STRUCTURAL INVESTIGATIONS OF THE CANCER-ASSOCIATED LAMININ BINDING PROTEIN AND NOS L, A NOVEL COPPER BINDING PROTEIN

by

Lara Marie Taubner

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

October 2005

©COPYRIGHT

by

Lara Marie Taubner

2005

All Rights Reserved

APPROVAL

of a dissertation submitted by

Lara Marie Taubner

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. Valérie Copié

Approved for the Department of Chemistry and Biochemistry

Dr. David Singel

Approved for the College of Graduate Studies

Dr. Joseph J. Fedock

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a doctorate's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Lara Marie Taubner
October 2005

DEDICATION

I would like to thank my mother, my father, and my sisters Kathleen and Sarah, for their unconditional love and support throughout these last years that has made this dissertation possible.

TABLE OF CONTENTS

1. INTRODUCTION 1

2. BACKGROUND ON THE LAMININ BINDING PROTEIN 4

Carcinogenesis 4
Extracellular Matrix-Cell Interactions and Tumorigenesis 9
Laminin-1 and Peptide 11 10
The Laminin Binding Protein 13
LBP in Normal Tissues 16
The Laminin Binding Protein and Integrin 17
Characterization of Native Laminin Binding Protein 20
Ribosomal Function of LBP 23
The Laminin Binding Domain of LBP 26
Hypothesis 28

2. MATERIALS AND METHODS FOR LBP 30

Cloning of Recombinant LBP in *E.coli* 30
Expression and Purification of Bacterial Recombinant LBP 30
Limited Proteolysis and Identification of Domains 31
Cloning and Expression of Proteolysis-derived Domains 32
ELISA Assay and Laminin-1 Sepharose Affinity Chromatography 33
CD and NMR Spectroscopy 34

3. RESULTS FOR LBP 36

Expression and Purification of Recombinant LBP 36
Biological Activity of Recombinant LBP 40
Limited Proteolysis of Recombinant LBP 40
Cloning and Expression of Proteolysis-derived Domains 47
Biological Activity Screen of LBP Domains 49
CD and NMR Spectroscopy of 200-295 51

4. DISCUSSION ON LBP 54

5. BACKGROUND ON NOS L 69

Denitrification 69
Nitrous Oxide Reductase 75
The *nos* Operon 83
NosL 87

TABLE OF CONTENTS CONTINUED

Metallochaperones 91
Hypothesis....96

6. MATERIALS AND METHODS FOR NOS L 98

Expression and Purification of Recombinant Apo-NosL 98
Expression and Purification of Recombinant Holo-NosL 98
NMR Spectroscopy....99
Structure Calculations....100

7. RESULTS FOR NOS L....106

Expression and Purification of Apo and Holo Nos L 106
Sequential Assignment and Secondary Structure of Apo NosL 106
Apo NosL solution structure....116
Structure homologs to apo NosL 129
Comparison of apo and holo ^{1}H-^{15}N HSQC spectra....131

10. DISCUSSION ON NOS L....139

11. CONCLUSIONS....148

REFERENCES CITED....151

APPENDICES 175

APPENDIX A. Complete Chemical Shift Assignments
for Apo NosL 176
APPENDIX B. Amide proton and ^{15}N data for apo vs
holo NosL....189
APPENDIX C. Chemical shift file generated by
NMRView....195
APPENDIX D. Cross peak file generated by
NMRView....197
APPENDIX E. Conversion file for converting
NMRView data into xml format....199
APPENDIX F. ARIA project file, setup and run....203

LIST OF TABLES

Table Page

1. Fragments of LBP retrieved from limited proteolysis experiments.....46
2. NMR experiments and parameters.....101
3. Restraints used in structure calculations and statistical analysis.....119
4. Residues exhibiting chemical shift differences between apo and holo NosL.....136

LIST OF FIGURES

Figure Page

1. Functional map of laminin 12
2. Focal adhesion schematic 19
3. Amino acid sequence of the laminin binding protein 23
4. A. Sequence similarity between the N-terminal region of LBP to prokaryotic ribosomal proteins 25

   B. Schematic representation of the modeled LBP structure 26
5. SDS-PAGE of full length recombinant LBP 37
6. MALDI-TOF MS of full length LBP 38
7. Native 15% polyacrylamide gel of recombinant LBP 39
8. ELISA assay of full-length recombinant LBP 41
9. Blot of fractions from laminin-Sepharose chromatography 42
10. Range-finding results 43
11. Limited proteolysis of recombinant LBP as visualized by SDS-PAGE 44
12. MALDI-TOF MS data collected on proteolysis fragments 45
13. The amino acid sequence of hamster LBP 46
14. Schematic representing the isolated domains 47
15. A. SDS-PAGE of 200-295 48

    B. SDS-PAGE of 1-220 and 137-230 48
16. MALDI-TOF MS of fragment 200-295 49

LIST OF FIGURES CONTINUED

Figure Page

17. ELISA assay of proteolysis-derived domains....50

18. CD spectrum collected on 200-295....52

19. ^{1}H-^{15}N HSQC NMR spectrum of construct 200-295....53

20. The global nitrogen cycle....70

21. The denitrification apparatus....74

22. Nitrous oxide reductase from *Pseudomonas nautica*....79

23. Crystal structure of the Cu_z cluster from *P. nautica*....81

24. Components of Cu_z assembly....87

25. Sequence alignment of full NosL sequences....89

26. Assigned ^{1}H-^{15}N HSQC spectrum of apo NosL....107

27. Enlarged region of the ^{1}H-^{15}N HSQC spectrum of apo NosL....108

28. Sequential stripplots from the CBCACONH and HNCACB....110

29. Summary of amide exchange data, patterns of sequential and short-range NOE's and secondary chemical shifts....113

30 A. and B. Strand topology of NosL....115

31. Ramachandran plot of apo NosL....118

32. Topology Diagram of NosL....121

33. Superposition of the two domains of NosL....121

34. Backbone wire diagram of the ensemble of 20 lowest energy structures....123

LIST OF FIGURES CONTINUED

Figure Page

35 A, B and C. Ribbon diagrams of apo NosL from different perspectives.....124

36 A and B. Molecular surface figures of apo NosL from two views.....125

37 A and B. Electrostatic surface potential of NosL from two views 127

38 A, B and C. Superposition of MerB with NosL.....130

39. Overlay of the ^{1}H-^{15}N-HSQC spectra of apo (black) and holo (red) NosL.....132

40. ^{1}H chemical shift differences between apo and holo NosL 134

41. ^{15}N chemical shift differences between apo and holo NosL.....134

42. Ribbon diagram of NosL with amide protons perturbed by the presence of copper highlighted 137

43. Superposition of NosL with MerB with metal-binding ligands shown 138

ABSTRACT

This thesis consists of two distinct projects, one on the metastasis-associated laminin binding protein and the other on the putative copper chaperone NosL, both related by the common aim of investigation of the relationship between protein structure and function using nuclear magnetic resonance techniques. In the first part of this dissertation, the role that the metastasis-associated laminin binding protein or LBP plays in the spread and development of cancer was investigated. Functional domains of LBP were delineated by limited proteolysis, overexpressed, and then assayed for their ability to bind to the previously identified *in vivo* binding partner laminin. These assays demonstrated that, at least under the conditions used in this assay, binding to laminin was localized to domain 137-230, a region that encompasses a previously identified binding site known as Peptide G. This protein, like the full-length recombinant laminin binding protein, aggregated under conditions used for nuclear magnetic resonance experiments and therefore could not be analyzed with this technique. Contrary to previous findings on a synthetic peptide corresponding to residues 205-229, this sequence within the context of the 200-295 construct demonstrated no laminin binding activity. Furthermore, the peptide lacked the predicted α-helical content and tertiary structure as ascertained by nuclear magnetic resonance and by circular dichroism spectroscopy. A potential role for the disorder exhibited by this region of LBP is proposed, and suggests possible new functions for the laminin binding protein in angiogenesis.

NosL, the subject of the second part of this thesis, is a highly conserved copper(I) binding lipoprotein encoded by the nitrous oxide reductase (*nos*) gene cluster of denitrifying bacteria. To identify functional features and structural homologues of this protein, the structure of apo NosL was solved using nuclear magnetic resonance techniques. The high-resolution structure of NosL consists of one four-strand antiparallel beta sheet, one three strand antiparallel beta sheet and two α-helices organized in a twisted butterfly-like fold that is structurally homologous to MerB, an alkyl mercury lyase. Chemical perturbation mapping performed on the copper(I)-protein defined regions of NosL potentially involved in copper binding, and thus allowed preliminary identification of the copper-binding ligand Met109.

INTRODUCTION

Tumor cell interactions with the extracellular matrix component laminin-1 are important determinants of cancer cells' ability to invade tissue compartments, degrade the extracellular matrix, and migrate during the processes of angiogenesis and metastasis. These processes are in part mediated by the cell-surface laminin binding protein or LBP. While there has been great interest in this protein due to the correlation between upregulated levels of LBP and the aggressiveness of tumor cells, attempts to obtain structural information on the lipidated 67 kDa protein have been disappointing thus far. In order to obtain the high concentrations of protein required for NMR work, the protein has been expressed in a bacterial over-expression system and proper folding verified by laminin-1 binding assays. Because the protein aggregated at the concentrations needed for NMR experiments, the construct was subjected to limited proteolysis in order to identify domains within the full-length sequence that may be more amenable to solution NMR studies. Once domains 1-136, 1-220, 200-295 and 137-230 were identified, they were assayed for biological activity. ELISA assays on laminin-1 binding to subdomains of LBP demonstrated that, at least under the conditions used in this assay, binding was localized to domain 137-230, a region that encompasses a previously identified binding site known as Peptide G. Contrary to previous findings on a synthetic peptide corresponding to residues 205-229, this sequence within the context of the 200-295 construct demonstrated no laminin-1 binding activity or the previously assumed α-helical content. This region demonstrated lack of secondary and tertiary structure by CD spectroscopy, 1D-, 2D- and ^{15}N-NOESY NMR techniques. A potential role for the

disorder exhibited by this region of LBP is proposed, and suggests possible new functions for the laminin binding protein.

NosL, the subject of the second part of this thesis, is a highly conserved outer membrane lipoprotein encoded by the nitrous oxide reductase (*nos*) gene cluster of denitrifying bacteria. This gene cluster is required for the bioassembly of nitrous oxide reductase or N_2OR, an enzyme that catalyzes the final step of denitrification, converting gaseous nitrous oxide into dinitrogen. Nitrous oxide reductase is able to activate the relatively inert substrate by virtue of a completely novel sulfide-bridged four copper cluster that requires the coexpression of the accessory proteins of the *nos* cluster. Unlike other components of this transcriptional unit, NosL is not strictly required for copper insertion into the catalytic site of N_2OR under the conditions examined and yet is present in every organism studied, suggesting and important but unidentified function for this protein. EXAFS data collected on recombinant NosL lacking the native thioether-linked lipid revealed that this protein binds copper(I) in a 1:1 stoichiometry via a unique ligand architecture, leading some to speculate that this protein may be a copper chaperone for the N_2OR system. In order to obtain information on potential functions of NosL and on the larger N_2OR assembly machinery, NMR investigations were undertaken to determine the three-dimensional solution structure of apo-NosL. The high resolution structure of NosL consists of two beta sheets, two α-helices, two 3_{10} helices and disordered termini. The sheets and α-helix α1 are orientated so as to create a wide, deep pocket surrounded by basic residues Arg53, Arg60, and Arg106 amidst a chiefly acidic potential surface. The only structural homolog of NosL is MerB, an alkyl mercury lyase that detoxifies

organomercury compounds by cleaving the carbon-mercury bond in a thiol-dependent manner. Chemical perturbation mapping experiments performed on the ^{15}N labeled holo protein defined regions of NosL involved in copper binding and therefore allowed preliminary identification of the copper binding ligands Met109 and Met26.

BACKGROUND ON THE LAMININ BINDING PROTEIN

Carcinogenesis

Cancer is a challenging condition to treat and to investigate because this illness is actually made up of many diseases, arising from multiple causes, displaying diverse cellular phenotypes, and responding to treatment in a variety ways. Fortunately, common patterns in the cellular biology and the underlying molecular mechanism of tumorigenesis have emerged, making a rational approach to this highly complex subject possible. Cancer originates from genetic damage to one progenitor cell, due to exposure to a physical, chemical or biological carcinogen. Upon the accumulation of other mutations and exposure to proliferative signals, subpopulations derived from the original transformed cell evolve and flourish, developing into increasingly malignant forms. Ultimately, the progeny escape their initial site and spread to distant tissues in a process known as metastasis, which is the major cause of cancer-related mortality. By the time metastases have formed, it may be impossible to determine the original tissue type of the cell due to the loss of all differentiation markers, reflecting the radical changes that have occurred to the cancer throughout the tumorigenic progression [1].

Mutations that cause cancer usually occur to genes known as proto-oncogenes and tumor suppressor genes. Proto-oncogenes are normal genes, that when mutated, give rise to protein products that deliver signals to proliferate—once mutated, these genes become oncogenes. These genes include those that code for extracellular growth factors, growth factor receptors, transcription factors, and intracellular signal transducers. *Ras*, a proto-

oncogene encoding an intracellular effector molecule, is one of the most commonly mutated genes, present in at least 30% of tumors [2, 3]. Ras is a small guanine nucleotide-binding protein that serves to couple cell-surface receptor signals to the mitogenic cascade. Mutated Ras, because it transduces a myriad of extracellular signals from many different types of receptor tyrosine kinases or RTK's, and because it triggers many downstream effectors such as the PI3K and Raf/MEK/MAPK pathways, is responsible for such diverse cancer capabilities as proliferation, motility, cell adherence, and protease production [2, 4].

Tumor suppressor genes, unlike proto-oncogenes, receive genetic damage that renders them inactive, leading to a decreased ability of the cell to constrain cell proliferation and to undergo apoptosis or programmed cell death. Tumor suppressor genes encode cell cycle control proteins, differentiation control proteins, proteins that mediate apoptosis, and transcription factors for cell cycle control factors. P53, a prototypical tumor suppressor gene, is a nuclear transcription factor that is exquisitely sensitive to DNA damage, hypoxia, oncogene expression, and abnormal cell proliferation, and mediates both growth arrest and apoptosis [5, 6]. Many observed mutations in the p53 allele occur on the DNA binding domain, and prevent proper transcription of a large number of DNA-repair and stress enzymes, as well as components of the apoptotic signaling machinery—the importance of this protein is demonstrated by the fact that it is the most frequently mutated gene in human cancers, present in greater than 50% of all tumors and up to 85% of breast, lung and colon cancers [1, 3, 6].

Genetic damage to both proto-oncogenes and tumor suppressor genes is critical to

the cancer initiation process because the net effect of these mutations is to uncouple cell growth and proliferation from the normally tightly regulated signaling from surrounding cells and the environment, as well as from the cell's innate ability to keep these processes in check. To highlight the importance of these initiating events, it has been found that damage to both types of genes is required and sufficient to transform human cells [1].

Once an initiated cell is stimulated to proliferate by an exogenous signal, the damaged cell begins a cascade of unchecked growth and the accompanying accumulation of genetic lesions, passing through a series of several discrete pre-malignant stages. Once damage to DNA repair enzymes occurs, the number of mutations and chromosomal abnormalities is allowed to accelerate, until these number in the thousands [7]. This progression of somatic mutations serves to confer the cells with a mitotic advantage over normal cells, following a Darwinian-type evolution towards an increasingly malignant phenotype, through which subpopulations derived from the original cancer ultimately display enhanced ability to recruit new vasculature and to metastasize.

While the specific path a cell takes to malignancy differs from individual to individual, characterization of highly malignant tumors reveals many underlying commonalities among them. Alongside the ability to proliferate uncontrollably, often by secreting their own growth hormones, cancer cells also consistently fail to respond to growth-inhibitory signals or their downstream effectors, such as those that trigger quiescence or terminal differentiation [3]. They secrete paracrine signals, received by the stroma, which then responds by releasing nutrients that furnish the tumor cells with additional growth factors. In the latter stages of cancer, tumor cells become

immortalized, most often by overexpressing telomerase; mutations in telomerase are found in over 85% of human cancers [8]. Telomerase is a ribonucleic protein responsible for repairing the telomeres that stabilize and protect the ends of chromosomes. It is believed that telomere length, which normally shortens with every cell division, as well as telomere structure, are responsible for cellular senescence, and thus act as a check on the unbridled proliferation that accompanies tumor growth [7]. Overexpression of telomerase inappropriately maintains telomere length, thereby preventing the cell death that should normally occur after a certain number of divisions. Throughout this malignant conversion, the cells must constantly evade the normal apoptotic processes that detect DNA damage and trigger cell death.

Highly aggressive subpopulations of the cancer eventually acquire the ability to manipulate their environment to both establish a blood supply and to metastasize. Tumors greater than 0.5 mm in diameter cannot be supplied with sufficient nutrients and oxygen by passive diffusion, and must recruit vessels to furnish a blood supply and remove waste [9]. To accomplish this, tumor cells secrete molecules like vascular endothelial growth factor (VEGF) and fibroblast growth factor (FGF) that stimulate the proliferation and migration of endothelial cells in order to extend a capillary sprout toward the tumor [10]. Activated endothelial cells secrete proteases to degrade the intervening stroma and ultimately lay down basement membrane, a continuous sheet of secreted interlacing proteins that line vessel walls, in order to complete tube formation [10]. The poorly regulated vascularization of tumors results in leaky, defective vessels, through which cells from the main tumor mass can escape into the blood supply. The

new blood supply allows the tumor population to grow and develop, such that the degree of vascularization of a tumor correlates with poor clinical prognosis. Not surprisingly, the upregulation of pro-angiogenic factors like VEGF and proteases, particularly matrix metalloproteinases and plasminogen activators, is a characteristic shared by many cancers [11].

In a related process, the tumor cell must also interact with and remodel its environment to escape the barriers between tissues and vasculature and to establish colonies in distant sites. To escape the primary tissue, the tumor cell secretes proteinases such as collagenase VI and uPA (urokinase-like plasminogen activator) to degrade the surrounding stroma and epithelium [12]. The migrating cell must then penetrate a blood or lymphatic vessel, where it is carried passively or aggregated with platelets and hematopoetic cells to a new site, often the next capillary bed. The detached cell must then adhere to the endothelium often via the inappropriate overexpression of cell adhesion molecules. Upon cell attachment, the endothelial cells retract, allowing tumor cells access to the underlying basement membrane [9]. The cell then extravasates or migrates through into the stroma, establishing a new colony, where it begins to multiply. These acquired capabilities rely on the misregulation of degradation/proteolytic enzymes, cell adhesion molecules, and motility factors; inhibitors of these molecules have been shown to reduce cell invasion and chemotaxis [12]. The complexity of the tumor cell-host interaction and host immune responses are responsible for the fact that metastasis is a highly inefficient process, with less than 0.01% circulating tumor cells successfully establishing a metastasis [12, 13]. Despite this inefficiency, the ability of a cancer cell to

establish secondary colonies is the most lethal aspect of this disease, leading to over 90% of all cancer-related deaths [14].

Extracellular Matrix-Cell Interactions and Tumorigenesis

Tumor cells must interact with the extra-cellular matrix or ECM at several points along the carcinogenic cascade in order to establish an adequate blood supply and to form metastases at distant locations. The extracellular matrix is a network of interlocking proteins, proteoglycans and accompanying polysaccharides secreted by epithelial, endothelial and mesynchymal cells, forming compartments between tissues and supports for these cells to grow and spread upon. As such, the ECM acts as a physical barrier to the spread of cancer. Ubiquitous throughout the body, variation in the different components of the ECM serve tissue- and development- specific needs in order to perform distinct functional roles [15]. Not only does the network of structural and adhesive proteins serve as a fine mechanical filter, but also as a reservoir for growth factors, anti-apoptic factors, and cytokines embedded within the matrix, regulating proliferation, differentiation, and migration of multiple cell types [16, 17]. Local degradation of the ECM by secreted and cell-bound proteinases exposes these cellular determinants, allowing dynamic interaction of the cell with its environment, required for processes such as tissue remodeling during wound-healing [17]. Not merely passive structural components, the proteins making up the ECM also directly bind to cell receptors and adhesion complexes, affecting downstream signaling pathways that govern cell behavior.

The ECM is chiefly composed of collagens, laminins, elastin, fibronectin, nidogen/entactin and several types of variably glycosylated and sulfated proteoglycans. The collagens and laminins self-assemble into dimers and trimers as well as associate laterally. The two types of structural proteins interact with each other via nidogen/entactin, providing a scaffold for the architectural support of cell layers. The three-dimensional organization of the ECM directs intercellular spacing, cell shape and cell location [18]. The superhelices adopted by collagens and laminins lend flexibility and strength to the framework. Displayed throughout this network are a variety of high and low affinity binding sites available for interaction with other ECM components and spreading cells. Heparan and heparin sulfate bind to both collagens and laminins and contain protein cores decorated with multiple and variably (10-200) sulfated saccharide moieties [19]. Themselves powerful biological effectors, this family of molecules have been implicated in such diverse processes as protein-protein interaction enhancement, concentration control, degradation resistance, and diffusion control [20]. Sulfate groups on the sugars possess large hydration spheres and are highly negatively charged, allowing little free water to pass and forming effective charge barriers [15].

Laminin-1 and Peptide 11

Laminin-1, a predominant component of the ECM, is a 950,000 kDa glycoprotein that directs the attachment, spreading, growth, differentiation and migration of cells through a variety of binding sites presented on the surface of the ECM [15]. Laminin-1 is one of seven known laminins, a structurally related family of glycoproteins composed of

three subunits, α,β,γ, that exist in different isoforms; different combinations of these subunits display disparate properties that vary with tissue type [21]. The three polypeptide chains (α1, β1, γ1) that make up laminin-1 are joined in a coiled-coil helix motif held together by disulfide bonds forming one long arm approximately 77 nm long, which separate into three shorter arms that consist of epidermal growth factor-like cysteine-rich repeats or EG modules each interrupted by a globular domain (see Figure 1) [21]. Several globular domains reside at the termini of each arm and midway along the short arms [22]. The protein contains 14-25% carboyhydrate and lies in an extended fashion forming polymers via Ca^{2+}-mediated self-assembly of the globular termini of the short arms [23].

In addition to binding to collagen and nidogen, laminin-1 contains several heparin binding sites, including one at the C-terminus at the end of the long arm [24, 25]. Like other components of the ECM, laminin directs cell behavior via interactions with cell-surface receptors. All of the laminins are believed to mediate attachment of cells chiefly through their interactions with integrins, omnipresent heterodimeric transmembrane receptors capable of both inside-out and outside-in signaling [26]. The integrins comprise a large family of adhesion molecules with different α and β transmembrane subunits, that mediate varying cell-cell and cell-ECM interactions depending on the combination of different types of subunit. These, together with the possible subunit isoforms of the laminins, generate a large number of possible interactions that allow for the exquisite specificity required for the multiple biological phenomena mediated by these molecules [21, 27].

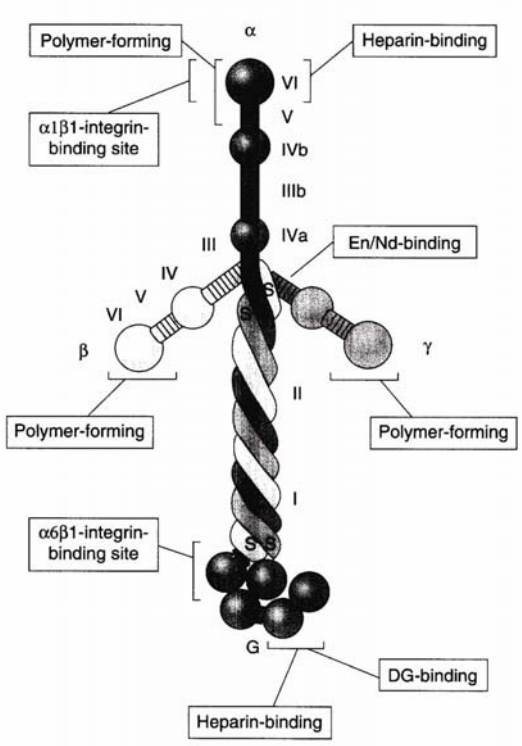


Figure 1. Functional map of laminin. Binding sites are displayed while α, β, and γ subunits are shown in different shading [21].

A synthetic nonapeptide (termed Peptide 11)—CDPGYIGSR--derived from a sequence contained in the EGF-like domain unique to one of the short arms (β1) of laminin-1 was found to stimulate the attachment and migration of cells when coated on plastic [28]. Synthetic Peptide 11 was also found to inhibit laminin-mediated attachment of human fibrosarcoma HT-1080 cells in a dose-dependent fashion, suggesting that this site in laminin-1 is chiefly responsible for adhesion of cells to laminin substrates. Later studies showed that the pentapeptide YIGSR is the minimum active sequence [29].

When combined with tumor cells prior to injection or when coinjected, peptide 11 (74%) and YIGSR (>90%) both reduced the number of lung colonies in mice while unrelated sequences from laminin-1 did not, presumably by competing with endogenous laminin for the cell-surface [30].

The Laminin Binding Protein

In the early 1980's a non-integrin protein that resides upon the cell surface and specifically binds to laminin-1 was discovered and termed the 67 kDa laminin receptor [31, 32]. The 67 kDa laminin receptor or laminin-binding protein (LBP) was initially isolated by the separation of a detergent extract of the plasma membranes of both murine melanoma cells and murine fibrosarcoma cells on a laminin affinity column [31]. LBP subsequently eluted from a laminin-Sepharose column was shown to bind ^{125}I-laminin-1 with a K_d of 2 X 10^{-9} M [31]. Detergent extracts of murine neuroblastoma cells were applied to a laminin-Sepharose column and the bound material eluted with Peptide 11. The Peptide 11 wash (1mg/mL) eluted the 67 kDa laminin binding protein, suggesting that this is the site on laminin-1 that the cell surface protein adheres to [28]. Attachment assays and *in vivo* experiments on retinal vascularization indicate that LBP may also bind to the epidermal growth factor-derived peptide (33-42)—CVIGYSGDRC—that bears homology to Peptide 11, and that this peptide inhibits LBP-mediated angiogenesis [33, 34].

Since the discovery of the laminin binding protein (LBP), a correlation between the metastatic potential and aggressiveness of a tumor and the upregulation of this protein on the cell surface has been established. As early as 1987, it was recognized that

antibodies to LBP stained the most aggressively invasive tumor cells more than surrounding normal tissues and that these antibodies inhibited laminin-mediated haptotaxis of human melanoma cells *in vitro* [35]. Likewise, tissue sections of breast, lung, prostate, colon, and ovarian carcinomas display reactivity to the LBP monoclonal antibody while adjacent normal and benign tissues did not [36-40].

The relationship between aggressive cancer and cell-surface expression does not end at the protein level. A nine-fold increase of LBP mRNA in poorly differentiated carcinomas over that of similar normal epithelium was found, suggesting a connection between the aggressiveness of colon tumors with high levels of LBP [36, 41]. Liver metastastes derived from these colon cancers contained 3 to 10 times greater mRNA levels compared to normal tissues, again demonstrating the correlation of LBP with high metastatic potential and poor prognosis [42]. Studies on pancreatic endocrine tumors show an association of LBP with high proliferative activity, large tumor diameter, and fatality [43]. Recent experiments identifying differentially expressed mRNA using microarray analysis and rapid subtractive hybridization of uveal (eye tumor) melanoma and malignant melanoma cells, respectively, pulled LBP out of a large pool of possible mRNAs, and the upregulation confirmed by independent techniques [44, 45]

A relationship between the angiogenic or neovascular capacity of cancers and LBP has also been demonstrated. Intra-tumor microvessel density, a measure of angiogenesis, correlates with LBP expression in node-negative breast cancers [46]. Lung cancer cells transfected with an LBP antisense RNA expression vector and subsequently injected into mice resulted in reduced vasculature in the tumors that formed as well as

reduced levels of VEGF, an angiogenic inducer, compared to control cells [32, 47]. Experiments that showed upregulation of LBP mRNA in malignant mesothelioma cells by microarray analysis (11.6 fold) followed by immunohistochemical validation demonstrated that at least in this particularly fatal cancer, LBP was overexpressed only in the vessels within the tumor, tying LBP to angiogenic processes [48].

While the correlative evidence linking LBP to metastasis and angiogenesis is powerful, the specifics of the functional significance of LBP in the process of tumorigenesis awaits further elucidation [21]. It has been postulated that increased levels of LBP enhance the ability of circulating cancer cells to attach to and migrate on laminin substrates, allowing them to metastasize and form vasculature more effectively. Along this vein, antibodies to LBP that prevent the migration of cancer cells on laminin-1 substrates are being studied as therapeutic targets for metastasis [49]. Other experimental evidence suggests that LBP binding to laminin-1 causes a conformational change in laminin-1 that enhances proteolysis by cathepsins that are released by surrounding cells [50]. The degradation of laminin exposes motility-promoting sites on the surface of the extracellular matrix as well as releasing bioactive fragments into the microenvironment that support tumor cell migration and growth, thus facilitating both metastasis and angiogenesis [50]. Increased levels of LBP have also been shown to correlate with increased matrix metalloproteinase-2 (MMP-2) expression and activity in stably transfected melanoma cells [51]. MMP-2 overexpression has also been linked to the tumorigenic phenotype, presumably by aiding in degradation of extracellular matrix components during tumor cell migration and invasion [51]. Finally, LBP has been

associated with multi-drug-resistant cancer cells mediated by an unkown mechanism [52]. Interestingly, a recent communication demonstrated that (-)-epigallocatechin-3-gallate or EGCG, the major polyphenol in green tea, binds to LBP with a K_d of 39.9 nM and can compete for binding to laminin, which is responsible for its antitumor activity [53].

LBP in Normal Tissues

While LBP has chiefly been studied in the context of cancer, less is known about the function that LBP serves in normal tissues. The protein has been found in lymphatic cells such as T cells, neural cells, embryonic cells, platelets, spermatogenic cells and cells involved in wound healing [54-58]. LBP is intimately associated with the basement membrane in the majority of tissues studied, epithelium or endothelium, including skin, liver, brain, muscle, intestine, stomach and lung [36]. Like other cell adhesion molecules, LBP plays a part in the morphogenesis and differentiation of the nervous system and of developing muscle, as well as in the implantation of embryos onto the basement membrane of the uterus [55, 56, 59]. In retinal mircrovascular endothelial cells, LBP is expressed predominately during rapid proliferation, not during quiescence [60].

Increasingly, it has been shown that the laminin binding protein plays a part in pathogenesis. LBP interacts with the sindbis virus, dengue virus, and tick-borne encephalitis virus, allowing the viruses both a foothold on the cell surface and a means of cell membrane penetration [61-63]. LBP also specifically binds to the cytotoxic necrotizing factor 1 (CNF1), a toxin secreted by meningitis-causing E. coli K1 that

becomes endocytosed, initiating cytoskeletal rearrangement and ultimately leading to the uptake of the pathogen [64]. Prions, infectious proteinaceous agents responsible for the development of transmissible spongiform encephalopathies such as Creutzfeldt-Jakob disease and scrapie, were also shown to interact with LBP at the cell surface [65]. The cellular prion protein, PrP^c, anchored by a glycosylphosphatidylinositol (GPI) tail to the surface, bound to LBP in a yeast 2-hybrid system. This interaction was confirmed by co-expression in insect, yeast, and mammalian cells as well as by co-immunoprecipitation [66]. Cell-surface expressed LBP also bound exogenous PrP, leading to the internalization of the complex, perhaps allowing exogenous PrP contact with endogenous PrP, demonstrating a role for LBP in the pathogenesis of prion proteins [65].

The Laminin Binding Protein and Integrin

In normal tissue, cells must attach to and detach from the extracellular matrix in a highly coordinated fashion in order to appropriately migrate, divide and differentiate during wound healing, embryogenesis, vascularization, neurogenesis, and even homeostasis [67] . The ECM exerts control over these processes by harboring binding sites for receptors and for adhesion molecules on the surface of cells. To respond to environmental stimuli, specialized macromolecular structures act as a physical bridge between the extracellular matrix and the actin cytoskeleton of the attached cell—these complexes, known as focal adhesion plaques, communicate extracellular information to the nucleus, where the appropriate genes are transcribed and the corresponding actions initiated.

In focal adhesion complexes, integrins, the most thoroughly studied family of adhesive proteins, cluster together to bind ECM components and recruit cytoplasmic adaptor proteins such as tensin, α-actinin, vinculin and talin which in turn interact with actin, coupling the cell surface to the cytoskeleton (see Figure 2) [68]. Actin then assembles into long filaments or stress fibers, triggering a myriad of downstream signals with pleiotrophic effects, causing changes in the cell cycle, further migration, proliferation or differentiation [69].

While integrins are the primary mediators of cell interaction with the ECM, the laminin binding protein has also been shown to participate in focal adhesion complexes [70]. Within these structures, LBP appears to colocalize with proteins such as vinculin, α-actinin, and $\alpha_6\beta_1$ integrin as well as indirectly with actin microfilaments [70, 71]. LBP is exported upon exposure to laminin, and shares common intracellular trafficking with integrin and the vitronectin receptor [72]. On migrating cells, significant amounts of LBP becomes detached from the cell surface and remains behind, attached to laminin on the extracellular matrix [70, 73].

The relationship between integrin and LBP does not end at the spatial proximity and functionality of the two proteins. While the specifics of the interaction between these proteins remain to be elucidated, it has been shown that $\alpha_6\beta_1$ integrin surface expression correlates with LBP expression, and together correlate with increased invasive activity in pancreatic carcinoma cells [36, 70, 74]. Within the cell cytoplasm, antibodies to the α_6

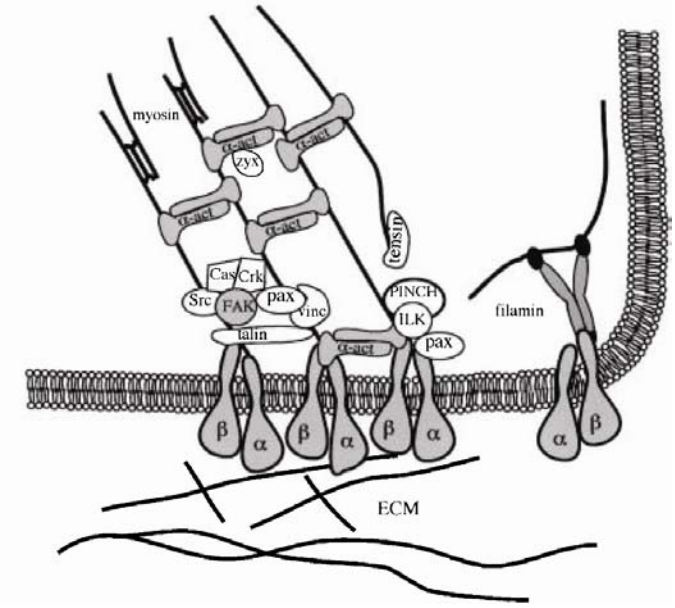


Figure 2. Focal adhesion schematic showing the interactions of integrin subunits with the extracellular matrix. The intracellular portions of integrin recruit talin, paxillin (pax), vinculin (vinc) and focal adhesion kinase (FAK), linking the ECM to actin filaments [75].

chain and the β_1 chain of integrin recognize antigens within the same structures harboring LBP and have been shown to share common intracellular trafficking pathways [76]. In epidermoid carcinoma cells, antisense oligonucleotides for the α_6 chain of integrin resulted in down-modulation of LBP levels [77]. The two proteins co-immunoprecipitate, demonstrating a direct physical association of the two molecules [77].

Not only do LBP and integrin colocalize with each other, but they also engage in synergistic binding to laminin within the ECM. By itself, integrin only displays weak affinity (micromolar) for laminin. Early experiments showed that when tumor cell detergent extracts were passed over a laminin-Sepharose column, LBP bound with high affinity. However, once this LBP was eluted from the Sepharose and reapplied under

identical conditions, most of the protein did not bind, strongly suggesting that an additional membrane component, possibly integrin, is necessary for avid laminin binding [78]. In other laminin-Sepharose affinity experiments, separation of neural cell extracts co-eluted an 120-140 kDa protein along with LBP, indicating a tertiary complex of LBP, laminin, and an unidentified protein [55]. Recently it was shown that in activated T cells, avid laminin binding requires high expression of the α_6 chain and high levels of LBP, aiding in the extravasation of these inflammation cells [79]. It appears that these two proteins act in concert, binding different regions of laminin and recruiting each other at the membrane-ECM interface [79].

Characterization of Native Laminin Binding Protein

Despite the intense interest in LBP throughout the last fifteen years, the elusive protein has defied fundamental characterization. Native LBP is extracted in only small quantities from plasma membranes using Triton X-100 or Nonidet P-40 detergents, making it difficult to obtain pure protein preparations for characterization [31, 32]. The active gene coding for LBP has been identified and consists of seven exons and six introns [80]. This DNA sequence encodes only a 33 kDa polypeptide, which is roughly half the mass of the native protein, and a single mechanism for the transformation of the gene product into the full cell-surface protein has yet to be accepted [81]. Native LBP lacks typical N-linked glycosylation consensus sequences that normally accompany carbohydrate modification. Treatment with O-glycocanase and neuraminidase do not affect the mass by SDS-PAGE, consistent with a paucity of large amounts of sugar on the

protein, so it seems unlikely that glycosylation is responsible for the molecular weight difference. While there is no evidence that LBP is glycosylated, the protein has been shown to be acylated with palmitate, stearate and oleate lipids [78].

To further define the role acylation plays in the transformation of the precursor to the mature protein, the fatty acid synthesis inhibitor and antibiotic cerulenin was added to carcinoma cell cultures that normally produce both the precursor and the 67 kDa products. A reduction in the amount of 67 kDa protein was accompanied by an increase in the amount of a 37 kDa protein observed by SDS-PAGE [82]. Furthermore, treatment of a purely 67 kDa preparation with hydroxylamine, which should release all thio and oxy ester fatty acid linkages, yielded several bands at 30, 37, and 67 kDa [82]. These experiments suggest that acylation may be important to the production of the mature cell surface protein from the precursor, perhaps aiding in targeting LBP to the membrane and in dimer formation.

Quantification of fatty acid content revealed that the amount of lipid modification cannot fully account for the dramatic weight difference between the isolated 67 kDa and the encoded 33 kDa. Since neither glycosylation nor lipidation can account for the molecular weight difference, it has been suggested that native 67 kDa LBP is either a homodimer or a heterodimer, associating with an unidentified polypeptide [78, 82, 83]. In support of the former hypothesis, transfection of hamster CHO cells with plasmid containing the 33 kDa gene results in expression of a 67 kDa protein on the cell surface, requiring no overexpression of an additional polypeptide.[84]. Furthermore, the amino acid composition of the full protein compared to that of the amino acid sequence of the

33 kDa precursor is almost identical [78]. Yeast two-hybrid studies on monomeric 33kDa LBP, on the other hand, failed to show an interaction between two monomers, proving that dimerization does not occur directly [85].

While there are two cysteine residues within the sequence, it is unknown whether or not they participate in disulfide linkages, which seems likely due to the relatively oxidizing environment outside the cell. Treatment of the 67 kDa protein with 1 mM DTT at 80º for one hour did not reduce the protein into smaller fragments by SDS-PAGE, so apparently disulfide bonds are not fully responsible for dimerization [78].

While it is widely accepted that the 67 kDa protein acts at the cell-surface, the precise mode of association with the plasma membrane remains unclear. No typical transmembrane domain (typically a stretch of 20 hydrophobic residues) can be found in the sequence of LBP, although a short hydrophobic stretch (residues 86-101) near the N-terminus has been suggested by some (see Figure 3) [86]. In support of this hypothesis, LBP mutants lacking the proposed transmembrane domain were secreted into the extracellular space in baby hamster kidney cells [65]. LBP could be a peripheral protein, but since the protein is not GPI-linked and contains no classic signal peptide sequence, the mechanism of processing and secretion is uncertain [78]. Several (T)E/DW(S) repeats, two of which are tandem repeats of TEDWSXAP, reside on the highly acidic C-terminus. This region of the protein has been shown to be available to antibodies under conditions that do not disrupt the plasma membrane, indicating that this region is extracellular [86-88].

```
  1 MSGALDVLQM KEEDVLKFLA AGTHLGGTNL DFQMEQYIYK RKSDGIYIIN 50
 51 LKRTWEKLLL AARAIVAIEN PADVSVISSR NTGQRAVLKF AAATGATPIA 100
101 GRFTPGTFTN QIQAAFREPR LLVVTDPRAD HQPLTEASYV NLPTIALCNT 150
151 DSPLRYVDIA IPCNNKGAHS VGLMWWMLAR EVRRMRGTIS REHPWEVMPD 200
201 LYFYRDPEEI EKEEQAAAEK AVTKEEFQGE WTAPAPEFTA AQPEVADWSE 250
251 GVQVPSVPIQ QFPTEDWSAQ PATEDWSAAP TAQATEWVGA TTEWS      295
```

Figure 3. Amino acid sequence of the laminin binding protein. Laminin-1 binding sites Peptide G (161-180), helix (200-295), and the (T)E/DW(S) repeats are underlined.

Ribosomal Function of LBP

Although LBP has chiefly been studied from the perspective of its interactions with the extracellular matrix, additional roles for the protein have recently emerged. The N-terminus of the sequence, roughly residues 1-200, displays significant homology to the S2 family of ribosomal proteins of prokaryotes (see Figure 4 A.) [80, 89-91]. A yeast homologue of LBP has been isolated and found to be an essential component of the small ribosomal subunit and to be involved in 20s rRNA processing [90, 92]. A 40 kDa protein isolated from rat 40S ribosomes was sequenced and found to be 99% homologous to LBP, conclusively demonstrating that vertebrate LBP is associated with ribosomal complexes and is therefore a bifunctional molecule [93]. These data explain the observation that significant amounts of LBP antigen is detected within the cytoplasm and in complex with ribosomes [91, 94].

In a yeast two-hybrid screen of a HeLa cDNA library, a truncated version (residues 1-207) of LBP bound ribosomal protein S21, a protein localized to the "protuberance" region of the 40S ribosome [95]. S21 is part of the 40S ribosomal complex and thought to be involved in the initiation of translation. Deletion mutants of

LBP showed that residues 17-194 are required for this interaction, corresponding to a large binding surface that almost exactly corresponds to the region of N-terminal homology to prokaryotic S2 proteins [95]. The role LBP plays in the protein synthesis machinery has led some to suggest that the link between cancer aggressiveness and LBP overexpression might be mediated by its ability to effect the synthesis of growth regulatory proteins [90]. Supporting a role for the overexpression of ribosomal proteins in cancer, microarray techniques that analyzed differences between normal and urothelial tumor cells found that 44% of upregulated RNA sequences identified were proteins involved in translation [96].

The crystal structure of the small ribosomal subunit of *Thermus thermophilus* has allowed comparative modeling of the N-terminus of LBP (residues 13 to 188 of LBP from *Homo sapiens*) based on its homology to the S2p proteins [89]. The predicted fold is formed of 5 parallel beta strands sandwiched between 4 helices (see Figure 4 B.). The putative transmembrane domain (residues 86-101) is part of the hydrophobic core and is distributed between a helix and a beta strand, leading Kazmin et al. to conclude that the deletion mutant lacking this domain secreted into the extracellular space was probably misfolded [65, 89]. The model lacks a two helix extension that serves to make extensive contact with 16S RNA and S3p in the 30S ribosome of *Thermus thermophilus*.
The C-terminus of LBP (200-295), in stark contrast to the N-terminus, appears to have evolved coincident with laminin expression. Highly conserved, diverging only 1.7% in mammals, the C-terminus in its present form has been preserved since the divergence of mammals from birds 300 million years ago [80]. Because of the

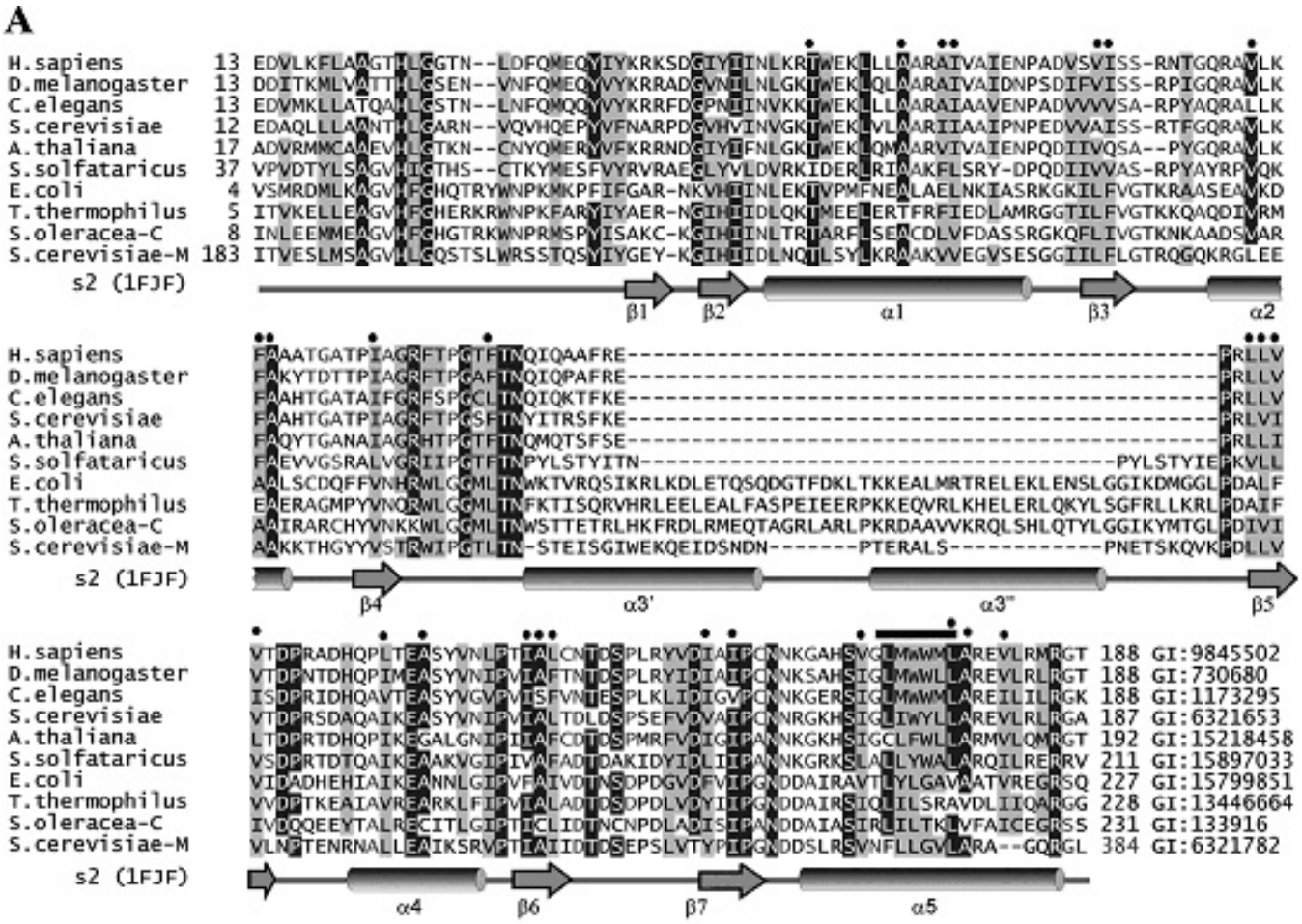


Figure 4 A. Sequence similarity between the N-terminal region of LBP to prokaryotic ribosomal proteins. Black letters On light gray background represent conserved hydrophobic amino acids. Conserved amino acids with the frequency greater than 80% are shown in white on black background. Gray rectangles represent α-helices while gray arrows represent β-strands.

