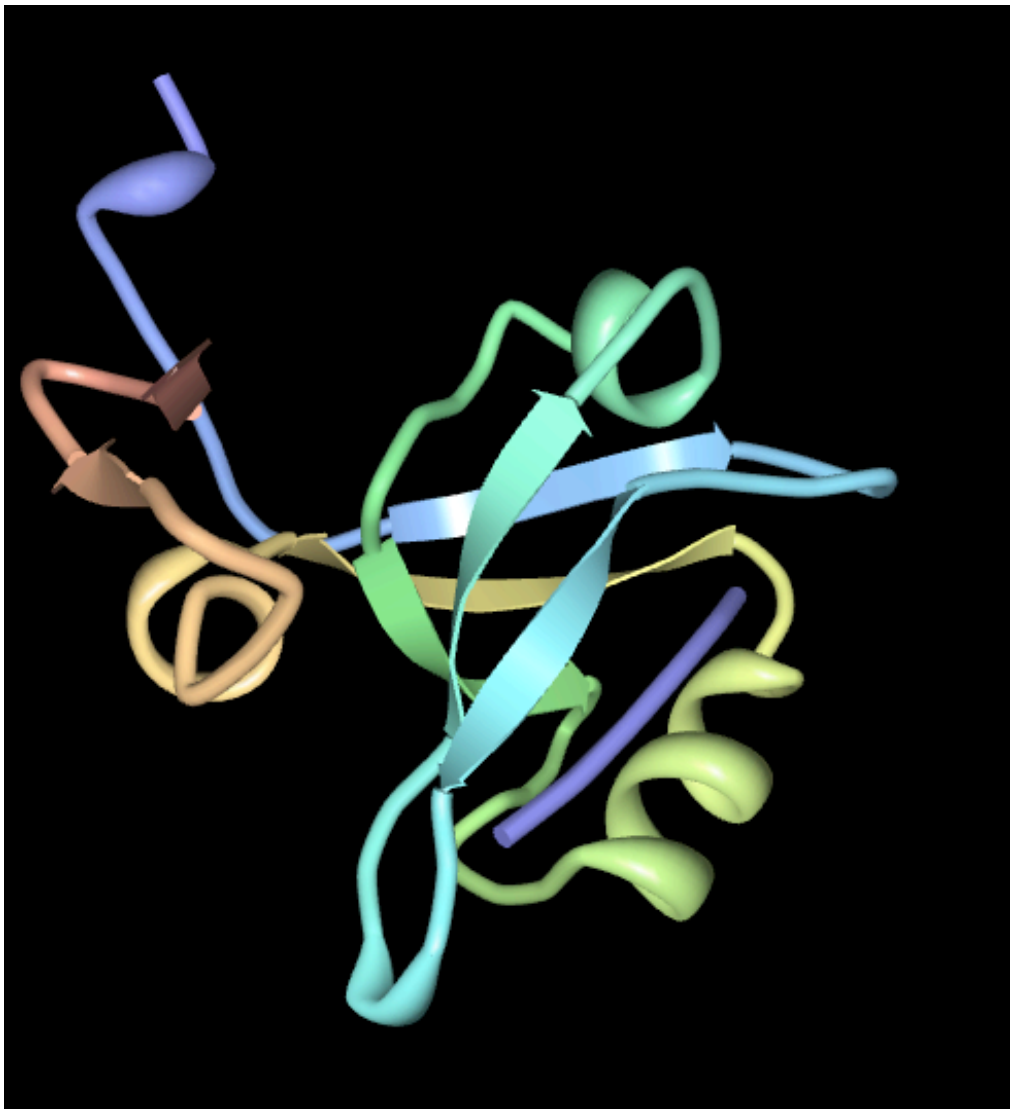

Figure 3.2 Ribbon diagram showing PDZ3 complexed with KKETWV, PDB code 1TP5. Ligand shown in dark blue. Image generated by Protein Workshop.

Crystals of PDZ3 soaked with KKETWV peptide diffracted at high resolution, 1.54 Å. The structure of the PDZ3/KKETWV complex represents the highest resolution structure of the PDZ3 of PSD-95 available, and one of the highest resolution structures of all PDZ domains solved to date. These structures reveal the specific interactions that play an important role in the PDZ3 recognition

of these ligands. Even though these peptides have similar sequences and identical residues at the critical “0” and “-2” positions, they still bind to PDZ3 domain with different thermodynamic properties and this is a result of the structural differences in each complex formation.

Structural studies have also discovered important biological functions of these domains. The differences in C-terminal peptide recognition between class I and II PDZ domains became clear with the information from the complex structures of PDZ domains of each class [88, 120, 173, 174]. The recognition of internal sequences by some PDZ domains were also revealed from structural data [121, 122]. The effect of chaperoning and intermolecular interaction of PDZ domains with each other [122, 174-176] is one aspect of PDZ function known from the structural data. Novel structural features of PDZ domains, like the presence of alternative binding pocket [177], or novel mode of peptide recognition by a PDZ domain [178], have also become clear from analysis of crystallographic and NMR structural information. Diverse biological functions of PDZ domains were discovered from structural studies as well. One example of this is the study of DegS stress sensor PDZ domain that recognizes misfolded proteins and activates the proteolytic activity of this enzyme [179]. To further explain this behavior, the function of PDZ domains in DegP (or also known as HtrA) protease might be very similar, as described by Krojer *et al*. [180], indicating a possible general mechanism of recognition and binding of misfolded proteins by PDZ domains.

Interestingly, the most precise crystal structure of a protein known so far is the PDZ2 domain of syntenin protein. This structure is defined at ultra-high resolution of 0.87 Å with crystallographic refinement values of R = 7.5% and R_{free} = 8.7% [181]. These refinement values are typically much higher for protein structures (in the range of 20 to 25% for well-refined structures)[182] and closer to the small molecule refinement, “bridging the gap” [181] between the two fields. Being an ultra-high resolution structure, the model of PDZ2 of syntenin also shows important features of protein structures like the flexibility of the peptide bond, and the presence and stereochemistry of weak C–H---O hydrogen bonds [181].

3.1.1.2 Dual or tandem PDZ domains. An important question for which the answer remains largely unknown and that we have tried to address is the manner of binding action for multiple PDZ domains. Specifically, this entails the form involving two vicinal PDZ domains, which are referred to as *tandem* or *dual domains.* Such tandem forms may be critical for the biological activity of the protein in which they are found. Examples include neuronal nitric oxide synthase (nNOS) [183] and the cystic fibrosis transmembrane conductance regulator (CFTR) protein [110, 184, 185].

In the case of PDZ1-2 from PSD-95, which is a focus of this dissertation research, isothermal titration calorimetry data yielded results that were difficult to interpret. Using different experimental conditions, it appeared that only one peptide ligand was binding to the dual PDZ1-2 domain. These finding were discussed in detail in Chapter I. Also, thermodynamic experiments have their

own limitations and could not answer some important questions: how do tandem domains bind to similar targets, and where does the difference in specificity come from? These questions are best answered by structural methods using a tandem PDZ1-2 domain as a target to crystallize in both the free and ligand bound forms.

Postsynaptic density protein-95 (PSD-95; also known as synapse associated protein 90(SAP 90))[85] found in high concentrations at the post synaptic densities of dendritic spines. All the members of this family consist of three PDZ domains, followed by one SH3 domain and one guanylate kinase homology (GK) domain Figure 3.3.

Figure 3.3 Domain organization of PSD-95 protein.

Interestingly enough, a NMR structure of tandem PDZ1-2 domain of PSD-95 was previously published [186]. However, this was not a true solution structure, but the result of modeling the two domains together, each of which had been solved independently. The PDZ1 domain NMR structure was also previously solved and deposited in the PDB (codes 1iu0 and 1iu2). Relying on other experimental results, PDZ2 domain was modeled. A mechanism of binding was hypothesized, and modeled as well. This study is the first to characterize the three-dimensional structure of a tandem (supramodular) array of PDZ domains.

Previously, the same group has solved the free structure of PDZ2 and deposited in the PDB (PDB code 1qlc) [187].

Although these modeled results are intriguing, it is unclear why an effort was not made to solve a true NMR structure of dual domain? One can only speculate that it was attempted, but for unreported reasons was not successful. In the absence of an explanation from the reporting group, it is not unreasonable to suppose that the NMR methodology did not yield solvable information on the actual tandem construct. Thus, there remains a large hole in our understanding of how the PDZ12 dual domain of PSD-95 is truly structured. The availability of this structure would add considerably to our understanding of the endogenous binding interactions that PDZ12 engages in. Besides helping to shed light on this fundamental character of PDZ12, it would also greatly aid efforts to design novel inhibitors that could specifically and simultaneously target both domains. As described in Chapter 2 there are profound biological—and potentially therapeutic—effects associated with small peptides that target PDZ1 and PDZ2. Therefore, a significant effort was expended in attempts to solve the structure of the first two PDZ domains of PSD-95 protein by X-ray crystallography.

3.2 Materials and Methods

3.2.1 Materials Crystallization plates were obtained from Nextal Biotechnologies Inc. (Qiagen). For fast setup of crystallization conditions screw-in, greaseless crystallization supports were used that also provide easier setup and crystal recovery. The EasyXtal Tool is supplied as an empty 24 well plate with a capacity of 1 mL of solution per well and 24 EasyXtal CrystalSupports. For initial screening of crystallization conditions the *Wizard™ I&II* (Emerald Biosystems) was used. *Wizard™ I&II* screens provided full 96 unique formulations of the *random sparse matrix* for the crystallization of biological macromolecules. A Leica GZ7 microscope was used to check the plates after set up and to observe the crystallization process. Diffraction data were collected from several crystals on a Rigaku/MSC FRD generator/R-AXIS HTC image plate system. Silicon Graphics and Monarch Computer Systems Unix box were used for computer calculations.

3.2.2 Methods

3.2.2.1 PDZ1-2 protein expression and purification The PDZ1-2 dual domain of PSD-95 was expressed and purified as described in section 2.3.2.1 of Chapter 2.

3.2.2.2 Crystallization of dual PDZ1-2 protein The PDZ1-2 samples were dialyzed into the final dialysis buffer (10 mM sodium phosphate, pH 8.0). The sample was then concentrated up to 15 mg/mL using Amicon centrifugal filter devices 3000 MWCO (Millipore) and stored at 4 °C until needed. Preliminary crystallization trials were performed using *Wizard™ I&II*. The hanging-drop

vapor-diffusion technique was utilized. For initial screening, concentrated PDZ1-2 protein (1 µL) was mixed with an aqueous buffer containing precipitant (1 µL). The cover slip with the prepared droplet was inverted over the same precipitant aqueous solution (0.7 mL) in the well, and after the set up the plates were incubated in the dark under controlled temperature (22 ºC) and humidity conditions. After 1-2 weeks micro crystals were observed in three out of 96 conditions; these were from *Wizard™II* crystallization screens listed below:

Solution 1: 10% (w/v) PEG-3000 , 100mM acetate(pH 4.5) , 200mM $Zn(OAc)_2$

Solution 19: 1.6 M NaH_2PO_4/0.4 M K_2HPO_4, 100mM phosphate-citrate (pH 4.2)

Solution 40: 20% (w/v) PEG-3000, 100mM imidazole (pH 8.0) ,200mM $Zn(OAc)_2$

These conditions were repeated, but now with varying protein to precipitant ratios (1:1, 1:2, 1:3, 1:4) and with the same ratios for precipitant to protein. Chemicals were purchased from following vendors: PEG-3000 (Fluka), $Zn(OAc)_2$, NaH_2PO_4, K_2HPO_4, imidazole, acetic acid, sodium acetate, citric acid (Sigma). Controls were run alongside with final dialysis buffer.

Microcrystals were observed only in the drops formed with 10% (w/v) PEG-3000, 100 mM acetate (pH 4.5), 200mM $Zn(OAc)_2$. These were further optimized resulting in good size crystals (0.2-0.4) mm obtained by mixing 4 µL of high purity protein sample (15 mg/mL) with 1 µL of precipitant. Crystals formed within one week. Hexagonal PDZ1-2 crystals were stable for approximately three months in mother liquor, but started to degrade afterwards. These crystals were

directly flash-frozen in liquid nitrogen and preliminary test showed very good diffraction. Therefore, it was concluded that the presence of PEG-3000 was sufficient and no additional cryoprotectant was needed.

3.2.2.3 PDZ1-2 data collection Diffraction data were collected from a single crystal on a Rigaku/MSC FRD generator/R-AXIS HTC image plate system at WSU Medical School. The temperature during data collection was -180 °C due to continuous liquid nitrogen stream over the crystal. The crystal-to-detector distance was 150 mm. Copper radiation (1.5418 Å) was used. For a complete data set, 360 frames with 0.5° oscillation were collected. The exposure time was 3 minutes for each frame. Diffraction data were collected and processed using Rigaku/MSC CrystalClear software.

3.3 Results

3.3.1 Crystallization properties of the free PDZ1-2 dual domain The conjoined first two PDZ domains (PDZ1-2) from PSD-95 were overexpressed as a GST fusion protein and purified using a slightly altered purification protocol. The 190-amino acid long tandem domain was separated from the GST affinity tag by trypsin cleavage. After further purification the isolated PDZ1-2 protein appeared very stable when stored at 4 ºC, even though it was highly concentrated (up to 15 mg/mL).

The crystallization setup using optimized conditions (10% (w/v) PEG-3000, 100mM acetate (pH 4.5), 200mM $Zn(OAc)_2$) as explained in detail in (3.2.2.2), provided for well shaped crystals as shown in Figures 3.4 and 3.5.

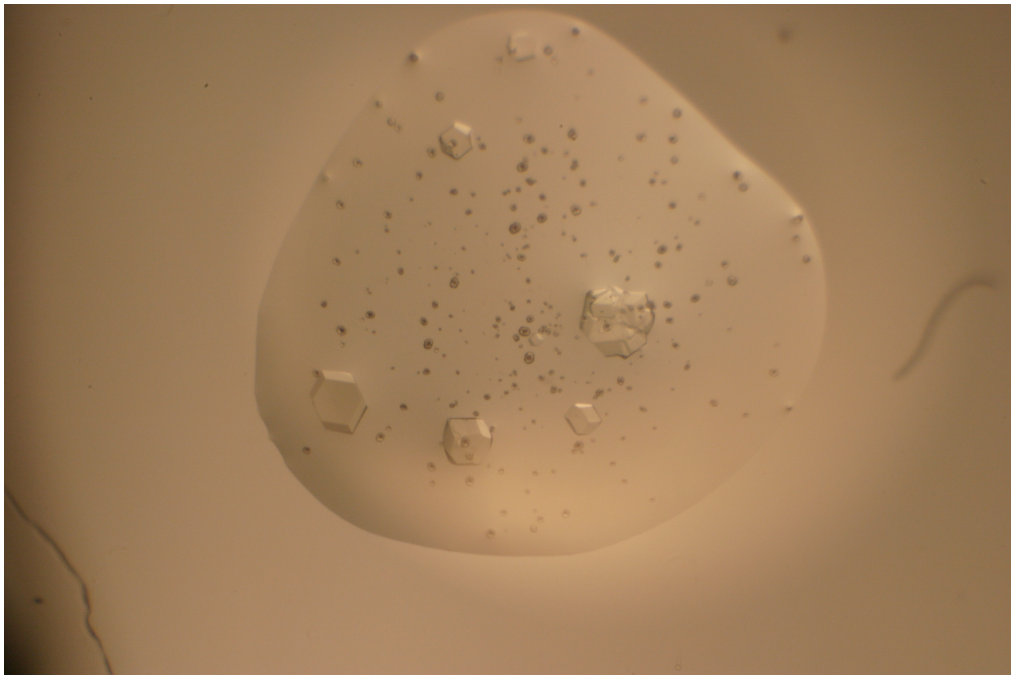

Figure 3.4 PDZ1-2 dual domain crystals in a 5 µL hanging drop (crystal diameter 0.4 mm).

Figure 3.4 is a photograph of a one week old crystals in the 5 µL drop. Nicely shaped hexagonal crystals, well dispersed in mother liquor can be easily separated. No skin formation on the drop also made the process of mounting the crystal much easier. Crystals were flash-frozen without the additional cryoprotectant.

