STRUCTURAL CHARACTERIZATION OF THE RECEPTOR FOR ADVANCED GLYCATION END PRODUCTS REVEALS A TWO DOMAIN MODULAR ARCHITECTURE

By

Brian Matthew Dattilo

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

August, 2007

Nashville, Tennessee

Approved:

Professor Walter J. Chazin

Professor Charles R. Sanders, II

Professor Martin Egli

Professor Billy G. Hudson

ACKNOWLEDGEMENTS

First and foremost I thank my advisor, Dr. Walter Chazin. Not only did he provide the opportunity to freely perform exciting research in his laboratory, but he stuck with me during the low points of my project and development as a scientist.

I also thank the other members of my continuing committee, Drs. Billy Hudson, Charles (Chuck) Sanders, Martin Egli, and Matthew Breyer, for advice and discussion. But mostly I thank them for one particular meeting that helped me make the leap to the problem solver I have become today.

I thank the other three founding members of the CLDT. They know who they are and what the acronym means.

Writing this dissertation would not have been as enjoyable without the presence of two great friends, Kelly and Mike. Thank you for being great distractions from the many hours of sitting, reading, and writing.

All of the work in this dissertation has been performed in close collaboration with Dr. Guenter Fritz from the University of Konstanz. Thanks for provided stimulating discussion throughout the development of this project in addition to his contributions as an expert protein crystallographer. In addition, I thank Dr. Estelle Leclerc for her valuable Biacore expertise.

Of all the past and present members of the Chazin I especially thank Dr. Craig Vanderkooi and Susan Meyn. Craig helped develop the initial ideas for this dissertation, taught me everything I know about NMR, and provided incredible discussion throughout.

I thank Susan, of course, for putting up with my shenanigans in the lab but also for helping me develop as a person and fight through the tough times during grad school.

Funding was provided by the molecular biophysics training grant and NIH research grants. The Diabetes Research Training Center provided initial pilot funding which helped launch this work.

Finally I thank my family. Thanks for the example that provided me with all the tools needed to achieve success. And thanks for the love and support that helped me maintain enough focus to climb out of the low points of my training.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Aβ	Amyloid β Peptide
AD	Alzheimer's Disease
AGE	Advanced Glycation End products
APP	Amyloid Precursor Protein
BBB	Blood Brain Barrier
BSA	Bovine Serum Albumin
CD	Circular Dichroism
CML	N-(epsilon)-(CarboxyMethyl)Lysine
DLS	Dynamic Light Scattering
DSC	Differential Scanning Calorimetry
dnRAGE	dominant negative RAGE
DTT	DiThioThreitol
EBP	Erythryopoietin Binding Protein
EMP	Erythryopoietin Mimetic Peptide
EPO	ErythroPOietin
EPOR	ErythroPOietin Receptor
ERK	Extracellular signal-Regulated Kinase
FRET	Fluorescence Resonance Energy Transfer
GH	Growth Hormone
GHR	Growth Hormone Receptor
GHbp	Growth Hormone binding protein

GST	Glutathione S-Transferase
HMG	High Mobility Group
HSA	Human Serum Albumin
HSQC	Heteronuclear Single Quantum Coherence
IC	IntraCellular domain
Ig	Immunoglobulin
IPTG	IsoPropyl β-d-1-ThioGalactopyranoside
IL-4R	InterLeukin-4 Receptor
JNK	JaNus Kinase
MAPK	Mitogen Activated Protein Kinase
MALDI-MS	Matrix Assisted Laser Desorption Ionizing Mass Spectrometry
NCAM	Neural Cell Adhesion Molecule
NF-κB	Nuclear Factor kappa B
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect SpectroscopY
PDB	Protein DataBank
RAGE	Receptor for Advanced Glycation End products
RMSD	Root Mean Square Deviation
SDS-PAGE	Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electrophoresis
SPR	Surface Plasmon Resonance
sRAGE	soluble RAGE (extracellular ligand binding region)
TM	TransMembrane helix
TOCSY	TOtal COrrelated SpectroscopY

CHAPTER I

INTRODUCTION

RAGE: A Multiligand Receptor for Advanced Glycation End Products

RAGE is a cell surface receptor initially discovered and named for its ability to bind proteins modified by glucose metabolites termed advanced glycation end products (AGEs) (1, 2). AGE signaling through RAGE is widely thought to play a role in the progression of diabetic complications and chronic inflammatory processes (3). In the past 15 years, RAGE has been shown to interact with a range of structurally and functionally diverse ligands including AGEs, amphoterin, amyloid-β peptides, and members of the S100 protein superfamily (4). Interaction with these ligands implicates RAGE in several physiological processes (neural growth, acute inflammation) and disease pathologies (diabetes, tumor growth/metastasis, Alzheimer's disease, chronic inflammation) implicating RAGE as a promising lead for therapeutic intervention (4, 5).

RAGE is a member of the immunoglobulin (Ig) superfamily with three predicted Ig-like domains in its N-terminal extracellular region, in addition to a single transmembrane helix and small intracellular domain (Figure 1.1). The extracellular Ig-like domains have been classified as an N-terminal V-type domain and two subsequent C-type domains (1). However, no direct structural data existed prior to the work in this dissertation to confirm or refute this assignment, specify RAGE quaternary structure, elucidate the basis for interactions with structurally diverse ligands, or generate hypotheses about receptor autoinhibition and activation. The issue of sRAGE quaternary

is important for understanding how an initial binding event causes a signal to be transferred across the cell membrane. The basis for interactions with structurally diverse ligands is critical when addressing the issue of pharmaceutical intervention. And lastly, addressing both of these issues can provide insight into mechanisms of receptor autoinhibition and activation. Thus, the desire to obtain a broader knowledge of how RAGE functions in the cell and malfunctions in disease set the stage for the structural, biochemical, and biophysical work presented in this dissertation.

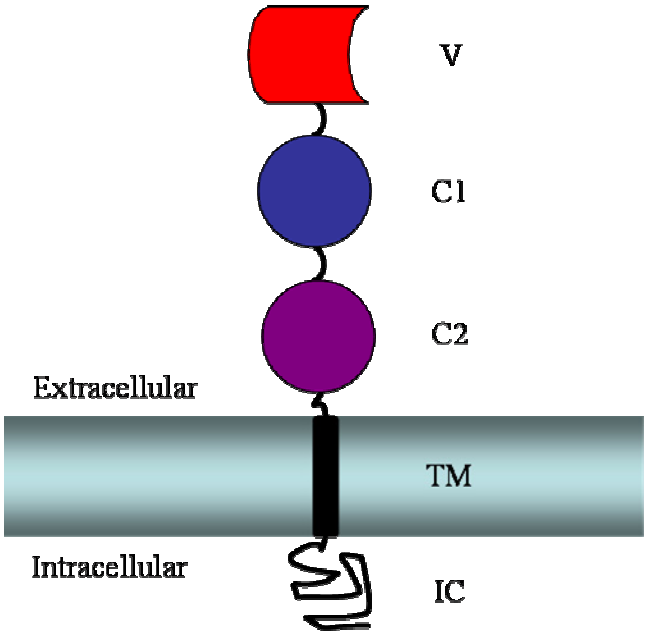


Figure 1.1: Schematic of RAGE domain organization. The membrane is shown as a horizontal rectangle with the extracellular and intracellular regions marked appropriately. The intracellular domain (IC) is shown as a random black sketch. The single transmembrane helix (TM) is a black vertical rectangle. The three predicted Ig-like domains of sRAGE are shown above the membrane. The predicted C-type Ig domains are depicted in dark blue (C1) and violet (C2). The predicted V-type Ig domain is shown in red.

RAGE Biology

In this section, I describe the current knowledge of RAGE function associated with normal physiology and cellular dysfunction. There is a wealth of information on the biology associated with RAGE as a signaling receptor. One fascinating aspect of RAGE biology is its recognition of structurally diverse ligands, which can be grouped into four classes: (1) AGEs, (2) amphoterin, (3) amyloid-β peptides (Aβ), and (4) S100 proteins. While multiligand recognition is not directly addressed, the conclusions and future directions in this dissertation are discussed with respect to signaling mechanisms specifically implicated by multiligand recognition. For scientific completeness, I also delve into the diverse biology associated with these ligands and some of their non-RAGE functions.

Advanced Glycation End Products (AGEs)

AGE Chemistry

Protein glycation is a non-enzymatic process initially discovered by Louis Maillard who observed chemical modifications to amino acids when heated in the presence of reducing sugars, a process often referred to as the Maillard reaction. During the early stages of glycation, sugars such as glucose form a Schiff base with free amino groups on proteins at the N-terminus or on lysine and arginine side chains (Figure 1.2) (6). The Schiff base undergoes a rearrangement involving proton transfer to the more stable and reversible Amadori Product (Figure 1.2). The late stages of glycation are irreversible and consist of a variety of chemical modifications. The best studied AGE moiety is N(epsilon)-(carboxymethyl)lysine (CML) which results in a charge reversal

(Figure 1.2) that could disrupt intra- or intermolecular salt bridges. Overall, the consequences of AGE formation can be as diverse as their chemistry.

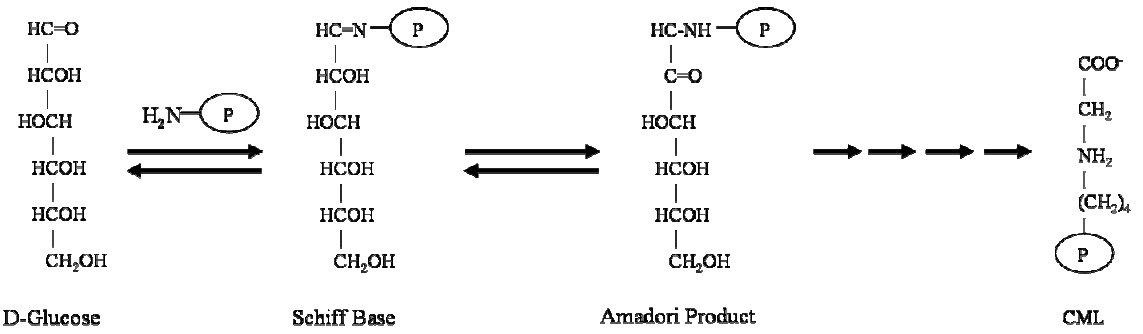


Figure 1.2: Chemical pathway towards AGE formation. In the first reversible step, D-glucose reacts with free amino groups on a protein forming a Schiff base (circle with letter "P" represents the protein). Reversible proton transfer leads to Amadori product formation in the second step. Finally, a series of irreversible non-enzymatic chemical reactions leads to AGE-modified protein (example shows CML modification to lysine side chain). Figure adapted from (6).

AGEs in Normal Aging and Diabetes

The accumulation of advanced glycation end products is dependent on two primary factors, the concentration of aldose sugars such as glucose and the half-life of the modified protein. Prolonged hyperglycemia associated with diabetes increases the concentration of glucose, a precursor to many AGEs. Observations in the late 1980s suggested AGEs played a role in the pathogenesis of cardiovascular disorders (7). These authors took advantage of the yellow-brown fluorescent pigmentation with characteristic UV absorption of AGEs. They compared the levels of an Amadori product to those of the highly fluorescent advanced glycation end product in rat aorta. Amadori product levels reached a maximum between 50 and 70 weeks whereas levels of the fluorescent advanced product reached maximum levels between 90 and 120 weeks old (7). This

observation implied a correlation between AGE formation and age. A similar study on humans showed increased Amadori levels between Type 1 diabetic patients versus control patients, but no correlation between duration or severity of diabetes (8). However, there was a direct correlation between the duration and severity of diabetic complications and the levels of fluorescent protein adducts, assumed to be AGEs as a result of prolonged hyperglycemia (9). Thus, it appeared that the irreversible AGE moieties and not the reversible Amadori products were associated with diabetes.

AGE formation has been directly observed on cellular components associated with complications of diabetes. Human cataract eye lens show characteristics of AGE formation compared to lens from non-cataract patients (10). Similar observations were made on insoluble fractions of collagen comparing both old to young patients and normal patients to those age-matched with Type 1 diabetes (11, 12). Additional work correlates AGE-β_2 microglobulin in patients on long-term hemodialysis suggesting AGE-related pathogenesis in patients being treated for renal failure (13). Lastly, *in vitro* formation of amyloid-β peptide (Aβ) aggregates, a major component of senile plaques, occurs rather slowly but is markedly enhanced when seeded with AGE-modified Aβ suggesting AGE modification may be a catalyst for aggregate formation in Alzheimer's brain (14). In a similar example, antibodies positively stain two AGEs, pentosidine and pyrraline, on senile plaques and neurofibrillary tangles (15). Thus, the high insolubility and resistance to denaturant of neurofibrillary tangles could be a byproduct of AGE induced covalent aggregation (16).

Initial work to discover potential receptors for AGEs characterized a 90 kDa protein from murine macrophages that bound a synthetic AGE compound with

nanomolar affinity (17, 18). Concurrent with work on macrophages were experiments on endothelium, tissue that lines the interior of blood vessels and whose dysfunction is connected with diabetic vascular complications. AGE-bovine serum albumin (AGE-BSA) specifically binds to cultured endothelial cells (19). Interestingly, monolayers of endothelial cells show increased vascular permeability, in addition to modulated coagulant properties, when in contact with AGE-BSA (19). Thus, the aggregate evidence suggests AGEs accumulate in long-lived proteins such as collagen and their localization leads to interaction with cell surface molecules causing cellular dysfunction. Studies of these diverse aspects of AGE biology lead to the search for and eventual discovery, isolation, cloning, and characterization of RAGE.

AGE Signaling Induces Oxidative Stress Though RAGE

The progression of diabetic hyperglycemia to cardiovascular disease, renal failure, etc. may arise from tissue damage or altered signaling pathways caused by increased oxidative stress in cells. Cultured capillary endothelial cells (a cell type present in the vasculature that is damaged in diabetes) show increased oxidative stress in response to treatment with AGE-BSA, which was blocked by antibodies to RAGE (20). In addition, electrophoretic mobility shift assays show AGE-BSA binding to RAGE induces specific translocation of nuclear factor kappa-B (NF-κB) to the nucleus (20). NF-κB is a well characterized transcription factor best studied for its role in gene expression in response to inflammatory stimuli in the immune system (21). In fact, oxidative stress alone in the form of hydrogen peroxide and superoxide radical induces NF-κB activation on T cells (22). Therefore one possible interpretation of the data is that AGE-BSA interacts with RAGE leading to increased oxidative stress which in turn induces

activation of NF-κB by a secondary mechanism. Another possibility is that AGE-BSA binding to RAGE directly leads to activation of signaling pathways causing oxidative stress on the target cell. Subsequent work has confirmed that activation of signaling pathways, and not binding alone, is necessary thereby establishing RAGE as a signaling receptor (23).

Signaling Pathways Triggered by AGEs

Many of the kinases involved in signaling between ligand binding and NF-κB activation have also been identified. In pulmonary artery smooth muscle cells treated with AGE-BSA, activation of NF-κB occurs through $p21^{ras}$ and the extracellular signal regulated kinase 1/2 (ERK1/2) (24). In human monocyte cells, both ERK and p38 mitogen activated protein kinase (MAPK) were activated by CML modified human serum albumin (HSA) (25). This study also showed no activation of the third kinase family, janus kinase. Thus, RAGE signaling in response to AGE exposure is a link to cell stress associated with diabetic vascular complications and stimulated interest in understanding RAGE biology.

RAGE Activation by Amphoterin: a Physiological Role for RAGE

Analysis of RAGE tissue distribution shows higher protein expression in developing tissue than in adult tissue (26). This observation suggests endogenous ligands may be present to activate RAGE during development. Amphoterin was identified as the first "natural" ligand for RAGE (AGE formation being an "unnatural" ligand due to the non-enzymatic chemistry). *In situ* hybridization and immunohistochemistry showed high expression and colocalization of RAGE and amphoterin in developing cerebral cortex,

hippocampus, and cerebellum of rat brains (27). *In vitro*, amphoterin binds to RAGE with the highest affinity of all ligands (K_D = 6.4 nM) and induces neurite outgrowth on cultured neurons (27). Receptor-dependent neurite outgrowth, therefore, was the first established native function for RAGE in the cell.

Amphoterin is one of several diffusible factors that induce neurite outgrowth on cultured neurons, a function vitally important during nervous system development (28, 29). Prior to its discovery as a neurite outgrowth factor described above involving RAGE, amphoterin (high mobility group 1, HMG1) had been characterized for its DNA binding activity and association with chromatin as a non-histone component (30, 31). In addition to DNA binding activity, HMG1 interacts with DNA-binding proteins and may modulate DNA processing events such as transcription (32, 33).

Amphoterin is described as containing two ~10 kDa HMG box motifs identified by sequence homology and shown to be structurally independent by limited proteolysis (34, 35). In addition, a highly acidic C-terminal extension exists adjacent to the two structured domains (34). Solution structures are available for each of the two HMG box motifs of rat amphoterin (Figure 1.3). The two domains are highly homologous with backbone RMSD of 2.0 Å (36, 37). The fold is predominately α-helical with helices 1 and 2 forming an approximate right angle with respect with helix 3 (Figure 1.3). This L-shaped α-helical structure is quite different from the small chemical modifications that make up AGEs. In addition, the name amphoterin is derived from the primary sequence showing two regions with completely different electrostatic properties; both HMG box domains are quite basic while the C-terminal extension is quite acidic. In fact, it appears that the basic region (pI ~10.0), part of which is α-helical in the HMG box B structure

(Figure 1.3), is capable of binding to RAGE (38). Thus, it is possible that molecules with different electrostatic properties are recognized by RAGE.

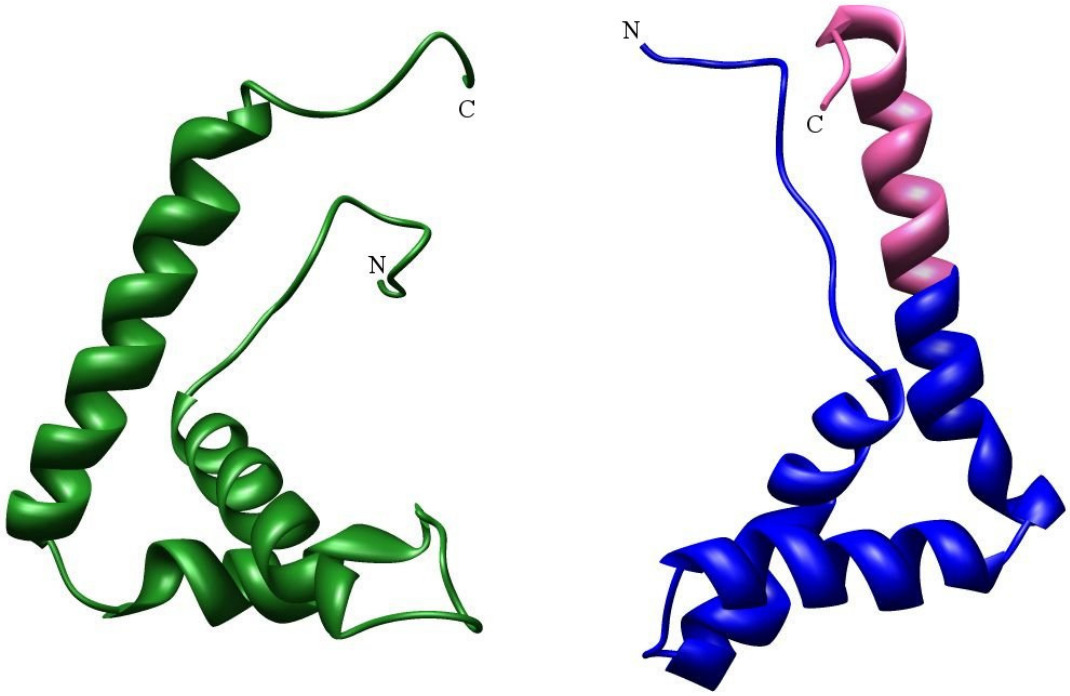


Figure 1.3: Structure of rat amphoterin. Ribbon diagrams representing one model of solution structures of the two HMG box domains of rat amphoterin (rat and human are >99% identical). The N-terminal "A" domain corresponding to residues 1-83 of native amphoterin (PDB 1AAB) is shown in green. The C-terminal "B" domain corresponding to residues 88-164 of native amphoterin (PDB 1HME) is shown in blue. N- and C-termini of each domain are marked accordingly. A portion of the amphoterin peptide known to bind and antagonize RAGE is colored in pink. The remainder of this peptide was not included in the construct used in structure determination. All protein structure images in this dissertation were made using Chimera (39).

RAGE Binding to Amyloid β: Implications in Neurotoxicity

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by lesions in regions of the brain important for intelligence and memory. Two primary morphologies are observed, neurofibrillary tangles and senile plaques. The major protein component of neurofibrillary tangles are tau protein aggregates while that for senile plaques are amyloid fibrils composed largely of amyloid β (Aβ) peptides from the amyloid precursor protein (APP). RAGE has been identified as a protein capable of binding Aβ peptides and does so with low nanomolar affinity (40).

APP is an integral membrane protein with a large extracellular domain, a single transmembrane helix, and a small intracellular domain. Normal cellular processing of APP in lysozomal compartments leads to proteolytic fragments that are secreted from various cells in culture (41). However, neither the function of APP nor the role of normal protein turnover is well understood. In senile plaques, Aβ peptides are heterogeneous in length but average around 40 amino acids. Nonetheless, isolated Aβ peptides offer functional and structural features unique from AGEs and amphoterin.

The interaction of soluble Aβ peptide with RAGE induces migration of microglial cells and oxidative stress in endothelial cells, both of which are RAGE dependent (40). Thus, RAGE may enhance neurotoxicity directly through Aβ peptide interaction or indirectly by inducing microglial infiltration. The role of RAGE as a signaling receptor in neurodegenerative disorders was strengthened by studying a transgenic mouse model for AD expressing a mutant form of APP (42). Transgenic mice also expressing a signaling deficient form of RAGE (dnRAGE, see below) show markers such as enhanced NF-κB activation and diminished learning and memory compared to mice only expressing wild-

type RAGE, consistent with its role as a signaling receptor in Aβ induced neuropathy (42).

S100 Proteins are RAGE Ligands

The continued search for native ligands of RAGE lead to the identification of a gene termed EN-RAGE (endogenous RAGE) (43). Analysis of the sequence revealed EN-RAGE was a member of the S100 superfamily of EF-hand proteins, S100A12. The precise function of S100 proteins is not well understood. Nonetheless, they are involved in diverse biological processes.

S100A12 (EN-RAGE) is one of the calgranulins, which are exclusively expressed and secreted from granulocytes of the immune system. Radioligand binding assays show S100A12 binds with K_D ~90 nM to both purified RAGE and RAGE on the endothelial cell surface (43). RAGE activation by S100A12 on endothelial cells induces nuclear translocation of NF-κB and the subsequent expression of vascular cell adhesion molecule-1 and intercellular adhesion molecule-1, both mediators of the inflammatory response. Furthermore, S100A12 clearly promotes cell migration of mononuclear phagocytes in a chemotaxis assay strongly suggesting a RAGE-S100 protein role in acute inflammation (43).

A second S100 protein, S100B, was shown to activate RAGE (44). S100B is known to regulate the extension of neurites *in vitro* (45, 46). At nanomolar concentrations of S100B, neurite outgrowth is promoted. This function appears to be coregulated in an additive manner by S100B, S100A1 and amphoterin up to 1 μM. At micromolar concentrations, however, both S100 proteins lead to cellular toxicity and apoptosis

through the caspase cascade (44). For all three proteins at neurotrophic concentrations, RAGE activation led to NF-κB upregulation (44).

S100A12 and S100B are the best studied RAGE ligands in the S100 superfamily. However, other S100 proteins are implicated in RAGE biology. S100P is overexpressed in many cancers. In both NIH3T3 and pancreatic cancer cells, S100P is coimmunoprecipitated with RAGE and activation of NF-κB is blocked with a signaling deficient RAGE (dnRAGE, see below) (47, 48). S100A4 leads to RAGE dependent activation of p38 MAPK, nuclear translocation of NF-κB, and overproduction of matrix metalloprotease 13 in cultured human chondrocytes (49). And lastly, S100A11, overexpressed in osteoarthritic cartilage, appears to activate chondrocytes by signaling through RAGE and p38 MAPK (50). S100 protein activation of RAGE is likely regulated by the subcellular localization and tissue-specific nature of S100 proteins since many in the superfamily bind to and activate RAGE.

S100s are small, acidic, α-helical proteins with high structural similarity across the family. A ribbon diagram of human S100B is shown in Figure 1.4. While S100 proteins share secondary structure elements with amphoterin, there is no shared tertiary structure nor is there any similarity to AGEs or Aβ. In addition, S100 proteins can be quite acidic with S100B having an isoelectric point of 4.6. The biochemistry of S100B as a RAGE ligand is studied in this dissertation. Therefore, a more detailed discussion of S100 protein structure is warranted and included in the next section.

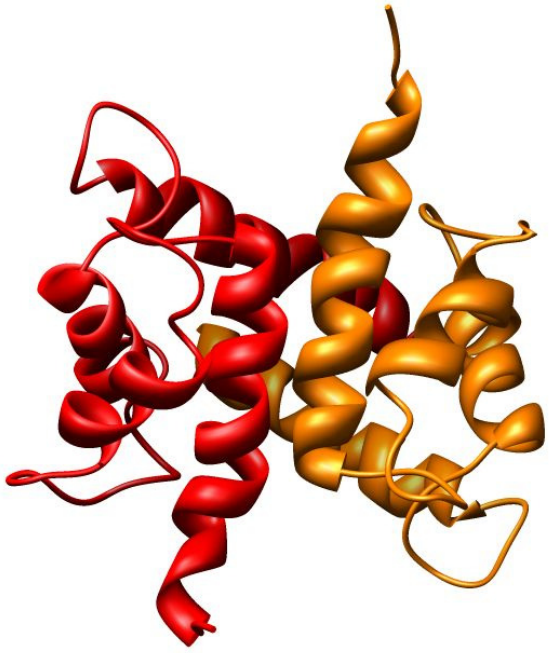


Figure 1.4: Solution structure of Ca^{2+}-loaded human S100B. The ribbon diagram (PDB 1UWO) shows one subunit of the symmetric homodimer in red and the other in orange. All S100 protein structures in this dissertation are shown in this orientation with the two H_{IV} helices in front and the two H_I helices in back.

Divergent Signaling Pathways and Essential Signaling Elements

Kinase Pathways

AGE binding to RAGE induces the activation of NF-κB through $p21^{ras}$ and ERK1/2. Much like AGEs, amphoterin induces the translocation of NF-κB to the nucleus in cultured neuroblastoma cells (23). However, rac and Cdc42 were used to signal to NF-κB instead of ras and ERK1/2. Working in a fibrosarcoma cell line (HT1080), Ishihara and coworkers showed amphoterin activates ERK1/2 (51). This cell line is more similar to the endothelial cells used in AGE studies than the neural cells used by Huttunen and coworkers. Thus, it is possible that the signaling pathways used by RAGE are cell type dependent but not ligand dependent. Interestingly, ERK1/2 coimmunoprecipitated with RAGE from HT1080 cells after activation with amphoterin. The interaction with ERK2 (ERK1 was not tested) was specifically localized to the first 18 amino acids of the intracellular domain in glutathione S-transferase (GST) pull downs from rat lung extracts (51).

Dominant-Negative RAGE

A very important observation was made during initial studies with amphoterin. A truncation mutant of RAGE was transfected into neuroblastoma cells that had endogenous expression of RAGE. This truncation mutant lacked the ~40 amino acid intracellular domain (IC). Translocation to the membrane was not prevented nor was amphoterin or AGE-BSA binding. Cell signaling, however, was deficient in the construct lacking the IC (23). Since wild-type RAGE is also present at the cell surface and ligand binding is not diminished, a dominant-negative form of RAGE (dnRAGE) implies that RAGE functions in some sort of oligomeric state. The likely role of RAGE in neural

growth and development and in progression of diabetes and chronic inflammatory disorders, coupled with utilization of divergent signaling pathways described in this section, has greatly stimulated interest in RAGE as a signaling molecule.

Positive Feedback Loop

Additional insights into RAGE signaling arose from analysis of its promoter. Two functional NF-κB-like binding sites have been characterized, both of which are necessary for basal RAGE expression and increased RAGE expression in response to stimuli (52). RAGE expression is shown to increase in diabetic versus normal vasculature, as well as other pathologic settings (53). Therefore, an interesting scenario has arisen in RAGE biology. Binding of ligand to RAGE activates NF-κB which in turn enhances RAGE expression increasing potential cell surface binding sites. This positive feedback loop has been hypothesized to be responsible for the RAGE induced amplification of chronic disorders such as diabetes and inflammation (54).

Potential for RAGE-Based Therapeutics

As described above, RAGE participates in a diverse array of biological and pathological processes each defined by the particular class of ligand activating the receptor. Due to its implied role in multiple disease states, RAGE is considered a therapeutic target (55). In fact, ligand binding in a variety of culture systems leads to activation of signaling pathways consistent with the associated disease states (5). This suggests blockade of the RAGE signaling axis could provide an avenue for pharmaceutical intervention. Exogenous sRAGE has been used to block ligand binding

and to prevent receptor activation in cell culture. These observations have been translated into mouse models for various diseases.

Diabetes

Apolipoprotein E-null mice naturally develop accelerated atherosclerosis and show a greater than 2 fold increase in plasma AGEs (56). Intraperitoneal administration of sRAGE significantly blocked the development of atherosclerotic lesions in the aorta (57). The mechanism is presumably through blockade of AGE-RAGE interaction since neither hyperglycemia nor insulinemia were affected.

Cancer

Transplanting C6 glioma cells into immunocompromised mice causes rapid tumor growth that is significantly suppressed by intraperitoneal administration of either sRAGE or antibodies against RAGE or amphoterin (58). Furthermore, stably transfected C6 glioma cells either expressing dnRAGE or secreting sRAGE showed neither tumor growth nor tumor metastasis. Thus, the leading edge colocalization of amphoterin and RAGE observed in culture (27) implies a mechanism for promoting (and blocking) tumor invasion.

In a similar study, B16-F1 melanoma cells were injected into the tail veins of immunocompromised mice and assessed for colonies indicative of lung metastasis (38). Two strategies suggested an amphoterin-RAGE signaling axis is important for formation of metastatic colonies in the lung. One, injecting cells transfected with dnRAGE showed 72% fewer colonies than injection of parental B16-F1 cells (38). The second involved coinjection of B16-F1 cells with an amphoterin peptide antagonist (amphoterin 150-183),

which caused a 64% reduction in lung metastatic colonies (38). Thus, multiple lines of evidence point to an amphoterin-RAGE axis important in tumor metastasis.

Alzheimer's Disease

The vascular theory of AD states that accumulation of Aβ in Alzheimer's brain may arise from circulating Aβ and dysfunction of the equilibrium of peptides across the blood brain barrier (BBB) (59). BBB transport would presumably require specific cell surface receptors for Aβ. Using an *in vitro* model for BBB, anti-RAGE antibodies prevented transport of peptides suggesting a role for RAGE (60). Transgenic mice expressing mutant APP function as a model for AD since Aβ containing plaques accumulate rapidly in the brain and correlate with impaired learning and memory (61). Treatment with sRAGE results in high levels of sRAGE-Aβ coimmunoprecipitates from plasma in addition to decreased levels of Aβ plaque formation (62).

Chronic Inflammation

Another potential therapeutic approach is based on the interaction of RAGE with S100A12 which is highly associated with chronic inflammation such as inflammatory bowel disease. Interleukin-10 deficient mice serve as a model for enterocolitis characterized by hyperactive bowel inflammation (63). In this model, administration of sRAGE prevented inflammatory infiltrates into the colon likely by blocking S100-RAGE interaction and signaling (43).

Concluding Remarks

There is a wealth of information on the biology associated with RAGE as a signaling receptor. Comparing the signaling pathways utilized by all ligands suggests that

RAGE universally signals through NF-κB. The biological processes involved range from neural growth and development and acute inflammatory responses to complications associated with diabetes and AD. Thus, knowledge gained by studying one aspect of RAGE function has the potential to transcend all aspects of RAGE biology.

S100 EF-Hand Calcium-Binding Proteins

S100 proteins, specifically S100B, are a major focus of this dissertation. These proteins have been well characterized structurally and biochemically (64) and reagents were readily available to pursue structural studies on S100-RAGE interaction. In this section I take a broader look at calcium signaling involving S100 proteins in addition to the current knowledge about S100 protein structure and function.

Structure of the EF-hand Motif

The calcium ion is known to function as a second messenger in many signaling cascades. Signal transduction from a physiological change in calcium concentration to a biochemical response in the cell is mediated by EF-hand calcium-binding proteins. The EF-hand is a helix-loop-helix motif in which the protein chelates calcium ions using a series of side chain and backbone oxygen atoms (65). The prototypical EF-hand protein is calmodulin, a single polypeptide with four EF-hand motifs organized into two structural domains separated by a flexible linker. Each domain consists of paired EF-hand motifs that fold into a stable four helix bundle. Each EF-hand is able to bind one ion of calcium. Thus, calmodulin binds four calcium ions, two in each structural domain.