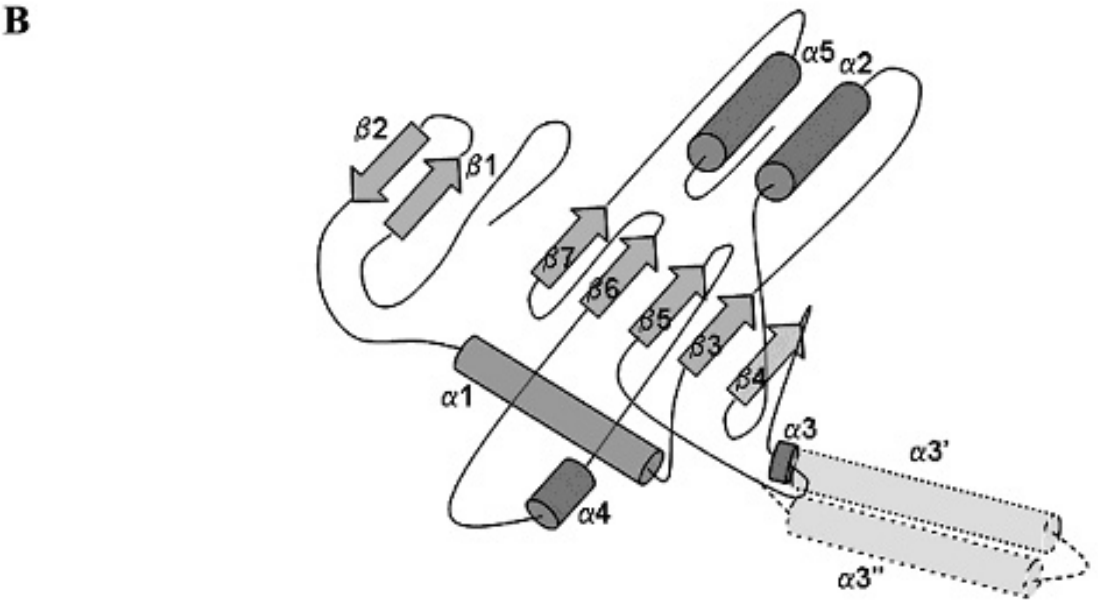


Figure 4 B. Schematic representation of the modeled LBP structure. The dashed lines and pale gray color denoting the unmodeled two-helix domain [89].

association of this region of LBP with its function in eukaryotes, it is believed that the structural determinants responsible for dimerization of the protein reside on this region as well. Finally, this high sequence similarity suggests steady evolutionary pressure to conserve this part of the molecule.

The Laminin Binding Domain of LBP

The emergence of the C-terminal region of LBP coincident with the appearance of laminin strongly implicates that this region is involved in laminin binding. In support of this conclusion, function-blocking antibodies raised to the 67 kDa LBP initially localized the laminin-1 binding site to residues 161-295 [88]. To further define the residues

involved in laminin binding, antibodies were raised to synthetic peptides corresponding to residues 263-282 and 165-184, which were predicted to exhibit a beta-turn motif associated with receptor-ligand binding domains. A haptotaxis assay was performed in which antisera to the first domain, but not the second, blocked migration of cancer cells by competing for cell surface LBP. This peptide, referred to as Peptide G (residues 161-180), containing the palindrome LMWWML, bound to laminin substrates with high affinity ($K_d = 5 \times 10^{-8}$M) and eluted native LBP from a laminin-affinity column [86].

Subsequent studies on Peptide G showed that the eicosapeptide interacts with heparin sulfate, which also binds to laminin-1, suggesting instead a divalent interaction between LBP and the ECM [97]. Unfortunately, because heparin sulfate is abundantly bound to the laminin extracted from Engelbreth-Holm-Swarm (EHS) tumors, the most common source of laminin-1, the effects of these two interactions have not been dissected out [97]. Adding to the confusion, heparin sulfate proteoglycans act as attachment sites for bacteria and viruses including alphaviruses like the sindbis and dengue viruses [98]. However, because LBP itself binds heparin sulfate, it is not clear if these viruses bind LBP alone, heparin sulfate alone, or the two in concert [61, 62, 85]. The Peptide G region (147-179) has also been implicated as the binding site for prion protein, which also interacts with heparin sulfate albeit versus a distinct site of the prion protein [66, 85].

A highly charged stretch of amino acids encompassing residues 205-229 of LBP initially attracted attention because it appears at the beginning of exon 6, associated with eukaryotic species and predicted to adopt an amphipathic alpha-helix [80]. Studies on a synthetic peptide corresponding to this sequence demonstrated interaction with both

laminin-1 and heparin sulfate in plate assays [84, 99]. Furthermore, when B16BL6 melanoma cells were coinjected with this peptide, metastasis was reduced by 40% relative to cells alone [84]. When the cells were incubated with the peptide prior to injection, however, lung colonization was enhanced by 150%, hinting at the complexity of the LBP-cancer link.

Phage display studies screening for peptide 11-responsive sequences confirmed the identification of the two identified laminin-1 binding sites while indicating a third site within the TE/DWS repeating motif of LBP [99]. These phage display results suggest that a relatively large surface of the protein, roughly residues 161-295, may be exposed to and interact with the extracellular matrix at multiple foci.

Hypothesis

Despite preliminary reports that structural data on diffraction-quality crystals of LBP was forthcoming, no structural information on this cell-surface molecule is currently available [63, 89]. It has been demonstrated that bacterially expressed recombinant LBP, which lacks posttranslational modification and is present as in the 33 kDa form, is able to bind laminin on a laminin-Sepharose affinity column and in a blot assay [100, 101]. These results suggest that recombinant LBP may be expressed in its active form, and that interactions between laminin and LBP may take place via protein-protein determinants, so that structural studies of the bacterial product will give us insight into the mechanisms of LBP activity [63]. To this aim, we cloned, purified, and over-expressed full length LBP (295 amino acids) in an *Escherichia coli* expression system and subjected the

construct to biophysical characterization. Early one dimensional proton Nuclear Magnetic Resonance (NMR) experiments on this construct revealed that the full protein was aggregated at concentrations needed for structure determination (0.5-1.0 mM). In an effort to overcome this difficulty, domains representing independent structural elements of monomeric LBP were identified by limited proteolysis. The resulting protein fragments were identified by N-terminal sequencing and mass spectrometry and subsequently subcloned, expressed, and characterized by NMR methods when applicable. To assay for native structure, the ability of the recombinant protein and isolated domains to bind to laminin-1 was evaluated by enzyme-linked immuno-absorbent assay (ELISA) and by laminin-1-Sepharose affinity chromatography. Information obtained from limited proteolysis experiments may aid in the isolation of a biologically relevant laminin-binding domain that will be amenable to Nuclear Magnetic Resonance methods, in an effort to better understand tumor cell-extracellular matrix interactions and prion infectivity.

MATERIALS AND METHODS FOR LBP

Cloning of Recombinant LBP in *E. Coli*

A cDNA isolated from hamster CHO cells encoding the 33 kDa LBP (kindly provided by Dr. James Strauss in a pcDNA I/neo vector (Invitrogen, Inc.)) was subcloned into the pET-15b vector (Novagen, Inc.) and the sequence subsequently verified by DNA sequencing (Iowa State DNA sequencing facility). The construct was designed to include a cleavable N-terminal polyhistidine tag for purification using nickel affinity methods. Plasmids were propagated in Top Ten F' *E. Coli* cells (Invitrogen, Inc.), and transformed into BL21(DE3)-C41 (Cambridge, U.K.) for protein expression.

Expression and Purification of Bacterial Recombinant LBP

BL21(DE3)-C41 cells transformed with the plasmid described above were grown to an OD_{600nm} of 0.60 (using the 635 Varian spectrometer)in 500 mL Luria-Bertani medium at 37°C under ampicillin selection, then induced with 1mM isopropylthiogalactoside. Cells were harvested by centrifugation after 6 hours of growth at 37°C.

Cell pellets were thawed on ice and resuspended in ice-cold binding buffer (50mM sodium phosphate/400mM NaCl/10mM imidazole/10mM 2-mercaptoethanol/1% CHAPS) at pH 8.0. Cells were lysed using a French Press pressure cell and the insoluble portion pelleted by centrifugation at 10,000g for 30 min at 4°C. The supernatant was combined with 1 mL of a 50% Ni-NTA slurry (Qiagen Inc.), stirred at 4°C for 1 hour,

and poured into a siliconized 1x15 cm column. Unbound proteins were washed off with lysis buffer and the protein eluted with 5-10 mL elution buffer (50mM sodium phosphate/400mM NaCl/200mM imidazole/10 mM 2-mercaptoethanol/1% CHAPS, pH = 8.0) into siliconized tubes. Fractions were analyzed by SDS-PAGE. Protein-containing fractions were further purified on a calibrated Superdex 200 gel filtration column equilibrated in PBS (pH=7.5). Any aggregated material was removed at this step. Correct molecular weight was verified by Matrix-Assisted Laser Desorption Ionization time-of-flight mass spectrometry or MALDI-TOF MS (Biflex II, Bruker Inc., Montana State University) using dihydroxy benzoic acid/50:50; acetonitrile:water for crystallization.

Limited Proteolysis and Identification of Domains

Bacterial recombinant LBP was subjected to limited proteolysis using the enzymes endoproteinase Glu-C (V8), bromelain, chymotrypsin, elastase, proteinase K, subtilisin, thermolysin, and trypsin (Sigma Inc.) following protocols described in Carey [102]. To gauge the appropriate dilution of stock enzymes for thorough digestion under the specific sample conditions ("range-finding" experiment), 30 μg of LBP was treated with 15 μL of a 10-fold dilution of a given stock enzyme (at ~1mg/mL), on ice. The digestions were incubated on ice for 30 minutes, and then analyzed by SDS-PAGE.

Once the activity of each enzyme under given sample conditions was evaluated, a timecourse over a 48 hour time period was conducted on ice to establish the presence of fragments that persisted for a longer period of time than for the parent protein, indicating greater stability. For the timecourse experiment, 50 μL of a ~0.005-.01 mg/mL of each

enzyme was used to digest 50 μg LBP, and a fraction of the digest analyzed at 0.5, 1, 2, 5, 10, 24, and 48 hours by SDS-PAGE.

Larger amounts of the desired fragments from the trypsin and the endoproteinase C digests were produced using the conditions established above. Protease cleavages were stopped by acidification with 1/10 of AcOH stock (25%) and, for trypsin, by adding 10mM PMSF, and then stored at -80°C. Fragments were then separated on a C18 reverse phase HPLC column using an H_2O-ACN/ 0.1%TFA solvent system, monitored by absorbance at 230 nm and fluorescence, (excitation at 340nm, emission at 280 nm). Fragment-containing fractions were lyophilized, confirmed by SDS-PAGE, and identified by matrix-assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry (Biflex II, Bruker Inc., Montana State University) and N-terminal amino acid sequencing (Protein Sciences Facility at the University of Illinois).

Cloning and Expression of Proteolysis-derived Domains

LBP fragments identified by limited proteolysis were subcloned using the full-length plasmid as template. The regions encompassing residues 137 to 230, residues 1 to 220, residues 1 to 136, and residues 200 to 295 were subcloned separately into pET-15b vectors using Xho I and BamH I restriction sites, and are referred to as 137-230, 1-220, 1-136, and 200-295, respectively. Segment 200-295 was cloned rather than the 221-295 fragment identified from limited proteolysis (see Table 1) to include the 205-229 sequence previously reported to bind laminin-1. All fragment identities were confirmed by DNA sequencing (Iowa State DNA sequencing facility).

E. Coli cells transfected with plasmids 200-295, 1-136, 1-220, and 137-230 were cultured and induced under conditions similar to the ones described for expression of the full-length protein, except that the cells expressing 200-295 and 137-230 were co-transformed with a plasmid containing the genes for GroEL and GroES under chloramphenicol selection [103]. Protein purification was carried out in a similar manner as for the full-length protein, except that the detergent CHAPS was not required during the isolation and purification of the 200-295 protein construct. Expression of 1-136 produced a highly insoluble product, precipitating at low concentrations within hours after purification, and so was not used in subsequent experiments. The expression of all constructs was confirmed by MALDI-MS (Biflex II, Bruker Inc., Montana State University). With the exception of 200-295, cleavage of the his_6-tag with thrombin reduced the solubility of the fragments.

ELISA Assay and Laminin-1-Sepharose Affinity Chromatography

The wells of a 96-well microtiter plate (Immulon 2, Dynatech Inc.) were incubated with 100 μL of a 100 nM laminin-1 solution (EHS/Sigma Inc.) in PBS overnight at 4°C. Wells were washed three times with DPBS/.1% Tween, then incubated with 5% dry milk in DPBS/.1% Tween for two hours at room temperature. The wells were subsequently washed twice with DPBS/.1% Tween, then once with DPBS, followed by incubation with 100μL of full-length LBP or proteolysis-derived domains at concentrations of 0, 20, 50, 100, 200 and 500 nM for two hours at room temperature. An unrelated polyhis-tagged protein was used as a negative control. The wells were again

washed with DPBS, and then treated with anti-polyHis antibody-HRP conjugate (Sigma Inc.) in PBS/0.05% Tween for thirty minutes. The wells were then washed four times and developed with HRP substrate (ABTS/hydrogen peroxide, Bio-Rad) according to the manufacturer's instructions. Plates were read at 415 nm after 20 minutes. Interaction of the anti-polyhis-HRP antibody with laminin-1 was negligible. Attempts to perform the assay by fragment inhibition of full-length LBP binding to laminin-1 were complicated by the tendency of these proteins to self-associate.

Laminin-1 (Sigma Inc.) was conjugated to CNBr-Activated Sepharose following the manufacturer's instructions. After pre-elution with DPBS /2M NaCl, the laminin-1-Sepharose slurry was equilibrated in DPBS, allowed to interact with LBP, and then washed with DPBS to elute any unbound protein. The bound LBP protein was eluted from the laminin-1-Sepharose using SDS and analyzed by SDS-PAGE. The polyhis-tagged control protein was not retained by the laminin-1 Sepharose.

CD and NMR Spectroscopy

1D ^{1}H and 2D ^{1}H-^{15}N correlation HSQC spectra of domain 200-295 was recorded at 298 K in PBS at a pH of 5.0 at Montana State University using a Bruker DRX600 four-channel NMR spectrometer. 2D ^{1}H-^{15}N HSQC spectra were typically acquired in States-TPPI mode, with 512 t_1 increments and 2048 data acquisition points, and a relaxation recovery delay of 1.2 sec, with spectral widths of 12 ppm for the ^{1}H acquisition dimension, and 30 ppm for the ^{15}N indirect dimension [18,19]. Spectra were processed using Xwinnmr, version 2.6 (Bruker, Inc.). All spectra were zero-filled once in both

dimensions, and sine bell apodization functions phase shifted by 45-90^{o} were applied in both dimensions prior to Fourier transformation. Circular dichroism spectra were acquired from 260 to 190 nm on a JASCO J-710 spectropolarimeter on samples of roughly 0.1mg/mL in a .1mm pathlength cell.

RESULTS FOR LBP

Expression and Purification of Recombinant LBP

Expression of full-length LBP in the pET-15b vector as described resulted in a protein that migrated at an apparent molecular weight of 43 kDa by SDS-PAGE, consistent with earlier observations that the protein travels anomalously (see Figure 5) [63]. The MALDI-TOF mass spectrum of the full protein fused to the polyhis tag confirms the correct molecular weight of the expression product (see Figure 6). A Western blot utilizing antibody raised to a synthetic peptide corresponding to residues 205-229 of the LBP sequence also confirmed the identity of the protein (data not shown). After nickel affinity purification, the protein was approximately 95% pure by SDS-PAGE (Figure 5). In initial purifications performed without the detergent CHAPS, the protein formed a fibrous precipitate within 24 hours that could not be resolubilized, even at low concentrations (< .1 mg/mL). CHAPS was included in the initial solubilization and in the following elution steps because it prevented precipitate formation, improved solubilization yields and enhanced separation of non-specifically bound proteins. CHAPS or 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate, a sulfobetaine-type detergent, is nondenaturing up to concentrations of 10mM, is electrically neutral, and disrupts protein-protein interactions, and as such is the ideal detergent for protein structural work [104]. Other detergents such as octyl glucoside, Triton-X 100 and NP-40 were also used, but did not disaggregate as well as CHAPS and interfered with analyses

such as protein concentration determination by absorbance at 280 nm and NMR spectroscopy.

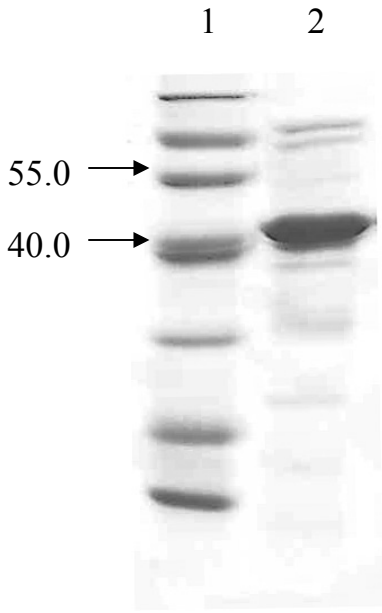


Figure 5. SDS-PAGE of full length recombinant LBP (15% polyacrylamide). Molecular weight marker with masses in kilodaltons as shown in lane 1, recombinant LBP in lane 2, migrating at roughly 43 kDa.

Gel filtration was performed as a second purification step in order to remove aggregated material and isolating the monodisperse species. The fractions eluted from the nickel column and then applied to the Superdex 200 gel filtration column contained two major species; roughly 35 to 50% high molecular weight aggregate (>400,000) and a second large peak corresponding to a molecular weight of roughly 68,600 Da; SDS-PAGE on these fractions demonstrated that both peaks contain LBP. While it is possible that bacterial product forms a dimer in solution, the presence of CHAPS, which disrupts protein-protein interactions, and the finding of only the monomeric species in other recombinant LBP systems [85], strongly suggests that in this case, either the monomer is

merely traveling differently than the globular protein standards, a phenomenon which has been observed for elongated and random coil proteins, or the protein is associated with a large number of detergent molecules [105]. Native gel electrophoresis of the soluble recombinant protein (Figure 7) illustrates the aggregation observed in the size exclusion experiments—there are several species present, including a lower molecular weight band corresponding to the monomer at roughly 35,000 kDa and several species of higher molecular weight aggregate, including some that do not penetrate the stacking gel.

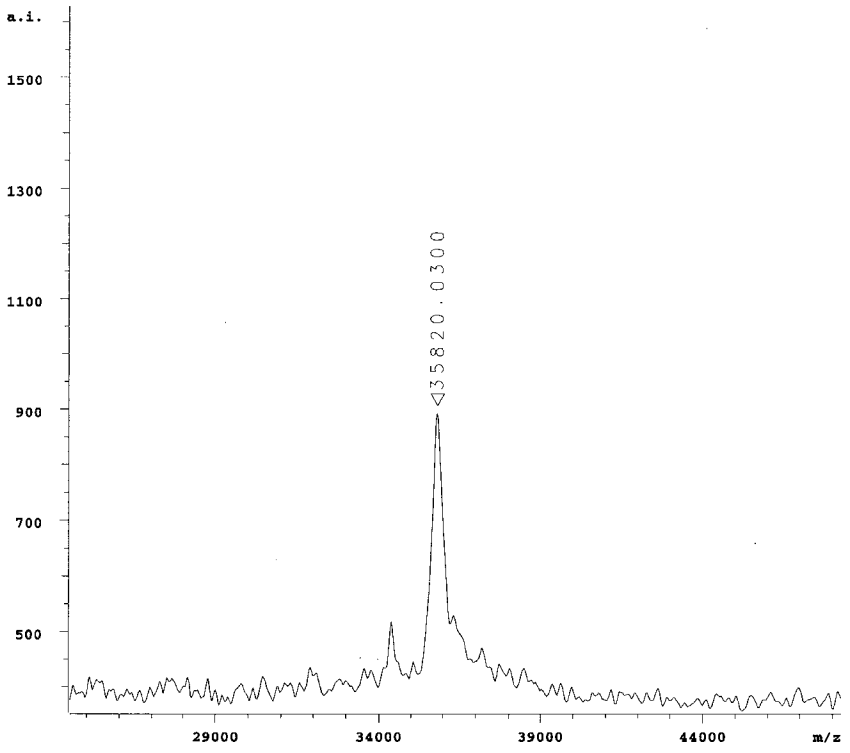


Figure 6. MALDI-TOF MS of full length LBP. The predicted mass from the pET-15b expression product is 35,176 Da, while the measured mass is 35,820 Da, within 2% error of measured mass.

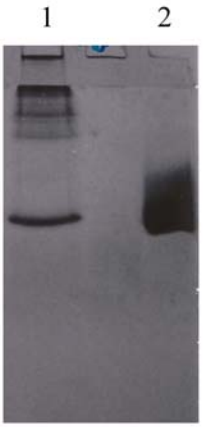


Figure 7. Native 15% polyacrylamide gel of recombinant LBP. LBP is shown in lane 1 and molecular weight standard with a mass of 37 kDa is shown in lane 2.

One dimensional proton NMR spectroscopy of the full protein in the presence of 1mM DTT both with and without the detergent CHAPS showed a highly aggregated protein, such that all resonances were broadened beyond detection. Aggregation increases the effective size and the relaxation behavior of a molecule [106]. The linewidth of an NMR resonance is proportional to the size and shape of the protein as well as the overall viscosity of a sample, creating a limit to the size of the molecule that can be studied by current NMR technology; apparently aggregated LBP falls outside of this limit.

To test whether or not additional cofactors or additives are required for deaggregation or protein folding, a variety of small molecules were added to NMR samples. Equimolar amounts of Peptide 11 and heparin sulfate, known binding partners of LBP, were added but had no effect on the one dimensional spectrum. Hydrophobic additives, sometimes added to samples in order to mimic a membranous environment,

were also added without effect; these included ethanol, butanol, detergents, and trifluoroethanol. Changes in pH, ionic strength, and cation (Ca^{2+} and Mg^{2+}) concentration exhibited little influence on the spectral properties of LBP.

Biological Activity of Recombinant LBP

In order to ensure that the full-length recombinant protein as isolated contains relevant structural determinants for binding for laminin, the ability of the monomeric LBP to bind to laminin-1 was assessed by ELISA (Enzyme-Linked Immuno-Sorbent Assay) assay. ELISA plate assays results are shown in Figure 8, and demonstrate that the expressed full-length LBP binds to EHS tumor cell-isolated laminin in a dose-dependent fashion. These data were confirmed by laminin-Sepharose affinity, which showed that laminin-1 immobilized on Sepharose beads retained LBP under the conditions used for our structural work (see Figure 9).

Limited Proteolysis of Recombinant LBP

In an effort to find biologically active domains suitable for NMR structure determination, limited proteolysis was performed on the monodisperse preparation of the full-length protein. Range-finding experiments were conducted in order to establish the enzyme concentration that results in limited digestion so that a timecourse could be performed (see Figure 10). In the timecourse experiments, digestion of LBP with all of the proteases revealed consistent reproduceable digest patterns, delineating a protein fragment core of approximately 20 to 33 kDa in molecular weight, and several lighter,

poorly staining (9-14 kDa) bands when visualized by Coomassie Blue/SDS-PAGE. Larger scale digestions were performed with the enzymes V8 and trypsin because these enzymes cleave at specific amino acids, E/D and K/R, respectively. Digestion of LBP by trypsin resulted in the appearance of one major band at an apparent molecular weight of 23 kDa and 4 smaller, but reproduceable, bands varying from 12.5 to 10.5 kDa (Figure 11 A.). In the V8 digest, a heavy band at 27 kDa and a band at 13 kDa were produced (Figure 11 B.).

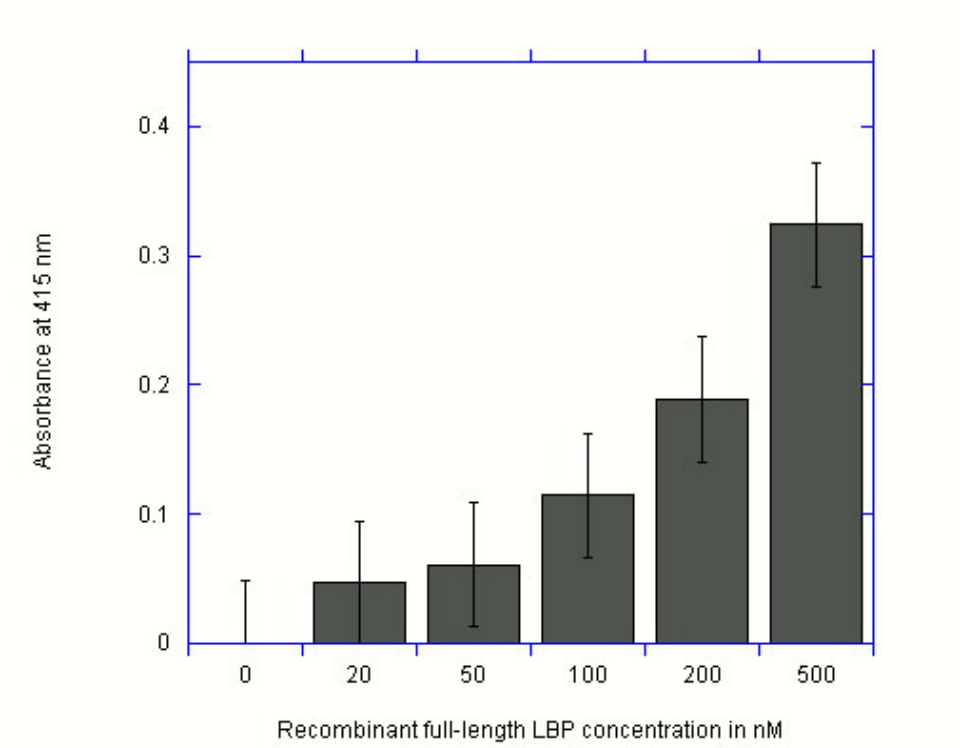


Figure 8. ELISA assay of full-length recombinant LBP. Increasing absorbance demonstrates dose-dependent binding of recombinant LBP to laminin-1 from EHS extracts.

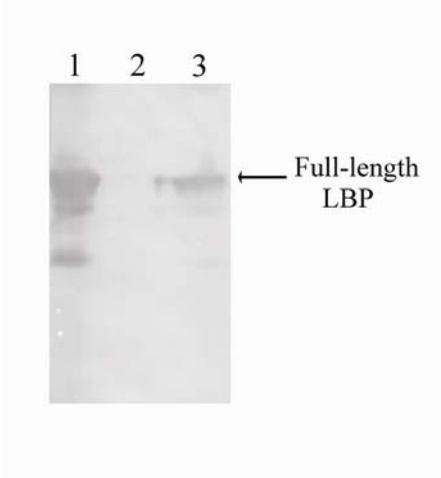


Figure 9. Blot of fractions from laminin-Sepharose chromatography. Proteins applied to a 12% polyacrylamide gel were blotted to nitrocellulose and probed with anti-polyhis-HRP antibody. Lane 1 shows the LBP sample prior to loading onto the laminin-Sepharose column, lane 2 shows the unbound proteins in the flow-through, and lane 3 corresponds to the fraction eluted with SDS, demonstrating that LBP is retained by laminin-1.

N-terminal sequencing and MALDI-TOF data of the fragments (see Table 1 and Figure 12) along with primary sequence information yielded the identities of the domains. Trypsin separated LBP into two major domains by cleavage at lysine residue 220, liberating a small 75 residue C-terminal region from the rest of the protein. V8 cleaved at two major sites, one after glutamic acid residue 136, and one after glutamic acid residue 244. V8 proteolysis generated three major domains, one N-terminal domain encompassing the very hydrophobic portion of the molecule postulated to be the transmembrane region, one domain consisting of both the peptide G region and the 205-229 sequence, and another extreme C-terminal domain containing the TE/DWS repeat (see Figure 13).

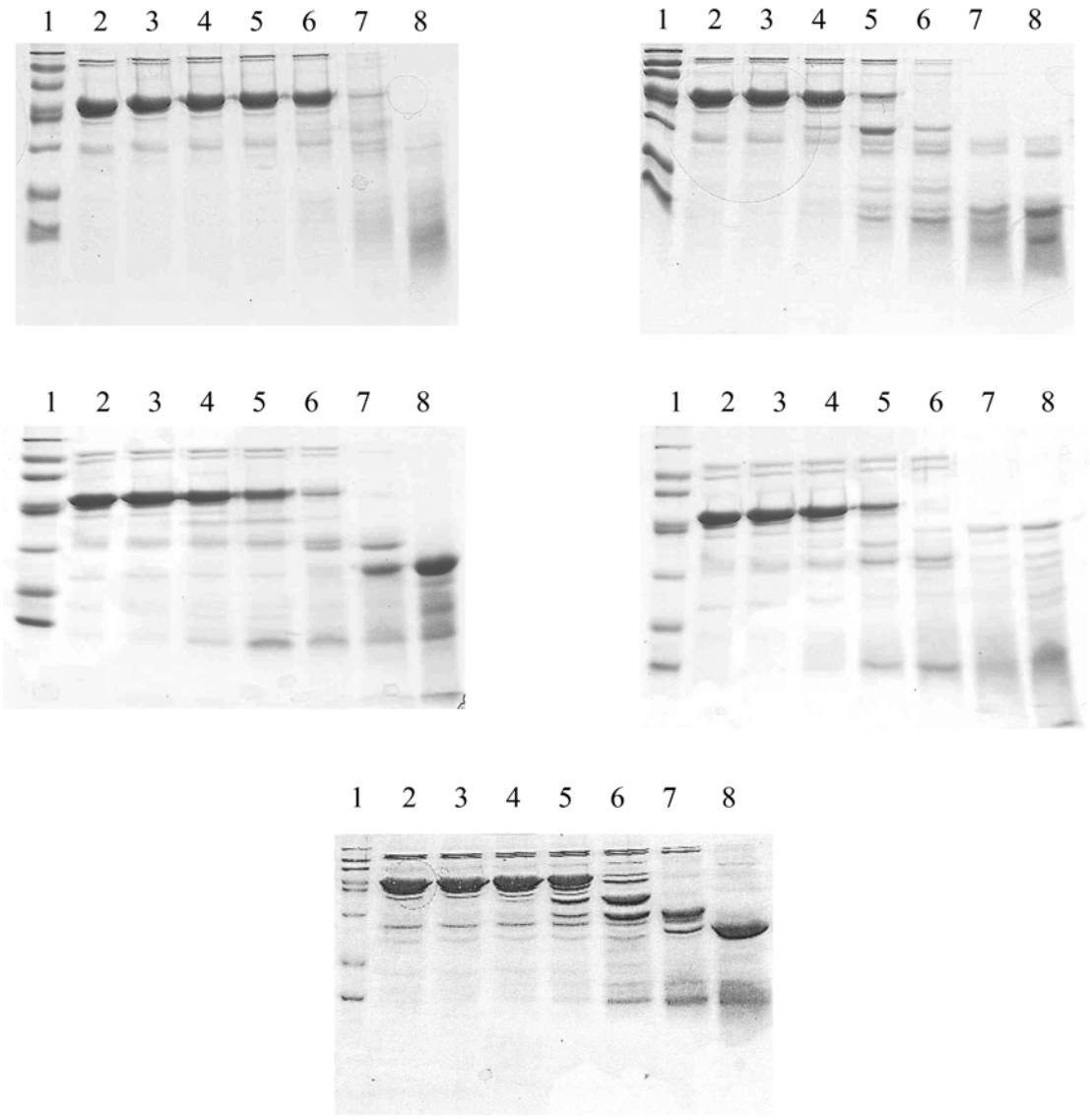


Figure 10. Range-finding results. SDS-PAGE gel analysis depicts the digestion that occurs with different enzyme concentrations. In each gel, lane 1 corresponds to the molecular weight marker with masses, starting top to bottom, of 95.0, 66.2, 55.0, 40.0, 31.0, 21.5, and 14.4 kilodaltons. Lane 2-8 show successive ten-fold dilutions of stock enzyme (see Materials and Methods). Enzymes used as follows: Top left: bromelain; Top right: Chymotrypsin; Middle left: trypsin; Middle right: thermolysin; Bottom: elastase.

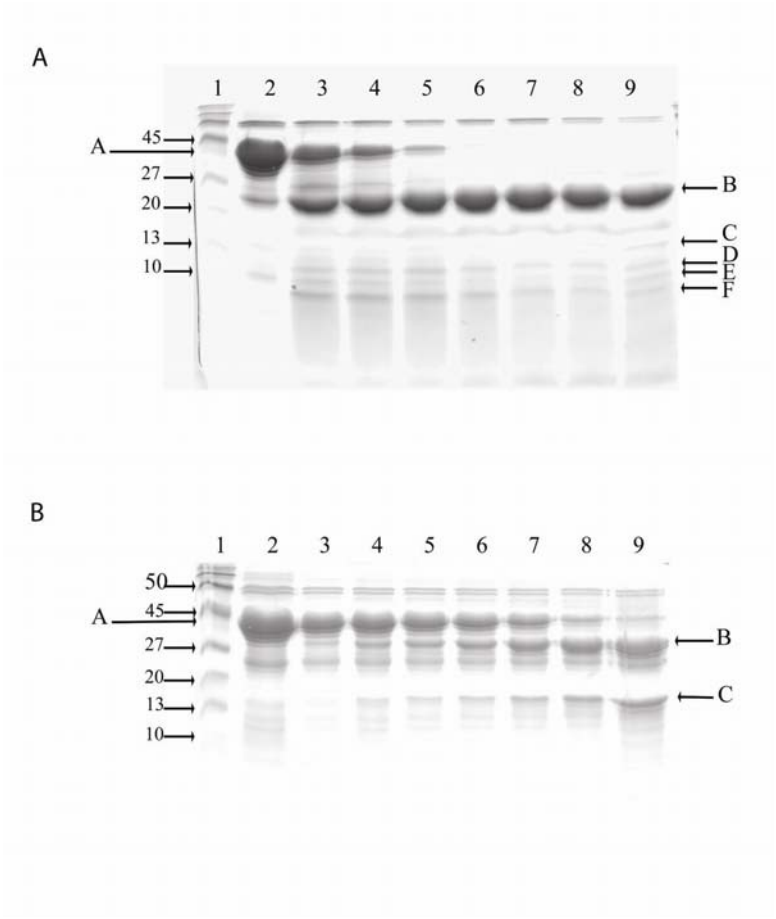


Figure 11. Limited proteolysis of recombinant LBP as visualized by SDS-PAGE. Recombinant LBP was subjected to limited proteolysis by the enzymes trypsin (Figure 11 A.) and endoproteinase Glu-C or V8 (11 B.). 50 µg LBP was treated with 50 µL .01mg/mL trypsin stock and 50 µL .008 mg/mL endoproteinase Glu-C stock separately and incubated on ice. A fraction of each digest was removed and boiled in Laemmli Sample Buffer for three minutes to stop digestion; aliquots were taken at .5 hour (lane 3), 1 hour (lane 4), 2 hour (lane 5), 5 hours (lane 6), 10 hours (lane 7), 24 hours (lane 8), and 48 hours (lane 9) and compared against a molecular weight marker (lane 1) and undigested LBP (lane 2) on SDS-PAGE stained with Coomassie Blue. In the trypsin digest (Figure 11. A.), band A corresponds to undigested LBP. Band B, C, D, E, and F are the fragments produced by digestion; band B was identified by N-terminal sequencing and MALDI-TOF MS to correspond to residues -7-220, band C corresponds to residues 221-295. In the endoproteinase Glu-C digest (Figure 11. B.), band A is undigested LBP, fragment B was identified as residues –22-136, and fragment C corresponds to residues 137-230.

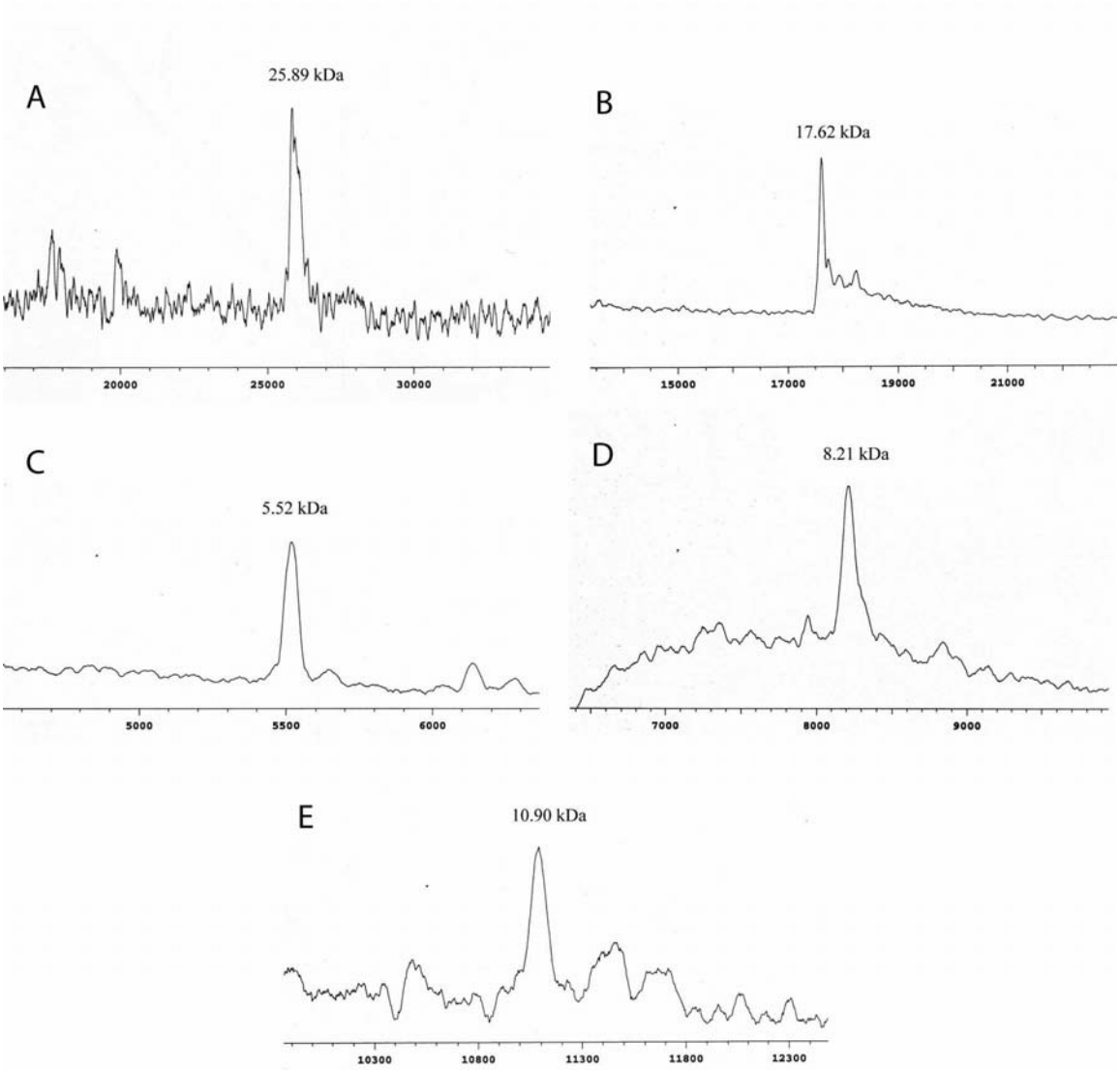


Figure 12. MALDI-TOF MS data collected on proteolysis fragments. Figure 12 A. corresponds to fragment –7-220; 12 B. to fragment –22-136; 12 C. to fragment 244-295; 12 D., to fragment 221-295, and 12 E. to fragment 137-230. Fragment identities are shown in Table 1 below.

Residue Number/ Enzyme	N-terminal Sequencing Results	Measured/Calculated Mass (kilodaltons)
A. -7-220/ trypsin	GSHML	25.89/ 25.60
B. 221-295/ trypsin	AVTKE	8.21/ 8.16
C. -22-136/ V-8	GS*X*HH	17.62/ 17.59
D. 137-230/ V-8	ASYVN	11.09/ 10.90
E. 245-295/ V-8	V*X*D	5.52/ 5.55

Table 1. Fragments of LBP retrieved from limited proteolysis experiments. Fragments were identified by N-terminal sequencing and mass spectrometry. *X* indicates amino acids that could not be unambiguously assigned in the N-terminal sequencing data.

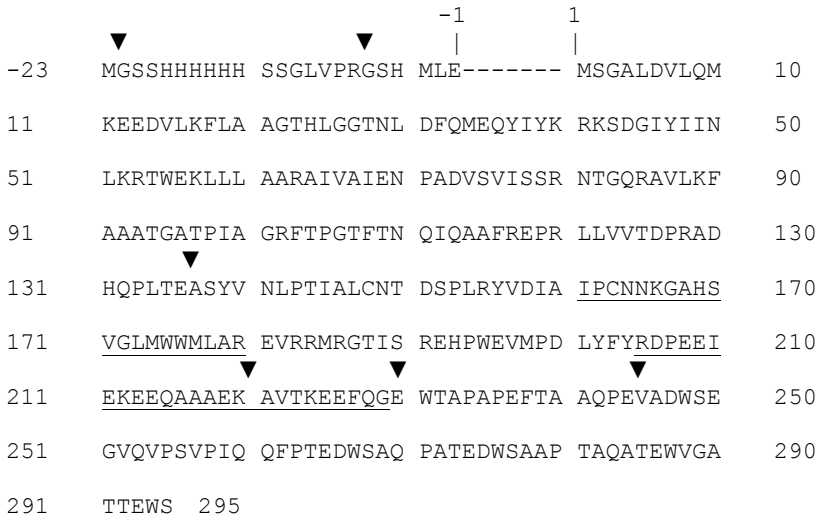


Figure 13. The amino acid sequence of LBP from hamster (97% sequence identity with human LBP). The negative numbers (-1 to –23) correspond to the extra 23 residues that have been added to the LBP amino acid sequence as a result of the cloning strategy used to produce recombinant LBP in the pET15b expression vector (see Materials and Methods). This numbering scheme permits direct comparisons with LBP residue stretches previously identified as being important for biological laminin-1 binding activity. Arrows indicate residues cleaved by trypsin and endoproteinase Glu-C/V8. The region of the sequence corresponding to peptide G is underlined (residues 161-180) as well as residues 205-229, both regions implicated in laminin binding activity.

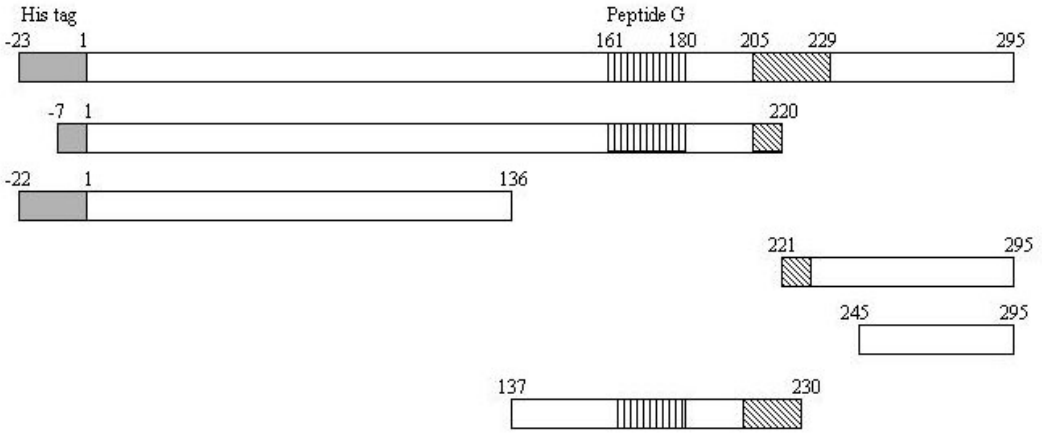


Figure 14. Schematic representing the isolated domains. Domains are shown relative to the entire sequence, to the fusion tag, and to the laminin-1 binding sites Peptide G (161-180) shown in vertical stripes and the proposed amphipathic helix (205-229) shown cross-hatched.

Cloning and Expression of Limited Proteolysis-Derived Domains

To evaluate the physiological functions of the three LBP domains identified by limited proteolysis, cDNA encoding residues 1-220, 1-136, and 137-230 were subcloned into pET-15b for protein expression. A cDNA encoding amino acids 200-295 was also cloned into pET-15b in an attempt to distinguish the contributions of residues 205-229 to the binding of LBP to laminin-1 from that of the Peptide G region. Large scale protein expression was achieved for all constructs made (see Figure 15 A., Figure 15 B.). However, co-expression of the chaperonins GroEL and GroES was required for expression of LBP 200-295 protein fragment. The 200-295 construct migrated at an anomalous molecular weight of roughly 26 kDa (Figure 15 A.), almost twice that of its predicted mass. MALDI-TOF MS confirms the correct molecular weight, suggesting that this is the region of LBP responsible for the anomalous migration of the full protein.

A.



B.

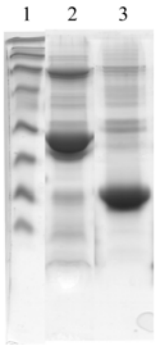


Figure 15. A. SDS-PAGE of 200-295. Expression of fragment 200-295 (lane 2) resulted in an expression product that migrated at an apparent molecular weight of 26 kDa by SDS-PAGE, perhaps due to its highly acidic composition. Figure 15. B. SDS-PAGE of 1-220 and 137-230. Lane 2 corresponds to the expressed domain 1-220 migrating at 26 kDa and domain 137-230 in lane 3 migrating at 14 kDa by SDS-PAGE analysis by comparison with the molecular weight marker in lane 1. These migrations are in good agreement with the predicted molecular weights of 27,272 and 13,437 Da, respectively. In both A. and B., lane 1, the molecular weights of the marker proteins, from top to bottom, are as follows: 184, 123, 85, 52, 41, 27, 21, 14, and 11 kDa.

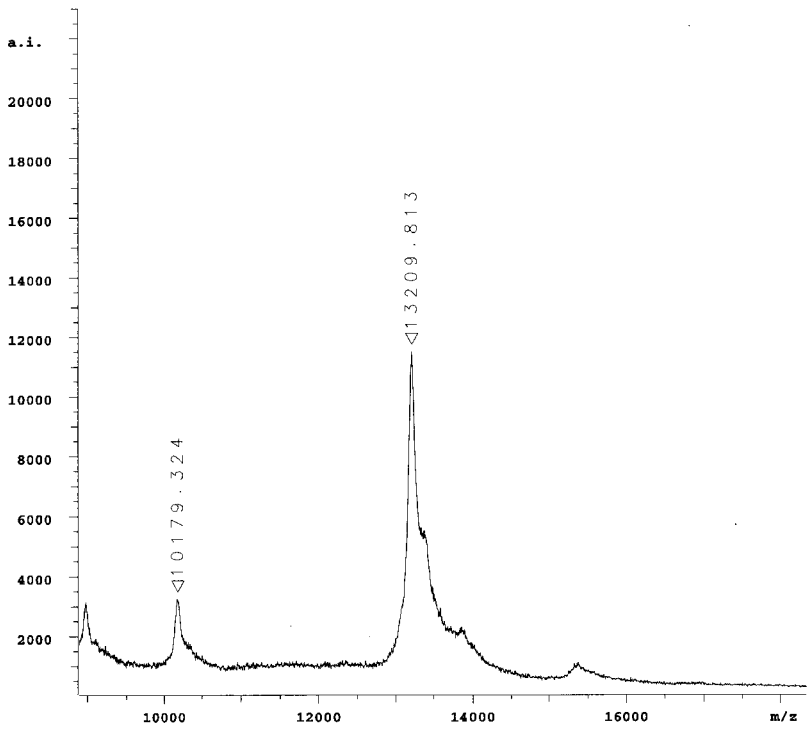


Figure 16. MALDI-TOF MS of fragment 200-295. The observed mass is within 200 mass units of the expected molecular weight of 12,901 Da.

While LBP domains 1-220, 1-136 and 137-230 exhibited severe aggregation at the millimolar (mM) concentrations required for NMR spectroscopy, the smaller C-terminal 200-295 construct was sufficiently soluble to allow for preliminary 1D ^{1}H and 2D ^{1}H-^{15}N-correlation NMR analyses.