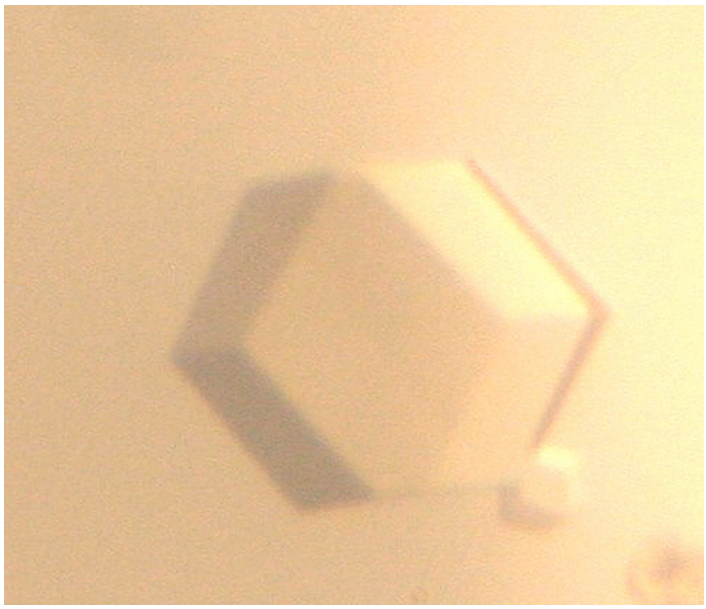

Figure 3.5 Single PDZ1-2 crystal. Hexagonal crystals grew in a week in 10% (w/v) PEG-3000, 100 mM acetate (pH 4.5), 200 mM $Zn(OAc)_2$. (crystals 0.4 mm).

3.3.2 Data collection. The crystals formed were of very good quality and stable while flash-freezing in liquid nitrogen. Upon assessing crystal diffraction quality by collecting initial frames, obtained diffraction pattern was very good and estimated a resolution was high. Therefore, the same crystal was left in the X-ray beam for the complete data set collection (360 frames, 0.5º oscillation and 3 minutes exposure time per frame). A typical diffraction pattern of PDZ1-2 dual domain protein is shown in Figure 3.6.

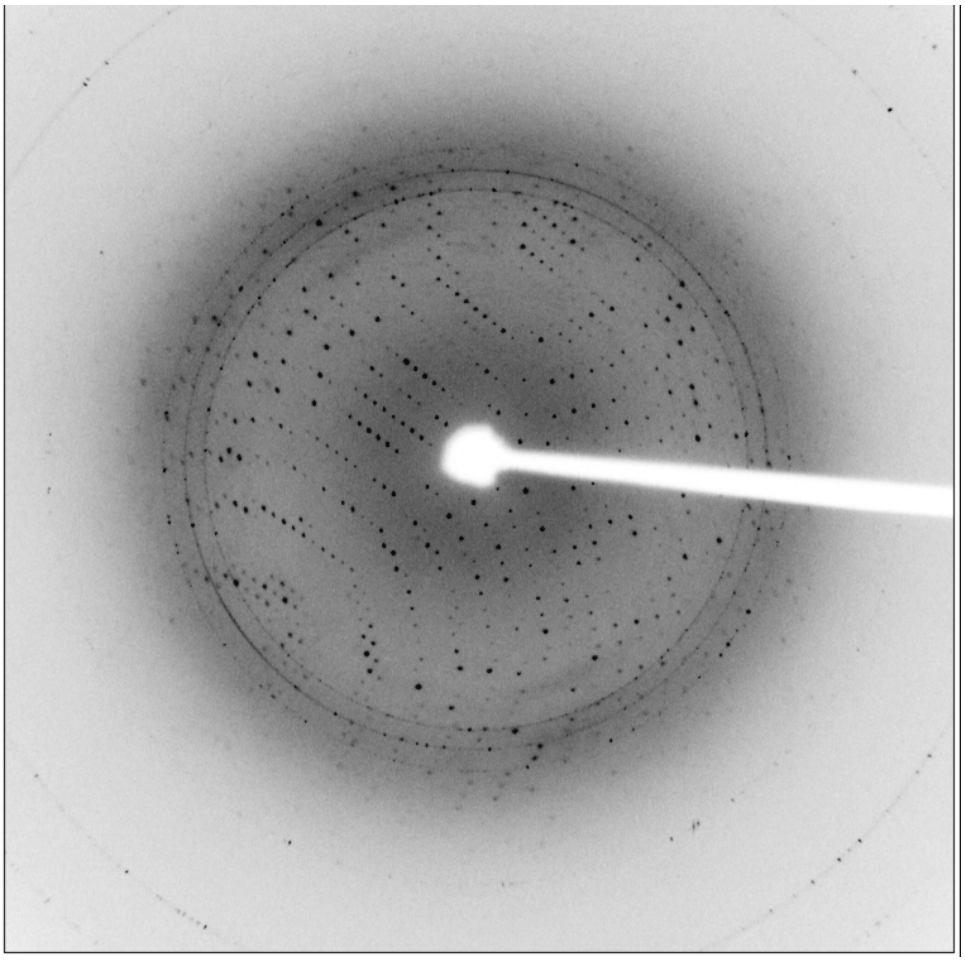

Figure 3.6 Diffraction pattern-single frame of free tandem PDZ1-2. Exposure time 3 minutes.

3.3.3 Data reduction. The crystal of PDZ1-2 dual domain protein diffracted at good resolution (2.6 Å). A complete data set of 360 frames were collected, 0.5º oscillation angle and exposure time of 3 minutes per frame. This data set was collected using a Rigaku/MSC FRD generator/R-AXIS HTC image plate system. Data collection and initial processing were conducted using CrystalClear for Rigaku. CrystalClear incorporates the concept that data acquisition and data processing should be combined. Further data reduction was performed utilizing MOSFLM software. MOSFLM processes diffraction images and gives a MTZ file

as an output, which contains reflection indices with their intensities and standard deviations. This MTZ file is then used by other programs of the CCP4 program suite for data reduction, scaling and merging. Table 3.1 lists the parameter values after processing data with SCALA from the CCP4 package. These results indicate that PDZ1-2 dual domain form cubic crystals that belong to the P23 space group. This is a high symmetry space group, which, unfortunately, has historically proven to be a problem when attempting to solve the structure using the molecular replacement method.

Table 3.1 Data collection statistics for the dual PDZ1-2 protein domain

Resolution range (Å)	29.09 – 2.60	(2.74 – 2.60)
Unit –cell parameters (Å)	$a = b = c = 87.2667$ Å	
	$\alpha = \beta = \gamma = 90°$	
Space group	P23	
R_{merge}	0.112	(0.630)
No. of total reflections	149876	
No. of unique reflections	7052	
Completeness (%)	99.9	(100.0)
Average $I/\sigma(I)$	29.0	(5.1)
Multiplicity	21.3	(21.4)

Values in () are for the last resolution shell

3.4 Discussion

A PDZ1-2 dual domain protein that was expressed and purified as discussed, was successfully crystallized. Purification protocol for this protein was altered and explained in detail in Chapter I of this report. This modified purification scheme resulted in a highly pure and stable protein. Crystallization set-up experiments using optimized conditions (10% (w/v) PEG-3000, 100 mM acetate (pH 4.5), 200 mM $Zn(OAc)_2$) explained in detail in (section 2.2.2), provided for well-shaped hexagonal crystals. PDZ1-2 crystallizes in cubic P23 group. Data reduction gave a R_{merge} value of 11.2 %, which proves that this was a high quality data set.

We tried to solve the PDZ1-2 dual domain protein structure using the molecular replacement method. The rationale for this was that structures of several PDZ domains are known and can be used as a model for this technique. The first model used was that of the crystal structure of PDZ3 domain of PSD-95 protein [88], complexed with a peptide ligand. This attempt failed to yield an acceptable solution, as this model represents only a single PDZ domain and we are looking for the structure of a dual PDZ domain containing protein. The structure of another dual PDZ domain protein syntenin was also used in molecular replacement calculations, but likewise did not result in a clear solution. [188]

The 1n99.pdb file of PDZ tandem of human syntenin was downloaded from the PDB and used as the model (Fig 3.7). The file was modified by removing solvent molecules, and used as a model during calculations. In order to

ensure that these changes to the template model really occurred, the structure was inspected using PyMol display software, and in fact it was confirmed to be a single polypeptide chain.

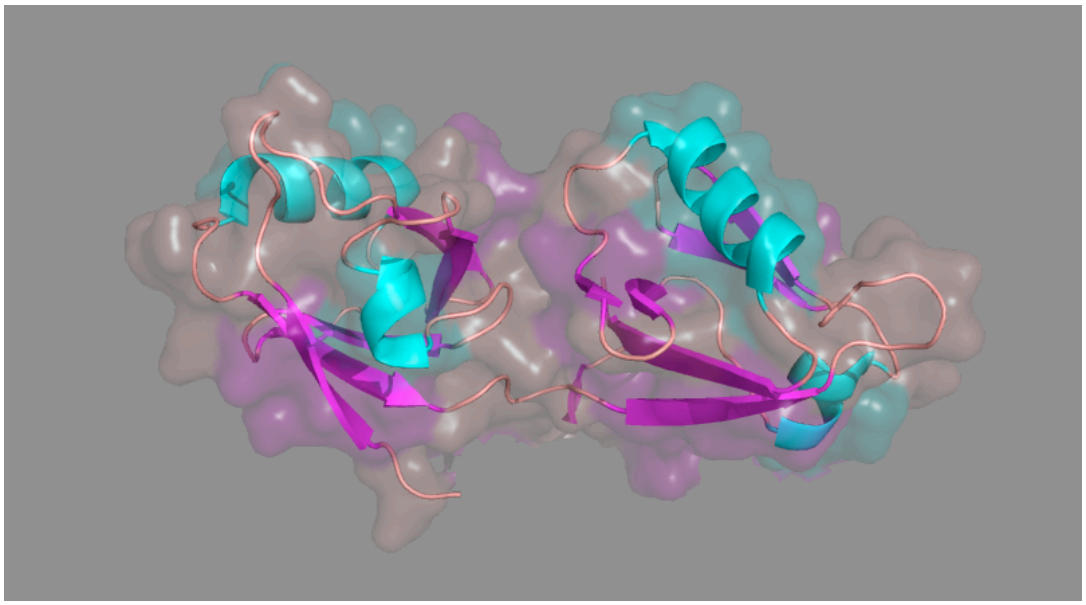

Figure 3.7 PDZ 1-2 of synthenin. PDB code: 1N99. Image generated with PyMol.

The programs utilized for solving the structure were from the CCP4 package; Refmac 5 was used. The template file was modified with rigid body refinement at 4 Å. After the first round of Refmac 5 was successful, a rigid body refinement at 3 Å was performed, followed by restrained refinement at 3 Å. This round was followed by restrained refinement to the limited resolution. In our case that was 2.6 Å. The Mtz output file was created, followed by twenty cycles of ARP/wARP and adding water molecules to the structure resulted in file 10_solvent.pdb. An image from this file is reproduced below in Figure 3.8.

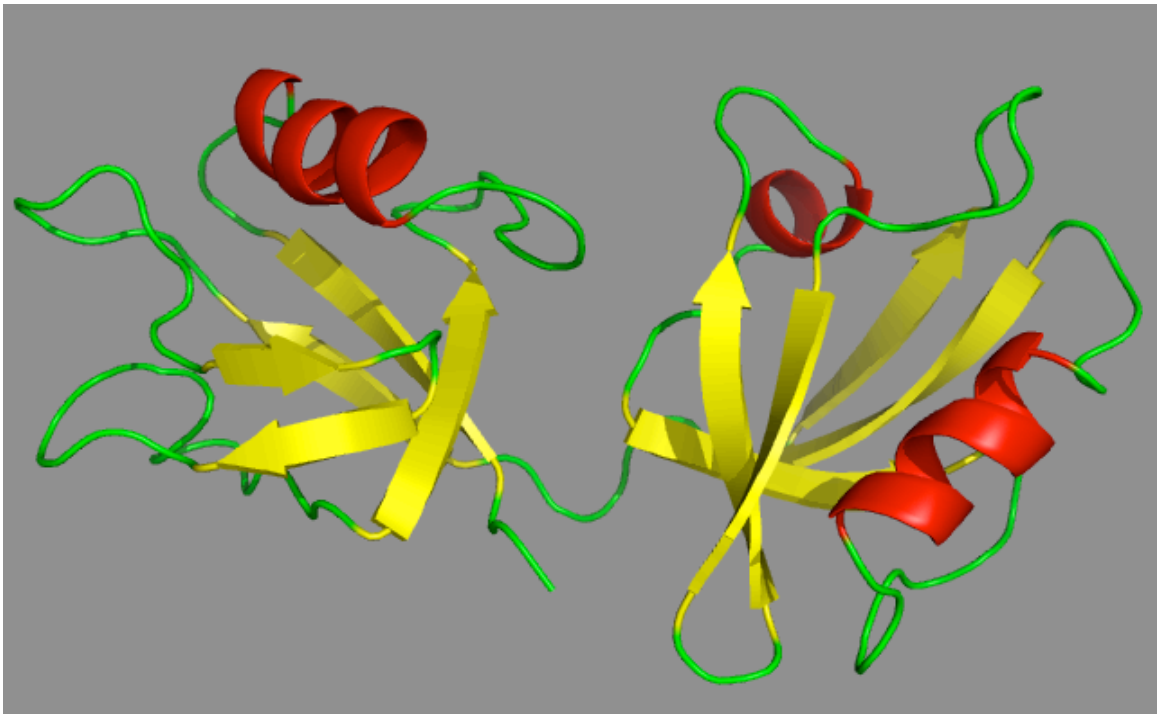

Figure 3.8 Image of the 10_solvent.pdb. Image generated with PyMol.

This image represents mostly a random coiled structure, but it can be clearly seen that there are some structured regions. However, this does not represent a true solution of the PDZ1-2 tandem domain structure, since R = 47.7 and Rfree = 52.8. As these values were very high (more than 50%), this indicates that the experimental data are not in good agreement with the model. A good starting point for R/Rfree values is around 40%, and further structure refinement can bring those values down to acceptable values of 20 – 30%.

Another view of the hypothetical structure denoted 10_solvent.pdb is depicted in Figure 3.9. The characteristic fold of PDZ domains of six β-strands and two α-helices can be easily visualized in both domains, although for the

reasons explained above, this rendering cannot be accepted with confidence as a true solution structure.

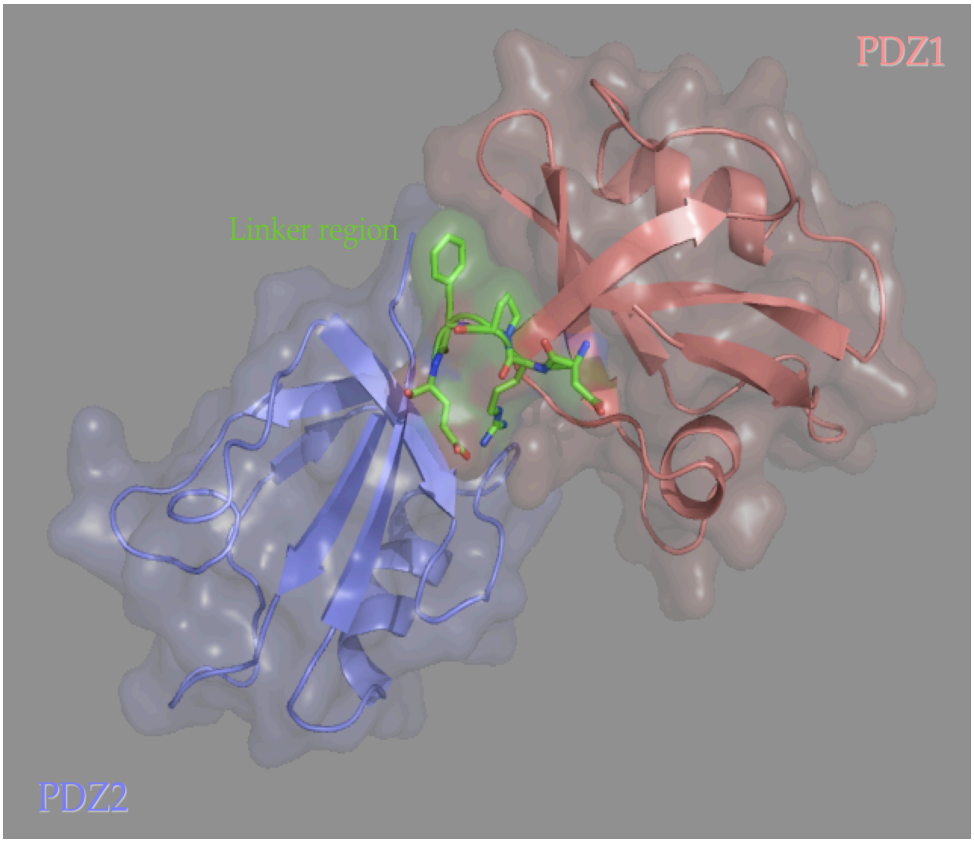

Figure 3.9 Image of the 10_solvent.pdb. Image generated with PyMol.

The reason the molecular replacement calculations failed relate to the different space groups of the model and our molecule. An alternative way to solve this problem would be to use AMoRe [189], another molecular replacement program that has the advantage of transforming both the model and the new structure into a hypothetical space group and then performing the searches. If the molecular replacement methods do not work, then the structure can still be solved by another method, like heavy atom derivative or MAD (multiwavelength anomalous diffraction).