Calcium Sensing by EF-Hand Proteins

The functional role of EF-hand proteins in calcium signal transduction is mediated by structural and biophysical responses to changes in intracellular calcium concentration (66). The prototypical EF-hand calcium sensor, calmodulin, binds calcium ions with K_Ds of ~10^{-6} and ~10^{-5} for the C- and N-domains, respectively, in the range of the intracellular calcium signal (67). Thus, at basal levels of calcium, calmodulin is in a magnesium-bound or the apo- (calcium-free) state. The release of calcium from intracellular stores during response to a stimulus leads to micromolar concentrations of calcium, with possible subcellular compartments exposed to millimolar concentrations. At mid-micromolar or higher concentrations of calcium, all four binding sites are filled and calmodulin is activated. The structural basis for calmodulin signaling activity is revealed by comparative analyses of apo- and calcium-bound structures. Upon binding calcium, both domains of calmodulin undergo large structural rearrangements exposing deep hydrophobic pockets frequently used to bind to α-helical target peptides (66).

Architecture of the S100 Protein

S100 proteins are EF-hand containing calcium binding proteins named for solubility in 100% saturated ammonium sulfate. The architecture of an S100 protein is quite different from calmodulin. The minimal structural and functional S100 unit is a highly integrated symmetric homodimer first seen in the structure of S100A6 (calcyclin) determined in our laboratory (68). Each subunit of the S100 dimer consists of two EF-hands connected by a hinge region similar to each of the domains of calmodulin. The C-terminal EF-hand motif is canonical (calmodulin-like) whereas the N-terminal EF-hand is

S100-specific (69). Unlike the canonical 12 residue EF-hand, the S100-specific EF-hand is 14 residues and displays weaker affinity for calcium. Much like calmodulin, affinities for calcium are in the range of the intracellular calcium signal classifying most S100 proteins as calcium sensors.

Calcium binding to S100 proteins induces a similar yet different conformational change than calcium binding to calmodulin for two reasons. One, helices H_I and H_{IV} form the tightly integrated dimer interface and are therefore restricted. Two, the S100-specific EF-hand resembles the calcium bound state even when calcium is absent (70, 71). It has been proposed that the S100-specific EF-hand is stabilized in the absence of calcium by the presence of a water molecule in the binding site. Such a water molecule showing similarity to the calcium coordination geometry was observed in the S100-specific binding loop of apo-S100A3 (72). Regardless of the origin of the effect, the major conformational change in S100 proteins occurs in the C-terminal EF-hand, specifically in helix H_{III} (Figure 1.5). Calcium binding to S100 proteins creates a shallow hydrophobic surface much different from the deep hydrophobic pocket formed by calcium binding to calmodulin.

Due to the high sequence homology among the S100 protein family, it is not surprising that they adopt similar structures in both the absence and presence of calcium (73). However, many subtleties exist. For example, S100A7 and S100A10 have amino acid substitutions or deletions that inactivate calcium binding at key glutamate/aspartate residues in the EF-hands (74). Remarkably, an x-ray crystal structure of S100A10 bound to a target peptide reveals that calcium is not required to stabilize the open conformation since it is sufficiently stabilized by the binding of target (75). It has yet to be determined

whether S100A10 adopts the calcium-bound state in the absence of target binding or is in dynamic exchange between the apo- and calcium-bound states.

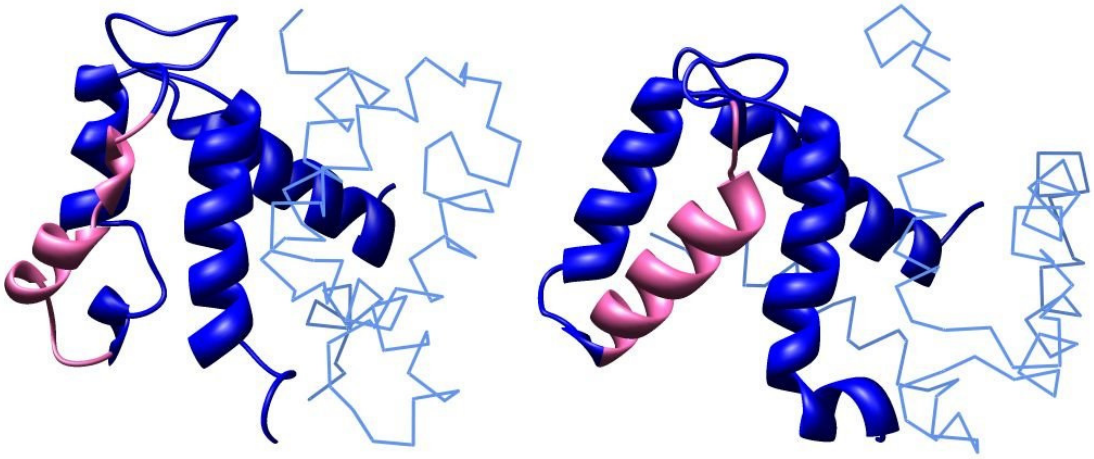


Figure 1.5: Calcium induced conformational change in S100A6. Solution structures of apo-S100A6 (left, PDB 1CNP) and Ca^{2+}-loaded S100A6 (right, PDB 1A03). For each homodimer, one subunit is shown as a light blue Cα trace with a dark blue ribbon diagram showing the 4 helix bundle. The two molecules have been aligned by helices H_I and H_{IV} at the dimer interface. Colored in pink is helix H_{III} emphasizing the calcium induced structural change.

S100 Proteins and Zinc

S100 proteins exhibit the ability to bind other metals, a characteristic that distinguishes them among EF-hand proteins. For example, S100A2 has at least two high affinity sites for zinc (II) that are independent from the calcium binding sites (76). Binding to zinc (II) occurs with nanomolar K_Ds, near the physiological levels of intracellular zinc (II). A solution structure of S100B bound to zinc (II) reveals utilization of sites separate from the calcium binding sites (77). Unlike calcium binding which induces a large conformational change in helix H_{III}, binding of zinc (II) to S100B causes very little structural perturbation (77).

The S100A8/S100A9 heterodimer has also been characterized as a zinc (II) binding protein (78). S100A8/S100A9 is highly expressed in monocytic cells of the immune system and is thought to play a role in the hummoral response to pathogen invasion (79). Addition of the heterodimer but not the individual homodimers inhibits growth of cultured *Candida albicans* (80). Furthermore, growth inhibition by the S100A8/S100A9 heterodimer is reversed by addition of excess zinc (II). Thus, the antimicrobial activity of the S100A8/S100A9 heterodimer may occur through chelation of essential zinc (II). A newly described human disorder characterized by aberrant zinc metabolism, called hyperzincaemia, is correlated with highly elevated levels of S100A8/S100A9 heterodimer (81).

Higher Order Oligomeric States of S100 Proteins

Although S100 proteins form highly integrated dimers, evidence has been gathering in support of S100 proteins functioning *in vivo* as higher order oligomers. Early

reports on the neurotrophic effects of S100B suggested disulfide-linked higher order oligomeric species were essential for function (45, 46, 82). More recent evidence arises primarily from crystal structures and other biophysical techniques. S100A12 has been crystallized in numerous forms, including a hexameric structure (83, 84). This hexamer was apparently stabilized by additional calcium binding sites between individual dimers (85). There is some biophysical support for higher order oligomer formation in the absence of lattice forces from crystal packing. Dynamic light scattering (DLS) reveals an increase in the hydrodynamic volume of S100A12 when calcium concentrations reach extracellular physiological levels (~1 mM) (85).

Gel filtration of S100A4 revealed multiple oligomeric states, which were correlated with potential to induce neurite outgrowth on cultured hippocampal cells. Fractions containing S100A4 tetramer or higher were necessary to induce neurite outgrowth (86). However, it is not clear whether the higher order oligomeric species are in equilibrium with the dimeric species. Thus, further studies involving mutations at key residues involved specifically in higher order oligomer formation are definitely required to test this proposal.

The x-ray crystal structure of S100A8/S100A9 heterodimer has three tetramers in the asymmetric unit (PDB 1XK4). Significant contacts are made between helices H_{IV} of one dimer and the calcium binding loops of the second dimer. This tetramer interface places the traditional S100 target interacting surfaces of opposite dimers in proximity which could generate novel target recognition surfaces. Chemical crosslinking and mass spectrometry suggest calcium dependent tetramer formation of S100A8/S100A9 (87, 88). The crystal structure of S100A10 in complex with a peptide from annexin II also reveals

disulfide linked tetramers (75). In addition, a recently described crystal form of S100B shows the presence of four tetramers (89). It is intriguing to consider S100 oligomerization in the context of receptor activation. Not only can a S100 dimer potentially dimerize a receptor (or alter the organization of a pre-formed receptor dimer), but higher order oligomers can multimerize or create novel binding surfaces for receptor recognition.

Target Recognition by S100 Proteins

Multiple structures of S100 proteins bound to peptide fragments of cellular targets are available and reveal differences in the details of target recognition (64). Crystal structures of S100A10 and S100A11 bound to peptides from annexin II and annexin I, respectively, provided early insight into S100-target recognition (75, 90). In both cases, the annexin peptides adopt α-helical secondary structure and are bound to the S100 protein in a calcium-dependent manner with a stoichiometry of two peptides per S100 dimer (Figure 1.6). Thus, S100 proteins could function by oligomerizing target proteins, by bridging two different molecules, or by inducing conformational changes in proteins by binding to two different regions of one polypeptide.

The interaction between annexin peptides and their respective S100 proteins is stabilized by a combination of hydrophobic and hydrophilic interactions primarily in the shallow pocket formed between helices H_{III} and H_{IV}. Unique to these two structures are contacts between the peptide and helix H_I of the opposite subunit. This type of contact is absent in structures of S100B bound to target peptides from p53 and NDR kinase (Figure 1.6) (91, 92). Nonetheless, the peptides adopt α-helical secondary structure and

contribute a combination of hydrophobic and hydrophilic side chains to stabilize the interaction. Earlier in this chapter I discussed multiligand recognition of structurally diverse molecules by RAGE. Ig domains do not contain a significant amount of α-helical secondary structure. Typically, 3-4 helical turns are required for efficient binding to S100 proteins (Figure 1.6). Thus, the work described in this dissertation sets the stage to study the potentially novel interface used by S100 proteins to bind and activate RAGE.

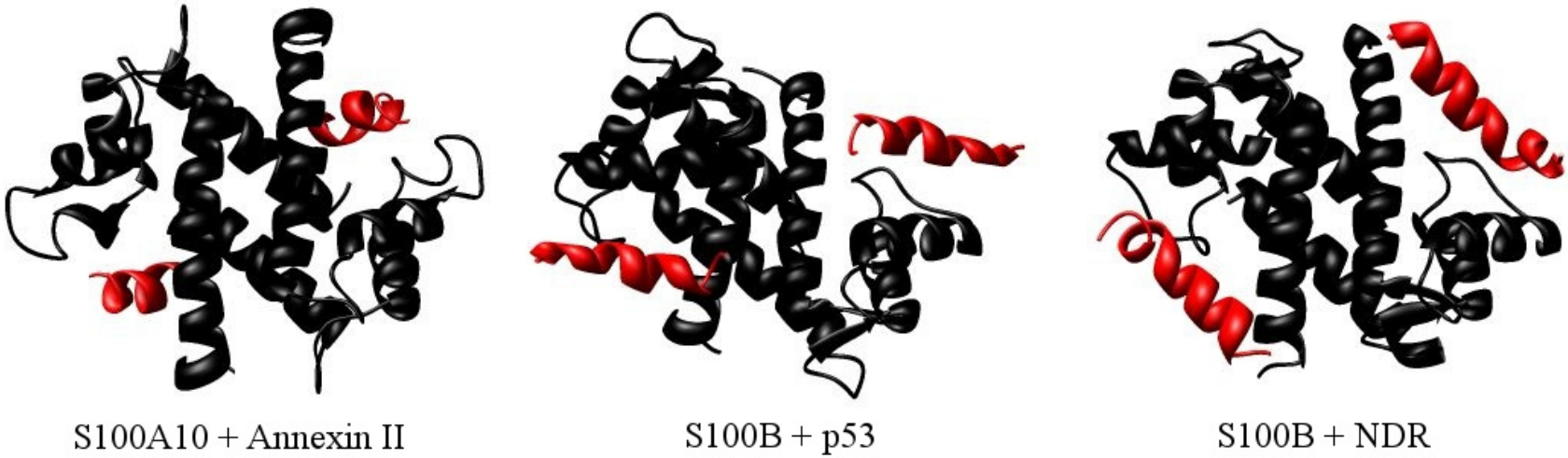


Figure 1.6: Variability in target recognition by S100 proteins. Displayed are the complexes of S100A10 with annexin II peptide (PDB 1BT6), S100B bound to a p53 peptide (PDB 1DT7), and S100B bound to a NDR kinase peptide (PDB 1PSB). Ribbon diagrams shown with both subunits of the S100 dimer in black and the target peptides in red. Figure adapted from (64).

Mechanisms of Cytokine Receptor Signaling and Autoinhibition

RAGE functions as a signaling receptor. The structural work in this dissertation allows me to propose hypotheses about receptor activation taking advantage of the current knowledge available in the literature. A common theme within receptor activation is the concept of ligand induced oligomerization. An increasing level of sophistication in autoinhibition and activation models evolved as more information became available suggesting oligomerization is necessary but not sufficient for activity. In this section, I examine the available data on two well studied systems, growth hormone receptor (GHR) and erythropoietin receptor (EPOR), summarizing structural mechanisms of receptor autoinhibition and activation.

Cytokine Receptor Classification

Cytokine receptors are integral membrane proteins sharing common features such as a large extracellular domain responsible for ligand binding, a single transmembrane helix, and an intracellular domain containing either intrinsic kinase activity or adaptor function to interact with kinases (93). The receptors are classified according to the size and complexity of their extracellular ligand binding domains. Traditionally, “simple” receptors contain only extracellular domains necessary for ligand binding and include GHR, EPOR, and the interleukin-4 receptor (IL-4R). “Complex” receptors contain structured domains not involved in ligand recognition but play a critical role in signal transduction, the best studied of which is the insulin receptor.

Scissor-like Activation of EPOR

Early studies in cytokine receptor function hypothesized the initial step towards activation required ligand-induced dimerization. For example, cysteine mutants in the extracellular region (containing two Ig-like domains D1 and D2) of EPOR (EPO binding protein, EBP) lead to disulfide linked dimers and constitutive activation (94). Production of a peptide that mimics receptor activation (EPO-mimetic peptide, EMP) and induces dimerization of EBP supported ligand-induced dimerization (95, 96). Additional EMPs that efficiently promote dimerization but lack the ability to activate the receptor have been characterized (97). Thus, the basic hypothesis was refined with the discovery that dimerization is necessary but not sufficient for activation.

The structure of unliganded dimeric EBP provides insight into the autoinhibition mechanism of EPOR shown schematically in Figure 1.7 (98). Analysis of this structure placed the membrane proximal D2 domains > 70 Å apart compared to the direct contacts between D2 domains observed in the x-ray crystal structure of erythropoietin bound to EBP (EPO-EBP) (99). Thus, the physical separation of the intracellular domains and bound kinases functions as the autoinhibited state. Ligand binding physically reorganizes the receptor bringing the bound kinases into proximity for trans-phosphorylation and signal activation. This scissor-like model is supported by a fluorescence complementation assay where two fragments of dihydrofolate reductase are fused to EPOR and transfected into Chinese hamster ovary cells. Intracellular binding of fluorescently conjugated methotrexate (which requires proximity of the two fragments) was only possible when the receptor was activated by ligand or when a 30 amino acid linker was present

providing extreme flexibility (the observations with this linker are also consistent preformed EPOR dimers) (100).

Comparisons of the x-ray crystal structures of EMP-EBP and EPO-EBP complexes revealed insights into subtle details of activation (96). The two EBP molecules show ~120° rotation in the crystal structure of the EPO-EBP complex (99). Two EMP-EBP structures show relative rotations different from 120°. Since one EMP was inactive and the other was partially active, a logical conclusion is that a specific receptor orientation is required for efficient activation. Therefore, both proximity of the bound kinases and a specific orientation are required for signal transduction.

Transmembrane Rotation Model for Activation of GHR

GHR was also thought to function through ligand-induced dimerization when a combination of titration calorimetry and gel filtration suggested a stoichiometry of 1:2 between growth hormone (GH) and the ligand binding region of GHR (GHbp) (101). This stoichiometry was confirmed in the crystal structure of GH-GHbp (102). Studies using a variety of monoclonal antibodies that target the GHR extracellular domain show that dimerization is not sufficient for activation but that specific structural orientation of the receptor is necessary (103). Unlike EBP, GHbp exists as a monomer despite a combination of coimmunoprecipitation and FRET studies showing unliganded GHR dimers in cell culture (104). A mechanism for GHR signal activation has also been proposed and is depicted schematically in Figure 1.7. A series of alanine insertions were made between the transmembrane helix and the intracellular domain. Constitutive activation was achieved when exactly four alanines were introduced, which is predicted

to produce an ~40° net rotation (total of ~400°) of the intracellular domain if they continue α-helical secondary structure (104). Thus, unlike EPOR, which dimerizes through the extracellular domain in an inactive conformation physically separating the kinase adapter intracellular domain, GHR appears to dimerize through the transmembrane helix with its kinase adapter domains rotated in an orientation that autoinhibits trans-phosphorylation.

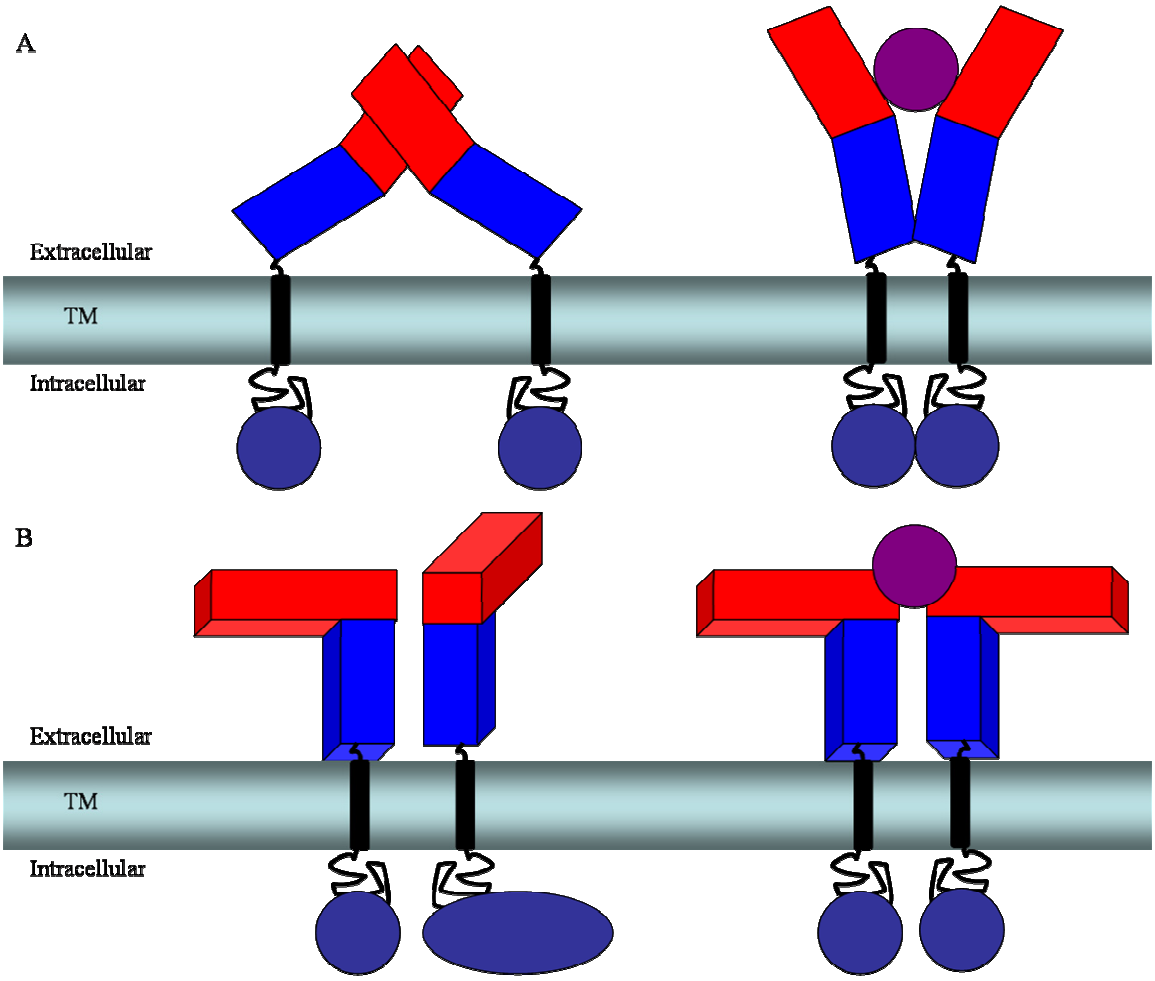


Figure 1.7: Two models for cytokine receptor autoinhibition and activation. (A) Scissor-like model for EPOR as described in the text. D1 and D2 domains of EPObp are shown in red and blue, respectively. JAK2 is shown as a blue circle associated with the intracellular domain. EPO is shown as a violet circle. Figure adapted from (99). (B) Rotation model for activation of GHR as described in the text. The two extracellular domains of GHbp are shown as red and blue 3D rectangles. One associated JAK2 is a blue circle and the other is a blue ellipse. GH binding (violet circle) induces rotation aligning the JAK2 molecules for activation. Figure adapted from (104).

Structural Studies on sRAGE and Ligand Interactions

The central focus of this dissertation is to understand the structure of RAGE and its function as a signaling receptor. Specifically, I describe the first detailed structural and biophysical characterization of sRAGE (Chapter II). The techniques used include limited proteolysis, a powerful approach for analyzing flexible regions of a protein (105). A combination of CD and NMR spectroscopy are used to determine the secondary and tertiary structure elements of sRAGE and single and tandem domain constructs. The combination of these experimental procedures revealed unique features of sRAGE quaternary structure. Further insights into sRAGE structure and function were obtained from a high resolution x-ray crystal structure of the tandem VC1 construct of sRAGE (Chapter III). Lastly, the interaction of sRAGE with two of its physiological ligands, S100B and CML, is characterized using a combination of protease protection, surface plasmon resonance, and NMR (Chapter IV). The combination of these results provided insight into sRAGE structure and RAGE-ligand interactions, the broader implications of which are discussed along with ideas about future directions for structure-based research on RAGE (Chapter V).

CHAPTER II

THE EXTRACELLULAR REGION OF THE RECEPTOR FOR ADVANCED GLYCATION END PRODUCTS IS COMPOSED OF TWO INDEPENDENT STRUCTURAL UNITS

Introduction

RAGE is a member of the immunoglobulin (Ig) superfamily containing three predicted Ig-like domains in its N-terminal ligand binding region (1). Further classification to an N-terminal V-type Ig domain and two subsequent C-type Ig domains was proposed based on sequence analysis; however, this assignment was based largely on homology to NCAM, whose β-strand topology has since been shown to differ slightly from V-type and C-type domains (106).

As described in Chapter I, a critical aspect to understanding mechanisms of receptor signaling is obtaining information on the natural or ligand-induced quaternary structure of the receptor. In fact, while the general tertiary structure of the domains in sRAGE is assumed to be a two-sheet β-sandwich stabilized by an internal disulfide bond, little information is known about sRAGE quaternary structure. In this chapter, production protocols for sRAGE and a series of domain constructs are developed to establish a foundation for structural studies. Subsequently, the issue of sRAGE quaternary structure is addressed using limited proteolysis, which can identify variability in structured vs. unstructured regions of the protein (105). Lastly, the relationship of the three Ig domains is assayed by characterizing their thermal denaturation midpoints and analyzing multidimensional NMR spectra and a homology model for the tandem VC1 domains.

Production of sRAGE and Domain Constructs

High level production and purification protocols were developed for sRAGE and the five single and tandem domain constructs (VC1, C1C2, V, C1, C2; see Figure 1.1). While the production of recombinant sRAGE has been described for both eukaryotic and prokaryotic expression systems (43, 107-109), the strategy described here provided consistent high level production of all six constructs without the need to refold protein from inclusion bodies. Bacterial expression vectors were constructed that enable purification via an N-terminal His_6-tag and nickel affinity chromatography (see Chapter VI). Use of cells providing an oxidizing environment and optimization of expression parameters (temperature, duration, IPTG concentration) produced soluble protein for each of the constructs. Protein yields after purification were similar for sRAGE, VC1, V, and C2 (15-20 mg/L) and lower for C1C2 and C1 (5-10 mg/L) as a result of their lower expression levels.

When performing structural analysis on isolated domains of a protein, it is important to properly define construct boundaries. Initial boundaries were based on sequence alignment with known Ig family sequences and secondary structure prediction. Ala23 represents the first amino acid of the native receptor after loss of the signaling peptide and was selected as the N-terminus for our sRAGE construct. The C-terminus of sRAGE, Gly327, represents the end of the C2 structural region, which is linked to the transmembrane helix. The N-terminus for the V and VC1 constructs and the C-terminus for the C1C2 and C2 constructs were the same as for sRAGE. The predicted linker region separating C1 and C2 was identified by multiple sequence alignments of four RAGE genes (human, bovine, rat, and mouse; Figure 2.1). The eleven amino acid insert in

bovine RAGE after Trp230 aided identification since regions of low identity and homology often represent loop and linker regions. The intervening boundaries between V and C1 proved to be more challenging to precisely identify and had to be manually optimized based on expression level, solubility, and stability. This analysis suggests there may be unique features relating V to C1.

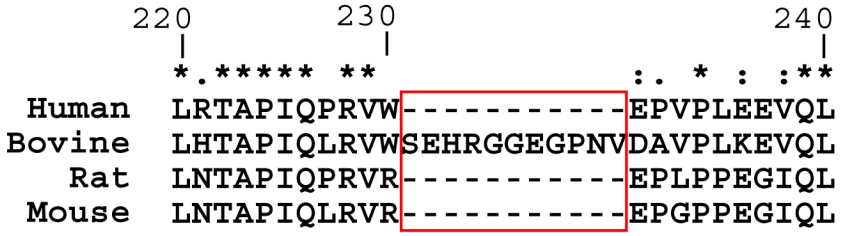


Figure 2.1: Multiple sequence alignment of RAGE proteins. From top to bottom, human (Q15109), bovine (Q28173), rat (Q63495) and mouse (Q62151) RAGE were aligned using ClustalX run on a Windows PC. The numbering on the top represents the residue number of the human sequence. The symbol (*) denotes 100% conservation while the symbols (:) and (.) represent high and moderate conservation, respectively. See Appendix A for full sequence alignment.

Analysis of purified constructs under reducing and non-reducing SDS-PAGE conditions showed a mobility shift consistent with formation of disulfide bonds. Observing a single band under non-reducing conditions suggested homogeneity in disulfide formation. Interestingly, ^{15}N-^{1}H HSQC spectra of V acquired in the presence of reducing agent showed a collapse of the chemical shift dispersion in the ^{1}H dimension, which is characteristic of an unfolded protein (Figure 2.2). This suggests any experiments in the presence of reducing agents could lead to artifacts due to protein unfolding.

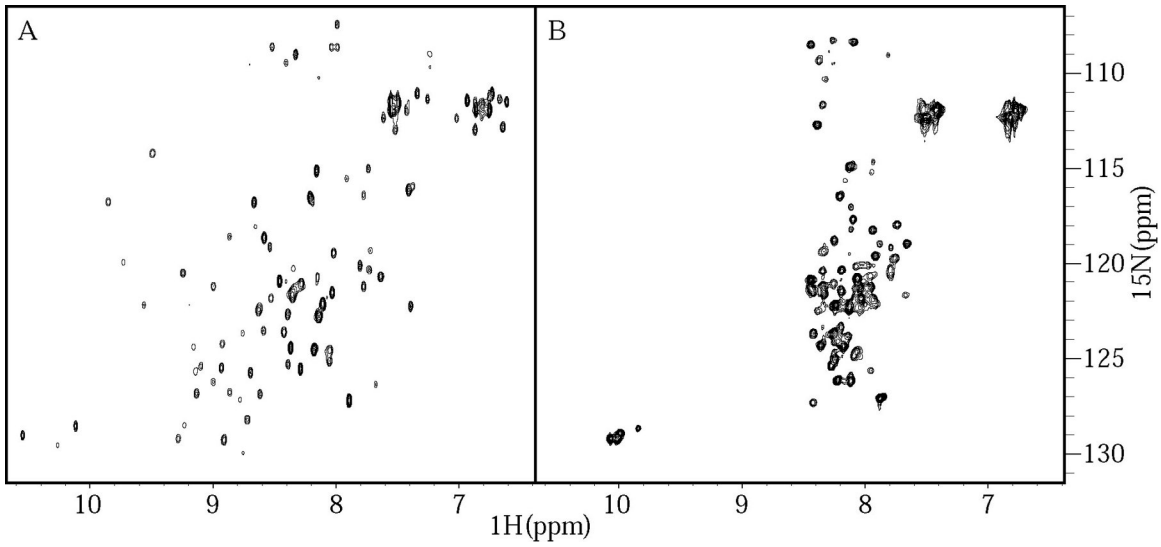


Figure 2.2: The effect of reducing agent on Ig domain tertiary structure. (A) ^{15}N-^{1}H HSQC spectrum of isolated V in the absence of DTT acquired according to Chapter VI. (B) ^{15}N-^{1}H HSQC spectrum of V in the presence of 1 mM DTT.

Stability of sRAGE Domains

A series of limited proteolysis experiments were performed in order to further characterize the domain structure of sRAGE. A time course for digestion of the protein by trypsin shows that sRAGE is completely digested within four hours (Figure 2.3A). Two major species are formed during early time points, one migrating at ~25 kDa (Figure 2.3A, arrow b) and another at ~12 kDa (Figure 2.3A, arrow c). In-gel digestion of the 25 kDa band with trypsin followed by MALDI-MS analysis showed this band contained peptides unique to the V and C1 domains. Figure 2.3B shows the sequence coverage obtained for both intact sRAGE (Figure 2.3A, arrow a) and the ~25 kDa fragment. Notably, coverage obtained for the V and C1 domains was nearly complete in the 25 kDa band (Figure 2.3A, arrow b), ranging from a peptide encoding a region of the N-terminus (23-29) to a peptide near the C-terminus of C1 (199-216).

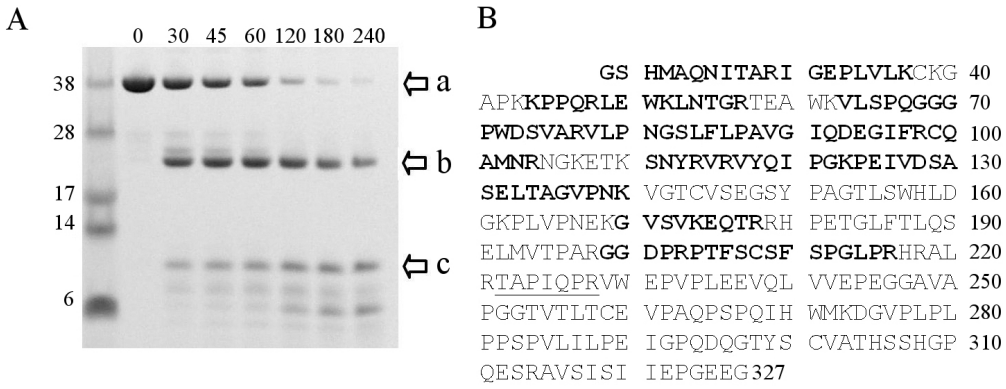


Figure 2.3: Limited proteolysis of sRAGE reveals two stable domains. (A) Reducing SDS-PAGE of sRAGE digested by trypsin (500:1, w/w) at room temperature. Arrows at right highlight the three main fragments. Each of these bands shift in mobility when gels are run under non-reducing conditions indicating they contain one or more disulfide bonds. Time points are given in minutes. (B) Sequence of recombinant sRAGE after removal of His_6-tag. The first four amino acids (GSHM) remain after thrombin cleavage. Ala23 is the first amino acid of the mature protein after the signal sequence is removed. Highlighted in bold are peptides observed in mass spectrometry analysis of the trypsinized 25 kDa band (band "b") in panel (B). The only peptide observed in the mass spectrometry analysis of band "a" that is not seen in band "b" is underlined. The numbering on the right corresponds to the last amino acid in each row.

To further confirm the origin of the 25 kDa band, limited proteolysis of VC1 (Figure 2.4) was performed under conditions identical to those used for sRAGE. Digestion of VC1 produced a small truncation to a band with the same mobility as the 25 kDa band generated from sRAGE digestion (arrow d, Figure 2.4). Both the intact VC1 and the 25 kDa fragment were digested at the same rates as those observed in the experiment with sRAGE. Notably, the 25 kDa fragment produced from either sRAGE or VC1 digestion was largely stable to further proteolysis despite the presence of both lysine and arginine residues in the expected linker region between the two domains. To further understand the apparent stability of the V-C1 linker we performed limited proteolysis experiments on isolated V and C1. Figure 2.4 shows that both intact V and especially C1 were much more susceptible to trypsin proteolysis than the 25 kDa fragment that contained both V and C1. Furthermore, we failed to observe a truncation to a stable fragment for either construct that would indicate the presence of a single stable V or C1 domain.