Biological Activity Screen of LBP Domains

In order to fully isolate the domain(s) involved in LBP binding to laminin-1, ELISA assays were performed on the proteolysis-derived expression construct, except for the 1-136 protein construct which was insoluble under the ELISA assay conditions

(Figure 17). Surprisingly, the 137-230 protein appears to be the most active domain, despite the fact that the 1-220 construct only differs by 10 amino acids. However, caution should be exerted when interpreting this result, because the 137-230 protein product is significantly more hydrophilic and soluble than the 1-220 construct, which is barely soluble under the assay conditions.

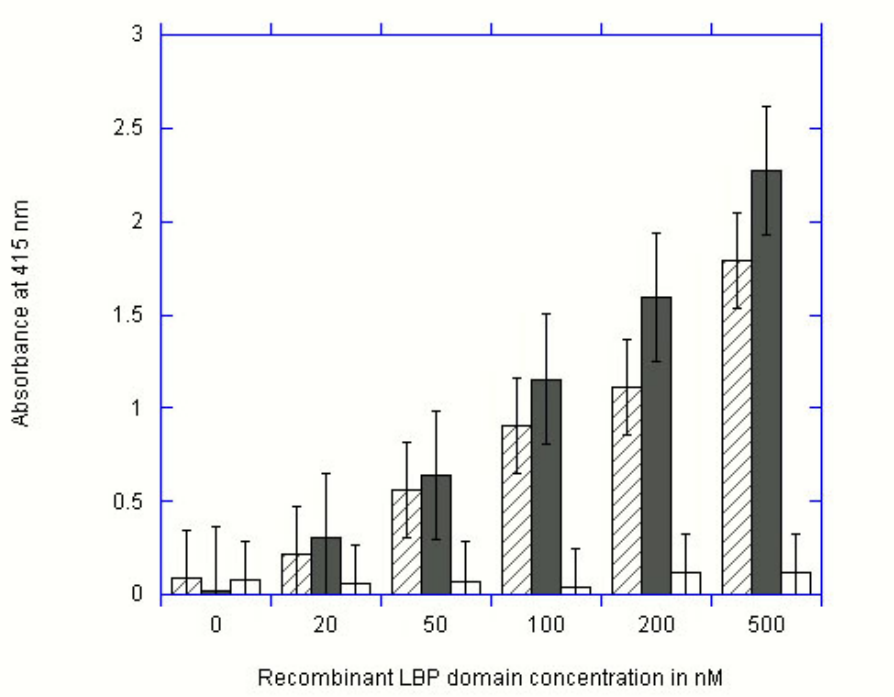


Figure 17. ELISA assay of proteolysis-derived domains. The binding of the proteolysis-derived domains 1-220 (cross-hatched bars), 137-230 (solid bars), and 200-295 (empty bars) to laminin-1 were assayed by ELISA. Microtiter plates (Immulon 2) were treated with 100 µL of 100nM laminin-1 from EHS (Sigma), blocked, and then incubated with different dilutions of each construct ranging from 0 to 500 nM. The amount of protein bound was determined by anti-polyhis-HRP antibody development of substrate, measured by absorbance at 415 nm. The control consisted of wells incubated with DPBS/Tween without laminin-1.

The 137-230 protein, which contains the Peptide G sequence implicated in LBP binding, is roughly as active as the 1-220 domain, indicating this region is all that is required for laminin-1 binding. Surprisingly, the 200-295 expression product exhibits very little activity, indicating that residues 200-230 may be involved in heparin sulfate or integrin binding rather than interaction with laminin-1.

CD and NMR Spectroscopy of 200-295

Circular dichroism (CD) spectroscopy was performed on construct 200-295 at a concentration of .06 mM (4.6mg/mL) in PBS (pH=7.0) in a .1 mm path-length cell to check for the presence of secondary structure elements (Figure 18). Like measurements were made at the same concentration of protein but in the presence of 25% trifluoroethanol (TFE), 50% TFE and finally 90% TFE. The spectrum without TFE demonstrates a random coil conformation, with increasing amounts of helicity with increasing TFE concentration up to 90%, as evidenced by the shifting minima towards 208 nm and an increase in the shoulder at about 225 nm [107].

The 2D ^{1}H-^{15}N HSQC spectrum of the 200-295 protein product is shown in Figure 19, displaying roughly 80 sharp NMR resonances but lacking the characteristic ^{1}H chemical shift dispersion of a well-folded protein. As can be seen in this figure, the majority of the $^{1}H/^{15}N$ amide resonances are clustered within the ~ 1 ppm ^{1}H chemical shift range typically observed in the NMR of poorly structured proteins [108]. Of the six tryptophan residues in the sequence, only two or three of the tryptophan indole $^{1}H/^{15}N$ NMR signals could be distinguished. The poorly dispersed signals and resonance overlap

are further evidence of the random coil character of the 200-295 fragment of LBP in solution. ^{15}N-NOESY experiments conducted on the protein showed few crosspeaks (data not shown), signals which correspond to magnetization transfer between two atoms close in space (5-6Å), further confirming the lack of three dimensional structure.

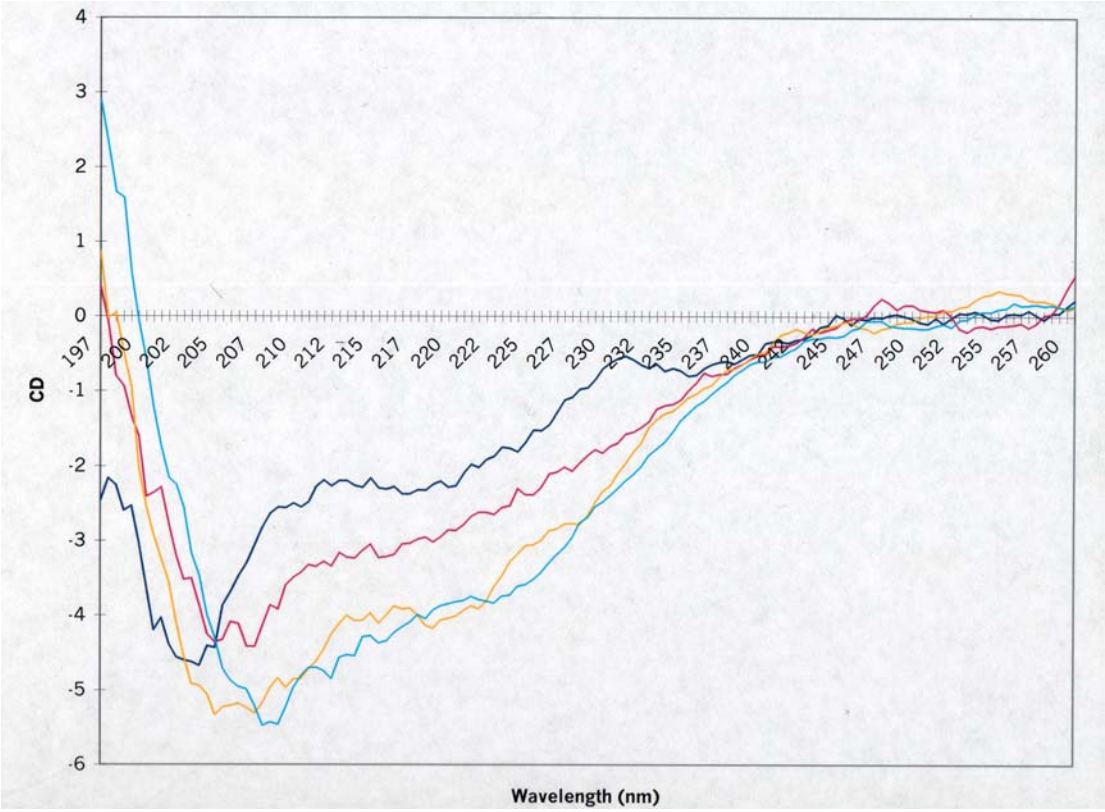


Figure 18. CD spectrum collected on 200-295. Data was collected from wavelengths of 265 nm to 198 nm at a protein concentration of 4.6mg/ml in PBS, along with the effect of increasing TFE concentrations. The curve in dark blue is 0% TFE, 25% TFE in red, 50% TFE in yellow and 90% TFE in light blue.

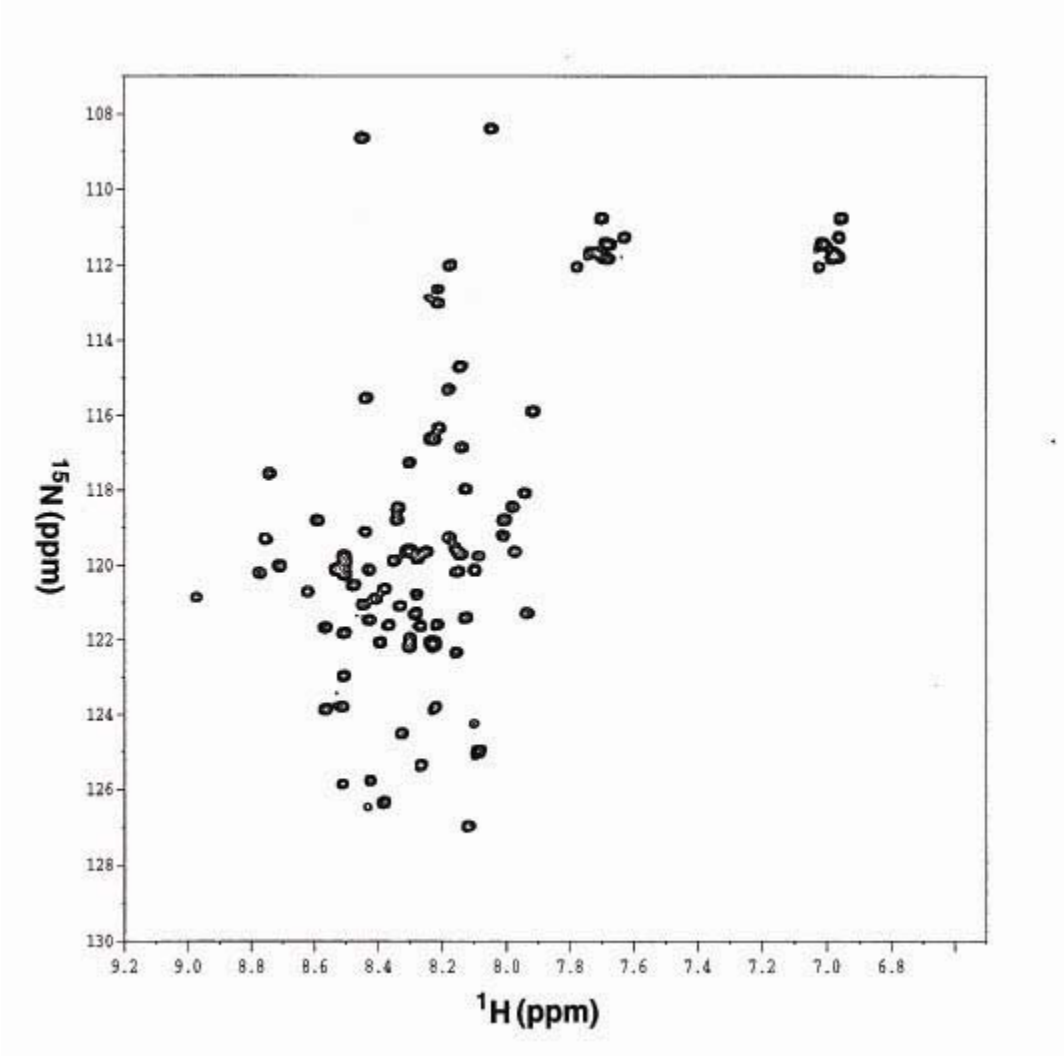


Figure 19. ^{1}H-^{15}N HSQC NMR spectrum of construct 200-295. The 2D ^{1}H-^{15}N HSQC was recorded on a Bruker DRX 600MHz spectrometer at 300K in PBS at a pH of 5.0 and concentration of 1 mM. Note the narrow dispersion of the proton chemical shifts indicative of a poorly structured protein fragment.

DISCUSSION ON LBP

Despite the fact that a suitable domain of LBP that bound laminin for structure determination was not isolated, some fundamental characterization of the recombinant bacterial protein product and *in vitro* function of the domains of LBP, both necessary steps toward structural insight into the functional aspects of LBP, was gained. Unfortunately, the full recombinant protein suffered from severe aggregation as measured by gel filtration and as visualized by native gel electrophoresis, increasing with time and concentration. The aggregation was observed by the formation of thread-like insoluble precipitates and, when an otherwise monodisperse preparation was concentrated, often formed a viscous gel. In addition, NMR experiments performed on the expressed full-length product under the conditions described produced no discernable resononances, indicating an apparent molecular weight far exceeding that of monomeric or even dimeric protein.

An explosion in the literature on aggregation has recently emerged, driven by the interest in insoluble plaque formations displayed by diseases such as spongiform encephalopathies, Alzheimer's disease, and Huntington's disease, just to name a few [109]. While the fibrils and "aggresomes" characteristic of these disorders are built upon specific interactions and therefore assume an ordered assembly, the nucleation event that spawns these larger complexes begins with an amorphous aggregate or micelle [110, 111]. Study of the initiation of these phenomena, therefore, provides insight into more general aggregation processes. Aggregation is essentially caused by the incorrect interactions of protein molecules, and arise from misfolding, unfolding, or exposure to

non-native conditions (i.e. extremes in pH, temperature, concentration, etc.) [109]. The latter is observed in membrane proteins, in which the chiefly hydrophobic residues in contact with the plasma membrane in their native environment require the use of nonpolar detergents or solvents for solubilization. In misfolded or unfolded proteins, hydrophobic residues are sequestered from the aqueous, polar environment via interaction with similar regions on another unfolded molecule, forming an aggregate [111]. Both of these processes are driven by the same thermodynamic forces of entropy and enthalpy [112].

Because LBP is a cell-surface protein with an unresolved relationship with the plasma membrane, it is a distinct possibility that the protein self-associates in order to bury hydrophobic residues that might interact with the plasma membrane or lipid moieties present in the native state. Alternatively, LBP may display a hydrophobic surface that forms a protein-protein interface, which at high concentrations forms nonspecific interactions such as that observed in the ribosomal protein eIF4E, that requires partial immersion within a CHAPS micelles for structure determination [113, 114]. To address these possibilities, LBP was solubilized by a number of detergents as well as glycerol and nonpolar solvents, and while detergents like CHAPS aided extraction of soluble protein and allowed significantly higher concentrations, no hydrophobic additives tested completely abolished aggregation. Proteolysis-derived domains 1-220, 1-136, and 137-230 all aggregated as well and required the use of detergent, raising the possibility that hydrophobic regions within these sequences may be involved in self-association.

Another potential cause of aggregation is misfolding. Heterologous expression often leads to misfolded proteins which are shuttled into inclusion bodies as amorphous complexes of proteinaceous and membranous material [115]. The chaperones, protein disulfide isomerases, peptidyl-prolyl isomerases and complement of cofactors that assist protein folding in the endoplasmic reticulum (ER) of eukaryotes are not present or differ in bacteria, which may lead to improper folding of mammalian proteins [112, 115]. Alternatively, many proteins are folded within the ribosome aided by chaperones, and so fold co-translationally [109, 116].

Exported to the cell-surface and modified by lipids, LBP undergoes much post-translational processing that appears to be required for dimerization, suggesting that exposure to ER components does occur during native expression [85]. A significant amount of LBP was found in inclusion bodies, detected as insoluble proteins after cell lysis that could not be refolded. However, the high degree of sequence homology of LBP to bacterial S2 proteins (roughly 25-30% identity) suggests that the cytoplasmic environment used for expression should allow proper folding, at least of the N-terminal two-thirds of the protein. It seems more likely that in this case inclusion body formation was due to the very high concentrations of LBP produced (60 mg LBP /L culture). Because disorder-mediated aggregation generally leads to functional inactivation [109], activity assays were performed on the recombinant product to demonstrate proper folding. Not only did these experiments establish binding to laminin-1 substrates, limited proteolysis of the recombinant product confirmed the presence of stable, protease-resistant domains, arguing against any misfolding.

In addition to improper folding, another complication in heterologous expression systems is disulfide formation. The bacterial cytoplasmic environment is reducing, and so proteins expressed in bacteria may lack the proper disulfide bonds, important structural determinants that aid in stabilizing the tertiary structure of proteins in extracellular environments [117]. As mentioned above, LBP has two cysteine residues that are most likely involved in disulfides, although whether these participate in intermolecular or intramolecular bonds is unknown. Comparison of migration rates on SDS-PAGE with and without reductant show that the cysteine residues in the LBP expression product are present as free thiols that required reduction in order to prevent intermolecular disulfide bond formation. While disulfide bonds may be important structurally, both cysteines are not strictly conserved among laminin-1 expressing organisms.

Several expression strategies were employed to address the proper disulfide formation. LBP was cloned into a *Pichia pastoris* over-expression system (pPICZαA, Invitrogen, Inc.), so that the protein would have the advantages of protein processing, protein folding and posttranslational modification accompanying eukaryotic expression that might be required by the C-terminal one-third of LBP and disulfide linkage. Unfortunately, no protein expression was observed in this system by SDS-PAGE or Western blot, possibly because LBP is highly homologous to the yeast LBP/p40 protein, and therefore may be subject to regulation by the host. In another attempt to allow correct disulfide formation, the LBP gene in the pET-15b construct was transformed into a commercially available cell line, AD494 (Novagen, Inc.), that harbors a mutation in the

thioredoxin reductase gene leading to a less reducing environment in the cytoplasm, enhancing disulfide formation. Expression levels in AD494 cells, however, were exceedingly low and the cultures grew poorly, preventing efficient isolation of purified protein.

The bacterial expressions described in the literature identified the cellular location of LBP as inclusion bodies which were either denatured by urea and subsequently dialyzed or resolubilized in 0.5% Triton X-100, a powerful detergent [63, 95, 100]. None of these sources mentioned performing size exclusion removal of potentially aggregated material. In our hands, refolding of denatured LBP by dilution or dialysis, even in the presence of detergent, resulted in a completely aggregated product by native gel and gel filtration. Resolubilization in Triton X-100 was not as effective as CHAPS, and left a significant portion of the total protein aggregated. Based on our observations of refolded and resolubilized LBP, as well as the requirement of reducing conditions to prevent intermolecular disulfide formation, it is likely that proteins used for experiments consisted of mostly aggregated material. Since aggregated proteins act as nucleation sites that recruit other proteins via inappropriately exposed hydrophobic regions [109], which might conceivably result in nonspecific protein-protein complexes, these observations may explain some of the conflicting results reported in the literature on LBP [100] and the myriad of new binding partners found [63, 64, 118].

All of the recombinant eukaryotic expressions in which LBP was isolated utilized Triton X-100 as a solubilizing agent. One of these described the over-expression of LBP as a C-terminal fusion with the small FLAG tag in baby hamster kidney cells, first

solubilized in 0.1% Triton X-100 and then analyzed by size exclusion. In these experiments the protein traveled as a monomer without aggregate [85]. These data suggest that aggregate formation observed in our preparations is associated with bacterial expression, again, possibly due to the high concentrations of LBP produced. The protein isolated from eukaryotic sources was not concentrated to the high concentrations required for structural work and whether or not eukaryotic expression leads to aggregated product remains unresolved.

In order to ensure that the full-length recombinant protein was folded properly, and therefore contained relevant structural determinants for laminin-1 binding, ELISA assays for detection of a complex were conducted. ELISA plate assays demonstrate that the expressed protein binds to EHS tumor cell-isolated laminin-1 in a dose dependent fashion. These data were confirmed by laminin-1 Sepharose affinity chromatography, and, like previous studies on bacterial recombinant LBP, indicate that integrins and are not strictly required for LBP binding.

Limited proteolysis was performed on the full length construct in an effort to identify an active, structurally accessible domain for NMR studies. Limited proteolysis has long been used as a powerful technique for probing solvent/surface exposure, functional domain structure, ligand-binding, and ligand-induced conformational changes in proteins [119]. Limited proteolysis is based on the observation, by comparison of cleavage sites with crystal structures, that digestion does not occur in the protein core or in regular secondary structural elements as readily as in structurally accessible regions such as loops, linkers, disordered regions and domain boundaries [119]. These

observations have been rationalized by the study of crystallographic data on protease/inhibitor complexes that demonstrate that local unfolding must occur at the cleavage site in order for the polypeptide backbone to be positioned properly in the binding pocket. Trypsin, for example, has been estimated to require the unfolding of roughly 10 residues and the accompanying polypeptide deformation for efficient cleavage. Secondary structure elements that contain significant hydrogen bond content and residues buried within the protein core make cleavage at these sites less energetically favorable than cleavage in those solvent accessible, flexible and weakly packed regions. There are three caveats to domain identification by limited proteolysis methods. First, surface loops within one domain may be cleaved, leading to an incorrect identification of a domain boundary [119]. Second, some domains are not formed by contiguous sequences, so the fragments produced do not correspond to a single domain. Finally, initial cleavage of a region crucial to the tertiary fold can lead to an unfolding or "falling apart" of molecule, which is then cleaved in ways unrelated to domain structure [119].

The small, 22 amino acid N-terminal fusion tag was cleaved twice in the limited proteolysis experiments conducted, consistent with the expected solvent accessibility accompanying the tag's ability to bind to the nickel affinity resin used for purification. Limited proteolytic cleavage occurred in 3 different sites within the C-terminus, digesting within a relatively large span of 24 residues. This suggests a lack of secondary structure and a solvent accessibility within these regions, indicative of a disordered C-terminus. One of these cleavage sites (Lys 220) falls within the helix predicted between residues 205 to 229. While cleavage at helices is more favorable than cleavage at β-strands by

virtue of the relatively more flexible conformation of the helix, only very little empirical evidence exists for ascribing helicity to this sequence (one CD spectrum not shown [99]). In fact, the NMR ^{1}H-^{15}N HSQC (see Figure 19)and ^{1}H-^{15}N NOESY data shows that the entire region between 200 and 295 is disordered, confirming these proteolysis results.

Limited proteolysis by V8 led to the identification of a domain consisting of residues 1 to 136 and a domain encompassing residues 137 to 230. The cleavage site at residue Glu 136 resides in helix 4 of the homology-modeled structure and does not appear, from the model, to be the most accessible cleavage site to V8 (see Figure 4 B.) [89]. Asp118, appearing on a long loop between beta strand 4 and 5 that replaces the two unmodeled helices in the S2p structure seems to be a much more likely candidate for cleavage [89]. Comparative modeling can be a powerful approach for predicting the three-dimensional fold of a protein when experimental data is lacking. In the approach used in modeling LBP, the template sequence from *Thermus thermophilus* was manually aligned with the sequence of LBP from *Homo sapiens* (residues 13 to 188) based on secondary structure prediction. The aligned protein was then modeled using the satisfaction of spatial restraints method. Sequence alignment, the most crucial and error-prone step in modeling, becomes difficult when there is less than 30% sequence identity between the template and target proteins [120]. In the study conducted by Kazmin et al., there is only 22% identity between the *H. sapiens* and *T. thermophilus* sequences. Misalignment is very common between sequences with less than 25% identity and errors arising from misalignment cannot be overcome by present refinement methods [120, 121].

One major region of difficulty in the model occurs mid-sequence, in which two long helices and two loops contained in the S2p structure are completely lacking in the LBP sequence (see Figure 4 A., 4 B.). Segments that differ by as few as 9 residues between target and template are extremely difficult to model and may have unrelated conformations [120]. Therefore, it is unlikely that regions modeled past this large gap in the model are aligned correctly, which would include the cleavage site at 136. Complicating analysis of the model, the S2p structure is bound within the context of the ribosome, while the construct we studied was free in solution. Differences between the ribosome-bound and free conformations are likely to be great, since the tightly interlaced complex of RNA and other ribosomal proteins are often required for the structural stabilization of ribosomal proteins [122]. Finally, the fitness of the LBP model to function within the eukaryotic ribosome was only briefly tackled and may not be possible at this point. Model evaluation is a critical stage in the modeling process, and currently remains unaddressed [120].

ELISA data obtained on the subcloned domains 1-220, 1-136 and 200-295 demonstrated that recombinant 1-220 and 1-136 constructs bind laminin-1, while 200-295 does not. 1-220 and 1-136 have roughly the same activity within experimental error, suggesting that, at least under the conditions used in this assay, regions encompassing amino acids Glu 136-Lys 220 mediate LBP binding to laminin-1. The consistently slightly lower average binding observed for domain 1-220 may be due to the poorer solubility and higher aggregation potential of this construct. These data are consistent with earlier results showing that the Peptide G region, amino acids Ile 161-Arg 180 are

biologically active, either by binding directly to laminin-1 or indirectly via bound heparin sulfate.

Surprisingly, construct 200-295 does not exhibit laminin-1 binding as would be suggested by studies on synthetic peptides containing the 205-229 sequence and phage display experiments that indicated that phage homologous to the 205-229 sequence can be eluted from laminin-1 by both peptide 11 and heparin sulfate [84, 99]. In the first experiment on peptide 205-229, the peptide was applied to a positively charged plate which was blocked, overlaid with tritiated laminin-1, washed and then subsequently treated with SDS to remove bound proteins and scintillation counting done to quantitate bound laminin-1. In the experiments we performed on laminin-1 it was found that this glycoprotein has a large affinity for plastic even in the presence of detergent, and, since, unfortunately, control data are not shown for the plate assay, we cannot account for this effect [84]. In an effort to demonstrate that the 205-229 sequence interacts with laminin-1-derived Peptide 11, 205-229 peptide bound to Sepharose was incubated with laminin-1 both with and without Peptide 11. Data from this experiment showed that Peptide 11 inhibited roughly 50% of laminin-1 binding, however, an unrelated sequence of similar charge inhibited binding by 15%, leading to questions as to how specific this interaction is [84]. Further studies that attempted to demonstrate Peptide 11 binding to Peptide 205-229 used a cross-linkable form of Peptide 11. Strangely, concentrations of 12.5 ug of Peptide 205-229 exhibited maximal Peptide 11 binding, while higher concentrations showed decreased binding down to near control levels [99]. This result was explained by aggregation of the peptide, which was apparently observed in an NMR spectrum (data not

shown). The study then goes on to demonstrate binding between this peptide and immobilized heparin sulfate, which increased in a dose-dependent manner, leveling off at 4 ug heparin sulfate/100ug of 205-229 peptide [99].

Based on these data, a model for the interaction of peptide 205-229 and the ECM was proposed in which this peptide adopts a helical conformation, forming a hydrophobic pocket that interacts with Peptide 11 on one side with a heparin sulfate binding domain on the "reverse" side [99]. The data on peptide 205-229 and the proposed model contrast with our results on both the structural content and the biological activity of construct 200-295, which encompasses this sequence and 69 additional residues. CD data and NMR data acquired on this construct was devoid of structure including α-helicity. Unlike peptide 205-229, construct 200-295 was highly soluble, not shown to aggregate under NMR conditions. These observations suggest that the behavior of the peptide 205-229 differs markedly from that sequence when in the context of the 200-295 construct, both in physicochemical characteristics and biological activity. It is conceivable that the disorder observed in the larger construct leads to greater flexibility which might effectively "unwind" or destabilize any latent helical propensity.

NMR is a technique exquisitely sensitive to structure. The dispersion of the amide resonances in an ^{15}N-HSQC spectrum is a reflection of the different magnetic environment an atom is surrounded in by virtue of the stable, three-dimensional arrangement about the atom. Dispersions less than 1 ppm indicate that the amide protons are all roughly within the same, averaged environment, exposed to solvent rather than a folded proteinaceous matrix [108]. The ^{15}N-NOESY experiment detects the transfer of

magnetization from one proton close in space (≤5 Å) to another, a close proximity made possible by the conformation of the protein. Construct 200-295 exhibited too few cross-peaks in the NOESY spectrum to account for three-dimensional structure. CD data on 200-295 supports this lack of structure, although addition of TFE does increase the helicity of this region, hinting at the helical propensity calculated by secondary structure prediction programs but not displayed in this construct. TFE also increased the dispersion of the resonances by NMR, but only by about 0.2 ppm, and, since TFE has been shown to stabilize non-native structure, was not pursued further.

While construct 200-295 contains many E, Q, P and K residues, disorder-promoting residues, it also exhibits several (6) W and V residues, known for promoting order [123]. There are three positively charged residues and 19 negatively charged residues, a dramatic charge imbalance that would cause electrostatic repulsion between residues on the polypeptide chain and destabilization of structure within this region [123, 124]. Limited proteolysis results also support the dynamic nature of the protein; three cleavage sites were identified within a short span of 24 residues. It should also be noted that polypeptides smaller than these were not selected during the separation of proteolytic fragments because large domain information was sought.

Interestingly, expression of the 200-295 construct was unobserved unless co-expressed with GroEL and GroES, chaperones that aid in protein folding [112]. Small, disordered polypeptides are often destroyed during overexpression, targeted by the housekeeping machinery for degradation. GroEL and GroES are known to bind unfolded proteins less than 60kDa in size [112]. While the protein was not folded by the

chaperones, it is conceivable that the chaperones bound the protein, effectively sequestering the protein from the proteolytic machinery so that it could be isolated before degradation.

Disorder and flexibility have been recognized as characteristics of functionally relevant native states for an increasing number of proteins. Inherent disorder is believed to play a role in molecular recognition/ligand binding,"linker" regions associated with domain movement and flexibility, regulation via proteolysis, dissociation from binding partners, as well as in entropic springs. Some "natively unfolded" proteins display lack of secondary and tertiary structure in regions up to 600 residues long [123]. Many proteins, including many ribosomal proteins, remain in an unfolded state until they bind to a target molecule, perhaps aiding in achieving maximal complementarity and regulation of complex formation [122, 124]. In addition, unfolded states may play a role in facilitating membrane passage, remaining denatured within the cytoplasm until membrane approach [124].

In an effort to induce structure in the C-terminal domain of LBP, equimolar amounts of Peptide 11 and heparin sulfate were added to the NMR sample in separate experiments and combined, but no changes in the spectrum were observed, not even a broadening of resonance which might indicate some interaction with these molecules, supporting the lack of laminin-1 binding observed in the ELISA data. Perhaps the peptide experiments demonstrating binding to these ligands involved a "freezing out" or selection of one of the ensemble of conformations adopted by this region that does not occur in the larger context of the protein due to thermal motion [123]. Because the C-terminus of

LBP is associated with laminin-1 binding abilities, and therefore might also be responsible for targeting to the membrane, the possibility that this construct may interact with membrane components was tested. Dodecylphosphocholine liposomes and SDS micelles were both mixed with samples of 200-295 but did not confer significant change in the structural content of this construct by CD or NMR methods.

The failure of proposed ligands and membrane components to induce structure in the 200-295 construct imply either that an undefined cofactor or posttranslational modification is required to induce structure or, alternatively, that disorder is important to the function of this region of LBP. While the state of native posttranslational modification within LBP remains unknown, another potential function of this region of the protein has been hinted at by recent experiments on the $\alpha_{IIb}\beta_3$ integrin. While searching for integrin-specific antagonists, in contrast to the well-charcacterized RGD peptides that interact with most integrins, a peptide that was shown to inhibit platelet aggregation and activity both *in vitro* and *in vivo* has been identified and consists of a PXXXDW motif, in which the free imino group on proline is essential for activity [125, 126]. Platelet aggregation is the first step in thrombus formation in response to tissue damage, forming a clot to prevent bleeding and to recruit endothelial cells to initiate wound healing. Activation of platelets exposes high affinity sites on the extracellular region of the $\alpha_{IIb}\beta_3$ integrin for the RGD site within fibrinogen, a plasma protein that is subsequently cleaved by thrombin into insoluble fibrin that is deposited on the ECM [127]. Once bound to the ECM fibrinogen directs the proliferation, adhesion, and

migration of incoming endothelial cells that have been stimulated by pro-angiogenic factors [128].

LBP is expressed on both platelets and endothelial cells, and it is likely that the disordered C-terminus, when exposed to circulation, would be cleaved into fragments by the aminopeptidases in the blood and on cell surfaces [126]. Of these fragments, the four repeats of the PXXXDW motif might be released, which may provide a short-lived regulation of platelet adhesion to fibrinogen, affecting the delicate balance between hemostasis and angiogenesis [127], and explain the expression of LBP on platelets and cells involved in wound healing. The peptides may also exert influence over endothelial cell migration that is subject to tight regulation during these processes. The interference between $\alpha_{IIb}\beta_3$ integrin and fibrinogen may promote the circulation of platelets, allowing dislodged tumor cells to remain in the circulation and leaving leaky vessels unoccluded by clot formation, providing another mechanism for LBP involvement in tumorigenesis.

BACKGROUND ON NOS L

Denitrification

Denitrification, the transformation of nitrates into innocuous dinitrogen, is a key process within the larger global nitrogen cycle. The biogeochemical interconversion of nitrogen balances the two major reservoirs of this element, the atmosphere and the much smaller store residing in organic matter, temporarily incorporated in the structures of amino acids, proteins and nucleic acids. Nitrogen can only be utilized by plants in specific forms and so, despite the large quantity of nitrogen in the atmosphere, is often the limiting nutrient in ecosystems, which has important consequences for the environment and the organisms that rely on plants as nitrogen sources. To make gaseous nitrogen available to biological systems, dinitrogen must first be fixed by chemical or biological processes and subsequently transformed into ammonium salts and nitrates, a process predominantly performed by microbes in the soil and in the water. Ammonium salts can then be oxidized to nitrites and then to nitrates through the process of nitrification, which are then assimilated by plants. Ultimately, organic material in turn is converted into gaseous nitrogen and nitrogen oxides through the denitrification pathway, thus completing the nitrogen cycle (see Figure 20)[129]. From a chemical perspective, this cycle involves a complex series of oxidation and reduction reactions, in which the individual steps are coupled to the electron transport systems of the diverse and highly adaptive organisms mediating these processes [130-132].

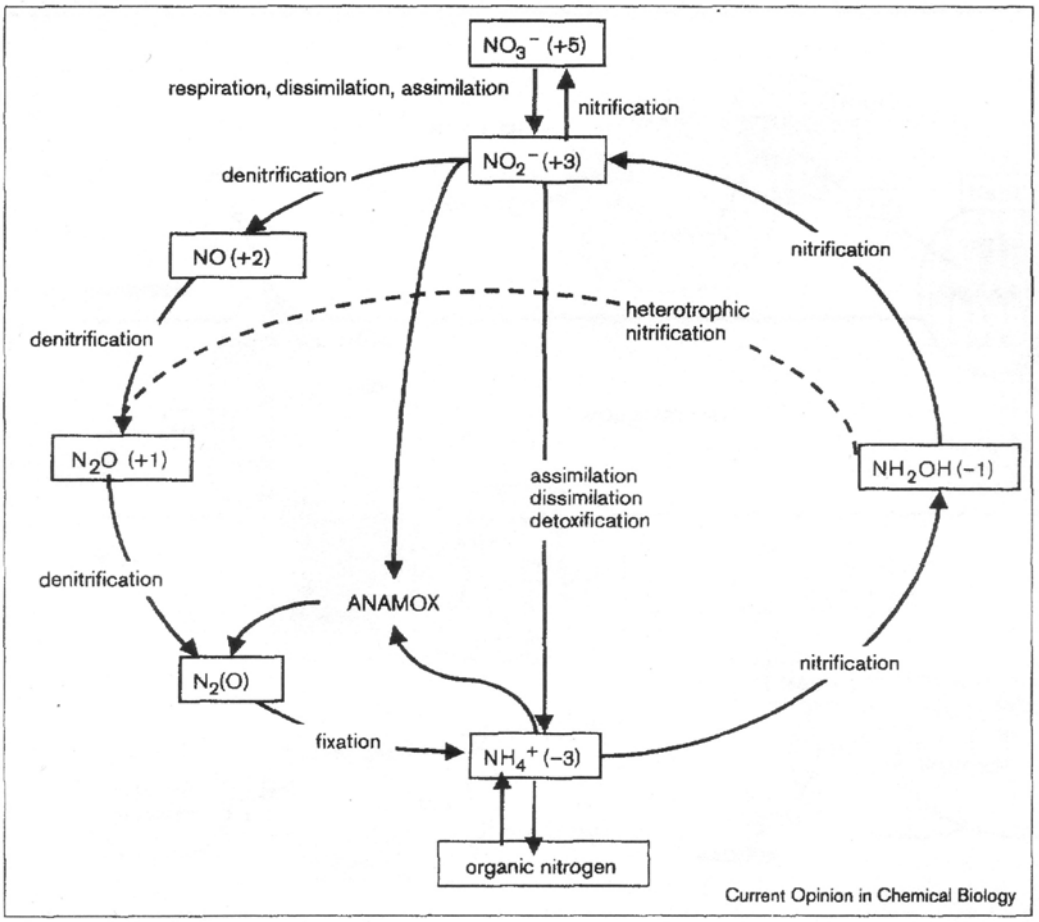


Figure 20. The global nitrogen cycle. Nitrogen is fixed from the atmosphere into ammonia. The conversion of ammonium salts to nitrate (NO_3^-) and nitrite (NO_2^-) is accomplished through the nitrification process, while transformation from nitrate to gaseous nitrogen via nitric oxide (NO) and nitrous oxide (N_2O) intermediates occurs through denitrification, completing the cycle [133].

The transformations that occur throughout the denitrification pathway are inextricably linked to greater environmental and biological processes. Denitrification is the chief means of removing excess nitrogen sources from water, thereby regenerating aqueous resources and recycling organic material. A plethora of species are involved in wastewater treatment, in which nitrogen must be removed prior to remediation [134].

The removal of nitrite, a highly toxic intermediary in sewage treatment processes, is particularly important since this compound has been associated with cancer [135]. Denitrifiers in the soil also rapidly degrade applied chemical fertilizers and so result in commercial agricultural losses. At the same time, increasing levels of agricultural, industrial, and sewage runoff have led to a significant excess of nutrients, particularly nitrate, in the environment. This nutrient overload is known as eutrophication and, when in water systems, often results in phytoplankton blooms. These blooms reduce the light and oxygen available to other organisms, and therefore threaten the overall health of the biosphere. Denitrification has been shown to be a significant mechanism for removing these excess nitrogen sources from lakes, oceans and streams, so that algal blooms are not allowed to harm fish and shellfish populations [136]. The complete conversion of nitrates to harmless dinitrogen, of which nitrous oxide reduction is the final step, is also required to prevent nitrogen oxide reactions with ozone and with rainwater that can ultimately form other toxins and acid rain. In fact, N_2O is an increasing threat to environmental health since this compound is the third most predominant component of greenhouse gas and persists in the environment for over 100 years and, unlike other greenhouse gases, continues to increase yearly [137].

The denitrification apparatus can be found within a wide, unrelated group of organisms including archaea and fungi, but the bulk of nitrate conversion is performed by proteobacteria. The stunning biodiversity exhibited by bacterial populations in soil and aquatic environs has only been hinted at by recent studies that estimate that each gram of soil harbors roughly 100,000 different bacterial genomes [130, 134, 138]. Even in the

relatively more characterized populations in bioreactors, only 1-15% of the microbes are culturable and therefore are poorly understood [134]. With the limited information available, it has been estimated that 5% of bacteria within the soil are capable of denitrification [130]. DNA hybridization studies using denitrification genes as probes found that denitrifying bacteria are found in the vicinity of plant roots within the upper 5 cm of soil. Denitrifying bacteria also inhabit the upper clines of freshwater and marine systems, often associated with biofilms and sediments.

While the faculty for denitrification appears to be an unstable trait, evidenced by organisms that carry the genetic information for proteins involved in this process but that are functionally inactive, the genes responsible for this process are for the most part chromosomal rather than plasmid-encoded [130]. The steps of the denitrification pathway are coupled to the generation of a proton-motive force, generating energy for these organisms, and as such nitrogen oxides provide an alternative to oxygen as terminal electron acceptor. The denitrification apparatus utilizes the same electron transport machinery as aerobic respiration, specifically ubiquinol and cytochrome bc_1, but is less tightly coupled and therefore less efficient. Anaerobic conditions, low oxygen tension and high N-oxide concentrations regulate the expression of the proteins involved in denitrification, to ensure that, if possible, oxygen respiration occurs preferentially [132]. Historically, denitrification has been associated with anaerobic conditions, but aerobic denitrifyers have been isolated and characterized [130]. Interestingly, the enzymes that carry out the steps of denitrification in these organisms appear to be associated with

functions other than energy production, including detoxification of nitrite and nitric oxide and maintenance of redox conditions within the cell [132].

The genes encoding the enzymes that effect denitrification are clustered with genes that produce accessory proteins, often organized on a single transcriptional unit. These genes code for proteins that aid in proper processing and maturation of the enzyme including electron transporters, cofactor importers, and cofactor processors. The conserved arrangement of the structural genes and the accessory genes suggests that the denitrification capability has been transferred among species by lateral gene transfer [132]. Phylogenetic analysis of denitrifying sequences within the gene database demonstrates the presence of very similar genes among genetically distinct groups, supporting this hypothesis. However, the occurrence of related denitrification genes in both prokaryotes and archaea suggest a common ancestor as well [139].

The conversion of nitrate to dinitrogen is performed by four distinct and relatively independent respiratory complexes that ultimately leads to ATP synthesis and therefore energy production for the organism (see Figure 21) [129]. Nitrate reduction can be performed by nitrate reductase within the periplasm (Nap) or on the inner face of the cytoplasmic membrane by the membrane-bound form (Nar) [129, 140]. All of the known nitrate reductases harbor a unique molybdopterin cofactor (MGD) at their active site with an accompanying iron sulfur center for electron donation nearby [140]. Nitrite reductase, of which there are two entirely distinct types, a copper enzyme (NirS) and a cd_1 cytochrome (NirK), reduce nitrite to nitric oxide in the periplasm. The respective membrane compartmentalization of these two enzymes necessitates a nitrate/nitrite

transporter across the inner membrane. While no transporter has been unequivocally identified thus far, a component of the *nar* operon (NarK) appears to be involved [129, 139].

The highly reactive NO radical species generated by nitrite reductase is subsequently reduced by nitric oxide reductase (Nor). Nor is a multisubunit membrane protein containing heme and non-heme iron, and is less well understood due to the difficulties in isolating and expressing this twelve membrane-span protein [139, 140]. Finally, nitrous oxide reductase catalyzes the reduction of N_2O to gaseous nitrogen, forming a triple bond between the two nitrogen atoms and releasing the oxygen as water. Interestingly, bacteria species have been found with only plasmid-encoded *nos* genes, in which N_2O respiration is separate from denitrification. This reduction has been shown to derive its electrons indirectly from cytochrome bc_1, with concomitant proton transfer across the cytoplasmic membrane, and therefore represents a respiratory process in itself [129].

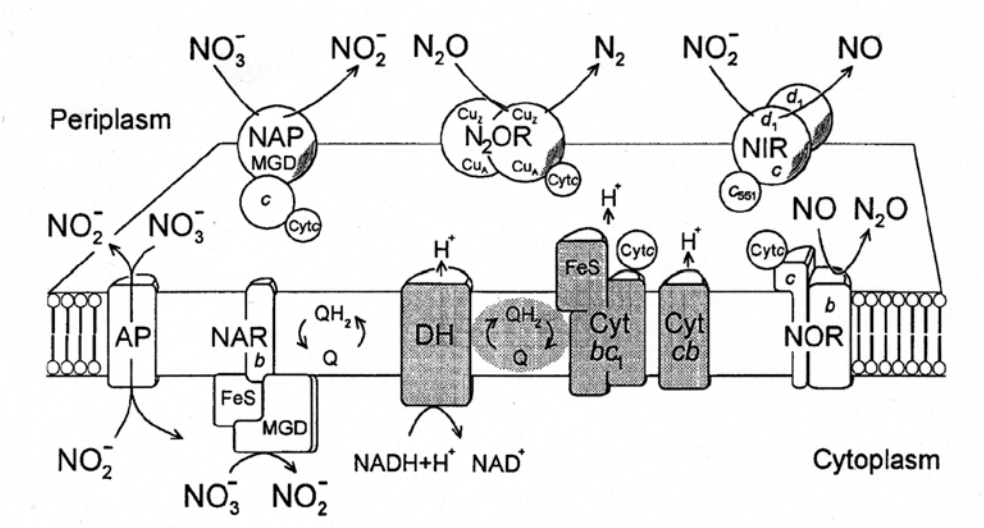


Figure 21. The denitrification apparatus [141].

Electron transport partners for the enzymes of denitrification, although required for electron delivery to these enzymes, have not been identified and may in fact be non-specific and readily interchangeable [129]. The multitude of encoded gene products involved in metal center assembly, electron transport, regulation, and metal transport of the denitrification pathway have yet to be characterized. In contrast, detailed structural information on all of the metalloenzymes of denitrification, with the exception of nitric oxide reductase, is currently available [140].

Like many of the enzymes involved in metabolism and particularly respiration, metal cofactors are utilized by the enzymes of the denitrification pathway in order bind the small inorganic substrates and perform the unique redox chemistry required [142]. These metallocenters take the form of copper clusters, iron-sulfur clusters, molybdopterin prosthetic groups, heme and non-heme iron. The variability of oxidation states along with the accompanying flexible valency exhibited by metals make metalloproteins particularly suitable for electron transfer reactions that are not satisfactorily performed by purely organic matrices [142]. At present, the details of these crucial processes are poorly understood, especially concerning the roles and specificity of different electron donors and acceptors which are in part responsible for the staggering diversity of respiratory processes.

Nitrous Oxide Reductase

Nitrous oxide reductase or N_2OR catalyzes the final step in the denitrification process, converting the toxic nitrous oxide into an innocuous gas. The conversion of N_2O to N_2 requires two electrons and two protons, producing water as a byproduct:

$$N_2O + 2H^+ + 2e^- \rightarrow N_2 + 2\,H_2O$$

The reaction itself is energetically demanding; while favored thermodynamically (ΔG = -25 kcal/mole), the reaction is characterized by a high activation energy (E_a = 59 kcal/mol) [143]. While this reduction is typically mediated by a copper-containing nitrous oxide reductase in the periplasm of anaerobic organisms, aerobic nitrous oxide reductase has also been described [129]. Prior to translocation into the periplasm, the N_2OR preprotein contains a 50 residue signal peptide consisting of a twin-arginine motif (S/TRRXFLK) [144]. This arginine motif is recognized by the Tat translocon (Twin arginine translocation) which is responsible for the secretion of many complex redox proteins via a Sec-independent mechanism [145]. This secretion sequence has been associated with periplasmic proteins that require processing and cofactor insertion, which may occur prior to translocation. This system is capable of translocating folded proteins on a 5-6 nm scale, differentiating it from the Sec pathway, in which an unfolded substrate is threaded through the cytoplasmic membrane [146]. In the case of nitrous oxide reductase, copper insertion occurs in the periplasm but the polypeptide appears to be correctly folded prior to Tat translocation [144, 147, 148].

While it has been recognized since the early 1980's that nitrous oxide reductase contains multiple copper atoms, only recently have the structural details about the two distinct copper centers that mediate the two-electron reduction of N_2O come to light. Three crystal structures of N_2OR, one lower resolution structure from *Pseudomonas nautica* (2.4Å) and two higher resolution structures from *Paracoccus denitrificans* (1.6 Å), have been solved (see Figure 22) [137, 149, 150]. Nitrous oxide reductase is a

homodimer of 130 kDa, in which each monomer consists of two domains. A cupredoxin domain harbors an electron transfer site similar to that found in cytochrome *c* oxidase, designated Cu_A. The second domain consists of a seven-bladed propeller that funnels substrate into the novel catalytic cluster, Cu_Z. The two monomers are arranged such that the Cu_A center of one subunit lies in close proximity to that of the catalytic site of the other subunit (10Å compared to the 40Å intramonomer distance), presumably to facilitate electron transfer and further suggesting that dimerization is required for efficient transport [143, 151]. Two residues, one methionine and one phenylalanine provided by the C-terminal domain, bridge the space between the two copper centers of different subunits, suggesting that they might form the electron pathway to the catalytic site. The two monomers exhibit a highly complementary interaction surface consisting of both polar and hydrophobic residues, as well as two tightly bound calcium ions at the monomer-monomer interface in the *P. denitrificans* structure [137].