The problem with molecular replacement methods not being able to solve the PDZ1-2 structure, despite the fact that there is good quality diffraction data available, might be that the PDZ1-2 dual domain crystallizes in a different way than the previously known structures. So even though structures of separate PDZ domains are known, their relative orientation in the PDZ1-2 dual domain might be different and that accounts for significant changes between the model and the experimental data that the molecular replacement method cannot resolve.

Since PDZ12 of PSD-95 is a novel unsolved protein, never crystallized before, further efforts to solve the structure should focus on using methods other then molecular replacement. One way is producing heavy metal derivatives. Anomalous scattering methods can be used to calculate phases based on the advantage that a heavy-atom can absorb X-rays of specified wavelengths. Also, multiwavelength anomalous dispersion (MAD) phasing can be used as an alternative to solving the phase problem in crystallography. For this purpose selenomethionine derivatives can be used. In this strategy, the protein is expressed in media that contains selenomethionine as the only source for amino acid methionine.

GSEYEEITLERGNSGLGFSIAGGTDNPHIGDDPSIFITKIIPGGAAAQDGRL
RVNDSILFVNEVDVREVTHSAAVEALKEAGSIVRLYVMRRKPPAEKVMEIK
LIKGPKGLGFSIAGGVGNQHIPGDNSIYVTKIIEGGAAHKDGRLQIGDKILAV
NSVGLEDVMHEDAVAALKNTYDVVYLKVAKPSNA

Figure 3.10. Sequence of PDZ1-2 with highlighted methionines.

As a result selenium is incorporated as a heavy atom containing amino acid in the protein that can be further purified and crystallized as the wild type. If

that is successful, than data collection using synchrotron radiation with multiple wavelengths can result is solving the phase problem and the crystal structure of the protein.

3.5 Conclusions

In this thesis chapter, crystallization of a new protein construct, PDZ1-2 of PSD-95 is reported. Identifying the conditions required to grow crystals is a major step forward in efforts to solve this tandem domain by X-ray methods. Significant efforts were made to solve the structure using the molecular replacement method, but due to specific space group that protein crystallizes in, thus far solved models of similar proteins could not be implemented. However, producing a seleno-methionine derivative would gain a diffractable crystal and would enable to solve the structure of the protein.

CHAPTER 4

CHARACTERIZING BINDING PROPERTIES OF Fe(II) BINDING TO FRATAXIN

Further exploring the idea of multivalency and applying ITC as a true solution method for assessing affinity of protein–mediated interactions; another project was carried out in collaboration with Dr. Timothy Stemmler's lab at WSU Medical School. The study provided greater in depth understanding of cellular processes, this time on the example of metal binding to the iron chaperone-Frataxin. The work presented in this chapter has been published in Biochemisty, 2006.

Friedreich's ataxia (FRDA) is caused by deficiency in the protein frataxin. FRDA is as an autosomal recessive cardio and neurodegenerative disorder that affects approximately 1 in 50 000 people, as reported by independent studies [190-192]. At this rate, it is believed to be the most common autosomal-recessive neurodegenerative disease of Caucasians [193]. Frataxin is nuclear-encoded protein subsequently localized in mitochondria at the matrix side of the inner membrane [194, 195]. Cells depleted of frataxin will have progressive mitochondrial accumulation of iron. However, iron deposits are biologically unavailable, which leads to further decrease of metal activities – its role in iron-sulfur cluster enzymes, oxidative damage to mitochondria [196, 197]. Clinical picture of iron deficiency in humans manifests through hypertrophic cardiomyopathy, degeneration of sensory neurons in the dorsal root ganglia, and diabetes mellitus [195].

4.1 Mitochondrial iron metabolism Transition metals play a key role in regulatory, structural and catalytic properties of the cell [198]. Metal content in prokaryotic cells of iron, zinc and calcium is very similar and ranges around 0.1 mM [198]. The concentration of iron in mitochondria of yeast cells is the same as for prokaryotes, which is consistent with its role within the organelle. Iron is required for two major pathways in the mitochondria: iron-sulfur (Fe-S) cluster and heme assembly. These cofactors are found in key proteins like aconitase (Krebs cycle), complex I of the respiratory chain (electron transport) that require Fe-S clusters [199] or cytochromes, membrane proteins that utilize heme. However, redox ability of the transition metals, iron and copper in particular can generate toxic organic and oxide based radicals [200]. Maintaining iron concentration within the mitochondria is crucial for cell survival.

Iron is found ubiquitously in all living systems, although both metal deficiency and metal overload are destructive to cells. Through its protein controlled mechanisms organisms have evolved, so that iron import can be quickly enhanced or suppressed, or finally stored for the future use [201]. In humans, Fe obtained from diet is reduced and acidified by the low pH environment of the stomach and potion is transported for production of heme and Fe-S clusters. More than 2/3 of Fe found in the body is incorporated into hemoglobin in developing erythroid precursor cells or in mature red blood cells [202]. Such an abundance of heme prosthetic group is absolutely necessary for controlling O_2 concentration, availability and reactivity. As previously outlined, Fe-s clusters are another Fe containing prosthetic groups that help control cell

viability by taking part in electron transfer events, substrate binding and activation, iron and sulfur storage, regulation of gene expression and enzyme activity [203].

4.2 Heme bisynthesis. Heme is an iron-containing tetra-pyrrole ring system. The heme biosynthetic pathway was described in detail during 1950's and 60's [204]. Heme biosynthesis consists of eight sequential steps [205], four of which (step 1 and steps 6-8) occur in mitochondria, while the other steps occur in the cytoplasm. The final step is the insertion of ferrous ion into porphyrin ring by enzyme ferrochelatase to yield a functional heme prosthetic group [206]. In prokaryotes, ferrochelatase is found within the cytoplasm or embedded in the cytoplasmic membrane [207], as in eukaryotes enzyme is part of the inner mitochondrial membrane. In was revealed that ferrochelatase can chelate divers divalent metal ions, most prominently Zn^{2+} [208].

4.3 Iron-sulfur cluster biosynthesis Fe-S clusters are ancient and complex prosthetic group. They are absolutely crucial in electron transport, catalysis, gene regulation and iron uptake [209]. In the recent studies various steps of their biosynthetic pathways have been revealed, but surprisingly research is still ongoing for in depth understanding of particular processes.

Iron sulfur cluster (ISC) assembly in yeast begins with the liberation of sulfur by the cysteine desulfurase, Nfs1 [210]. Nfs1 is an enzyme that contains pyridoxal phosphate (PLP) as cofactor, through which it forms Schiff base with free cysteine. This is followed by a nucleophilic attack on the cysteine-PLP complexby the active site thiolate, releasing alanine [211]. Most recently a small

protein Isd11 is identified, believed to act as an adaptor between Nfs1 and scaffold proteins thus helping stabilize the complex and promote sulfur release [212-214]. In the next step, sulfur is transferred to the cluster assembly scaffold proteins, ~ 28 kDa protein dimer constructed of Isu1 (and/or) 2, usually through heterotetrameric complexes of the scaffold and the desulfurase [211, 215-217] This protein–controlled mechanism prevents toxic effect of iron while enabling bioavailability for delivery to the Isu.

4.4 Disruption of iron homeostasis and Friedreich's ataxia Friedreich's ataxia (FRDA) is a disease related to the disruption of cellular iron homeostasis [201]. First described by a physician Nicholaus Friedreich, FRDA is an autosomal cardio and neurodegenerative disorder, and it is the most prevalent inherited ataxia. FRDA affects 1 in 50,000 people and is caused by deficiency of the protein frataxin (Ftx). The void of Ftx leads to progressive iron accumulation and a deficiency in heme and Fe-S clusters, resulting in loss of cell viability and eventually patient death. The patients with FRDA (96%) have extensive trinucleotide repeat expansions in the first intron of the gene encoding Ftx, while small some patients have frataxin point mutations [218]. The expanded GAA repeat adopts abnormal DNA structures that impair Ftx transcription; longer the repeat, the more profound is the frataxin deficiency and the earlier the onset and greater the intensity of the disease [219-221].

Early symptoms of the disease like loss of coordination, muscle weakness and scoliosis usually begin between ages 5 to 15 years. About 20% of patients develop diabetes mellitus. Disease is progressive and death occurs during early

thirties and mostly (91%) involves cardiomyopathy. At present, there is no cure for Friedreich's ataxia. The treatment is limited to 3 possible approaches [222]. One strategy involves antioxidants targeted to mitochondria and recent trials include idebenone (proven to delay cardiac dysfunction in mice) and drugs like VitQ and MitoQ. Second approach uses iron chelators to protect from peroxide induced toxicity (deferoxamine). At last, third approach uses DNA sequence specific polyamides and histone deacetylase inhibitors that reverse gene silencing [222].

4.5 Role of frataxin in the mitochondria Frataxin is a nuclear encoded protein localized in the mitochondria. This protein is highly conserved from prokaryotes to eukaryotes and amino acid sequence identity of yeast frataxin (Yfh1) compared to bacterial (CyaY) is 28.1% and to human (Hftx) is 37.8%. Frataxin is abundant in tissues with high metabolic rate such as kidneys, liver, brown fat neurons and heart [195]. Initial studies of frataxin's role in maintaining iron homeostasis were done on yeast homolog (Yfh1)[223]. Deletion of frataxin gene results in accumulation of mitochondrial Fe deposits [223], followed by aconitase and general Fe-S cluster protein deficiencies [196, 224]. The presence of zinc suppresses iron accumulation phenotypes in yeast [225]. Cells depleted of Yfh1 were hypersensitive to H_2O_2, iron and copper levels [223, 224]. However, it was also noted that after reintroducing Yfh1 expression under depleted conditions, recovery and export of original mitochondrial iron deposits was promoted. Therefore, it was proposed that frataxin has a role of directly controlling mitochondrial iron efflux [226].

Frataxin's ability to directly bind iron was first identified in the yeast system [227]. The same property has been confirmed for bacterial and human orthologs [228-233]. Initial hypothesis was that the iron binding in the yeast system induces frataxin to oligomerize under unique solution conditions. Numerous reports after that have proven that oligomerization is not essential when protein act as a chaperone, but it might be necessary when protein participates in helping control ROS production under iron overload conditions [234]. Metal to protein stoichiometries, measured for monomeric bacterial, yeast and N-terminally truncated human frataxin show these proteins will tightly bind 2, 2 and 7 Fe(II) atoms, respectively [230, 231, 233]. Iron dissociation constants, measured for bacterial (CyaY), yeast (Yfh1) and N-terminally truncated human frataxin (HsFtx) give average K_d values of 3.8, 3.0 and 55.0 μM, respectively [230, 233].

However, the exact role of frataxin is still unclear. Initially, it was believed to be an iron regulator, since metal accumulation occurs as a result of its depletion [226]. It was also shown that frataxin is involved in energy conversion and ATP synthesis [235], also protecting aconitase $[4Fe\text{-}4S]^{2+}$ clusters against disassembly and inactivation [236, 237]. Frataxin is involved in the *in vivo* production of heme [238-240] and Fe-S clusters [216, 241, 242].

In order to characterize binding affinity of monomeric yeast frataxin to ferrous ion, isothermal titration calorimetry was used. The research conducted is a collaboration project with Dr.Stemmler's group (WSU, Medical School). In order to better understand how frataxin binds iron, besides ITC measurements, nuclear magnetic resonance (NMR) spectroscopic analysis was used to further identify

amino acids affected by the presence of iron. Finally, X-ray absorption spectroscopic (XAS) studies were used to determine the oxidation and spin-state of iron bound to frataxin, as well as help provide structural detail of the binding event [233].

ITC experiments performed so far were done using yeast frataxin (Yfh1), wild type protein. Since, there is a high sequence homology and structural similarities between protein family members, these results provide additional direct insight into general properties that dictate how frataxin orthologs bind iron. Besides wild type protein, additional mutant proteins were used. Mutations were engineered in order to knock out the assumed binding points for ferrous ion.

This project was worth more than 80 ITC experiment. Unfortunately, for a lot of them good curve fitting was impossible to achieve. Working under anaerobic conditions, with the presence of a reducing agent requires extreme caution; since iron solution readily oxidizes. Experiments were very lengthy due to a long equilibration period between injections.

4.6 Structure of frataxin Human, yeast and bacterial frataxin orthologs have their structures deposited in PDB [243-248]. All frataxin variant share structural similarity, as expected based on the high sequence overlap. Frataxin structure is comprised of two terminal α-helices that make one plane, five antiparallel β-strands that build a second plane of the protein and a sixth β-strand (even seventh in case of HsFtx) that intersects the planes to give an overall planer α-β sandwich structural motif (Fig. 4.1).

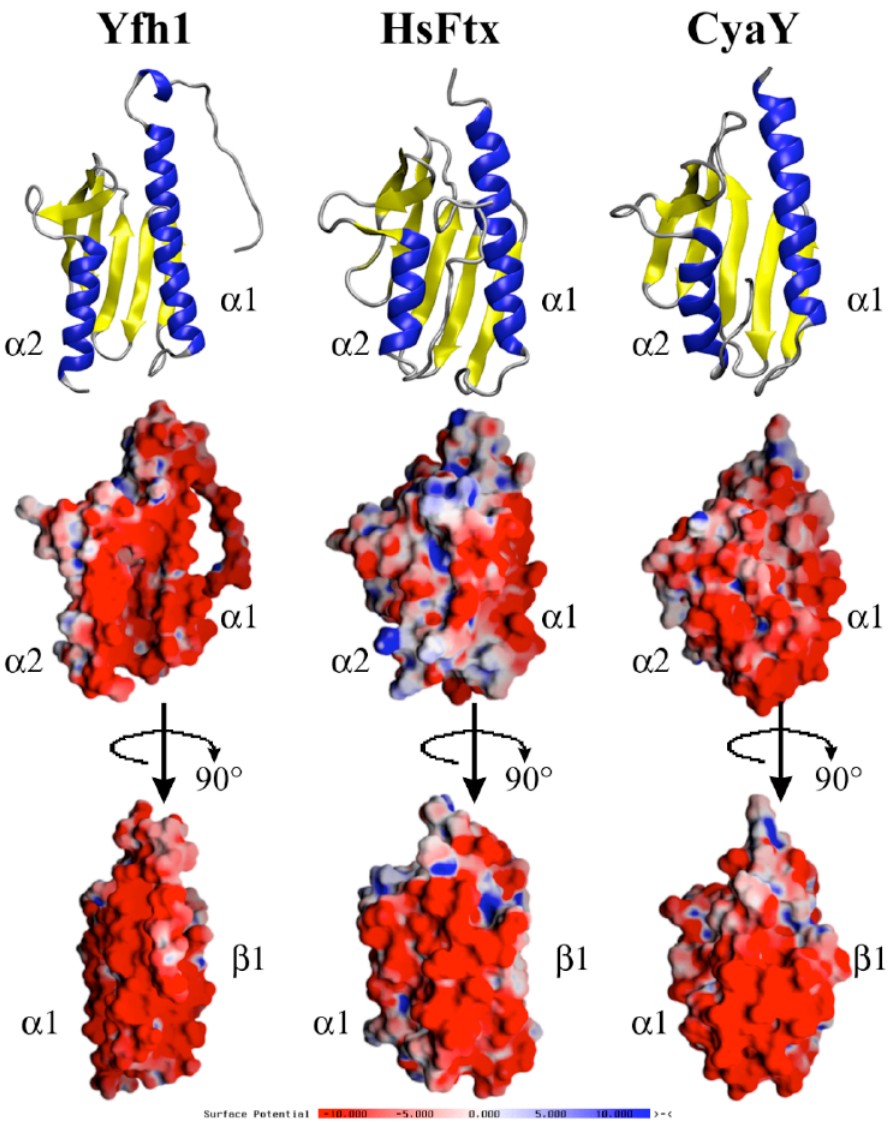

Figure 4.1. Top: ribbon diagram of yeast, human and bacterial frataxin. Middle: electropotential plots for proteins in same orientation. Bottom: electropotential plots for proteins rotated -90 degrees around the y-axis compared to top display. Structure figures made using solution structures of Yfh1 (PDB ID# 2GA5), HsFtx (PDB ID# 1LY7) and CyaY (PDB ID# 1SOY) frataxins.

As a illustration of the structure similarities between orthologs, RMSDs of HsFtx versus CyaY and Yfh1 in regions of secondary structure are 1.34 and 0.65 respctively [201]. Very interesting property of frataxin structure is a large number

of Asp and Glu residues in the N-terminal region. These residues are primarily located in the α1 and β1 regions of the protein [201]. In the electrostatic potential plots (Fig 4.1), these conserved acidic residues form negatively charged surface that covers almost a quarter of frataxin's total accessible surface [244]. Carboxylate Asp and Glu serve as ligands for metal binding, so it is believed to be the case for frataxin as well.