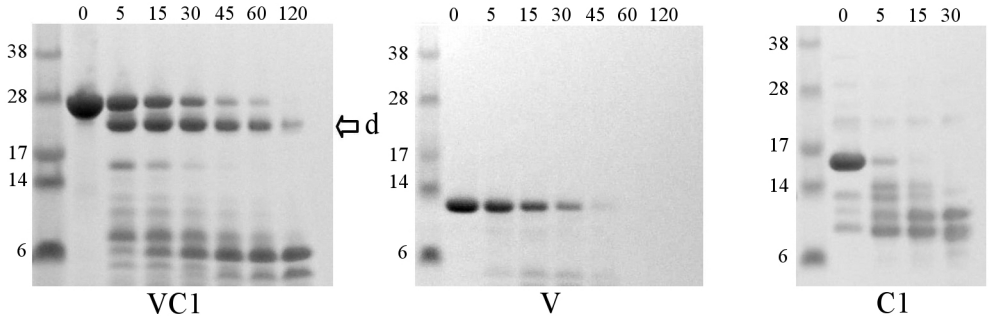


Figure 2.4: Resistance to proteolysis of V and C1 is enhanced by covalent attachment. Reducing SDS-PAGE analysis of VC1, V, and C1 proteolyses by trypsin (500:1, w/w). The mobility of band "d" and that for the intact protein shift when run under non-reducing conditions. Time points are given in minutes.

Unlike sRAGE and VC1, the C1C2 construct was completely digested by trypsin in 5 minutes, and only a single 12 kDa band was observed thereafter (Figure 2.5, arrow e). This stable band had identical mobility to the 12 kDa band seen in the sRAGE digestion (Figure 2.3, arrow c). The presence of this band in sRAGE and C1C2 digestions, but not for VC1, implies the band belongs to the C2 domain. Exposure of isolated C2 to trypsin showed it is completely stable to digestion by this protease (Figure 2.5). This observation is identical to that made for the 12 kDa band in the sRAGE (Figure 2.3, arrow c) and C1C2 digestions. The assignment of the 12 kDa band to C2 and the absence of other bands in C1C2 digestions indicate that the C1 domain on its own is not stable.

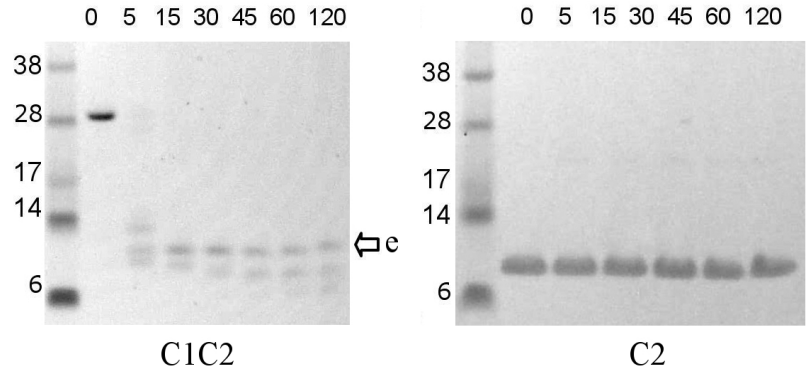


Figure 2.5: The C2 domain is stable to trypsin proteolysis. Reducing SDS-PAGE analysis of C1C2 and C2 by trypsin (500:1, w/w) at room temperature. The mobility of band "e" corresponds to the ~12 kDa band "c" in Figure 2.3A. Time points are given in minutes.

Differential scanning calorimetry (DSC) experiments were performed to better understand the stability and interdependence of sRAGE domains. Intact sRAGE showed a single thermal transition with a T_m at 55.1 °C, which was surprising given the much higher T_m value of isolated C2 (vide infra). Unfortunately, heating beyond 60 °C leads to aggregation and precipitation (Figure 2.6, solid line), which precludes further analysis. VC1 responded in a very similar manner, with a single transition corresponding to T_m at 55.1 °C followed by aggregation and precipitation (Figure 2.6, dashed line).

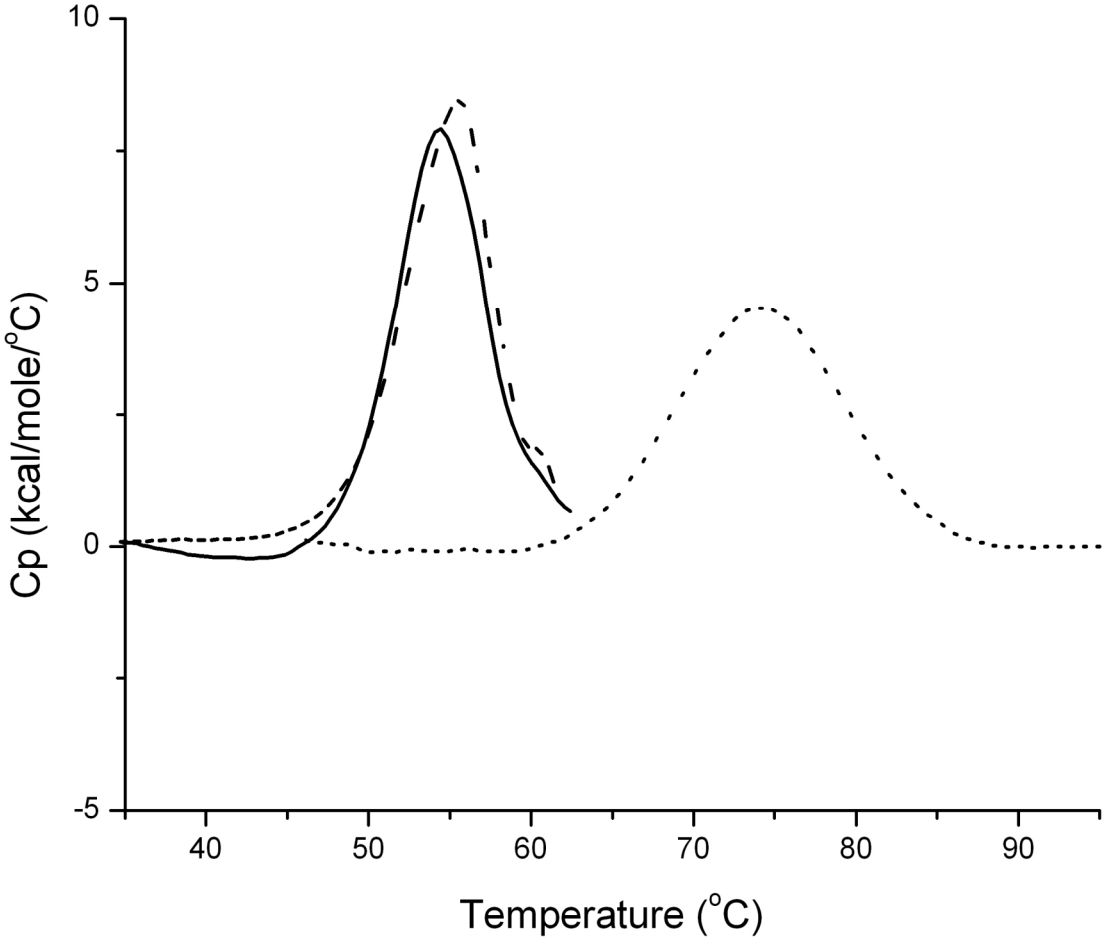


Figure 2.6: VC1 undergoes a complex thermal transition. DSC thermograms are shown for sRAGE (−), VC1 (--) and C2 (··). Experimental details are described in Chapter VI.

In an attempt to determine if the thermal transition derives specifically from one domain, DSC experiments were performed on isolated V and C1. However, although the proteins remained soluble, no discrete transitions were observed for either of the isolated domains (Figure 2.8A). These observations are consistent with the limited proteolysis experiments, which suggested that isolated V and C1 are not stably folded independent domains. Based on the evidence from the thermal denaturation and limited proteolysis data, we attribute the unfolding event at 55 °C to uncoupling of the V and C1 domains. Remarkable parallels are found in a very thorough study of the unfolding of a tandem (one V-type and one C-type domain) Ig protein from a multiple myeloma κI light chain, from which the authors concluded that the phenomena observed did not correspond to unfolding of independent structural domains (110). As an alternative to the DSC approach, the temperature dependence of the CD spectrum of VC1 was examined. The rationale is that the degree of unfolding of the protein would be reflected in the degree of loss of CD ellipticity. Since the protein precipitates at higher temperatures, the maximum temperature was restricted to 55 °C. Figure 2.7 shows a reversible (and incomplete) loss of CD ellipticity is observed as the temperature is raised. The reversibility of the thermal melt rules out precipitation as the only contributing factor to the CD transition. In addition, the retention of significant secondary structure (~60%) is consistent with uncoupling of V and C1, and not global unfolding, as interpreted for the DSC results.

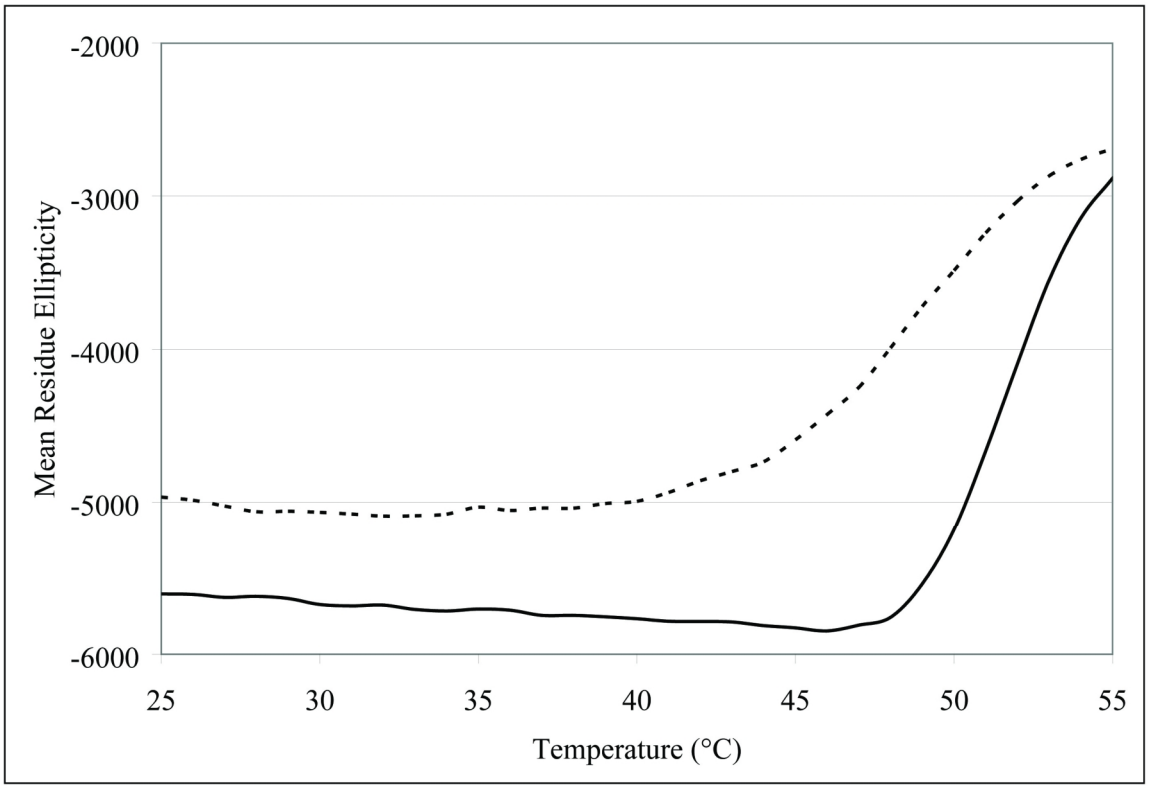


Figure 2.7: CD thermal melt of VC1. Shown are the melting (−) and subsequent cooling (··) of VC1 from 25 °C to 55 °C. Raw data were converted to mean residue ellipticity according to Chapter VI. The difference in mean residue ellipticity at 25 °C is consistent with a small amount of protein precipitation during the experiment.

Analysis of C1C2 and C2 constructs showed a single unfolding transition with T_m values of 72.7 °C and 74.3 °C (Figure 2.8B; Figure 2.6, dotted line), respectively. In both cases this transition was nearly 100% reversible. This distinct thermal transition is attributable to unfolding of the C2 domain, which supports our limited proteolysis data suggesting that C2 is independent from the rest of the sRAGE molecule. Note that the unfolding transition for C2 in the context of full sRAGE was not observed because it was masked by the aggregation/precipitation associated with VC1 that occurs well below the T_m of C2. Taken together, the limited proteolysis and DSC data imply that the three predicted domains are present. However, whereas C2 is a completely independent and highly stable domain, V and C1 are seen to have a significant influence on each other's stability.

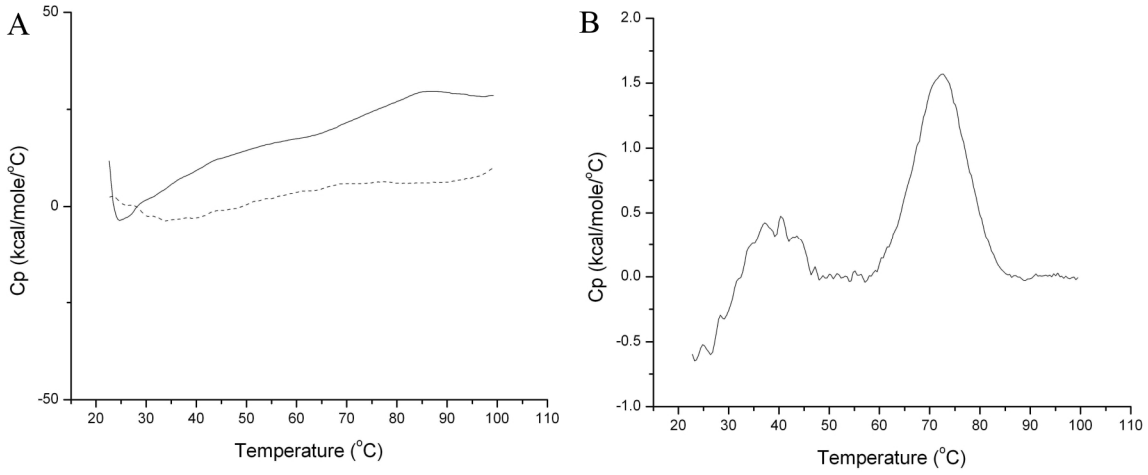


Figure 2.8: Thermal stability of V, C1, and C1C2 constructs. (A) DSC thermograms for V (−) and C1 (··). Shown is one of many inconsistent results. (B) DSC thermogram for C1C2. Experimental details are described in Chapter VI.

sRAGE is Composed of Coupled V and C1 Domains With an Independent C2

Circular dichroism, dynamic light scattering, and NMR were used to characterize the secondary and tertiary structure and the oligomeric state of sRAGE and domain constructs. CD provided distributions of secondary structure elements for each of the six protein constructs, all of which showed a combination of β-sheet and random coil (Figure 2.9). Remarkably, sRAGE, VC1, C1C2, and C2 all were assigned 47% β-sheet, 48% random coil, and 5% α-helix, and only slightly reduced β-sheet and increased random coil was observed for isolated V (40% β-sheet, 51% coil) and C1 (43% β-sheet, 48% coil). These results show all six protein constructs contain secondary structure highly consistent with one or more Ig-like folds.

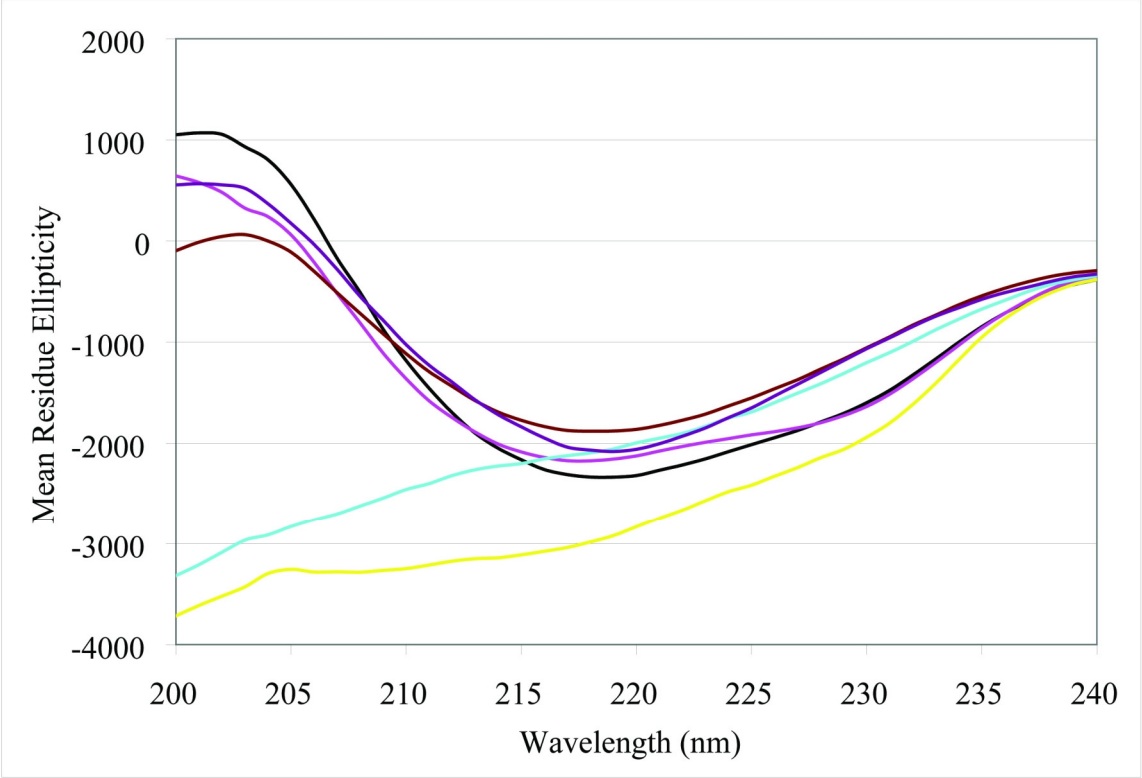


Figure 2.9: Circular dichroic spectra of sRAGE and domain constructs. CD spectra are shown for sRAGE (black), VC1 (pink), C1C2 (dark red), V (yellow), C1 (turquoise), and C2 (violet). Data shown are representative of at least three accumulations. All data were collected, converted to mean residue ellipticity, and secondary structure content estimated as described in Chapter VI.

The tertiary structure of sRAGE and each of the tandem and single domain constructs was analyzed by ^{15}N-^{1}H HSQC NMR (Figure 2.10). Despite the relatively small size of the three domains, it was surprisingly difficult to obtain high quality NMR spectra. Experimental conditions (temperature, pH, and ionic strength) for each construct had to be optimized independently. Dynamic light scattering was also used to complement the NMR experiments, in particular to assess whether the domain constructs were monomeric under the conditions of the NMR experiments. Hence, the light scattering data were collected at the same protein concentrations as those used in NMR experiments.

In the ^{15}N-^{1}H HSQC spectrum of sRAGE, less than 100 crosspeaks were observed (Figure 2.10A). The linewidths of these signals were relatively narrow, as would be expected for a single Ig domain. Moreover, the signals correspond precisely to the peaks observed in the corresponding spectrum of isolated C2 (Figure 2.10F). Since the NMR chemical shift is an exquisitely sensitive probe of structure, these observations indicate that the observed peaks in the sRAGE spectrum arise from C2. Of note, signals in both spectra have uniform intensity and line width and are well dispersed, consistent with a well-folded globular domain. In addition, light scattering data on isolated C2 were fit to a mean radius of 2.21 ± 0.27 nm with a polydispersity of 12%, fully consistent with a monomeric non-spherical β-sandwich.

An interesting question is why the V and C1 peaks from sRAGE are so much broader than the C2 peaks that they are not observed in a standard HSQC spectrum. Since previously published analytical ultracentrifugation experiments showed that sRAGE is monomeric (107), the likely explanations are the two domains tumble in solution with the

characteristics of a larger particle or that they are both conformationally heterogeneous. To obtain further insight, light scattering experiments were performed on sRAGE to verify that the protein is monodisperse. Unfortunately, the data were ambiguous because the level of polydispersity assigned (33.5%) is above the threshold for which reliable conclusions can be drawn. [I believe sRAGE is not poly-disperse, but rather that the fitting of light scattering data is not straightforward because sRAGE contains two independent, flexibly linked structural modules.] The smaller VC1, V and C1 constructs were then used to obtain further insights.

The ^{15}N-^{1}H HSQC spectrum of VC1 showed at least 150 distinct backbone peaks out of a maximum of ~200 (Figure 2.10B), as expected for a protein of this size. Not every peak will be discernable in a 2D experiment because those in the crowded central region are likely to overlap. Importantly, the relatively sharp and well dispersed peaks from C2 observed in the spectrum of sRAGE were clearly not present in the spectrum of VC1. Moreover, the linewidths of the peaks in the VC1 spectrum were significantly larger than those observed for the C2 domain in the spectrum of sRAGE. Light scattering data on VC1 were fit to an average radius of 3.43 ± 0.33 nm with a low level of polydispersity (9.5%). The radius suggests an elongated molecule consistent with the homology model generated for VC1 described below in addition to the ellipsoid structure modeled from analytical ultracentrifugation (107), and indicates that the VC1 protein is monomeric. The ability to observe a rather complete and well-dispersed ^{15}N-^{1}H HSQC spectrum for VC1 with relatively uniform peak intensities, combined with the light scattering analysis and the lack of signals for VC1 residues in the spectrum of sRAGE,

strongly implies that the V and C1 domains are well-folded and suggests that they tumble as an integrated structural unit.

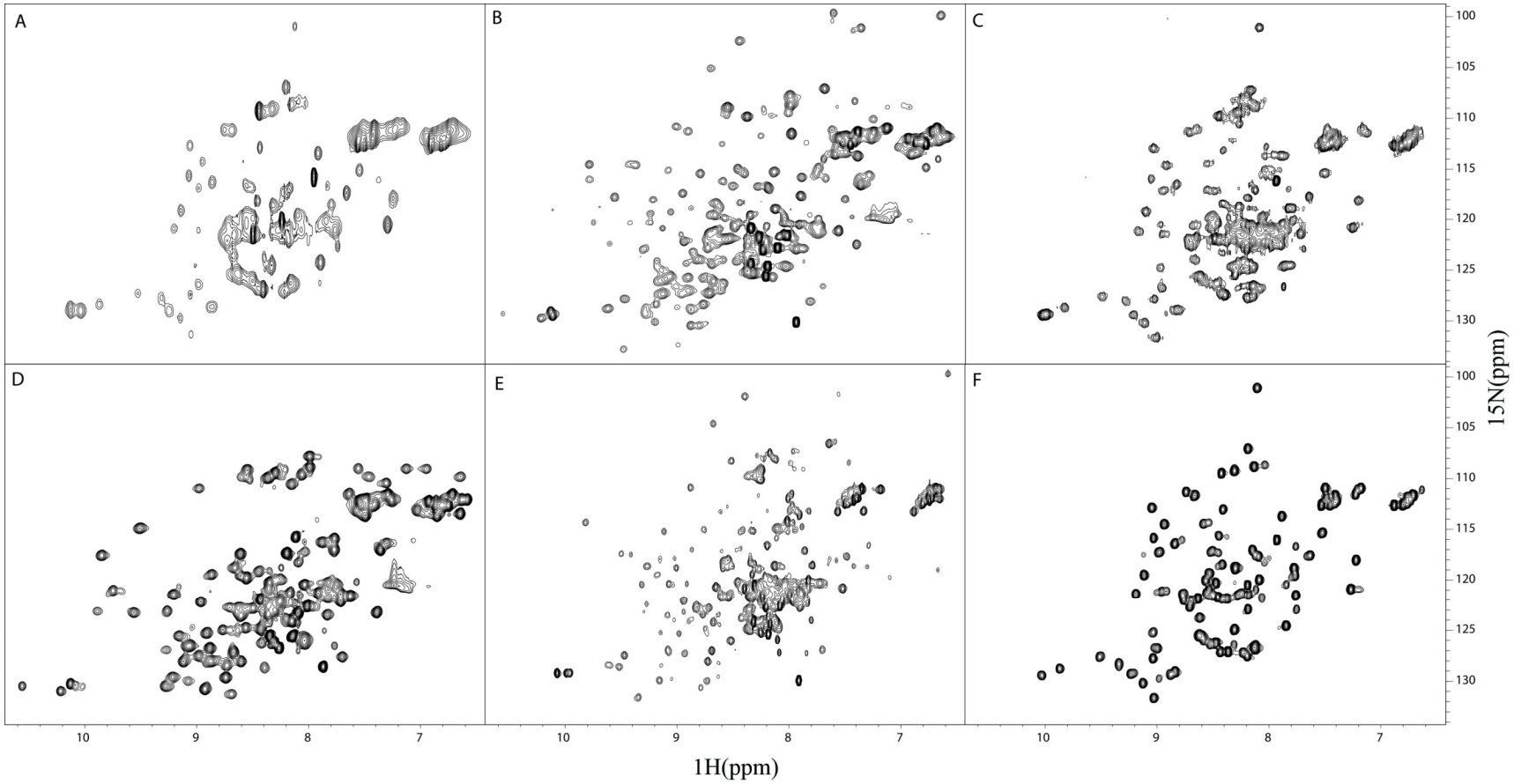


Figure 2.10: Heteronuclear NMR spectra of RAGE constructs. ^{15}N-^{1}H HSQC spectra are shown for sRAGE (A), VC1 (B), C1C2 (C), V (D), C1 (E), and C2 (F). The data were collected at 600 (C1C2, C1) or 800 (sRAGE, VC1, V, C2) MHz at 25 °C. Solution conditions are provided in Chapter VI.

To further verify the conclusions from the analysis of VC1, the use of specialized techniques to improve the sRAGE spectrum including TROSY and perdeuteration of the protein was investigated (111, 112). Perdeuteration in particular provided a substantial improvement (Figure 2.11), and resulted in well over 200 discrete peaks in the spectrum. Interestingly, there are two populations of peaks in the spectrum of perdeuterated sRAGE with different relative intensities and linewidths. Many peaks from the subset with lower intensity/wider linewidths were well dispersed into both upfield and downfield regions and corresponded well with peaks in the spectrum of VC1 (Figure 2.11, red circles). The subset with higher intensity/narrower linewidths overlap well with the C2 peaks observed in the spectra of protonated sRAGE, C1C2, and C2 (Figure 2.10A, 2.10C, 2.10F, Figure 2.11, green circles). Together, these observations show that V and C1 are indeed folded in the context of sRAGE and that the lower intensity/wider linewidth peaks in the spectrum of sRAGE arise from these domains. The fact that linewidths of peaks from the V and C1 domains are the same and significantly larger than those of C2 strongly implies that V and C1 form an integrated structural unit.

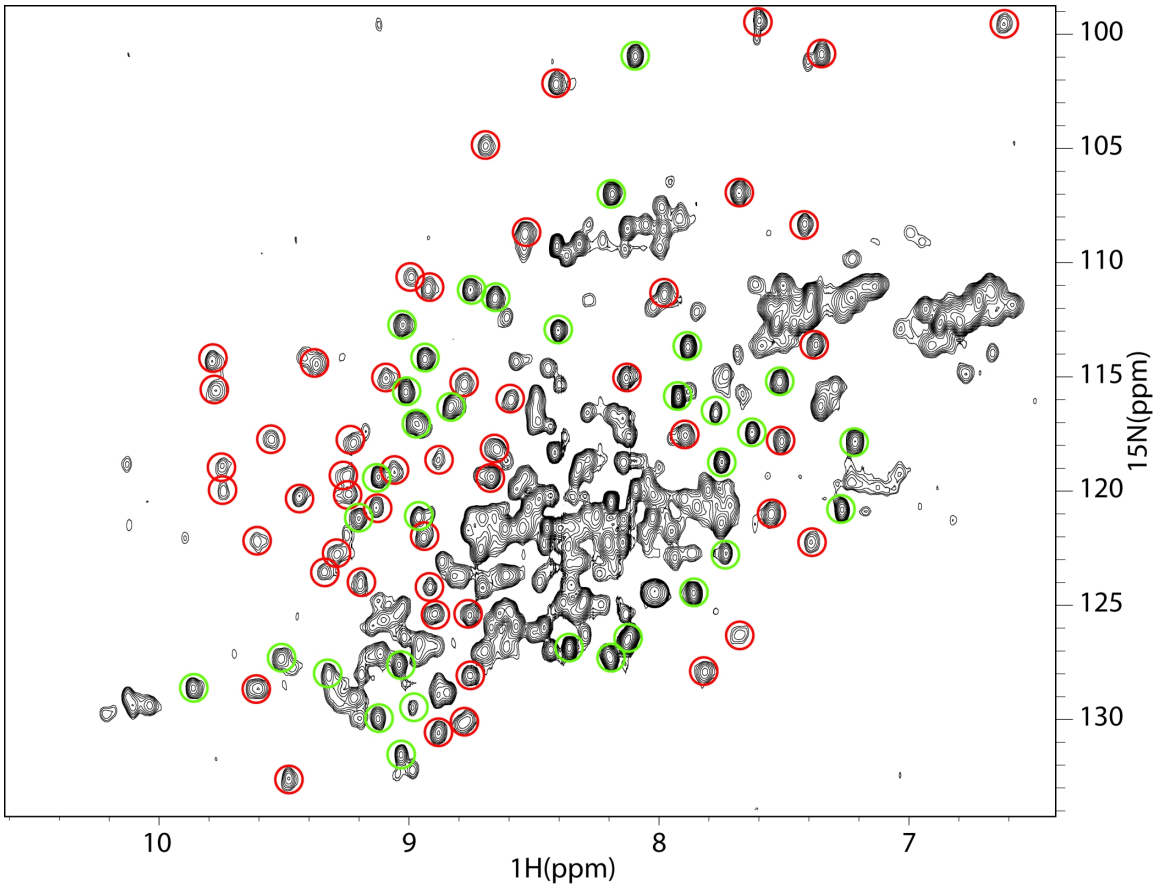


Figure 2.11: VC1 and C2 are separate structural units. ^{15}N-^{1}H HSQC spectrum of perdeuterated sRAGE is shown with peaks labeled based on direct comparison to the spectra of VC1 and C2. Red circles denote crosspeaks uniquely assigned to VC1 and green circles denote crosspeaks uniquely assigned to C2. Unmarked crosspeaks represent those that are overlapped or are not sufficiently unique to be confidently assigned by comparison to one of the two sub-spectra.

Further support for the structural integration of the V and C1 domains was obtained from comparisons to the chemical shifts of isolated V and C1 (Figure 2.10). The ^{15}N-^{1}H HSQC spectrum of C1 is highly heterogeneous, with a large excess of peaks. In addition to a number of well dispersed peaks, a large number of overlapped peaks in the central region are observed along with a significant variation in signal intensity (Figure 2.10E). Light scattering experiments revealed a very high 39.2% level of polydispersity. These observations are indicative of a heterogeneous C1 domain with substantial unfolded or partially folded regions. The ^{15}N-^{1}H HSQC spectrum of isolated V (Figure 2.10D) has about the expected number of peaks but linewidths are significantly larger than the peaks observed in the spectrum of the well-folded C2 domain. The larger linewidths of V are presumably the result of conformational heterogeneity as observed for C1 or possibly compounded by some self-association, both as a byproduct of separation from C1. The high level of polydispersity (34.5%) in light scattering on this domain is consistent with a lack of structural stability and a tendency to aggregate.

Despite the complications of working with the isolated V and C1 domains, detailed comparison of NMR chemical shifts of V, C1, and VC1 could be performed to obtain domain-specific assignments. This was possible because many peaks in the spectrum of VC1 could be assigned to the V or C1 domain since they appeared in a nearly identical and distinctive location in the spectrum of both VC1 and the corresponding isolated domain. In addition to these peaks that correspond very closely, there were a number of characteristic peaks in the spectra of isolated V and C1 that appear in clearly different locations in the spectrum of VC1, as shown in Figure 2.12. For the V domain, approximately 50 peaks outside the crowded 8.0-8.5 ppm ^{1}H region could

be distinguished. Among these, approximately 20 were clearly shifted relative to their position in the spectra of VC1 and sRAGE. A number of readily distinguishable C1 peaks could also be identified, but the analysis was not nearly as complete due to the heterogeneity of C1. Regardless, there is sufficient data to conclude that there are perturbations of peaks in the downfield dispersed region, which indicates that the structure of V is altered by the presence of C1 and vice versa, i.e. there is a specific structural interface between V and C1. In summary, these data show the V and C1 domains of sRAGE form an integrated structural unit that is structurally independent of and dynamically linked to the C2 domain.