The Cu_A site consists of a dinuclear copper cluster bound within the C-terminal domain which is comprised of roughly 100 residues and bearing homology to the Cu_A domain of bovine cytochrome *c* oxidase [149]. This cupredoxin structure is formed from a nine-stranded antiparallel β-sandwich with an alpha helix on one side. Tightly packed hydrophobic sidechains join the two facing beta-sheets, forming a structurally strong central core. The copper binding site includes residues from two loops and resides on one edge of the cupredoxin domain. Each copper is liganded by the δ nitrogen atom of a histidine residue and is bridged by two cysteine residues. Two other residues, tryptophan and methionine, complete the copper coordination and serve to fine-tune the electronic

properties of the center. The dinuclear center consists of a metal-metal bond of 2.47Å and demonstrates a mixed valence site in which both coppers are believed to possess an oxidation state of +1.5 [137, 151]. The specific geometry of the cluster allows delocalization of charge across the two sulfur atoms and is arranged such that changes in oxidation state and therefore changes in the preferred coordination of the copper centers occur readily.

The Cu_A site within N_2OR has been shown to accept electrons from both pseudoazurin and cytochrome c_{551} *in vitro*, as well as from artificial electron donors such as methylene blue [129]. Electron carriers within the periplasm have been shown to substitute for each other, so that the physiologic donor may be difficult to determine. Interestingly, the nondenitrifying organism *Wolinella succinogenes* contains a nitrous oxide reductase that, along with the two copper clusters normally found in other N_2OR's, bears a C-terminal monoheme cytochrome *c* domain that presumably serves as a covalently attached electron donor [152].

Unlike the C-terminal electron transfer domain, the N-terminal domain of N_2OR (consisting of 480 residues) exhibits no significant homology to other proteins of known structure. Linked via 33 residues to the C-terminal cupredoxin domain, the catalytic domain adopts a seven-bladed propeller structure, in which each blade consists of a four-stranded, twisted antiparallel β-sheet arranged about a central channel. Beta-propellers

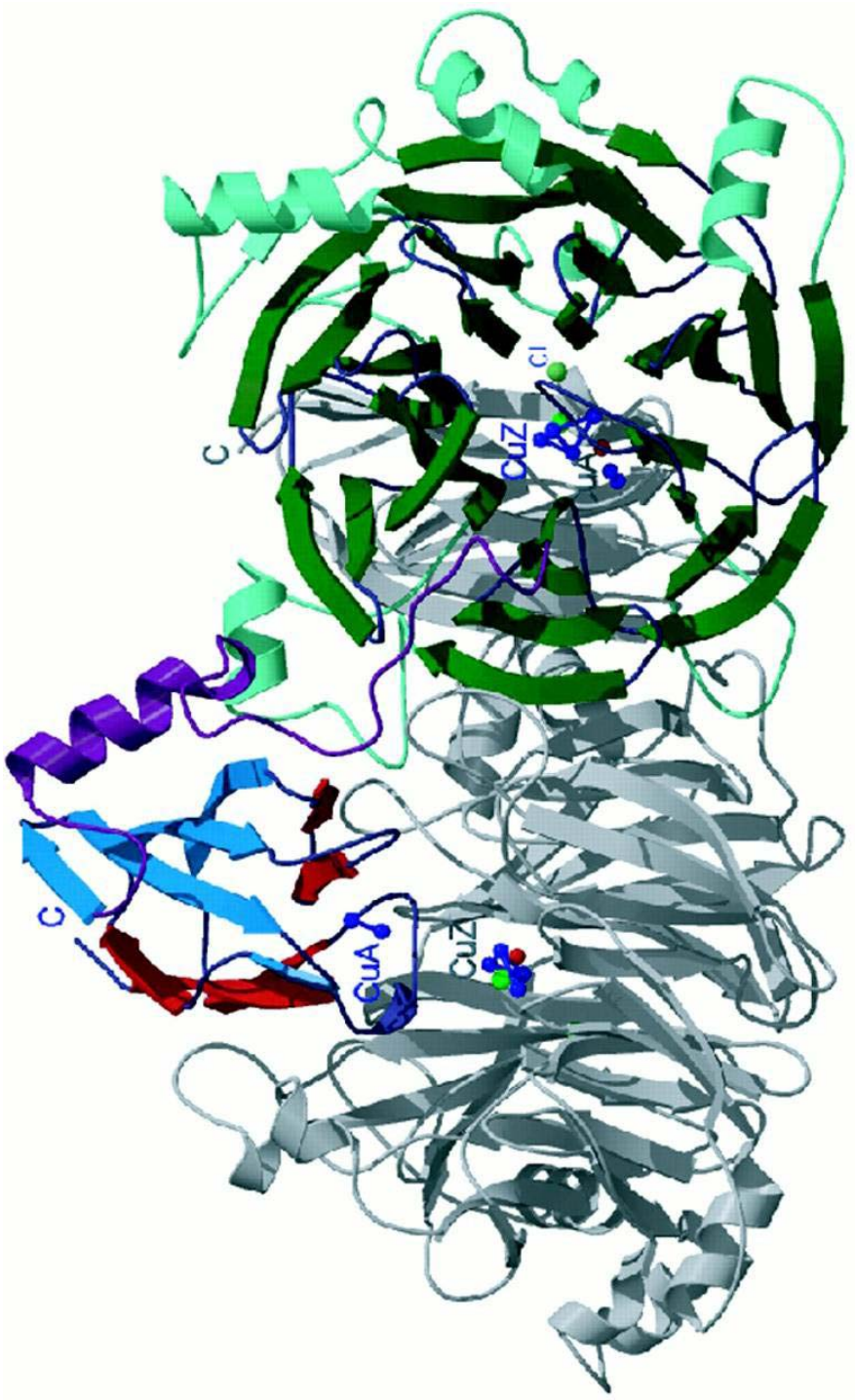


Figure 22. Nitrous Oxide Reductase from *Pseudomonas nautica*. The crystal structure of the dimer is shown; one monomer is uniformly gray while the propeller domain from the other dimer is green and the cupredoxin fold is shown in blue, red and purple [150].

are found in a wide variety of proteins that function predominantly in catalysis, appearing to play a scaffolding role in these types of proteins [153]. In the case of N_2OR, this domain creates a cone-shaped solvent channel well-suited for directing substrate into the μ_4-sulfide bridged copper cluster residing at the axis of symmetry in the propeller domain [149].

The Cu_Z cluster housed within the active site of the N-terminal domain is a novel metallocenter specifically organized to facilitate nitrous oxide binding and reduction (see Figure 23). The unprecedented metallocluster consists of four copper atoms arranged in a distorted tetrahedron, in which Cu1 is slightly displaced away from the other copper ions, averaging a distance of 3.10 Å away from Cu3 and Cu4 versus the average distance between the other copper atoms (Cu3/Cu4 to Cu2) of 2.55 Å in the *P. nautica* structure [151]. The sulfide is positioned on top of this tetrahedron, bridging each copper in a relatively equidistant fashion--this unique and unexpected mode of sulfur coordination is consistent with magnetic circular dichroism, Resonance Raman and EPR data previously collected [150, 154]. The ligand organization about the μ_4-sulfide bridged cluster forms a "butterfly" pattern that is relatively symmetric with the exception of Cu4, which possesses one histidine ligand in contrast to the two histidine ligands forming the coordination sphere of the other copper centers. The ligands for the Cu_Z cluster are provided by conserved histidine residues belonging to the innermost strands of the propeller blades as well as on loops on the top of the propeller domain, held in specific orientations by an intricate network of hydrogen bonds. To complete the coordination of Cu4, a weakly bound unidentified ligand (H_2O, O_2 or OH) is found along the Cu4-Cu1

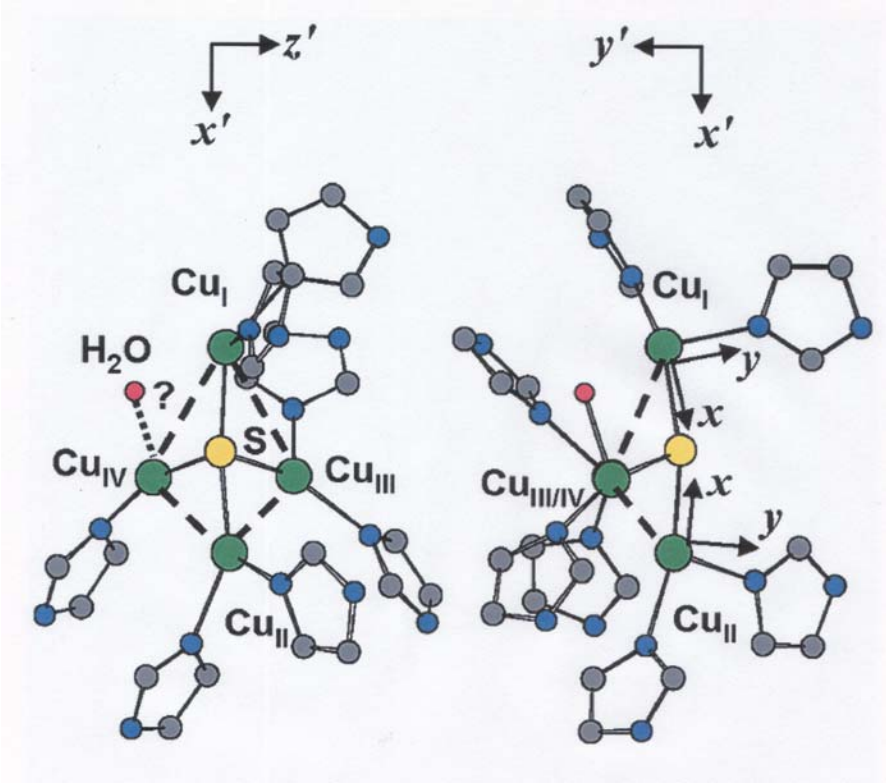


Figure 23. Crystal structure of the Cu_z cluster from *Pseudomonas nautica*. The front view is shown on the left, while a side view is displayed to the right [143].

edge of the tetrahedron in the higher resolution *P. denitrificans* structure [137, 149]. The histidine ligands of Cu1 are arranged so as to accommodate Cu1 binding to this ligand as well. This 4 coordinate geometry about the Cu1 center is consistent with calculations that suggest that this copper is oxidized (and therefore favors 4 ligands over 3) in the +2 state while the remaining coppers exist in the +1 oxidation state. Furthermore, the positioning of this water moiety suggests that this site is available to incoming substrate and therefore might be the binding site for nitrous oxide.

Spectral data acquired on N_2OR, coupled to density functional calculations, have suggested a model for the bonding of the cluster and hints at a mechanism for N_2O

reduction. The +2 oxidation state of Cu1 is determined by the slight outward displacement of Cu1 from the rest of the copper ions, which exhibit a +1 oxidation state. The substrate, which appears to have access to the Cu1-Cu4 edge of the cluster according to crystallographic data, most likely bridges the two copper atoms in a μ-1,2 bridging mode. The sulfur atom, which can effectively accommodate electron density, acts as a pathway between Cu1 and Cu2 for the electron that is required for reduction. Electron donation by the relatively reduced and proximal Cu4 atom may occur simultaneously, thereby effectively overcoming the high energy barrier to N_2O activation. As in the copper architecture of the Cu_A cluster, the delocalization provided by surrounding copper atoms and bound sulfur contribute to low reorganization energy and stabilization of the cluster at different oxidation states, and therefore lowers the activation energy [143, 155].

Recent EPR experiments on N_2OR from *Pseudomonas nautica* have shown that the Cu_Z cluster can be further reduced [155]. The four Cu (I) form of the tetranuclear cluster is more active and therefore kinetically more consistent with observed substrate turnover rates, leading the researchers to suggest that this is in fact the redox active, catalytically relevant form of the nitrous oxide reductase cluster [156]. Computational methods were used to model the binding of the nitrous oxide substrate to this reduced cluster, in which each copper is in the +1 oxidation state. In the lowest energy structure, the terminal nitrogen of N_2O binds to Cu1 while the oxygen atom interacts with Cu4. This binding mode induces a “bend” in the normally linear gaseous molecule, altering its electronic properties in order to make it a better electron acceptor and therefore activating

it for reduction [156]. While the results of these calculations are intriguing, further chemical characterization of this reduced state is needed.

The *nos* Operon

Given the complexity of the Cu_Z center, it is perhaps not surprising that early studies of nitrous oxide reductase revealed that redox activity requires the co-expression of additional factors for metal cluster assembly, much like other metalloenzymes of the denitrification pathway. Early experiments demonstrated that reconstitution of N_2OR with exogenous copper results in an incomplete Cu_Z metal center. Further studies showed that expression of N_2OR in *Escherichia coli* yields the apo enzyme despite the fact that the protein is successfully exported into the periplasm [141]. Genetic lesions introduced by insertional mutagenesis localized the Cu_Z assembly defect to specific genes located adjacent to *nosZ*, the structural gene encoding N_2OR. The gene encoding N_2OR is part of a larger gene cluster, the *nos* operon (*nosRZDFYLX*), which codes for gene products associated with N_2OR metal cofactor assembly, transcriptional regulation, copper transport and perhaps folding, and therefore confer activity when co-expressed.

Little is known of the precise functions these gene products perform, but some insight into the mechanisms governing copper cluster assembly of and insertion into N_2OR can be gained from sequence analysis. The gene components *nosDFYL* bear topological and functional similarity to the ABC transporter superfamily. ATP binding cassette or ABC transporters are multi-component assemblies that are used ubiquitously by archaeal, eukaryotic, and prokaryotic organisms for small molecule transport. Most

ABC transporters couple the hydrolysis of ATP and transport of molecules across a membrane against a concentration gradient and can serve as both importers and exporters [157, 158]. ABC transporters consist of a hydrophobic integral membrane protein that forms a channel through the membrane, a conserved, hydrophilic cytoplasmic ATP/GTPase for energy production, and one or two periplasmic components that serve varying substrate-specific functions. These systems are implicated in such diverse cellular processes as detoxification, ion transport, DNA repair, regulation of gene expression and cofactor transport [158]. ABC transporters have been associated with cytochrome *bd* biogenesis, iron-sulfur cluster assembly, and cytochrome *c* heme synthesis, and so their appearance in a cluster coding for metal cluster assembly is consistent with previously identified functions [157].

In the *nos* operon, the proteins *nosDFYL* are believed to provide the machinery designed to couple ATP/GTP hydrolysis to copper and/or sulfur insertion within the periplasm. NosF is a peripheral cytoplasmic protein that contains an ATP/GTP binding motif (G-XXXX-GKT) and bears sequence homology to other proteins involved in metal transport [141]. *In vitro* studies of recombinant NosF demonstrated ABC-type ATPase activity [159]. NosY is predicted to contain six transmembrane helices, a common characteristic of the membrane components of other bacterial transport systems. This protein may transduce energy generated from ATP hydrolysis in the cytoplasm (provided by NosF) and couple it to the metal or sulfur insertion process that takes place in the periplasm. NosD is a 45 kDa hydrophilic periplasmic constituent predicted to contain a characteristic carbohydrate-binding and sugar hydrolase (or CASH) domain consisting of

a three strand beta sheet that forms the turns of a B helix [160]. NosD, F, and Y have all been demonstrated to be essential for the formation of catalytically active nitrous oxide reductase through insertional mutagenesis experiments.

The *nos* cluster also encodes accessory proteins that are not strictly required for Cu_Z formation or, alternatively, that can be functionally replaced by other factors. NosL is a highly soluble, 18.5 kDa periplasmic lipoprotein that exhibits little sequence similarity to proteins of known function [161]. NosX is a peripheral membrane protein proposed to have one transmembrane helix and contains a twin-arginine secretion sequence that targets it for export by the Tat translocon [162]. NosX is homologous to RnfF, an iron-sulfur protein involved in electron transport, but lacks the conserved [4Fe-4S] cysteine motif and so does not appear to be an iron-sulfur protein [163]. While NosX and its homologue NirX have been shown to be required for holo enzyme biogenesis in the organisms *Sinorhizobium meliloti* and *P. denitrificans*, it is not present in the *nos* cluster of every organism studied and therefore may not be essential for cluster assembly [148].

Three additional accessory proteins have been identified that often appear on separate transcription units adjacent to the *nos* operon. NosR is a regulatory protein predicted to contain five transmembrane helices spanning the cytoplasmic membrane [162, 164]. This protein harbors two [4Fe-4S] clusters, at least one of which may function in the cytoplasm, perhaps acting as an oxygen sensor to direct NosZ transcription under oxygen-deficient conditions [164]. This sensing activity is reminiscent of FNR (fumarate nitrate reduction), a global regulator of anaerobic

metabolism bearing a [4Fe-4S] cluster [165]. The iron sulfur cluster of FNR and FNR-like transcriptional regulators is responsible for stabilizing the dimeric form of the protein that binds to DNA with high affinity and directs transcription. Upon exposure to oxygen, the cluster disassembles, the dimer dissociates, and the protein loses its ability to bind DNA. NosR also binds a flavin moiety, which when deleted from a recombinant construct, resulted in loss of nitrous oxide activity despite the fact that the Cu_Z center was intact, suggesting an electron transfer role for this cofactor [164]. NosR expression is required for transcription of both NosD and NosZ, and is subject to regulation by a DnrD transcription factor similar to FNR, which in turn responds to nitric oxide derived from upstream denitrifying activities [148, 159, 164, 166].

NosA is a copper-binding outer membrane protein that is believed to form a channel to the extracellular milieu and is involved in copper uptake and insertion into N_2OR when the exogenous copper supply is low [148, 167]. While it has been shown to be required for copper center assembly and therefore N_2O reductase activity in *Pseudomonas stutzeri*, [168] NosA is not consistently found in all denitrifying organisms. In *P. stutzeri*, a *tatE* operon which encodes TatE, a member of the translocation machinery for nitrous oxide reductase, is found downstream of the *nos* operon and is responsible for complete export and maturation of N_2OR [159].

Based on the similarities with the ABC transporter systems and sequence-based predictions, a model for the locations and interactions of the Nos proteins has been proposed (see Figure 24). NosL, N_2OR, NosD, and NosX are all soluble components of the periplasm, while the energy adapting proteins NosF and NosY are shown at or in the

inner membrane (IM) as is NosR. NosA, believed to be a copper transporter, is shown in the outer membrane (OM). It has been postulated that the ABC transporter machinery is utilized for sulfur (shown as [S]) transport and incorporation.

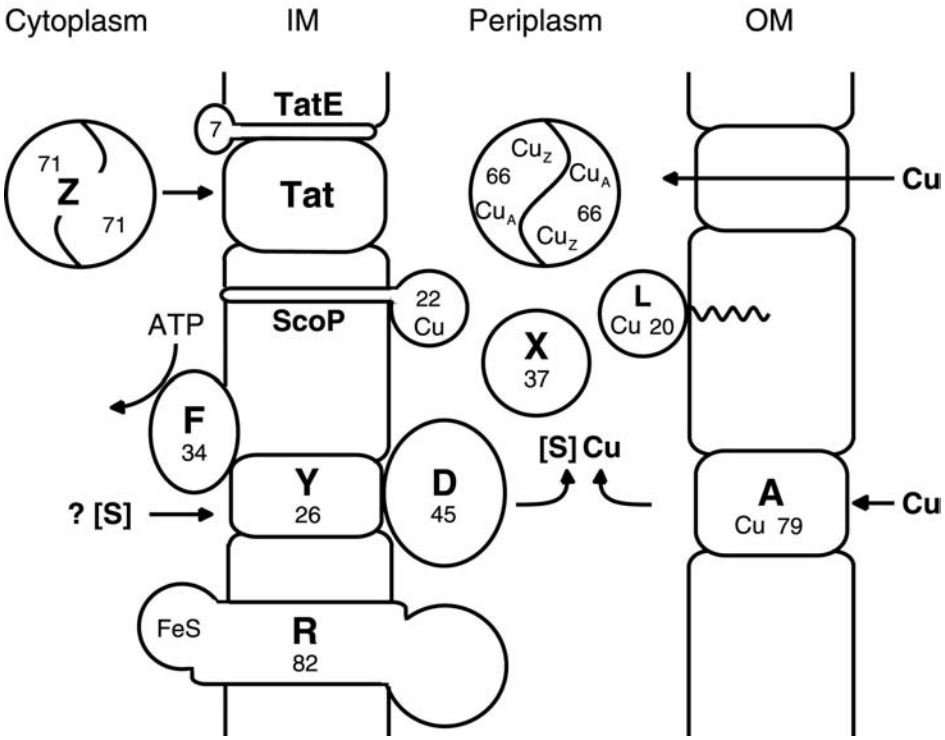


Figure 24. Components of Cu_Z assembly. The numbers correspond to predicted masses. [S] is the unidentified sulfur donor. IM refers to the cytoplasmic membrane while OM refers to the outer membrane [164].

NosL

As a preprotein, NosL contains a lipobox, a signal peptide (L(A/S)(G/A)C) that targets the protein for translocation through the inner membrane, where it may remain or be subsequently shuttled to the outer membrane, where most lipoproteins reside [169]. Upon translocation, the Cys residue is modified via a thioether linkage to diacylglycerol,

the signal leader is cleaved, and the amino-terminal of the Cys residue is acylated. Mutagenesis experiments demonstrated that the identity of the second position (Asp) after the Cys residue determines whether or not the lipoprotein avoids the machinery that serves to release lipoproteins from the inner membrane and therefore leads to incorporation into the outer membrane. In all NosL sequences analyzed, only one sequence *(Pseudomonas aeruginosa*) contains an Asp residue at position two, suggesting that NosL is an outer membrane protein in most organisms (see Figure 25) [169]. The common cellular localization and potential requirement of NosA and NosL under copper-limiting conditions suggest that these two proteins might interact, providing the apparatus for copper delivery/insertion into soluble N_2OR under biologically relevant conditions. In order to fully understand the function of NosL, copper center assembly may need to be studied in the presence of chelators to mimic these conditions. Furthermore, successful *in vitro* assembly of N_2OR may need to be accomplished in order to discover the presence of a "rescue" copper chaperone that may be capable of delivering Cu(I) in lieu of NosL, leading to the Nos+ phenotype observed in NosL mutants *in vivo.*

In a surprising discovery, recombinant NosL from the facultative anaerobe *Achromobacter cycloclastes* as expressed in an *E. coli* host was found to contain one copper atom per monomer in the +1 oxidation state, despite the fact that the protein exhibited no known copper-binding motif [161]. Reaction of the copper-loaded protein with 2,2-biquinoline, a Cu^{+1}-selective chromophoric reagent, confirmed this finding [170].

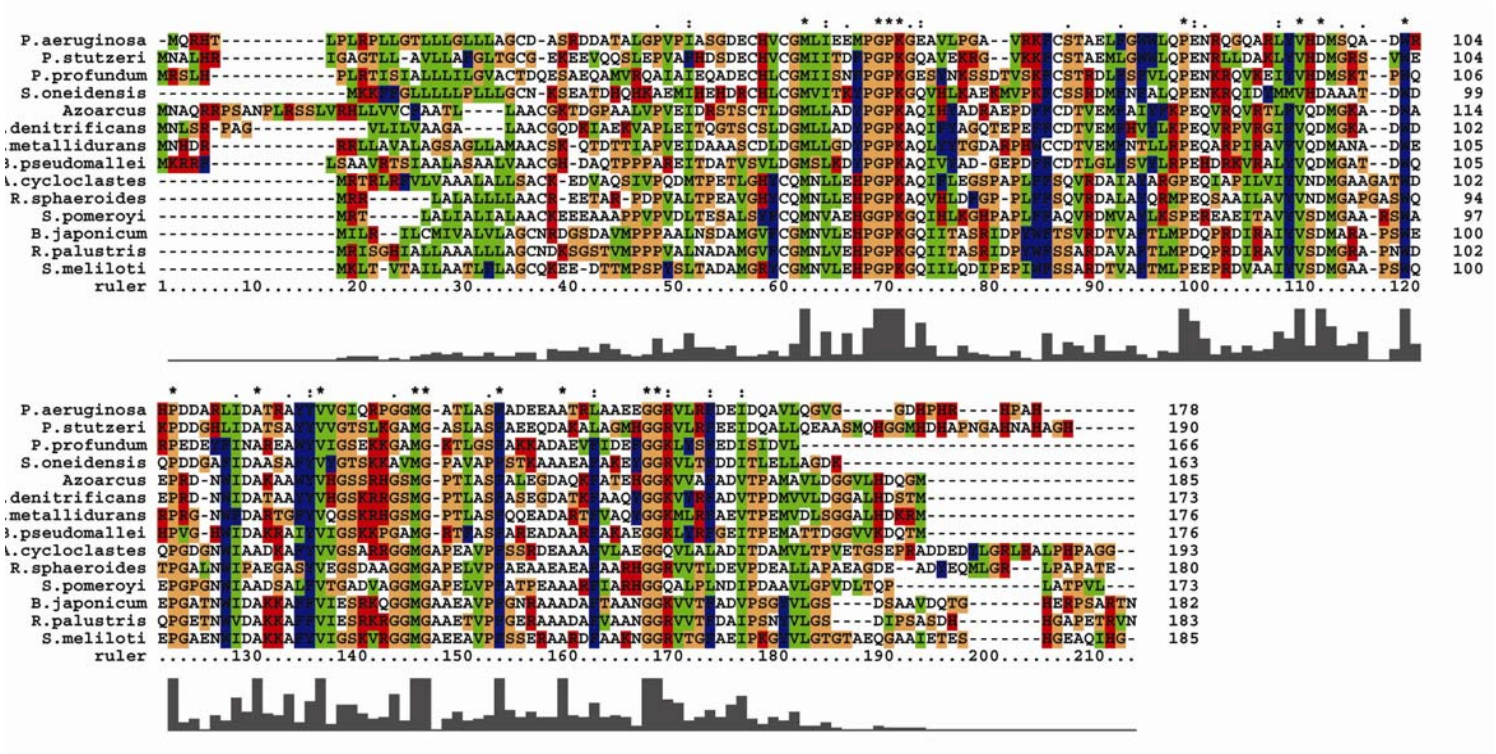


Figure 25. Sequence alignment of full NosL sequences in the BLAST database performed by Clustal X [171].

Cu EXAFS absorption spectra showed that the copper is bound by three ligands, two of which are likely to be sulfur ligands and one of which is believed to be a nitrogen/oxygen ligand.

A clustal alignment of the 14 fully sequenced NosL primary structures (see Figure 25) shows that there are two conserved methionine residues that may serve as potential copper(I) ligands and one cysteine residue aside from the lipid modified cysteine that, while not positionally conserved, is present in all sequences [171]. There are no conserved histidine residues, but one strictly conserved lysine that may act as a nitrogen ligand, and a conserved aspartic acid and tyrosine that may serve as oxygen donors. The ligand content of this mononuclear copper center is unprecedented, making NosL a unique metalloprotein and suggesting a highly specialized role for this motif in Cu_Z assembly. NosL failed to incorporate Mn(I), Co(II), and Co(III), and so appears to be specific for copper. Prolonged exposure to aerobic conditions resulted in the loss of copper and dimerization via an intermolecular disulfide in the recombinant *A. cycloclastes* product.

The ability of NosL to bind copper has opened up new possibilities for the function of this highly conserved protein. Because many electron transfer proteins of the periplasm are mononuclear copper(I) proteins, *in vitro* studies were undertaken in order to determine if NosL is capable of functioning as an electron transfer partner to N_2OR. In the presence of N_2O and N_2OR, the EPR spectrum of NosL was monitored to detect the conversion of copper from the +1 to the EPR-active +2 oxidation state and therefore indicate electron donation to the Cu_A site. The EPR spectrum remained unchanged,

strongly suggesting that the protein is not an electron donor. These results are supported by the lack of sequence homology between NosL and other copper-containing electron transfer proteins and the observation that blue copper proteins must retain affinity for Cu^{+2} after electron delivery.

Metallochaperones

Cells have developed a great capacity to chelate copper in order to minimize the toxicity of this highly reactive metal, such that while it is estimated the total cell concentration of copper is 10μM, there is far less than one free copper atom per cell, highlighting the extraordinarily tightly regulated nature of copper homeostasis [172, 173]. Copper incorporation into metalloenzymes, therefore, is a highly orchestrated event, requiring specifically dedicated systems for copper sequestration, protection, transport, and insertion. Metallochaperones are a relatively new class of proteins that deliver and insert the copper cofactor into larger clusters of a target protein, often into membrane-bound or compartmentalized metalloenzyme targets, and are distinct from detoxification proteins. The flexible two and three ligand geometry exhibited by copper metallochaperones allows the approach of incoming ligands from a target protein in order to facilitate copper transfer while simultaneously binding the metal tightly enough to protect it from oxidation and scavenging by competing thiols. This distinct coordination environment is unique to this class of proteins, in direct contrast to other copper (I) proteins that utilize copper as cofactors, by saturating, burying and constraining the copper center [172, 173]. The emerging importance of these metal trafficking is underscored by the association of disorders such as Menkes disease, Wilson disease,

Alzheimer's, Parkinson's and Creutzfeld-Jakob disease with abnormal copper transport [173].

The Atx1 and Atx1-like metallochaperones, including a human homologue HAH1 or ATOX1, are the most well-understood and prototypical copper delivery systems. After copper import by Ctr1, these proteins chelate copper(I) and deliver it to the ATPase copper transporters ATP7A and ATP7B in humans and CCC2 in yeast, both of which pump copper into the lumen of the Golgi or into vesicles upon ATP hydrolysis, where it is incorporated into secreted copper proteins (ceruloplasmin in human and Fet3, a multipcopper oxidase, in yeast) [172]. The Atx1-like chaperones harbor a MXCXXC copper-binding motif near the N-terminus, in which the two cysteine residues bind the copper ion and the methionine residue extends into the hydrophobic core of the protein, perhaps serving as a sensitive tether between the core and the copper-binding loop [174]. A lysine residue within this loop neutralizes the negative charge of the bis-thiolate center and is conserved among eukaryotic homologues.

Structures of Atx1 and the structurally homologous copper chaperones CopZ and ATOX1, and the related mercury chaperone MerP are available and reveal a conserved structure consisting of a ferredoxin-like βαββαβ fold in which two helices are superimposed onto a four strand β-sheet [175]. Notably, the copper atom is bound within a relatively flexible solvent-exposed loop to accommodate changes in coordination number [176]. Similar to NosL, the metal dissociates from these chaperones upon oxidation to the +2 state, a feature that contributes to facile copper transfer.

Comparison of the structures of both the apo and holo forms of Atx1-like proteins reveal striking conformational changes upon copper-loading and release. Upon loss of copper, the ligand-bearing loop becomes more disordered and longer, reflected in chemical shift and relaxation behavior differences in NMR spectra between the apo and holo forms. Both the cysteine residues and the methionine residue within the MXCXXC motif become more solvent-exposed in the apo protein [174, 177]. It is believed that transfer of the copper atom occurs via a facile ligand exchange process involving a series of two to three coordinate intermediates, made possible by the flexible copper binding loop [177]. Dissociation of the copper from the chaperone may be in part driven by the increased mobility of the copper-binding loop observed in the apo protein [178]. In this family of chaperones, protein-target recognition is achieved chiefly through electrostatic interactions with often structurally homologous target domains, bearing slight amino acid differences between them, as well as by residues in the copper-binding loop [174].

Despite the fact that NosL lacks the conserved metal binding motif MXCXXC, the coordination and oxidation state of the copper center of NosL are consistent with a role as a copper chaperone, more so than with any other family of known copper(I)-binding proteins [173]. Like the metallochaperones, the geometry of the NosL copper center stabilizes the +1 oxidation state even within the relatively oxidizing environment of the periplasm, shielding it from disproportionation into Cu(0) and Cu(II) complexes. Recent NMR structure determination of the cyanobacterial copper chaperone ScAtx1 revealed three coordinate copper-binding by two cysteine residues within the CXXC motif and by an $N^{\varepsilon 2}$ histidine 46 residues away in the primary sequence, perhaps an

ancestral relic of a stabilizing interaction supplanted by the Lys residue in analogous eukaryotic organisms [179]. This emerging copper-binding motif demonstrates that other ligand arrangements can be successfully accommodated by metallochaperones.

In vivo incorporation of copper into Cu, Zn superoxide dismutase (SOD), an enzyme that protects cells from oxidative damage by neutralizing superoxide, requires the chaperone CCS. The crystal structure of CCS includes two of three domains. The first domain adopts a similar structure to Atx1 containing the MXCXXC copper-binding motif, while the second domain is structurally similar to the target SOD and mediates recognition via dimerization with its target. A third domain, required for SOD activation, contains a potential metal binding site (CXC) but is disordered in the crystal structure [180]. It is believed that copper is acquired and protected by the Atx1-like domain, and then transferred to the third domain after target recognition via the SOD-like C-terminus [181]. This small, mobile domain is believed to be responsible for the actual insertion of copper into the buried active site of SOD, in which there are four histidine ligands to Cu (II), demonstrating that copper donation to a histidine-rich active site by a metallochaperone has precedent [180]. Interestingly, CCS and domain I in particular appear to be critical for SOD function only when copper concentrations are limiting in the growth media, perhaps analogous to NosL in the N_2OR system [172, 181]. Interestingly, CCS has been shown to activate SOD1 exclusively under aerobic conditions, and thus may also serve a more complex regulatory role in shifting the pool of SOD from apo- into holo-SOD when required [182].

CCS is not the only copper chaperone that exhibits structural features not found in the more simple Atx1-like metallochaperones. Cox17 and Cox11 are both believed to be involved in copper delivery to membrane-bound cytochrome *c* oxidase, the enzyme that catalyzes the reduction of oxygen to water thereby generating a proton gradient across the mitochondrial inner membrane. The multisubunit protein requires two copper clusters, one Cu_A dinuclear cluster similar to that in N_2OR and one mononuclear cluster Cu_B, as well as a heme. The newly available structures of these two chaperones display folds other than the βαββαβ fold exhibited by Atx1-like proteins, and, given the very different target metalloclusters and protein partners these copper chaperones interact with, is perhaps not surprising [179, 183]. Cox11, which is believed to be involved in the biogenesis of the Cu_B copper cluster, adopts an Ig-like β-sandwich fold in which the copper is believed to be held in a loop between two β-strands on the surface, creating a unique variation on copper presentation [179]. The protein dimerizes via this metal-binding region of the monomer, producing a two copper-4 cysteine cluster at the domain interface that may be important for transfer to the all-histidine coordination of the Cu_B site. The Cox17 structure is a small 69 residue protein consisting of two helices and a relatively large unstructured region (comprising 40% of the protein) that may be involved in target recognition. The single copper(I) ion is bound via two cysteines in a loop at one end of the N-terminal helix in an analogous manner to the Atx1 binding site. Interestingly, the requirement for the chaperone Cox17 in Cu_A cluster assembly suggests that under physiological conditions a copper chaperone may be required to aid in the

biogenesis of the Cu_A center of N_2OR as well, hinting at an as yet unidentified function for NosL [184].

The nitrous oxide reductase assembly system to which NosL belongs is more complex than the systems associated with previously studied copper chaperones, particularly in the requirement for energy transduction and in the novel architecture of the polynuclear sulfide-bridged Cu_Z center. While the copper binding characteristics associated with the NosL metal center are in general consistent with the features exhibited by other copper chaperones, the ligand environment is quite distinct, perhaps serving a highly specific role in Cu_Z and/Or Cu_A assembly.

Hypothesis

Early studies on recombinant apo NosL lacking the thioether-linked lipid demonstrated that the protein was well-folded and amenable to NMR structural studies. Determination of the three dimensional solution structure of apo NosL was undertaken in order to shed light on the potential function of NosL within the N_2OR system. Structural data on the apo form of this unique copper-binding protein, in concert with the increasing amount of available structural information and sophisticated algorithms for searching structural databases, should allow the identification of structural homologues to NosL, potentially revealing functional and evolutionary relationships. Functional features of NosL that reflect its capacity to deliver copper and/or assemble copper centers of the nitrous oxide reductase system will also be evaluated. Preliminary NMR experiments were also performed on holo NosL that allowed chemical shift mapping of copper-

specific changes upon ligand binding, providing valuable information on the unique copper center of NosL and suggesting that more global effects occur upon copper binding than those observed in previously characterized copper chaperones.

MATERIALS AND METHODS FOR NOS L

Expression and Purification of Recombinant Apo-NosL

Nucleotides 7479-8000 of NosL from *Achromobacter cycloclastes* were cloned as described previously into the pET-20b+ vector (Novagen Inc., Madison, WI) [161]. Residues encoding the native secretion signal and the N-terminal cysteine thought to be lipidated *in vivo* were omitted to aid in purification. ^{15}N and ^{13}C uniformly labeled recombinant NosL was produced by growing bacterial cell cultures in MOPS media enriched with [^{15}N]NH_4Cl (99% ^{15}N-enriched, CIL, Cambridge, MA) or [^{15}N]NH_4Cl /D-Glucose-$^{13}C_6$ (99% ^{13}C-enriched, CIL, Cambridge, MA) as the sole nitrogen and carbon sources respectively. Cultures were grown to an OD_{600nm} of approximately 0.7 at 37ºC (measured on a Varian 635 spectrometer), and then induced with 1mM IPTG and allowed to express at 20 ºC for 10 hours. Cells were harvested and the periplasmic protein extracted via osmotic shock. The protein was purified using DEAE-FF (Sigma Co., St. Louis, MO) anion exchange, followed by final purification on a Resource Q (Amersham Pharmacia Biotech Inc., Piscataway, NJ) column. Final NMR samples consisted of 1.0 mM protein in 100mM sodium phosphate, 1 mM d_{10}-DTT, 1mM EDTA and 10% D_2O at a pH of 6.5.

Expression and Purification of Recombinant Holo NosL

Expression of ^{15}N labeled Cu(I)-NosL was achieved using the same expression and purification conditions as the apo protein, except that once NosL was purified, the protein was dialyzed against argon-purged 100mM sodium phosphate, pH = 6.5, and then

dialyzed against this same buffer containing 200 μM $(CH_2CN)_4Cu(I)PF_6$ (Sigma Inc.) to introduce copper (I). Unbound copper was removed by subsequent dialysis against three changes of argon-purged buffer. The sample was then concentrated under argon. Copper content was assayed using the Cu+1 selective reagent 2,2'-biquinoline which verified a 1:1 stoichiometry of NosL to copper [170]. NMR sample conditions were identical to those used for the apo protein, but without DTT and EDTA and maintained under an argon atmosphere.

NMR Spectroscopy

All NMR spectra were acquired at 30 ^{o}C on a four-channel Bruker DRX-600 spectrometer, with a triple ^{15}N, ^{13}C, ^{1}H inverse resonance probe equipped with triple axis gradients. Data were processed and analyzed using NMRPipe [185], PIPP, and Xwinnmr version 3.1 (Bruker Inc.) software packages. Two dimensional ^{1}H-^{15}N HSQC spectra were acquired with a spectral width of 14.0 ppm in t_2 and 40.0 ppm in t_1, with the proton carrier frequency set at 4.7 ppm and the nitrogen carrier set at 116.0 ppm. Data were collected with 1024 complex points in t_2 and 128 complex points in t_1, using Waltz-16 for ^{15}N decoupling during data acquisition. Apodization was performed using a sine bell squared function shifted by 0.35 π radians in t_2, and a sine bell function shifted by 0.40 π radians in t_1.

Sequential and intra-residue ^{1}H, ^{15}N, and ^{13}C backbone and side-chain chemical shift assignments were extracted from a series of double and triple resonance NMR experiments (HNCA, HNCO, HNCACB, CBCA(CO)NH, C(CO)NH, HCC(CO)NH, ^{15}N-edited ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C-CT HSQC and HCCH-TOCSY) acquired with spectral

widths, number of points and carrier frequencies shown in Table 2. ^{1}H-^{13}C-CT HSQC and HCCH-TOCSY experiments were performed in D_2O. Side chain atoms were not assigned stereospecifically. The DIPSI scheme was utilized for ^{1}H decoupling during carbon evolution and the Waltz-16 scheme for ^{15}N decoupling during data acquisition. Similar apodization functions were used in all spectral dimensions, using shifted sine bell functions. ^{1}H, and ^{13}C chemical shift dimensions were indirectly referenced to DSS while ^{15}N chemical shifts were indirectly referenced to liquid ammonia.

Structure Calculations

NOE data collected in order to derive secondary structure and three-dimensional structural restraints were obtained from 3D ^{15}N NOESY experiments at mixing times of 100, 120, and 140 msec and from 3D HCHC-NOESY (MT=140msec) spectra acquired at 600 MHz. ^{15}N NOESY-HSQC data at a mixing time of 120 msec and ^{13}C NOESY-HSQC data at a mixing time of 120 msec were also acquired on a Varian 800 INOVA (800 MHz) instrument at the Environmental Molecular Sciences Laboratory (EMSL), sponsored by the Department of Energy's Office of the Biological and Environmental Research and located at the Pacific Northwest National Laboratory (PNNL).

Hydrogen bond restraints were extracted by monitoring the disappearance of amide resonances upon exchange with deuterium in the 2D ^{1}H-^{15}N-HSQC spectrum. To perform this experiment, a highly concentrated sample of apo NosL was rapidly diluted with $^{2}H_2O$-containing buffer and reconcentrated, followed by acquisition of the spectrum

after roughly 30 minutes. Inferred nitrogen-oxygen distances were given the lower distance bound of 2.5 Å and an upper distance bound of 3.3 Å. The hydrogen bonds

Experiment	Sweep Width/Carrier Offset (ppm) F1	F2	F3	Acquired Data (Complex pointes) F1	F2	F3	Mixing Time	Number of Scans
^{1}H-^{15}N HSQC	40/116	12.5/4.69		100	512			32
CBCA(CO)NH	36/116	67/42	12.5/4.69	48	58	512		16
HNCACB	36/116	67/42	12.5/4.69	40	58	512		32
HNCO	36/116	14/175	12.5/4.69	32	32	512		16
HNCA	36/116	23/52	12.5/4.69	64	64	512		32
^{15}N edited ^{1}H ^{1}H TOCSY	36/116	9.6/4.69	12.5/4.69	32	64	512	30 msec	16
HCCONH	36/116	10/4.69	12.5/4.69	36	68	512		32
CCONH	36/116	67/42	12.5/4.69	40	64	512		16
^{1}H-^{13}C-CT HSQC	60/40	12.5/3.0		512	1024			32
HCCH-TOCSY	60/40	6.51/3.0	8.0/3.0	116	128	512		16
3D ^{15}N NOESY	36/116	9.6/4.69	12.5/4.69	32	96	512	120 msec	16
3D HCHC-NOESY	60/40	8.0/3.0	6.7/3.0	116	128	512	140 msec	16
^{15}N NOESY-HSQC (EMSL/ 800 MHz)	36/116	12.5/4.69	13.0/4.69	30	90	512	120 msec	16
^{13}C NOESY-HSQC (EMSL/ 800 MHz)	24/35	12.5/4.69	12.5/4.69	36	106	512	120 msec	16

Table 2. NMR experiments and parameters used to obtain chemical shift assignments and structural restraints.

between the amide proton and oxygen were given a lower bound of 1.5 Å and an upper bound of 2.3 Å. The program TALOS generated phi and psi dihedral angle restraints based on observed chemical shifts for Cα, Cβ, CO, N and Hα [186].

^{13}C- and ^{15}N- edited NOESY spectra were processed by NMRPipe and imported into NMRView 5.0 [185, 187]. A list for each spectrum consisting of the NOE cross peaks picked by NMRView using the standard peak-picking algorithm was generated (see Appendix D for the first page of the cross peak file). Cross peaks that corresponded to intraresidue, sequential, and secondary structure information were manually assigned. A total of 112 NOE's were manually assigned in this fashion and entered into the NOE cross peak list. At this stage, peaks picked by NMRView but believed to be artifacts were deleted from the list. The resulting NOE cross peak files and chemical shift information (see Appendix C for the chemical shift/ppm.out file), along with the primary sequence of NosL were input into ARIA 2.0 (Ambiguous Restraints for Iterative Assignment), a software package designed to automatically assign NOESY data and calculate structures based on ambiguous distance restraints (see Appendix E for the NMRView-ARIA conversion file) [188].

ARIA proceeded in iterations of NOESY spectra calibration and assignment followed by calculation of an ensemble of 20 structures (see Appendix F for the project file used to run ARIA). For the structure calculation cycle, ARIA invoked the software suite Crystallography and NMR System (*CNS*) using the automatically assigned NOE-derived distance information, hydrogen bond data, and dihedral angle data as the empirical restraints for structure calculation [189]. In order to streamline calculations,

residues Met1-Pro34 and residues Glu161-Gly175 were omitted from the calculation, based on the observation that several of these residues lacked chemical shift information and that the remaining residues demonstrated little if any NOE information, an indication of flexible, disordered termini. For each iteration, the average distance information from the 7 lowest energy structures from the ensemble of 20 was used as the basis for NOE assignment and calibration in the next cycle. For each run, a total of nine iterations were performed followed by a final refinement in explicit solvent [190].

Within the NOE assignment protocol, unassigned cross-peaks were matched to each possible shift (within a frequency window of 0.05 ppm for proton and 0.50 ppm for heteronuclei) in the chemical shift list (chemical shift values within 0.05 ppm), generating several potential assignments. These assignments were then treated by ARIA as ambiguous distance restraints or ADR's [188]. Each ADR is an effective distance restraint that contains the contribution of all possible NOE pairs and the corresponding peak volumes based on the distance between the potentially interacting spins. This ADR is treated as a restraint with an upper and lower bound during structure calculation, although this restraint is actually a distance between more than two interacting spins.

After cross peaks were assigned and merged with each other in order to remove degenerate information, structure calculations were performed. In the first iteration of ARIA, input structures began in an extended polypeptide conformation, containing the bond lengths, bond angles and nonbonded (or van der Waals) interactions generated by *CNS* from the given primary sequence of the protein and containing the idealized geometry information. For the simulated annealing protocol in *CNS*, a combination of

Cartesian and torsion angle dynamics and energy minimization was utilized to generate structures [191]. Torsion angle space sampling, in which bond angles and bond lengths remain fixed while the bond between the atoms is rotated, was incorporated into the molecular dynamics because it is a more efficient search of conformational space than Cartesian and distance geometry methods.

The simulated annealing structure calculation protocol run by *CNS* consisted of several torsion angle dynamics stages. In the first stage, the extended polypeptide is subjected to 90 ps of torsion angle molecular dynamics at 10,000K in order to randomize the positions of the atoms in the polypeptide and to allow a search of torsion angle space stein. The protein was then slowly "cooled" from 10,000K to 1000K, during which the protein sampled torsion angle conformational space for another 45 ps. This was followed by a second cooling stage from 1000K to 50K for 90 ps. During the first stage, the NOE hydrogen bond, and dihedral energy terms are all weighted to 10, in the first cooling stage the parameters are ramped from 10 to 50, and in the second cooling stage, the energy factors are weighted to 50.

After this first iteration, an ensemble of 20 structures was generated and the 7 lowest energy structures then selected and analyzed for restraint violations. Restraints that systematically violated (i.e. violate in > 85% of the ensemble) were removed, and a restraint set with fewer assignment possibilities was used for a second round of structure calculation [192, 193]. Also, restraints contributing to ADR's corresponding to large distances in the structures were discarded, increasing the efficiency of calculation [188]. This entire protocol starting from NOE assignment to structure calculation was repeated

using the modified ADR's from the previous iteration, and the entire cycle repeated for a total of 9 iterations. In the last step, the structures were refined in an explicit shell of water [190].