4.7 Mapping iron binding sites of yeast frataxin NMR spectroscopy was used to identify frataxin amino acids perturbed by the presence of ferrous iron under solution conditions that stabilize the monomeric protein. The goal was to locate the region of the monomeric protein affected by the presence of iron with the intent to determine which amino acids may be directly involved in iron binding [233]. Subsequently, these residues will be subjected to mutations in order to further investigate the nature of Fe binding. Ferrous iron titrations performed anaerobically with the ^{15}N labeled frataxin orthologs show distinct alterations in α1 and β1 amino acid NMR resonances. General trends in amide resonance perturbations upon addition of iron can be divided into two categories: amide resonances that significantly line broaden (often beyond detection) and amide chemical shifts that are shifted in the presence of metal. Yfh1 resonances broadened by addition of up to 2 iron atoms include alpha-1 residues Asp28, Asp31, His32 and Asp35, the β1 residue Glu61 and the β2 residue Glu61 (Fig 4.2).

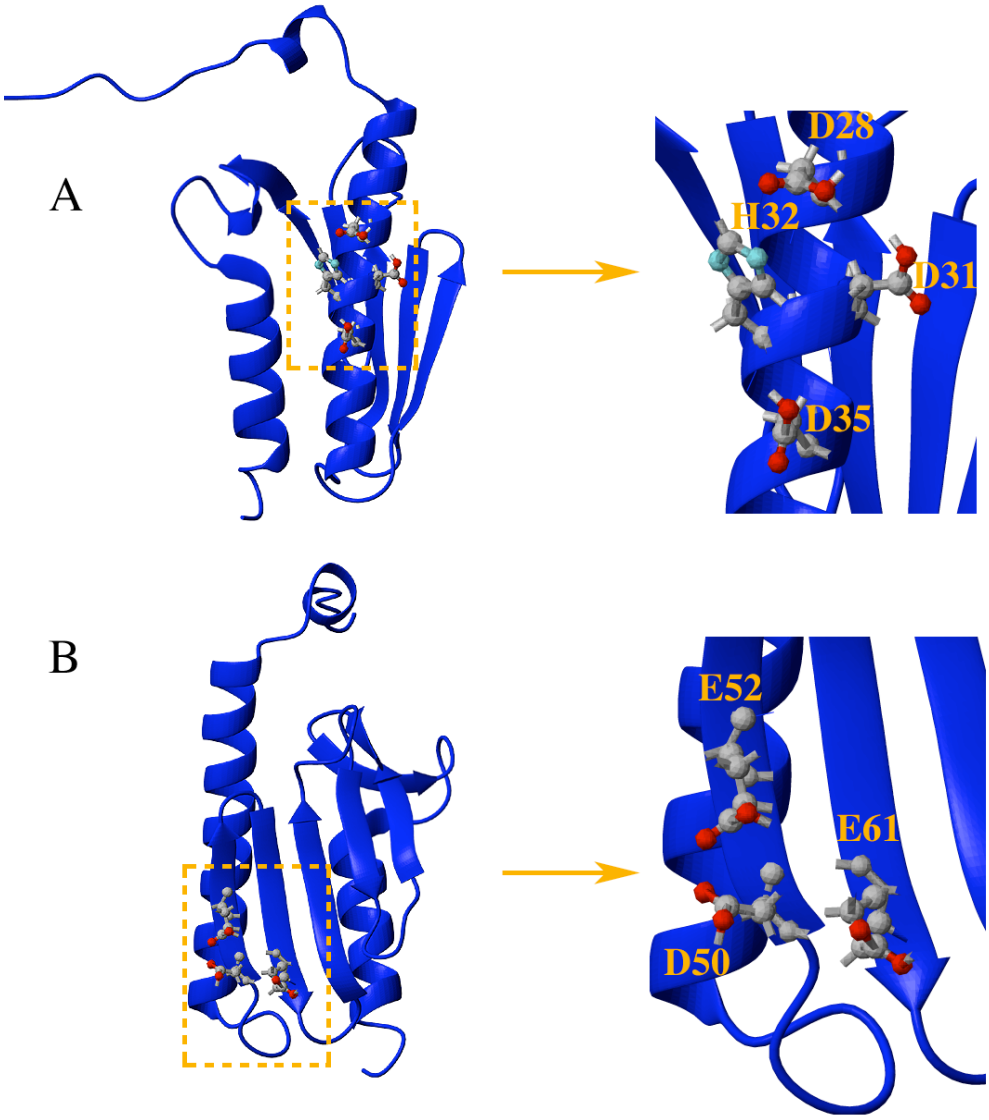

Figure 4.2 Mapping binding sites on the frataxin**.** A) Apo-protein structure with helix-1 amino acid side chain atoms from line broadened amide resonances in the yeast frataxin NMR iron titration. Expanded view to the right of arrows. B) Apo-protein structure with β-sheet amino acid side chain atoms from line broadened amide resonances in the yeast frataxin NMR iron titration and D50. PDB accession number 2GA5.

Amide resonances for additional α1, β1 and β2 residues undergo chemical shift perturbations in the presence of iron [233, 245]. Ferric iron titrations were

performed on CyaY with similar results to those obtained from ferrous iron binding.

Characterizing the metal-ligand coordination geometry and electronic properties of bound iron provide additional insight into the metal binding ability of the monomeric frataxin orthologs. X-ray absorption spectroscopy (XAS) provides a powerful tool for characterizing structural and electronic properties of metals bound to proteins in solution [249].This technique has been used to characterize iron bound to Yfh1. Analysis of the iron 1s→3d transitions in the X-ray absorption near edge structure (XANES) portion of the XAS spectrum are consistent with ferrous iron existing in the high-spin state and coordinated in a highly centro-symmetric metal-ligand coordination geometry when bound to monomeric HsFtx and Yfh1 [201, 233]. Structural analysis of the Fe(II)-ligand coordination geometry for monomeric Yfh1 and HsFtx from the extended x-ray absorption fine structure (EXAFS) portion of the XAS spectrum are consistent with iron bound in a highly symmetric 6 coordinate ligand environment. Ligands coordinating Fe(II) are exclusively oxygen and/or nitrogen based, in agreement with α1 and β1 Asp, Glu and His residues identified in the multiple NMR titrations as residues that possibly interact with iron. Additional ligands to the protein bound metal surely also come from water or hydroxide ions. There is no evidence for any metal•••metal interaction for the ferrous iron bound to monomeric frataxin, even when up to 2 metals are bound to Yfh1.

4.8 Results-ITC measurement of Fe binding to yeast frataxin ITC measurements were performed to determine the metal binding affinity and

maximum metal to protein stoichiometry of monomeric yeast frataxin. Experiments were performed at 30 °C on a VP-ITC titration microcalorimeter (Micro Cal Inc.). Protein and aqueous ferrous ammonium sulfate solutions were prepared anaerobically in 20 mM HEPES (pH = 7.0) and 10 mM $MgSO_4$. Experiments were designed to keep a ca. 20 fold excess of iron in the syringe and data were collected at three independent protein concentrations (100 μM, 70 μM and 50 μM) corresponding to aqueous iron solutions of 2.5 mM, 1.4 mM and 0.8 mM, respectively. Protein samples were loaded anaerobically into the sample chamber and after an initial 2 μL iron solution injection, 29 additional injections of 10 μL each were titrated into the protein sample. Spacing between injections was 10 minutes and the stirring speed of the syringe was held constant at 500 rpm. All experiments were repeated in triplicate on independent protein and iron samples Isothermal titration calorimetry (ITC) measurements were performed to measure the to ensure data reproducibility. The Origin 5.0 software package supplied by Micro Cal Inc. was used for data analysis. Origin 5.0 uses non-linear least squares curve fitting algorithm to determine the stoichiometric ratio, dissociation constant and change in enthalpy of the reaction.

4.8.1 Yeast Frataxin Binds Two Iron Atoms at Micro-Molar Affinity. ITC was used to provide insight into monomeric yeast frataxin's iron binding affinity and the maximal number of ferrous iron atoms that the protein can bind. ITC measurements were performed anaerobically under solution conditions that stabilized the monomeric iron loaded protein. Changes in heat resulting from iron binding are displayed in μcal/sec on the raw ITC data (Figure 4.3A). These

data show a single strong exothermic reaction that follows exponential behavior in the early injection points. Heats at the end of the titration equal those obtained from Fe(II) alone, indicating metal binding is complete following the early titration points. Initially, data sets were treated using a model with one set of binding sites (Origin 5.0). The calculated stoichiometry from the fitted curve provided for an average of 1.4-1.6 iron atoms per monomeric frataxin. However, upon analyzing the diagram, it was apparent that complete saturation is achieved for a 2:1 Fe(II) to protein ratio. Considering also the nature of binding sites in yeast frataxin, an advanced curve fitting of two sets of binding sites was employed. The theoretical curve for this model resulted in a better fit to the original data represented in the metal binding isotherm (Figure 4.3B), and used to determine the protein's metal binding affinity. After curve fitting analysis, binding affinities for both sites was determined to be in the micromolar range (K_{d1} = 3.0 ± 0.4 μM and K_{d2} = 2.0 ± 0.2 μM). These results indicate the two metal binding sites are independent, although they have nearly identical binding affinities.

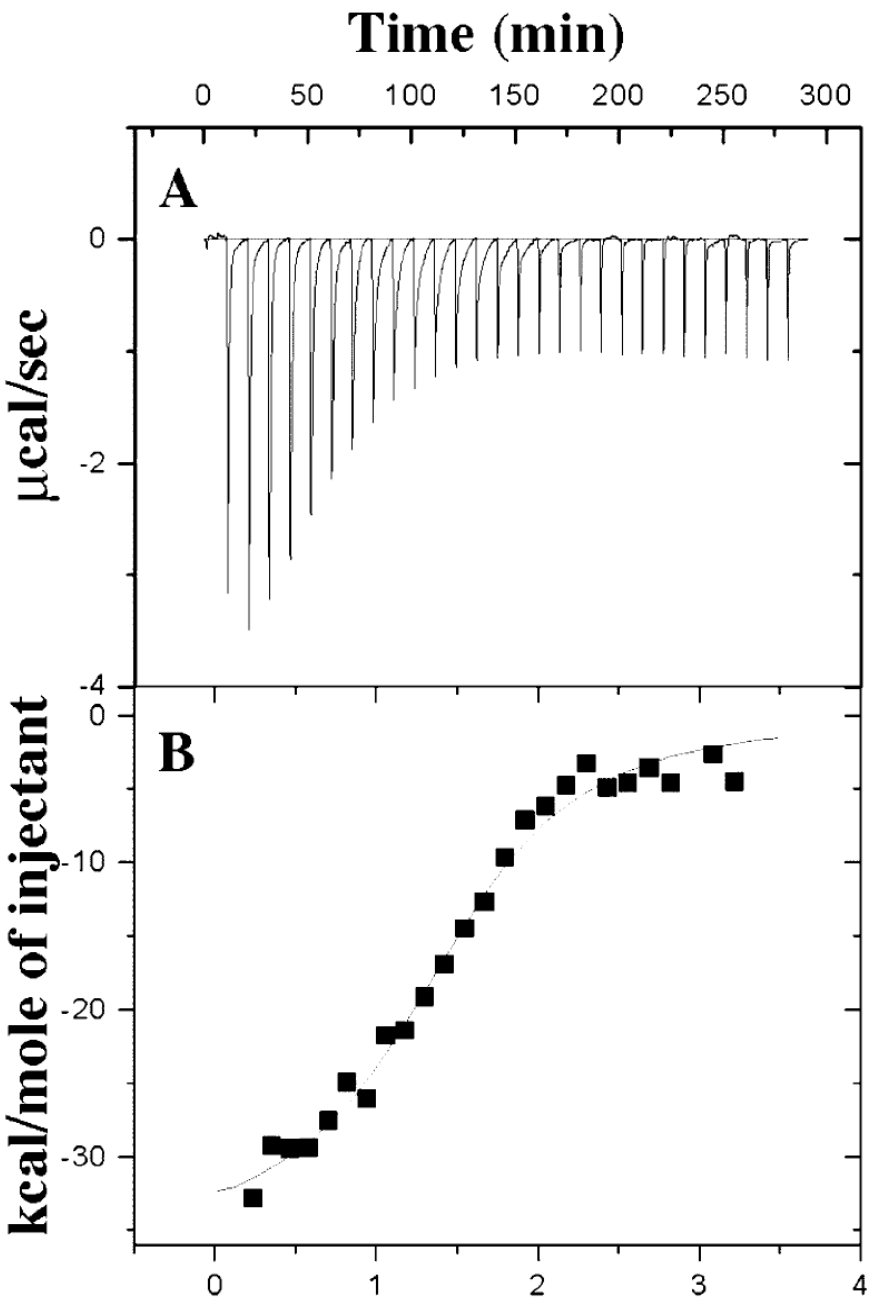

Figure 4.3. Raw isothermal titration calorimetry (A) and binding isotherm data (B) for ferrous iron to yeast frataxin. Grey line in panel (A) shows baseline and in panel (B) shows the simulated fit to the binding isotherm data. Data were collected anaerobically at 30 °C in 20 mM HEPES (pH 7.0) and 10 mM $MgSO_4$. These data were collected by injecting small aliquots of a 0.8 mM ferrous iron solution into a 50 µM Yfh1 sample. Spacing between injections was 10 minutes. Syringe stirring speed was kept at 500 rpm [233].

4.8.2 Binding Affinity of Mutated Yeast Frataxin Homologs In order to investigate the role of individual residues involved in iron binding, different mutations were made to the original sequence. Work on the double mutants was done in collaboration with Andrew Dancis group at University of Pennsylvania. N-terminal double mutant (H32A/D35A), showed binding of two iron atoms, per molecule of protein. Binding isotherm is shown in figure 4.4.

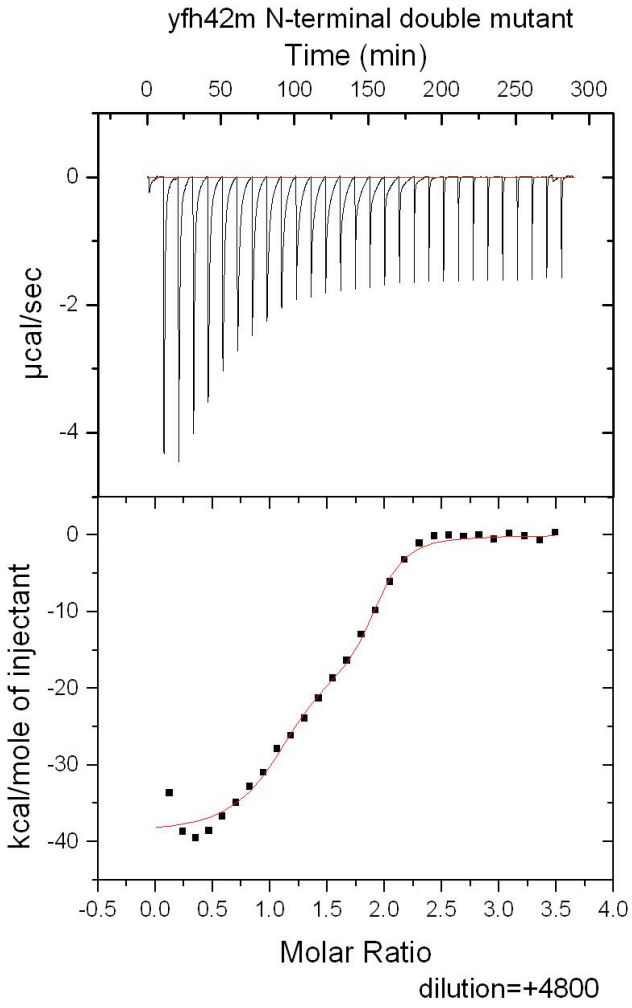

Figure 4.4 Binding of frataxin double –mutant to Fe(II).

These results are easily interpreted if the original NMR structure is examined [233]. There are a lot of residues in helix-1 being perturbed by the presence of iron (shown in red on Figure 4.4), which implies that there are more residues that can facilitate metal binding instead of H32 and D35.

4.9 Discussion

Frataxin has been hypothesized to play a variety of roles in helping maintain cellular iron homeostasis, most of which involve direct binding to iron. Frataxin facilitates *in vivo* heme and Fe-S cluster production and the monomeric protein binds to enzyme partners that complete the assembly of these iron-containing cofactors, suggesting monomeric protein may play a direct role in both processes by acting as the iron chaperone [238, 239]. Since a single iron atom is transferred during heme and Fe-S cluster assembly, it seems further unlikely that frataxin oligomers would participate as the source of metal during delivery. Given this rationale, having a structural and biochemical understanding of iron bound to monomeric frataxin would help propel the general understanding of how frataxin functions towards controlling iron bioavailability in the cell.

Frataxin's metal binding affinity is similar to those seen for the copper chaperones suggesting a common theme in cellular metal delivery if frataxin is in fact acting as an iron chaperone. The micromolar dissociation constants measured for yeast frataxin, with respect to ferrous iron (K_d's of 2.0 and 3.0 μM), resembles those obtained for the human and bacterial orthologs (K_d's of 55.0 and 3.8 μM, respectively). Ferrous iron binding affinities measured for ferrochelatase are also within the micromolar range indicating frataxin's metal binding affinity is

closely matched by its protein partner that receives metal. It is therefore reasonable to suggest frataxin can act as an iron chaperone during Fe-cofactor production.