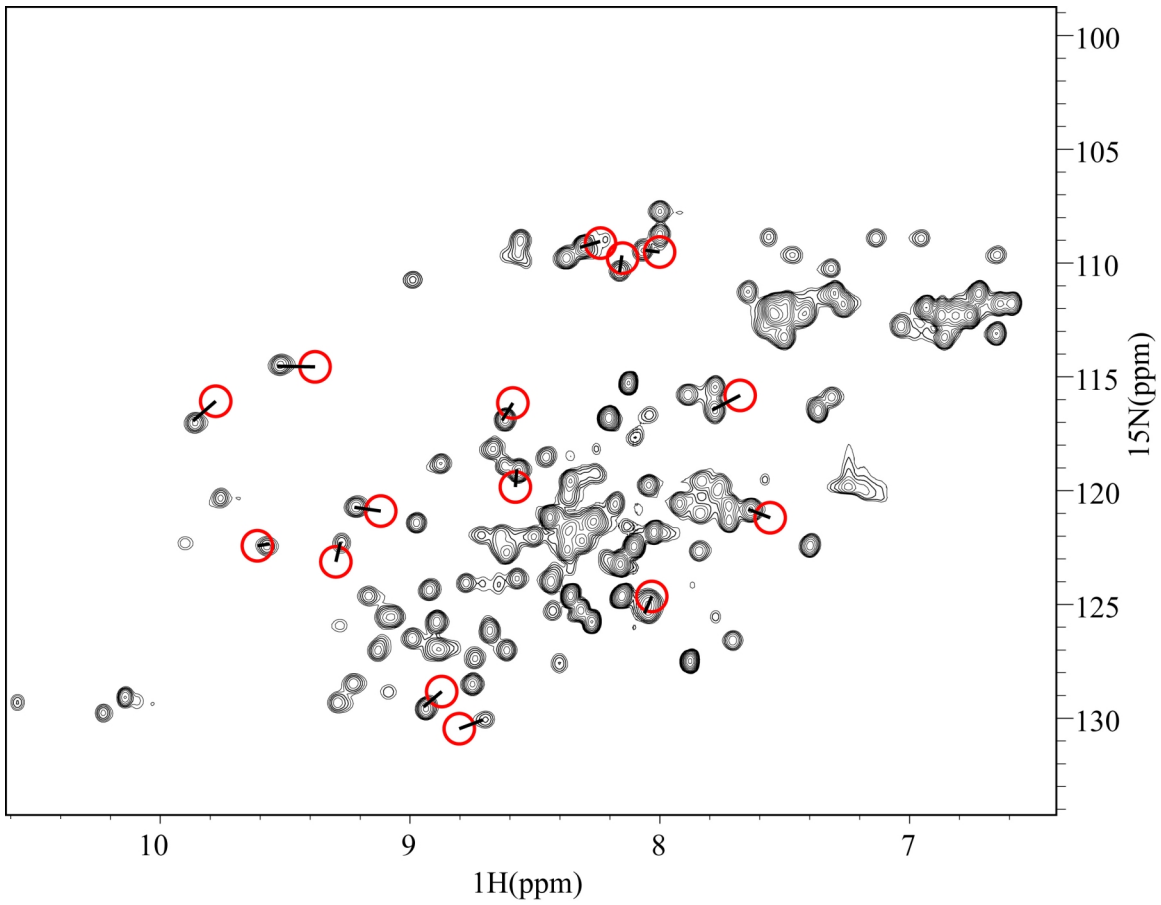


Figure 2.12: The V and C1 domains form a discrete structural interface. Comparison of the heteronuclear NMR spectra of isolated V construct and the tandem VC1 construct. The ^{15}N-^{1}H HSQC spectrum of isolated V is shown. Red circles mark the position of same crosspeaks in the spectrum of VC1. Lines are drawn between a crosspeak in the V spectrum and the crosspeak of nearest proximity in the VC1 spectrum. Unmarked crosspeaks either do not experience significant chemical shift perturbations or cannot be analyzed due to overlap.

Homology Modeling of sRAGE

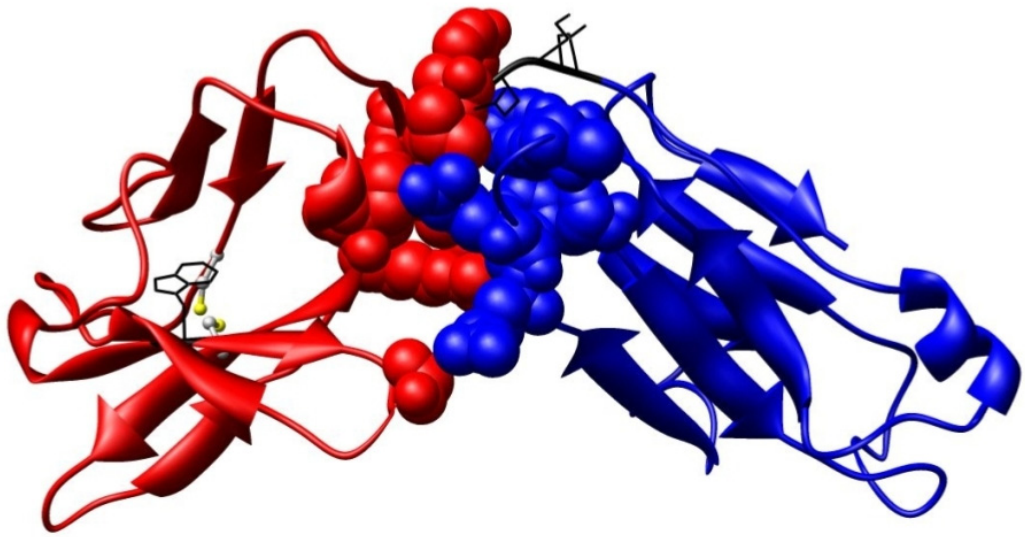
A homology model of the three-domain sRAGE protein was constructed using MODELLER software via the ESyPred3D web server in order to facilitate analysis of the structural data and better understand the implications for cellular signaling. An extensive database of Ig-like domain sequences and structures are available, which implied a high quality model could be generated. The automated pairwise alignment module selected NCAM (PDB 1QZ1) as the template protein because it has 19% identity with sRAGE after alignment. However, the structure of NCAM has no significant linkers between any of the domains, which differs from our analysis of sRAGE. Thus, models of VC1 and C2 were generated separately.

The pairwise alignments against VC1 resulted in highest homology with an Fab light chain (PDB 1IQD), which has 17.6% identity, so this structure was used to generate the model (Figure 2.13A). Figure 2.13B shows the final pairwise alignment of VC1 with the template, which reveals that many of the residues involved in interdomain packing are not conserved. Hence, the intimate details of the interdomain packing and orientation in the model are viewed as an approximation. On the other hand, the global features of the model have a very high likelihood of being accurate. The high quality of the VC1 model is reflected in an analysis with PROCHECK, e.g. 87% of residues are in most favorable conformations, 11% in the additionally allowed region, and 2% in the generously allowed region. Similar high quality results were obtained for the C2 domain, which was modeled from human APEP-1 and is presented in Chapter V (PDB 1U2H, Figure 5.1).

The VC1 homology model predicts globular structure for residues Asn25-Thr222. Thus, in the VC1 construct (Ala23-Glu243) used in our experimental studies, most of the

C-terminal 21 residues are predicted to be unstructured because they span the C1-C2 linker and into the C2 domain. In the model, the linker between V and C1 is predicted to be only three residues in length, Gln119-Pro121, since Tyr118 forms backbone hydrogen bonds to a β-strand in the V domain and Gly122 forms backbone hydrogen bonds to a β-strand in the C1 domain (Figure 2.13A). A critical aspect of the model is that these three linker residues are predicted to contribute to interdomain packing along with several residues from the V domain (Thr27, Glu94, Arg114, Arg116, Val117, Tyr118) and the C1 domain (Tyr150, Ala152, Arg178, Arg179, Thr183, Gly184, Phe186). The solvent accessible surface of V and C1 that is buried in the VC1 model is ~500 $Å^2$ in each subunit. Although the relatively small size of this interface suggests that the intrinsic affinity of the two domains is relatively weak, the interaction between the domains will be enforced by the short length and structural integration of the linker. This interpretation is supported by the lack of chemical shift perturbations in NMR experiments acquired for mixtures of isolated V and C1 domains. Thus, the model is fully consistent with the observation that V and C1 have a structural interface and tumble as an integrated structural unit.

A



B

```
VC1    25   30        40        50        60        70        80
--------NITARIGEPLVLKCKGAPKKPPQRLEWKLNTGRTEAWKVLSPQGGGPWDSVARVLPNG-----
1iqd.A
IALTQSPGTLSLSPGERATLSCRASQSFSSSYLAWYQQKPGQAPRLLIYGASTRATGIPDRFSGSGSGTDF

VC1   90       100       110       120          130       140       150
SLFLPAVGIQDEGIFRCQAMNRNGKETKSNYRVRVYQIPGKPEIV---DSASELTAGVPNKVGTCVSEGSYP-
1iqd.A
TLTISRLEPEDFAVYYCQKYGTSAITFGQGTRLEIKGTVAAPSVFIFPPSDEQLKSGTA-S-VVCLLNNFYPR

VC1   160       170       180       190       200       210       220
AGTLSWHLDGKPLVPNEKGVSVKEQTRRHPETGLFTLQSELMVTPARGGDPRPTFSCSFSPGLPRHRALRT---
1iqd.A
EAKVQWKVDNAL----QSGNSQESVTEQDSKDSTYSLSSTLTLSKADYE-KHKVYACEVTHQGL---VTKSFNR
```

Figure 2.13: Homology modeling of VC1. (A) Ribbon diagram of the homology model of VC1 with the V domain (25-118) shown in red, the linker (119-121) in black, and the C1 domain (122-222) in blue. Side chains for the linker residues are shown as black wires. Side chains for V and C1 involved in interdomain packing are shown in space filling representation in the respective colors for the domains. The side chain of Trp51 is shown as black wire. The side chains for the two cysteine residues in the V domain are shown in ball and stick colored according to atom (grey, carbon; yellow, sulfur). (B) Pairwise alignment of VC1 with the A chain of PDB molecule 1IQD. Residues in the linker region are in bold. Residues involved in interdomain packing are highlighted in the respective colors used in (A) (V, red; C1, blue; linker, black).

Summary

As described in Chapter I, RAGE is involved in numerous aspects of biology and pathology ranging from neural growth and development to tumor growth and metastasis. After establishing an *E. coli* expression system that takes advantage of OrigamiB cells that have a cytoplasm with an oxidizing environment and allows disulfide bond formation during protein expression, milligram quantities of sRAGE and domain fragments could be produced for subsequent biochemical, biophysical, and structural characterization.

A series of biochemical analyses including limited proteolysis and characterization of thermal denaturation were used to characterize sRAGE. Limited proteolysis initially suggested that the highest degree of flexibility was between C1 and C2. Additionally, powerful observations were made that V and C1 were much more stable to digestion when covalently attached. Differential scanning calorimetry further suggested structural independence since VC1 and C2 appear to experience separate unfolding transitions. These observations led to the use of structural biology to further characterize sRAGE.

First, qualitative analysis of NMR spectra was used characterize the quaternary structure of sRAGE. These studies revealed that the C1-C2 linker has a high degree of flexibility. The structural independence of the VC1 region from the C2 domain was clearly evident from the similarity of C2 chemical shifts observed for all sRAGE constructs containing C2. In contrast, the V and C1 domains have a discrete structural interface as reflected by chemical shift differences between VC1 and the isolated V and C1 domains. Second, a computational homology model of VC1 provided insight into the details of the interdomain interface explaining the observation that they behave as an

integrated structural unit. The implications of these results are discussed with respect to mechanisms of cell signaling in Chapter V. Finally, this work provided novel insight into sRAGE quaternary structure and defined the proper constructs needed for high resolution structure determination of VC1, which is analyzed in the next chapter.

CHAPTER III

HIGH RESOLUTION CRYSTAL STRUCTURE OF VC1

Introduction

RAGE is a member of the immunoglobulin (Ig) superfamily containing three predicted Ig-like domains in its large extracellular ligand binding region (1). In Chapter II, I described the first biochemical and structural characterization of RAGE which revealed that the V and C1 domains tumble as an integrated structural unit with the C2 domain flexibly attached. A homology model of VC1 revealed a significant structural interface between V and C1. However, due to lack of sequence homology to the template structure at the interface I could not conclude that either the interdomain angle or the specific side chain packing between V and C1 was accurate.

From the results reported in Chapter II, the linker between C1 and C2 was estimated to be at least 10 residues. High sequence homology (except for the bovine protein which has a large insert that could increase flexibility) suggests the structural flexibility of the C1-C2 linker is highly conserved across species (Figure 2.1). Since the accuracy of a model is dependent on homology to the template, important details can be different once the structure is determined. For example, the V-C1 interdomain angle and the precise length of the C1-C2 linker could be critical in autoinhibition and activation of RAGE. In order to better understand the basis for RAGE signaling, Dr. Guenter Fritz used constructs designed from the biochemical analysis to determine the x-ray crystal structure of the tandem VC1 construct to 1.8 Å (statistics in Appendix B). This high

resolution structure allowed a detailed analysis of the V-C1 interface and sets the stage for understanding ligand-induced signal transduction and the structural determinants of interdomain flexibility, both of which are presented in this chapter.

Crystallization and Structure Determination

The VC1 construct and purification protocol were generated as described in Chapter VI. Initial protein samples were prepared at Vanderbilt University, but eventually the expression vector was sent to Germany along with a detailed protocol. Purified VC1 was prepared at a concentration of 10 mg/ml and buffer exchanged into 10 mM sodium acetate at pH 4.5. Crystals were grown at 298 K by mixing 2 μl of the protein solution with 2 μl of 0.1 M sodium cacodylate at pH 6.5, 0.2 M zinc (II) acetate, 11% PEG 8000. Crystals grew to a size of 200 μm x 100 μm x 200 μm after 10-20 days.

For cryoprotection, the crystals were soaked for 1 min in three consecutive steps in mother liquor progressively containing 4, 8 and 12% glycerol and flash frozen in liquid nitrogen. Native crystals diffracted to a resolution of 1.8 Å. As zinc (II) turned out to be essential for crystallization it was assumed that zinc (II) specifically interacts with VC1 and therefore could be used for experimental phasing. Multiple anomalous dispersion data were collected at the Swiss Light Source (Villigen, CH) beamline X06SA. The data sets were reduced and scaled with the XDS package (113). Crystals belonged to the space group $P2_12_12$ with unit cell parameters of 74.63 119.68 28.82 as determined by XPREP (Bruker-AXS, 2005). The zinc substructure was solved using SHELXD (114). Refinement of the zinc sites and phase calculation was done with SHARP (115) followed by solvent flattening and initial fragment building using RESOLVE (116). The final

model was built manually in COOT (117) followed by several cycles of restraint and TLS refinement as implemented in REFMAC5 (118, 119). All stages of crystallization and structure determination were performed by Dr. Guenter Fritz at the University of Konstanz in Germany. I performed the analysis of the structure described in the following sections.

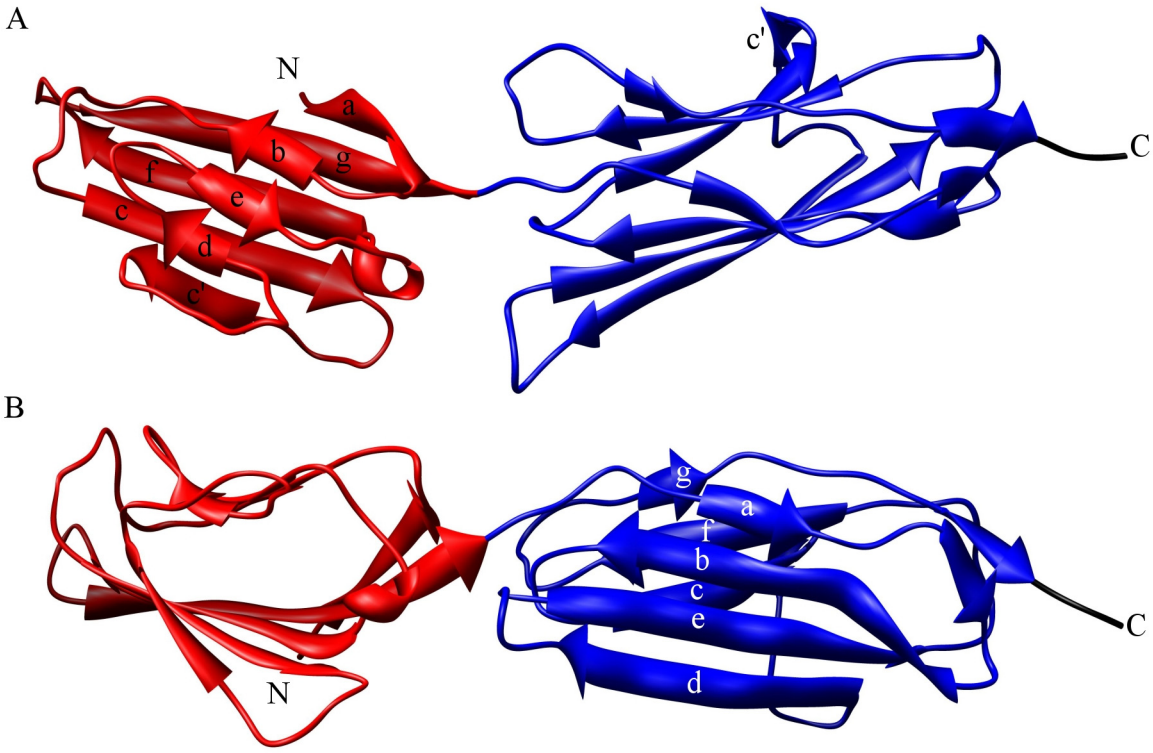


Figure 3.1: The x-ray crystal structure of VC1. Ribbon diagram of VC1 colored according to figures in Chapter II. The V domain (23-118) is in red and the C1 domain (119-233) is in blue. A portion of the C-terminal extension (231-233) is shown in black. Panel (B) is a 90 degree rotation of panel (A) towards the top of the page. The individual strands are labeled for the V domain (A) and C1 domain (B) according to topology shown in Figure 3.2. The *c′* strand of the C1 domain is labeled in panel (A).

Structural Characteristics of VC1

The structure of VC1 is composed of two integrated Ig-like domains (Figure 3.1). The 2D topology diagrams for each domain are shown in Figure 3.2. The N-terminal domain has characteristic V-type topology exemplified by sheet swapping of the first β-strand (strand *a*) forming a parallel β-sheet with the last β-strand (strand *g*; Figure 3.1A). Strands *a*, *c*, *c*', *f*, and *g* form one sheet while strands *b*, *d*, and *e* form the second sheet. Another characteristic of V-type Ig domains is additional residues between strands *c* and *d* labeled *c*' and *c*". The *c*' strand clearly extends the anti-parallel β-sheet while *c*" functions merely as a long loop leading into strand *d*. The lack of the *c*" strand sometimes classifies the Ig domain as a C/V hybrid; however, strand *a* swapping is unique amongst V-type Ig domains and confirms this assignment (120).

The C-terminal domain forms an apparent H-type fold with strands *a, b*, *c*', *d*, and *e* forming one sheet and strands *c*, *f*, and *g* forming the second sheet of the β-sandwich (120). Strand *c*' is very short only containing residues 162-163. As anticipated, disulfide bonds are present between residues Cys38 and Cys99 of the V domain and Cys144 and Cys208 of the C1 domain. While C1 should be classified as an H-type Ig domain, the nomenclature 'C1' will continue to be used for consistency with previous chapters.

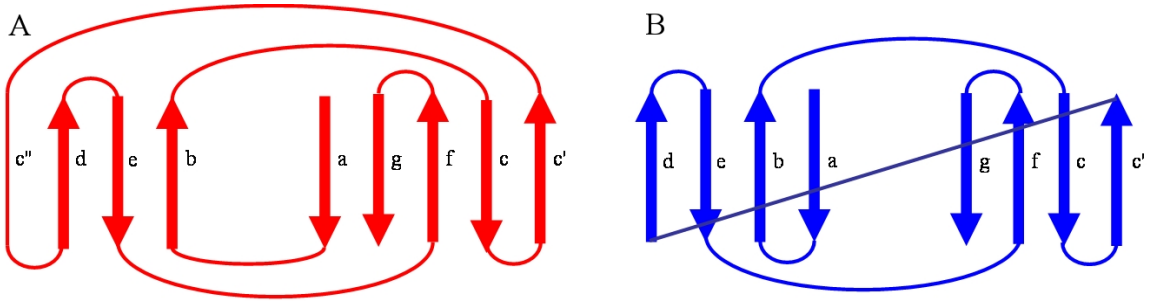


Figure 3.2: Topology diagram of the x-ray crystal structure of VC1. (A) Topology diagram of the V domain in VC1. Each strand is denoted as an arrow with the strand label on the right where strand *a* and strand *g* are the N- and C-termini respectively. (B) Topology diagram of the C1 domain in VC1. Strands depicted and labeled as in (A).

Overall the topology and hydrogen bonding pattern of both domains is quite standard, although two deviations stood out. Strand *g* of the C1 domain contains residues Leu220 through Gln226, however, there is a bulge at Ala223 and Pro224 (Figure 3.1B, directly below the strand *g* label). The typical backbone hydrogen bond between the carbonyl oxygen of Ala223 and the amide proton of Cys208 on the adjacent β-strand is absent and is replaced by hydrogen bonds between the Thr222 hydroxyl side chain and backbone atoms from Ala223 and Cys208 (Figure 3.3). The residues of strands *f* and *g* in C1, which includes the bulge, are 100% conserved across species except for Arg221 (Appendix A).

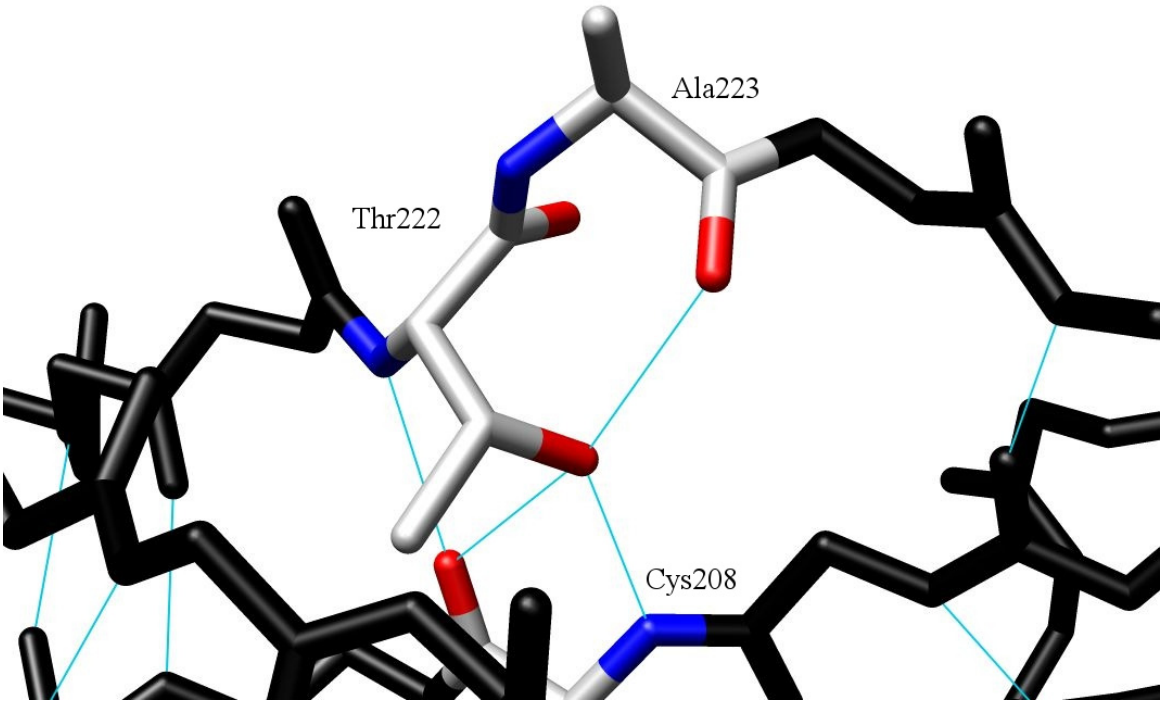


Figure 3.3: Side chain hydroxyl stabilizes a bulge in a C1 β-strand. Backbone and side chain heavy atoms are shown for the region of the C1 domain containing a bulge in hydrogen bonding pattern. Thr222, Ala223, and Cys208 are colored according to atom type (C, grey; N, blue; O, red) with the rest of the atoms and bonds in black. Potential hydrogen bonds are shown as thin cyan lines. The side chain of Thr222 is in proximity to form hydrogen bonds most likely with backbone oxygen and nitrogen atoms of Ala223 and Cys208 respectively.

A second interesting feature is present at the C-terminus of C1. Residues 228 through 231 of strand *g* swap and form a short parallel β-sheet with strand *a* near the $C1_{A-B}$ loop. This causes a break in β-sheet hydrogen bonding and an exaggerated rolling of the sheet at the N-terminus of strand *b* (Figure 3.4). In conjunction with this break, the β-sheet hydrogen bonding between strands *f* and *g* is broken at Pro227. The proline is replaced by leucine in both the bovine and mouse sequences, in which the amide nitrogen of Leu227 would be in register to hydrogen bond to the carbonyl oxygen of Pro204 on strand *f*. One anticipates a different conformation in these proteins. Other forces such as side chain packing could drive the formation of this short parallel β-sheet even in the bovine and mouse proteins; computational analysis or structure determination would be required to resolve this question.

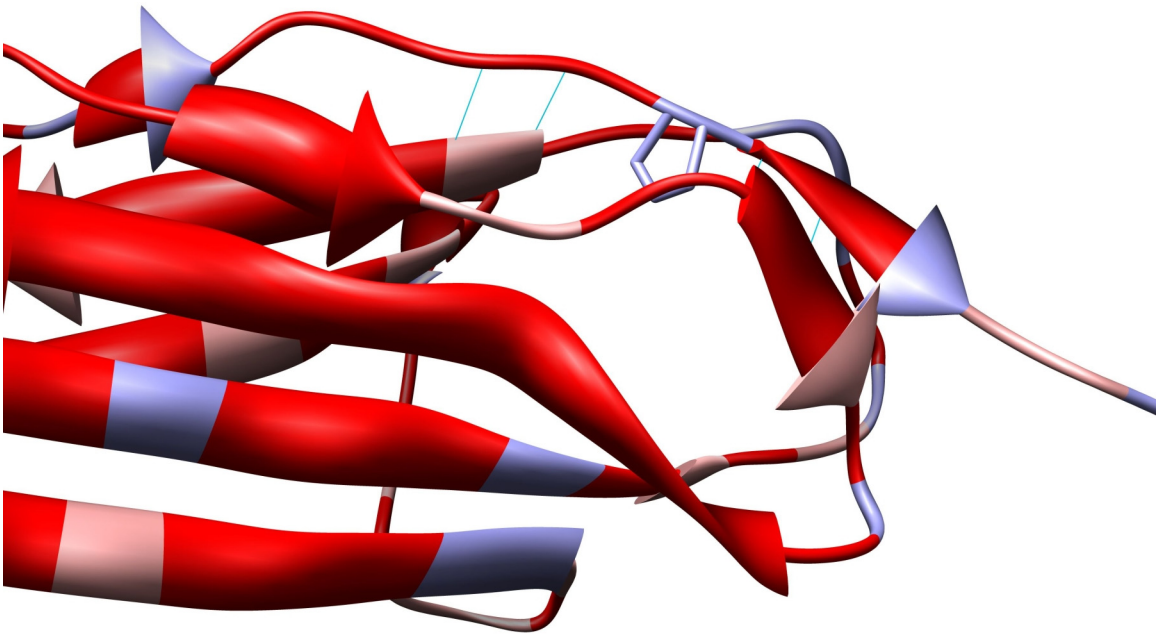


Figure 3.4: Strand swap at the C-terminus of the C1 domain. Expanded view of the C-terminal region of the C1 domain in the orientation shown in Figure 3.1B. Ribbon diagram is colored according to average % conservation from unconserved (dark blue, none shown) to fully conserved (red, majority of structure). Potential hydrogen bonds are shown as thin cyan lines. The side chain for Pro227 (conserved in 2 of 4 species, Appendix A) is depicted in stick representation.

Comparing the individual domains provided significant insight into the accuracy of the homology model presented in Chapter II. The orientation and topology of β-strands of the V domains in the homology model and x-ray crystal structure are similar (Figure 3.5A). However, the two structures have a RMSD from the mean structure of 2.2 Å when aligned according to the β-strands. The most dramatic differences are in the loops opposite to the V-C1 interface. For example, the $V_{B\text{-}C}$ loop in the crystal structure is flipped out of the plane of the β-sandwich. Consequently, the $V_{F\text{-}G}$ loop in the homology model is flipped into the β-sandwich plane to maintain side chain packing between the loops.

The orientation and topology of β-strands of the C1 domains from the homology model and x-ray crystal structure also appear to be similar (Figure 3.5B). However, the two structures have a RMSD from the mean structure of 2.9 Å when aligned according to the β-strands, slightly higher than the alignment of the V domains. Like the V domains, the most dramatic differences are in loop regions at the ends of the β-sandwich opposite to the interdomain interface (Figure 3.5B, circled). Specifically, significant differences are observed in the $C1_{A\text{-}B}$ loop (Figure 3.5B). The loop no longer extends beyond the β-sandwich due to introduction of approximately 1.5 turns of α-helix. The $C1_{E\text{-}F}$ loop likewise adopts a different orientation since side chains pack between the $C1_{A\text{-}B}$ and $C1_{E\text{-}F}$ loops.

In summary, the x-ray crystal structure of VC1 is consistent with the initial sequence analysis and homology modeling. An N-terminal V-type Ig domain and a C-terminal H-type Ig domain are found. The details of the experimental structure suggest a short sequence of β-strand (*c*') between strands *c* and *e* which is more consistent with an

H-type Ig fold as opposed to the originally assigned C-type Ig fold. A more detailed analysis of the x-ray crystal structure will be made in the next section to determine if the hypotheses generated in Chapter II about sRAGE quaternary structure are supported.

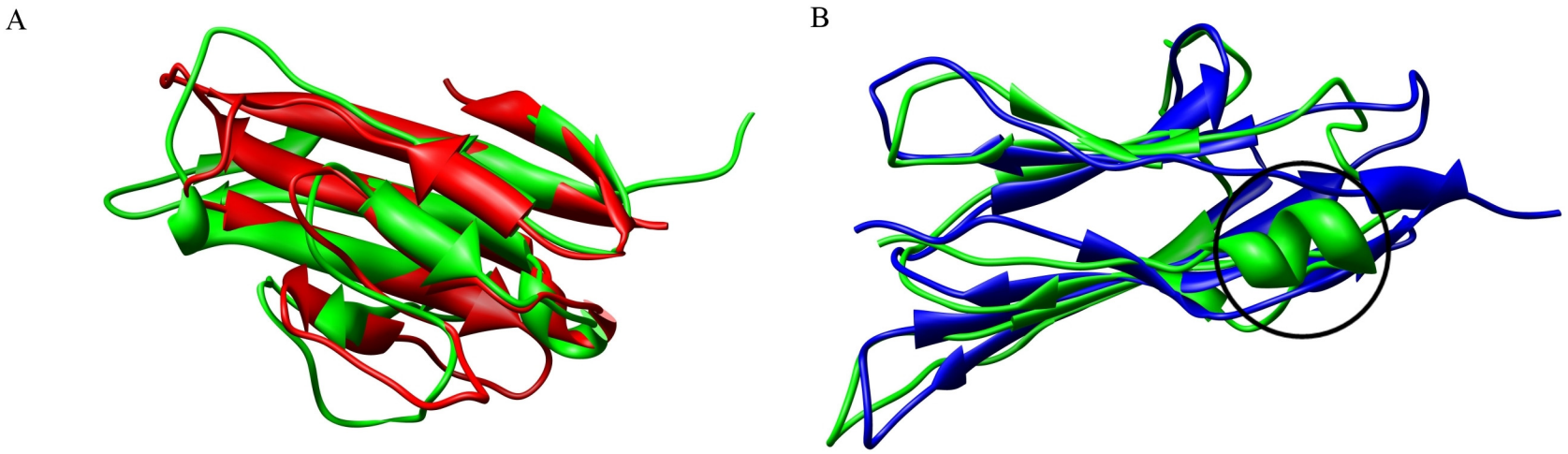


Figure 3.5: Comparison of the homology model and x-ray crystal structure of VC1. (A) Overlay of the V domain from the homology model (green) and the x-ray crystal structure (red). The orientation is the same as in Figure 3.1A. (B) Overlay of the C1 domain from the homology model (green) and the x-ray crystal structure (blue). The orientation is the same as in Figure 3.1A. Overlays were made by superimposing backbone atoms and minimizing the RMSD using Chimera (39).

The V-C1 interface

In Chapter II, a series of biochemical and modeling data were used to generate hypotheses about sRAGE quaternary structure. The critical observation was that V and C1 tumble as an integrated structural unit while C2 has significant structural flexibility from VC1. Aligning the VC1 tandem domain structures through the C1 domain shows there is a striking difference in the V-C1 interdomain angle between the model and the experimental structure (Figure 3.6). As described above, the overall topology and orientation of the loops at the V-C1 interface are quite similar. Thus, the subtle details of the interface likely contribute to the observed difference in interdomain angle.