After each run in ARIA, the resulting structures were manually inspected and the remaining violations analyzed. At this step, large violations were removed (those yielding energies of roughly > 100 kcal/mol), and ARIA restarted, repeating the process using the assignments generated from the first run. For the final ARIA calculation, 100 structures were generated and the 20 lowest energy structures selected for analysis using the programs Quanta (Molecular Simulations Inc.), MOLMOL [194], Grasp [195] and PROCHECK-NMR [196]. Structural homologs were identified by submitting the three lowest energy structures to the DALI server (www..ebi.ac.uk/dali) as a web request [197].

RESULTS FOR NOS L

Expression and Purification of Apo and Holo Nos L

Expression of recombinant apo NosL lacking the lipobox signal peptide resulted in milligram quantities of labeled protein sufficient for structure determination. Expression levels of roughly 3-4 mg/L culture for the ^{15}N labeled protein and 1-2 mg/L culture for the $^{13}C/^{15}N$ samples were obtained, requiring the growth of 6L of culture per sample. Uniform isotopic enrichment, while required to replace the normally spectroscopically silent ^{14}N and ^{12}C nuclei and thereby allow the use of the full manifold of heteronuclear experiments required for structure determination, resulted in relatively poor yields, especially for ^{13}C samples. An ^{15}N-labeled sample of holo NosL was also produced by anaerobic dialysis against buffer containing a cu(I) complex. While holo NosL had previously been produced by supplementing *Escherichia coli* cultures with $CuNO_3$, this report is the first describing the post-expression incorporation of copper(I) [161]. A 1:1 stoichiometry of copper(I) to NosL was verified by interaction with the copper(I)-selective reagent 2,2'-biquinoline.

Sequential Assignment and Secondary Structure of Apo NosL

The two-dimensional ^{1}H-^{15}N-HSQC experiment correlates the backbone amide protons with the attached ^{15}N nuclei, producing the best resolved type of resonances within a protein and therefore allowing preliminary assessment of the feasibility of structure determination. Spectra of this type (shown in Figure 26) performed on apo

NosL exhibited very large chemical shift dispersion of the amide protons (Δ^1H_N of ~4.0 ppm), indicating that the protein is well-folded and rich in β strand content. Despite the considerable proton chemical shift dispersion, many resonances were overlapped, particularly between 7.6 and 8.4 ppm in the proton dimension, as shown in the enlarged view in Figure 27. Without including NosL's fourteen proline residues, which are silent in the 2D ^1H-^{15}N-HSQC experiment, 11 out of a total of 161 amide resonances remain unassigned, corresponding to a sequential assignment of ~93%. Of the eleven unassigned H_N resonances, two correspond to Met 1 and Asp 2, while the others include the missing

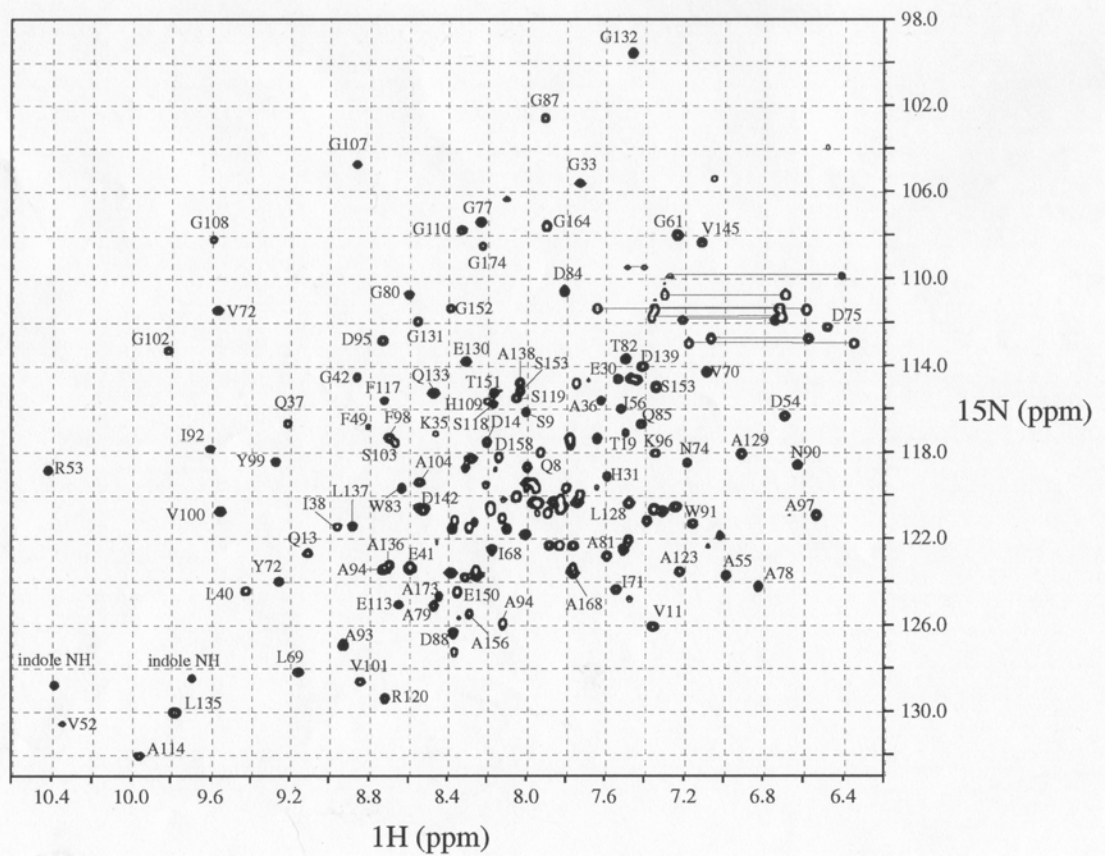


Figure 26. Assigned ^1H-^{15}N HSQC spectrum of apo NosL at 303K collected on the Bruker DMX600 MHz instrument.

$^{1}H/^{15}N$ amides of Thr 16, Leu 20, Tyr 23–Leu 29, and Ser 50. Residues missing amide resonances are most likely involved in conformational exchange processes on the NMR timescale (msec) that broaden signals beyond detection and therefore prevent reliable assignment.

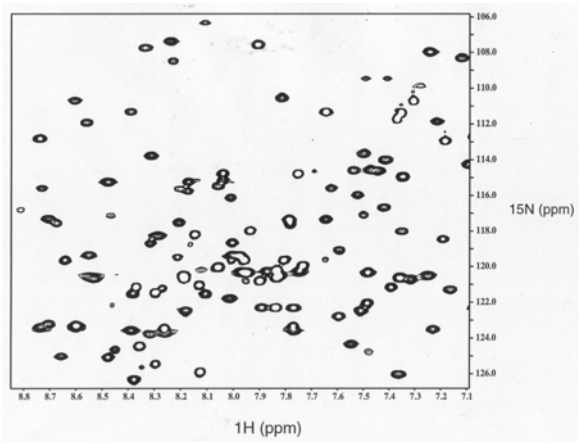


Figure 27. Enlarged region of the ^{1}H-^{15}N HSQC spectrum of apo NosL, showing the overlap of resonances within this chemical shift region.

A combination of 2D and 3D heteronuclear NMR experiments were used to sequentially assign the backbone resonances of apo NosL, the first step in structure determination. The pulses and delays in the CBCA(CO)NH experiment [198] serve to transfer magnetization from the alpha and beta carbon nuclei to the amide proton of the succeeding residue via the intervening ^{13}CO nucleus, and therefore provide sequence-specific connectivities between these nuclei. The HNCACB experiment correlates the amide proton to intra- and inter-residue Cα and Cβ nuclei of the preceding residue [199]. Using the spectra from these two experiments, stretches of amide resonances were

matched by the chemical shifts of Cα and Cβ atoms to interresidue Cα and Cβ chemical shifts, generating sequential stretches of neighboring pairs as depicted in Figure 28, interrupted only by proline residues which lack amide protons. The Cα and Cβ chemical shifts were then assigned an amino acid-type based on empirically-determined Cα and Cβ chemical shift ranges [200]. For example, threonine residues have Cβ resonances shifted far downfield of the other Cβ frequencies, resonating at 69.8 ppm, which distinguishes them from other amino acids allowing ready amino acid-type assignment of this residue. Glycines, on the other hand, exhibit upfield-shifted Cα frequencies between 43 and 46 ppm and possess no Cβ, allowing facile assignment (other useful residues with diagnostic chemical shift patterns were A, S, G, I, V and P). Building on tentative identification of the amino acid type of a particular amide resonance, stretches of sequences were generated, until the primary sequence of the protein could be aligned with these segments. HNCO and HNCA experiments, which correlate amide protons with the attached ^{15}N and the preceding Cα and CO atoms, respectively, verified the sequence connectivities and allowed measurement of carbonyl chemical shifts [201].

Once the sequential backbone assignments were made, side-chain assignments could then be derived for each residue. Many side-chain assignments were achieved using the ^{15}N-edited ^{1}H-^{1}H TOCSY experiment, which correlates the amide proton with intraresidue spin-system resonances. The C(CO)NH and HC(CO)NH experiments, which correlate all of the aliphatic side-chain carbon or proton residues, respectively, with the amide resonance of the succeeding residue, allowed assignment of most sidechain nuclei [202]. HCCH-TOCSY experiments [203], used in concert with ^{1}H-^{13}C

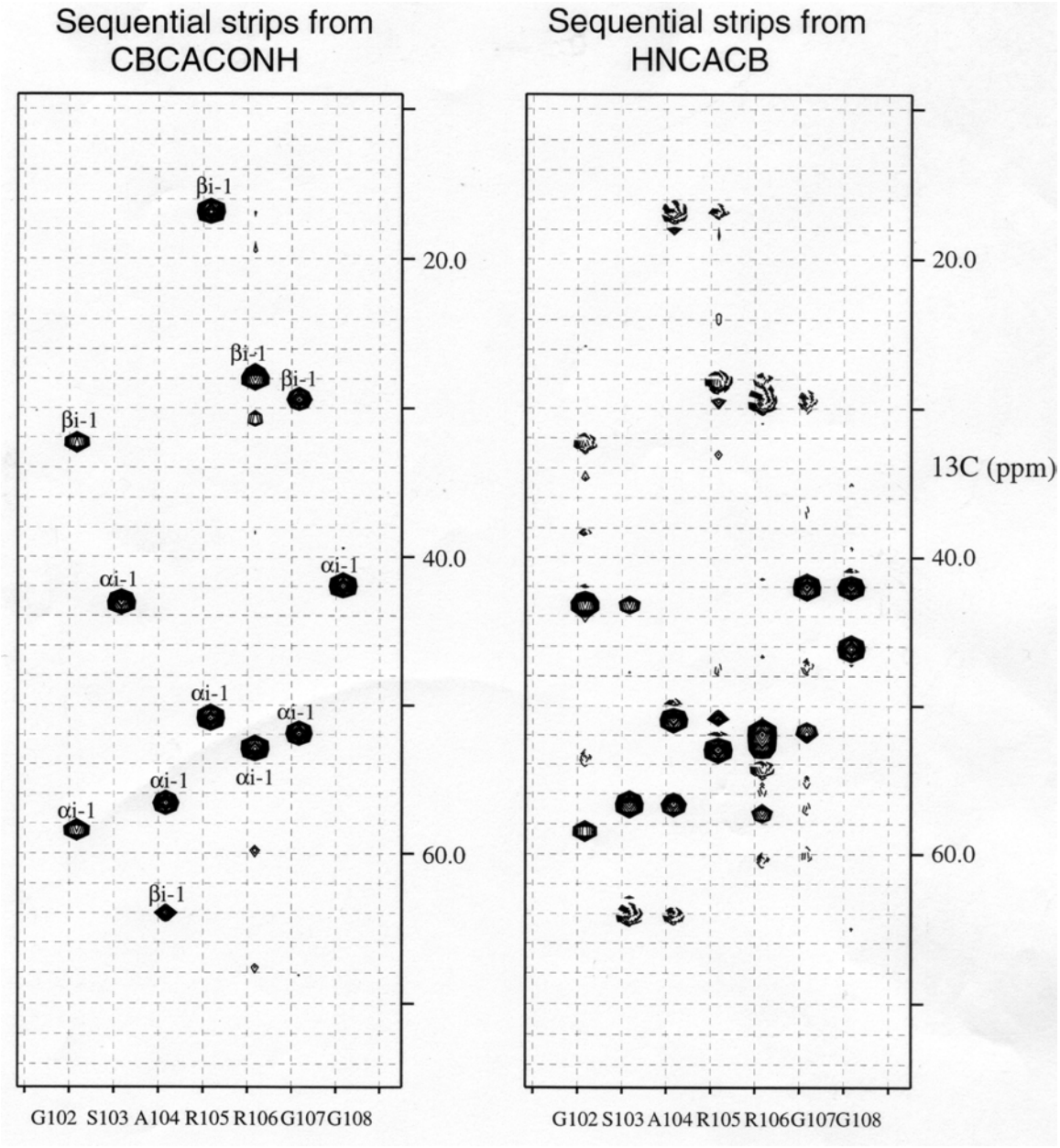


Figure 28. Sequential stripplots from the CBCACONH (shown on left) and the HNCACB (shown on right) experiments. Strips corresponding to amide resonances as labeled were extracted from the 3D experiments and arranged according to the primary sequence. Plot generated by the plot_sequence program (D. Garrett, NIH).

CT HSQC spectra [204] and performed in D_2O, allowed correlation of the Cα and Cβ nuclei with rest of the side-chain proton and carbon resonances, completing side-chain assignments.

In this manner, the Cα/Cβ and CO chemical shifts were determined for 166 and 141 residues of the protein's 175 amino acids, respectively. ~94% of the Hα/Hβ resonances were assigned and complete assignment of sidechain ^{13}C and ^{1}H resonances were made for the vast majority of the residues that were identified. Backbone and sidechain (^{1}H, ^{15}N ^{13}C) resonance assignments have been deposited in the BioMagResBank in Madison, (WI, USA) under the accession number 5595 [205] and are tabulated in Appendix A.

Aside from providing amino acid-type information, chemical shift values were also used to predict secondary structure. Hα, Cα, Cβ, and CO chemical shifts, in particular, are strongly dependent on secondary structure conformation [186]. Statistical analysis of existing structures along with directly measured shifts from peptides representing random coil values showed that, depending on the type of secondary structure, chemical shifts were found to be consistently either upfield or downfield of random coil shifts [206]. Carbonyl and Cα carbons resonate upfield when in β strands and downfield when in helices. Alpha proton chemical shifts are typically lower in value (upfield) compared to random coil shifts in helices and higher in value when in beta strands. Cβ carbons are shifted higher or downfield of random coil shifts when in β sheets, but cannot be used to predict helices [206]. The deviation of the $^{13}Cα$, $^{13}Cβ$ and Hα chemical shifts of apo NosL from random coil chemical shift values are shown in

Figure 29 and were qualitatively used in concert with other data to preliminarily assign secondary structure elements in apo NosL [207].

Chemical shift information was also used in a more quantitative manner to generate psi and phi dihedral angle restraints using the software package TALOS [186]. TALOS uses Cα, Cβ, Hα, CO, and N chemical shift information to predict the phi and psi angles for each residue in the protein. Along with chemical shift information, the program includes the amino acid type of the preceding and following amino acids to search a database of 78 proteins for sequence and chemical shift homology. Once matches for each triplet are found, phi and psi dihedral angles derived from each crystal structure are compared, and, if similar, yield a prediction for the average dihedral angle along with a predicted distribution for direct use as restraints. 91 psi and 91 phi dihedral angle restraints were generated in this manner. Hydrogen/deuterium exchange experiments were used to determine amide protons in slow exchange and therefore most likely to be involved in hydrogen bonds, complementing the TALOS, CSI, and NOE data (see Figure 29).

To complete the secondary structure identification, NOE correlations, which indicate through-space magnetization transfer, were obtained from the 3D ^{13}C HCCH-NOESY and the 3D ^{15}N-NOESY experiments. Along with sequential information, NOE's also provide crucial interactions that accompany stable helical and beta strand elements, as well as the strand topology. In particular, valuable NOE's used for identifying alpha helices are the $d_{\alpha N}$ (i, i+3), $d_{\alpha\beta}$ (i, i+3) and $d_{\alpha N}$ (i, i+4), d_{NN} (i, i+1), and d_{NN} (i, i+2) correlations [208]. The interactions most useful for designation of

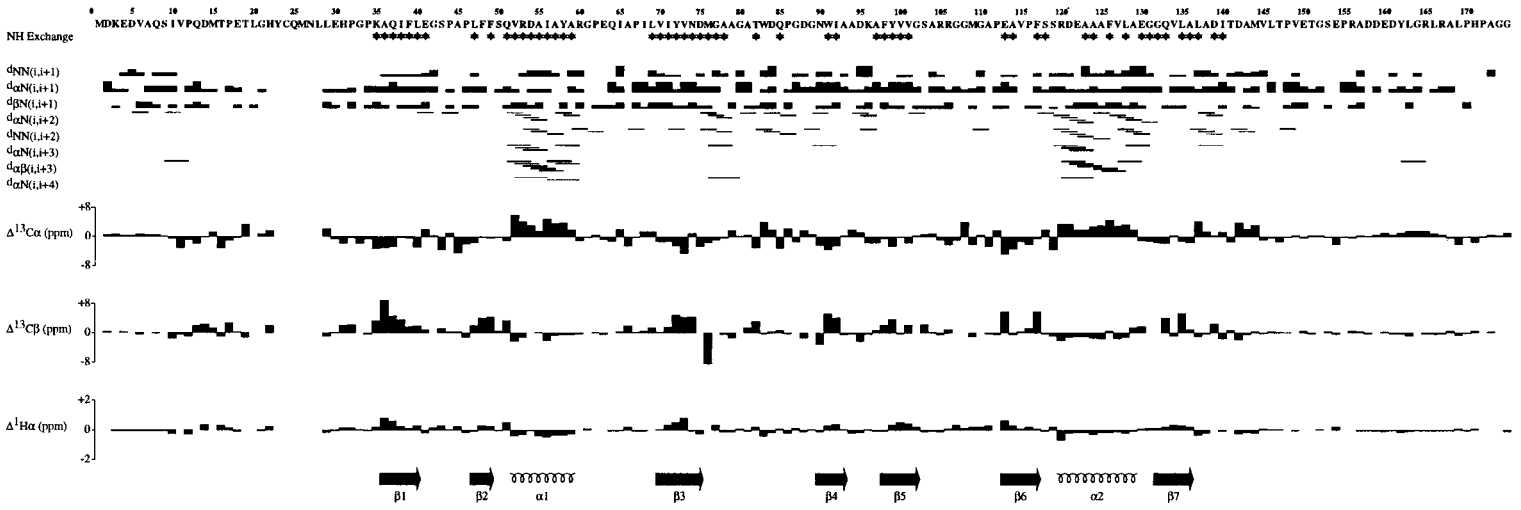
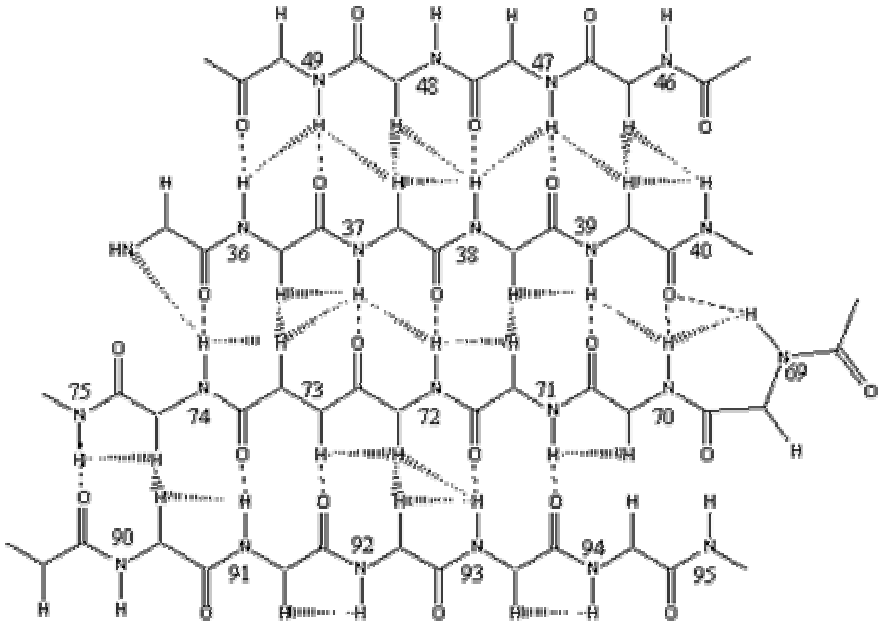


Figure 29. Summary of amide exchange data, patterns of sequential and short-range NOE's and secondary chemical shifts for residues 1-175 of apo NosL. The height of the box of the NOE connectivities reflect the intensity of the cross-peak. Differences in ppm in $^{13}C\alpha$, $^{13}C\beta$, and $^{1}H\alpha$ relative to random coil values at 25° C are shown. The last row shows the consensus secondary structure elements.

antiparallel beta sheets were the $d_{\alpha\alpha}$ (i,j), $d_{\alpha N}$ (i,j), and d_{NN} (i,j). Some of these diagnostic NOE correlations are shown in Figure 29, along with the secondary structure determined from short-range NOE's, hydrogen bond information, and chemical shift values.

The structural topology of apo NosL consists of two antiparallel beta sheets, two α-helices, two 3_{10} helices, and two relatively lengthy, disordered termini. Beta sheet 1 is made up of four strands extending from residues Ala36 to Leu40 (strand β1), Leu47 to Phe49 (strand β2), Val70 to Asn74 (strand β3), and Asn90 to Ala93 (strand β4) (see Figure 30 A). Strand 3 of sheet 1 has a beta-bulge centered at Leu40, the carbonyl of which interacts with the amide proton of residues Ile68 and Leu69. The other beta sheet, sheet 2, is composed of three strands consisting of residues Phe98 to Gly102 (strand β5), Glu113 to F117 (strand β6) and finally Gly132 to Ala136 (strand β7) (Figure 30 B). The protein contains two α helices, one nine-residue helix (helix α1) encompassing residues Val52 to Ala59 following strand 2 after a short linker region, and helix α2 between strand 6 and strand 7, consisting of residues Arg120 to Ala129. The N-terminus and C-terminus of apo NosL, extending from residues Met1-Pro34 and Val145-Gly175 respectively, appear to be disordered, consistent with the lack of NOE's, hydrogen bonds and chemical shift deviations in these regions. The secondary structure elements of apo NosL are punctuated by loop regions of varying size and encompassing more residues than the number of amino acids involved in regular secondary structure elements. Residues Leu137-Ile140 and residues Ala94-Ala97 each form a 3_{10} helix, as determined by the characteristic $d_{\alpha N}$(i, i+2) NOE's and hydrogen bonding pattern.

A.



B.

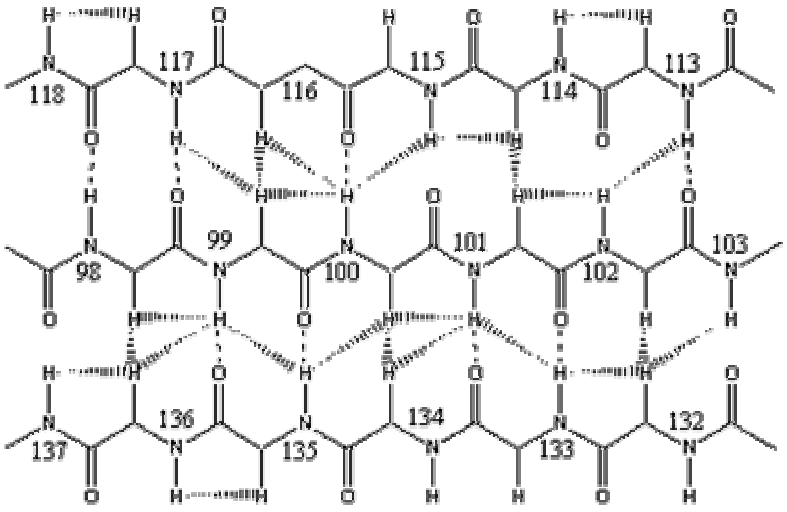


Figure 30. A. and B. Strand topology of Apo NosL. A. shows the first 4-stranded antiparallel beta sheet, B. shows the second three-stranded antiparallel beta sheet. Hydrogen bonds are shown in dotted lines and NOE interactions are shown as dashed wedges.

Apo NosL Solution Structure

Once the secondary structure elements of apo NosL were identified, the global fold of apo NosL was then solved using the software package ARIA 2.0 [188]. NOE's providing distance restraints for structure calculation were derived from the HCHC-NOESY spectrum acquired at 600 MHz at MSU and the ^{15}N NOESY-HSQC spectrum acquired at EMSL, sponsored by the DOE and located at the Pacific Northwest National Laboratory (PNNL) with the assistance of Dr. David Hoyt. Secondary structure NOE's and intraresidue NOE's were then partially manually assigned using NMRView and then imported into ARIA along with the primary sequence of the protein and the chemical shifts. A total of 112 manual assignments were entered into ARIA in this manner. ARIA then automatically assigned the remaining NOE cross-peaks utilizing the concept of ambiguous distance restraints. The use of ambiguous distance restraints for NOE assignment allows the determination of NMR structures without unambiguous NOE data and is based on the observation that most erroneous assignments and noise peaks are inconsistent with the rest of the fold and therefore can be filtered out in subsequent iterations [188, 192]. ARIA was particularly useful for the structure determination of apo NosL because of the lack of sequence similarity of this protein to others of known structure and because very few unambiguous NOE's could be assigned based on chemical shift values alone. In addition, many resonances were overlapped due to large regions of disorder in the protein, making the use of multiple assignments to the same peak beneficial [193]. For the purposes of structure calculation, only residues Lys35 to Glu160 were modeled in order to greatly simplify the list of possible chemical shift

values and therefore the potential assignments made by ARIA. Truncation of the sequence also made convergence of the structures more efficient in both manual and ARIA calculations. As described above, the majority of the region left out of the calculation exhibited few non-sequential NOE's suggesting that they most likely contribute little structural information, while residues Thr19-Leu20/ Tyr23-Leu28 could not be assigned and therefore would not be recognized by ARIA. The manually assigned and automatically generated ambiguous distance restraints assigned by ARIA were utilized in combination with hydrogen bond data and dihedral angle data for structure calculation by *CNS*. For the final apo NosL structure calculation, a total number of 1677 NOE's (780 intraresidue restraints and 897 interresidue restraints), 146 dihedral angle restraints and 76 hydrogen bond restraints were used as empirical data.

The backbone rmsd for secondary structure elements of apo NosL is 0.62 Å, and 1.61 Å overall. The ensemble contains an average of 5 NOE violations greater than 0.5 Å, making up less than 0.3% of the total number of NOE's contained in the structure. Ramachandran mapping (see Figure 31 for the Ramachandran plot of the lowest energy structure of the ensemble) and PROCHECK-NMR analysis demonstrated that less than 2.3% of residues fall within disallowed regions of phi/ psi angle space. The residues in disallowed regions (Arg106, Arg156, and Ile65) are located in loops or in ill-defined regions of apo NosL. The conformational and energetic analyses on the ensemble of 20 structures performed by PROCHECK-NMR are given in Table 3. The backbone atoms of the ensemble of structures are displayed in Figure 33, which allows visualization of the rmsd variation throughout different regions of the protein.

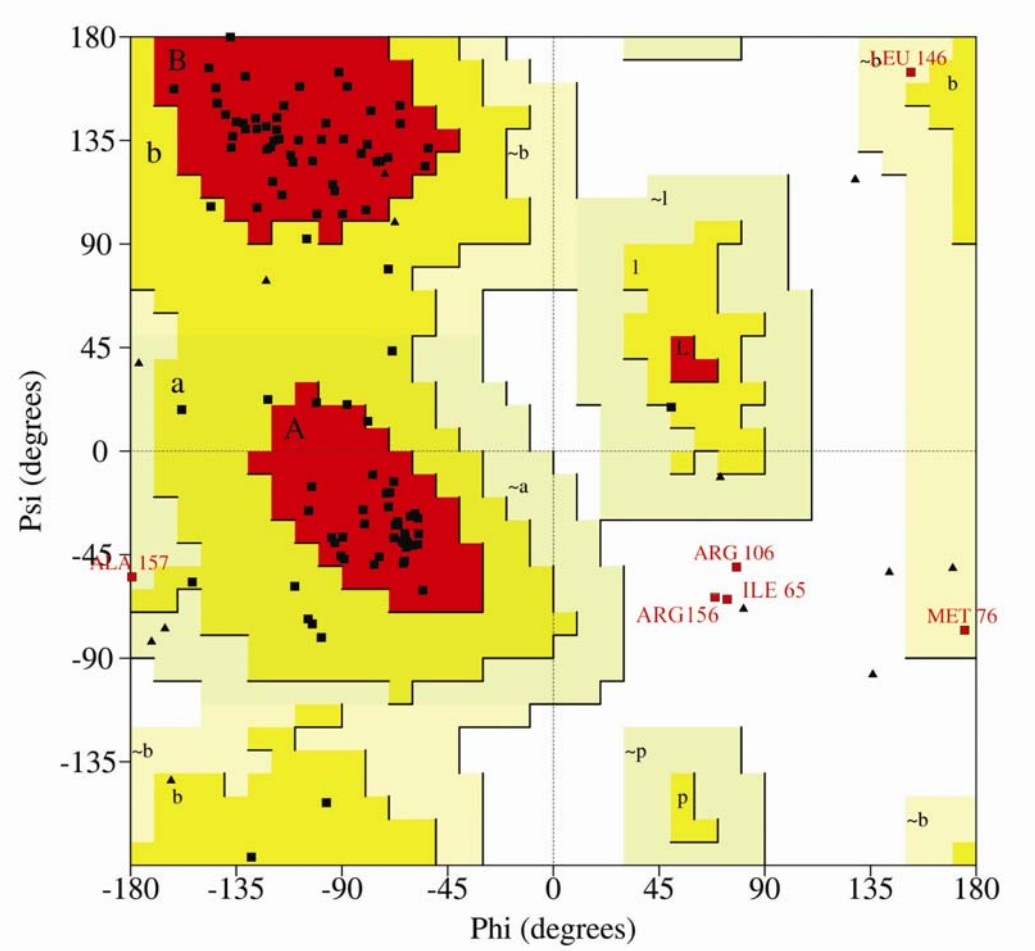


Figure 31. Ramachandran plot of apo NosL. Red regions (A,B,L) correspond to highly favored regions, yellow regions (a,b,l,p) to additionally allowed regions, tan regions (~a,~b,~l,~p) to generously allowed regions and white to disallowed regions. Residues in disallowed and generously allowed regions are labeled in red and correspond to residues in loops (Ile65 and Arg106) or in the disordered C-terminus (Arg156). The Ramachandran map was generated for the lowest energy structure from the ensemble and plotted by PROCHECK-NMR.

Experimental Restraints used for Structure Calculations

Total NOE distance restraints	1677
Intraresidue NOE restraints	780
Interresidue sequential NOE's (|i-j| = 1)	320
Interresidue medium-range NOE's (1 < |1-j| < 5)	201
Interresidue long-range NOE's (|i-j| > 4)	376
Dihedral angle restraints	182
H-bond restraints	76

Structural Statistics

Root-mean-square-deviations relative to average structure (Å)	
Backbone (C^{α}, C', N) atoms in 2^{nd} structure	0.62 ± 0.13
Backbone all residues	1.61 ± 0.74
Heavy atoms in 2^{nd} structure	1.00 ± 0.15
Heavy atoms in all residues	2.03 ± 0.77
Restraint violations	
nOe distances with violations > 0.5Å	5.2 ± 1.5
dihederal with violations > 5^{o}	0.0 ± 0.0
H-bond with violations > 0.5Å	2.3 ± 0.6
Final energies from simulated annealing (kcal/mol)	
F_{vdw}	-992 ± 29
F_{ele}	-4323 ± 74
Deviation from idealized geometry	
Bonds (Å)	0.0044 ± 0.001
Angles (deg)	0.60 ± 0.02
Impropers (deg)	1.84 ± 0.11
Ramachandran analysis (% of all residues)	
Residues in most favored regions	81.1
Residuesin additional allowed regions	13.2
Residues in generously allowed regions	3.4
Residues in disallowed regions	2.3

Table 3. Restraints used in structure calculations and statistical analysis of the 20 lowest energy structures. Rmsd's calculated for core regions of defined secondary structure include residues belonging to beta strands β1, β2, β3, β4, β5, β6, β7, and helices α1 and α2.

The overall structure of apo NosL consists of two structurally similar and relatively non-interacting domains comprised of an α + β topology in which each domain contains a relatively novel ββαβ motif (see Figure 32). The second and third β strands within each motif are connected by an α-helix via a right-handed crossover, positioning the helix against the face of the β sheet as shown in Figure 32. The first three β strands and α helix of each motif superpose with an rmsd of 1.81 Å, despite the fact that the sequence identity of the aligned residues within the secondary structure elements is only 11% and yield a Z score of only 1.7 from a DALI pairwise alignment (see Figure 33). Not surprisingly, loops β1/β2 and β5/β6 differ significantly both in length and in conformation. Despite the low Z score and the low percent of sequence similarity between the two domains, it seems likely that the two domains are the result of an ancient gene duplication event, but have diverged significantly such that functional conservation is not a given [209].

Unlike many α + β structures, the face of each β sheet of NosL interacts with only one helix, lacking a second helix on the opposite face that characterizes the flavodoxin fold. In fact, the only proteins found by extensive database searches (VAST, SCOP, DALI, and FATCAT) to contain the ββαβ topology of the two domains of NosL are the actin-binding domain of cofilin and cofilin-like proteins, which contain an additional parallel β strand and three other α helices that pack on each face of the beta sheet, differentiating this protein from the NosL domain [210]. The first three elements of the domain (ββα) are topologically similar to the zinc finger fold adopted by nucleic-acid binding proteins, but lack the C2H2 zinc coordination ligands Cys and His [211]. Given

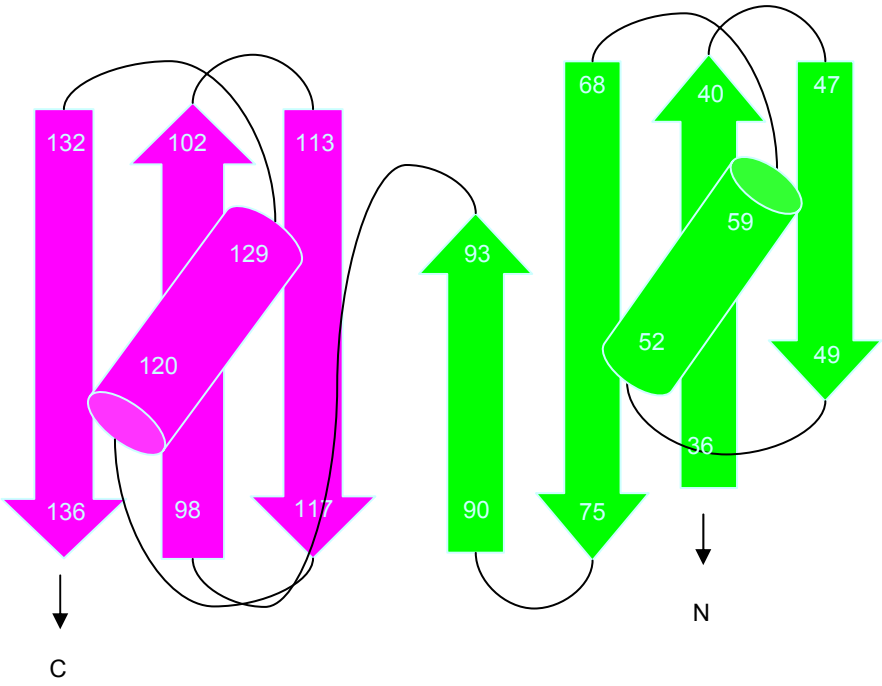


Figure 32. Topology of NosL. The two structurally homologous domains of NosL are color-coded in green for the first domain and magenta for the second.

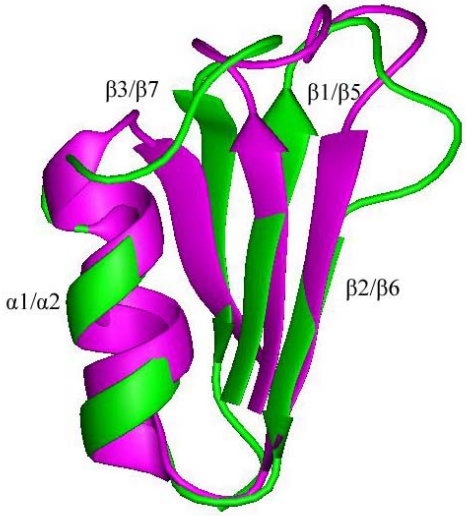


Figure 33. Superposition of the two domains of NosL. Domain 1 (residues Ala36-Asp75) is displayed in magenta with strands β1, β2, β3 and helix α1 labeled while domain 2 (residues Phe98-Ala136) is shown in green with strands β5, β6, β7 and α2 labeled. The backbone secondary structure elements align with an rmsd of 1.81 Å, varying chiefly in loop β1/β2 and loop β5/β6.

the lack of significant sequence and structural similarity of both the zinc finger and actin-binding domains to NosL, these proteins are most likely unrelated or distantly related to the domain.

The two domains are connected by a short 3_{10} helix that forms a hinge between them and that projects outwards into solvent (see Figure 35 B). The two domains are almost perpendicular to each other and rotated roughly 80º apart, forming a twisted butterfly-like configuration (see Figure 35 C). Helix α1 forms the interface between the two lobes, resulting in a large cleft that bisects the bulk of the protein and is solvent-accessible (see Figure 35 A).

One other short 3_{10} helix is found on the region C-terminal to sheet 2 in the otherwise poorly defined region. All of the four loops in apo NosL are solvent-exposed except for several residues in the loop encompassing residues Met76-Gly89 (Loop β3/β4), in which several of the amide protons are deuterium exchange-protected and at least two involved in hydrogen bonds between Thr82 and Glu85. Due to the relatively independent relationship of the two subdomains, the structure of apo NosL exhibits a relatively small region of interaction between them, and, along with the many loops, generate an irregular surface with many available sites for potential ligand and protein-protein interactions (see Figure 36 A and B).

The C-terminal end of helix 1 and one of the faces of sheet 1 interact with sheet 2 in an edge-on manner, partially burying the N-terminal end of the helix between the two sheets. The sidechains of Leu137, Val52, Ile71, Ile56, Phe117, Ala36, Ile38, and Leu47 make up the network of hydrophobic interactions that hold the two subdomains together.

The orientations of the two sheets and α1 create a wide, deep pocket in the center of the protein (see Figure 36 A and B) that is solvent-accessible. The residues lining the pocket and constituting the hydrophobic core of NosL are semi-conserved residues (Phe117, Leu137, Ile71, Val73, Ala36). Unlike α1, α2 is on the periphery of the structure,

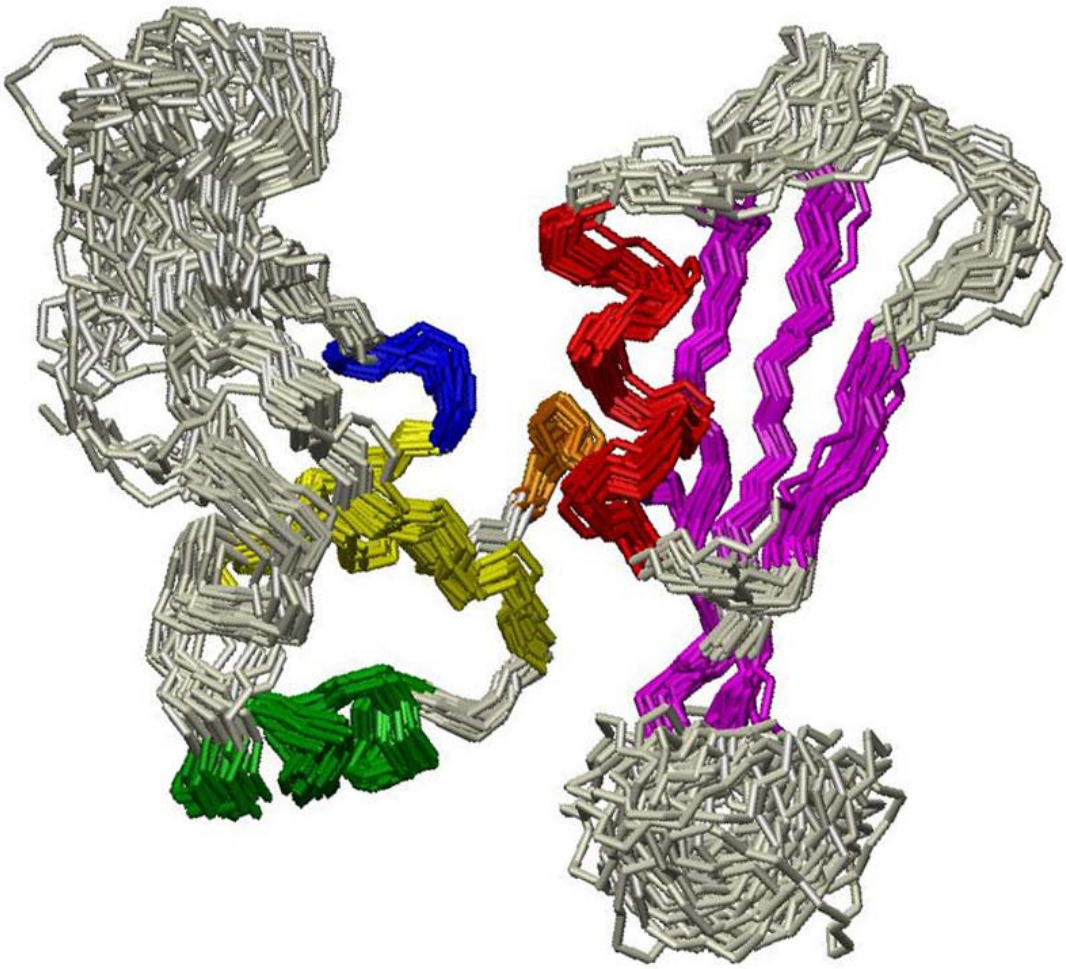


Figure 34. Backbone wire diagram of the ensemble of 20 lowest energy structures out of 100. Sheet 1 is colored in purple, helix 1 in red, sheet 2 in yellow, and helix2 in green. The two 3_{10} helices are colored in orange and blue.

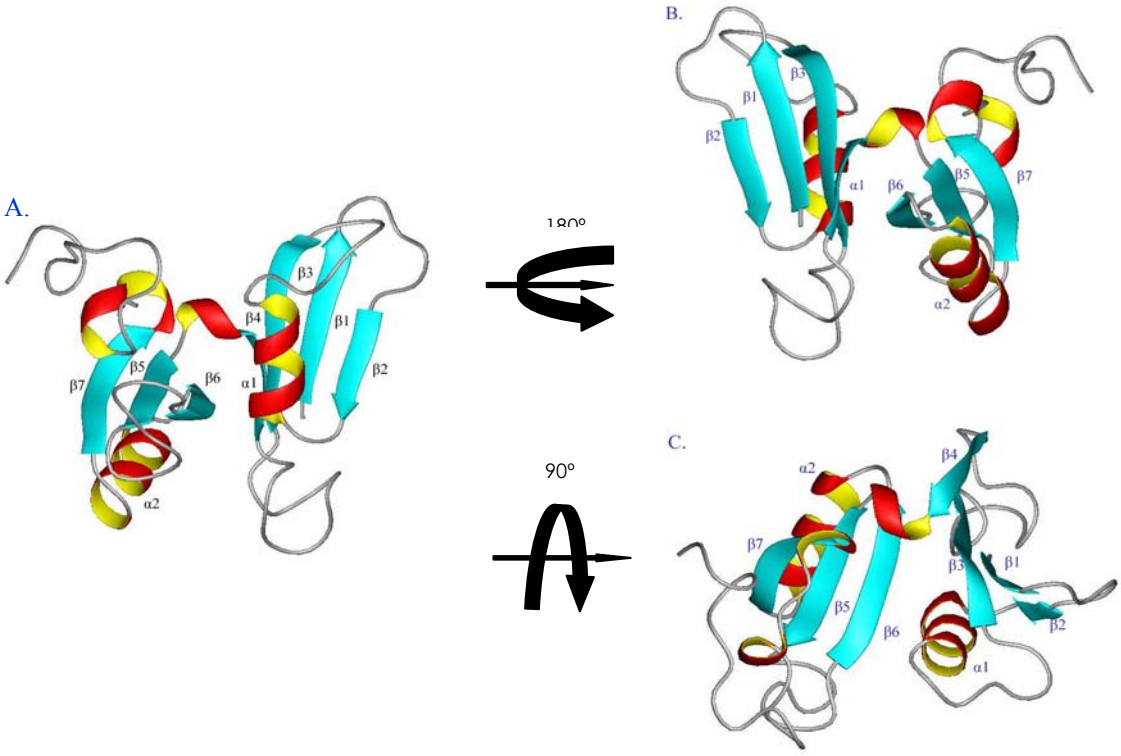


Figure 35 A, B, and C. Ribbon diagrams of apo NosL from different perspectives. Beta sheets are shown in cyan, helices in red and yellow. Beta strands β1, β2, β3, β4,β5, β6, and β7, and helices α1, and α2 as described in the text are labeled. Figure B. shows the view relative to ribbon Figure A. rotated by 180º as shown, in which the 3_{10} "hinge" is projecting towards the viewer. Figure C. has been rotated 90º as shown relative to the first ribbon figure and shows the perpendicular relationship between the two beta sheets.

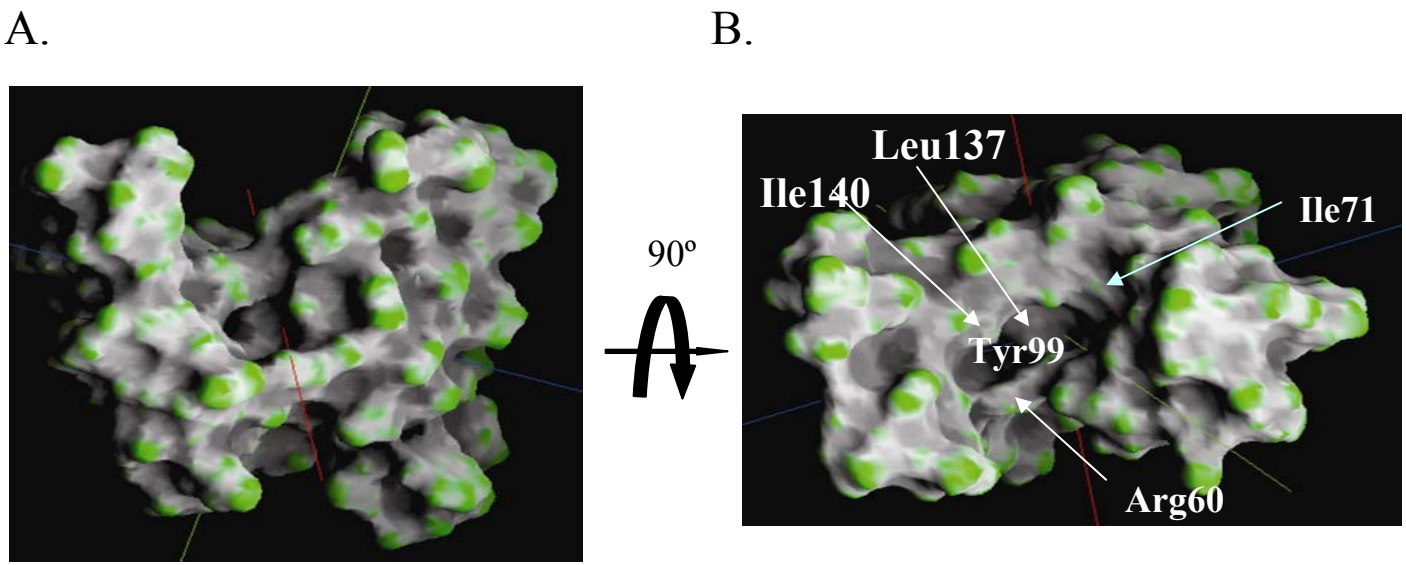


Figure 36 A. and B. Molecular surface figures of apo NosL. Figure A. shows the view of the large cleft between the two subdomains and the irregular surface created by the multiple solvent-exposed loops. Figure B. shows the view down the deep pocket created by the interaction between α1, β1, β3, and β6. Figures were generated by Grasp and contoured according to surface curvature to emphasize depth.

interacting exclusively with sheet 2 via Ala123 and Phe126 from the helix and Val100 from sheet2.