4.9 Conclusions

All three structurally characterized frataxin orthologs bind multiple iron atoms with a micromolar binding affinity when the protein exists in the monomeric state, which is consistent with a functional role of monomeric protein being involved in iron binding and delivery. The weak nature of the protein's metal binding affinity (K_d's of 2.0 and 3.0 μM) and the ligand environment utilized by the protein when binding metal would be consistent with frataxin acting as an iron chaperone; the protein must be able to form a favorable interface with its protein partners and also easily release the metal. Additional structural characterization will be critical to help elucidate how frataxin binds iron and docks with its protein partners to promote metal delivery.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Our understanding of multivalent effects in biological molecular recognition has been extensively broadened by emerging new research involving biochemical studies that seek to address the topic. Whether this pertains to the binding of two or more peptide ligands to a single or dual domain protein, or the binding of multiple metal ions to a protein, there are common themes that unite the two. Conceptually, there are binding theories and data treatments that can be applicable to both. Operationally, what is also unfortunately in common is the degree of experimental difficulty that often accompanies studying such systems. But such investigations are necessary if we not only want to understand the fundamental molecular recognition properties of these types of systems, but also want to apply that knowledge to the discovery and development of cellular probes and therapeutic agents that take multivalency into account. This could yield ligands that have higher affinity, and improved selectivity, for a given biomolecular receptor. The projects outlined in this dissertation, encompassing the peptide-protein interactions of PDZ domains, and the iron-binding interactions of frataxin, are steps towards understanding multivalency for both fundamental knowledge and biomedical application.

5.1 Protein-Peptide Interactions: PDZ Domains

Regarding the first project area, PDZ domains have proven to be an ideal subject for exploring multivalency, given the incidence of the phenomenon as reported in the literature, especially in proteins involved in critical cellular

processes. This domain family often appears in proteins situated at neuronal synapses, and involved in organizing transduction complexes, clustering membrane receptors, and maintaining cell-polarities [96, 250-252]. These properties of PDZ domains have been extensively discussed in Chapter 1 of this dissertation. In these proteins, PDZ domains are usually present in multiple, but not identical, copies. The protein that we have centered our research upon, the tandem (PDZ1-2) domain of the N-terminal region of postsynaptic density protein-95 (PSD-95), is connected by a five-residue peptide linker between the two domains [186]. Our choice for such a spatially proximal domain may allow our studies to contribute to understanding other domain constructs so configured. For example, a four amino acid residue linker is observed in tandem repeats in mammalian syntenins [253] and also in X11s (also known as Mints) [254]. Besides these dual repeats, clustering of the first three (PDZ123) and second three (PDZ45) domains has been observed in GRIP (glutamate receptor-interacting protein) and ABP (AMPA receptor-binding protein) [255-257]. Evidence is emerging which indicates that tandem PDZ repeats show distinct target-binding properties from the individual PDZ domains or from a simple sum of the two PDZ domains in the repeats [137, 185, 186, 258-260], although the molecular basis of such supramodular properties remains largely unclear.

The research efforts in our laboratory yielded monovalent [136, 167] and cyclic [128] peptides that provided insight into nature of interaction of a single PDZ domain (PDZ3 of PSD-95). In brief, we found that relative short peptides, with exposed C-termini, are capable of binding to PDZ domains with dissociation

constants in the low micromolar range, provided they possess certain sequence motifs. Further on, this knowledge led us to validation of a PDZ domain-directed molecular probe that functions within human cells [130]. Based on this efforts homobivalent ligands were designed and probed against the PDZ3 [261] and proven was the formation of the ternary complex.

The above-mentioned results provided us with the chemical tools to tackle the problem of the dual PDZ1-2 domain, although immediately it became clear that the nature of the questions we would ask were based on exactly how we constructure our bivalent participants; the possibilities for the two-binding site architecture is outlined in the schematic Fig 5.1. Our initial focus was to pursue studies that would correlate with cases (3) and (4) in that diagram.

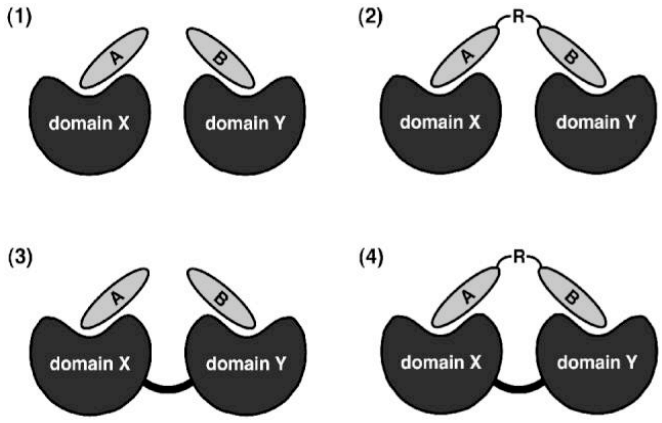

Fig 5.1 Protein domain-ligand combinations for bivalent interactions. Reprinted by permission.

Based on the previous experience in the group, in the reported dissertation studies we continue to rely on the extensively used technique of ITC, which had been applied towards linear [131, 167] and macrocyclic [128, 129]

ligands for the third PDZ domain (PDZ3) of the PSD-95. These studies benefited greatly from the advantages of ITC, a pure solution method that provides for each of the thermodynamic binding parameters, as well as the stoichiometry of the interaction. We have employed the same approach to our protein domains PDZ1, PDZ2 and PDZ1-2 tandem.

Challenging the individual PDZ1 and PDZ2 and dual domains with short peptides derived form the C-terminal region of the domains' natural binding partners, we found that only *one* of the domains remains functional when expressed separately. Binding of PDZ1 and tandem PDZ1-2 we were able to characterize, whereas PDZ2, despite extensive efforts, could not be stabilized under solution conditions to yield a valid binding curve. The PDZ1-2 dual domain of PSD-95 therefore functions in the same manner as hypothesized for PDZ1-2 of GRIP1 [262] or PDZ4-5 of the same protein [263]. We suggest that PDZ2 induces the folding of PDZ1, stabilizing its structure. The two domains can pack closely to each other and form a compact supramodule, which is the case in GRIP1 [263]. Alternatively, they might be partially restrained to promote synergistic target binding, as in the case of syntenin [176]. Our ITC data for the binding of PDZ1-2 to natural binding partners supports the hypothesis that PDZ2 binding groove is distorted and probably cannot bind C-terminal peptides, which is in good agreement with the published work for syntenin [169].

At the same time, in an ongoing effort to develop new cellular probes for *in vivo* investigation of multiprotein assembly, we studied modified ligand CN2180 (designed and synthesized by Dr. Chamila Rupansinghe) which targets the

tandem PDZ1-2 of PSD-95. This cyclic peptide showed striking biological activity, which was thoroughly described in section 2.9.5.1 of Chapter 2. Our *in vitro* testing by ITC of the same peptide after integrating the binding isoterm, yielded a stoichiometry of n=2, indicating that both binding sites are fully occupied. Affinity for the interaction is K_{d1} = 6 μM and K_{d2} = 9 μM, for each domain. These values are comparable to the previously obtained result for PDZ3 (K_{d1} = 8 μM), and thus make this compound a reasonably effective inhibitor of the PSD-95 interaction with the NMDA receptor (NMDAR).

Although two different values are obtained, the model fitting does *not* differentiate between the binding sites. Although ITC is a very powerful technique, it is often hard to deconvolute the heat signal. Therefore, based on this data we cannot make the final determination on the precise mode of PDZ1-2 binding; this will have to await for more structural information to be obtained.

In terms of progress on that structural front, we have successfully crystallized a new protein construct, PDZ1-2 of PSD-95. Identifying the conditions required to grow crystals is a major step toward the goal of solving the structure of this tandem domain by X-ray methods. Significant efforts were made to solve the structure using the molecular replacement method, but due to specific space group that protein crystallizes in, thus far solved models of similar proteins could not be implemented. We are confident that producing a seleno-methionine derivative of PDZ1-2 would gain a diffractable crystal and would enable us to eventually solve the structure of the protein. Further on, these protein crystals should be co-crystallized with ligands already tested by ITC.

These results should be valuble in revealing the mode of protein-protein interactions mediated by PSD-95.

5.2 Protein-Metal Interactions: Frataxin

Exploring the phenomenon of mutivalency and its possible applications, we have looked into the nature of protein-metal interaction on the example of frataxin binding to ferrous ion. This project provided valuable knowledge on possibilities of ITC when analyzing more complexed interactions beyond standard 1:1 host-ligand titration.

Iron deficiency still represents one of the medical concerns of the modern world. It has been speculated that nearly 9 milion people are iron deficient in the United States. Iron deficiency is associated with pregnancy problems and impaired cognitive development in the newborns. On the other side, excess iron (ca. ten times the nutritional need) causes severe mitochondrial dysfunction that is associated with several prominent neurodegenerative, hematological and other related disorder [222, 264]. Iron dysregulation is proposed to be important in the pathophysiology of human brain diseases during Parkinson's, FRDA, and Alzheimer's [222]. Another iron overload disorder, hemochromatosis leads to severe health consequences characterized by diabetes, cirrhosis of the liver and bronze pigmentation of the skin [222]. Therefore, a deeper understanding of iron metabolism and characterization of proteins involved in iron homeostasis is essential to tackle these diseases.

It has been proposed that the cellular iron metabolism can be divided into three sequential steps. First is the synthesis of proteins that are modulated by

iron levels in the cell. Second is the regulation of the activities of the proteins that transport iron in and out of the cell, store cytoplasmic iron and direct iron to various cellular compartments. Third is the compartmental trafficking of soluble ionic iron for activation of essential cofactors. The protein that is the center of our research efforts is implicated in the third biochemical activity.

Frataxin is a nuclear encoded protein targeted to the mitochondrial matrix. This protein is essential for the effective regulation of cellular iron homeostasis. Inability to produce frataxin leads to mitochondrial iron accumulation and loss of cell viability. This results in a severe neurodegenerative disease, Friedreich's ataxia, that affects 1 in 50, 000 humans. Frataxin is essential for the cellular iron control and is believed to serve as an iron chaperone that delivers mitochondrial Fe(II) to the enzymes ferrochelatase and the ISU apparatus for completion of heme and Fe-S cluster biosynthesis.

ITC measurements were performed to measure the metal binding affinity and maximum metal to protein stoichiometry of monomeric yeast frataxin. The micromolar dissociation constants measured for yeast frataxin, with respect to ferrous iron (K_d's of 2.0 and 3.0 μM) were obtained from ITC titrations. The weak interaction is consistent with the hypothesis of frataxin acting as an iron chaperone. Frataxin must be able to form a favorable interface with its protein partners and also easily release the metal. Additional structural characterization will be critical to help elucidate how frataxin binds iron and docks with its protein partners to promote metal delivery.

Frataxin has been suggested to play a variety of roles in helping to maintain cellular iron homeostasis, most of which involve direct binding to iron. Frataxin facilitates *in vivo* heme and Fe-S cluster production and the monomeric protein binds to enzyme partners that complete the assembly of these iron-containing cofactors, suggesting monomeric protein may play a direct role in both processes by acting as the iron chaperone [238]. Fluctuations in mitochondrial iron or Fe-cofactor levels, and the differences in binding affinities between partners, have been proposed to serve as the trigger for which cofactor assembly system is delivered iron [231].

Metal binding site(s) on monomeric frataxin must accommodate the binding of multiple iron atoms. Additional structural characterization will be critical to help elucidate how frataxin binds iron and docks with its protein partners to promote metal delivery.

APPENDIX A

SINGLE AND THREE LETTER CODES FOR AMINO ACIDS

	Codes	
Amino acids	Single letter	Three letter
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cysteine	C	Cys
Glutamic acid	E	Glu
Glutamine	Q	Gln
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

APPENDIX B

GENERAL CHARACTERISTICS OF PDZ1-2 DUAL DOMAIN OF PSD-95

Protein PDZ1-2M PROTEIN ONLY
Local object
Created: 01/17/02 05:55PM
Last Modified: 01/17/02 05:55PM
length: 190 aa

Standard Fields

Analysis

Analysis	Entire Protein
Length	190 aa
Molecular Weight	20112.61 m.w.
1 microgram =	49.720 pMoles
Molar Extinction coefficient	6400
1 A[280] corr. to	3.14 mg/ml
A[280] of 1 mg/ml	0.32 AU
Isoelectric Point	5.75
Charge at pH 7	-3.77

Amino Acid(s)	Number count	% by weight	% by frequency
Charged (RKHYCDE)	56	36.08	29.47
Acidic (DE)	25	14.92	13.16
Basic (KR)	21	14.01	11.05
Polar (NCQSTY)	34	18.72	17.89
Hydrophobic (AILFWV)	71	35.89	37.37
A Ala	18	6.82	9.47
C Cys	0	0.00	0.00
D Asp	12	6.79	6.32
E Glu	13	8.13	6.84
F Phe	4	2.81	2.11
G Gly	23	7.34	12.11
H His	5	3.30	2.63
I Ile	18	10.04	9.47
K Lys	13	8.08	6.84
L Leu	13	7.25	6.84
M Met	3	1.90	1.58
N Asn	9	5.06	4.74
P Pro	8	3.92	4.21
Q Gln	3	1.86	1.58
R Arg	8	5.93	4.21
S Ser	11	4.92	5.79
T Thr	6	3.04	3.16
V Val	18	8.97	9.47
W Trp	0	0.00	0.00
Y Tyr	5	3.85	2.63
B Asx	21	11.85	11.05
Z Glx	16	10.00	8.42
X Xxx	0	0.00	0.00

PDZ1-2 sequence

GSEYEEITLERGNSGLGFSIAGGTDNPHIGDDPSIFITKIIPGGAAAQDGRLRV
NDSILFVNEVDVREVTHSAAVEALKEAGSIVRLYVMRRKPPAEKVMEIKLIKG
PKGLGFSIAGGVGNQHIPGDNSIYVTKIIEGGAAHKDGRLQIGDKILAVNSVGL
EDVMHEDAVAALKNTYDVVYLKVAKPSNA

REFERENCES

1. Cooper, A. and C.M. Johnson, *Isothermal titration microcalorimetry.* Methods Mol Biol, 1994. **22**: p. 137-50.
2. Doyle, M.L., *Characterization of binding interactions by isothermal titration calorimetry.* Curr Opin Biotechnol, 1997. **8**(1): p. 31-5.
3. Jelesarov, I. and H.R. Bosshard, *Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition.* J Mol Recognit, 1999. **12**(1): p. 3-18.
4. Ladbury, J.E., *Counting the calories to stay in the groove.* Structure, 1995. **3**(7): p. 635-9.
5. Ladbury, J.E. and B.Z. Chowdhry, *Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions.* Chem Biol, 1996. **3**(10): p. 791-801.
6. Pierce, M.M., C.S. Raman, and B.T. Nall, *Isothermal titration calorimetry of protein-protein interactions.* Methods, 1999. **19**(2): p. 213-21.
7. Holdgate, G.A., *Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics.* Biotechniques, 2001. **31**(1): p. 164-6, 168, 170 passim.
8. Leavitt, S. and E. Freire, *Direct measurement of protein binding energetics by isothermal titration calorimetry.* Curr Opin Struct Biol, 2001. **11**(5): p. 560-6.
9. Perozzo, R., G. Folkers, and L. Scapozza, *Thermodynamics of protein-ligand interactions: history, presence, and future aspects.* J Recept Signal Transduct Res, 2004. **24**(1-2): p. 1-52.
10. Cliff, M.J. and J.E. Ladbury, *A survey of the year 2002 literature on applications of isothermal titration calorimetry.* J Mol Recognit, 2003. **16**(6): p. 383-91.

11. Cliff, M.J., A. Gutierrez, and J.E. Ladbury, *A survey of the year 2003 literature on applications of isothermal titration calorimetry.* J Mol Recognit, 2004. **17**(6): p. 513-23.
12. Ababou, A. and J.E. Ladbury, *Survey of the year 2005: literature on applications of isothermal titration calorimetry.* J Mol Recognit, 2007. **20**(1): p. 4-14.
13. Ababou, A. and J.E. Ladbury, *Survey of the year 2004: literature on applications of isothermal titration calorimetry.* J Mol Recognit, 2006. **19**(1): p. 79-89.
14. MicroCal. http://www.microcal.com/.
15. O'Brien, R., Chowshury, B.Z. & Ladbury, J.E., *Isothermal titration calorimetry of biomolecules*. Oxford University Press, Oxford, UK 2001.
16. Liu, Y. and J.M. Sturtevant, *Significant discrepancies between van't Hoff and calorimetric enthalpies. II.* Protein Sci, 1995. **4**(12): p. 2559-61.
17. Fisher, H.F. and N. Singh, *Calorimetric methods for interpreting protein-ligand interactions.* Methods Enzymol, 1995. **259**: p. 194-221.
18. Zhang, Y.L. and Z.Y. Zhang, *Low-affinity binding determined by titration calorimetry using a high-affinity coupling ligand: a thermodynamic study of ligand binding to protein tyrosine phosphatase 1B.* Anal Biochem, 1998. **261**(2): p. 139-48.
19. Sigurskjold, B.W., *Exact analysis of competition ligand binding by displacement isothermal titration calorimetry.* Anal Biochem, 2000. **277**(2): p. 260-6.
20. O'Brien, R., Chowshury, B.Z. & Ladbury, J.E., *Isothermal titration calorimetry of biomolecules.* Oxford University Press, Oxford, UK 2001
21. Holdgate, G.A. and W.H. Ward, *Measurements of binding thermodynamics in drug discovery.* Drug Discov Today, 2005. **10**(22): p. 1543-50.