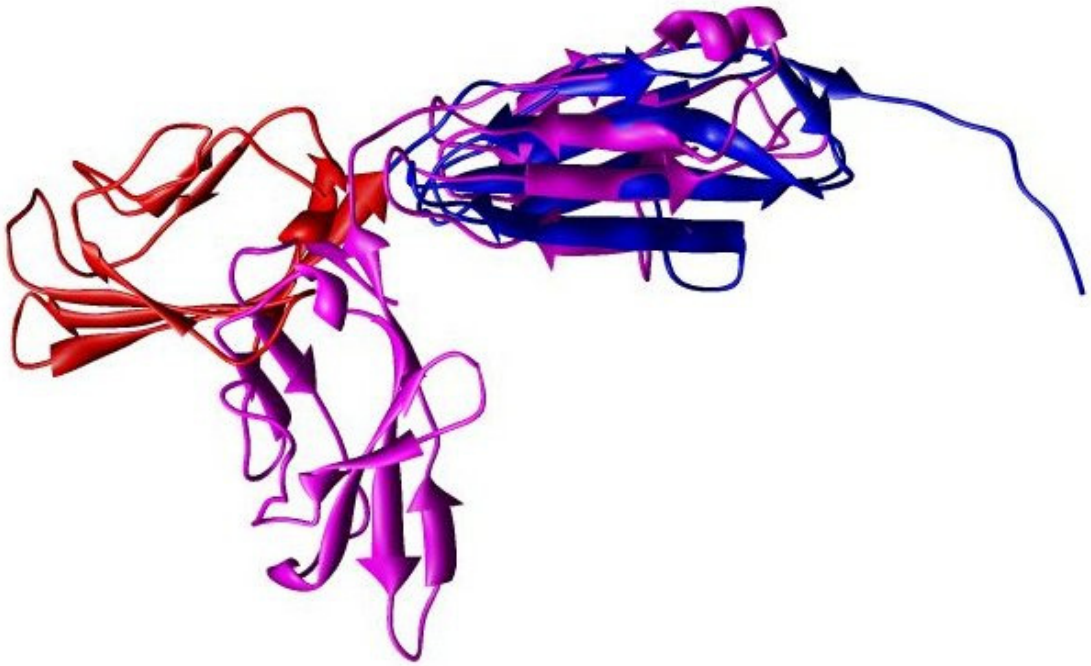


Figure 3.6: The V-C1 interdomain angle is different in the x-ray crystal structure. Overlay of ribbon diagrams of the x-ray crystal structure (colored as in Figure 3.1) and the VC1 homology model (colored in violet). The two models are overlaid using to the C1 domains. Superpositions were made by minimizing the RMSD in Chimera (39).

Analysis of the VC1 homology model suggested numerous side chains from each domain pack against each other creating the interface that fixes the orientation of V and C1 (Figure 2.12). The V-C1 interface in the x-ray crystal structure shows neither the three amino acid linker nor the same dense packing of side chains as seen in the homology model. Specifically, the buried surface area is 300 $Å^2$ in the x-ray crystal structure compared to 500 $Å^2$ in the homology model. Additionally, both Tyr118 and Gln119 show β-sheet hydrogen bonding patterns within the V and C1 domains respectively (Figure 3.7A). Therefore, there is no linker between V and C1 in the x-ray crystal structure.

Despite the absence of intervening residues, it is important to characterize the interface between V and C1. Two interdomain contacts were identified by selecting a zone of 3.5 Å from all atoms in the V and C1 domains. The first is packing of Pro215 side chain atoms from the $C1_{F\text{-}G}$ loop with the side chain of Tyr118 (Figure 3.7B). The second is a potential hydrogen bond between a carboxylate oxygen atom of Glu94 in the V domain and the hydroxyl hydrogen on Tyr150 in the C1 domain (Figure 3.7B). Notably, Glu94, Tyr118, Tyr150, and Pro215 are absolutely conserved across species supporting their functional role in aligning V and C1 in a relatively fixed orientation (Appendix A). These interdomain contacts in combination with the steric exclusion of conformational space cause the two domains to tumble together in solution.

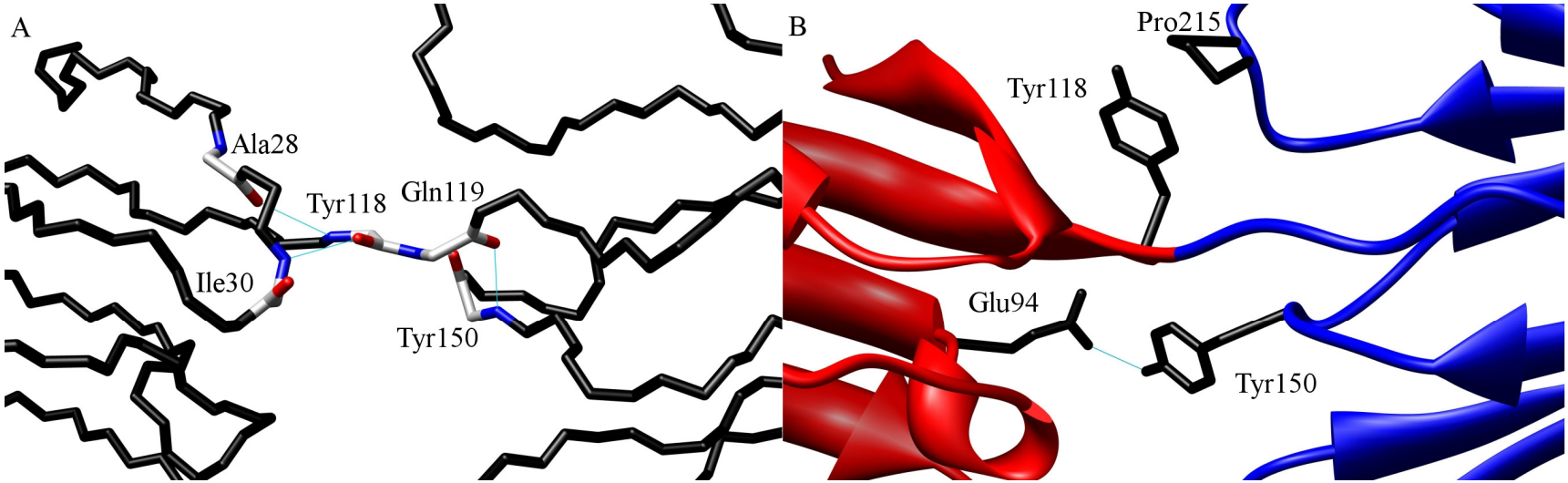


Figure 3.7: Interdomain contacts in the VC1 x-ray crystal structure. (A) Stick representation of backbone atoms. Ala28, Ile30, Tyr118, Gln119, and Tyr150 are colored according to atom type (C, grey; N, blue; O, red). Potential hydrogen bonds are shown as thin cyan lines. (B) Ribbon diagram of VC1 showing the V domain colored in red and the C1 domain colored in blue. Side chain atoms from the V domain (Glu94 and Trp118) and the C1 domain (Trp150 and Pro215) are colored as black sticks. The potential hydrogen bond between a carboxylate oxygen of Glu94 and the hydroxyl hydrogen of Try150 is shown as a thin cyan line.

Summary

Determining the high resolution x-ray crystal structure of VC1 provided useful insight into the function of RAGE as a signaling molecule. Aside from minor differences in the topology of C1, analysis of the structure confirmed hypotheses generated in Chapter II from homology modeling about the quaternary structure of sRAGE. Clearly, there is an interface (albeit small) between the V and C1 domains that enforces the structural integration of the two domains. This interface is stabilized by packing and hydrogen bonding of side chain atoms (Figure 3.7B). The lack of intervening residues between the two domains also serves to inhibit relative mobility through steric limits on the relative positioning of the two domains. Our ability to optimize construct lengths based on accurate knowledge of the structure could dramatically aid in future biochemical and structural studies. In addition, the structural studies presented in this chapter and in Chapter II set the stage to study RAGE-ligand interactions, which are presented in the next chapter. The implications and broader conclusions towards receptor signaling drawn from the knowledge of VC1 behaving as an integrated structural unit are presented in Chapter V.

CHAPTER IV

LIGAND BINDING TO SRAGE

Introduction

S100s are EF-hand containing Ca^{2+}-binding proteins known to have both intracellular and extracellular functions (121). As described in Chapter I, S100 proteins typically bind to short α-helical peptide segments of their target protein. RAGE is a member of the immunoglobulin (Ig) superfamily containing three predicted Ig-like domains in its N-terminal ligand binding region (1). Since both the x-ray crystal structure of VC1 described in Chapter III and sequence analysis of C2 suggest all β-sheet secondary structure, the recognition of RAGE by S100 proteins will either use a novel mechanism for binding or induce significant structural changes in RAGE.

N(epsilon)-(carboxymethyl)lysine (CML) is a well characterized advanced glycation end product (AGE) and a specific, high affinity (K_D ~76 nM) ligand for RAGE (122). CML is a small AGE moiety characterized by a charge reversal on lysine side chains suggesting electrostatics play an important role in recognition by RAGE (Figure 1.2). CML represents approximately 30% of AGEs in diabetic blood plasma (123). Therefore, knowledge of RAGE recognition of CML could provide structural information crucial to designing antagonists to lessen the complications of diabetes.

While affinities have been estimated for many RAGE ligands, little information is available on which Ig domains are used in binding. Only the V domain has been implicated in direct binding to CML (122). While S100A12 has been shown to bind

RAGE with low nM affinity (43), similar measurements have not been made for S100B. In this chapter, the limited proteolysis experiments described in Chapter II are extended to gain insight into the domain specificity of S100B binding to sRAGE. S100B was chosen due to the large amount of structural information available including complete backbone assignments from numerous NMR structures in both apo- and Ca^{2+}-loaded states in addition to multiple ligand-bound states, e.g. (124). This information allowed preliminary characterization of S100B complexes using NMR. Binding affinities were estimated between S100B and sRAGE using surface plasmon resonance. SPR experiments were performed by Estelle Leclerc and Dr. Guenter Fritz. Lastly, chemical shift perturbation experiments were performed in attempt to map the interaction between CML and sRAGE. The summary of these data are combined with the structural characterization to generate hypotheses about multiligand receptor activation.

S100B Binding Induces Structural Change in sRAGE

Having established limited proteolysis conditions for sRAGE and the six domain constructs, proteolysis experiments could be used to characterize the interaction of S100B with RAGE. Initial studies with wild-type S100B led to disulfide cross-linking between S100B and the V domain. While the S100B-sRAGE disulfide cross-linking could be reversed by addition of DTT, the reducing agent also caused the cleavage of the disulfide bond essential to maintain the structure of V (Figure 2.2). Consequently, a series of three Cys→Ser mutants in S100B were prepared (C68S, C84S, C68S/C84S) and purified. Analysis of the V construct in the presence of the three mutants by SDS-PAGE under non-reducing and reducing conditions confirmed that the reactive cysteine was Cys84, so

the C84S mutant was used for all further experiments (Figure 4.1). Notably, this cysteine to serine mutant was shown previously to be functional (125).

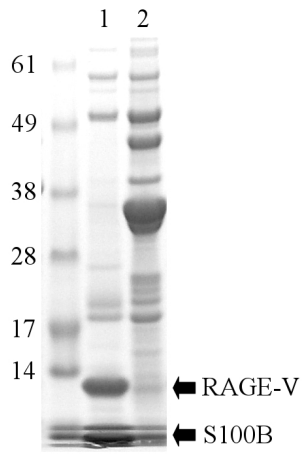


Figure 4.1: Cys84 of S100B covalently attacks the V domain disulfide bond. Non-reducing SDS-PAGE of S100B Cys→Ser mutants in the presence of V. Lane 1, S100B(C84S). Lane 2, S100B(C68S). Absence of bands at indicated arrow positions and the presence of high molecular weight aggregates indicate covalent aggregation.

Studies of isolated sRAGE showed that trypsin digestion results in complete loss of the intact protein within 4 hours (Figure 2.3A). The overall rate of digestion was similar in the presence of Ca^{2+}-S100B (Figure 4.2A), although sRAGE was slightly, yet reproducibly, better protected in the presence of S100B than in isolation. As expected, this observation was dependent on the presence of calcium. Interestingly, the 25 kDa band identified as VC1 in the digestion of free sRAGE was completely absent when Ca^{2+}-S100B was present (Figure 4.2A, see arrows). This observation was also calcium-dependent. While protection from proteolysis is more common, an *increased* rate of proteolysis is also possible when structural perturbations are induced by protein-protein interactions. The expedited degradation of the VC1 fragment in sRAGE digests could not be explained by the presence of a contaminating protease in our S100B sample since this would be inconsistent with the observation that degradation of the intact protein was slowed. Thus, the interaction between Ca^{2+}-S100B and sRAGE appears to cause a conformational change in the VC1 region of sRAGE that enhances proteolysis by trypsin.

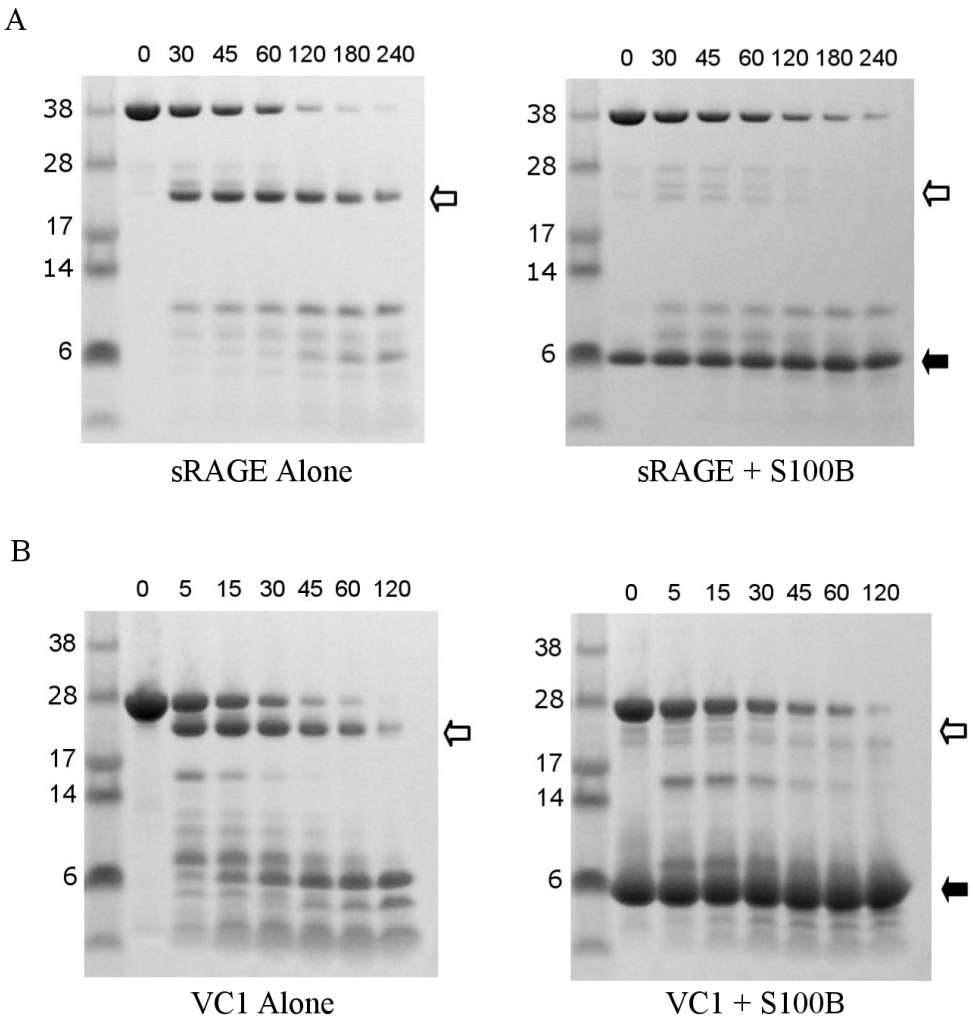


Figure 4.2: S100B binding causes a conformational change in sRAGE. Protease protection of sRAGE by Ca^{2+}-S100B. Reducing SDS-PAGE of sRAGE (A) and VC1 (B) incubated with trypsin in the absence (left) or presence (right) of Ca^{2+}-S100B. Solution conditions were the same in the absence and presence of S100 protein and are described in Chapter VI. Filled arrows mark the S100B band. Open arrows mark the stable 25 kDa VC1 fragment most affected by S100B binding. Time points are given in minutes.

To corroborate these findings, the experiments were repeated with the various sRAGE domain constructs. Much like intact sRAGE, isolated VC1 was reproducibly protected in the presence of Ca^{2+}-S100B, but the 25 kDa fragment was very rapidly degraded (Figure 4.2B). The isolated V domain was also more rapidly degraded in the presence of Ca^{2+}-S100B consistent with the increased rate of proteolysis on VC1 (Figure 4.3). Corresponding experiments with C1C2 and isolated C1 and C2 revealed no interaction with Ca^{2+}-S100B as the C1 domain continues to be degraded extremely rapidly under all conditions, and the C2 domain remains totally resistant to digestion by trypsin under these experimental conditions. Together, these limited proteolysis results indicate that the interaction of S100B with RAGE is calcium-dependent and localized primarily to the V domain.

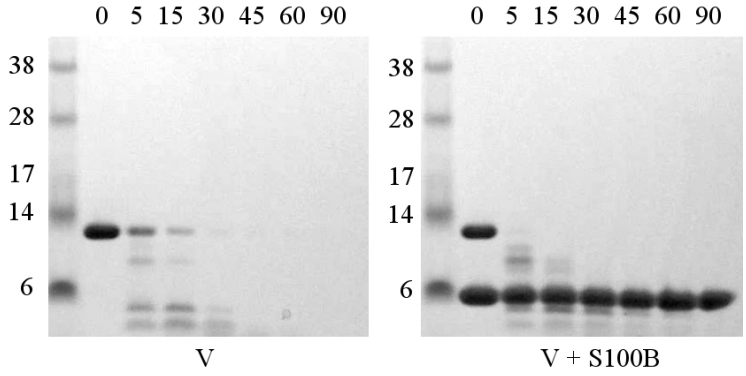


Figure 4.3: S100B binding is localized to the V domain. Protease protection of RAGE-V by Ca^{2+}-S100B. Reducing SDS-PAGE of V domain incubated with trypsin in the absence (left) or presence (right) of Ca^{2+}-S100B. The band migrating at ~ 6 kDa is a single S100B polypeptide. Solution conditions were the same in the absence and presence of S100 protein. Time points are given in minutes.

NMR Characterization of S100B Binding to sRAGE

S100B-sRAGE binding was further investigated using NMR spectroscopy. As mentioned in Chapter II, the NMR chemical shift is exquisitely sensitive to the local chemical environment and can be used to monitor structural changes induced by ligand binding. The optimal conditions for acquiring NMR spectra for sRAGE required the use of high ionic strength buffer (>250 mM NaCl). However, the complex could be formed successfully only at lower ionic strength (<100 mM NaCl), conditions under which, remarkably, free sRAGE had only very limited solubility. The fact that binding of S100B solubilized sRAGE is, in and of itself, confirmatory evidence of an interaction between S100B and sRAGE. The ^{15}N-^{1}H HSQC and TROSY-HSQC spectra of ^{2}H,^{15}N-sRAGE in the presence of excess Ca^{2+}-S100B showed only peaks arising from C2 (Figure 4.4). The absence of signals from VC1 is interpreted as a result of binding of S100B to one or both domains, since it is expected that a large molecular complex is formed containing an S100B dimer and presumably two sRAGE molecules, although no conclusive experimental evidence is currently available. The observation of C2 signals, completely unperturbed, indicates that C2 retains its structural independence from the rest of the complex.

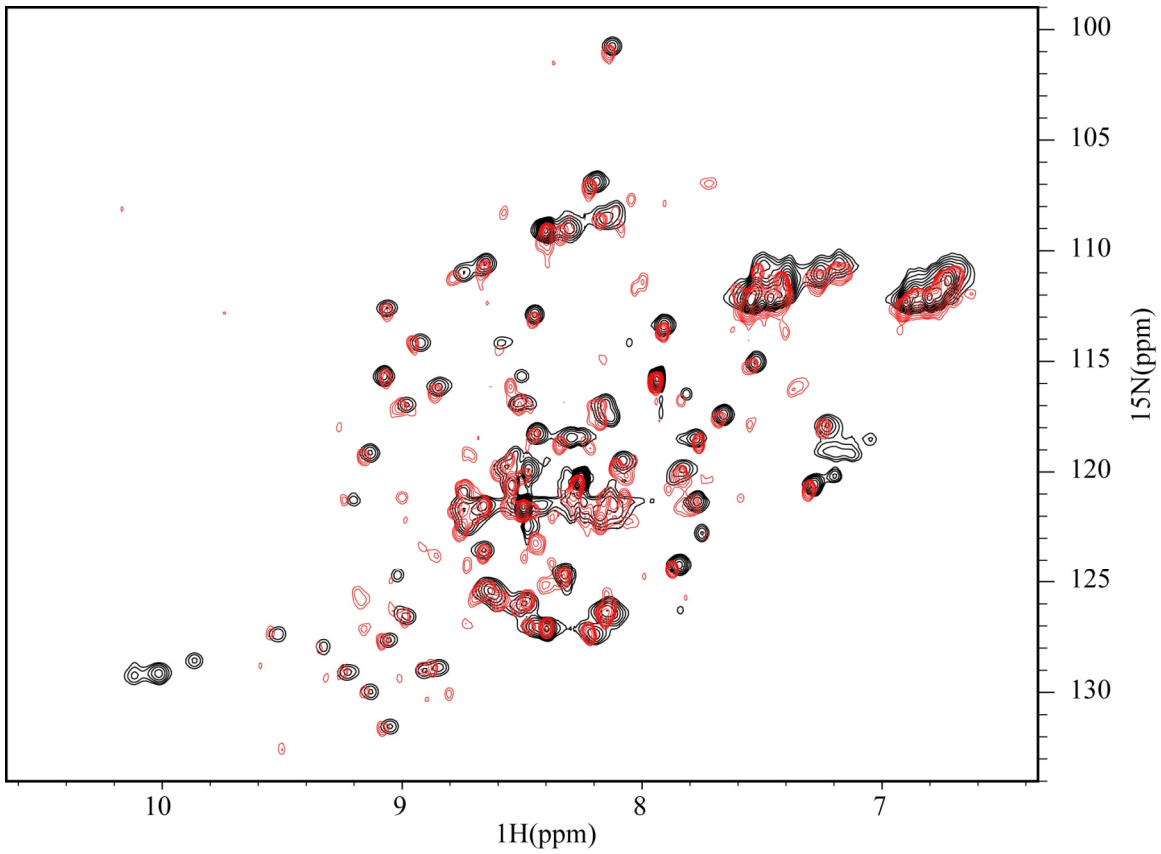


Figure 4.4: C2 retains its structural independence when in complex with Ca^{2+}-S100B. ^{15}N-^{1}H HSQC spectra are overlaid for either protonated (black) or perdeuterated (red) sRAGE in the presence of S100B. Some of the red contours that do not have corresponding black contours are excessive noise due to the extremely low protein concentrations used in that sample (~ 40 μM).

To complement the study of S100B binding to sRAGE, NMR was used to probe the interaction with the V and VC1 constructs. In titrations of ^{15}N-enriched V and VC1 with Ca^{2+}-S100B, a uniform decrease in signal intensity was observed for all except ten of the crosspeaks in VC1 as S100B was added (Figure 4.5). The few VC1 signals that remained observable had chemical shifts and linewidths consistent with being part of the flexible, unstructured tail at the C-terminus of this construct. Control experiments on C1C2, C1, and C2 showed no changes in ^{15}N-^{1}H HSQC spectra induced by the presence of Ca^{2+}-S100B. Thus, consistent evidence was obtained for relatively strong interaction of S100B with sRAGE, and that the V domain is clearly involved.

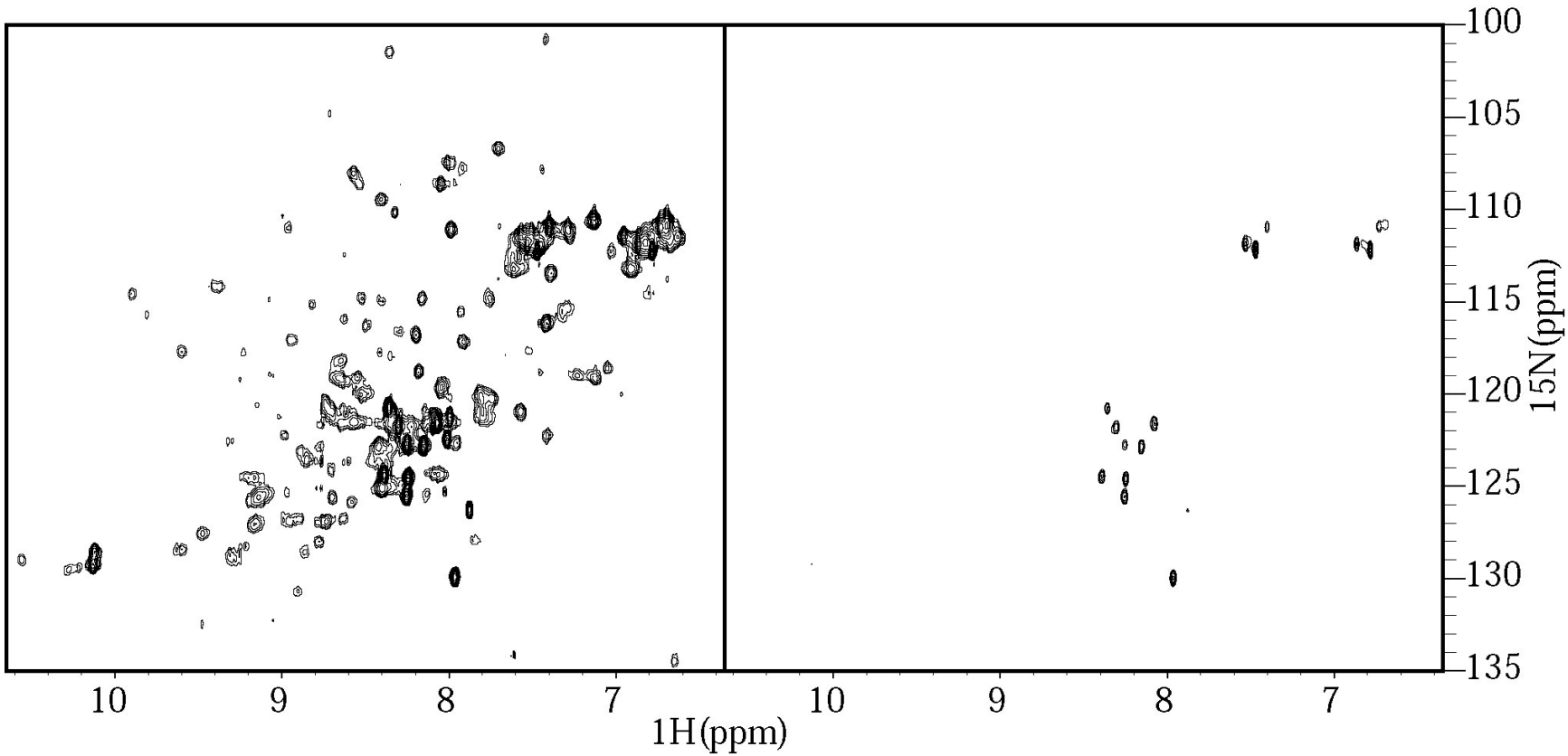


Figure 4.5: Both domains of VC1 are broadened by S100B. Shown are ^{15}N-^{1}H HSQC spectra of VC1 in the absence (left) or presence (right) of a 2:1 stoichiometric amount of VC1 to S100B dimer.

S100B Binds to RAGE with nM Affinity

A detailed analysis of surface Plasmon Resonance (SPR) measurements for S100B binding to immobilized sRAGE revealed affinity constants in the nM range (Estelle Leclerc, Guenter Fritz, Claus Heizmann, in preparation). SPR was used here to further probe the binding of S100B to sRAGE using VC1, V, and C2 constructs. In these experiments, binding was observed to VC1 and V, but no interaction with the C2 domain was detected.

Typical S100B binding curves for VC1 and V chips are shown in Figure 4.6. In the case of VC1, the data were fit with the parallel reaction model for heterogeneous ligand:

$$A + B1 \leftrightarrow AB1 \text{ and } A + B2 \leftrightarrow AB2$$

where one analyte molecule (A = S100B) can bind independently to two ligand molecules B1 or B2 (VC1) with K_{D1} the dissociation constant of the first equilibrium, and K_{D2} the corresponding parameter for the second event. This analysis gave two binding constants: K_{D1} = 11.1 nM and K_{D2} = 244 nM (Figure 4.6A). These results are largely consistent with the affinity determined for intact sRAGE. The S100B binding curves for the V domain (Figure 4.6B) were also fit with the parallel reaction model. Fitting of the binding data resulted in K_{D1} = 550 nM and K_{D2} = 470 nM. Although the quality of the fit for K_{D1} was much poorer for VC1 than for the isolated V domain, the affinity of S100B for the immobilized VC1 construct is clearly higher than that for the isolated V domain. In summary, these data indicate that S100B binds to the V domain of sRAGE and while V contains a site sufficient for binding of S100B dimers, additional contributions are made either directly or allosterically by the C1 domain.

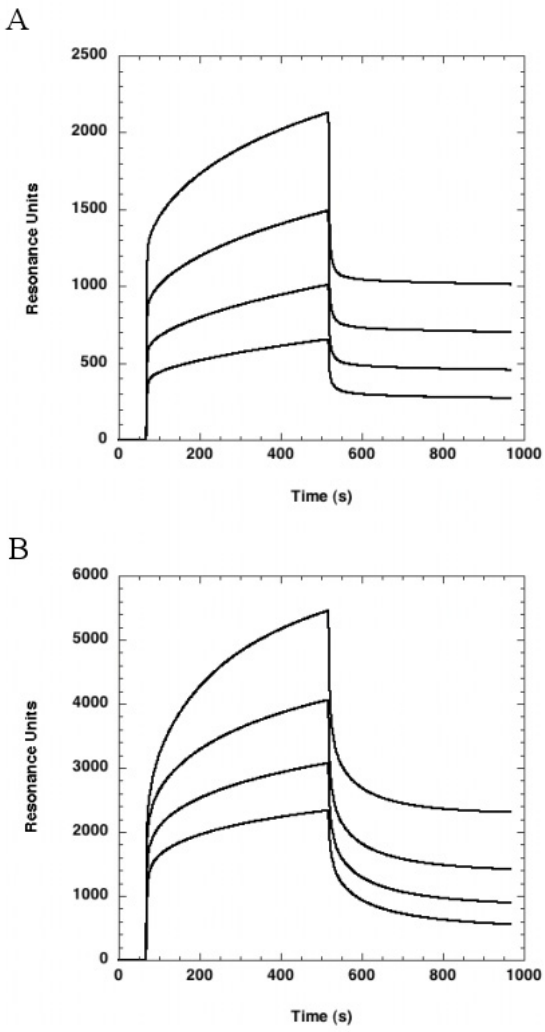


Figure 4.6: SPR binding of S100B to sRAGE domains. Representative binding sensorgrams for S100B binding to (A) immobilized VC1 and (B) immobilized V. Injection concentrations from top to bottom are 12.6 μM, 6.3 μM, 3.15 μM, and 1.57 μM for both (A) and (B).

CML Binding to RAGE

CML modified human serum albumin was first shown to be a ligand of RAGE by Kislinger and coworkers in 1999 (122). However, isolated CML was unable to compete for binding to CML-HSA. More recently, Ostendorp and coworkers showed direct binding of purified CML to recombinant sRAGE purified from yeast suggesting isolated CML could be used for chemical shift mapping studies (109). In order to determine if NMR could be used to monitor the binding of CML to sRAGE and its domain constructs, initial titrations were performed using ^{15}N-enriched V. Although reasonably strong binding has been reported, titration points up to 20 times the concentration of V domain failed to produce any observable NMR chemical shift perturbations (Figure 4.7). Several experimental conditions were tested and varied in ionic strength (0-300 mM NaCl), buffer composition (Tris-HCl, phosphate), and pH (6.0-7.5) with identical results. Repeating these CML titrations using ^{15}N-enriched VC1 provided very similar results.