The residues surrounding the channel give a positive electrostatic surface potential due to the presence of Arg60, Arg53, and Arg106 while the inside of the cavity itself is relatively neutral (see Figure 37 A and B). The rest of the molecule exhibits an overwhelmingly negative surface potential especially in the C-terminus due to the clustering of acidic residues Asp142, Asp139, Asp158, Glu150, Glu154, Glu160, Asp159 consistent with the low pI of the protein. Amid the pronounced negative electrostatic potential of the surface, Lys96, Arg105, Lys35, and Arg120 do provide some accessible positive charges.

The C-terminal residues after the short 3_{10} helix encompassing residues Leu137, Ala138 and Asp139 appears to form a small hydrophobic cluster composed of the sidechains of Ile140, Leu137, and Tyr99, perhaps explaining the conserved nature of these residues. These contacts extend over the top of sheet 2 in an irregular conformation. On the end of sheet 1 furthest from the interior of the structure, the beta-bulge formed by residues Ile 68 and Leu 69 projects towards α1, followed by a long loop consisting of residues Arg60-Pro67 (Loop α1/β3). Despite the length of the loop, its conformation appears to be relatively well-constrained like that of the shorter loop Glu41-Pro46 (Loop β1/β2), presumably in part due to the presence of two proline residues per loop. Pro62 is conserved in all NosL sequences and Glu63 is highly conserved, suggesting that the conformation of this loop is important for activity, perhaps via interactions with other proteins due to the solvent-exposed nature of these residues.

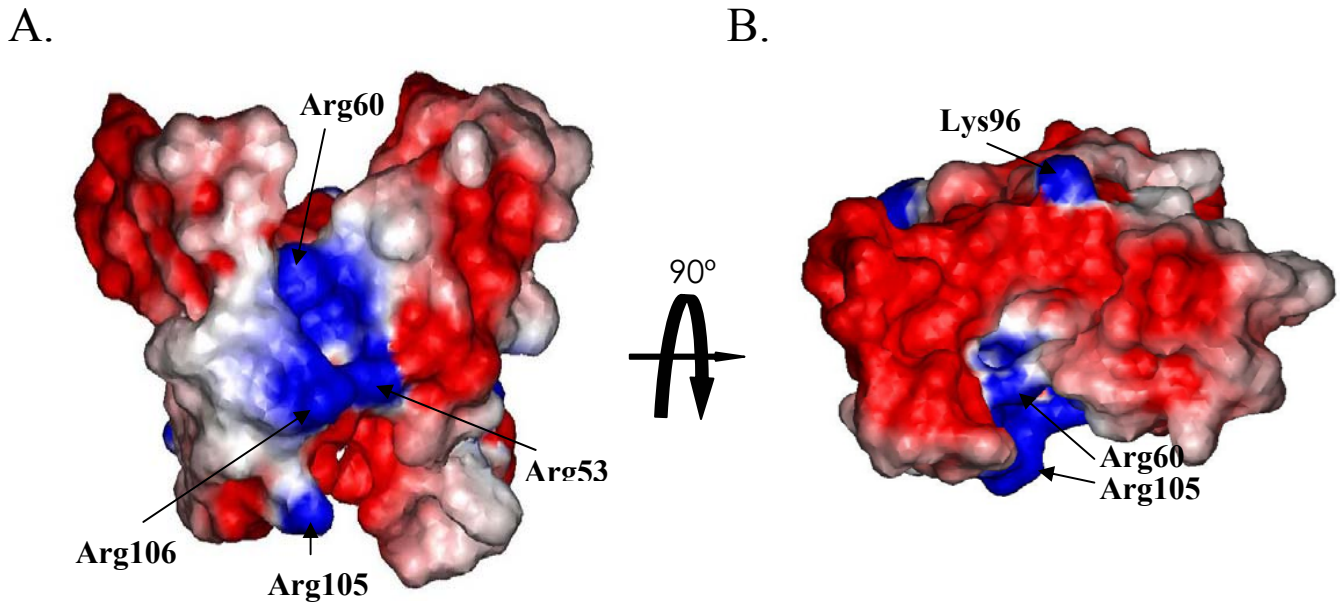


Figure 37 A. and B. Electrostatic surface potential of NosL from two views. Figure A. shows the basic patch surrounding the pocket created by residues Arg106, 53 and 60. Figure B. shows the view down the cleft. Figures were generated by MOLMOL.

On the opposite end of sheet 1 between strands β3 and β4, the 13 residue loop (Loop β3/β4) consisting of Met76 to Gly89 is restrained by NOE's and hydrogen bonds between Thr82 and Gln85, resulting in a short beta sheet-like interaction. The amide protons of Met76, Gly77, and Ala78 are also deuterium exchange-protected, presumably sequestered from solvent by the irregular structure adopted by this loop. The rmsd of this loop is the greatest, and in the highest energy structure, the loop is flipped by roughly 180º away from α2. It is unclear whether or not the alternative loop conformation is due to lack of available restraints within this region or due to genuine sampling of both locations.

The hinge region joining the two beta sheets is composed of a 3_{10} turn in which the slowly exchanging amide of Ala97 is most likely involved in a hydrogen bond with the carbonyl of Ala94, both of which are highly conserved residues. This hinge could conceivably flex to bring to the two subdomains of NosL in closer contact.

The fourth solvent-exposed loop (Loop β5/β6) consists of residues Ser103-Pro112, and contains one of the few motifs present in the family of NosL sequences, G/A-G/A-MG. This loop extends out towards the C-terminal helix-like turn near residues Met144 and Val145, leaving Met109 highly solvent exposed. Interestingly, two strictly conserved residues, Gly131 and Gly132, are in proximity to residues in this loop, particularly Ser103. The first three residues of the only other conserved motif, PGPK-G/A, were not calculated in these structures, but residues Lys35, Ala36, and Gln37 can be visualized on β1 near the interaction zone between α1 and sheet 1.

Structural Homologs to Apo NosL

A search of the Protein Data Bank using the DALI server identified one structural homolog to apo NosL with a significant similarity score (Z=4.0) [197, 212]. This protein, MerB, was aligned with apo NosL using MOLMOL (see Figure 38 A). The rmsd between the two proteins is 3.74 Å, demonstrating that, aside from a few differences, the bulk of the apo NosL structure superposes well with that of apo MerB despite the lack of sequence similarity between the two [213]. MerB contains three separate beta sheets and six α-helices. Three of the helices combine with two short beta strands to form a winged helix αααββ fold that is proposed to function in protein-protein interactions. The structure of MerB adopts a unique fold, and regions of similarity between MerB and other proteins in the Protein Data Bank encompass only the winged helix region, which does not align with apo NosL. Instead, the apo NosL structure superposes with the two C-terminal beta sheets of MerB and with helix (α5), while the lengths of the loops on the sheets varies between the two. In the place of the disordered C-terminus of NosL, MerB contains a long α-helix (α6) that serves to cover one of the faces of sheet 2. Helix α1 of NosL is completely absent in MerB, although a coil runs through the helical axis.

MerB is encoded on a plasmid-carrying operon (*MerR,A,P,T,B*) responsible for mercury resistance in bacteria. MerB is an organomercurial lyase capable of binding a variety of organometallic compounds, cleaving the carbon-mercury bond and releasing Hg(II) for further detoxification by MerA [214]. The two mercury binding residues are

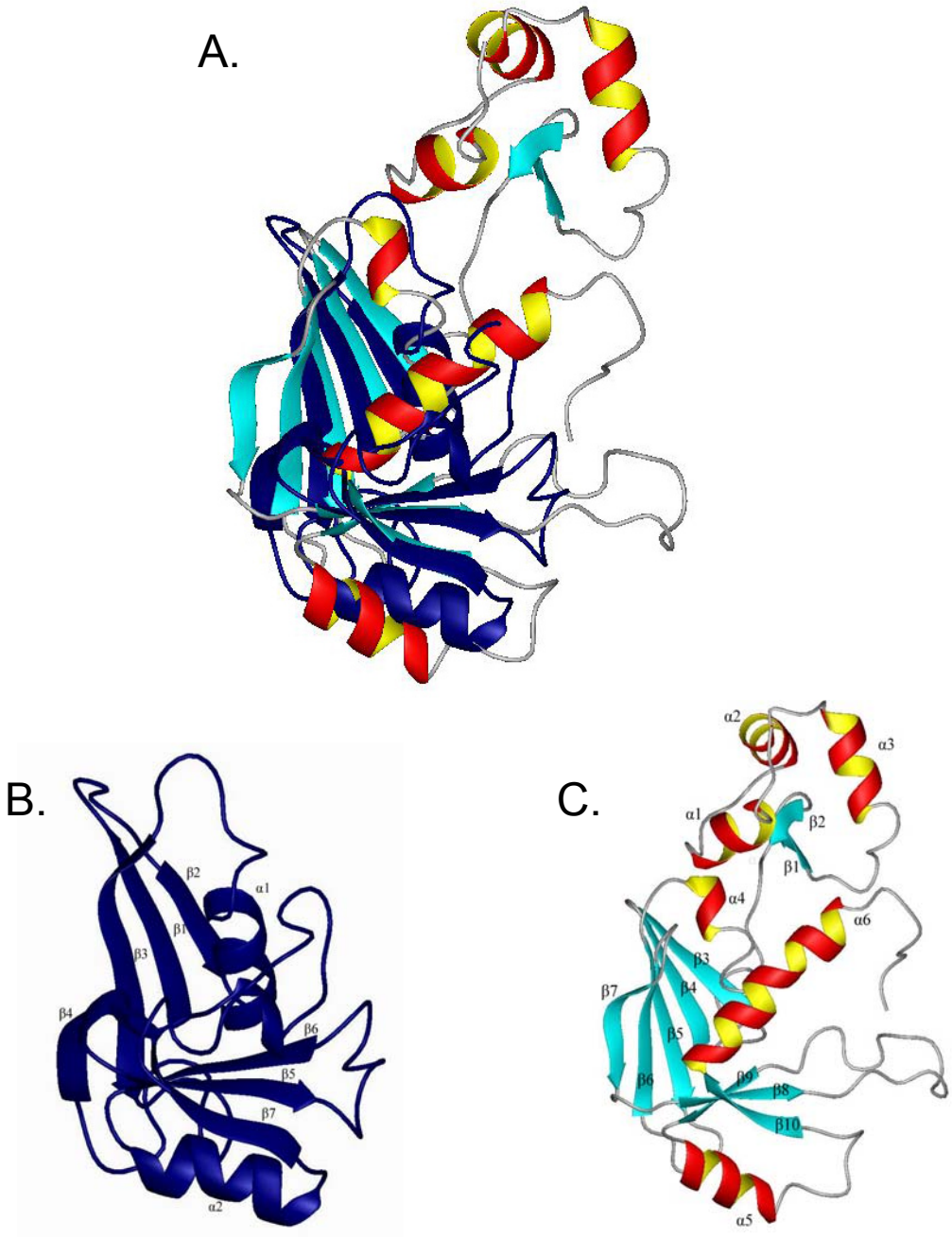


Figure 38 A, B, and C. Superposition of MerB with NosL. MerB is shown is cyan for the beta strands, red and yellow for the helices, while NosL is shown in dark blue. Sheets 1 and 2 and helix α2 align closely with MerB, while helix α1 is overlaid with a coil. MerB has a long helix in place of the relatively disordered C-terminus of NosL. B. Ribbon figure of Apo NosL with the strands and helices labeled in same view as A. C. Ribbon figure of MerB with the strands and helices labeled in the same view as A.

Cys96, originating from the second β sheet, and Cys159, residing on a solvent-exposed loop extending from the strands β8 and β9. The ligands are in relatively poorly-defined regions of the protein that become stabilized upon mercury binding, suggesting that flexibility of the mercury binding site in the apo protein may be important to the activity of MerB [213]. MerB is believed to bind the metal of a thiol-organomercury substrate via Cys159, releasing the thiol. Cys96 is then proposed to form a bond to the mercury via its sulfur, giving up its proton and protonolysizing the organic moiety. Finally, a thiol from the buffer, either glutathione or cysteine, then displaces the two Hg(II) ligands, releasing the detoxified metal. Based on chemical shift mapping on the mercury-bound protein in complex with the inhibitor DTT, the large hydrophobic binding pocket binds one molecule of DTT but is larger than the ligand requires. The functional significance of the structural resemblance between this protein and NosL will be discussed in the next chapter.

Comparison of Apo and Holo NosL ^{1}H-^{15}N-HSQC Spectra

The 2D ^{1}H-^{15}N-HSQC spectra of apo and holo NosL were compared in order to gain preliminary information on the feasibility of copper(I)-loaded NosL structure determination and to identify residues influenced by the incorporation of copper. Figure 38 is an overlay of the two spectra on identical scales, demonstrating that while most of the amide proton resonances overlap, a few cross-peaks exhibit dramatic chemical shift differences between the two. In addition, there are at least four backbone amide protons

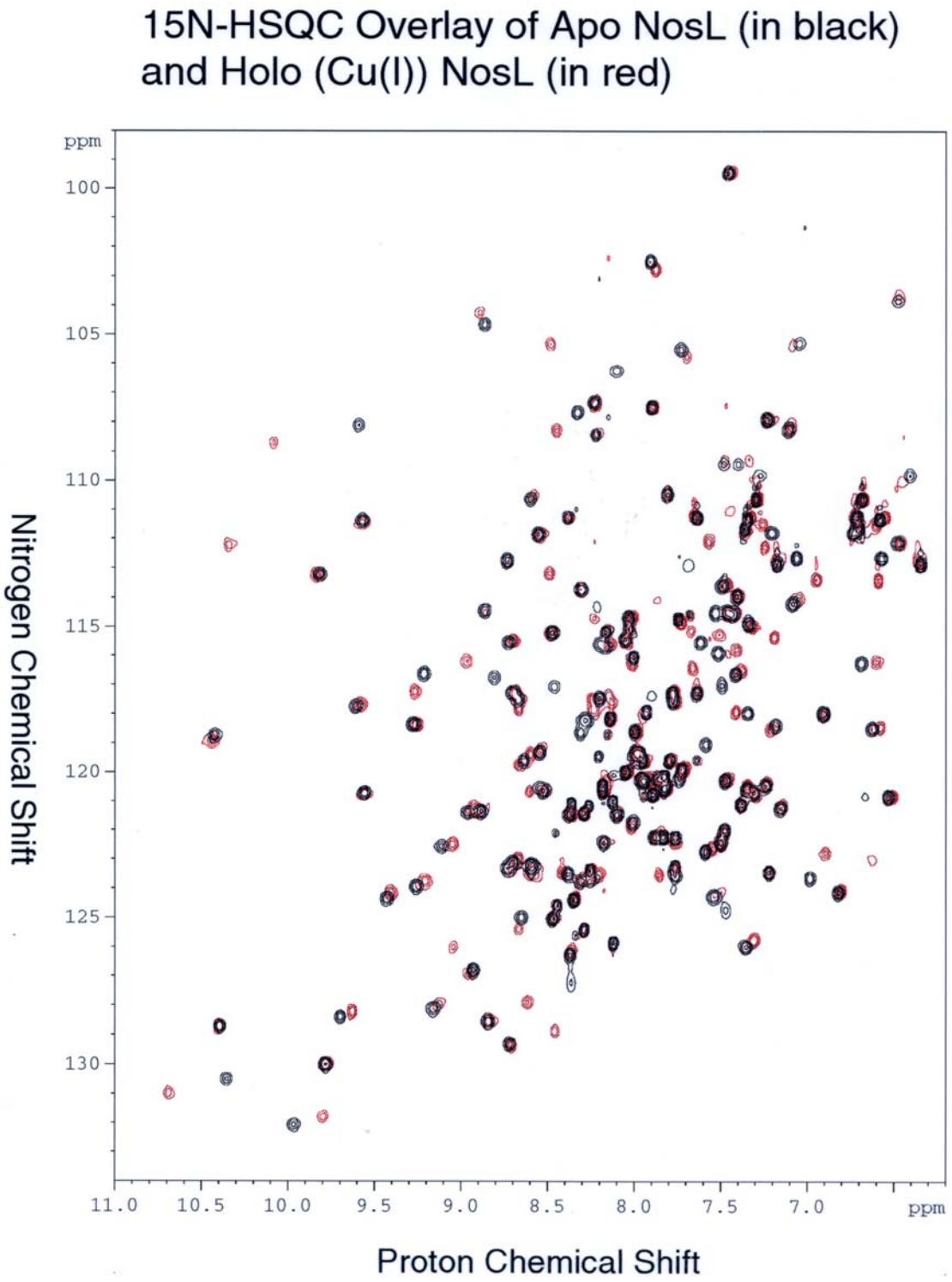


Figure 39. Overlay of the ^{1}H-^{15}N-HSQC spectra of apo (black) and holo (red) NosL.

present in the holo spectrum that are not observed in that of the apo, suggesting that some of the chemical exchange broadening leading to the inability to assign amide protons to residues Thr16 and Tyr23-Leu28 occurs to a lesser degree in the holo form. While it is tempting to speculate that copper binding induces more ordering of this region of the protein by stabilizing a specific conformation and thereby reducing exchange broadening, other factors such as changes in solvent accessibility (and therefore amide exchange rates) and altered relaxation behavior can also account for these observations.

The chemical shift differences between assigned residues in the two proteins as a function of residue number are depicted in Figures 40 and 41, showing the variation in ^{15}N and amide ^{1}H chemical shifts, respectively, upon copper binding. Many of the chemical shift differences induced by the presence of copper are quite large, consistent with the presence of a positively charged ion which would be expected to exert a large effect on the magnetic shielding environment experienced by nearby nuclei.

Residues exhibiting significant chemical shift changes (greater than two standard deviations from the mean, +/- .14 ppm for proton and +/- 0.40 ppm for nitrogen) upon copper binding are shown in Table 4. The truncation of the N-terminus for calculation purposes prevents visualization of the shifted residues Thr19, Leu20 and Gly21--however, given the inability to assign the intervening residues and the disorder in the N-terminus, this region would most likely remain ill-defined and yield little interpretive value.

The residues that can be pinpointed on the structure are distributed throughout the secondary structure elements of the first 114 residues in the primary sequence and form

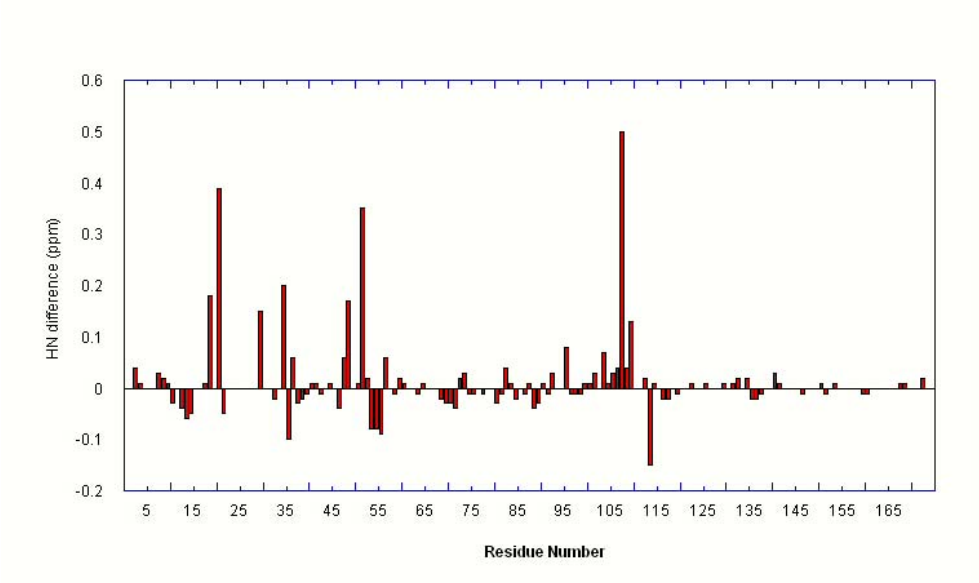


Figure 40. Amide ^{1}H chemical shift differences between apo and holo NosL. Chemical shift differences of the amide proton are plotted on the vertical axis in ppm, the horizontal axis is the position of the residue in the primary sequence.

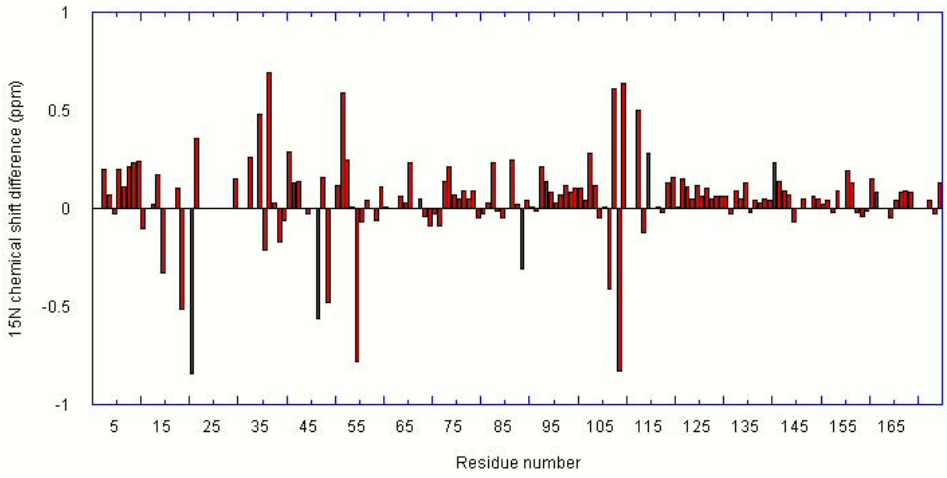


Figure 41. ^{15}N chemical shift differences between apo and holo NosL. Chemical shift differences of the amide nitrogen are plotted on the vertical axis in ppm, the horizontal axis is the position of the residue in the primary sequence.

Residue	Amide proton Chemical Shift difference (ppm)	Nitrogen Chemical Shift difference (ppm)
Thr19	0.18	-0.51
Gly21	0.39	-0.84
Glu30	0.15	0.15
Lys35	0.20	0.48
Gln37	0.06	0.69
Phe49	0.17	-0.48
Val52	0.35	0.59
Ala55	-0.08	-0.78
Gly107	0.04	-0.41
Gly108	0.50	0.61
Met109	0.04	-0.83
Gly110	0.13	0.64
Glu113	0.02	0.50
Ala114	-0.15	-0.12

Table 4. Residues exhibiting chemical shift differences between apo and holo NosL greater than two standard deviations from the mean (+/- .14 ppm for ^{1}H and +/- 0.40 ppm for ^{15}N).

three spatially distinct, but contiguous, regions. Residues Lys35, Gln37, Phe49, Val52 and Ala55 are clustered about one end of the sheet 1-α1 interface. Residues Gly107, Gly108, Met109 and Gly110 form the locus of another distinct region of the protein that is perturbed by the introduction of copper. Between these two regions are residues Glu113 and Ala114, forming a continuous zone of highly shifted amide protons

extending from the loop between strands β5 and β6 on sheet 2 all the way to strands β1 and β2 (see Figure 42).

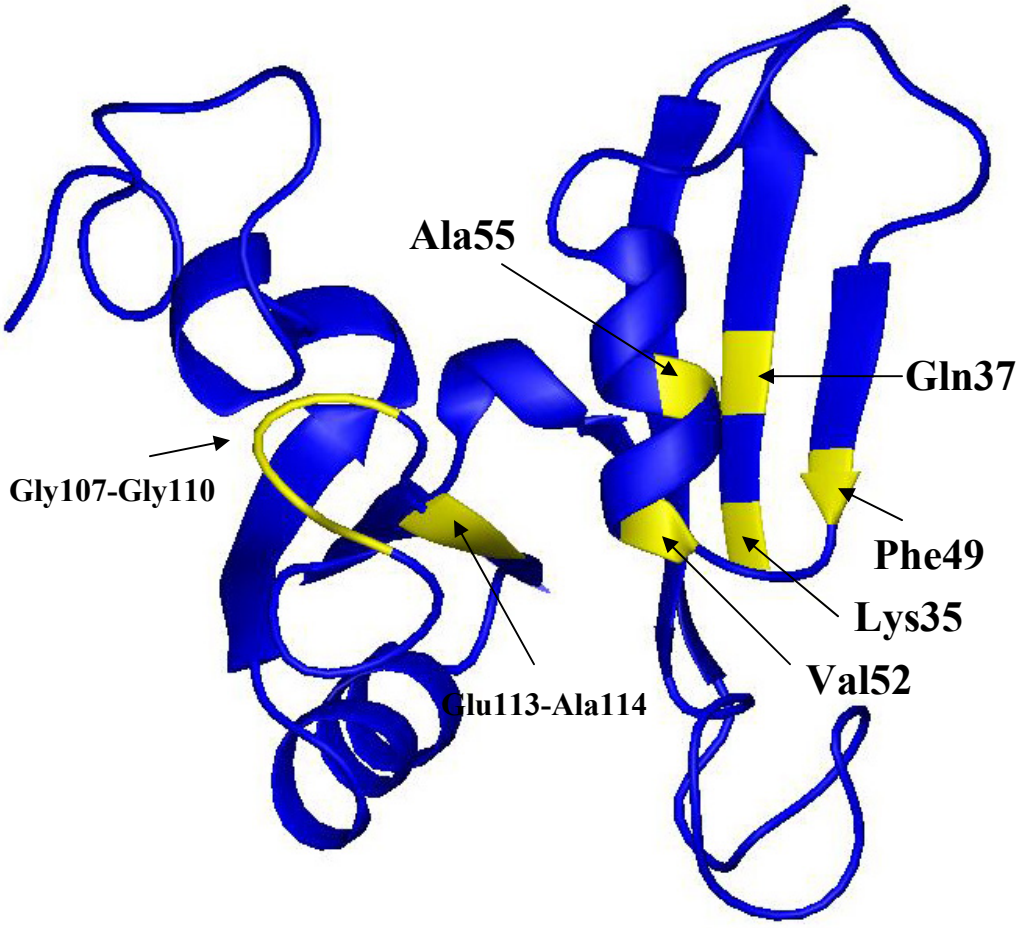


Figure 42. Ribbon diagram of NosL with amide protons perturbed by the presence of copper highlighted. Amide protons shifted by two standard deviations are shown in yellow while the rest of the structure is shown in dark blue.

Because chemical shift effects can be both local as well as a reflection of long-range conformational changes, interpretation of the data is not straightforward without more structural details on the holo protein. Although this is only preliminary data, the importance of the residues exhibiting large chemical shift perturbations is underscored by

the observation that Lys35, Gln37, Phe49, Gly107, Gly108, Met109, and Gly110 are all highly conserved residues.

Comparison of apo NosL with its structural homolog MerB reveals some similarities in the location of mercury ligands on MerB with the location of residues perturbed by the presence of copper in NosL (see Figure 43). Cys159 is located on the β8/β9 loop (residues 146-161) of MerB, corresponding to the β5/β6 loop of NosL (residues Ser103-Pro112) containing the strictly conserved G/A-G/A-MG motif. The residues within this motif show demonstrable chemical shift differences upon copper incorporation, and, coupled to the EXAFS data identifying a sulfur ligand to the copper, strongly suggest that Met109 is one of the metal ligands in NosL. Cys96, the second mercury ligand in MerB, resides on the loop between strands β3 and β4 of the MerB structure, which is spatially homologous to the α1 helix of NosL. While this region of NosL is perturbed by the presence of copper, particularly residues Val52 and Ala55, and to a lesser extent Arg53 and Asp54, none of these residues are completely conserved. Regions of NosL proximal to this part of α1 include the lower half of sheet 1, in which residues Gln37, Phe49, and Lys35 are strongly affected by the presence of copper. Of these residues, only Lys35 is a potential although unlikely copper ligand and is strictly conserved.

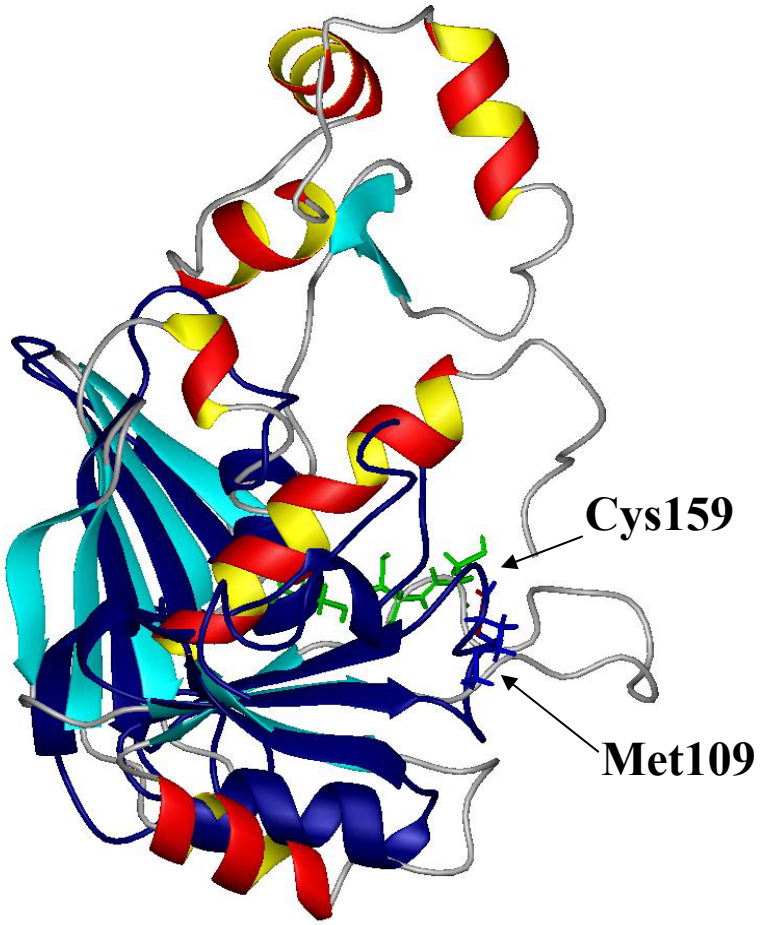


Figure 43. Superposition of NosL with MerB with metal-binding ligands shown. MerB is shown in cyan, yellow and red while NosL is shown in dark blue. One of the proposed copper ligands for NosL, Met109, is shown in dark blue, while the mercury binding ligands for MerB, Cys159 and Cys96 are shown in green.

DISCUSSION ON NOS L

The ^{1}H-^{15}N-HSQC data collected on apo NosL immediately suggested a high degree of β-sheet content by the high degree of spectral dispersion, which has been born out by the structure. The protein is comprised of two domains, each beta sheet flanked by an α-helix, arranged in a novel twisted butterfly-like fold in which the core is made up of hydrophobic residues residing on α1, β1, β3 of sheet 1 and β6 of sheet 2. The orientation of these structural elements results in the formation of an intriguing large cleft, surrounded by basic residues Arg53, Arg60, and Arg106 amidst a predominantly acidic solvent-accessible surface.

The spectra collected on the protein also exhibited several sharp resonances (indicating long relaxation rates) within the random coil chemical shift range, indicating the presence of a number of disordered residues, many of which were overlapped preventing reliable NOE assignment. Indeed, residues within the N-terminus (1-34) and the C-terminus (145-175) appeared to be disordered as determined by resonance linewidths, by the lack of NOE's and hydrogen bonding, and by the random coil chemical shift values. Consistent with these results, previous N-terminal sequencing combined with mass spectrometry performed on NosL demonstrated that the last ten residues on the C-terminus were proteolyzed during isolation [161]. For these reasons, the structure was truncated to include only residues Lys35 to Glu160 in the calculations. The use of this reduced sequence aided the automated assignment process and increased the efficiency of convergence of structures generated by ARIA. Residues Val145-Glu160, although

included in the calculation, are also disordered and form a patch of highly negative electrostatic potential on the surface of the protein.

Residues Thr16 and Tyr23-Leu28 within the N-terminus were exchange-broadened on the NMR timescale, indicating that these residues sample at least two distinct conformations within roughly a millisecond time period. Many proteins, including the apo forms of the AtxI-like copper chaperones, exhibit exchange broadening of residues about the metal binding sites and imply that the flexibility of these regions in the apo form is biologically relevant. NMR Spectra revealed additional conformational exchange processes exhibited by apo NosL, specifically by residues Asn74 and Ala55, both of which reside within well-ordered structural elements.

Other residues within the N-terminus, specifically Ile10, exhibited non-random coil chemical shift values as well as some undefined long-range NOE interactions. The amide proton of residue Leu69 and the γ proton of Val70 were close in space to the δ protons of Ile10 as evidenced by the NOE data, suggesting that there may be some stable interaction of the N-terminus with β3. Unfortunately, because the chemical shifts of intervening residues (Thr16-His31) could not be ascertained, inclusion of this region would have complicated ARIA automated assignment, in which greater than 90% chemical shift completeness is desirable [192]. Spectra acquired on the holo protein exhibited several additional amide resonances compared to the apo form, suggesting that at least part of the protein becomes more ordered upon metal binding.

Perhaps because the sequencing of denitrifying organisms has been hindered by the difficulty in culturing these organisms, there were no significant sequence homologs

to NosL other than closely related NosL sequences. The NosL sequences are highly divergent with few completely conserved residues and appear to form their own family of proteins (P. Babbitt, personal communication). There are two discernable motifs, PGPK-G/A and G/A-G/A-MG, conserved throughout this family but not identified with any known functions.

Despite the few number of conserved residues, it was impossible to definitively identify the copper binding ligands of NosL based on sequence analysis alone. EXAFS data collected on holo NosL showed that the ligands to copper are two sulfur donors and one nitrogen or oxygen ligand. Typical copper ligands include Asp, Glu, His, Met and Cys, while unusual but possible ligands include mainchain oxygen and nitrogen and the amino acids Tyr, Ser, Thr, Gln, Asn, Lys and Arg [215]. NosL has two conserved sulfur donors, Met26 and Met109, and several conserved potential oxygen/nitrogen donors including Lys35, Glu/Gln37, Glu/Asp63, and Tyr99, not to mention the mainchain atoms (although these are typically poor copper ligands).

Lacking sequential homologs, structural homologs to NosL were sought in an effort to identify functionally related proteins. MerB, an alkylmercury lyase, was determined by DALI to be the only significant structural homolog to NosL. The structure of NosL almost completely aligns with the structure of MerB, differing most noticeably in the length and orientation of loops and in the presence of the long helix (α6) in MerB in place of the loosely defined C-terminus of NosL. The large cleft observed in apo NosL is also found in MerB and was found to be a binding site for the inhibitor DTT by

chemical shift mapping performed on the complex. It is presumed that this cleft binds the physiologically relevant thiol donor as well.

Chemical shift perturbation mapping was performed by comparing the ^{1}H-^{15}N-HSQC spectra of apo NosL with that of the holo protein. Chemical shift comparison is a simple but very informative technique for identifying short-range and more global conformational changes upon ligand binding. Chemical shifts are rich in information content and are exquisitely sensitive to changes in the electronic environment that reflect the sum of electrostatic contributions, ligand interactions, changes in packing, long-range conformational effects, and ring-current effects. As a result, direct interpretation of chemical shift mapping is fraught with difficulty and requires independent verification. Despite these difficulties, key information on the copper-loaded form of NosL was obtained by mapping the shifted residues onto the structure of apo NosL and by comparison with the location of the mercury-binding ligands of MerB.

Chemical shift mapping demonstrated that a relatively large region that extends to both of the domains of NosL is affected by the presence of copper, extending from the β5/β6 loop, along the edge of sheet 2, to the intervening helix α1 and to residues residing on the N-terminal end of sheet 1. These long-range effects are distinct from those observed upon similar chemical shift perturbation experiments performed on apo vs. holo forms of the AtxI-like metallochaperones, in which only residues within the copper-binding loop are affected by copper incorporation. Interestingly, amide protons shifted upon metal binding include those that are not solvent accessible in the apo structure,

including Val52, Ala55, and Ala114, suggesting that some conformational exchange occurs upon copper binding.

Of the residues that are shifted by more than 2 standard deviations in ppm by the presence of copper, only Lys35, Gly107, Met109, and Gly110 are strictly conserved. Of these, only Lys35 and Met109 are possible copper ligands, although the deprotonated state of the lysine εN is relatively inaccessible under physiologic conditions and therefore seems an unlikely ligand. If we expand the search for copper-binding residues to those shifted by one standard deviation in the amide or nitrogen chemical shift, other potential ligands can be identified and include Asp54, Gln37, Ser103, Thr141, Lys96, Glu41, Glu113, and Arg53. Of these, there are no strictly conserved residues, although, intriguingly, Asp/Glu54 is present in every NosL sequence except for one, which contains a nearby Asp (Asp51 in *Burkholderia pseudomallei*).

Strikingly, the spatial location of the potential copper ligand Met109 is highly similar to that of the mercury-binding ligand Cys159 in the MerB structure. Cys159 resides on the loop that corresponds to Loop β5/β6 (Ser103-Pro112) in NosL. The two loops are of different length and conformation, but the identification of Met109 within a conserved motif and along with observed chemical shift changes upon metal incorporation are consistent with its role as a copper ligand. Met26 is the only other completely conserved sulfur-containing residue, and, although it has no analogue in the MerB structure, is likely to be the third copper ligand. Interestingly, this residue exhibited the severe exchange-broadening that characterized the mercury ligands of MerB, preventing assignment of this residue.

The other mercury ligand of MerB, Cys96, resides in a coil region that superposes with the axis of α1 in the NosL structure that intervenes between the two domains. While there are no conserved copper-binding residues in this region of NosL, Asp/Glu54 is located on this helix and, as mentioned above, is conserved in all NosL sequences except for one, suggesting that this residue may be the second copper ligand. This region of MerB is also spatially proximal to the lower half of sheet 1 and to the conserved PGPK-G/A motif. Lys35 is 100% conserved and is shifted by the presence of copper, although it is an unlikely ligand and may serve as a positively charged analog for copper, stabilizing the apo state of NosL analogous to the conserved lysine of the AtxI-like metallochaperones. Interestingly, this residue exhibits some exchange broadening effects parallel to that observed for Cys96.

Inspection of the structure of apo NosL reveals that the potential copper-binding residues Met109 and Lys35/Asp54 are located on different subdomains. While these residues are likely to be too far apart in the apo structure to accommodate copper binding without change in conformation, the chemical shift perturbation data showed that extensive reorganization of the interceding residues is highly likely upon copper incorporation and could conceivably serve to bring the two sidechains close to each other. Specifically, residues Glu113, Ala114, Val52, Ala55, and to a lesser extent Ala93, Arg53, and Val115, and Lys96 exhibit demonstrable chemical shift changes upon metal binding, suggestive of a large conformation change that alters the interactions of α1 with sheet 1 and sheet 2 as well regions about the hinge connecting the two domains. Intriguingly, this reorganization would most likely result in changes to the deep cleft of

NosL that correlates with the DTT-binding pocket in the MerB structure. Because Met26 was not modeled in these structures, the spatial proximity of this potential copper ligand to the others cannot be ascertained.

The discovery of a relatively close structural homolog, an alkylmercury lyase, opens up unanticipated possibilities for the function of NosL. While the similarity between the metal binding ligand Met 109 of NosL to Cys159 of MerB is remarkable, many key differences between the two proteins does not allow for the direct identification of the function of NosL. MerB binds a sulfur-organomercury complex via two mercury-binding cysteine residues, that along with a thiol provided by the buffer, appear to be responsible for the catalytic cleavage of the alkylmercury bond in order to detoxify the metal compound and deliver it to MerA [213, 214]. In contrast, only the incorporation of copper(I) and other metals into the apo protein has been studied *in vitro* and no catalytic ability has been investigated. Furthermore, NosL binds copper via two sulfur and one nitrogen or oxygen donor, such that the coordination of this metal differs from that of mercury and is likely to be specific for the role NosL plays in N_2OR assembly. Finally, MerB functions in the cytoplasm while NosL resides in the outer membrane of the periplasm, which has a distinct redox environment incompatible with free thiol, perhaps explaining the different sulfur ligands used by the two proteins.

While the differences between the two proteins are considerable, some of the parallels between NosL and MerB hint at potential roles for NosL in the N_2OR assembly machinery. MerB harbors a large binding pocket, most likely suited for binding the physiologic thiol donors cysteine and/or glutathione found in the cytoplasm. NosL

displays a very similar pocket created by the specific orientations adopted by the beta pleated sheets and the α1 helix, suggesting that NosL may also be capable of binding another ligand in complex with copper. Like copper, free sulfide is highly toxic to cells and must also be tightly regulated, requiring a sulfur donor to generate the unique Cu_Z cluster of N_2OR. Perhaps NosL, like MerB, interacts with a copper-sulfur complex, generating a Cu-S cluster precursor for ultimate incorporation into the Cu_Z metallocenter. In either case, the potential of NosL to bind a biologically available copper complex and an additional ligand may explain the difficulties in producing subsequent samples of holo NosL by simple dialysis against a non-native copper compound.

The relationship between a component of the nitrous oxide reductase assembly system and one from the mercury detoxification machinery underscores a recurring theme reinforced by our recent understanding of these systems. Metal detoxification requires extensive substrate channeling—the proteins involved must specifically bind and sequester the toxic compounds and deliver them to their target without release, but must also be able to deliver the metal to the appropriate targets under specific conditions. It is believed that these systems developed in response to the extreme environment that organisms first evolved in, and most likely predate the copper clusters used for metabolism in an oxidizing environment [216]. It seems likely, therefore, that the pre-existing mercury detoxification proteins have been adapted by the N_2OR system for copper transport, modification, and finally insertion.

Although NosL does not appear to be a simple copper transfer/delivery protein like the AtxI-like matellochaperones, there are some analogies that can be drawn with

this family of proteins. NosL appears to bind copper in a relatively solvent-exposed location and somewhat incomplete coordination sphere, similar to chaperones, although studies on the holo protein are still needed to verify this conclusion. This suggests that the function of NosL is in part to deliver the copper or a copper complex to a target enzyme by maintaining the accessibility of the site. Unlike the metallochaperones, however, extensive reorganization of NosL appears to occur upon copper binding, inconsistent with the facile uptake and delivery of a metal. The potential processing of copper and/or sulfur by NosL based on the identification of the structural homolog MerB may occur with the aid of an additional, unidentified substrate that binds within the NosL pocket, distinguishing this protein from the more simple chaperones and opening up new lines of investigation for the role that this protein plays in nitrous oxide reductase assembly.

CONCLUSIONS

The two projects described in this thesis, one of which was amenable to NMR techniques and the other not, demonstrated the utility of this spectroscopic method as well as its limitations. For the first project, regions of LBP responsible for the aggregation that has hampered structure determination by both NMR and crystallography techniques have been identified. Bacterially-expressed recombinant mammalian LBP was isolated in its soluble form and separated as a monomeric species, the first step towards developing an expression system suitable for ^{15}N and ^{13}C labeling required for NMR structural work. Structural domains roughly located between residues 1 to 136 and 137-220 were identified, and ELISA data results show that recombinant LBP binding to laminin-1 is mediated by the 137-230 domain. A disordered region between residues 200 and 295 was established, supported by limited proteolysis, CD and NMR data, arguing against the model proposed by Kazmin et al [99]. This construct lacked biological activity under these conditions in both ELISA and NMR experiments, contrasting with the experiments performed on synthetic peptide 205-229, suggesting that the behavior of this disordered domain and that of the peptide are strikingly different. Finally, an alternative function for the C-terminus of LBP has been proposed in which the disorder of this region provides a mechanism for displaying cleavage sites that release peptides involved in modulating platelet activation.

Unlike the laminin binding protein, apo NosL was highly suited for NMR structural investigations despite the low yields obtained from recombinant expression. The structure of NosL, calculated by the automated assignment program ARIA in

combination with CNS, exhibits two domains, each consisting of a beta sheet and a flanking α-helix. The two domains are joined by a short 3_{10} helix and twisted relative to each other such that the two sheets are almost completely perpendicular to one another. The orientation of the two sheets generates a deep pocket between the two domains surrounded by basic residues. The termini of the protein were determined to be disordered, and therefore left out of the calculation. MerB, an alkylmercury lyase involved in mercury detoxification, was identified as the only significant structural homolog to NosL. Superposition of the two proteins demonstrated that the two structures are highly similar except for the presence of an N-terminal domain in MerB not present in NosL and a long C-terminal helix that takes the form of an irregular, relatively disordered region of NosL. The alignment of the two proteins, in concert with chemical shift perturbation experiments performed to identify changes upon copper incorporation, allowed the identification of possible copper ligands for NosL as Met109 and Met26 for the two sulfur ligands, while the third nitrogen/oxygen ligand remains unknown. Although a definitive function for NosL within the N_2OR system cannot be identified, it is clear that NosL is not a simple copper chaperone but may exhibit some catalytic activity as well.

The next logical step in the study of NosL is to determine the specific function of this protein by identification of substrates and conditions under which the possible catalytic activity can be studied. Mutation of the proposed copper ligands, along with the solution structure of the holo protein, should provide details on the mode of copper binding and potential modification of the metal/metal complex by additional factors. The

methods described herein for the introduction of copper via dialysis should allow the preparation of the holo protein for NMR samples, although more scrupulous maintenance of anaerobic conditions and optimization of copper incorporation will need to be performed. Finally, interactions between NosL and other members of the N_2OR system need to be addressed. Overexpression of the members of the Nos cluster and the addition of these components to NMR samples of NosL could allow for the identification of binding partners, even in weak complexes. Although the structure of apo NosL has generated more possibilities than answers, the role that this potential catalyst plays in the creation of the novel Cu_Z cluster of nitrous oxide reductase opens up a new chapter in the assembly of complex metalloclusters.

REFERENCES CITED

1. Bishop, J.M., Weinberg, R.A., ed. *Scientific American Molecular Oncology*. 1996, Scientific American Inc.: New York, New York.

2. Campbell, P.M., Channing, J.D., *Oncogenic Ras and its role in tumor cell invasion and metastasis.* Seminars in Cancer Biology, 2004. **14**: p. 105-114.

3. Hanahan, D., Weinberg, R.A., *The Hallmarks of Cancer.* Cell, 2000. **100**: p. 57-70.

4. Shields, J.M., Pruitt, K., McFall, A., Shaub, A., Der, C.J., *Understanding Ras: 'it ain't over 'til it's over'.* Trends in Cell Biology, 2000. **10**: p. 147-154.

5. Vousden, K.H., *Activation of the p53 tumor suppressor protein.* Biochimica et Biophysica Acta, 2002. **1602**: p. 47-59.

6. Collot-Teixeira, S., Bass, J., Denis, F., Ranger-Rogez, S., *Human tumor suppressor p53 and DNA viruses.* Reviews in Medical Virology, 2004. **14**: p. 301-319.

7. Masutomi, K., Hahn, W.C., *Telomerase and tumorgenesis.* Cancer Letters, 2003. **194**: p. 163-172.

8. Brunori, M., Luciano, P., Gilson, E., Geli, V., *The telomerase cycle: normal and pathological aspects.* Journal of Molecular Medicine, 2005.

9. Liotta, L., Stetler-Stevenson, W.G., *Cancer: Principles and Practices of Oncology*, ed. H.S. DeVita VT, Rosenberg SA. 1993, Philadelphia: J B Lippincott Co.

10. Fox, S.B., Gasparini, G., Harris, A.L., *Angiogenesis: pathological, prognostic, and growth-factor pathways and their link to trial design and anticancer drugs.* The Lancet, Oncology, 2001. **2**: p. 278-289.