22. Horn, J.R., et al., *Van't Hoff and calorimetric enthalpies from isothermal titration calorimetry: are there significant discrepancies?* Biochemistry, 2001. **40**(6): p. 1774-8.
23. Horn, J.R., J.F. Brandts, and K.P. Murphy, *van't Hoff and calorimetric enthalpies II: effects of linked equilibria.* Biochemistry, 2002. **41**(23): p. 7501-7.
24. Naghibi, H., A. Tamura, and J.M. Sturtevant, *Significant discrepancies between van't Hoff and calorimetric enthalpies.* Proc Natl Acad Sci U S A, 1995. **92**(12): p. 5597-9.
25. Sciences, C. *http://www.calscorp.com/*.
26. MicroCal, *VP-ITC MicroCalorimeter User's Manual.*
27. Ward, W.H. and G.A. Holdgate, *Isothermal titration calorimetry in drug discovery.* Prog Med Chem, 2001. **38**: p. 309-76.
28. Bains, G. and E. Freire, *Calorimetric determination of cooperative interactions in high affinity binding processes.* Anal Biochem, 1991. **192**(1): p. 203-6.
29. MicroCal. *http://www.microcal.com/index.php?id=350*.
30. Origin, *ITC Data Analysis in Origin Tutorial Guide version 5.0.* 2002.
31. Origin, *ITC Data Analysis in Origin Tutorial Guide version 7.0* 2004.
32. Khalifah, R.G., et al., *Thermodynamics of binding of the CO2-competitive inhibitor imidazole and related compounds to human carbonic anhydrase I: an isothermal titration calorimetry approach to studying weak binding by displacement with strong inhibitors.* Biochemistry, 1993. **32**(12): p. 3058-66.
33. Hu, D.D. and M.R. Eftink, *Thermodynamic studies of the interaction of trp aporepressor with tryptophan analogs.* Biophys Chem, 1994. **49**(3): p. 233-9.
34. Sigurskjold, B.W., C.R. Berland, and B. Svensson, *Thermodynamics of inhibitor binding to the catalytic site of glucoamylase from Aspergillus niger determined by displacement titration calorimetry.* Biochemistry, 1994. **33**(33): p. 10191-9.

35. Velazquez-Campoy, A., Y. Kiso, and E. Freire, *The binding energetics of first- and second-generation HIV-1 protease inhibitors: implications for drug design.* Arch Biochem Biophys, 2001. **390**(2): p. 169-75.
36. Williams, D.H., et al., *Toward an estimation of binding constants in aqueous solution: studies of associations of vancomycin group antibiotics.* Proc Natl Acad Sci U S A, 1993. **90**(4): p. 1172-8.
37. Eftink, M.R., A.C. Anusiem, and R.L. Biltonen, *Enthalpy-entropy compensation and heat capacity changes for protein-ligand interactions: general thermodynamic models and data for the binding of nucleotides to ribonuclease A.* Biochemistry, 1983. **22**(16): p. 3884-96.
38. Lumry, R. and S. Rajender, *Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water.* Biopolymers, 1970. **9**(10): p. 1125-227.
39. Pearce, K.H., Jr., et al., *Structural and mutational analysis of affinity-inert contact residues at the growth hormone-receptor interface.* Biochemistry, 1996. **35**(32): p. 10300-7.
40. Connelly, P.R., et al., *Enthalpy of hydrogen bond formation in a protein-ligand binding reaction.* Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1964-8.
41. Dunitz, J.D., *Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions.* Chem Biol, 1995. **2**(11): p. 709-12.
42. Ringe, D. and G.A. Petsko, *Protein Structure and Function*. 2003: New Science Press, Ltd., London
43. Minor, D.L., Jr., *The neurobiologist's guide to structural biology: a primer on why macromolecular structure matters and how to evaluate structural data.* Neuron, 2007. **54**(4): p. 511-33.
44. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
45. Jhoti, H., *High-throughput structural proteomics using x-rays.* Trends Biotechnol, 2001. **19**(10 Suppl): p. S67-71.
46. Schmid, M.B., *Structural proteomics: the potential of high-throughput structure determination.* Trends Microbiol, 2002. **10**(10 Suppl): p. S27-31.

47. Yee, A., et al., *Structural proteomics: toward high-throughput structural biology as a tool in functional genomics.* Acc Chem Res, 2003. **36**(3): p. 183-9.
48. Stryer, L., *Implications of X-ray crystallographic studies of protein structure.* Annu Rev Biochem, 1968. **37**: p. 25-50.
49. Eisenberg, D. and C.P. Hill, *Protein crystallography: more surprises ahead.* Trends Biochem Sci, 1989. **14**(7): p. 260-4.
50. Rose, D.R., *Crystallographic determination of protein structure.* Biotechnology, 1990. **14**: p. 111-29.
51. Palmer, R.A. and H. Niwa, *X-ray crystallographic studies of protein-ligand interactions.* Biochem Soc Trans, 2003. **31**(Pt 5): p. 973-9.
52. Whittle, P.J. and T.L. Blundell, *Protein structure--based drug design.* Annu Rev Biophys Biomol Struct, 1994. **23**: p. 349-75.
53. Davis, A.M., S.J. Teague, and G.J. Kleywegt, *Application and limitations of X-ray crystallographic data in structure-based ligand and drug design.* Angew Chem Int Ed Engl, 2003. **42**(24): p. 2718-36.
54. Jackson, R.C., *Contributions of protein structure-based drug design to cancer chemotherapy.* Semin Oncol, 1997. **24**(2): p. 164-72.
55. Rhodes, G., *Crystallography made crystal clear, second addition*. 2000: Elsevier.
56. Perutz, M.F., *New x-ray evidence on the configuration of polypeptide chains.* Nature, 1951. **167**(4261): p. 1053-4.
57. Kendrew, J.C., et al., *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis.* Nature, 1958. **181**(4610): p. 662-6.
58. Ringe, D. and G.A. Petsko, *New Perspectives in Drug Design*. 1995: Academic Press.
59. Tomasselli, A.G., R.L. Heinrikson, and K.D. Watenpaugh, *Recombinant DNA technology and crystallography. A new alliance in unraveling protein structure-function relationships.* Bioprocess Technol, 1991. **12**: p. 285-315.

60. Georgiou, G. and L. Segatori, *Preparative expression of secreted proteins in bacteria: status report and future prospects.* Curr Opin Biotechnol, 2005. **16**(5): p. 538-45.
61. Terpe, K., *Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems.* Appl Microbiol Biotechnol, 2006. **72**(2): p. 211-22.
62. Goulding, C.W. and L.J. Perry, *Protein production in Escherichia coli for structural studies by X-ray crystallography.* J Struct Biol, 2003. **142**(1): p. 133-43.
63. McPherson, A., *Crystallization of Biological Macromolecules*. 1999, Cold Spring Harbor, NY Cold Spring Harbor Press.
64. Geerlof, A., et al., *The impact of protein characterization in structural proteomics.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 10): p. 1125-36.
65. Ferre-D'Amare, A.R. and S.K. Burley, *Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies.* Structure, 1994. **2**(5): p. 357-9.
66. Wen, J., T. Arakawa, and J.S. Philo, *Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions.* Anal Biochem, 1996. **240**(2): p. 155-66.
67. Wilson, W.W., *Light scattering as a diagnostic for protein crystal growth--a practical approach.* J Struct Biol, 2003. **142**(1): p. 56-65.
68. McPherson, A., *Introduction to protein crystallization.* Methods, 2004. **34**(3): p. 254-65.
69. McPherson, A., *Macromolecular crystals.* Sci Am, 1989. **260**(3): p. 62-9.
70. Raman, P., V. Cherezov, and M. Caffrey, *The Membrane Protein Data Bank.* Cell Mol Life Sci, 2006. **63**(1): p. 36-51.
71. Radaev, S., S. Li, and P.D. Sun, *A survey of protein-protein complex crystallizations.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 6): p. 605-12.

72. Garman, E., *`Cool' crystals: macromolecular cryocrystallography and radiation damage.* Current Opinion in Structural Biology, 2003. **13**(5): p. 545-551.
73. Garman, E.F. and R.L. Owen, *Cryocooling and radiation damage in macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 32-47.
74. Rodgers, D.W., *Cryocrystallography.* Structure, 1994. **2**(12): p. 1135-40.
75. Dauter, Z., M. Dauter, and E. Dodson, *Jolly SAD.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 3): p. 494-506.
76. Dodson, E., *Is it jolly SAD?* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 11): p. 1958-65.
77. Ealick, S.E., *Advances in multiple wavelength anomalous diffraction crystallography.* Current Opinion in Chemical Biology, 2000. **4**(5): p. 495-499.
78. Blow, D., *Outline of Crystallography for Biologists*. 2002, New York: Oxford University Press.
79. Jaenicke, R., *Protein folding: local structures, domains, subunits, and assemblies.* Biochemistry, 1991. **30**(13): p. 3147-61.
80. Pawson, T. and J.D. Scott, *Signaling through scaffold, anchoring, and adaptor proteins.* Science, 1997. **278**(5346): p. 2075-80.
81. Mayer, B.J. and H. Hanafusa, *Mutagenic analysis of the v-crk oncogene: requirement for SH2 and SH3 domains and correlation between increased cellular phosphotyrosine and transformation.* J Virol, 1990. **64**(8): p. 3581-9.
82. Drubin, D.G., et al., *Homology of a yeast actin-binding protein to signal transduction proteins and myosin-I.* Nature, 1990. **343**(6255): p. 288-90.
83. Sudol, M., et al., *Characterization of a novel protein-binding module--the WW domain.* FEBS Lett, 1995. **369**(1): p. 67-71.
84. Zhou, M.M. and S.W. Fesik, *Structure and function of the phosphotyrosine binding (PTB) domain.* Prog Biophys Mol Biol, 1995. **64**(2-3): p. 221-35.

85. Cho, K.O., C.A. Hunt, and M.B. Kennedy, *The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumor suppressor protein.* Neuron, 1992. **9**(5): p. 929-42.
86. Woods, D.F. and P.J. Bryant, *ZO-1, DlgA and PSD-95/SAP90: homologous proteins in tight, septate and synaptic cell junctions.* Mech Dev, 1993. **44**(2-3): p. 85-9.
87. Kim, E., et al., *Clustering of Shaker-type K+ channels by interaction with a family of membrane-associated guanylate kinases.* Nature, 1995. **378**(6552): p. 85-8.
88. Doyle, D.A., et al., *Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ.* Cell, 1996. **85**(7): p. 1067-76.
89. Gee, S.H., et al., *Cyclic peptides as non-carboxyl-terminal ligands of syntrophin PDZ domains.* J Biol Chem, 1998. **273**(34): p. 21980-7.
90. Ponting, C.P. and R.R. Russell, *The natural history of protein domains.* Annu Rev Biophys Biomol Struct, 2002. **31**: p. 45-71.
91. Kornau, H.C., et al., *Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95.* Science, 1995. **269**(5231): p. 1737-40.
92. Niethammer, M., E. Kim, and M. Sheng, *Interaction between the C terminus of NMDA receptor subunits and multiple members of the PSD-95 family of membrane-associated guanylate kinases.* J Neurosci, 1996. **16**(7): p. 2157-63.
93. Shieh, B.H. and M.Y. Zhu, *Regulation of the TRP Ca2+ channel by INAD in Drosophila photoreceptors.* Neuron, 1996. **16**(5): p. 991-8.
94. Nourry, C., S.G. Grant, and J.P. Borg, *PDZ domain proteins: plug and play!* Sci STKE, 2003. **2003**(179): p. RE7.
95. Saras, J. and C.H. Heldin, *PDZ domains bind carboxy-terminal sequences of target proteins.* Trends Biochem Sci, 1996. **21**(12): p. 455-8.
96. Sheng, M. and C. Sala, *PDZ domains and the organization of supramolecular complexes.* Annu Rev Neurosci, 2001. **24**: p. 1-29.

97. Fanning, A.S. and J.M. Anderson, *Protein-protein interactions: PDZ domain networks.* Curr Biol, 1996. **6**(11): p. 1385-8.
98. Ranganathan, R. and E.M. Ross, *PDZ domain proteins: scaffolds for signaling complexes.* Curr Biol, 1997. **7**(12): p. R770-3.
99. Tsunoda, S., et al., *A multivalent PDZ-domain protein assembles signalling complexes in a G-protein-coupled cascade.* Nature, 1997. **388**(6639): p. 243-9.
100. Koo, T.H., et al., *Syntenin is overexpressed and promotes cell migration in metastatic human breast and gastric cancer cell lines.* Oncogene, 2002. **21**(26): p. 4080-8.
101. Boukerche, H., et al., *mda-9/Syntenin: a positive regulator of melanoma metastasis.* Cancer Res, 2005. **65**(23): p. 10901-11.
102. Kang, S., et al., *PCD1, a novel gene containing PDZ and LIM domains, is overexpressed in several human cancers.* Cancer Res, 2000. **60**(18): p. 5296-302.
103. Furuya, M., et al., *A novel gene containing PDZ and LIM domains, PCD1, is overexpressed in human colorectal cancer.* Anticancer Res, 2002. **22**(6C): p. 4183-6.
104. Sasaki, M., et al., *PCD1, a novel gene containing PDZ and LIM domains, is overexpressed in human breast cancer and linked to lymph node metastasis.* Anticancer Res, 2003. **23**(3B): p. 2717-21.
105. Chaib, H., et al., *Activated in prostate cancer: a PDZ domain-containing protein highly expressed in human primary prostate tumors.* Cancer Res, 2001. **61**(6): p. 2390-4.
106. Akabas, M.H., *Cystic fibrosis transmembrane conductance regulator. Structure and function of an epithelial chloride channel.* J Biol Chem, 2000. **275**(6): p. 3729-32.
107. Anderson, M.P., et al., *Demonstration that CFTR is a chloride channel by alteration of its anion selectivity.* Science, 1991. **253**(5016): p. 202-5.
108. Hall, R.A., et al., *A C-terminal motif found in the beta2-adrenergic receptor, P2Y1 receptor and cystic fibrosis transmembrane conductance*

regulator determines binding to the Na+/H+ exchanger regulatory factor family of PDZ proteins. Proc Natl Acad Sci U S A, 1998. **95**(15): p. 8496-501.

109. Short, D.B., et al., *An apical PDZ protein anchors the cystic fibrosis transmembrane conductance regulator to the cytoskeleton.* J Biol Chem, 1998. **273**(31): p. 19797-801.
110. Raghuram, V., H. Hormuth, and J.K. Foskett, *A kinase-regulated mechanism controls CFTR channel gating by disrupting bivalent PDZ domain interactions.* Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9620-5.
111. Fallon, L., et al., *Parkin and CASK/LIN-2 associate via a PDZ-mediated interaction and are co-localized in lipid rafts and postsynaptic densities in brain.* J Biol Chem, 2002. **277**(1): p. 486-91.
112. Dev, K.K., et al., *Part I: parkin-associated proteins and Parkinson's disease.* Neuropharmacology, 2003. **45**(1): p. 1-13.
113. Takamiya, K., et al., *A direct functional link between the multi-PDZ domain protein GRIP1 and the Fraser syndrome protein Fras1.* Nat Genet, 2004. **36**(2): p. 172-7.
114. Bladt, F., et al., *Epidermolysis bullosa and embryonic lethality in mice lacking the multi-PDZ domain protein GRIP1.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 6816-21.
115. Tomita, S., et al., *Interaction of a neuron-specific protein containing PDZ domains with Alzheimer's amyloid precursor protein.* J Biol Chem, 1999. **274**(4): p. 2243-54.
116. Songyang, Z., et al., *Recognition of unique carboxyl-terminal motifs by distinct PDZ domains.* Science, 1997. **275**(5296): p. 73-7.
117. Bezprozvanny, I. and A. Maximov, *Classification of PDZ domains.* FEBS Lett, 2001. **509**(3): p. 457-62.
118. SMART. Available from: http://smart.embl-heidelberg.de/.
119. Harris, B.Z. and W.A. Lim, *Mechanism and role of PDZ domains in signaling complex assembly.* J Cell Sci, 2001. **114**(Pt 18): p. 3219-31.