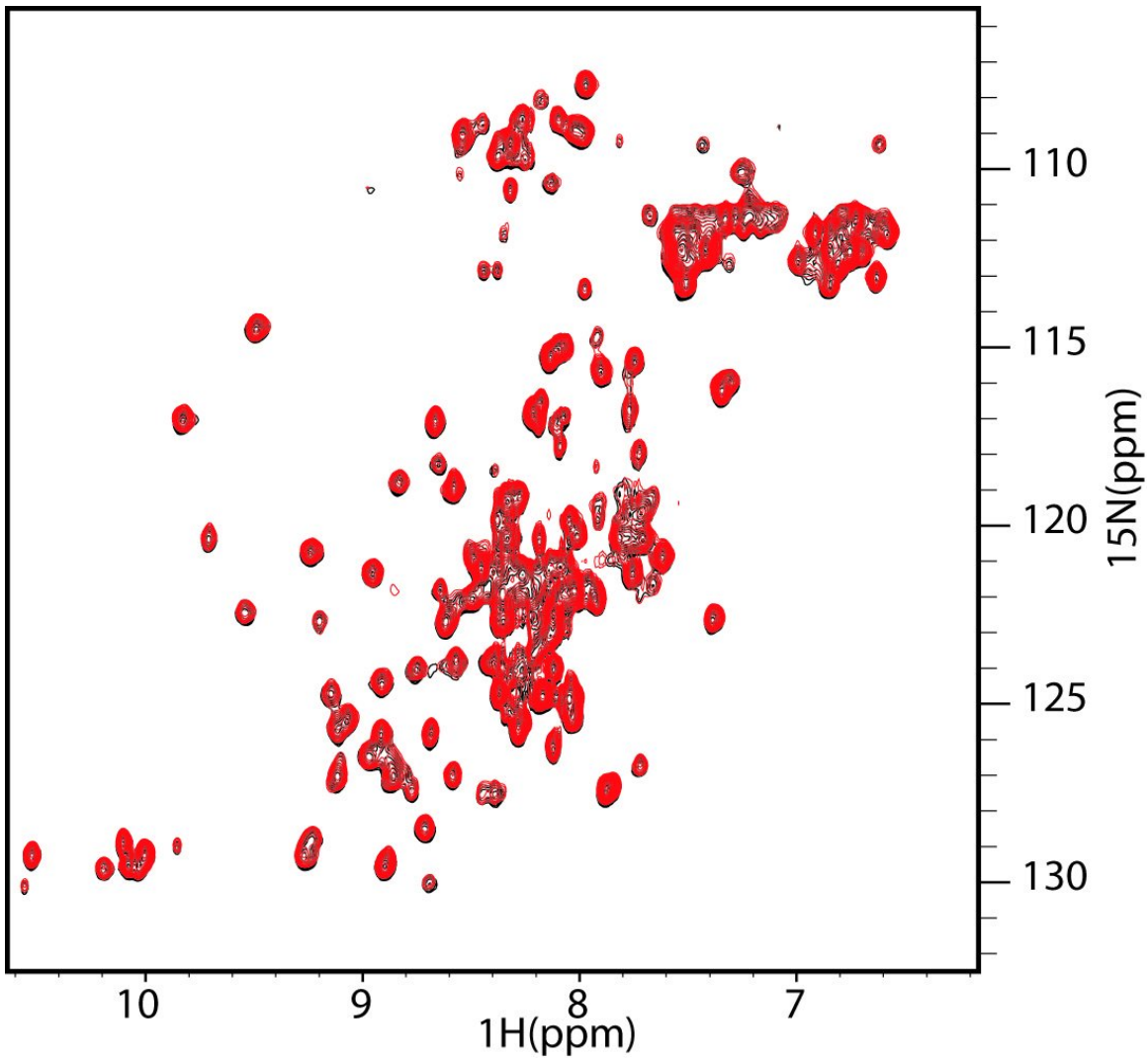


Figure 4.7: CML titration into ^{15}N-enriched V. Overlay of ^{15}N-^{1}H HSQC spectra of isolated V domain in the absence (black) and presence (red) of 20x excess CML.

A separate analysis was then performed where a ^{1}H spectrum of CML was observed as unlabeled V was titrated. This experiment would take advantage of the fact that the signals arising from CML would be much more narrow and intense than signals arising from V. However, no significant changes in the linewidth or chemical shift of CML peaks were observed (Figure 4.8). Slight intensity changes are observed on CML peaks between 1.0 and 2.0 ppm, presumably arising from the additive effect of signals from the V spectrum (Figure 4.8). Since monitoring by NMR proved unsuccessful, alternative strategies were attempted to characterize CML binding. Titrations were performed monitoring either tryptophan fluorescence or heat evolution by isothermal titration calorimetry. However, neither experiment produced a binding curve that would be consistent with the strong binding reported by Ostendorp and coworkers (109). The interpretation of these results is that CML binds dynamically to RAGE, i.e. the small ligand associates tightly with the receptor overall but interacts with multiple binding sites. An elaboration of this concept is described in Chapter V.

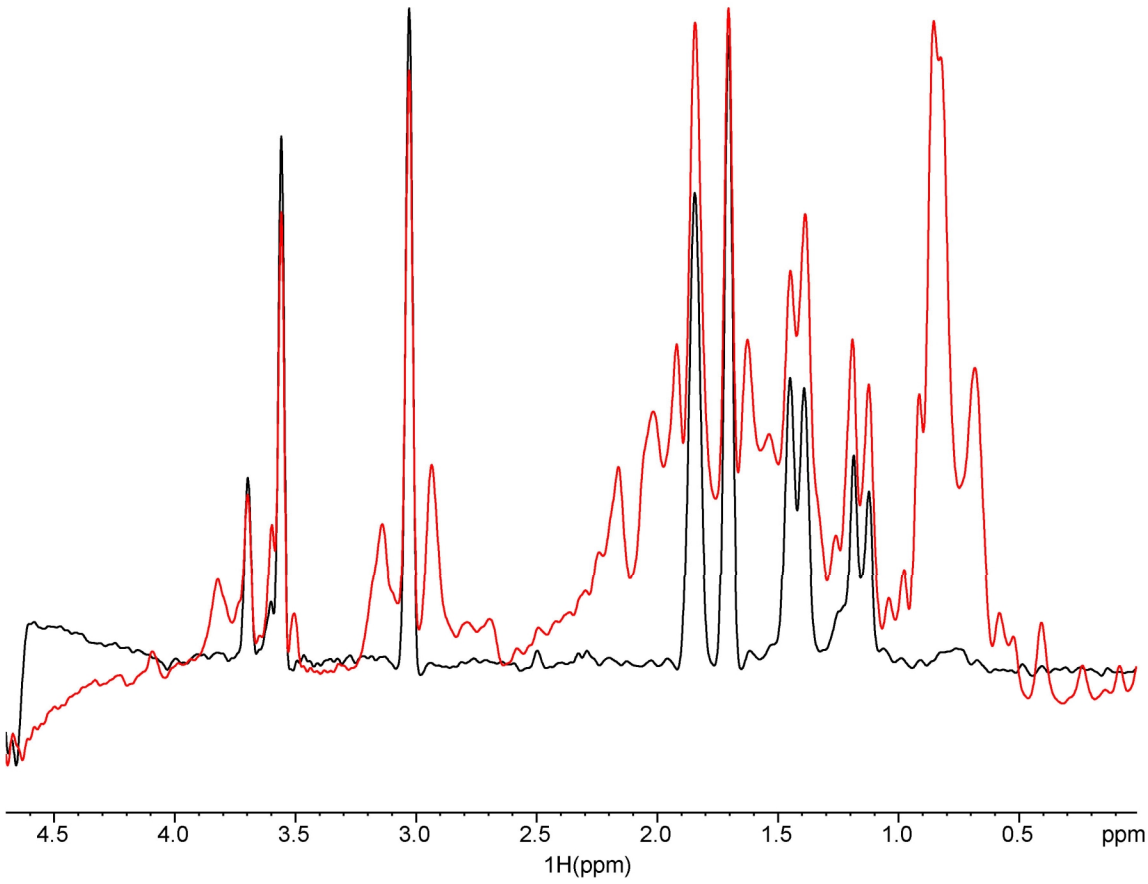


Figure 4.8: CML is unperturbed in the presence of the isolated V domain. Overlay of ^{1}H NMR spectra of CML in the absence (black) and presence (red) of isolated V domain.

Summary

The interaction of S100B with sRAGE was characterized using a protease protection assay. The proteolysis data suggest S100B binding is localized primarily to the V domain since no interaction is observed to isolated C1 or C2. However, C1 does play a secondary role, either through stabilization of V or an allosteric structural effect. In addition, binding appears to cause a conformational change in sRAGE as evident by the increased rate of trypsin digestion.

Attempts to use NMR chemical shift perturbation assays to map binding surfaces between S100B and sRAGE were plagued by complications. The one fact that remained clear was the C2 domain remained flexible in the complex of S100B and sRAGE. Hence, these results support the model for sRAGE structure described in Chapter II and suggest C2 remains structurally independent during S100B binding.

Binding of S100B to sRAGE was quantified and found to be in the range of nM affinity. An important aspect of these experiments was the structural integration of the V and C1 domains. Specifically, S100B binds to VC1 with V as the major interaction site. Consistent with the results from limited proteolysis experiments, binding to isolated V occurs with much weaker affinity than to tandem VC1, presumably due to the reduced stability of V in the absence of C1. Thus, integration of V and C1 is necessary for both structural integrity and ligand binding. The implications and broader conclusions of these data are discussed in Chapter V.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

Implications and Conclusions

The Modular Structure of sRAGE

The generation of efficient protein expression systems was critical to the studies reported in this dissertation and opens the door to use these reagents to probe RAGE function. Biophysical and structural characterization showed that C2 exhibits the characteristics of a stable well-folded structural domain. In contrast, V and C1 do not form fully independent structural domains but rather behave as an integrated structural unit. Thus, in the model for the structure of sRAGE (Figure 5.2), the C2 domain retains significant structural flexibility, whereas V and C1 are structurally interdependent.

There is ample evidence for V and C1 forming an integrated structural unit. From the initial stages of analysis, the C1 domain in particular appeared to be rather unstable without V. Many constructs of C1 and C1C2 were designed in the attempt to obtain expression of soluble protein in *E. coli* and it was only after extensive optimization that even limited success was obtained. Even the most stable of the C1 constructs exhibits clear signs of heterogeneity, is digested rapidly in limited proteolysis experiments, and readily degrades over time. The ^{15}N-^{1}H HSQC spectrum of our optimized C1 construct contains some signals with chemical shift dispersion, but the existence of a large excess of peaks with variable intensities and the crowding of many signals in the central region

suggests that the isolated domain is structurally heterogeneous and in fact a significant portion of this construct may be unstructured. The isolated V domain is more stable and structured than isolated C1. For example, it was digested more slowly by trypsin and the ^{15}N-^{1}H HSQC spectrum of V has far fewer peaks than C1, although it too shows signs of conformational heterogeneity. Importantly, both the V and C1 domains become significantly more stable to trypsin digestion and more structurally homogeneous when they are linked to each other in the VC1 and sRAGE constructs.

The x-ray crystal structure of VC1 allows refinement of the model for the structure of sRAGE. Trp230 is the C-terminal residue participating in secondary structure in C1. Two potential hydrogen bonds, Glu231 (side chain oxygen) to Gly199 (amide proton) and Pro232 (carbonyl oxygen) to Arg198 (side chain N-H), could effectively shorten the C1-C2 linker. However, crystal packing involving the C-terminal extension could stabilize these hydrogen bonds which may not be highly populated in solution. The homology model for C2 shows the first amino acid involved in β-sheet hydrogen bonding is Pro234 (Figure 5.1). Thus, the aggregate data suggest the C1-C2 linker is three residues (ignoring side chain contacts described above).

Therefore, it is intriguing to understand the structural determinants of C2 flexibility. As in all Ig folds the region of highest variability is between the *c* and *e* strands. C2 shows a truncation in sequence between strands *c* and *e* that contains a short strand (*c*') and a linker that is nearly perpendicular to the plane of the β-sandwich (Figure 5.1). This amounts to a reduction in the number of interstrand loops in proximity to the C1-C2 linker. In fact, analysis of the C2 homology model actually shows characteristics of an S-type Ig fold (120). The C1 domain has three large loops directly facing the V

domain (Figure 3.1) while the C2 domain only has two loops directly facing the C1 domain. Thus, it is possible that the lack of side chain density and surface complementarity at the C-terminal face of C1 and the N-terminal face of C2, despite the likelihood that the C1-C2 linker is ~3 residues, enables the C2 domain to remain structurally independent of VC1.

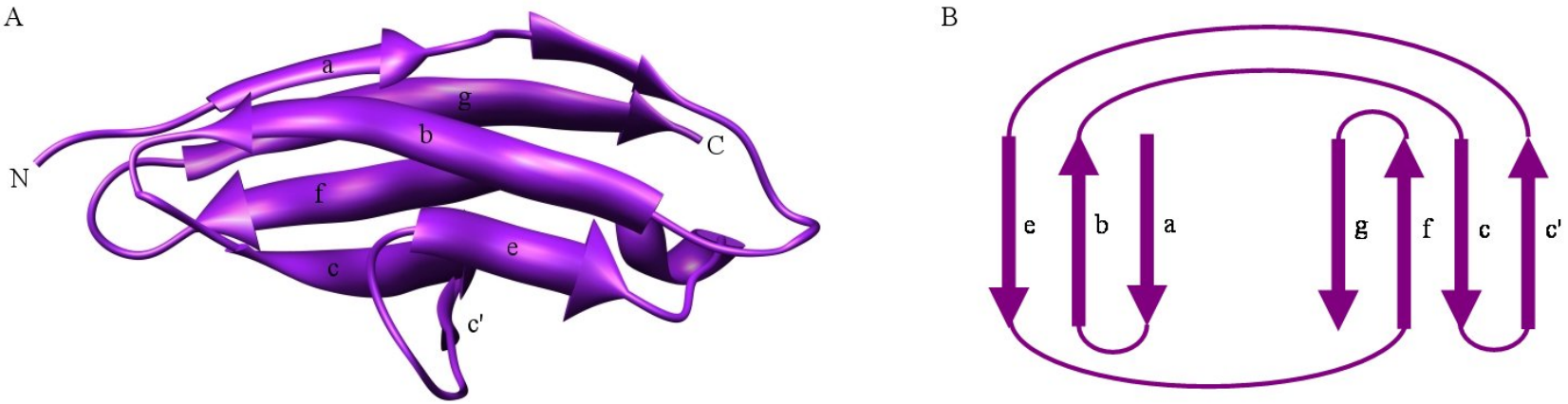


Figure 5.1: Homology model of C2 suggests S-type Ig fold. (A) Ribbon diagram of the homology model of C2 with each strand labeled according to the topology diagram in (B). The N- and C-termini are labeled. The B-C and F-G loops, in spatial proximity to C1, are near the N-terminus.

There are likely two functions to Ig interdomain linkers. One is to enable non-linear interdomain packing, as in the case of the Fab light chain used to model VC1. A short linker like in VC1 promotes a near linear interdomain angle (Figure 3.1). Increasing the linker length could therefore enable an increased angle between the two domains. In addition, longer linkers could allow conformational changes when ligand binds. The absence of a linker in VC1 will likely inhibit significant conformational changes upon ligand binding. Thus, the conformational change in VC1 induced by S100B binding that is inferred from protease protection assays (Chapter IV) may be localized to interstrand loops and not a reorientation of the interdomain angle. A second function for interdomain linkers is to allow Ig domains to remain structurally independent of each other. This seems to be the case for C2, where structural flexibility is apparently enabled by in the C1-C2 linker.

The x-ray crystal structure also helps to explain several key experimental observations. For example, the observation of an ~2 kDa truncation of VC1 in limited proteolysis experiments on the RAGE(23-243) VC1 construct can be correlated with the prediction of an ~10 residue unstructured C-terminal tail. This also explains the ~10 extremely sharp peaks observed in the HSQC spectrum of V, which the structure predicts arise because the V construct (23-132) extends 14 residues beyond the predicted C-terminus of the domain. The NMR signals from these sharp resonances change significantly when they are incorporated into the VC1 construct. This can be attributed to their shift from being unstructured in isolated V to participating in β-sheet secondary structure when the C1 domain is included (Figure 2.11). This has been confirmed through the generation of new VC1(23-233) and V(23-119) constructs whose ^{15}N-^{1}H HSQC

spectra show retention of the overall fold while lacking the ~10 sharp peaks (Appendix C). The x-ray crystal structure also suggests at least one large C1 sidechain (Tyr150) with significant hydrophobic character could be involved in interactions with V; it is possible the exposure of this residue to solvent could be the source of heterogeneity and/or instability. Hence, the overall conclusion supported by the biochemical and structural data is that V and C1 form an integrated structural unit that is fully independent of C2.

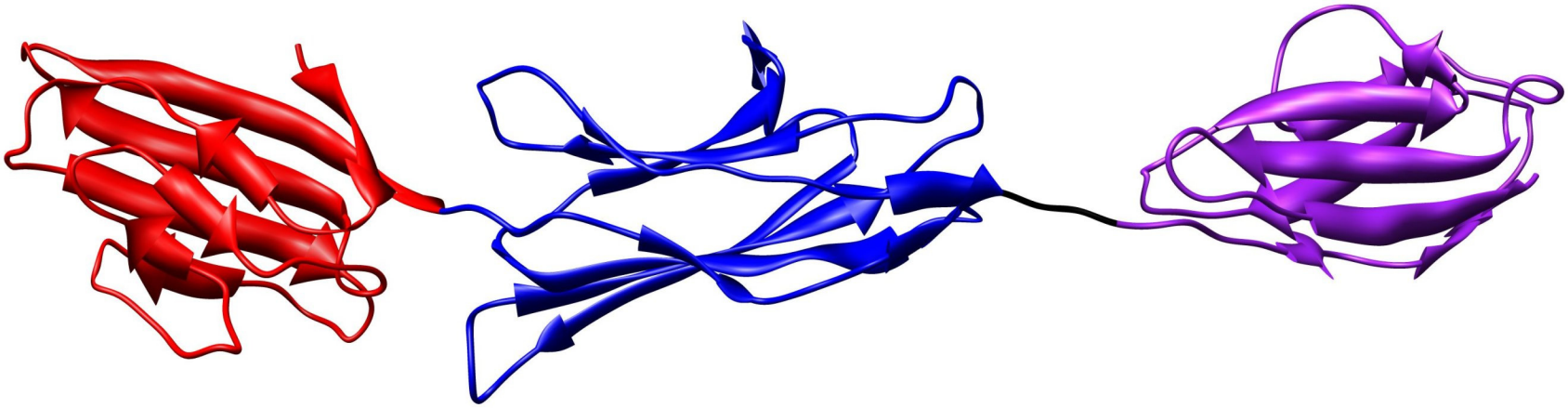


Figure 5.2: Model for the structure of sRAGE. This model was created by joining the x-ray crystal structure of VC1 (colored with the V domain in red and the C1 domain in blue) with a homology model of C2 (violet). The three residue C1-C2 linker is connected to scale and colored black. While the image is static, the C1-C2 linker has structural flexibility and the C2 domain could adopt many different orientations with respect to VC1.

Implications for Ligand Binding

The modular structure of sRAGE differs from the initial assumption based on sequence analysis that there would be three independent domains. It has been proposed that over the course of evolution, Ig-like superfamily proteins developed a modular assembly of domains. Ideally, each domain provides unique functionality, which results in the generation of proteins that can respond to multiple ligands (126). Central to this hypothesis is the concept of one-domain, one-ligand. However, RAGE does not seem to fit this paradigm as all of its ligands identified to date, including S100B, are found to interact with the N-terminal V domain of the protein. In this respect, a recent publication suggesting S100A12 binds to the C1 domain stands out as unusual (127). However, the conclusions drawn in that study were all based on use of a C1C2 construct with an N-terminus at Ala130. The x-ray crystal structure suggests this construct eliminates the entire first β-strand of C1, which is anticipated to be highly destabilizing. Indeed the results in this dissertation show that C1 does not maintain a proper fold in the absence of V. Since C1 requires V for full stability and structural integrity, the construct used in that study may well be structurally unstable and might exhibit aberrant function. The conclusion that S100A12 binds to the C1 domain may need to be re-examined with alternate reagents and these issues in mind.

The nM dissociation constant (K_D) determined for S100B binding to sRAGE and VC1 is consistent with values published for other sRAGE ligands. Dissociation constants of 60 nM, 6.4 nM, 56.8 nM, and 91 nM have been reported for AGE-BSA, amphoterin (HMG1), amyloid-β peptide, and EN-RAGE (S100A12), respectively (1, 27, 40, 43). Recently, Wilder and coworkers showed binding of a peptide from the V domain

(residues 42-59) to S100B and reported a K_D of 11 μM (128). Notably, that result is 2-3 orders of magnitude weaker than the K_Ds for other RAGE ligands. The authors proposed an important role for Trp51 in sRAGE for binding of S100B (128). In the x-ray crystal structure of VC1, this residue is buried in the hydrophobic core, which implies it is not involved in S100B binding. The structure predicts residues adjacent to Trp51 that are part of one or more loops are much more likely to contribute to S100B-RAGE binding. This hypothesis is consistent with our results on the binding of S100B to the V domain, which revealed K_Ds of 550 nM and 470 nM. Although clearly weaker than the affinity for VC1 and sRAGE, binding of the V domain is much stronger than for the V(42-59) peptide. Nonetheless, the constructs based on V alone lack critical binding elements relative to VC1, which weakens the interaction with S100B.

Attempts to monitor the interaction between synthesized CML and bacterial sRAGE did not produce observable chemical shift perturbations. One possible explanation is that native RAGE is N-glycosylated at two potential sites in the V domain, but the bacterially expressed proteins are not glycosylated. Favorable effects from glycosylation on solubility and stability are anticipated, but the effect on the structure of folded proteins is generally believed to be quite small. There is significant evidence in support of the latter in the case of RAGE. First, RAGE can be readily deglycosylated with Peptide N-Glycosidase F. This indicates that the carbohydrates are solvent accessible, which in turn implies they do not interact significantly with, or alter the structure of, the folded protein domains (129). Second these authors showed that RAGE deglycosylation has a minimal effect on binding of the ligands amphoterin and AGE-BSA. Third, similar observations have been recently reported for the bacterially

expressed protein (107). While bacterial sRAGE was capable of binding AGE-BSA (heterogeneous AGEs), Wilton and coworkers reported that bacterial sRAGE did not bind CML-BSA (107). The direct binding of synthesized CML to yeast sRAGE (109) therefore suggests some glycosylation may be necessary for maximal binding affinity.

An alternate explanation is that CML binds dynamically with high affinity and interacts via multiple binding sites. In this model, the highly acidic CML binds to the large basic surface of the V domain in a non-specific manner. The small size of CML allows it to interact at multiple sites on V and the high net positive charge of the surface ensures that CML does not effectively diffuse away allowing it to switch to a different site. The intrinsically dynamic nature of electrostatic-based ligand binding would minimize the entropic penalty associated with binding thereby enabling an intrinsically weak association for specific sites (at the microscopic scale) to be effective at providing the much stronger affinity observed overall at the macroscopic level. This is consistent with the lack of precise chemical shift perturbations and saturation of binding described in Chapter IV. Nonetheless, the accumulated evidence supports the notion that these constructs serve as valid models for the study of sRAGE structure and ligand binding.

Implications for RAGE Signaling

The characteristics of sRAGE structure relate to its mechanism of signaling. The results show that the C2 domain of RAGE has significant structural flexibility from the VC1 region. Specifically, NMR spectra of sRAGE-S100B complexes show C2 is flexibly attached to VC1 and remains unperturbed when S100B binds to VC1 suggesting there is no signal passed directly between VC1 and C2. Hence, a mechanism where ligand

binding in the VC1 region is transduced via C2 and the transmembrane helix to the intracellular domain is not anticipated for RAGE.

As described in Chapter I, RAGE belongs to a class of receptors that contain a single transmembrane helix and signal through kinases. While most of these receptors dimerize, the degree of oligomerization varies for different members of this class. In the case of EPOR, the D1 and D2 domains form an integrated structure that dimerizes in an autoinhibitory orientation (Figure 1.7). Ligand binding causes a conformational change in the interdimer angle, which is believed to promote signaling by enabling trans-phosphoryation of the kinases bound to the intracellular domain (98, 99). GHR also exists in an autoinhibited state but dimerizes via its transmembrane helix (Figure 1.7). Ligand binding to GHR results in rotation of the transmembrane helix, which in turn is transmitted to the intracellular domain because all of the structural elements are coupled (104). Thus, ligand binding causes reorientation into an architecture where trans-phosphorylation can occur. The flexibility between the VC1 and C2 domains implies neither of these allosteric mechanisms could be operative in the case of RAGE.

Signaling via RAGE occurs through the ERK1/2 kinase. Remarkably, the intracellular (IC) region of RAGE is very small and characterization of this region by CD and NMR showed it has little persistent secondary or tertiary structure (Figure 5.3). Despite being so small and unstructured, this domain binds to both phosphorylated and unphosphorylated ERK1/2 in a *ligand-dependent manner* (51). The absence of any structure in the intracellular domain implies the most likely mechanism for RAGE signaling is via ligand-induced oligomerization and resultant localization of multiple kinase molecules. A proposal for RAGE signaling through oligomerization has been

reported based on the crystal structure of S100A12 (85). In this model three dimers (six subunits) of S100A12 bind three molecules of RAGE, but the mechanistic implications for RAGE are not described.

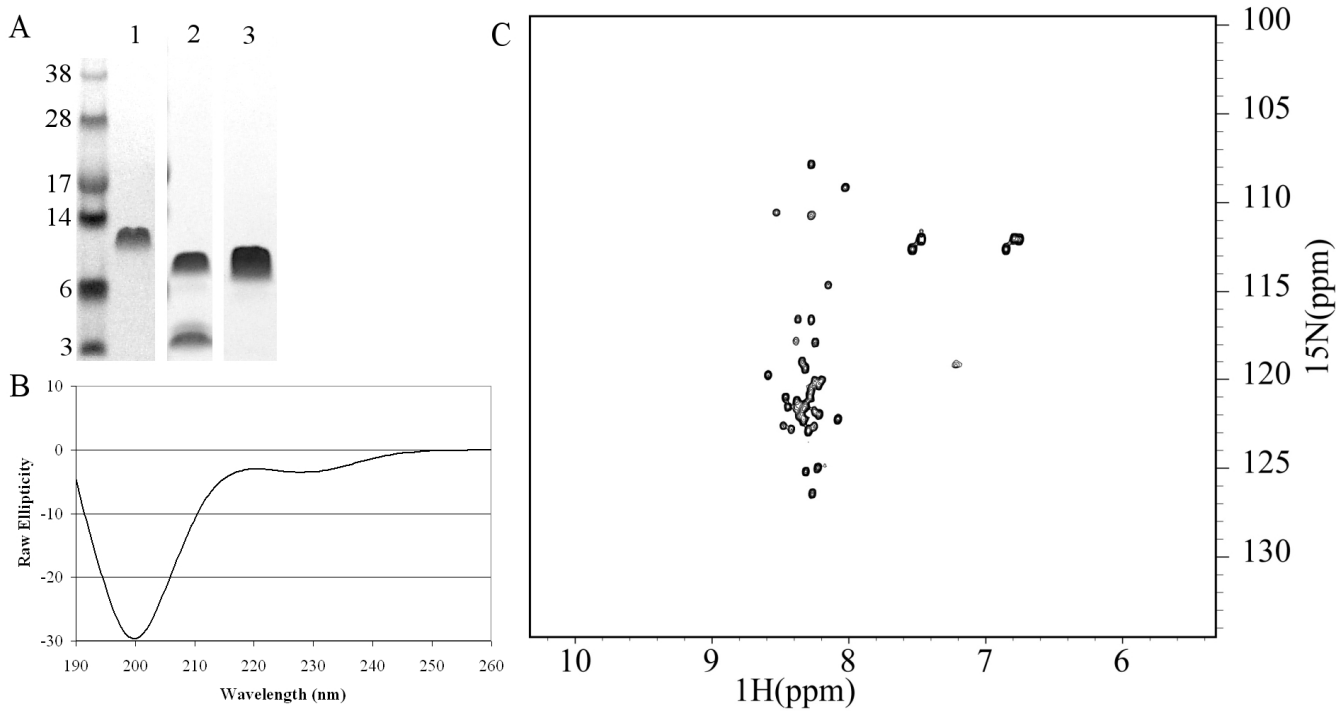


Figure 5.3: Structural characterization of the intracellular domain (IC) of RAGE. (A) SDS-PAGE showing purification of RAGE-IC (residues 364-404). Lane 1, after Ni-affinity. Lane 2, after thrombin cleavage. Lane 3, after SourceQ. Aberrant protein band migration is likely due to high acidic pI of the intracellular domain. (B) CD spectrum of RAGE-IC. (C) ^{15}N-^{1}H HSQC of RAGE-IC acquired at 500 MHz and 25 °C in 20 mM Tris-HCl, 10 mM $CaCl_2$ at pH 7.0.

Figure 5.4 shows schematic diagrams of simple models for RAGE signaling that incorporate the multiligand functionality of RAGE, receptor oligomerization, and the flexibility of the C1-C2 linker. In particular, free rotation around the C1-C2 linker allows the VC1 domains to utilize different surfaces to interact with different ligands without disrupting the orientation of the signaling components (C2, transmembrane helix, and/or intracellular domain). Figure 5.4A shows two hypothetical signaling complexes where ligand-induced dimerization occurs on different VC1 surfaces. In this scenario, unliganded RAGE would be autoinhibited, perhaps due to electrostatic repulsion of the highly charged intracellular kinase binding domains. Ligand binding dimerizes the receptor and increases localization of ERK1/2 to initiate the signaling cascade. This model could also be adapted if it turns out that RAGE dimerizes via the transmembrane helix (or via as yet uncharacterized ligands). Figure 5.4B shows a schematic diagram of one such scenario where ligand binding leads to formation of a network of RAGE receptors. One important feature of this model is the VC1 domains would still have rotational freedom to present different binding surfaces for different ligands. Thus, a ligand-induced network of RAGE receptors is formed, increasing the localization of ERK1/2 to promote signaling.

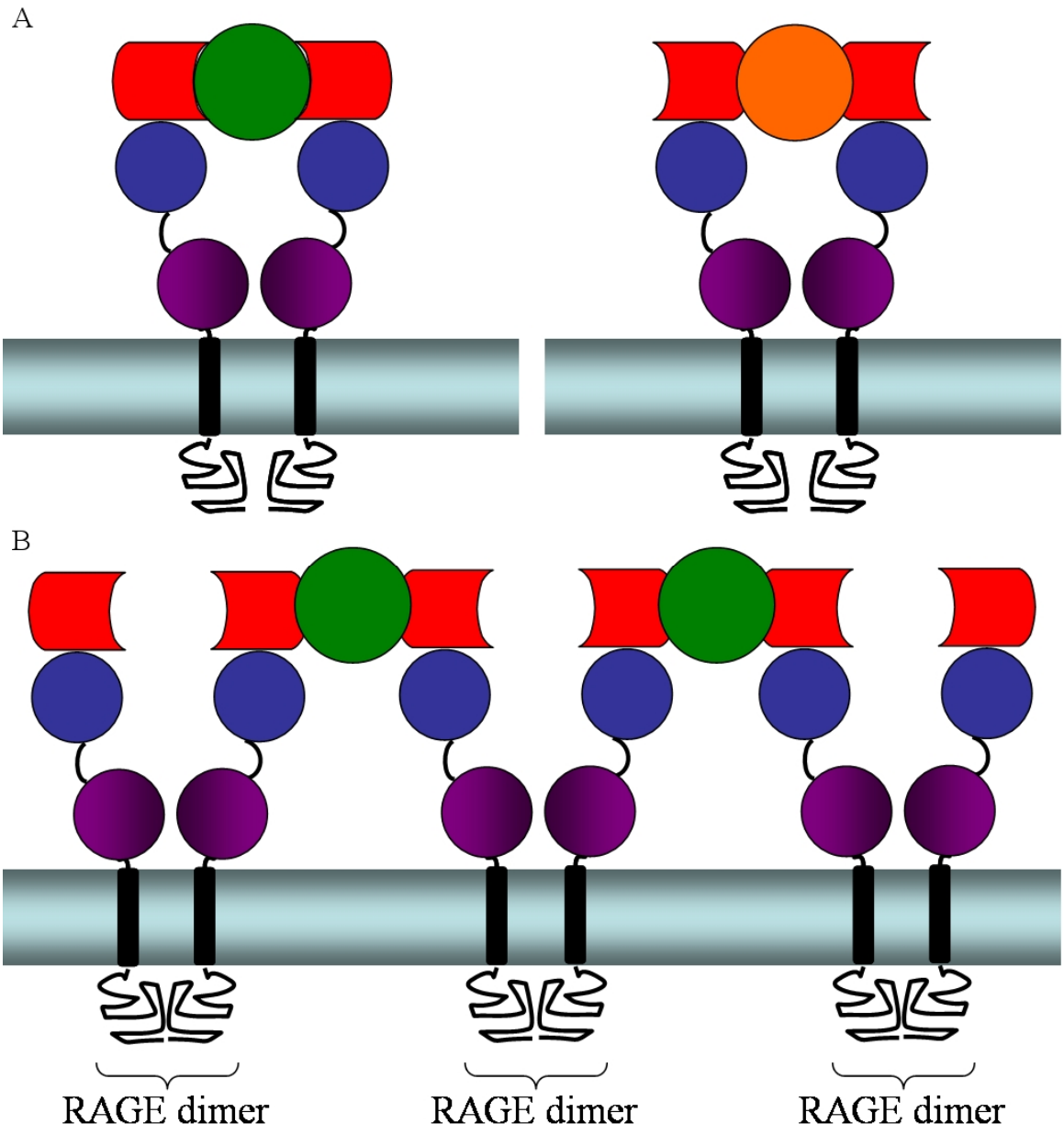


Figure 5.4: Models for ligand-induced activation of RAGE through oligomerization. (A) Schematic diagram showing ligand-induced dimerization is possible using different surfaces of VC1. (B) Schematic diagram of ligand-induced network of RAGE dimers. The ligand is shown as a green or orange circle. Coloring for sRAGE domains is the same as in Figure 1.1.