11. Gerwins P, S.E., Claesson-Welsh L, *Function of fibroblast growth factors and vascular endothelial growth factors and their receptors in angiogenesis.* Critical Reviews in Oncology/Hematology, 2000. **34**: p. 185-194.

12. Liotta, L.A., Stetler-Stevenson, W.G., *Tumor Invasion and Metastasis: An Imbalance of Positive and Negative Regulation.* Cancer Research (Suppl.), 1991. **51**: p. 5054s-5059s.

13. Weiss, L., *Metastatic Inefficiency.* Advances in Cancer Research, 1990. **54**: p. 159-211.

14. Woodhouse, E.C., Chuaqi, R.F., Liotta, L.A., *General Mechanisms of Metastasis.* Cancer, 1997. **80**: p. 1529-37.

15. Yurchenco, P.D., Schittny, J.C., *Molecular architecture of basement membranes.* FASEB J, 1990. **4**: p. 1577-1590.

16. Sotille, J., *Regulation of angiogenesis by extracellular matrix.* Biochimica et Biophysica Acta, 2004. **1654**: p. 13-22.

17. Streuli., C., *Extracellular matrix remodelling and cellular differentiation.* Current Opinion in Cell Biology, 1999. **11**: p. 634-640.

18. Lukashev, M., Werb, Z., *ECM signalling: orchestrating cell behavior and misbehavior.* Cell Biology, 1998. **8**: p. 437-441.

19. Sasisekharan, R., Venkatarman, G., *Heparin and heparan sulfate: biosynthesis, structure, and function.* Current Opinion in Cell Biology, 2000. **4**: p. 626-631.

20. Turnbull, J., Powell, A., Guimond, S., *Heparan sulfate: decoding a dynamic multifunctional cell regulator.* Trends in Cell Biology, 2001. **11**(2): p. 75-82.

21. Mercurio, A.M., *Laminin receptors: achieving specificity through cooperation.* Trends in Cell Biology, 1995. **5**: p. 419-423.

22. Ott, U., Odermatt, E., Engel, J., Furthmayr, H., Timpl, R., *Protease Resistance and Conformation of Laminin.* Eur J Biochem, 1982. **123**: p. 63-72.

23. Fujiwara, S., Shinkai, H., Deutzmann, R., Paulsson, M., Timpl, R., *Structure and distribution of N-linked oligosaccharide chains on various domains of mouse tumour laminin.* Biochem J, 1988. **252**: p. 453-461.

24. Charonis, A.S., Skubitz, A.P.N., Koliakos, G.G., Reger, L.A., Dege, J., Vogel, A.M., *A Novel Synthetic Pepitde from the B1 Chain of Laminin with Heparin-binding and Cell Adhesion-promoting Activities.* The Journal of Cell Biology, 1988. **107**: p. 1253-1260.

25. Beck, K., Hunter, I., Engel, J., *Structure and Function of laminin: anatomy of multidomain glycoprotein.* FASEB J, 1990. **4**: p. 148-160.

26. Ruoslahti, E., Pierschbacher, M.D., *New Perspectives in Cell Adhesion RGD and Integrins.* 1987.

27. Dogic D, R.P., Aumailley M, *Cell adhesion to laminin 1 or 5 induces isoform-specific clustering of integrins and other focal adhesion components.* Journal of Cell Science, 1998. **111**: p. 793-802.

28. Graf, J., Iwamoto, Y., Sasaki, M., Martin, G.R., Kleinman, H.K., Robey, F.A., Yamada, Y., *Identification of an amino acid sequence in laminin mediating cell attachment, chemotaxis, and receptor binding.* Cell, 1987. **48**(6): p. 989-996.

29. Graf, J., Ogle, R.C., Robey, F.A., Sasaki, M., Martin, G.R., Yamada, Y., Kleinman, H.K., *A pentapeptide from the laminin B1 chain mediates cell adhesion and binds the 67,000 laminin receptor.* Biochemistry, 1987. **26**(22): p. 6896-6900.

30. Iwamoto, Y., Robey, F.A., Graf, J., Sasaki, M., Kleinman, H.K., Yamada, Y., Martin, G.R., *YIGSR, a Synthetic Laminin Pentapeptide, inhibits experimental metastasis formation.* Science, 1987. **238**: p. 1132-1138.

31. Malinoff, H., Wicha, M.S., *Isolation of a Cell Surface Receptor Protein for Laminin from Murine Fibrosarcoma Cells.* The Journal of Cell Biology, 1983. **96**: p. 1475-1479.

32. Rao, C.N., Barsky, S.H., Terranova, V.P., Liotta, L.A., *Isolation of a tumor cell laminin receptor.* Biochemical and Biophysical Research Communications, 1983. **111**(3): p. 804-808.

33. Gebarowska, D., Stitt, A.W., Gardiner, T.A., Harriott, P., Greer, B., Nelson, J., *Synthetic peptides interacting with the 67-kD laminin receptor can reduce retinal ischemia and inhibit hypoxia-induced retinal neovascularization.* American Journal of Pathology, 2002. **160**(1): p. 307-313.

34. Nelson, J., Scott, W.N., Allen, W.E., Wilson, D.J., Harriott, P., McFerran, N.V., Walker, B., *Murine epidermal growth factor peptide (33-42) binds to a YIGSR-specific laminin receptor on both tumor and endothelial cells.* The Journal of Biochemistry, 1996. **271**(42): p. 26179-26186.

35. Wewer, U., Taraboletti, G., Sobel, ME., Albrechtsen, R., Liotta, LA., *Role of the laminin receptor in tumor cell migration.* Cancer Research, 1987. **47**(21): p. 5691-5698.

36. Hand, P.H., Schlom, T.J., Rao, C.N., Liotta, L., *Expression of Laminin Receptor in Normal and Carcinomatous Human Tissues as Defined by a Monoclonal Antibody.* Cancer Research, 1985. **45**(June): p. 2713-2719.

37. Martignone, S., Pellegrini, R., Villa, E., Tandon, N.N., Mastroianni, A., Tagliabue, E., Menard, S., Colnaghi, M.I., *Characterization of two monoclonal antibodies directed against the 67 kDa high affinity laminin receptor and application for the study of breast carcinoma progression.* Clinical &Experimental Metastasis, 1992. **10**: p. 379-386.

38. Waltregny, D., de Leval, L., Menard, S., Castronovo, V., *Independent prognostic value of the 67-kd laminin receptor in human prostate cancer.* Journal of the National Cancer Institute, 1997. **89**(16): p. 1224-1227.

39. van den Brule, F.A., Berchuck, A., Bast, R.C., Liu, F., Gillet, C., Sobel, M.E., Castronovo, V., *Differential expression of the 67-kD laminin receptor and 31-kD*

human laminin-binding protein in human ovarian carcinomas. European Journal of Cancer, 1994. **30A**(8): p. 1096-1099.

40. Menard, S., Tagliabue, E., Colnaghi, M.I., *The 67 kDa laminin receptor as a prognostic factor in human cancer.* Breast Cancer Research and Treatment, 1998. **52**: p. 137-145.

41. Yow, H., Wong, J.M., Chen, H.S., Lee, C., Steele, G.D., Chen, L.B., *Increased mRNA expression of a laminin-binding protein in human colon carcinoma: Complete sequence of a full-length cDNA encoding the protein.* Proceedings of the National Academy of Science, 1988. **85**: p. 6394-6398.

42. Mafune, K., Ravikumar, T.S., Wong, J.M., Yow, H., Chen, L.B., Steele, G.D., *Expression of a Mr 32,000 laminin-binding protein messenger RNA in human colon carcinoma correlates with disease progression.* Cancer Research, 1990. **50**(13): p. 3888-3891.

43. Pelosi, G., Pasini, F., Bresaola, E., Bogina, G., Pederzoli, P., Biolo, S., Menard, S., Zamboni, G., *High-affinity monomeric 67-kd laminin receptors and prognosis in pancreatic endocrine tumours.* Journal of Pathology, 1997. **183**: p. 62-69.

44. Zuidervaart, W., van der Velden, P.A., Hurks, M.H., van Nieuwpoort, F.A., Out-Luiting, C.J.J., Singh, A.D., Frants, R.R., Jager, M.J., Gruis, N.A., *Gene expression profiling identifies tumour markers potentially playing a role in uveal melanoma development.* British Journal of Cancer, 2003. **89**: p. 1914-1919.

45. Boukerche, H., Su, Z., Kang, D., Fisher, P.B., *Identification and cloning of genes displaying elevated expression as a consequence of metastatic progression in human melanoma cells by rapid subtraction hybridization.* Gene, 2004. **343**: p. 191-201.

46. Gasparini, G., Barbareschi, M., Boracchi, P., Bevilacqua, P., Verderio, P., Dalla Palma, P., Menard, S., *67-kDa laminin-receptor expression adds prognostic information to intra-tumoral microvessel density in node-negative breast cancer.* International Journal of Cancer, 1995. **60**: p. 604-610.

47. Tanaka, M., Narumi, K., Isemura, M., Abe, M., Sato, Y., Abe, T., Saijo, Y., Nukiwa, T., Satoh, K., *Expression of the 37-kDa laminin binding protein in*

murine lung tumor cell correlates with tumor angiogenesis. Cancer Letters, 2000. **153**: p. 161-168.

48. Singhal, S., Wiewrodt, R., Malden, L.D., Amin, K.M., Matzie, K., Friedberg, J., Kucharczuk, J.C., Litzky, L.A., Johnson, S.W., Kaiser, L.R., Albelda, S.M., *Gene expression profiling of malignant mesothelioma.* Clinical Cancer Research, 2003. **9**: p. 3080-3097.

49. Narumi, K., Inoue, A., Tanaka, M. Isemura, M., Shimo-Oka, T., Abe, T., Nukiwa, T., Satoh, K., *Inhibition of experimental metastasis of human fibrosarcoma cells by anti-recombinant 37 kDa laminin binding protein antibody.* Japan Journal of Cancer Research, 1999. **90**: p. 425-431.

50. Ardini, E., Sporchia, B., Pollegioni, L., Modugno, M., Ghirelli, C., Castiglioni, F., Tagliabue, C., Menard, S., *Identification of a Novel Function for 67 KDa laminin Receptor: Increase in Laminin Degradation Rate and Release Motility Fragments.* Cancer Research, 2002. **62**: p. 1321-1325.

51. Givant-Horwitz, V., Davidson, B., Reich, R., *Laminin-induced signaling in tumor cells: the role of the Mr 67,000 laminin receptor.* Cancer Research, 2004. **64**: p. 3572-3579.

52. Shi, Y., Zhai, H., Wang, X., Wu, H., Ning, X., Han, Y., Zhang, R., Xiao, B., Wu, K., Fan, D., *Multidrug-resistance-associated protein MGr1-Ag is identical to the human 37-kDa laminin receptor precursor.* Cellular and Molecular Life Science, 2002. **59**: p. 1577-1583.

53. Tachibana, H., Koga, K., Fujimara, Y., Yamada, K., *A receptor for green tea polyphenol EGCG.* Nature Structural and Molecular Biology, 2004. **11**(4): p. 380-381.

54. Fulcher, K.D., Welch, J.E., Davis, C.M., O'Brien, D.A., Eddy, E.M., Characterization of Laminin Receptor Messenger Ribonucleic Acid and Protein Expression in Mouse Spermatogenic Cells, 1993. **48**: p. 674-682.

55. Douville, P.J., Harvey, W.J., Carbonetto, S., *Isolation and Partial Characterization of High Affinity Laminin Receptors in Neural Cells.* The Journal of Biological Chemistry, 1988. **263**(29): p. 14964-14969.

56. Hara, K., Satoh, K., Ide, H., *Apical Ectodermal Ridge-Dependant Expression of the Chick 67 kDa Laminin Binding Protein Gene(cLbp) in Developing Limb Bud.* Zoological Science, 1997. **14**: p. 969-978.

57. Gloe, T., Riedmayr, S., Sohn, H., Pohlk, U., *The 67-kDa Laminin-binding Protein Is Involved in Shear Stress-dependent endothelial Nitroc-oxide Synthase Expression.* The Journal of Biological chemistry, 1999. **274**(23): p. 15996-16002.

58. Tandon, N.N., Holland, E.A., Kralisz, U., Kleinman, H.K., Robey, F.A., Jamieson, G.A., *Interaction of human platelets with laminin and identification of a 67 kDa laminin receptor on platelets.* The Journal of Biochemistry, 1991. **274**: p. 535-542.

59. Zhang, C., Duan, E., Cao, Y., Jiang, G., Zeng, G., *Effect of 32/67 kDa laminin binding protein antibody on mouse embryo implantation.* Journal of Reproduction and Fertility, 2000. **119**: p. 137-142.

60. Donaldson, E.A., McKenna, D.J., McMullen, C.B., Scott, W.N., Stitt, A.W., Nelson, J., *The expression of membrane-associated 67-kDa laminin receptor (67LR) is modulated in vitro by cell-contact inhibition.* Molecular Cell BIology Research Communications, 2000. **3**(53-59).

61. Thepparit, C., Smith, D.R., *Serotype-specific entry of dengue virus into liver cells: identifiication of the 37 kilodalton/67 kilodalton high-affinity laminin receptor as a dengue virus serotype 1 receptor.* Journal of Virology, 2004. **78**(22): p. 12647-12656.

62. Wang, K.S., Kuhn, R.J., Strauss, E.G., Ou, S. Strauss, J.H., *High-affinity laminin receptor is a receptor for Sindbis virus in mammalian cells.* Journal of Virology, 1992. **66**(8): p. 4992-5001.

63. Sorokin, A.V., Mikhailov, A.M., Kachko, A.V., Protopopova, E.V., Konovalova, S.N., Andrianova, M.E., Netesov, S.V., Kornev, A.N., Loktev, V.B., *Human recombinant laminin-binding protein: Isolation, purification, and crystallization.* Biochemistry (Moscow), 1999. **65**(5): p. 546-553.

64. Chung, J.W., Hong, S.J., Kim, K.J., Goti, D., Stins, M.F., Shin, S., Dawson, V.L., Dawson, T.M., Kim, K.S., *37 kDa laminin receptor precursor modulates*

cytotoxic necrotizing factor 1-mediated RhoA activation and bacterial uptake. The Journal of Biological Chemistry, 2003. **278**(19): p. 16857-16862.

65. Gauczynski, S., Peyrin, J., Haik, S., Leucht, C., Hundt, C., Rieger, R., Krasemann, S., Deslys, J., Dormont, D., Lasmezas, C.I., Weiss, S., *AThe 37-kDa/67-kDa laminin receptor acts as the cell-surface receptor for the cellular prion protein.* The EMBO Journal, 2001. **20**(21): p. 5863-5875.

66. Rieger, R., Edenhofer, F., Lasmezas, C.I., Weiss, S., *The human 37-kDa laminin receptor precursor interacts with the prion protein in eukaryotic cells.* Nature Medicine, 1997. **3**(12): p. 1383-1388.

67. Bissell, M.J., Nelson, W.J., *Cell-to-cell contact and extracellular matrix Integration of form and function: the central role of adhesion molecules.* Current Opinion in Cell Biology, 1999. **11**: p. 537-539.

68. Dogic, D., Rousselle, P., Aumailley, M., *Cell adhesion to laminin 1 or 5 induces isoform-specific clustering of integrins and other focal adhesion components.* Journal of Cell Science, 1998. **111**: p. 793-802.

69. Adams, J.C., *Cell-matrix contact structures.* Cellular and Molecular Life Science, 2001. **58**: p. 371-392.

70. Starkey, J.R., Uthayakumar, S., Berglund, D.L., *Cell surface and substrate distribution of the 67-kDa laminin-binding protein determined by using a ligand photoaffinity probe.* Cytometry, 1999. **35**: p. 37-47.

71. Massia, S.P., Rao, S.S., Hubbell, J.A., *Covalently immobilized laminin peptide YIGSR supports cell spreading and co-localization of the 67-kilodalton laminin receptor with alph-actinin and vinculin.* The Journal of Biological Chemistry, 1993. **268**(11): p. 8053-8059.

72. Singer, I.I., Scott, S., Kawka, D.W., Kazazis, D.M., *Adhesomes: Specific Granules Containing Receptors for Laminin, C3bi/Fibrinogen, Fibronectin, and Vitronectin in human Polymorphonuclear Leukocyes and Monocytes.* The Journal of Cell Biology, 1989. **109**(6): p. 3169-3182.

73. Karpatova, M., Tagliabue, E., Castronove, V., Magnifico, A., Ardindi, E., Morelli, D., Belotti, D., Calnaghi, M.I., Manard, S., *Shedding of the 67 kD laminin receptor by human cancer cells.* Journal of Cell Biochemistry, 1996. **60**(2): p. 226-234.

74. Halatch, M.E., Hirsch-Ernest, K.I., Kahl, G.F., Weinel, R.J., *Increased expression of alpha 6-integrin receptors and of mRNA encoding the putative 37 kDa laminin receptor precursor in pancreatic carcinoma.* Cancer Letters, 1997. **118**: p. 7-11.

75. van der Flier, A., Sonnenberg, A., *Function and interactions of integrins.* Cell Tissue Research, 2001. **305**: p. 285-298.

76. Romanov, V.I., Wrathall, L.S., Simmons, T.D., Pinto de Silva, P., Sobel, M.E., *Protein Synthesis is required for laminin-induced expression of the 67 kDa laminin receptor and its 37 kDa precursor.* Biochemical and Biophysical Research Communications, 1995. **208**(2): p. 637-643.

77. Ardini, E., Tagliabue, E., Magnifico, A., Buto, S., Castronovo, V., Colnaghi, M.I., Menard, S., *Co-regulation and Physical Association of the 67-kDa Monomeric Laminin Receptor and the alpha6Beta4 Integrin.* The Journal of Biological Chemistry, 1997. **272**(4): p. 2342-2345.

78. Landowski, T.H., Dratz, E.A., Starkey, J.R., *Studies of the Structure of the Matatasis-Associated 67 kDa laminin Binding Protein: Fatty Acid Acylation and Evidence Supporting Dimerization of the 32 KDa Gene Product to form the Mature Form.* Biochemistry, 1995. **34**(35): p. 11276-11286.

79. Canfield, S.M., Khakoo, A.Y., *The Nonintegrin Laminin Binding Protein (p67 LBP) Is Expressed on a Subset of Activated Human T Lymphocytes and , Together with the Integrin Very Late Activation Antgen-6, Mediates Avid Cellular Adherence to Laminin.* The Journal of Immunology, 1999: p. 793-802.

80. Clausse, N., Jackers, P., Jares, P., Joris, B., Sogel, M.E., Catronovo, V., *Identification of the Active Gene Coding for the Metastasis-Associated 37LRP/p40 Multifunctional Protein.* DNA and Cell Biology, 1996. **15**(12): p. 1009-1023.

81. Grosso, L.E., Park, P.W., Mecham, R.P., *Characterization of a Putative Clone for the 67 Kilodalton Elastine/Laminin Receptor Suggests That It Encodes a Cytoplasmic Protein Rather than a Cell Surface Receptor.* Biochemistry, 1991. **30**: p. 3346-3350.

82. Buto, S., Ardini, E., Ghirelli, C., Tagliabue, E., Ardini, E., Magnifico, A., van den Brule, F., Castronovo, V., Sobel, M.E., Colnaghi, M.I., Menard, S., *Formation of the 67 kDa laminin Receptor by acylation of the Precursor.* Journal of Cellular Biochemistry, 1998. **69**: p. 244-251.

83. Castronovo, V., Claysmith, A.P., Barker, K.T., Cioce, V., Krutzsch, H.C., Sobel, M.E., *Biosynthesis of the 67 kDa high affinity laminin receptor.* Biochemical and Biophysical Research Communications, 1991. **177**(1): p. 177-183.

84. Landowski, T.H., Uthayakumar, S., Starkey, J.R., *Control Pathways of the 67 kDa laminin binding protein: Surface expression and activity of tha new ligand binding domain.* Clinical &Experimental Metastasis, 1995. **13**: p. 357-372.

85. Hundt, C., Pyren, J., Haik, S., Gauczynski, S., Leucht, C., Rieger, R., Riley, M.L., Deslys, J., Dormont, D., Lasmezas, C.I., Weiss, S., *Identification of interaction domains of the prion protein with its 37-kDa/67-kDa laminin receptor.* The EMBO Journal, 2001. **20**(21): p. 5876-5886.

86. Castronovo, V., Taraboletti, G., Sobel, M.E., *Functional domains of the 67-kDa laminin reptor precursor.* The Journal of Biological Chemistry, 1991. **266**(30): p. 20440-20446.

87. Rao, C.N., Catronovo, V., Schmitt, C., Wewer, U.M., Claysmith, A.P., Liotta, L.A., Sobel, M.E., *Evidence for a Precursor of the High-Affinity Metastasis-Associated Murine Laminin Receptor.* Biochemistry, 1989. **28**(18): p. 7476-7486.

88. Wewer, U.M., Liotta, L.A., Jaye, M., Ricca, G.A., Drohan, W.N., Claysmith, A.P., Rao, C.N., Wirth, P., Coligan, J.E., Albrechtsen, R., Mudry, M., Sobel, M.E., *Altered levels of laminin receptor mRNA in various human carcinoma cells that have different abilities to bind laminin.* Proceedings of the National Academy of Science, 1986. **83**: p. 7137-7141.

89. Kazmin, D.A., Chinenov, Y., Larson, E., Starkey, J.R., *Comparative modeling of the N-terminal domain of the 67 kDa laminin-binding protien: implicaiton for putative ribosomal function.* Biochemical and Biophysical Research Communications, 2003. **300**: p. 161-166.

90. Ford, C.L., Randal-Whitis, L., Ellis, S.R, *Yeast Proteins Related to the p40/Laminin Precursor Are Required for 20S Ribosomal RNA Processing and the Maturation of 40S Ribosomal Subunits.* 59, 1999(Feb 1): p. 704-710.

91. Keppel, S., Schaller, H.C., *A 33 kDa protein with a sequence homology to the 'laminin binding protein' is assocatied with the cytoskeleton in Hydra and in mammalian cells.* Journal of Cell Science, 1991. **100**: p. 789-797.

92. Davis, S., Tzagoloff, A., Ellis, R., *Characterization of Yeast Mitochondrial Ribosomal protein Structurally Related to the Mammalian 68 kDa High Affinity Laminin Receptor.* The Journal of Biological Chemistry, 1991. **267**(8): p. 5508-5514.

93. Tohgo, A., Takasawa, S., Munakata, H., Yonekura, H., Hayashi, N., *Structural determination and characterization of a 40 kDa protein isolated from a 40S ribosomal subunit.* FASEB J, 1994. **340**(13715): p. 133-138.

94. McCaffery, P., Neve, R.L., Drager, U.C., *A dorso-ventral asymmetry in the embryonic retina defined by protien conformation.* Proceedings of the National Academy of Science, 1990. **87**: p. 8570-8574.

95. Sato, M., Saeki, Y., Tanaka, K., Kaneda, Y., *Ribosome-Associated Protein LBP/p40 Binds to S21 Protein of 40S Ribosome: Analysis Using a Yeast Two-Hybrid System.* Biochemical and Biophysical Research Communications, 1999. **256**: p. 385-390.

96. Diggle, C.P., Cruickshank, S., Olsburgh, J.D., Pellegrin, S., Smith, B., Banks, R.E., Selby, P.J., Knowles, M.A., Southgate, J., Harnden, P., *Identification of genes up-regulated in urothelial tumors.* American Journal of Pathology, 2003. **163**(2): p. 493-504.

97. Guo, N., Krutzch, H.C., Vogel, T., Roberts, D.D., *Interactions of Laminin-binding Peptide from a 33-kDa Protein Related to the 67-kDaLaminin Receptor with*

Laminin and Melanoma Cells are Heparin-dependent. The Journal of Biological chemistry, 1992. **267**(25): p. 17743-17747.

98. Smit, J.M., Waarts, B., Kimata, K., Klimstra, W.B., Bittman, R., Wilschut, J., *Adaptation of alphaviruses to heparan sulfate: interaction of the sindbis virus and semliki forest viruses with liposomes containing lipid-conjugated heparin.* Journal of Virology, 2002. **76**(20): p. 10128-10137.

99. Kazmin, D.A., Hoyt, T.R., Taubner, L., Teitze, M., Starkey, J.R., *Phage Display Mapping for Peptide 11 Sensitve Sequences Binding to Laminin-1.* Journal of Molecular Biology, 2000. **298**: p. 431-445.

100. Narasimhan, S., Armstrong, M.Y.K., Rhee, K., Edman, J.C., Richards, F.F., Spicer, E., *Gene for an extracellular matrix receptor protein from Pneumocystis carinii.* Proceedings of the National Academy of Science, 1994. **91**: p. 7440-7444.

101. Siyanova, E.Y., *Expression of LBP 32/67 kD human gene in E. coli and analysis of its binding with laminin.* Byulleten' Eksperemental'noi Biologii i Meditsini, 1992. **113**: p. 70-72.

102. Carey, J., *A systematic and general proteolytic method for defining structural and functional domains of proteins.* Methods in Enzymology, 2000. **328**: p. 499-514.

103. Thormahlen, M., Marx, A., Muller, S.A., Song, Y.-H., Mandelkow, E.M., Aebi, U., Mandelkow, E., *Interaction of monomeric and dimeric kinesin with microtubules.* Journal of Molecular Biology, 1998. **275**: p. 795-809.

104. Hjelmeland, L.M., *A nondenaturing zwitterionic detergent for membrane biochemistry: Design and synthesis.* Proceedings of the National Academy of Science, 1980. **77**(11): p. 6368-6370.

105. le Maire, M., Viel, A., Moller, J.V., *Size exclusion chromatography and universal calibration of gel columns.* Analytical biochemistry, 1989. **177**: p. 50-56.

106. Cavanaugh, J., Fairbrother, W.J., Palmer III, A.G., Skelton, N.J., *Protein NMR Spectroscopy: Principles and Practices*. 1996, San Diego, CA: Academic Press Inc.

107. van Holde, K.E., Johnson, W.C., Ho, P.S., *Principles of Physical Biochemistry*. 1998, New Jersey: Prentice-Hall Inc.

108. Huth, J.R., Bewley, C.A., Jackson, B.M., Hinnebusch, A.G., Clore, G.M., Gronenborn, A.M., *Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR.* Protein Science, 1997. **6**: p. 2359-2364.

109. Markossian, K.A., Kurganov, B.I., *Protein folding, misfolding and aggregation. Formation of inclusion bodies and aggresomes.* Biochemistry (Moscow), 2004. **69**(9): p. 971-984.

110. Grune, T., Jung, T., Merker, K., Davies, K.J.A., *Decreased proteolysis caused by protein aggregates, inclusion, bodies, plaques, lipfuscicn, ceroid, and aggresomes during oxidative stress, aging, and disease.* The Internation Journal of Biolchemistry and Cell Biology, 2004. **36**: p. 2519-2530.

111. Dobson, C.M., *Principles of protein folding, misfolding and aggregation.* Seminars in Cell and Developmental Biology, 2004. **15**: p. 3-16.

112. Walter, S., Buchner, J., *Molecular Chaperones--Cellular machines for protein folding.* Angewandte Chemie, 2002. **41**: p. 1098-1113.

113. Matsuo, H., Li, H., McGuire, C.M., Gingras, A., Sonen berg, N., Wagner, G., *Structure of translation factor elF4E bound to m7GDP and interaction with 4E-binding protein.* Nature Structural Biology, 1997. **4**(9): p. 717-722.

114. McGuire, A.M., Matsuo, H., Wagner, G., *Internal and overall motions of the translation factor elF4E: Cap binding and insertion in a CHAPS detergent micelle.* Journal of Biomolecular NMR, 1998. **12**: p. 73-88.

115. Maachupalli-Reddy, J., Kelley, B.D., De Bernardez Clark, E., *Effect of inclusion body contaminants on the oxidative renaturation of hen egg white lysozyme.* Biotechnology Progress, 1997. **13**: p. 144-150.

116. Kramer, G., Rauch, T., Rist, W., Vorderwulbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, El, Bukau, B., *L23 protein functions as a chaperone docking site on the ribosome.* Nature, 2002. **419**: p. 171-174.

117. Wilkinson, B., Gilbert, H.F., *Protein disulfide isomerase.* Biochimica et Biophysica Acta, 2004. **1699**: p. 35-44.

118. Salama, R.H., Muramatsu, H., Zou, K., Inui, T., Kimura, T. Muramatsu, T., *Midkine binds to 37 kDa laminin binding protein precursor, leading to nuclear transport of the complex.* Experimental Cell Research, 2001. **270**: p. 13-20.

119. Hubbard, S.J., *The structural aspects of limited proteolysis of native proteins.* Biochimica et Biophysica Acta, 1998. **1382**: p. 191-206.

120. Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., Sali, A., *Comparative protein structure modeling of genes and genomes.* Annual Review in Biophysical and Biolomolecular Structure, 2000. **29**: p. 291-325.

121. Schonbrun, J., Wedemeyer, W.J., Baker, D., *Protein structure prediction in 2002.* Current Opinion in Structural Biology, 2002. **12**: p. 348-354.

122. Ban, N., Nissen, P., Hansen, J., Moore, P.B., Steitz, T.A., *The complete atomic structure of the large ribosomal subunit at 2.4 A resolution.* Science, 2000. **289**: p. 905-920.

123. Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang. C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., Obradovic, Z., *Intrinsically disordered protein.* Journal of Molecular Graphics and Modelling, 2001. **19**: p. 26-59.

124. Wright, P.E., Dyson, H.J., *Instrinsically unstructured proteins: re-assessing the protein structure-function paradigm.* Journal of Molecular Biology, 1999. **293**: p. 321-331.

125. Hayashi, Y., Katada, J., Sato, Y., Igarashi, K., Takiguchi, Y., Harada, T., Muramatsu, M., Yasuda, E., Uno, I., *Discovery and structure-activity relationship*

studies of a novel and specific peptide motif PXXXDX as a platelet fibrinogen receptor antagonist. Bioorganic and Medicinal Chemistry, 1998. **6**: p. 353-364.

126. Katada, J., Hayashi, Y., Sato, Y., Muramatsu, M., Takiguchi, Y., Harada, T., Fujiyoshi, T., Uno, I., *A novel peptide motif for platelet fibrinogen receptor recognition.* Journal of Biological Chemistry, 1997. **272**(12): p. 7720-7726.

127. Browder, T., Folkman, J., Pirie-Shepherd, S., *The hemostatic system as a regulator of angiogenesis.* The Journal of Biological Chemistry, 2000. **275**(3): p. 1521-1524.

128. Daly, M.E., Makris, A., Reed, M., Lewis, C.E., *Hemostatic regulators of tumor angiogenesis: a source of antiangiogenic agents for cancer treatment?* Journal of the National Cancer Institute, 2003. **95**(22): p. 1660-1673.

129. Zumft, W.G., *Cell biology and molecular basis of denitrification.* Microbiology and Molecular Biology Reviews, 1997. **61**(4): p. 533-616.

130. Bothe, H., Jost, G., Schloter, M., Ward, B.B., Witzel, K., *Molecular analysis of ammonia oxidation and denitrification in natural environments.* FEMS Microbiology Reviews, 2000. **24**: p. 673-690.

131. Ferguson, S.J., *Nitrogen cycle enzymology.* Current Opinion in Chemical Biology, 1998. **2**: p. 182-193.

132. Berks, B.C., Ferguson, S.J., Moir, J.W.B., Richardson, D.J., *Enzymes and associated electron transport systems that catalyze the respiratory reduction of nitrogen oxides and oxyanions.* Biochimica et Biophysica Acta, 1995. **1232**: p. 97-173.

133. Richardson, D.J., Watmough, N.J., *In organic nitrogen metabolism in bacteria.* Current Opinion in Chemical Biology, 1999. **3**: p. 207-219.

134. Hiraishi, A., Khan, S.T., *Application of polyhydroxyalkoanoates for denitrification in water and wastewater treatment.* Applied Microbiological Biotechnology, 2003. **61**: p. 103-109.

135. Philips, S., Laanbroek, H.J., Verstraete, W., *Origin, causes and effects of increased nitrite concentrations in aquatic environments.* Reviews in Environmental and Biotechnology, 2002. **1**: p. 115-141.

136. Pinckney, J.L., Paerl, H.W., Tester, P., Richardson, T.L., *The role of nutrient loading and eutrophication in estuarine ecology.* Environmental Health Perspectives, 2001. **109**(5): p. 699-706.

137. Haltia, T., Brown, K., Tegoni, M., Cambillau, C., Saraste, M., Mattila, K., Djinovic-Carugo, K., *Crystal structure of nitrous oxide reductase from Paracoccus denitrificans at 1.6A resolution.* Biochemical Journal, 2003. **369**: p. 77-88.

138. Rosch, C., Mergel, A., Bothe, H., *Biodiversity of denitrifying and dinitrogen-fixing bacteria in an acid forest soil.* Applied and Environmental Microbiology, 2002. **68**(8): p. 3818-3829.

139. Phillipot, L., *Denitrifying genes in bacterial and Archaeal genomes.* Biochimica et Biophysica Acta, 2002. **1577**: p. 355-376.

140. Moura, I., Moura, J.J.G., *Structural aspects of denitrifying genes.* Current Opinion in Chemical Biology, 2001. **5**: p. 168-175.

141. Zumft, W.G., Viebrock-Sambale, A., Braun, C., *Nitrous oxide reductase from denitrifying Pseudomonas stutzeri.* European Journal of Biochemistry, 1990. **192**: p. 591-599.

142. Rees, D.C., *Great metalloclusters in enzymology.* Annual Review in Biochemistry, 2002. **71**: p. 221-246.

143. Chen, P., Cabrito, I., Moura, J.J.G., Moura, I., Solomon, E.I., *Spectroscopic and electronic structure studies of the u4-sulfide bridged tetranuclear Cuz cluster in N2O reductase: Molecular insight into the catalytic mechanism.* Journal of the American Chemical Society, 2002. **124**: p. 10497-10507.

144. Dreusch, A., Burgisser, D.M., Heizmann, C.W., Zumft, W.G., *Lack of copper insertion into unprocessed cytoplasmic nitrous oxide reductase by an R20D*

substitution in the arginine consensus motif of the signal peptide. Biochimica et Biophysica Acta, 1997. **1319**: p. 311-318.

145. Voordouw, G., *A universal system for the transport of redox proteins: early roots and latest developments.* Biophysical Chemistry, 2000. **86**: p. 131-140.

146. Robinson, C., Bolhuis, A., *Protein targeting by the twin-arginine translocation pathway.* Nature Reviews Molecular Cell Biology, 2001. **2**: p. 350-356.

147. Heikilla, M.P., Honisch, U., Wunsch, P., Zumft, W.G., *Role of the Tat transport system in nitrous oxide reductase translocation and cytochrome cd1 biosynthesis in Pseudomonas stutzeri.* Journal of Bacteriology, 2001. **183**(5): p. 1663-1671.

148. Wunsch, P., Herb, M., Wieland, H., Schiek, U.M., Zumft, W.G., *Requirements for Cua and CU-S center assembly of nitrous oxide reductase deduced from complete periplasmic enzyme maturation in the nondenitrifier Pseudomonas putida.* Journal of Bacteriology, 2003. **185**(3): p. 887-896.

149. Brown, K., Tegoni, M., Prudencio, M.,. Pereira, A.S., Besson, S., Moura, J.J., Moura, I., Cambillau, C., *A novel type of catalytic copper cluster in nitrous oxide reductase.* Nature Structural Biology, 2000. **7**(3): p. 191-195.

150. Brown, K., Djinovic-Carugo, K., Haltia, T., Cabrito, I., Saraste, M., Moura, J.J.G., Moura, I., Tegoni, M., Cambillau, C., *Revisiting the catalytic Cuz cluster of nitrous oxide reductase.* Journal of Biological Chemistry, 2000. **275**(52): p. 41133-41136.

151. Rosenzweig, A.C., *Nitrous oxide reductase from CuA to CuZ.* Nature Structural Biology, 2000. **7**(3): p. 169-171.

152. Simon, J., Einsle, O., Kroneck, P.M.H., Zumft, W.G., *The unprecendented nos gene cluster of Wolinella succinogenes encodes a novel respiratory electron transfer pathway to cytochrome c nitrous oxide reductase.* FEBS Letters, 2004. **569**: p. 7-12.

153. Jawad, Z., Paoli, M., *Novel sequences propel familiar folds.* Structure, 2002. **10**: p. 447-454.

154. Rasmussen, T., Berks, B.C., Sanders-Loehr, J., Dooley, D.M., Zumft, W.G., Thomson, A.J., *The catalytic center in nitrous oxide reductase, Cuz, is a copper-sulfide cluster.* Biochemistry, 2000. **39**: p. 12753-12756.

155. Ghosh, S., Gorelsky, S.I., Chen, P., Cabrito, I., Moura, J.J.G., Moura, I., Solomon, E.I., *Activation of N2O reduction by the fully reduced µ4-sulfide-bridged tetranuclear Cuz cluster active site.* Angew. Chemie, 2003. **43**: p. 15708-15709.

156. Chen, P., Gorelsky, S.I., Ghosh, S., Solomon, E.I., *N2O reduction by the u4-sulfide-bridged tetranuclear Cuz cluster active site.* Angewandte Chemie, 2004. **43**: p. 4132-4140.

157. Bouige, P., Laurent, D., Piloyan, L., Dassa, E., *Phylogenetic and funtional classificaiton of ATP-binding cassette (ABC) systems.* Current Protein and Peptide Science, 2002. **3**: p. 541-559.

158. Martinoia, E., Klein, M., Geisler, M., Bovet, L., Forestier, C., Kolukisaoglu, U., Muller-Rober, B., Schulz, B., *Multifunctionality of plant ABC-transporters--more than just detoxifiers.* Planta, 2002. **214**: p. 345-355.

159. Honisch, U., Zumft, W.G., *Operon structure and regulation of the nos gene region of Pseudomonas stutzeri, encoding an ABC-type ATPase for maturation of nitrous oxide reductase.* Journal of Bacteriology, 2003. **185**(6): p. 1895-1902.

160. Ciccarelli, F.D., Copley, R.R., Doerks, T., Russell, R.B., Bork, P., *CASH--a beta-helix domain widespread among carbohydrate-binding proteins.* TRENDS in Biochemcial Sciences, 2002. **27**(2): p. 59-63.

161. McGuirl, M.A., Bollinger, J.A., Cosper, N., Scott, R.A., Dooley, D.M., *Expression, purification, and characterization of NosL, a novel Cu(I) protein of the nitrous oxide reductase (nos) gene cluster.* Journal of Biological Inorganic Chemistry, 2001. **6**: p. 189-195.

162. Velasco, L., Mesa, S., Xu, C., Delgado, M.J., Bedmar, E.J., *Molecular characterization of nosRZDYFLX genes coding for denitrifying nitrous oxide reductase of Bradyrhizobium japonicum.* Antoine van Leeuwenhoek, 2004. **85**: p. 229-235.

163. Chan, Y., McCormick, W.A., Watson, R.J., *A new nos gene downstream from nosDFY is essential for dissimilatory reduction of nitrous oxide by Rhizobium (Sinorhizobium) meliloti.* Microbiology, 1997. **143**: p. 2817-2824.

164. Wunsch, P., Zumft, W.G., *Functional domains of NosR, a novel transmembrane iron-sulfur flavoprotein necessary for nitrous oxide respiration.* Journal of Bacteriology, 2005. **187**(6): p. 1992-2001.

165. Beinert, H., Kiley, P.J., *Fe-S proteins in sensing and regulatory functions.* Current Opinion in Chemical Biology, 1999. **3**: p. 152-157.

166. Arai, H., Mizutani, M., Igarashi, Y., *Transcriptional regulation of the nos gene for nitrous oxide reductase in Pseudomonas aeruginosa.* Microbiology, 2003. **149**: p. 29-36.

167. Lee, H.S., Hancock, R.E., Ingraham, J.L., *Properties of a Psuedomonas stutzeri outer membrane channel-forming protein (NosA) required for production of copper-containing N2O reductase.* Journal of Bacteriology, 1989. **171**(4): p. 2096-2100.

168. Mokhele, K., Tang, Y.J., Clark, M.A., Ingraham, J.L., *A Pseudomonas stutzeri outer membrane protein inserts copper into N2O reductase.* Journal of Bacteriology, 1987. **169**(12): p. 5721-5726.

169. Narita, S., Matsuyama, S., Tokuda, H., *Lipoprotein trafficking in Escherichia coli.* Archives of Microbiology, 2004. **182**: p. 1-6.

170. Hanna, P.M., Tamilarasan, R., McMillin D.R., *Cu(I) analysis of blue copper proteins.* Biochemical Journal, 1988. **256**: p. 1001-1004.

171. Higgins, D.G., Sharp, P.M., *CLUSTAL: a package for performing multiple sequence alignment on a microcomputer.* Gene, 1988. **73**: p. 237-244.

172. O'Halloran, T.V., Culotta, V.C., *Metallochaperones, an intracellular shuttle service for metal ions.* The Journal of Biological Chemistry, 2000. **275**(33): p. 25057-25060.

173. Finney, L.A., O'Halloran, T.V., *Transition metal speciation in the cell: Insights from the chemistry of metal ion receptors.* Science, 2003. **300**: p. 40-71.

174. Huffman, D.L., O'Halloran, T.V., *Function, structure, and mechanism of intracellular copper trafficking proteins.* Annual Review in Biochemistry, 2001. **70**: p. 677-701.

175. Serre, L., Rossy, E., Pebay-Peyroula, E., Cohen-Addad, C., Coves, J., *Crystal structure of the oxidized form of the periplasmic mercury-binding protein MerP from Ralstonia metallidurans CH34.* Journal of Molecular Biology, 2004. **339**: p. 161-171.

176. Rosenzweig, A.C., O'Halloran, T.V., *Structure and chemistry of the copper chaperone proteins.* Current Opinion in Chemical Biology, 2000. **4**: p. 140-147.

177. Arnesano, F., Banci, L., Bertini, I., *Solution structure of the Cu(I) and apo forms of the yeast metallochaperone, Atx1.* Biochemistry, 2001. **40**: p. 1528-1539.

178. Pufahl, R.A., Singer, C.P., Peariso, K.L., Lin, S.-J., Schmidt, P.J., Fahrni, C.J., Culotta, V.C., Penner-Hahn, J.E., O'Halloran, T.V., *Metal ion chaperone function of the soluble Cu(I) receptor Atx1.* Science, 1997. **278**: p. 853-856.

179. Banci, L., Bertini, I., Cantini, R., Ciofi-Baffoni, S., Gonnelli, L., Mangani, S., *Solution structure of Cox11, a novel type of immunoglobulin-like fold involved in CuB site formation of cytochrome c oxidase.* The Journal of Biological Chemistry, 2004. **279**(33): p. 34833-34839.

180. Lamb, A.L., Wernimont, A.K., Pufahl, R.A., Culotta, V.C., O'Halloran, T.V., Rosenzweig, A.C., *Crystal structure of the copper chaperone for superoxide dismutase.* Nature Structural Biology, 1999. **6**(8): p. 724-729.

181. Rae, T.D., Torres, A.S., Pufahl, R.A., O'Halloran, T.V., *Mechanism of Cu, Zn-superoxide dismutase activation by the human metallochaperone hCCS.* The Journal of Biological Chemistry, 2001. **276**(7): p. 5166-5176.

182. Brown, N.M., Torres, A.S., Doan, P.E., O'Halloran, T.V., *Oxygen and the copper chaperone CCS regulate posttranslational activation of Cu,Zn superoxide*

dismutase. Proceedings of the National Academy of Science, 2004. **101**(15): p. 5518-5523.

183. Abajian, C., Yatsunyk, L.A., Ramirez, B.E., Rosenzweig, A.C., *Yeast Cox17 solution structure and copper(I) binding.* The Journal of Biological Chemistry, 2004. **279**(51): p. 53584-53592.

184. Punter, F.A., Glerum, D.M., *Mutagenesis reveals a specific role for Cox17p in copper transport to cytochrome oxidase.* The Journal of Biological Chemistry, 2003. **278**(33): p. 30875-30880.

185. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., Bax, A., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* Journal of Biomolecular NMR, 1995. **6**(3): p. 277-293.

186. Cornilescu, G., Delaglio, F., Bax, A., *Protein backbone angle restraints from searching a database for chemical shift and sequence homology.* Journal of Biomolecular NMR, 1999. **13**: p. 289-302.

187. Johnson, B.A., Blevins, R.A., *NMRView: a computer program for the visualization and analysis of NMR data.* Journal of Biomolecular NMR, 1994. **4**: p. 603-614.

188. Nilges, M., Macias, M.J., O'Donoghue, S.I., Oschkinat, H., *Automated NOESY interpretation with ambiguous distance restraints: The refined NMR solution structure of the pleckstrin homology domain from beta-spectrin.* Journal of Molecular Biology, 1997. **269**: p. 408-422.

189. Brunger, A.T., Adams, P.D., Clore, M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jian, J.,Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L., *Crystallography and NMR system: A new software suite for macromolecular structure determination.* Acta Crystallography, 1998. **D54**: p. 905-921.

190. Linge, J.P., Williams, M.A., Spronk, C.A.E.M., Bonvin, A.M.J.J., Nilges, M., *Refinement of protein structures in explicit solvent.* Proteins: Structure, function, and genetics, 2003. **50**: p. 496-506.

191. Stein, E.G., Rice, L.M., Brunger, A.T., *Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation.* Journal of Magnetic Resonance, 1997. **124**: p. 154-164.

192. Downing, A.K., ed. *Protein NMR Techniques*. NOE Assignment with ARIA 2.0, ed. M. Habeck. Vol. 278. 2004, Humana Press: Totowa, New Jersey.

193. Linge, J.P., Habeck, M., Rieping, W., Nilges, M., *ARIA: automated NOE assignment and NMR structure calculation.* Bioinformatics, 2003. **19**(2): p. 315-316.

194. Korandi, R., Rullmann, J.A., Wuthrich, K., *MOLMOL: a program for display and analysis of macromolecular structures.* Journal of Molecular Graphics, 1996. **14**: p. 51-55.

195. Nicholls, A., Sharp, K., Honig, B., *Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.* Proteins, 1991. **11**(4): p. 281-296.

196. Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., Thornton, J.M., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* Journal of Biomolecular NMR, 1996. **8**(4): p. 477-486.

197. Holm, L., Sander, C., *Alignment of three-dimensional protein structures: network server for database searching.* Methods in Enzymology, 1996. **266**: p. 653-662.

198. Grzesiek, S., and Bax, A., *Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR.* Journal of the American Chemical Society, 1992.

199. Wittekind, M., Mueller, L., *HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha and beta carbon resonances in proteins.* Journal of Magnetic Resonance Series B, 1993. **101**: p. 201-205.

200. Grzesiek, S., and Bax, A., *Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins.* Journal of Biomolecular NMR, 1993. **3**: p. 185-204.

201. Kay, L.E., Ikura, M., Tschudin, R., Bax A., *Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins.* Journal of Magnetic Resonance, 1990. **89**: p. 496-514.

202. Grzesiek, S., Anglister, J. and Bax, A., *Correlation of backbone amide and aliphatic side-chain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization.* Journal of Magnetic Resonance Series B, 1993. **101**: p. 114-119.

203. Kay, L.E., Xu, G., Singer, A.U., Muhandram, D.R., Forman-Kay, J.D., *A gradient-enhanced HCCH-TOCSY experiment for recording side-chain 1H and 13C correlations in H2O samples of proteins.* Journal of Magnetic Resonance Series B, 1993. **101**: p. 333-337.

204. Vuister, G.W., Bax, A., *Resolution enhancement and spectral editing of uniformly 13C-enriched proteins by homonuclear broadband 13C decoupling.* Journal of Magnetic Resonance, 1992. **98**: p. 428-435.

205. Taubner, L.M., McGuirl, M.A., Dooley, D.M., Copie, V., *Letter to the Editor: 1H, 13C, 15N backbone and sidechain resonance assignments sof apo-NosL, a novel copper(I) binding protein from the nitrous oxide reductase gene cluster of Achromobacter cycloclastes.* Journal of Biomolecular NMR, 2004. **29**: p. 211-212.