120. Daniels, D.L., et al., *Crystal structure of the hCASK PDZ domain reveals the structural basis of class II PDZ domain target recognition.* Nat Struct Biol, 1998. **5**(4): p. 317-25.
121. Hillier, B.J., et al., *Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex.* Science, 1999. **284**(5415): p. 812-5.
122. Tochio, H., et al., *Solution structure of the extended neuronal nitric oxide synthase PDZ domain complexed with an associated peptide.* Nat Struct Biol, 1999. **6**(5): p. 417-21.
123. Niethammer, M., et al., *CRIPT, a novel postsynaptic protein that binds to the third PDZ domain of PSD-95/SAP90.* Neuron, 1998. **20**(4): p. 693-707.
124. Tochio, H., et al., *Formation of nNOS/PSD-95 PDZ dimer requires a preformed beta-finger structure from the nNOS PDZ domain.* J Mol Biol, 2000. **303**(3): p. 359-70.
125. Gee, S.H., et al., *Single-amino acid substitutions alter the specificity and affinity of PDZ domains for their ligands.* Biochemistry, 2000. **39**(47): p. 14638-46.
126. Novak, K.A., N. Fujii, and R.K. Guy, *Investigation of the PDZ domain ligand binding site using chemically modified peptides.* Bioorg Med Chem Lett, 2002. **12**(17): p. 2471-4.
127. Harris, B.Z., et al., *Role of electrostatic interactions in PDZ domain ligand recognition.* Biochemistry, 2003. **42**(10): p. 2797-805.
128. Li, T., D. Saro, and M.R. Spaller, *Thermodynamic profiling of conformationally constrained cyclic ligands for the PDZ domain.* Bioorganic & Medicinal Chemistry Letters, 2004. **14**(6): p. 1385-1388.
129. Udugamasooriya, G., D. Saro, and M.R. Spaller, *Bridged Peptide Macrocycles as Ligands for PDZ Domain Proteins*. 2005. p. 1203-1206.
130. Piserchio, A., et al., *Targeting specific PDZ domains of PSD-95; structural basis for enhanced affinity and enzymatic stability of a cyclic peptide.* Chem Biol, 2004. **11**(4): p. 469-73.

131. Saro, D., et al., *Thermodynamic analysis of a hydrophobic binding site: probing the PDZ domain with nonproteinogenic peptide ligands.* Org Lett, 2004. **6**(20): p. 3429-32.
132. Madsen, K.L., et al., *Molecular determinants for the complex binding specificity of the PDZ domain in PICK1.* J Biol Chem, 2005. **280**(21): p. 20539-48.
133. Wang, L., A. Piserchio, and D.F. Mierke, *Structural characterization of the intermolecular interactions of synapse-associated protein-97 with the NR2B subunit of N-methyl-D-aspartate receptors.* J Biol Chem, 2005. **280**(29): p. 26992-6.
134. Sun, C. and D.F. Mierke, *Characterization of interactions of Na+/H+ exchanger regulatory factor-1 with the parathyroid hormone receptor and phospholipase C.* J Pept Res, 2005. **65**(3): p. 411-7.
135. Lin, D., et al., *The carboxyl terminus of B class ephrins constitutes a PDZ domain binding motif.* J Biol Chem, 1999. **274**(6): p. 3726-33.
136. Saro, D., et al., *Thermodynamic Analysis of a Hydrophobic Binding Site: Probing the PDZ Domain with Nonproteinogenic Peptide Ligands*. 2004. p. 3429-3432.
137. Grootjans, J.J., et al., *Syntenin-syndecan binding requires syndecan-synteny and the co-operation of both PDZ domains of syntenin.* J Biol Chem, 2000. **275**(26): p. 19933-41.
138. Marfatia, S.M., et al., *The PDZ domain of human erythrocyte p55 mediates its binding to the cytoplasmic carboxyl terminus of glycophorin C. Analysis of the binding interface by in vitro mutagenesis.* J Biol Chem, 1997. **272**(39): p. 24191-7.
139. Anzai, N., et al., *The multivalent PDZ domain-containing protein PDZK1 regulates transport activity of renal urate-anion exchanger URAT1 via its C terminus.* J Biol Chem, 2004. **279**(44): p. 45942-50.
140. Duffy, H.S., et al., *Kinetics of protein-protein interactions of connexins: use of enzyme linked sorbent assays.* Cell Commun Adhes, 2003. **10**(4-6): p. 207-10.

141. Myszka, D.G., *Improving biosensor analysis.* J Mol Recognit, 1999. **12**(5): p. 279-84.
142. Nguyen, J.T., et al., *Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors.* Science, 1998. **282**(5396): p. 2088-92.
143. Matsuda, S., S. Mikawa, and H. Hirai, *Phosphorylation of serine-880 in GluR2 by protein kinase C prevents its C terminus from binding with glutamate receptor-interacting protein.* J Neurochem, 1999. **73**(4): p. 1765-8.
144. Jelen, F., et al., *PDZ domains - common players in the cell signaling.* Acta Biochim Pol, 2003. **50**(4): p. 985-1017.
145. Cao, T.T., et al., *A kinase-regulated PDZ-domain interaction controls endocytic sorting of the beta2-adrenergic receptor.* Nature, 1999. **401**(6750): p. 286-90.
146. Cohen, N.A., et al., *Binding of the inward rectifier K+ channel Kir 2.3 to PSD-95 is regulated by protein kinase A phosphorylation.* Neuron, 1996. **17**(4): p. 759-67.
147. Hurd, T.W., et al., *Direct interaction of two polarity complexes implicated in epithelial tight junction assembly.* Nat Cell Biol, 2003. **5**(2): p. 137-42.
148. Peterson, F.C., et al., *Cdc42 regulates the Par-6 PDZ domain through an allosteric CRIB-PDZ transition.* Mol Cell, 2004. **13**(5): p. 665-76.
149. Fuentes, E.J., C.J. Der, and A.L. Lee, *Ligand-dependent dynamics and intramolecular signaling in a PDZ domain.* J Mol Biol, 2004. **335**(4): p. 1105-15.
150. Smith, D.B. and K.S. Johnson, *Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.* Gene, 1988. **67**(1): p. 31-40.
151. Saxena, V.P. and D.B. Wetlaufer, *Formation of three-dimensional structure in proteins. I. Rapid nonenzymic reactivation of reduced lysozyme.* Biochemistry, 1970. **9**(25): p. 5015-23.

152. Sheng, M., *Molecular organization of the postsynaptic specialization.* Proc Natl Acad Sci U S A, 2001. **98**(13): p. 7058-61.
153. Brenman, J.E., et al., *Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.* Cell, 1996. **84**(5): p. 757-67.
154. Dawson, V.L., et al., *Nitric oxide mediates glutamate neurotoxicity in primary cortical cultures.* Proc Natl Acad Sci U S A, 1991. **88**(14): p. 6368-71.
155. Migaud, M., et al., *Enhanced long-term potentiation and impaired learning in mice with mutant postsynaptic density-95 protein.* Nature, 1998. **396**(6710): p. 433-439.
156. Sattler, R., et al., *Specific Coupling of NMDA Receptor Activation to Nitric Oxide Neurotoxicity by PSD-95 Protein*. 1999. p. 1845-1848.
157. Simon, R.P., et al., *Blockade of N-methyl-D-aspartate receptors may protect against ischemic damage in the brain*. 1984. p. 850-852.
158. Fix, A.S., et al., *Neuronal Vacuolization and Necrosis Induced by the Noncompetitive N-Methyl-D-Aspartate (Nmda) Antagonist Mk(+)801 (Dizocilpine Maleate) - a Light and Electron-Microscopic Evaluation of the Rat Retrosplenial Cortex.* Experimental Neurology, 1993. **123**(2): p. 204-215.
159. Davis, S.M., et al., *Selfotel in Acute Ischemic Stroke : Possible Neurotoxic Effects of an NMDA Antagonist*. 2000. p. 347-354.
160. Morris, G.F., et al., *Failure of the competitive N-methyl-D-aspartate antagonist Selfotel (CGS 19755) in the treatment of severe head injury: results of two Phase III clinical trials.* Journal of Neurosurgery, 1999. **91**(5): p. 737-743.
161. Manabe, S. and S.A. Lipton, *Divergent NMDA signals leading to proapoptotic and antiapoptotic pathways in the rat retina.* Invest Ophthalmol Vis Sci, 2003. **44**(1): p. 385-92.

162. Sucher, N.J., et al., *N-methyl-D-aspartate receptor subunit NR3A in the retina: developmental expression, cellular localization, and functional aspects.* Invest Ophthalmol Vis Sci, 2003. **44**(10): p. 4451-6.
163. Goebel, D.J. and B.S. Winkler, *Blockade of PARP activity attenuates poly(ADP-ribosyl)ation but offers only partial neuroprotection against NMDA-induced cell death in the rat retina.* J Neurochem, 2006. **98**(6): p. 1732-45.
164. Zhang, J., et al., *Nitric oxide activation of poly(ADP-ribose) synthetase in neurotoxicity.* Science, 1994. **263**(5147): p. 687-9.
165. Ha, H.C. and S.H. Snyder, *Poly(ADP-ribose) polymerase is a mediator of necrotic cell death by ATP depletion.* Proc Natl Acad Sci U S A, 1999. **96**(24): p. 13978-82.
166. Sheng, M. and C. Sala, *PDZ domains and the organization of supramolecular complexes.* Annual Review of Neuroscience, 2001. **24**: p. 1-29.
167. Saro, D., et al., *A Thermodynamic Ligand Binding Study of the Third PDZ Domain (PDZ3) from the Mammalian Neuronal Protein PSD-95*. 2007. p. 6340-6352.
168. Saro, D., et al., *A thermodynamic ligand binding study of the third PDZ domain (PDZ3) from the mammalian neuronal protein PSD-95.* Biochemistry, 2007. **46**(21): p. 6340-52.
169. Cierpicki, T., J.H. Bushweller, and Z.S. Derewenda, *Probing the supramodular architecture of a multidomain protein: the structure of syntenin in solution.* Structure, 2005. **13**(2): p. 319-27.
170. Lim, I.A., D.D. Hall, and J.W. Hell, *Selectivity and promiscuity of the first and second PDZ domains of PSD-95 and synapse-associated protein 102.* J Biol Chem, 2002. **277**(24): p. 21697-711.
171. McRee, D.E., *Practical Protein Crystallography; 1-st edition ed.* Academic Press
Inc.: San Diego, 1993.
172. Drenth, J., *Principles of Protein X-ray Crystallography.* 1994.

173. Karthikeyan, S., et al., *Crystal structure of the PDZ1 domain of human Na(+)/H(+) exchanger regulatory factor provides insights into the mechanism of carboxyl-terminal leucine recognition by class I PDZ domains.* J Mol Biol, 2001. **308**(5): p. 963-73.
174. Im, Y.J., et al., *Crystal structure of GRIP1 PDZ6-peptide complex reveals the structural basis for class II PDZ target recognition and PDZ domain-mediated multimerization.* J Biol Chem, 2003. **278**(10): p. 8501-7.
175. Zhang, Q., J.S. Fan, and M. Zhang, *Interdomain chaperoning between PSD-95, Dlg, and Zo-1 (PDZ) domains of glutamate receptor-interacting proteins.* J Biol Chem, 2001. **276**(46): p. 43216-20.
176. Kang, B.S., et al., *PDZ tandem of human syntenin: crystal structure and functional properties.* Structure, 2003. **11**(4): p. 459-68.
177. Feng, W., et al., *PDZ7 of glutamate receptor interacting protein binds to its target via a novel hydrophobic surface area.* J Biol Chem, 2002. **277**(43): p. 41140-6.
178. Feng, W., et al., *Homer regulates gain of ryanodine receptor type 1 channel complex.* J Biol Chem, 2002. **277**(47): p. 44722-30.
179. Wilken, C., et al., *Crystal structure of the DegS stress sensor: How a PDZ domain recognizes misfolded protein and activates a protease.* Cell, 2004. **117**(4): p. 483-94.
180. Krojer, T., et al., *Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine.* Nature, 2002. **416**(6879): p. 455-9.
181. Kang, B.S., et al., *The PDZ2 domain of syntenin at ultra-high resolution: bridging the gap between macromolecular and small molecule crystallography.* J Mol Biol, 2004. **338**(3): p. 483-93.
182. Vitkup, D., et al., *Why protein R-factors are so large: a self-consistent analysis.* Proteins, 2002. **46**(4): p. 345-54.
183. Christopherson, K.S., et al., *PSD-95 assembles a ternary complex with the N-methyl-D-aspartic acid receptor and a bivalent neuronal NO synthase PDZ domain.* J Biol Chem, 1999. **274**(39): p. 27467-73.

184. Wang, S., et al., *Accessory protein facilitated CFTR-CFTR interaction, a molecular mechanism to potentiate the chloride channel activity.* Cell, 2000. **103**(1): p. 169-79.

185. Raghuram, V., D.O. Mak, and J.K. Foskett, *Regulation of cystic fibrosis transmembrane conductance regulator single-channel gating by bivalent PDZ-domain-mediated interaction.* Proc Natl Acad Sci U S A, 2001. **98**(3): p. 1300-5.

186. Long, J.F., et al., *Supramodular structure and synergistic target binding of the N-terminal tandem PDZ domains of PSD-95.* J Mol Biol, 2003. **327**(1): p. 203-14.

187. Tochio, H., et al., *Solution structure and backbone dynamics of the second PDZ domain of postsynaptic density-95.* J Mol Biol, 2000. **295**(2): p. 225-37.

188. Grembecka, J., et al., *The binding of the PDZ tandem of syntenin to target proteins.* Biochemistry, 2006. **45**(11): p. 3674-83.

189. Navaza, J., *On the computation of structure factors by FFT techniques.* Acta Crystallogr A, 2002. **58**(Pt 6): p. 568-73.

190. Skre, H., *Friedreich's ataxia in western Norway*, in *Clin.Genet*. 1975. p. 287-298.

191. Winter, R.M., Harding, A.E., Baraitser, M., and Bravery, M.B., *Intrafamilial correlation in Friedreich's ataxia.* Clin. Genet., 1981. **20**: p. 419-427.

192. Pandolfo M, S.G., Antonelli A, Weitnauer L, Ferretti L, Leone M, Dones I, Cerino A, Fujita R, Hanauer A, et al., *Friedreich ataxia in Italian families: genetic homogeneity and linkage disequilibrium with the marker loci D9S5 and D9S15.* Am J Hum Genet., 1990. **47**(2): p. 228-235.

193. Puccio, H. and M. Koenig, *Recent advances in the molecular pathogenesis of Friedreich ataxia.* Hum Mol Genet, 2000. **9**(6): p. 887-92.

194. Campuzano, V., et al., *Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes.* Hum Mol Genet, 1997. **6**(11): p. 1771-80.

195. Koutnikova, H., et al., *Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin.* Nat Genet, 1997. **16**(4): p. 345-51.
196. Rotig, A., et al., *Aconitase and mitochondrial iron-sulphur protein deficiency in Friedreich ataxia.* Nat Genet, 1997. **17**(2): p. 215-7.
197. Bradley, J.L., et al., *Clinical, biochemical and molecular genetic correlations in Friedreich's ataxia.* Hum Mol Genet, 2000. **9**(2): p. 275-82.
198. Finney, L.A. and T.V. O'Halloran, *Transition metal speciation in the cell: insights from the chemistry of metal ion receptors.* Science, 2003. **300**(5621): p. 931-6.
199. Pierrel, F., P.A. Cobine, and D.R. Winge, *Metal Ion availability in mitochondria.* Biometals, 2007. **20**(3-4): p. 675-82.
200. Kaplan, J. and T.V. O'Halloran, *Iron metabolism in eukaryotes: Mars and Venus at it again.* Science, 1996. **271**(5255): p. 1510-2.
201. Bencze, K.Z., et al., *The structure and function of frataxin.* Crit Rev Biochem Mol Biol, 2006. **41**(5): p. 269-91.
202. Andrews, N.C., *Disorders of iron metabolism.* N Engl J Med, 1999. **341**(26): p. 1986-95.
203. Johnson, D.C., et al., *Structure, function, and formation of biological iron-sulfur clusters.* Annu Rev Biochem, 2005. **74**: p. 247-81.
204. Labbe, R.F., H.J. Vreman, and D.K. Stevenson, *Zinc protoporphyrin: A metabolite with a mission.* Clin Chem, 1999. **45**(12): p. 2060-72.
205. Tait, G.H., *The Biosynthesis and degradation of heme.In: F. De Matteis and N.W. Aldridge, Editors.* Heme and Hemoproteins, Springer-Verlag, Berlin 1978: p. pp. 1–35.
206. Taketani, S., *Aquisition, mobilization and utilization of cellular iron and heme: endless findings and growing evidence of tight regulation.* Tohoku J Exp Med, 2005. **205**(4): p. 297-318.
207. Dailey, H.A., et al., *Ferrochelatase at the millennium: structures, mechanisms and [2Fe-2S] clusters.* Cell Mol Life Sci, 2000. **57**(13-14): p. 1909-26.