These data and observations in the literature such as the analytical ultracentrifugation study by Wilton and co-workers (107) suggest the sRAGE molecule in isolation is monomeric. These results in turn point strongly to a ligand-mediated oligomerization mechanism for RAGE signaling. This hypothesis is supported by data from a variety of RAGE splice variants including a variant lacking the intracellular domain, which was found to be dominant-negative *in vivo* (23, 130). Another splice variant lacking a portion of the V-domain was not dominant-negative, presumably because ligand binding was weakened but not abolished (131). Determination of the oligomeric states of sRAGE-ligand complexes and the natural oligomeric state of the unliganded full receptor will be required in order to develop an understanding of the molecular basis for RAGE signaling. These and other future directions are proposed in the following section.

Future Directions

High Resolution Structure of the C2 domain

The studies in this dissertation describe the quaternary structure of sRAGE which includes a high resolution x-ray crystal structure of the VC1 region. However, the high resolution structure of C2 is still lacking. The homology model described above suggests C2 is an S-type Ig domain. Given the important differences between the experimental and modeled structure of VC1, determination of the C2 structure is warranted. However, numerous attempts at crystallization have failed to produce diffraction quality crystals.

The boundaries for this construct appear to be based on the homology model, so no simple solution to crystallization comes to mind.

In contrast, the prospects for structure determination by NMR spectroscopy are quite promising. The ^{15}N-^{1}H HSQC spectrum of C2 shows excellent chemical shift dispersion and resonance linewidths (Figure 5.5). In fact, nearly complete ^{13}C, ^{15}N, and ^{1}H backbone resonance assignments have been made for C2 using standard 3D triple resonance experiments including HNCA, HN(CO)CA, CBCA(CO)NH, and HNCACB (Figure 5.5). The few incomplete assignments arise from two serine and two leucine residues (Figure 5.5, Appendix D). All four of these residues are flanked by prolines at both the N- and C-terminal positions. Thus, standard triple resonance experiments cannot differentiate the two serines (and likewise the two leucines) from each other. Analysis of NOESY and TOCSY spectra should enable completion of these assignments. Hence, NMR solution structure determination is quite feasible. Key issues that remain to be resolved are the exact topology of C2 and the precise length of the C1-C2 linker, both of which are of vital importance to understand signaling mechanisms.

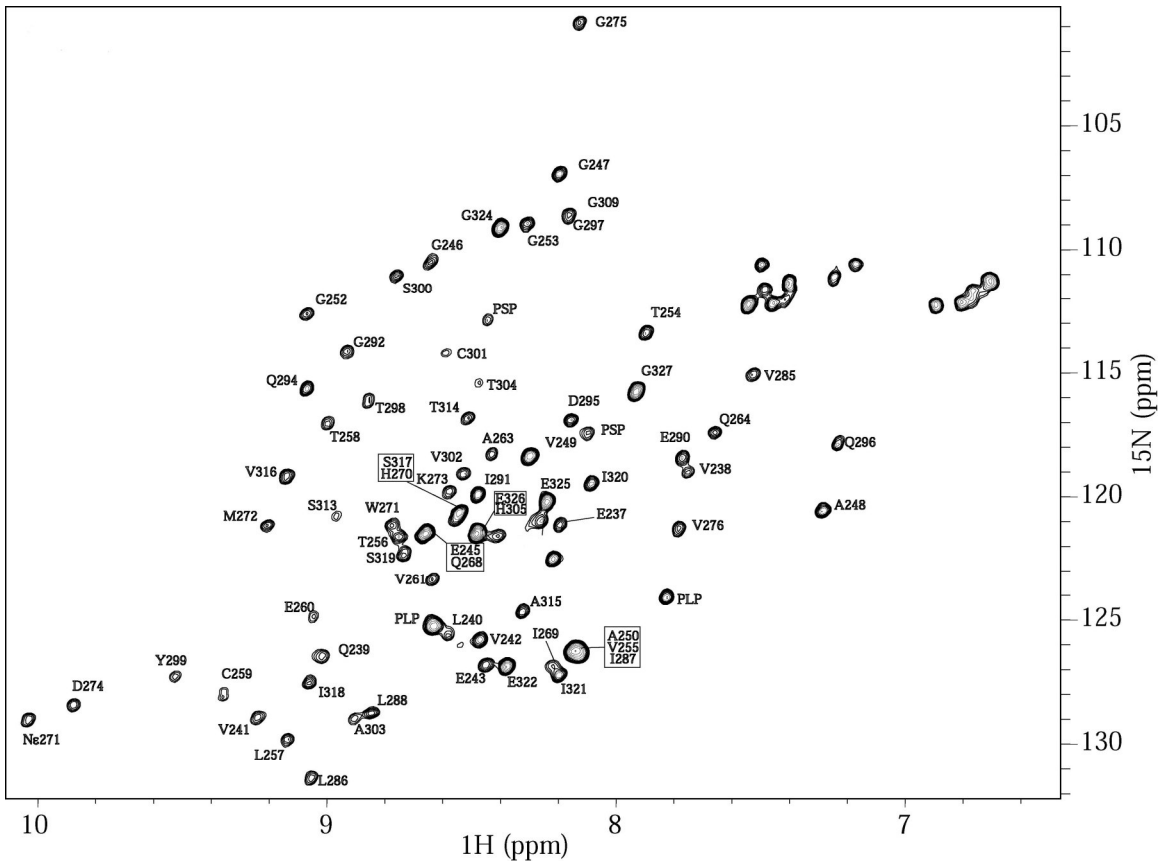


Figure 5.5: ^{15}N-^{1}H HSQC spectrum of C2 with backbone assignments. The underlined residues in the following sequence have been assigned:

GSHMLEEVQLVVEPEGGAVAPGGTVTLTCEVPAQPSPQIHWMKDGVPLPLPPSP
VLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISIIEPGEEG

A complete list of backbone ^{1}H, ^{15}N, and ^{13}C chemical shifts is presented in Appendix D.

The Oligomeric State of Intact RAGE

All analyses thus far have indicated that sRAGE exists primarily in a monomeric structural state (1, 107, 109, 132). The data in this dissertation is fully consistent with these studies. In a critical functional assay using intact RAGE, deletion of the cytoplasmic domain leads to a dominant-negative receptor (dnRAGE), which clearly implies RAGE functions in an oligomeric state. Thus, the natural oligomeric state of the intact receptor is critical for understanding the structural mechanisms of autoinhibition and activation.

Numerous approaches can be made to address this problem. RAGE constructs with different affinity tags can be cotransfected in cultured cells and analyzed with and without ligand stimulation. Immunoprecipitation with one antibody followed by detection with the other antibody could assess natural oligomerization and ligand dependence of oligomerization since physical contact between of the differently tagged proteins would be required for coimmunoprecipitation. Additionally, fusions to fluorescent probes would allow FRET experiments to assess both oligomerization and relative proximity of the intracellular domains in addition to ligand-induced changes. Fascinating experiments have been performed on EPOR using dihydrofolate reductase subunits to show both unliganded dimerization and a ligand-induced structural change (100).

The experiments described above, however, would not establish the structural features of RAGE that promote oligomerization. Assays have been developed in *E. coli* to assess the potential for transmembrane helices to oligomerize. One assay called TOXCAT utilizes a transcriptional activator that is dependent on dimerization to confer antibiotic resistance to chloramphenicol (133). The transformed plasmid encodes for

periplasmic localized MBP, fused to the transmembrane helix sequence in question, which in turn is fused to the transcriptional regulator. Thus, not only can this assay establish if the transmembrane helix (RAGE-TM) dimerizes, but it can be used to assay mutations to disrupt oligomerization. Finally, these mutations can be reincorporated into cell based assays to assess the functional importance of RAGE-TM oligomerization. This information is critical to determine if oligomerization is necessary and sufficient for activation or if specific structural orientations are required.

A separate question involving receptor oligomerization is generated by the observation that ERK1/2 coimmunoprecipitates with RAGE and that the interaction is *ligand-dependent* (51). This implies either a direct structural change or oligomerization is necessary to present a binding surface for the kinase to interact with the intracellular domain (RAGE-IC). While the interaction in cell culture was ligand-dependent, purified RAGE-IC was also able to bind ERK2 (ERK1 was not tested). However, it is of note that the approach for purification and immobilization in the pull down assays involved GST. GST is a known dimer and the specific plasmid used has a significant linker that could allow the fused RAGE-ICs to come into proximity. This interaction should be tested *in vitro* using purified proteins with no tag or with fusion proteins that are monomeric. Differences in binding to monomeric RAGE-IC vs. GST-fused RAGE-IC could provide insight into mechanisms of receptor activation, especially if a dimeric GST is required for interaction. Receptor oligomerization is also discussed later with respect to the stoichiometry of sRAGE-ligand complexes.

Variable RAGE Surfaces Used in Ligand Recognition

RAGE is activated by structurally diverse ligands, which suggests that different surfaces of RAGE extracellular domain are utilized. Multiple approaches can be used to determine the molecular details of protein-protein interactions including x-ray crystallography. In fact, preliminary crystals of the S100B-VC1 complex have been obtained by Dr. Guenter Fritz and diffract to ~7 Å. These crystals may be improved using optimized constructs based on the x-ray crystal structure presented above (Appendix C). Nonetheless, it is essential to pursue structure determination of this complex or use other techniques to elucidate the surface residues necessary for binding because the details of ligand recognition are central to the hypothesis introduced above that the flexibility of the C1-C2 linker is required to allow activation by different ligands.

NMR provides an alternative strategy to characterize the structure of the S100B-VC1 complex. Overcoming the large molecular size and long correlation time is typically done using perdeuteration and TROSY (111, 112). However, this system has still been difficult to study despite these technical advances, presumably due to non-binary binding kinetics (consistent with the complex binding model from SPR data discussed in Chapter IV). Therefore, simplifying the analysis by using less direct structure-function strategies can be taken to obtain structural insights.

For example, one recent study identified a short peptide from the V-domain of RAGE that exhibited an 11 μM dissociation constant for binding to S100B (128). An S100B consensus binding motif has been previously identified (134) and points to a key role for Trp51, whose side chain lies buried in the hydrophobic core of the V domain. Therefore, it is possible that surface exposed residues adjacent to Trp51 play a more

important role in interaction with the peptide. Similar approaches involving mutation of amino acids with surface exposed side chains (controlling for overall structural integrity) can be used to assay whether this segment of the V domain is involved in binding to S100B (Figure 5.6).

A 34 residue peptide from amphoterin has been shown to bind RAGE (38). Half of this peptide lies within the C-terminal α-helix of the second HMG box domain of amphoterin (Figure 1.3), with the other half existing in the unstructured C-terminal region. This peptide functions as a RAGE antagonist *in vitro* and *in vivo* and could be used to assess the molecular details of RAGE-amphoterin interaction. Analyzing a complex of a peptide and a ^{15}N-enriched protein is typically easier than two intact proteins due to smaller changes in the size of and hence the correlation time of the complex.

Another structure-function study involved the minimal Aβ sequence that shows strong affinity for sRAGE. The Aβ sequence is 11 residues in length and encompasses positions 25-35 of the canonical Aβ peptide (40). Generally, fibril formation is a slow process in the absence of seeding suggesting NMR titrations could be performed, especially at lower protein concentrations, taking advantage of the high S/N of cryoprobes. This might provide similar information to the above described analysis with amphoterin.

All information gained from the above experiments would be combined with our current knowledge of AGE recognition by RAGE. Yeh and coworkers showed that a 17 amino acid peptide from the V domain (shown mapped onto the structure of V in Figure 5.6) is able to compete for binding of CML-HSA (25). However, this finding is

controversial since some studies suggest RAGE binding to CML requires N-glycosylation (107). Nonetheless, a critical aspect to understanding the structural mechanisms of RAGE activation lie in the molecular details of multiligand recognition. Therefore a broad approach should be taken utilizing structural techniques in addition to carefully controlled functional studies to ascertain these critical pieces of information.

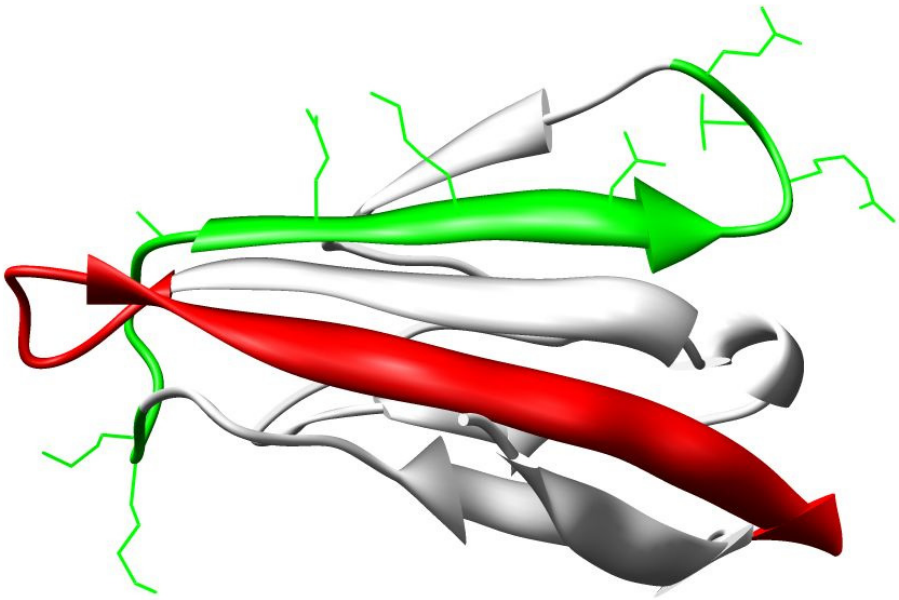


Figure 5.6: Regions of ligand recognition by RAGE. Ribbon diagram of the x-ray crystal structure of RAGE-V (C1 has been removed for clarity) showing segments implicated in ligand recognition. Colored in red is the peptide used to compete for binding to CML-HSA. Colored in green is the peptide shown to bind *in vitro* to S100B. Surface exposed side chains that could contribute to S100B binding are shown as wires.

Stoichiometries of sRAGE-Ligand Complexes

Earlier I proposed experiments to assess the oligomeric state of the intact receptor to gain insight into signaling mechanisms. Equally important in receptor signaling are changes in oligomeric state caused by ligand binding. In the absence of ligand, the RAGE extracellular domain is predominately monomeric as shown by numerous groups in addition to the studies described within this dissertation (1, 107, 109, 132). Based on structural characteristics of RAGE ligands and cytokine receptors described in Chapter I, multiple copies of sRAGE are predicted to be involved in RAGE signaling. Ideally, techniques such as gel filtration and light scattering could be useful in this context. However, the information from these approaches tends to be incomplete in the absence of details on the precise shape of the complex. Ig domains have an elongated structure, a property that is further accentuated in tandem constructs like VC1 and intact sRAGE, so other complementary experimental approaches are required. For example, analytical ultracentrifugation should be the primary technique used to estimate RAGE-ligand stoichiometries. A combination of sedimentation velocity and sedimentation equilibrium can assess both the molar mass of these complexes and equilibrium dissociation constants, respectively. Comparing the stoichiometries of different sRAGE-ligand complexes would provide information on potential differences in RAGE activation proposed at the beginning of this chapter.

Potential Mechanisms for AutoInhibition of RAGE

Preventing the transient activation of a cell surface receptor is important for tight regulation of signaling pathways. Autoinhibition of cytokine receptors was described in

Chapter I. In one model, a preformed receptor dimer exists in a fixed orientation that requires ligand binding to physically bring the intracellular domains into proximity and promote trans-phosphorylation of the associated kinases. In a second model, the associated kinases are in proximity in the preformed dimer but are rotated in an autoinhibited conformation.

A common structural feature within the extracellular domains of cytokine receptors, the WSXWS motif (one-letter amino acid code where X is any amino acid), is necessary for proper receptor regulation. The crystal structure of IL-4R reveals the two Trp side chains from this motif form a tryptophan-arginine zipper as shown in Figure 5.7. A recent proposal suggests the WSXWS tryptophan side chains can flip out of the zipper and embed in the membrane creating an autoinhibited conformation of the receptor (135). Free energies were calculated for solvent exposure of the WSXWS tryptophan side chains; the energetic penalty of solvent exposure could be compensated for by the free energy gained through membrane interactions. Thus, ligand binding could cause the release of the extracellular domain from its interactions with the membrane facilitating receptor activation.

RAGE does not have a defined tryptophan-arginine zipper; however, RAGE does have a number of surface accessible aromatic side chains. Two in particular are Trp61 and Trp72, of the V domain, whose side chains have 46% and 57% solvent accessible surface area, respectively (Figure 5.7). Both residues are 100% conserved across species (Appendix A). The functional significance of these surface aromatic side chains can be tested with point mutagenesis. Trp→Arg or Trp→Lys mutations could satisfy the hydrophobic contacts while placing charged groups at the surface. Mutations in the full

receptor can be assayed in cell culture for transient and ligand-induced activation. "Leaky" activation of NF-κB in mutant vs. wild-type RAGE would imply a disruption in autoinhibition.

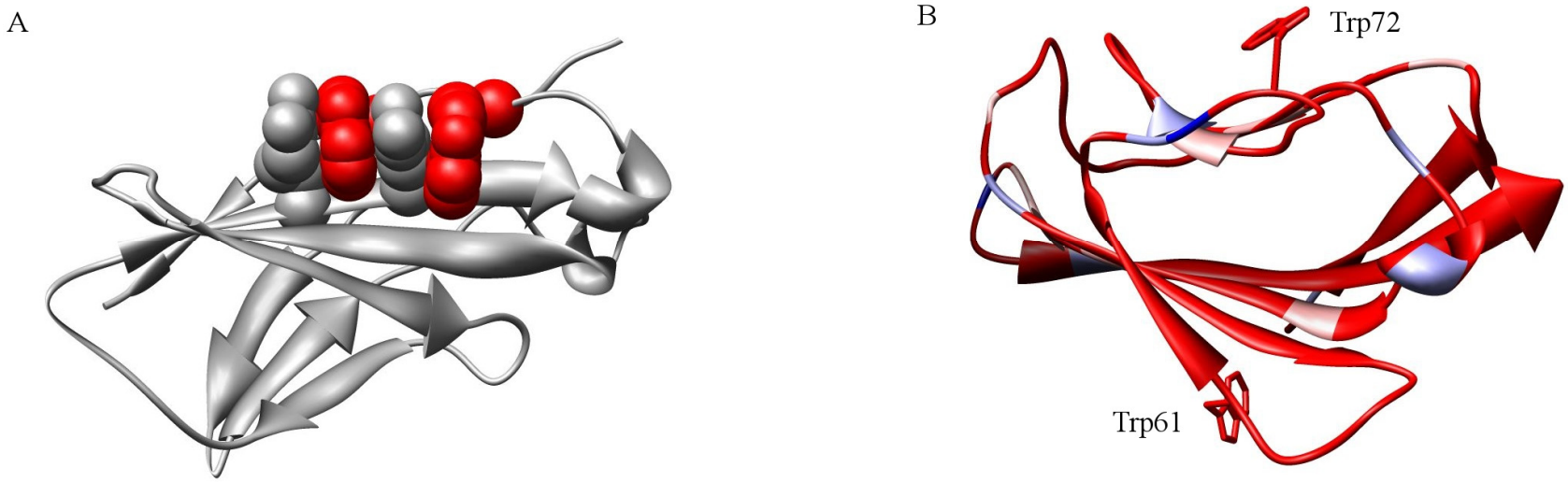


Figure 5.7: Solvent accessible side chains as a mechanism for autoinhibition. (A) A ribbon diagram shows the D2 domain (D1 removed for clarity) in the structure of IL-4R (PDB 1IAR). The two Trp side chains in the tryptophan-arginine zipper are shown as red spheres. The two arginine side chains in the tryptophan-arginine zipper are shown as grey spheres. (B) Ribbon diagram of the V domain in the VC1 x-ray crystal structure. Structure is colored according to % conservation as in Figure 3.4. The side chain atoms are shown for Trp61 and Trp72.

Functional Role of the C2 Domain

A gap in the knowledge of RAGE structure that is critical to function is the relative flexibility of the C2 domain with respect to the membrane. Based on the homology model of C2, there is an ~20 residue linker between the end of the C2 structured region and the start of the transmembrane helix. The flexibility of C2 can be tested *in vitro* by preparing ^{15}N-enriched protein in detergent micelles. An ideal construct would contain the C2 domain through the IC. The transmembrane helix would be solubilized by the detergent while the IC and the C2 domain would be exposed to solvent.

In fact, numerous constructs have been prepared to address this issue and preliminary expression and purification suggests sample preparation is feasible. The first strategy was to design a construct that would drive expression into the periplasm to potentially trap the transmembrane helix in the inner membrane. A second strategy is to express protein only with an N-terminal His_6 tag. This would eliminate the need to separate the fusion from the native protein. In fact, a C2 to IC construct with only a His_6 tag has readily been solubilized in β-octylglucoside, the detergent used in initial RAGE isolation (1).

NMR linewidths are sensitive to many factors, one being the correlation time of the molecule which is proportional to its size. Thus, a solubilized protein would have broad lines since the detergent micelle will have a long correlation time due to its size. However, regions of a protein that are flexible from the micelle could have shorter correlation times and hence narrower linewidths. This property was described in Chapter II with the linewidth differences between VC1 and C2 in full sRAGE ^{15}N-^{1}H HSQC spectra and equally applies to protein-detergent complexes.

A series of deletions and substitutions have shown that the membrane proximal domains of the complex receptor gp130 (containing 6 domains) are essential for efficient receptor activation (136). Thus, it is possible that C2 partially embeds in or interacts with the membrane and to some degree orients the VC1 region in either an autoinhibited conformation or a conformation necessary for activation. If C2 is flexible from the transmembrane helix, I anticipate observing signals in a ^{15}N-^{1}H HSQC spectrum that correspond to isolated C2. If C2 (or the linker between C2 and the transmembrane helix) interacts or embeds with the detergent then perdeuteration and TROSY will be required to overcome the broad linewidths caused by the long correlation time. Thus, the relative flexibility can be tested by 2D ^{15}N-^{1}H NMR experiments and provide this critical piece of information.

Prospects for sRAGE Based Therapeutics

Numerous studies described in Chapter I have shown that exogenous sRAGE functions as a decoy to block RAGE-ligand signaling in numerous pathologies. This is possible due to the absence of an immune response since sRAGE exists in blood plasma in an alternatively spliced form (131, 137). Models for advanced atherosclerosis (56), tumor growth and metastasis (38, 58), Alzheimer's disease (60), and chronic inflammation (63) have shown exogenous treatment with sRAGE has powerful potential for treatment.

Due to structural independence of VC1 from C2, in addition to ligand binding localization to this region of RAGE, VC1 may provide a more stable alternative to working with intact sRAGE. For example, the VC1 construct used in this dissertation has

better solubility properties than the sRAGE construct. Low solubility makes drug formulation challenging and can diminish shelf-life. The feasibility of using VC1 as a therapeutic should be tested in the mouse models described in Chapter I (using sRAGE and isolated C2 as controls) for its potential to augment disease progression.

Closing Remarks

Overall, the structural studies reported in this dissertation have provided important insight in to the function of RAGE. While many questions have been answered with regards to tertiary and quaternary structure, many more need to be answered including intact RAGE quaternary structure and the details of ligand recognition. Most importantly, this work has established the foundation upon which these future questions can be addressed thereby providing a broader understanding of RAGE as a signaling receptor in normal physiology and disease.

CHAPTER VI

MATERIALS AND METHODS

Molecular Biology

Bacterial expression vectors for RAGE constructs (GenBank accession no. NM_001136) were produced by three steps. DNA was extracted from baculovirus-infected insect cells generously provided by Dr. Ann Marie Schmidt (Columbia University, New York, NY) or purchased from Invitrogen. DNA fragments were PCR amplified with primers containing the appropriate restriction sites and stop codons. A list of constructs including precise sequences, the vector and restriction sites used, and minor details of protein expression and purification are shown in Table 6.1. Constructs subcloned into the pET15b vector (Novagen) produced N-terminal His_6 tagged fusion proteins separated by a thrombin cleavage site. Constructs subcloned into pET40b produced N-terminal DsbC fusions for periplasmic localization. Creation of a human S100B expression vector in pGEMEX, which produced protein of wild-type sequence, was described previously (89). Three Cys→Ser mutants of human S100B (C84S, C68S, C68S/C84S) were produced in the pGEMEX expression vector using the QuickChange strategy (Strategene).

Table 6.1: Summary of expression vectors with expression and purification strategies of recombinant proteins.

Protein	Vector	Sequence	Restriction	Cell Line	Purification[‖]
sRAGE[†]	pET15b	23-327	NdeI/XhoI	OrigamiB(DE3)	Ni, S
VC1-a[†]	pET15b	23-243	NdeI/XhoI	OrigamiB(DE3)	Ni, S
VC1-b[†§]	pET15b	23-233	NdeI/XhoI	OrigamiB(DE3)	Ni, S
C1C2	pET15b	122-327	NdeI/XhoI	OrigamiB(DE3)	Ni, S
V-a[†]	pET15b	23-132	NdeI/XhoI	OrigamiB(DE3)	Ni, S
V-b[†§]	pET15b	23-119	NdeI/XhoI	OrigamiB(DE3)	Ni, S
C1-a	pET15b	122-243	NdeI/XhoI	OrigamiB(DE3)	Ni, S
C1-b	pET15b	119-233	NdeI/XhoI	OrigamiB(DE3)	Ni, S
C2	pET15b	235-327	NdeI/XhoI	OrigamiB(DE3)	Ni, Q
IC-a	pET15b	364-404	NdeI/XhoI	BL21(DE3)	Ni, Q
IC-b	pET15b	364-403	NdeI/XhoI	BL21(DE3)	Ni, Q
RAGE	pET15b	23-404	NdeI/XhoI	OrigamiB(DE3)	Ni
C2IC	pET15b	235-404	NdeI/XhoI	OrigamiB(DE3)	Ni
RAGE	pET40b	23-404	NcoI/XhoI	BL21(DE3)	Ni
RAGE	pET40b	23-404	ScaI/XhoI	BL21(DE3)	Ni
C2IC	pET40b	235-404	NcoI/XhoI	BL21(DE3)	Ni
C2IC	pET40b	235-404	ScaI/XhoI	BL21(DE3)	Ni
TMIC	pET40b	364-404	NcoI/XhoI	BL21(DE3)	Ni
S100B[⊥]	pGEMEX	wild-type	N/A	BL21(DE3)	DEAE, S75

[†] G82S mutants were prepared for these constructs but were not characterized.
[§] These constructs were prepared by inserting a stop codon after the appropriate sequence by the QuickChange mutagenesis strategy.
[‖] Symbols used are Ni (Ni^{2+}-affinity), S (MonoS), Q (SourceQ), DEAE (diethyl aminoethane), and S75 (Superdex 75).
[⊥] C68S, C84S, and C68S/C84S mutants were prepared and purified the same.

Protein Expression and Purification

RAGE fragments overexpressed in *E. coli* strain OrigamiB(DE3) (Novagen) were grown at 37 °C to OD_{600} ~0.8, adjusted to 20 °C for 30 minutes, induced with 0.5 mM IPTG, and allowed to express for 4-6 hours. Constructs overexpressed in BL21(DE3) were grown the same but lacking the shift to 20 °C. Cells were lysed at 4 °C in 20 mM Tris-HCl, 20 mM imidazole, 300 mM NaCl at pH 8.0 in the presence of lysozyme (5mg/ml), followed by sonication (5 min with a 50% duty cycle). Clarified lysate was initially purified on His-Select (Sigma) resin equilibrated in the lysis buffer and eluted with a 4 column volume (CV) linear gradient to 20 mM Tris-HCl, 300 mM imidazole, 300 mM NaCl at pH 8.0. Following dialysis, the His_6 tag was removed by thrombin cleavage (1-2 units per mg of protein) incubated at room temperature for 1-2 hours followed by separation over SourceQ (C2) using 20 mM Tris-HCl at pH 7.7 and a 18 CV linear gradient from 0 to 1 M NaCl or MonoS (sRAGE, VC1, C1C2, V, C1) using 20 mM sodium phosphate at pH 6.0 and a 18 CV linear gradient from 150 to 850 mM NaCl (GE Healthcare). The His_6 tag was not cleaved from C1 or C1C2 due to observation of secondary cleavage. The amino acids Gly-Ser-His-Met remained at the N-terminus after thrombin cleavage.

The expression system for human S100B in pGEMEX vector was described previously (89) except luria broth (LB) was used in place of DYT. Cells were resuspended in 20 mM Tris-HCl at pH 7.7 and lysed at 4 °C in the presence of lysozyme (5mg/ml) followed by sonication (5 min with a 50% duty cycle). Purification of samples involved an initial step of DEAE ion exchange chromatography, equilibrated in 20 mM Tris-HCl at pH 7.7 and eluting bound protein using a linear gradient from 0 to 1 M NaCl.

This was followed by Superdex 75 gel filtration in 20 mM Tris-HCl, 0.1 mM EDTA at pH 7.7. C84S, C68S, and a double C68S/C84S mutant of S100B were purified the same as wild-type.

Samples of ^{15}N-enriched sRAGE and S100B proteins were produced as described above except for the use of M9 minimal media containing $^{15}NH_4Cl$ as the sole nitrogen source. $^{2}H,^{15}$N-enriched protein (~85% as estimated by MALDI mass spectrometry) was produced in a similar manner using cells adapted to growth in $^{2}H_2O$. The integrity of each protein sample was verified by SDS-PAGE and MALDI mass spectrometry.

Limited Proteolysis/Protease Protection

Limited proteolysis was performed at room temperature on sRAGE and domain fragments that had been dialyzed against 10 mM HEPES-NaOH, 75 mM NaCl at pH 7.0. Typically, 50-100 μg of purified protein were incubated with trypsin at an enzyme to protein ratio of 1:500 (w/w) and a volume of 100 μl. The reaction was stopped at various time points by mixing 14 μl of reaction mix with 14 μl 2x SDS loading buffer and heating at 90 °C for 5 minutes. For protection assays, S100B was mixed with sRAGE or domain fragments at a 2:1 molar stoichiometry of RAGE to S100B dimer, anticipating two binding sites per S100B dimer. $CaCl_2$ was added to a final concentration of 1 mM prior to the addition of enzyme. Control experiments with S100B alone and sRAGE or domain fragments were performed under identical conditions (i.e. in the presence of Ca^{2+}).

Mass Spectrometry

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) analysis was performed on tryptic peptides derived from gel-excised Coomassie stained bands. Samples were prepared by the dried-droplet method using α-cyano-4-hydroxycinnamic acid [dissolved in water/acetonitrile/trifluoroacetic acid (39.2:60:0.1) at a concentration of 5 mg/mL, supplemented with 1 mg/mL ammonium citrate] as a matrix. The peptides were initially identified by comparing the experimental masses of each peak with computer-predicted masses of tryptic peptides from the RAGE sequence. Amino acid sequence identity was definitively established by inducing ion fragmentation using tandem TOF-TOF MS.

Circular Dichroism

All samples were buffer exchanged into 50 mM potassium phosphate at pH 7.5. Protein concentration was typically around 30 μM. Data ranging from 200-240 nm (1 nm increment) were acquired on a Jasco J-810 CD spectrophotometer (Easton, MD). Baseline was adjusted during acquisition from data acquired on the solution alone. All data were converted to mean residue ellipticity using the following equation:

$$[\Theta] = 100*\text{signal} / C\ n\ l$$

where [Θ] is mean residue ellipticity in units $deg \cdot cm^2 \cdot dmol^{-1}$, signal is the raw signal output in mdeg, C is protein concentration in millimolar, n is the number of residues, and l is the cell pathlength in centimeters. Fitting using the K2d web server was used to estimate secondary structure content (138).

Thermal melts were performed using a scan rate of 60 °C per/h and monitoring β-sheet ellipticity at 218 nm. Data were acquired in 1 °C increments with a 5 min equilibration prior to cooling.

Dynamic Light Scattering

All data were acquired using a DynaPro ProteinSolutions molecular sizing instrument (Wyatt Technology Corporation). Typically, 150 μM protein in 10 mM HEPES-NaOH, 75 mM NaCl at pH 7.0 in a volume of 60 μL were filtered and centrifuged at >15,000 x g to remove any particulate matter. Twenty scans were accumulated for each sample at room temperature. The data were analyzed using the Dynamics 5 software provided by the manufacturer and are represented as mean radius ± polydispersity.

Differential Scanning Calorimetry

Samples of sRAGE, VC1, C1C2, V, C1, and C2 (50-100 μM) were buffer exchanged into 50 mM sodium citrate at pH 6.7 using a 50 mL desalting column (GE Healthcare) to ensure identically matched buffer components. Data were collected on a VP-DSC (Microcal; Northampton, MA) with a scan from 20 °C to 100 °C at a rate of 90 °C/hr. 50 mM sodium citrate at pH 6.7 was used as the reference solution. All samples and buffers were thoroughly filtered and degassed prior to analysis. Data were analyzed using the Origin7 software package provided by the manufacturer, baseline corrected for buffer alone, and fit with a non 2-state equation.