206. Wishart, D.S., Sykes, B.D., *The 13C chemical shift index: A simple method for the identification of protein secondary structure using 13C chemical-shift data.* Journal of Biomolecular NMR, 1994. **4**: p. 171-180.

207. Wishart, D.S., Nip A.M., *Protein chemical shift analysis: a practical guide.* Biochemical and Cell Biology, 1998. **76**: p. 153-163.

208. Wuthrich, K., *NMR of proteins and nucleic acids*. 1986, New York: John Wiley and Sons, Inc.

209. Orengo, C.A., Todd, A.E., Thornton, J.M., *From Protein Structure to Function.* Current Opinion in Structural Biology, 1999. **9**: p. 374-382.

210. Quintero-Monzon, O., Rodal, A.A., Strokopytov, B., Almo, S.C., Goode, B.L., *Structural and functional dissection of the Abp1 ADFH actin-binding domain reveals versatile in vivo adapter functions.* Molecular Biology of the Cell, 2005. **16**: p. 3128-3139.

211. Brown, R.S., *Zinc finger proteins: getting a grip on RNA.* Current Opinion in Structural Biology, 2005. **15**: p. 94-98.

212. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.G. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M., *The protein data bank: a computer-based archival file for macromolecular structures.* Journal of Molecular Biology, 1977. **112**: p. 535-542.

213. DiLello, P.D., Benison, G.C., Valafar, H., Pitts, K.E., Summers, A.O., Legault, P., Ominchinski, J.G., *NMR structural studies reveal a novel protein fold for MerB, the organomercurial lyase involved in the bacterial mercury resistance system.* Biochemistry, 2004. **43**: p. 8322-8332.

214. Benison, G.C., DiLello, P., Shokes, J.E., Cosper, N.J., Scott, R.A., Legault, P., Ominchinski, J.G., *A stable mercury-containing complelx of the organomercurial lyase MerB: Catalysis, product release, and direct transfer to MerA.* Biochemistry, 2004. **43**: p. 8333-8345.

215. Lippard, S.J., Berg, J.M., *Principles of Bioinorganic Chemistry*. 1994, Mill Valley, CA: University Science Books.

216. Solioz, M., Stoyanov, J.V., *Copper homeostasis in enterococcus hirae.* FEMS Microbiology Reviews, 2003. **27**: p. 183-195.

APPENDICES

APPENDIX A
Complete Chemical Shift Assignments for Apo NosL

APPENDIX A

Complete Chemical Shift Assignments for Apo NosL. Carbon shifts are referenced to DSS and nitrogen shifts are referenced to liquid ammonia.

Residue	H_N	^{15}N	CO	Cα	Hα	Hα	Cβ	Hβ	Hβ	Cγ	Cγ
Met 1											
Asp 2			175.9	54.64	4.65		41.4	2.71	2.6		
Lys 3	8.38	121.5	176.59	56.81	4.23		33.12	1.79		24.51	
Glu 4	8.48	121.42	176.33	56.9	4.22		30.12	1.97		36.39	
Asp 5	8.24	121.48		54.46	4.56		41.08	2.62			
Val 6	7.95	120.48	176.31	62.84	4.03		32.57	2.08		23.77	20.85
Ala 7	8.24	126.15	177.9	53.03	4.19		18.95	1.35			
Gln 8	8.05	118.26	175.97	56.1	4.22		29.27	2		33.84	
Ser 9	8.12	116.43	174.44	58.42	4.35		63.76	3.79			
Ile 10	7.88	122.57	175.03	60.57	3.65		37.49	1.43		27.06	13.37
Val 11	7.47	126.28		59.58	4.11		32.63	1.89			20.76
Pro 12			176.8	62.62	3.78		31.32	0.92	0.54	26.48	
Gln 13	9.22	122.88	175.1	54.15	4.37		31.02	2.18	1.68	34.19	
Asp 14	8.32	117.82	175.38	54.13	5.33		43.05	2.85	2.42		
Met 15	8.41	118.64		56.37			33.95				
Thr 16				59.2	4.9		69.05	4.7		21.1	
Pro 17			176.18	62.39	4.72		34.4	2.08	2.34	24.76	
Glu 18	8.57	122.42	176.66	56.45	4.19		30.09	2.07	1.97	36.86	
Thr 19	7.61	117.36		64.66			68.84				
Leu 20			177.53								
Gly 21	8.22	106.67	174.77	45.79	3.87						
His 22	7.8	119.64		56.43	5.2		30.7				
Tyr 23											
Cys 24											

APPENDIX A CONTINUED

Residue	HN	15N	CO	Cα	Hα	Hα	Cβ	Hβ	Hβ	Cγ	Cγ
Gln 25											
Met 26											
Asn 27											
Leu 28											
Leu 29			176.33	56.92	4.08		41.67	1.62	1.45	27.06	
Glu 30	7.65	114.88	175.73	55.97	4.21		29.96	2.04	1.81	36.46	
His 31	7.71	119.38		53.54	5.06		30.71	3.1	3.25		
Pro 32			175.55	63.02	4.71		33.89	2.5	2.04	27.05	
Gly 33	7.85	105.94		43.45	4.02	3.98					
Pro 34			174.35	62.75	4.3		31.73	1.81		26.99	
Lys 35	8.58	117.42	174.8	53.37	4.74		35.94	2.43	1.9	29.19	
Ala 36	7.74	115.89	175.21	50.05	5.9		26.49	1.38			
Gln 37	9.33	116.98	173.16	53.39	5.52		33.3	0.96	0.84	34.91	
Ile 38	9.07	121.7	174.52	60.97	4.69		41.81	1.53	1.48	18.27	14.71
Phe 39	8.36	124.12	175.06	57.41	4.77		40.98	2.81			
Leu 40	9.54	124.64	177.64	52.67	4.95		43.99	1.62	1.31	27.1	
Glu 41	8.71	123.56	177.53	58.34	3.99		30.59	1.94	1.86	36.42	
Gly 42	8.98	114.82	174.17	45.43	4.22	3.69					
Ser 43	7.46	115.26		55.29	5.11		64.74	3.9	3.83		
Pro 44			175.81	64.16	4.49		32.33	2.3	2.04	26.88	
Ala 45	7.47	120.92		48.78	4.84		19.27	1.25			
Pro 46			175.41	61.59	4.1		31.06	0.41	-0.25	25.76	
Leu 47	8.41	118.6	176.12	53.71	4.15		44.09	1.27			
Phe 48	8.67	120.9	175.02	57.57	5.19		42.98	2.74	2.57		
Phe 49	8.92	117.1		57.32	5.09		43.24	2.62			
Ser 50			173.33	58.19	4.39		64.18	3.81	3.05		
Gln 51	8.06	121.15	177.06	54.8	5.32		32.18	2.53	2.1	36.44	
Val 52	10.46	130.78	180.22	67.21	3.37		31	2.61		23.84	22.27
Arg 53	10.54	119.1	176.81	59.49	3.71		29.88	2.09	1.84	26.42	

APPENDIX A CONTINUED

Residue	HN	15N	CO	Cα	Hα	Hα	Cβ	Hβ	Hβ	Cγ	Cγ
Asp 54	6.81	116.62	177.93	56.78	4.59		41.07	3.23	3		
Ala 55	7.11	123.98	178.17	53.87	2.57		19.1	0.9			
Ile 56	7.63	116.27	176.61	65.16	3.25		37.12	1.58			17.58
Ala 57	7.99	120.62	180.7	55.64	3.7		18.49	1.48			
Tyr 58	8.22	120.49	178.58	61.92	3.86		38.27	3.11	2.86		
Ala 59	7.86	120.55	177.98	54.1	3.88		18.62	1.37			
Arg 60	7.57	114.88	176.79	55.13	4.31		30.69	1.94	1.59	26.81	
Gly 61	7.35	108.32		44.88	4.17	3.82					
Pro 62				63.83	4.38		32.23	2.19	1.96	26.96	
Glu 63	8.09	119.72	175.58	56.04	4.29		29.6	1.91		36.36	
Gln 64	8.01	122.56	176.53	54.75	4.26		29.42	2.01	1.75	33.15	
Ile 65	8.22	121.8	175.28	62.8	4.03		39.06	1.89		27.57	17.69
Ala 66	7.61	122.34		50.37	4.7		20.58	1.48			
Pro 67			175.87	63.15	4.33		32.31	2.26	1.74	27.83	
Ile 68	8.29	122.75	177.22	62.27	4.03		39.15	1.74			17.59
Leu 69	9.27	128.4	176.76	56.29	4.36		43.59	1.34	1.46	26.61	
Val 70	7.21	114.57	171.06	61.12	4.22		33.09	2.11		20.29	
Ile 71	7.66	124.59	175.17	59.96	4.83		40.19	1.58		19.14	15.5
Tyr 72	9.37	124.25	175.45	56.45	5.71		42.77	3.19	2.54		
Val 73	9.68	111.74	175.5	58.36	5.75		36.5	2.33		23	18.94
Asn 74	7.31	118.74	172.36	53.89	4.62		42.62	2.23	1.92		
Asp 75	6.61	112.49	178.17	52	4.16		41.11	3.51	2.28		
Met 76	8.43	124.06	176.39	54.03	4.47		25.95	1.82	0.91	33.21	
Gly 77	8.35	107.73	174.13	44.36	4.69	4					
Ala 78	6.94	124.45	178.11	52.18	4.14		18.75	1.33			
Ala 79	8.59	125.35	179.24	53.97	4.13		17.87	1.35			
Gly 80	8.72	111.02	173.94	45.25	4.11	3.69					
Ala 81	7.71	123.07	177.85	53	4.25		20.3	1.25			
Thr 82	7.61	113.96	173.6	59.31	4.7		72.38	4.6		21.83	

APPENDIX A CONTINUED

Residue	HN	15N	CO	Cα	Hα	Hα	Cβ	Hβ	Hβ	Cγ	Cγ
Trp 83	8.75	119.92	176.54	60.99	3.85		29.52	3.36	3.17		
Asp 84	7.92	110.86	176.4	55.91	4.35		41.37	2.65	2.52		
Glu 85	7.53	117		53.15	4.55		30.01	1.82			
Pro 86			177.44	65.23	4.07		32.03	1.75	1.48	27.96	
Gly 87	8.02	102.95	174.64	43.99	4.28	3.42					
Asp 88	8.49	126.58	177.94	55.73	4.4		39.93	2.59	2.39		
Gly 89	10.51	112.97	173.96	45.79	4.01	4.13					
Asn 90	6.74	118.83	172.95	51.06	4.53		36.41	2.53	2.33		
Trp 91	7.27	121.57	175.54	54.6	5.24		33.98	2.71	2.57		
Ile 92	9.72	118.1	175.15	59.05	4.99		42.29	1.99			18.63
Ala 93	9.05	127.12	179.96	52.83	4.48		18.58	1.43			
Ala 94	8.84	123.63	180.59	54.17	3.98		18.85	1.31			
Asp 95	8.85	113.14	176.55	55.2	4.37		39.24	2.79	2.57		
Lys 96	7.46	118.32	175.41	54.76	4.43		32.82	2.01	1.77	24.73	
Ala 97	6.65	121.19	175.06	51.13	4.27		19.73	0.84			
Phe 98	8.82	117.63	174.26	57.34	4.53		41.41	2.69	2.46		
Tyr 99	9.39	118.7	176.02	56.38	5.34		41.77	2.75	2.24		
Val 100	9.67	121.04	177.73	62.06	5.19		32.68	2.19		23.27	
Val 101	8.96	128.83	176.21	60.82	4.97		34.71	2.05		20.99	
Gly 102	9.93	113.62	175.51	45.43	4.31	3.93					
Ser 103	8.79	117.85	174.41	58.92	4.27		65.65	3.89			
Ala 104	8.66	119.66	178.31	53.3	4.23		19.22	1.52			
Arg 105	7.76	117.65	175.8	55.34	4.12		30.48	1.43	1.52	26.52	
Arg 106	8.51	123.84	176.9	54.2	4.56		31.74	1.83	1.48	27.61	
Gly 107	8.98	105.08	176.77	44.31	4.64	3.95					
Gly 108	9.71	108.5	175	48.48	4.22	4.03					
Met 109	8.32	115.96	177.48	53.64	4.88		32.12	1.85	2.34	32.7	
Gly 110	8.45	108.1	173.67	45.62	4.35	3.71					
Ala 111	7.88	123.84		50.34	4.92		18.89	1.38			

APPENDIX A CONTINUED

Residue	HN	15N	CO	Cα	Hα	Hα	Cβ	Hβ	Hβ	Cγ	Cγ
Pro 112			175.31	64.76	4.42		31.8	2.52	1.88	28.32	
Glu 113	8.77	125.3	174.93	52.74	5.66		34.77	1.82	2.06	37.06	
Ala 114	10.08	132.32	174.5	49.67	4.95		18.63	1.01			
Val 115	8.71	123.57		61.06	4.31		33.23	2.2			
Pro 116			175.8	61.58	4.79		33.2	1.87	1.63	28.99	
Phe117	8.84	115.89	175.91	57.29	4.84		44.53	3.35	2.49		
Ser 118	8.28	116.04	175.26	59.88	4.4		63.64	3.82	4		
Ser 119	8.17	115.82	175.21	55.35	4.97		64.47	4.07	3.85		
Arg 120	8.84	129.59	177.72	59.02	3.05		29.18	0.86	1.19	27.53	
Asp 121	7.89	117.93	179.27	57.11	4.25		40.2	2.57	2.38		
Glu 122	7.43	121.02	178.79	58.29	4.06		29.1	2.14	1.92	35.83	
Ala 123	7.34	123.76	178.35	54.18	4.01		18.16	1.03			
Ala 124	8.11	118.95	180.53	54.9	3.77		17.84	1.38			
Ala 125	7.51	121.44	180.11	55.05	4.03		17.73	1.46			
Phe126	7.62	122.79		61.57	4.32		39.97	3.18	2.95		
Val 127	8.08	120.56	179.92	64.71	3.88		31.65	2.06		24.7	21.74
Leu 128	7.59	120.62	177.98	57.88	3.94		41.54	1.8	1.57	26.59	
Ala 129	7.03	118.35	179.45	54.1	4.26		20.28	1.36			
Glu 130	8.42	114.08	177.8	55.68	4.56		31.44	0.89	1.96	33.55	
Gly 131	8.67	112.25	171.16	44.1	4.28	3.71					
Gly 132	7.57	99.92	173.42	43.71	4.25	3.53					
Gln 133	8.59	115.56	173.56	54.17	4.76		32.84	1.97		33.49	
Val 134	8.49	121.79	175.98	62.46	4.81		32.22	1.97		22.35	
Leu 135	9.89	130.26	176	54.08	4.98		46.91	1.82	1.55	27.41	
Ala 136	8.82	123.46	178.41	51.11	4.77		19.77	1.51			
Leu 137	9.01	121.67	178.97	58.58	3.64		41.6	1.72	1.68	26.56	
Ala 138	8.15	115.07	178.28	53.67	3.94		18.72	1.31			
Asp 139	7.53	114.33	175.55	54.36	4.61		43.17	2.92	2.81		
Ile 140	7.36	120.79	177.05	62.19	4.07		37.55	1.77		28.08	18

APPENDIX A CONTINUED

Residue	HN	15N	CO	Cα	Hα	Hα	Cβ	Hβ	Hβ	Cγ	Cγ
Thr 141	8.12	122.07	175.53	60.65	4.39		70.44	4.6		22.1	
Asp 142	8.64	120.95	178.12	57.38	4.12		39.5	2.64	2.57		
Ala 143	8.11	119.66	177.98	54.32	4.06		18.57	1.35			
Met 144	7.55	114.92	177.26	58.12	4.08		33.22	2.24	1.89	33.4	
Val 145	7.23	108.66		61.49	4.35		32.84	2.38		21.34	18.65
Leu 146	7.95	120.48	177.25	54.82	4.41		42.71	1.61	1.47	26.91	
Thr 147	7.91	117.67		60.57	4.45		69.97	4.08			21.47
Pro 148			176.77	63.1	4.42		32.2	2.27	1.87	27.44	
Val 149	8.17	120.32	176.22	62.5	4.04		32.87	2.02		23.95	21.06
Glu 150	8.47	124.72	176.67	56.34	4.38		30.31	2.05	1.92	36.29	
Thr 151	8.28	115.55	175.31	62.09	4.36		69.61	4.25		21.6	
Gly 152	8.51	111.66	174.18	45.41	4.02						
Ser 153	8.15	115.42	174.44	58.13	4.47		64.08	3.82			
Leu 166	7.95	120.99	177.89	55.81	4.21		42.09	1.6		26.92	
Arg 167	7.92	119.92	175.85	56.35	4.24		30.75	1.79		27.41	
Ala 168	7.88	123.59	177.21	52.12	4.3		19.35	1.34			
Leu 169	7.95	122.56		53.29	4.52		41.81	1.8	1.55		
Pro 170			176.38	63.06	4.36		32.02	2.2	1.77	27.41	
His 171	8.32	119.77		53.69	4.91		29.49	4.48	3.12		
Pro 172			176.82	63.24	4.39		32.13	2.26	1.91	27.43	
Ala 173	8.56	124.9	178.34	52.9	4.33		19.18	1.4			
Gly 174	8.34	108.84	173.73	45.34	3.95						
Gly 175	7.86	115.1		45.94	3.75						

Residue	Hγ1	Hγ1	Hγ2	Cδ	Cδ	Hδ	Hδ	4H	Cε	Hε
Met 1										
Asp 2										
Lys 3		1.41		29.12		1.67			42.17	2.99
Glu 4		2.23								
Asp 5										
Val 6		0.9								
Ala 7										
Gln 8		2.32								
Ser 9										
Ile 10	1.2	0.82	0.55	13.37		-0.06				
Val 11		0.9	0.83							
Pro 12		0.62	1.56	50.02		3.41	3.12			
Gln 13		2.49								
Asp 14										
Met 15										
Thr 16			1.12							
Pro 17	1.78	1.86		50.24		3.51				
Glu 18		2.17								
Thr 19										
Leu 20										
Gly 21										
His 22										
Tyr 23										
Cys 24										
Gln 25										
Met 26										
Asn 27										
Leu 28										
Leu 29				24.38	24.85	0.74				

APPENDIX A CONTINUED

Residue	Hγ1	Hγ1	Hγ2	Cδ	Cδ	Hδ	Hδ	4H	Cε	Hε
Glu 30		2.13								
His 31										
Pro 32		2.23	2.06	50.47		3.95	3.81			
Gly 33										
Pro 34		2.49	2.02			3.36				
Lys 35		1.55	2.34							3.11
Ala 36										
Gln 37		1.6	1.3							
Ile 38	0.81	1.05	0.61	12.73		0.63				
Phe 39										
Leu 40		1.56		24.97	25.53	0.97	0.87			
Glu 41		2.2	2.1							
Gly 42										
Ser 43										
Pro 44	2.21	2.05		51.13		3.97	3.81			
Ala 45										
Pro 46		1.14		49.83		3.5	3.35			
Leu 47		0.7		24.44		0.31				
Phe 48										
Phe 49										
Ser 50										
Gln 51		2.8	2.5							
Val 52		0.99	1.05							
Arg 53		1.84	1.61	43.66		3.2	3.13			
Asp 54										
Ala 55										
Ile 56	1.02	1.44	0.57	12.46		0.23				
Ala 57										
Tyr 58										

APPENDIX A CONTINUED

Residue	Hγ1	Hγ1	Hγ2	Cδ	Cδ	Hδ	Hδ	4H	Cε	Hε
Ala 59										
Arg 60		1.58		43		3.08	2.91			
Gly 61										
Pro 62		1.97		49.94		3.63	3.7			
Glu 63		2.12								
Gln 64		2.16								
Ile 65	1.5	1.22	0.92	13.47		0.9				
Ala 66										
Pro 67		2.08	1.89	50.88		3.8	3.53			
Ile 68			0.74	14.31		1.01				
Leu 69		1.55		26.07	22.78	0.79				
Val 70		0.59	0.45							
Ile 71	1.05	1.35	0.62	13.21		0.83				
Tyr 72										
Val 73		1.06	0.62							
Asn 74										
Asp 75										
Met 76		2.26	2.15						19.87	2.1
Gly 77										
Ala 78										
Ala 79										
Gly 80										
Ala 81										
Thr 82		1.19								
Trp 83										
Asp 84										
Glu 85		2.12								
Pro 86	1.43	1.28		49.68		3.49	3.28			
Gly 87										

APPENDIX A CONTINUED

Residue	Hγ1	Hγ1	Hγ2	Cδ	Cδ	Hδ	Hδ	4H	Cε	Hε
Asp 88										
Gly 89										
Asn 90										
Trp 91										
Ile 92	1.33		1	13.84		0.54				
Ala 93										
Ala 94										
Asp 95										
Lys 96	1.38	1.45		28.59		1.68			42.27	3.02/3.08
Arg 105		1.12	1.08	41.57		2.27	1.82			
Arg 106		1.64	1.5	42.67		3.08	2.99			
Gly 107										
Gly 108										
Met 109		2.3	2.72							
Gly 110										
Ala 111										
Pro 112		2.18	1.85	50.84		3.9	3.5			
Glu 113		2.27								
Ala 114										
Val 115		0.9								
Pro 116		2.43	1.83	51.05		3.3	3.26			
Phe117										
Ser 118										
Ser 119										
Arg 120		1.18	0.85	42.89		2.81				
Asp 121										
Glu 122		2.53	2.35							
Ala 123										
Ala 124										

Residue	Hγ1	Hγ1	Hγ2	Cδ	Cδ	Hδ	Hδ	4H	Cε	Hε
Ala 125										
Phe126										
Val 127		0.92	0.7							
Leu 128		1.76		24.82	23.37	0.86	0.82			
Gly 132										
Gln 133		2.15	2.33							
Val 134		0.85	0.8							
Leu 135		1.53		27.07	23.75	0.94	0.83			
Ala 136										
Leu 137		1.41		25.92	22.38	0.67	0.5			
Ala 138										
Asp 139										
Ile 140			0.68	12.78		0.18				
Thr 141		1.26								
Asp 142										
Ala 143										
Met 144		2.77	2.55							
Val 145		0.7								
Leu 146		1.83		25.22	23.33	0.8				
Thr 147		1.22								
Pro 148		2.01	1.95	51.08		3.67	3.84			
Val 149		0.92								
Glu 150		2.23								
Thr 151		1.19								
Gly 152										
Ser 153										
Glu 154		2.28								
Pro 155		1.98		50.59		3.78	3.65			
Arg 156		1.64	1.8	43.34		3.17				

APPENDIX A CONTINUED

Residue	Hγ1	Hγ1	Hγ2	Cδ	Cδ	Hδ	Hδ	4H	Cε	Hε
Ala 157										
Asp 158										
Asp 159										
Glu 160		2.23								
Gly 164										
Arg 165		1.6		43.49		3.15				
Leu 166				24.93	23.34	0.82				
Arg 167		1.61		43.42		3.14				
Ala 168										
Leu 169										
Pro 170		1.97	1.97	50.46		3.79	3.58			
His 171								7.31		
Pro 172		1.73	1.94	50.57		3.45	3.71			

APPENDIX B
Amide proton and ^{15}N data for apo vs holo NosL

APPENDIX B

Amide proton and ^{15}N data for apo vs holo NosL. ^{15}N shifts are referenced to liquid ammonia.

Residue	H_N apo	^{15}N apo	H_N holo	^{15}N holo	H_N chemical shift difference (ppm)	^{15}N chemical shift difference (ppm)
Met 1						
Asp 2						
Lys 3	8.38	121.5	8.42	121.7	0.04	0.2
Glu 4	8.48	121.42	8.49	121.49	0.01	0.07
Asp 5	8.24	121.48	8.24	121.45	0	-0.03
Val 6	7.95	120.48	7.95	120.68	0	0.2
Ala 7	8.24	126.15	8.24	126.26	0	0.11
Gln 8	8.05	118.26	8.08	118.47	0.03	0.21
Ser 9	8.12	116.43	8.14	116.66	0.02	0.23
Ile 10	7.88	122.57	7.89	122.81	0.01	0.24
Val 11	7.47	126.28	7.44	126.18	-0.03	-0.1
Pro 12						
Gln 13	9.22	122.88	9.18	122.9	-0.04	0.02
Asp 14	8.32	117.82	8.26	117.99	-0.06	0.17
Met 15	8.41	118.64	8.36	118.31	-0.05	-0.33
Thr 16						
Pro 17						
Glu 18	8.57	122.42	8.58	122.52	0.01	0.1
Thr 19	7.61	117.36	7.79	116.85	0.18	-0.51
Leu 20						
Gly 21	8.22	106.63	8.61	105.79	0.39	-0.84
His 22	7.8	119.64	7.75	120	-0.05	0.36
Tyr 23						
Cys 24						
Gln 25						
Met 26						
Asn 27						
Leu 28						
Leu 29						
Glu 30	7.65	114.88	7.8	115.03	0.15	0.15
His 31	7.71	119.38				
Pro 32						
Gly 33	7.85	105.94	7.83	106.2	-0.02	0.26
Pro 34						
Lys 35	8.58	117.42	8.78	117.9	0.2	0.48
Ala 36	7.74	115.89	7.64	115.68	-0.1	-0.21
Gln 37	9.33	116.98	9.39	117.67	0.06	0.69
Ile 38	9.07	121.7	9.04	121.73	-0.03	0.03

APPENDIX B CONTINUED

Residue	H_N apo	^{15}N apo	H_N holo	^{15}N holo	H_N chemical shift difference (ppm)	^{15}N chemical shift difference (ppm)
Phe 39	8.36	124.12	8.34	123.95	-0.02	-0.17
Leu 40	9.54	124.64	9.53	124.58	-0.01	-0.06
Glu 41	8.71	123.56	8.72	123.85	0.01	0.29
Gly 42	8.98	114.82	8.99	114.95	0.01	0.13
Ser 43	7.46	115.26	7.45	115.4	-0.01	0.14
Pro 44						
Ala 45	7.47	120.92	7.48	120.89	0.01	-0.03
Pro 46					0	0
Leu 47	8.41	118.6	8.37	118.04	-0.04	-0.56
Phe 48	8.67	120.9	8.73	121.06	0.06	0.16
Phe 49	8.92	117.1	9.09	116.62	0.17	-0.48
Ser 50						
Gln 51	8.06	121.15	8.07	121.27	0.01	0.12
Val 52	10.46	130.78	10.81	131.37	0.35	0.59
Arg 53	10.54	119.1	10.56	119.35	0.02	0.25
Asp 54	6.81	116.62	6.73	116.63	-0.08	0.01
Ala 55	7.11	123.98	7.03	123.2	-0.08	-0.78
Ile 56	7.63	116.27	7.54	116.2	-0.09	-0.07
Ala 57	7.99	120.62	8.05	120.66	0.06	0.04
Tyr 58	8.24	120.49				
Ala 59	7.86	120.55	7.85	120.49	-0.01	-0.06
Arg 60	7.57	114.88	7.59	114.99	0.02	0.11
Gly 61	7.35	108.32	7.36	108.33	0.01	0.01
Pro 62						
Glu 63	8.09	119.72				
Gln 64	8.01	122.56	8	122.62	-0.01	0.06
Ile 65	8.22	121.8	8.23	121.83	0.01	0.03
Ala 66	7.61	122.34	7.61	122.57	0	0.23
Pro 67						
Ile 68	8.29	122.75	8.29	122.8	0	0.05
Leu 69	9.27	128.4	9.25	128.36	-0.02	-0.04
Val 70	7.21	114.57	7.18	114.48	-0.03	-0.09
Ile 71	7.66	124.59	7.63	124.56	-0.03	-0.03
Tyr 72	9.37	124.25	9.33	124.16	-0.04	-0.09
Val 73	9.68	111.74	9.7	111.88	0.02	0.14
Asn 74	7.31	118.74	7.34	118.95	0.03	0.21
Asp 75	6.61	112.49	6.6	112.56	-0.01	0.07
Met 76	8.43	124.06	8.42	124.11	-0.01	0.05
Gly 77	8.35	107.73	8.35	107.82	0	0.09
Ala 78	6.94	124.45	6.93	124.5	-0.01	0.05

APPENDIX B CONTINUED

Residue	H_N apo	^{15}N apo	H_N holo	^{15}N holo	H_N chemical shift difference (ppm)	^{15}N chemical shift difference (ppm)
Ala 79	8.59	125.35	8.59	125.44	0	0.09
Gly 80	8.72	111.02	8.72	110.97	0	-0.05
Ala 81	7.71	123.07	7.68	123.04	-0.03	-0.03
Thr 82	7.61	113.96	7.6	113.99	-0.01	0.03
Trp 83	8.75	119.92	8.79	120.15	0.04	0.23
Asp 84	7.92	110.86	7.93	110.85	0.01	-0.01
Gln 85	7.53	117	7.51	116.95	-0.02	-0.05
Pro 86					0	0
Gly 87	8.02	102.95	8.01	103.2	-0.01	0.25
Asp 88	8.49	126.58	8.5	126.6	0.01	0.02
Gly 89	10.51	112.97	10.47	112.66	-0.04	-0.31
Asn 90	6.74	118.83	6.71	118.87	-0.03	0.04
Trp 91	7.27	121.57	7.28	121.58	0.01	0.01
Ile 92	9.72	118.1	9.71	118.09	-0.01	-0.01
Ala 93	9.05	127.12	9.08	127.33	0.03	0.21
Ala 94	8.84	123.63	8.84	123.77	0	0.14
Asp 95	8.85	113.14	8.85	113.22	0	0.08
Lys 96	7.46	118.32	7.54	118.35	0.08	0.03
Ala 97	6.65	121.19	6.64	121.26	-0.01	0.07
Phe 98	8.82	117.63	8.81	117.75	-0.01	0.12
Tyr 99	9.39	118.7	9.38	118.78	-0.01	0.08
Val 100	9.67	121.04	9.68	121.14	0.01	0.1
Val 101	8.96	128.83	8.97	128.93	0.01	0.1
Gly 102	9.93	113.62	9.96	113.66	0.03	0.04
Ser 103	8.79	117.85	8.79	118.13	0	0.28
Ala 104	8.66	119.66	8.73	119.78	0.07	0.12
Arg 105	7.76	117.65	7.77	117.6	0.01	-0.05
Arg 106	8.51	123.84	8.54	123.85	0.03	0.01
Gly 107	8.98	105.08	9.02	104.67	0.04	-0.41
Gly 108	9.71	108.5	10.21	109.11	0.5	0.61
Met 109	8.32	115.96	8.36	115.13	0.04	-0.83
Gly 110	8.45	108.1	8.58	108.74	0.13	0.64
Ala 111	7.88	123.84				
Pro 112						
Glu 113	8.77	125.3	8.79	125.8	0.02	0.5
Ala 114	10.08	132.32	9.93	132.2	-0.15	-0.12
Val 115	8.71	123.57	8.72	123.85	0.01	0.28
Pro 116						
Phe117	8.84	115.89	8.82	115.9	-0.02	0.01
Ser 118	8.28	116.04	8.26	116.02	-0.02	-0.02

APPENDIX B CONTINUED

Residue	H_N apo	^{15}N apo	H_N holo	^{15}N holo	H_N chemical shift difference (ppm)	^{15}N chemical shift difference (ppm)
Ser 119	8.17	115.82	8.17	115.95	0	0.13
Arg 120	8.84	129.59	8.83	129.75	-0.01	0.16
Asp 121	7.89	117.93	7.89	117.94	0	0.01
Glu 122	7.43	121.02	7.43	121.17	0	0.15
Ala 123	7.34	123.76	7.35	123.87	0.01	0.11
Ala 124	8.11	118.95	8.11	119	0	0.05
Ala 125	7.51	121.44	7.51	121.56	0	0.12
Phe126	7.62	122.79	7.63	122.85	0.01	0.06
Val 127	8.08	120.56	8.08	120.66	0	0.1
Leu 128	7.59	120.62	7.59	120.67	0	0.05
Ala 129	7.03	118.35	7.03	118.41	0	0.06
Glu 130	8.42	114.08	8.43	114.14	0.01	0.06
Gly 131	8.67	112.25	8.67	112.31	0	0.06
Gly 132	7.57	99.92	7.58	99.89	0.01	-0.03
Gln 133	8.59	115.56	8.61	115.65	0.02	0.09
Val 134	8.49	121.79	8.49	121.84	0	0.05
Leu 135	9.89	130.26	9.91	130.39	0.02	0.13
Ala 136	8.82	123.46	8.8	123.44	-0.02	-0.02
Leu 137	9.01	121.67	8.99	121.71	-0.02	0.04
Ala 138	8.15	115.07	8.14	115.1	-0.01	0.03
Asp 139	7.53	114.33	7.53	114.38	0	0.05
Ile 140	7.36	120.79	7.36	120.83	0	0.04
Thr 141	8.12	122.07	8.15	122.3	0.03	0.23
Asp 142	8.64	120.95	8.65	121.09	0.01	0.14
Ala 143	8.11	119.66	8.11	119.75	0	0.09
Met 144	7.55	114.92	7.55	114.99	0	0.07
Val 145	7.23	108.66	7.23	108.59	0	-0.07
Leu 146	7.95	120.48				
Thr 147	7.91	117.67	7.9	117.72	-0.01	0.05
Pro 148						
Val 149	8.17	120.32	8.17	120.38	0	0.06
Glu 150	8.47	124.72	8.47	124.77	0	0.05
Thr 151	8.28	115.55	8.29	115.57	0.01	0.02
Gly 152	8.51	111.66	8.5	111.7	-0.01	0.04
Ser 153	8.15	115.42	8.15	115.4	0	-0.02
Glu 154	8.37	123.72	8.38	123.81	0.01	0.09
Pro 155						
Arg 156	8.41	121.75	8.41	121.94	0	0.19
Ala 157	8.41	125.69	8.41	125.82	0	0.13
Asp 158	8.26	118.58	8.26	118.56	0	-0.02

APPENDIX B CONTINUED

Residue	H_N apo	^{15}N apo	H_N holo	^{15}N holo	H_N chemical shift difference (ppm)	^{15}N chemical shift difference (ppm)
Asp 159	8.07	120.05	8.07	120.01	0	-0.04
Glu 160	8.31	120.98	8.3	120.97	-0.01	-0.01
Asp 161	8.31	120.82	8.3	120.97	-0.01	0.15
Tyr 162	8.01	121.08	8.01	121.16	0	0.08
Leu 163	8.06	120.56				
Gly 164	8.02	107.92	8.02	107.92	0	0
Arg 165	7.84	120.24	7.84	120.19	0	-0.05
Leu 166	7.95	120.99	7.95	121.03	0	0.04
Arg 167	7.92	119.92	7.92	120	0	0.08
Ala 168	7.88	123.59	7.89	123.68	0.01	0.09
Leu 169	7.95	122.56	7.96	122.64	0.01	0.08
Pro 170						
His 171	8.32	119.77				
Pro 172						
Ala 173	8.56	124.9	8.58	124.94	0.02	0.04
Gly 174	8.34	108.84	8.34	108.81	0	-0.03
Gly 175	7.86	115.1	7.86	115.23	0	0.13

APPENDIX C
Chemical shift file generated by
NMRView

APPENDIX C

```
Chemical shift file generated by NMRView
/home/taubner/nosL_vc/13C_ppm.out

  2.CA      54.640 1
  2.HA       4.650 1
  2.CB      41.400 1
  2.HB2      2.600 2
  2.HB1      2.710 2
  2.C      175.900 1
  3.N      121.500 1
  3.HN       8.380 1
  3.CA      56.810 1
  3.HA       4.230 1
  3.CB      33.120 1
  3.HB1      1.790 2
  3.CG      24.510 1
  3.HG1      1.410 2
  3.CD      29.120 1
  3.HD1      1.670 2
  3.CE      42.170 1
  3.HE2      2.990 2
  3.C      176.590 1
  4.N      121.420 1
  4.HN       8.480 1
  4.CA      56.900 1
  4.HA       4.220 1
  4.CB      30.120 1
  4.HB1      1.970 2
  4.CG      36.390 1
  4.HG1      2.230 2
  4.C      176.330 1
  5.N      121.480 1
  5.HN       8.240 1
  5.CA      54.460 1
  5.HA       4.560 1
  5.CB      41.080 1
  5.HB1      2.620 2
  6.N      120.480 1
  6.HN       7.950 1
  6.CA      62.840 1
  6.HA       4.030 1
  6.CB      32.570 1
  6.HB       2.080 1
  6.CG2     23.770 1
  6.CG1     20.300 1
  6.HG11     0.900 3
  6.C      176.310 1
  7.N      126.150 1
  7.HN       8.240 1
```

APPENDIX D
Cross peak file generated by
NMRView

APPENDIX D

```
Cross-peak file generated by NMRView
/home/taubner/nosL_vc/13C_run10.xpk

label dataset sw sf
H1 H C13
13C_noesy_ext2.nv
{3783.40 } {3825.98 } {9054.75 }
{600.1318 } {600.1318 } {150.9100 }
 H1.L  H1.P  H1.W  H1.B  H1.E  H1.J  H1.U  H.L  H.P  H.W  H.B  H.E  H.J
H.U  C13.L  C13.P  C13.W  C13.B  C13.E  C13.J  C13.U  vol  int  stat
comment  flag0
1  {82.HB}   4.565   0.000   0.000   ++   0.000   {?}   {82.HG1}
1.174   0.000   0.000   ++   0.000   {?}   {82.CB}   72.260   0.000
0.000   ++   0.000   {?}  0.222E+01 0.000 0 {?} 0
1  {141.HB}   4.563   0.000   0.000   ++   0.000   {?}   {141.HA}
4.397   0.000   0.000   ++   0.000   {?}   {141.CB}   70.268   0.000
0.000   ++   0.000   {?}  0.222E+01 0.000 0 {?} 0
2  {141.HB 141.HB 141.HB}   4.564   0.000   0.000   ++   0.000   {?}
{141.HG22 141.HG23 141.HG21}   1.270   0.000   0.000   ++   0.000   {?}
{141.CB 141.CB 141.CB}   70.259   0.000   0.000   ++   0.000   {?}
0.222E+01 0.000 0 {?} 0
3  {141.HB 141.HB 141.HB}   4.554   0.000   0.000   ++   0.000   {?}
{140.HA 142.HA 143.HA}   4.090   0.000   0.000   ++   0.000   {?}
{141.CB 141.CB 141.CB}   70.047   0.000   0.000   ++   0.000   {?}
0.222E+01 0.000 0 {?} 0
4  {151.HB}   4.233   0.000   0.000   ++   0.000   {?}   {158.HA}
4.533   0.000   0.000   ++   0.000   {?}   {151.CB}   69.828   0.000
0.000   ++   0.000   {?}  0.222E+01 0.000 0 {?} 0
5  {151.HB}   4.228   0.000   0.000   ++   0.000   {?}   {150.HA}
4.401   0.000   0.000   ++   0.000   {?}   {151.CB}   69.742   0.000
0.000   ++   0.000   {?}  0.222E+01 0.000 0 {?} 0
6  {151.HB}   4.227   0.000   0.000   ++   0.000   {?}   {152.HA1}
4.016   0.000   0.000   ++   0.000   {?}   {151.CB}   69.726   0.000
0.000   ++   0.000   {?}  0.222E+01 0.000 0 {?} 0
7  {151.HB}   4.227   0.000   0.000   ++   0.000   {?}   {151.HG2%}
1.183   0.000   0.000   ++   0.000   {?}   {151.CB}   69.767   0.000
0.000   ++   0.000   {?}  0.222E+01 0.000 0 {?} 0
8  {147.HB}   4.069   0.000   0.000   ++   0.000   {?}   {147.HA}
4.462   0.000   0.000   ++   0.000   {?}   {147.CB}   69.668   0.000
0.000   ++   0.000   {?}  0.222E+01 0.000 0 {?} 0
9  {147.HB}   4.069   0.000   0.000   ++   0.000   {?}   {148.HD2}
3.818   0.000   0.000   ++   0.000   {?}   {147.CB}   69.604   0.000
0.000   ++   0.000   {?}  0.222E+01 0.000 0 {?} 0
10  {147.HB}   4.073   0.000   0.000   ++   0.000   {?}   {148.HD1}
3.655   0.000   0.000   ++   0.000   {?}   {147.CB}   69.611   0.000
0.000   ++   0.000   {?}  0.222E+01 0.000 0 {?} 011  {147.HB 147.HB
147.HB}   4.070   0.000   0.000   ++   0.000   {?}   {147.HG22 147.HG21
147.HG23}   1.220   0.000   0.000   ++   0.000   {?}   91   0.000
0.000   ++   0.000   {?}   {40.HG 128.HB2}   1.599   0.000   0.000
```

APPENDIX E
Conversion file for converting
NMRView data into xml format

APPENDIX E

```
Conversion file for converting NMRView data into xml format
/home/taubner/nosL_vc/conversion.xml
Commands to generate template and then convert:
aria2 --convert -t conversion.xml
aria2 --convert.xml

<!DOCTYPE conversion SYSTEM "conversion1.0.dtd">

<!--====================== Conversion ================================

  Template xml file for data conversion. Fill in the necessary fields
  between the quotation marks "". Leave optional fields that you do
  not want unchanged.

-->

<conversion>

<!--==================== Project definition ==========================

  If you want ARIA to automatically create a project XML file, fill
  in the field "filename". Your data files specified below (the
  respective XML files) will be referenced in the project-xml. The
  project XML file can be edited using ARIAs GUI, a XML editor or a
  text editor.

  Alternatively, you can also leave the field empty and create a
  project-xml file later invoking ARIA from the command line. You would
  then have to fill in the filenames of your converted data by hand.

  Note: a project XML file can only be created if all three data-types
        (molecule definition, chemical-shifts and spectra) could be
        successfully exported into XML.


  Optional fields:

  - filename

-->

  <project>
     <output
       filename="nosl_vc.xml"/>
  </project>

<!--=================== Molecule definition ==========================

  Allowed formats for the input data are "seq" (a sequence of three
  letter codes for the residues) or "pdb". If the input data are in
```

APPENDIX E CONTINUED

```
  PDB format a naming convention for the residues and atoms has to
  be specified; possible choices are: "iupac", "dyana", "cns". Further-
  more, the "molecule_segid" entry will be ignored and the segids
  will be read from the PDB-file. The type of the molecule can be
either
  "PROTEIN", "DNA" or "RNA". The additional tag "first_residue_number"
  specifies where the residue numbering starts (in case of SEQ format).

  Optional fields:

  - molecule_type        (optional in case of PDB input files)
  - molecule_name
  - molecule_segid       (will not be considered in case of PDB input
files)
  - first_residue_number (will not be considered in case of PDB input
files)

-->

  <molecule
    molecule_type="PROTEIN"
    molecule_name="nosl"
    molecule_segid=""
    first_residue_number="1">
    <input
      filename="full_nosL.seq"
      format="seq"
      naming_convention="cns"/>
    <output
      filename="seq.xml"/>
  </molecule>

<!--=================== Spectrum definition ==========================

  Each spectrum is defined by a list of chemical shift assignments
  and a list of NOE cross-peaks. The files can be either in
  "xeasy", "ansig" or "nmrview" format. For conversion of the cross-
  peaks the dimensions that correspond to the resonances have to be
  specified. The entry for the 1st and the 2nd proton should always
  be a number (1, 2, 3, or 4). If the NOESY experiment has been
  three- or four-dimensional the corresponding hetero dimensions have
  to be set as well.
  If you want to specify the segment(s) whose atoms have assigned
  chemical shifts or cross-peaks, use the "segid"-field in the spectrum
  header; in case of multiple segments the corresponding segids are
  separated by a slash (e.g., segids="A" or segids="A/B").

  Optional fields:

  - spectrum_name
```

APPENDIX E CONTINUED

```
  - segids

-->

<!-- ==== Copy the block <spectrum> ... </spectrum> if you want ====
          to convert multiple spectra. I.e. for n spectra you
          need n blocks. -->

  <spectrum
    spectrum_name="13c_noesy"
    segids="">
    <chemical_shifts>
      <input
        filename="13C_ppm.out"
        format="nmrview"/>
      <output
        filename="13C_shifts.xml"/>
    </chemical_shifts>
    <cross_peaks>
     <input
        filename="13C_run10.xpk"
        format="nmrview"
        proton1="1"
        hetero1="3"
        proton2="2"
        hetero2=""/>
      <output
        filename="13C_noe.xml"/>
    </cross_peaks>
  </spectrum>
  <spectrum
<!-- ============================================================== -->

</conversion>
```

APPENDIX F
ARIA project file, setup and run

APPENDIX F

```
Project file generation, setup and run
/home/taubner/nosL_vc/nosL_vc.xml
Command to set up template .xml file:
aria2 --project_template nosL_vc.xml
Command to set up directory tree:
aria2 -s nosL_vc.xml
Command to run aria:
aria2 nosL_vc.xml

<!DOCTYPE project SYSTEM "project1.0.dtd">
<project name="nosL_vc" version="1.0" author="" date="" description=""
comment="" references="" working_directory="/r2d2/taubner/nosL_vc"
temp_root="/r2d2/taubner/nosL_vc" run="1" file_root="nosL_vc"
cache="yes" cleanup="yes">
  <data>
    <molecule file="/r2d2/taubner/nosL_vc/seq.xml" format="xml">
      <linkage_definition name="automatic" filename=""/>
      <parameter_definition name="automatic" filename=""/>
      <topology_definition name="automatic" filename=""/>
    </molecule>
    <spectrum enabled="yes" use_assignments="yes"
trust_assigned_peaks="no">
      <shifts file="/r2d2/taubner/nosL_vc/13C_shifts.xml" format="xml"
default_shift_error="0.05"/>
      <peaks file="/r2d2/taubner/nosL_vc/13C_noe.xml" format="xml"
peak_size="volume" freq_window_proton1="0.05" freq_window_hetero1="0.5"
freq_window_proton2="0.05" freq_window_hetero2="0.5">
        <lower_bound_correction value="0.0" enabled="no"/>
        <upper_bound_correction value="6.0" enabled="no"/>
      </peaks>
    </spectrum>
    <spectrum enabled="yes" use_assignments="yes"
trust_assigned_peaks="no">
      <shifts file="/r2d2/taubner/nosL_vc/15N_shifts.xml" format="xml"
default_shift_error="0.05"/>
      <peaks file="/r2d2/taubner/nosL_vc/15N_noe.xml" format="xml"
peak_size="volume" freq_window_proton1="0.05" freq_window_hetero1="0.5"
freq_window_proton2="0.05" freq_window_hetero2="0.5">
        <lower_bound_correction value="0.0" enabled="no"/>
        <upper_bound_correction value="6.0" enabled="no"/>
      </peaks>
    </spectrum>
    <hbonds file="/r2d2/taubner/nosL_vc/H_bond.tbl" format="tbl"
enabled="yes" data_type="standard"/>
    <dihedrals file="/r2d2/taubner/nosL_vc/dihed.tbl" format="tbl"
enabled="yes" data_type="standard"/>
    <initial_structure file="" format="iupac" enabled="no"/>
  </data>
  <structure_generation engine="cns">
```

APPENDIX F CONTINUED

```
    <cns local_executable="/usr/local/cns_solve_1.1/intel-i686-
linux_g77/bin/cns" keep_output="yes" keep_restraint_files="yes"
create_psf_file="yes" generate_template="yes"
nonbonded_parameters="PROLSQ">
      <annealing_parameters>
        <unambiguous_restraints first_iteration="0" k_hot="10.0"
k_cool1_initial="10.0" k_cool1_final="50.0" k_cool2="50.0"/>
        <ambiguous_restraints first_iteration="0" k_hot="10.0"
k_cool1_initial="10.0" k_cool1_final="50.0" k_cool2="50.0"/>
        <hbond_restraints first_iteration="0" k_hot="10.0"
k_cool1_initial="10.0" k_cool1_final="50.0" k_cool2="50.0"/>
        <dihedral_restraints k_hot="5.0" k_cool1="25.0"
k_cool2="200.0"/>
```