208. Dailey, H.A.a.D., T.A., *Ferrochelatase. In The Porphyrin handbook (Kadish,K.M. et al., eds)* 2003: Elsevier Science. 93-121.
209. Chen, O.S., et al., *Transcription of the yeast iron regulon does not respond directly to iron but rather to iron-sulfur cluster biosynthesis.* J Biol Chem, 2004. **279**(28): p. 29513-8.
210. Kispal, G., et al., *The mitochondrial proteins Atm1p and Nfs1p are essential for biogenesis of cytosolic Fe/S proteins.* Embo J, 1999. **18**(14): p. 3981-9.
211. Lill, R. and U. Muhlenhoff, *Iron-sulfur-protein biogenesis in eukaryotes.* Trends Biochem Sci, 2005. **30**(3): p. 133-41.
212. Muhlenhoff, U., et al., *Components involved in assembly and dislocation of iron-sulfur clusters on the scaffold protein Isu1p.* Embo J, 2003. **22**(18): p. 4815-25.
213. Wiedemann, N., et al., *Essential role of Isd11 in mitochondrial iron-sulfur cluster synthesis on Isu scaffold proteins.* Embo J, 2006. **25**(1): p. 184-95.
214. Adam, A.C., et al., *The Nfs1 interacting protein Isd11 has an essential role in Fe/S cluster biogenesis in mitochondria.* Embo J, 2006. **25**(1): p. 174-83.
215. Schilke, B., Voisine,C., Beinert,H., Craig,E., *Evidence for a conserved system for iron metabolism in the mitochondria of Saccharomydes cervisiae.* Proc. Natl. Adad. Aci, USA, 1999. **96**: p. 10206-10211.
216. Gerber, J., U. Muhlenhoff, and R. Lill, *An interaction between frataxin and Isu1/Nfs1 that is crucial for Fe/S cluster synthesis on Isu1.* EMBO Rep, 2003. **4**(9): p. 906-11.
217. Garland, S.A., et al., *Saccharomyces cerevisiae ISU1 and ISU2: members of a well-conserved gene family for iron-sulfur cluster assembly.* J Mol Biol, 1999. **294**(4): p. 897-907.
218. Campuzano, V., et al., *Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion.* Science, 1996. **271**(5254): p. 1423-7.

219. Bidichandani, S.I., T. Ashizawa, and P.I. Patel, *The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure.* Am J Hum Genet, 1998. **62**(1): p. 111-21.

220. Ohshima, K., et al., *Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo.* J Biol Chem, 1998. **273**(23): p. 14588-95.

221. Pandolfo, M., *The molecular basis of Friedreich ataxia.* Adv Exp Med Biol, 2002. **516**: p. 99-118.

222. MacKenzie, E.L., K. Iwasaki, and Y. Tsuji, *Intracellular iron transport and storage: from molecular mechanisms to health implications.* Antioxid Redox Signal, 2008. **10**(6): p. 997-1030.

223. Babcock, M., et al., *Regulation of mitochondrial iron accumulation by Yfh1p, a putative homolog of frataxin.* Science, 1997. **276**(5319): p. 1709-12.

224. Foury, F. and O. Cazzalini, *Deletion of the yeast homologue of the human gene associated with Friedreich's ataxia elicits iron accumulation in mitochondria.* FEBS Lett, 1997. **411**(2-3): p. 373-7.

225. Knight, S.A., Sepuri,N.B., Pain,D., Dancis,A., *Mt-Hsp70 Homolog, Scs2p, Required for Maturation of Yeast Frataxin and Mitochondrial Iron Homeostasis.* J.B.C., 1998. **273**: p. 18389-18393.

226. Radisky, D.C., M.C. Babcock, and J. Kaplan, *The yeast frataxin homologue mediates mitochondrial iron efflux. Evidence for a mitochondrial iron cycle.* J Biol Chem, 1999. **274**(8): p. 4497-9.

227. Adamec, J., et al., *Iron-dependent self-assembly of recombinant yeast frataxin: implications for Friedreich ataxia.* Am J Hum Genet, 2000. **67**(3): p. 549-62.

228. Cavadini, P., et al., *Assembly and iron-binding properties of human frataxin, the protein deficient in Friedreich ataxia.* Hum Mol Genet, 2002. **11**(3): p. 217-27.

229. Adinolfi, S., et al., *A structural approach to understanding the iron-binding properties of phylogenetically different frataxins.* Hum. Mol. Genet., 2002. **11**: p. 1865-77.

230. Bou-Abdallah, F., et al., *Iron binding and oxidation kinetics in frataxin CyaY of Escherichia coli.* J Mol Biol, 2004. **341**(2): p. 605-15.

231. Yoon, T. and J.A. Cowan, *Frataxin-mediated iron delivery to ferrochelatase in the final step of heme biosynthesis.* J Biol Chem, 2004. **279**(25): p. 25943-6.

232. Yoon, T. and J.A. Cowan, *Iron-sulfur cluster biosynthesis. Characterization of frataxin as an iron donor for assembly of [2Fe-2S] clusters in ISU-type proteins.* J Am Chem Soc, 2003. **125**(20): p. 6078-84.

233. Cook, J.D., et al., *Monomeric yeast frataxin is an iron-binding protein.* Biochemistry, 2006. **45**(25): p. 7767-77.

234. Aloria, K., et al., *Iron-induced oligomerization of yeast frataxin homologue Yfh1 is dispensable in vivo.* EMBO Rep, 2004. **5**(11): p. 1096-101.

235. Ristow, M., et al., *Frataxin activates mitochondrial energy conversion and oxidative phosphorylation.* Proc Natl Acad Sci U S A, 2000. **97**(22): p. 12239-43.

236. Bulteau, A.L., et al., *Frataxin acts as an iron chaperone protein to modulate mitochondrial aconitase activity.* Science, 2004. **305**(5681): p. 242-5.

237. Bulteau, A.L., et al., *Reversible redox-dependent modulation of mitochondrial aconitase and proteolytic activity during in vivo cardiac ischemia/reperfusion.* Proc Natl Acad Sci U S A, 2005. **102**(17): p. 5987-91.

238. Lesuisse, E., et al., *Iron use for haeme synthesis is under control of the yeast frataxin homologue (Yfh1).* Hum Mol Genet, 2003. **12**(8): p. 879-89.

239. Schoenfeld, R.A., et al., *Frataxin deficiency alters heme pathway transcripts and decreases mitochondrial heme metabolites in mammalian cells.* Hum Mol Genet, 2005. **14**(24): p. 3787-99.

240. Zhang, Y., et al., *Frataxin and mitochondrial carrier proteins, Mrs3p and Mrs4p, cooperate in providing iron for heme synthesis.* J Biol Chem, 2005. **280**(20): p. 19794-807.
241. Stehling, O., et al., *Iron-sulfur protein maturation in human cells: evidence for a function of frataxin.* Hum Mol Genet, 2004. **13**(23): p. 3007-15.
242. Muhlenhoff, U., et al., *The yeast frataxin homolog Yfh1p plays a specific role in the maturation of cellular Fe/S proteins.* Hum Mol Genet, 2002. **11**(17): p. 2025-36.
243. Cho, S.J., et al., *Crystal structure of Escherichia coli CyaY protein reveals a previously unidentified fold for the evolutionarily conserved frataxin family.* Proc Natl Acad Sci U S A, 2000. **97**(16): p. 8932-7.
244. Dhe-Paganon, S., et al., *Crystal structure of human frataxin.* J Biol Chem, 2000. **275**(40): p. 30753-6.
245. He, Y., et al., *Yeast frataxin solution structure, iron binding, and ferrochelatase interaction.* Biochemistry, 2004. **43**(51): p. 16254-62.
246. Lee, M.G., et al., *Crystallization and preliminary X-ray crystallographic analysis of Escherichia coli CyaY, a structural homologue of human frataxin.* Acta Crystallogr D Biol Crystallogr, 2000. **56 (Pt 7)**: p. 920-1.
247. Musco, G., et al., *Towards a Structural Understanding of Friedreich's Ataxia: The Solution Structure of Frataxin.* Structure, 2000. **8**(7): p. 695-707.
248. Nair, M., et al., *Solution structure of the bacterial frataxin ortholog, CyaY: mapping the iron binding sites.* Structure (Camb), 2004. **12**(11): p. 2037-48.
249. Teo, B.K., *EXAFS: Basic Principles and Data Analysis*. 1986, New York: Springer-Verlag.
250. Zhang, M. and W. Wang, *Organization of signaling complexes by PDZ-domain scaffold proteins.* Acc Chem Res, 2003. **36**(7): p. 530-8.
251. Craven, S.E. and D.S. Bredt, *PDZ proteins organize synaptic signaling pathways.* Cell, 1998. **93**(4): p. 495-8.

252. Garner, C.C., J. Nash, and R.L. Huganir, *PDZ domains in synapse assembly and signalling.* Trends Cell Biol, 2000. **10**(7): p. 274-80.
253. Zimmermann, P., et al., *Characterization of syntenin, a syndecan-binding PDZ protein, as a component of cell adhesion sites and microfilaments.* Mol Biol Cell, 2001. **12**(2): p. 339-50.
254. Okamoto, M. and T.C. Sudhof, *Mints, Munc18-interacting proteins in synaptic vesicle exocytosis.* J Biol Chem, 1997. **272**(50): p. 31459-64.
255. Dong, H., et al., *GRIP: a synaptic PDZ domain-containing protein that interacts with AMPA receptors.* Nature, 1997. **386**(6622): p. 279-84.
256. Srivastava, S., et al., *Novel anchorage of GluR2/3 to the postsynaptic density by the AMPA receptor-binding protein ABP.* Neuron, 1998. **21**(3): p. 581-91.
257. Wyszynski, M., et al., *Association of AMPA receptors with a subset of glutamate receptor-interacting protein in vivo.* J Neurosci, 1999. **19**(15): p. 6528-37.
258. Lau, K.F., et al., *X11 alpha and x11 beta interact with presenilin-1 via their PDZ domains.* Mol Cell Neurosci, 2000. **16**(5): p. 557-65.
259. Jannatipour, M., et al., *Schwannomin isoform-1 interacts with syntenin via PDZ domains.* J Biol Chem, 2001. **276**(35): p. 33093-100.
260. Geijsen, N., et al., *Cytokine-specific transcriptional regulation through an IL-5Ralpha interacting protein.* Science, 2001. **293**(5532): p. 1136-8.
261. Klosi, E., D. Saro, and M.R. Spaller, *Bivalent peptides as PDZ domain ligands.* Bioorganic & Medicinal Chemistry Letters, 2007. **17**(22): p. 6147-6150.
262. Long, J., et al., *Supramodular nature of GRIP1 revealed by the structure of its PDZ12 tandem in complex with the carboxyl tail of Fras1.* J Mol Biol, 2008. **375**(5): p. 1457-68.
263. Feng, W., et al., *Tandem PDZ repeats in glutamate receptor-interacting proteins have a novel mode of PDZ domain-mediated target binding.* Nat Struct Biol, 2003. **10**(11): p. 972-8.

264. Walter, P.B., et al., *Iron deficiency and iron excess damage mitochondria and mitochondrial DNA in rats.* Proc Natl Acad Sci U S A, 2002. **99**(4): p. 2264-9.

ABSTRACT

ISOTHERMAL TITRATION CALORIMETRY STUDIES OF PROTEIN-MEDIATED INTERACTIONS AND PRELIMINARY STRUCTURAL STUDIES OF TANDEM PDZ1-2 DOMAIN OF PSD-95 PROTEIN

by

ANA D. JANKOVIC

August 2010

Advisor: Dr. Mark R. Spaller

Major: Chemistry

Degree: Doctor of Philosophy

Protein-mediated interactions that involve multiple ligands in their binding mechanisms are critical for many cellular functions. The primary focus of this dissertation research was to investigate such interactions for two proteins, the PDZ domain and frataxin, involving peptide and metal binding ligands, respectively. The three component projects of this work comprised (1) thermodynamic analysis of PDZ domain binding using calorimetry; (2) X-ray crystallographic structural studies of a PDZ dual domain; and (3) thermodynamic analysis of frataxin binding to iron. The specific goal of the research conducted with the PDZ domains was to understand the mechanism of action of multiple/tandem protein domains. The protein chosen for study was the dual domain PDZ1-2 of postsynaptic density-95 kDa (PSD-95) protein, a mammalian neuronal protein. The individual constructs, PDZ1 and PDZ2, were also prepared

for comparative studies. Challenging the individual PDZ1 and PDZ2 and dual domains with short peptides derived form the C-terminal region of the domains' natural binding partners (8-mer to 6-mer) sequences, it was found that only *one* of the domains remains functional when expressed separately. Binding of PDZ1 and tandem PDZ1-2 could be characterized, whereas PDZ2 could not be stabilized under solution conditions to yield a valid binding curve. The use of bivalent peptides that would bind with 1:1 stoichiometry to the tandem PDZ1-2 domain was problematic due to ligand insolubility. In an ongoing effort to develop new cellular probes for *in vivo* investigation of multiprotein assembly, we studied modified ligand CN2180 which targets the tandem PDZ1-2 of PSD-95. Our *in vitro* testing by ITC of the same peptide after integrating the binding isoterm, yielded a stoichiometry of n=2, indicating that both binding sites are fully occupied, and dissociation constants for each of the two interaction sites were in the single digit micromolar range.

In terms of progress on that structural front, the dual PDZ domain protein construct PDZ1-2 of PSD-95 was successfully crystallized. Identifying the conditions required to grow crystals is a major step toward the goal of solving the structure of this tandem domain by X-ray methods. Optimization of the purification protocol resulted in high purity protein that allowed for increased volume of protein crystallization attempts. The optimized conditions yielded well-shaped hexagonal crystals for PDZ1-2, which crystallizes in cubic P23 group. Data reduction resulted in a R_{merge} value of 11.2 %. A final structure could not be

solved by molecular replacement methods, and additional work will be needed to accomplish this.

In the last project, calorimetry was employed to investigate the binding of frataxin to iron ions. This protein is essential for the effective regulation of cellular iron homeostasis. Frataxin, is essential for cellular iron control and is believed to serve as an iron chaperone that delivers mitochondrial Fe(II) to the enzymes ferrochelatase and the ISU apparatus for completion of heme and Fe-S cluster biosynthesis. Monomeric frataxin protein has a high affinity for ferrous ion, completely saturating at 2:1 iron to protein ratio. The micromolar dissociation constants measured for yeast frataxin, with respect to ferrous iron (K_d's of 2.0 and 3.0 μM) were obtained from calorimetric titrations. The weak interaction is consistent with the hypothesis of frataxin acting as an iron chaperone. Frataxin must be able to form a favorable interface with its protein partners and also easily release the metal. Although this thermodynamic binding study increases understanding of the underlying metal-binding behavior, additional structural characterization will be critical to help elucidate how frataxin binds iron and docks with its protein partners to promote metal delivery.

AUTOBIOGRAPHICAL STATEMENT

Education

2001-2009	Ph.D. Chemistry (Biochemistry), Wayne State University, Detroit, MI
1992-2000	Bachelor of Science in Biochemical Engineering, University of Belgrade, Serbia

Professional Experience

2008-2009	Research Assistant in Prof. Timothy Stemmler's laboratory, Department of Biochemistry and Molecular Biology, School of Medicine, WSU
2001-2007	Graduate Research Assistant in Prof. Mark Spaller's laboratory, Chemistry Department, WSU
2001-2005	Graduate Teaching Assistant, Chemistry Department, WSU
2000-2001	Teaching Assistant, Department of General and Inorganic Chemistry, Faculty of Technology and Metallurgy, University of Belgrade, Serbia

Awards

- Wayne Sate University Excellence in Teaching Award, 2005
- Wayne Sate University Excellence in Teaching Award, 2006

Manuscripts

1. Cook, J.D.; Bencze, K.Z; **Jankovic, A.D.**; Crater, A.K.; Busch, C.N.; Bradley, P.B.; Stemmler,A.J.; Spaller, M.R.; Stemmler, T.L. "Monomeric Yeast Frataxin is an Iron Binding Protein*", Biochemistry,* 2006 Jun 27;45(25):7767-77.
2. **Jankovic, A.**; Yedidi, R.; Kovari L.C.; Spaller, M.R. "Purification, Crystallization and Preliminary X-ray Crystallographic Study of Tandem PDZ1-2 Domain of PSD-95", *manuscript in preparation.*
3. Rupasinghe, C. N**.; Jankovic, A.D.**; Saro, D.; Goebel, D.J.; Spaller, M.R.; "Development of PDZ Domain-Targeting Cellular Probes and their in vivo activity", *manuscript in preparation*