Surface Plasmon Resonance

CM5 sensor chips, coupling reagents (NHS, EDC and ethanolamine), and P20 were purchased from Biacore International SA (Freiburg, Germany). CM5 sensor chips were preactivated using the NHS:EDC chemistry according to previously published procedures (139, 140). 50 μg/ml V, VC1, C2, or yeast sRAGE in sodium acetate buffer at pH 5.5 was injected over the preactivated surface in order to obtain about 5000 RU. The surface was then blocked with ethanolamine. In each experiment, the fourth flow-cell of the sensor chip was kept empty and used as an internal reference. A series of various concentrations, ranging from 12.6 μM to 1.57 μM for S100B, were injected over the flow-cells. The running buffer was 50 mM Tris-HCl, 150 mM NaCl, 5 mM $CaCl_2$, 0.005% P20 at pH 7.4. The surface was regenerated each time by 1 min contact with 0.5 M EDTA following by 1 min contact with 50 mM borate, 1M NaCl at pH 8.5.

Binding data were analyzed by global analysis using BiaEvaluation 3.1 software from the manufacturer (Biacore), in which association and dissociation data for the series of S100B concentrations were fitted simultaneously (141).

Nuclear Magnetic Resonance Spectroscopy

Standard ^{15}N-^{1}H HSQC and TROSY-HSQC (111) spectra were acquired at 298 K on Bruker *Avance* 600 and 800 MHz spectrometers equipped with triple resonance cryoprobes. Typical acquisition parameters were 128 scans with 1024 points in the direct (^{1}H) dimension and 128 points in the indirect (^{15}N) dimension. Fourier transform, zero filling, and a 90 degree shifted squared sine window function were applied using XWINNMR (Bruker) and analyzed using Xeasy (142). The spectrum for sRAGE was

obtained in a solution containing 200 μM protein in 20 mM sodium phosphate, 250 mM NaCl at pH 6.0. Initial spectra for the other constructs were obtained under these conditions, and then optimized specifically for each. Deviations from the sRAGE conditions were: V, 10 mM sodium citrate at pH 6.0; C2, 20 mM sodium phosphate at pH 6.0. Protein concentrations were determined using the UV absorbance at 280 nm under denaturing and reducing conditions and ranged from 200 to 500 μM.

For studying sRAGE-S100B interactions, ^{15}N-V and unlabeled S100 proteins were buffered in 10 mM HEPES-NaOH at pH 7.0, 75 mM NaCl and 10 mM $CaCl_2$. Spectra of isolated proteins and protein-protein complexes were acquired under identical conditions with a 2:1 ratio of RAGE to S100B dimer. Samples of ^{2}H,^{15}N-sRAGE were buffer exchanged into 20 mM sodium phosphate and 500 mM NaCl at pH 6.0 in 100% H_2O. Sample volume was ~160 μL in a 3 mm NMR tube, which was placed within a 5 mm NMR tube filled with >99.9% $^{2}H_2O$.

Sequence-specific backbone assignments for C2 were made using 600 μM ^{15}N,^{13}C-enriched protein using standard triple-resonance experiments. This included the following 3D spectra: HNCA, HN(CO)CA, HNCO, HNCACB, CBCA(CO)NH. Typical acquisition parameters included 2048 points in the direct dimension (^{1}H) with 40 (^{15}N) and 96 (^{13}C) increments in the indirect dimensions. Deviations included 128 and 64 increments (^{13}C) for HNCACB and HNCO, respectively.

Homology Modeling and Structure Analysis

Sequences for human sRAGE and domain constructs were submitted to the ESyPred3D web server (143). The program performs all steps of homology modeling

including homolog identification (PSI-BLAST of NCBI NR databank then scored against the PDB), multiple sequence and pairwise alignments (combination of ClustalW, Dialign2, Match-Box, Multalin, and PRRP), and model building (MODELLER). Models were generated using the following proteins: VC1, Fab light chain in complex with human factor 8 (PDB 1IQD); V, mouse junction adhesion molecule (PDB 1F97); C1, human B7-1 (PDB 1DR9); C2, human APEP-1 (PDB 1U2H). Side chains were repacked and energy minimized using SCWRL 3.0 (144). PROCHECK was used to assess the stereochemical properties of each model (145). To analyze the surface area buried in the V-C1 contacts, the model was split after the linker. Fast Connolly surfaces were generated in the program Sybyl (Tripos). Surfaces for the homology model included VC1 (25-222), V domain (25-121), and C1 domain (122-222). Surfaces for the x-ray crystal structure included VC1 (22-233), V domain (22-118) and C1 domain (119-233). Residue specific solvent accessible surface area was calculated with Naccess.

CML Synthesis

Synthesis of N(alpha)-Boc-N(epsilon)-(carboxymethyl)lysine

The following syntheses were performed by Dr. Guenter Fritz at the University of Konstanz. N(alpha)-tert-Boc-L-lysine (Figure 6.1A) (985 mg, 4 mmol) and sodium glyoxylate (Figure 6.1B) (684 mg, 6 mmol) were dissolved in 8 ml 0.5 M sodium citrate. The solution was adjusted to pH 6 with 1M NaOH and stirred for 1 h at 65 °C. Subsequently, $NaCNBH_3$ (384 mg, 6.12 mmol) was added and the solution was stirred continuously for 12.5 h at 65 °C. Samples were taken regularly and analyzed using thin

layer chromatography (TLC). After 14 h of reaction time TLC analysis could not detect any remaining N(alpha)-tert-Boc-L-lysine in the solution. After completion of the reaction, the solution was concentrated in a rotary evaporator. The remaining solution was dissolved in methanol and desalted using a short column packed with cellulose. After washing with additional methanol, the collected filtered solution was dried. In a second desalting step, the residue was dissolved in methanol and separated through a long column packed with cellulose. Yield is 1086 mg (3.56 mmol, 88.9%).

Synthesis of N(epsilon)-(carboxymethyl)-lysine

N(alpha)-Boc-N(epsilon)-(carboxymethyl)lysine (Figure 6.1D) (1086 mg, 3.56 mmol) was dissolved in 10 ml of 3 M HCl and heated for 1 h at 70 °C. After removal of HCl in a vacuum concentrator, the residue was re-crystallized in water ethanol. Yield is 5,728 mg (2.62 mmol; 73.79%).

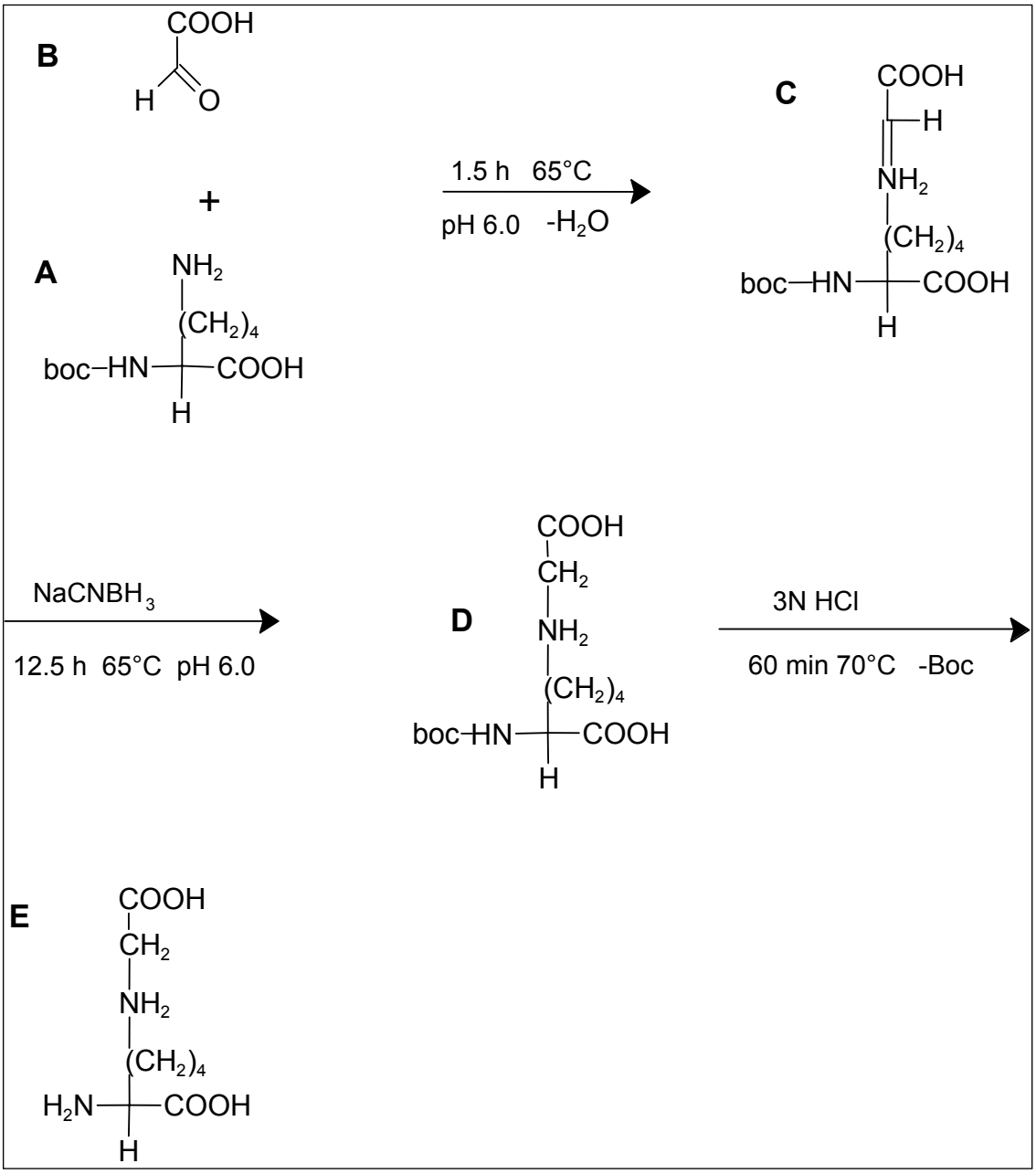


Figure 6.1: Reaction scheme for synthesis of N(epsilon)-(carboxymethyl)lysine. Synthesis steps are described in the text.

APPENDIX A

MULTIPLE SEQUENCE ALIGNMENT OF SRAGE PROTEINS

Figure A.1 shows the full sequence alignment of sRAGE proteins from human, bovine, rat, and mouse. Sequences for human (Q15109), bovine (Q28173), rat (Q63495) and mouse (Q62151) sRAGE were aligned using ClustalX run on a Windows PC. This is the alignment for which discussion in this dissertation are based.

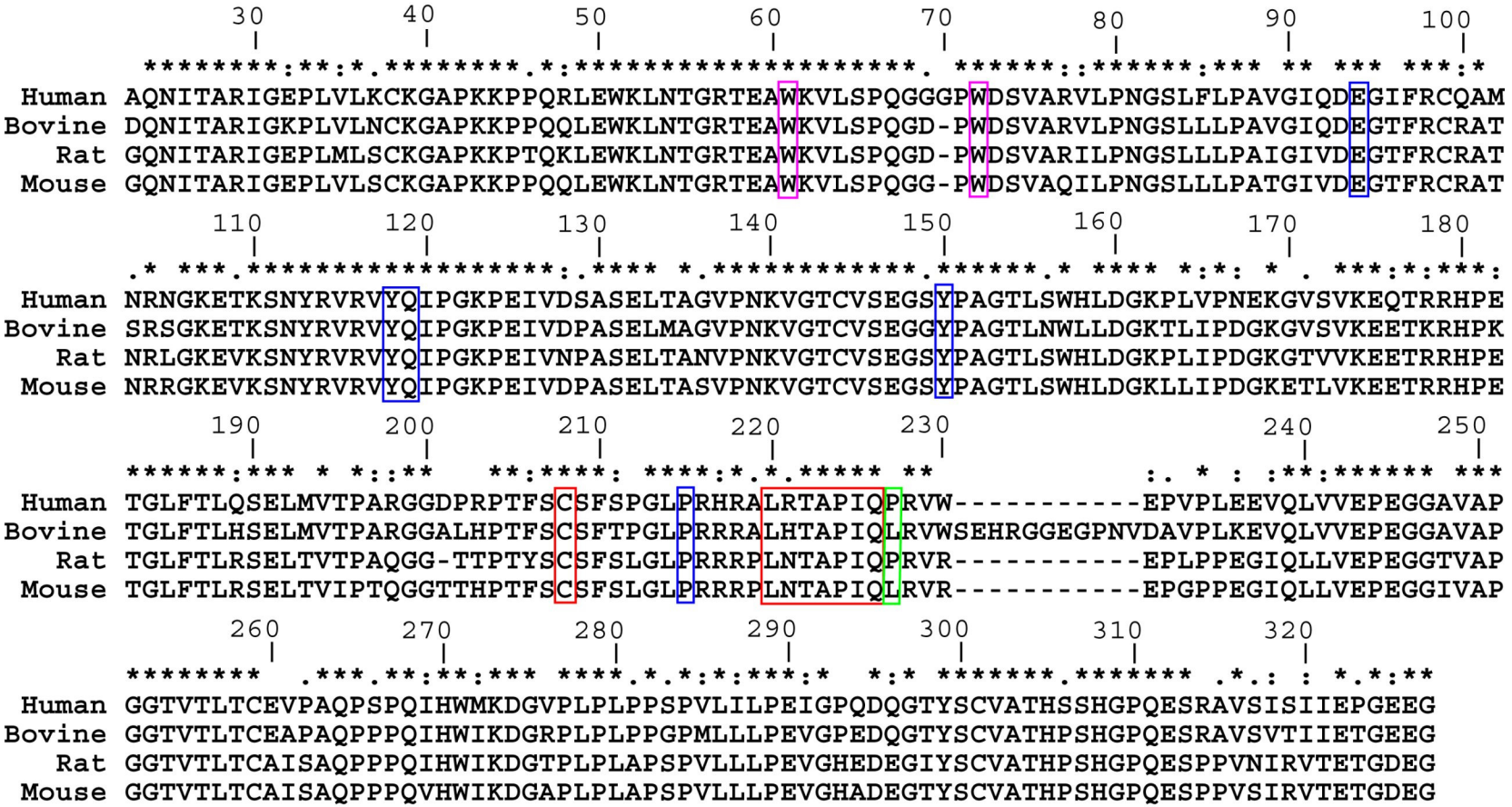


Figure A.1: Multiple sequence alignment of sRAGE from four species. The numbering corresponds to the human sequence. The residues implicated in the strand *g* bulge as described in Chapter III are contained within red boxes. Position 227 which may contribute to strand *g* swapping as described in Chapter III is contained within a green box. Residues implicated in the V-C1 linker as described in Chapter III are contained within blue boxes. Tryptophan residues discussed in Chapter IV are contained within violet boxes.

APPENDIX B

SUMMARY OF CRYSTALLIZATION STATISTICS FOR VC1

The following tables for statistics on the crystal structure of VC1 were provided by Dr. Guenter Fritz from the University of Konstanz.

Table B.1: Statistics of data collection and Zn MAD phasing.

Data set	**Peak**	**Inflection**	**Remote High**
Wavelength (Å)	1.282	1.283	1.008
Resolution limits (Å)	30-2.0 (2.1-2.0)	30-2.0 (2.1-2.0)	30-2.0 (2.1-2.0)
Unique reflections	33711 (4577)	33708 (4581)	33660 (4588)
Redundancy	3.7 (3.3)	3.7 (3.3)	3.8 (3.7)
Completeness (%)	99.9 (99.7)	99.9 (99.8)	99.7 (100.0)
I/σ	9.4 (2.1)	9.7 (2.2)	10.7 (3.8)
R_{sym} (%)	9.2 (55.5)	8.4 (51.5)	7.7 (31.9)
R_{meas} (%)	10.4 (64.8)	9.8 (61.4)	8.9 (37.3)
B-factor (Wilson plot)	35.1	35.1	33.2

Table B.2: Statistics for highest resolution dataset.

Resolution limits (Å)	30-2.0 (2.0-1.8)
Wavelength (Å)	1.008
Reflections	174312 (47326)
Unique reflections	24993 (6640)
Completeness (%)	99.7 (99.9)
I/σ	11.2 (2.8)
R_{merge} (%)	9.7 (43.9)

APPENDIX C

15N-1H HSQC SPECTRA OF OPTIMIZED SRAGE CONSTRUCTS

The initial sRAGE constructs used in Chapters II, III, and IV were designed based on a combination of multiple sequence alignments and secondary structure prediction as described in Chapter II. ^{15}N-^{1}H HSQC spectra of both VC1 and V revealed ~10 resonances with very sharp linewidths consistent with high flexibility. The x-ray crystal structure presented in Chapter III confirmed the existence of highly flexible C-terminal extensions. Based on analysis of the x-ray crystal structure, optimized VC1, V, and C1 constructs were designed and are listed in Table 6.1 as VC1-b, V-b, and C1-b. Overall, both the VC1 and V constructs behaved exactly as the originals through all steps of purification. The ^{15}N-^{1}H HSQC spectra of the new VC1 (Figure C.1) and V (Figure C.2) constructs show these are valid reagents for further structural and macromolecular interaction studies. The C1 construct showed lower expression and its ^{15}N-^{1}H HSQC spectrum was characteristic of a completely unfolded protein. However, the His_6 tag had been removed by thrombin cleavage. Previous work on the original C1 construct (122-243) revealed significant secondary cleavage during thrombin incubation. While preliminary work on the new C1 construct (119-233) suggested it was stable to secondary cleavage, the final NMR sample showed signs of degradation. Thus, these data should be reacquired with this issue in mind. Lastly, Figure C.4 compares the new VC1 and V constructs showing significant chemical shift perturbations, confirming the structural interface described in Chapters II and III.

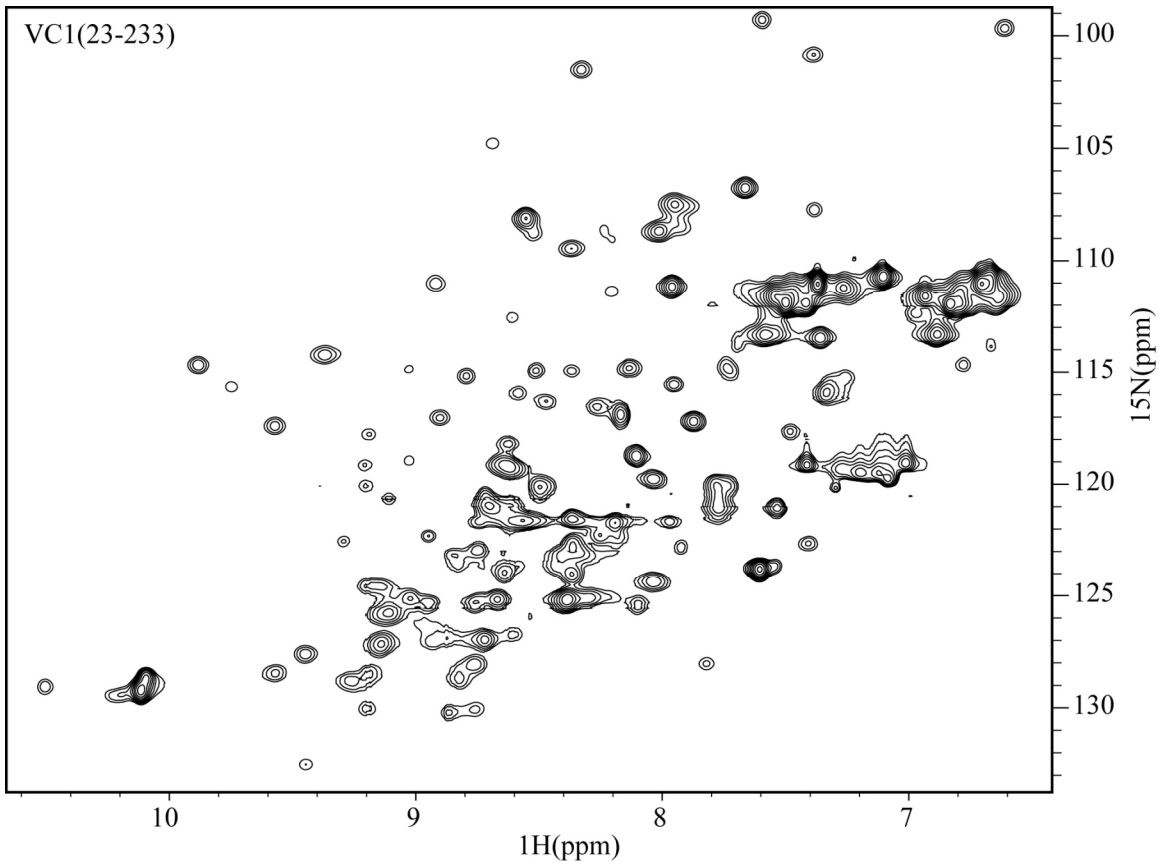


Figure C.1: ^{15}N-^{1}H HSQC spectra of VC1(23-233). Data were collected at 600 MHz and 298K in 30 mM potassium phosphate at pH 6.0.

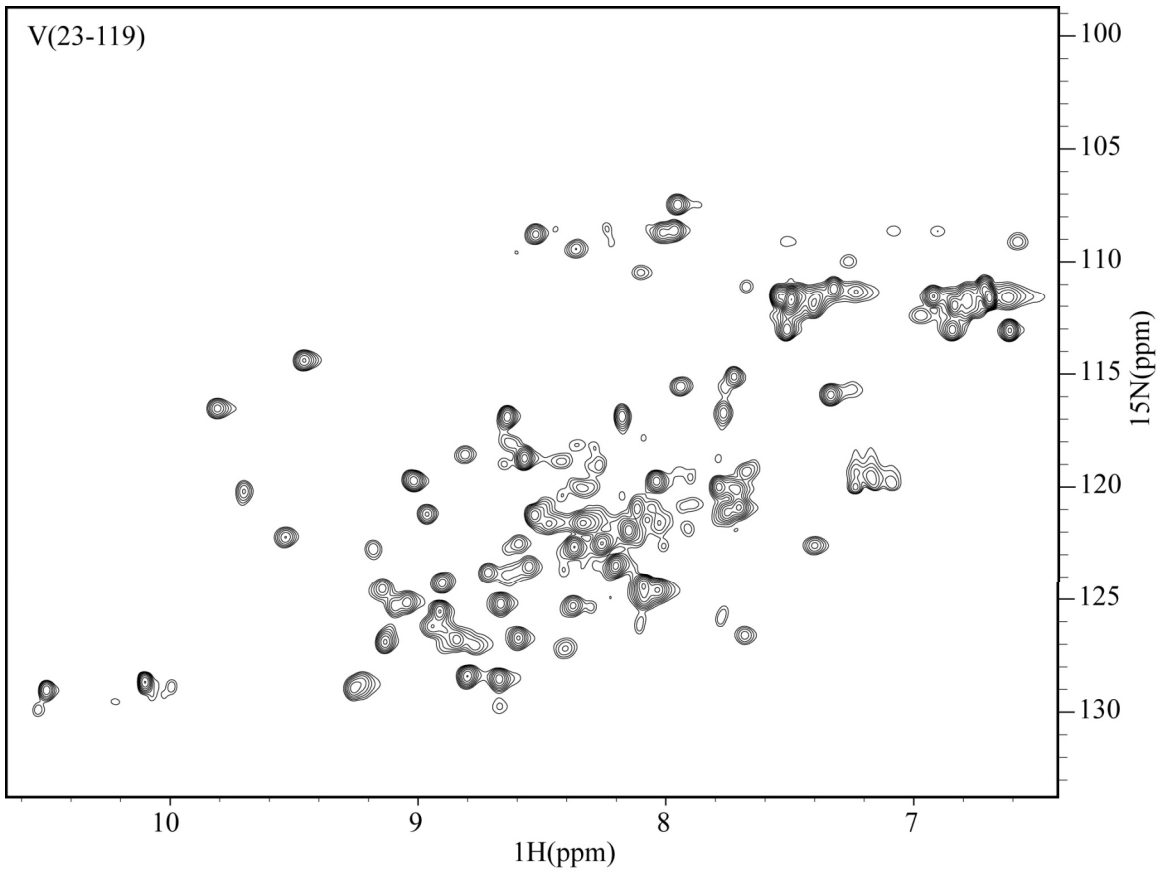


Figure C.2: ^{15}N-^{1}H HSQC spectra of V(23-119). Data were collected at 600 MHz and 298K in 30 mM potassium phosphate at pH 6.0.

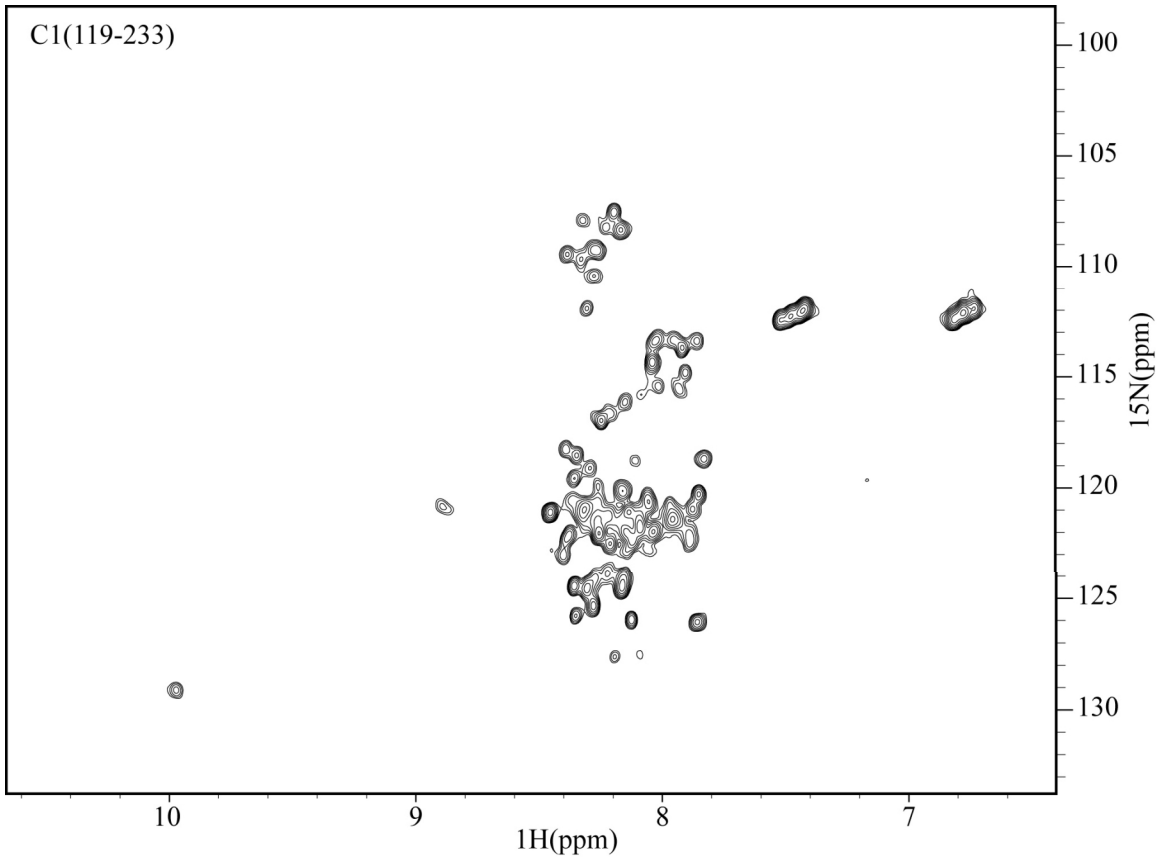


Figure C.3: ^{15}N-^{1}H HSQC spectra of C1(119-233). Data were collected at 600 MHz and 298K in 30 mM potassium phosphate at pH 6.0.

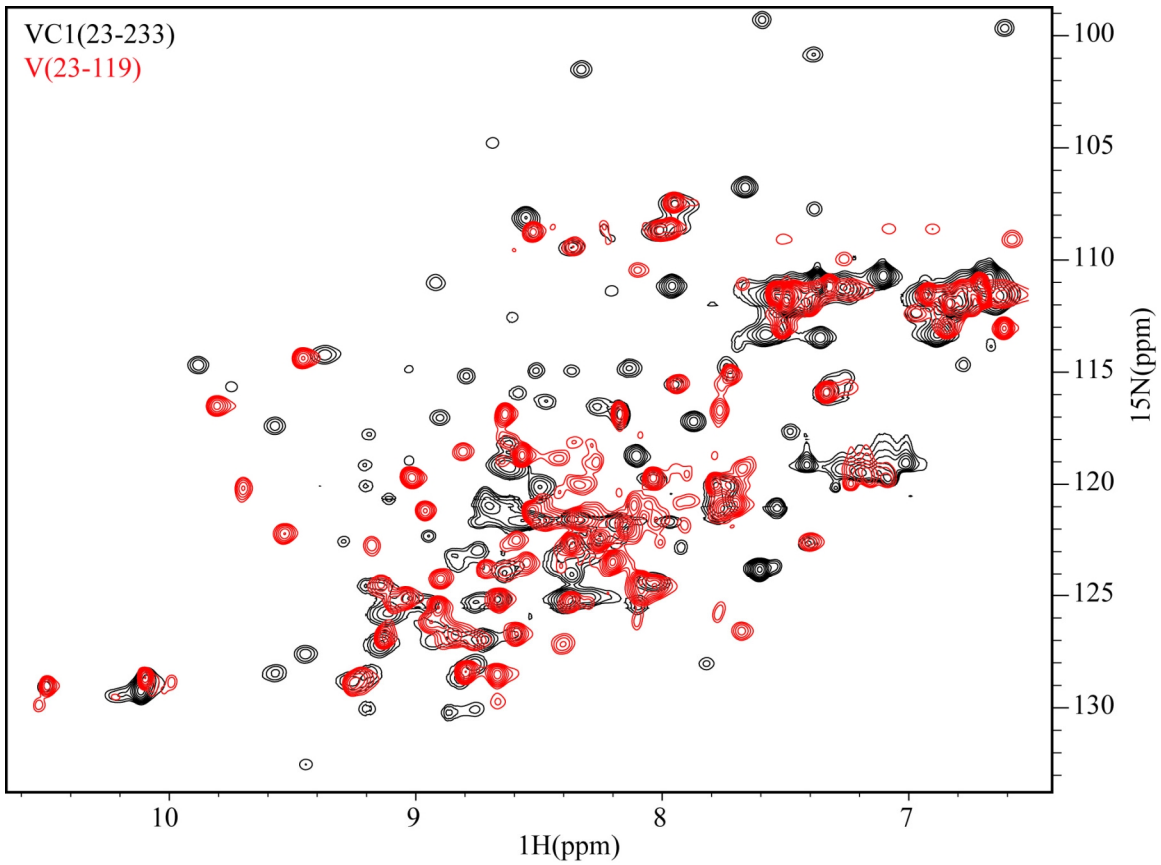


Figure C.4: Optimized V shows structural changes in the context of VC1. Overlay of 15N-1H HSQC spectra of V(23-119) in red and VC1(23-233) in black. The same peaks perturbed in Figure 2.11 are perturbed in this analysis.

APPENDIX D

TABLE OF BACKBONE ATOM CHEMICAL SHIFTS FOR C2

Table D.1: ^{1}H, ^{15}N, ^{13}C backbone chemical shifts for C2.

Res Num	N	NH	CA	CB
E236			53.85	27.92
E237	120.97	8.19	55.03	28.60
V238	118.74	7.75	59.40	30.74
Q239	126.22	9.01	52.17	28.90
L240	125.34	8.58	51.74	42.21
V241	128.63	9.23	59.18	31.79
V242	125.59	8.47	57.47	32.56
E243	126.59	8.44	50.21	30.63
P244			60.37	32.07
E245	121.28	8.65	56.35	27.24
G246	110.49	8.64	43.01	
G247	106.98	8.19	44.02	
A248	120.36	7.28	47.90	17.52
V249	118.23	8.30	56.58	33.98
A250	126.03	8.15	48.13	14.82
P251			61.57	28.76
G252	112.57	9.07	43.06	
G253	108.95	8.30	42.17	
T254	113.29	7.90	58.90	69.24
V255	126.22	8.12	57.81	32.67
T256	121.46	8.75	59.74	67.97
L257	129.55	9.13	51.79	41.40
T258	116.94	8.99	59.48	68.38
C259	127.76	9.35	52.63	43.25
E260	124.67	9.04	53.04	29.87
V261	123.13	8.64	55.90	
P262			62.20	29.60
A263	118.10	8.43	50.63	16.47
Q264	117.27	7.66		27.96
P265				
S266				

Table D.1: continued

Res Num	N	NH	CA	CB
P267			60.42	29.84
Q268	121.27	8.67	53.04	27.00
I269	126.58	8.23	56.52	36.59
H270	120.81	8.55	51.62	29.50
W271	120.96	8.77	55.29	30.73
M272	120.98	9.20	51.21	33.05
K273	119.72	8.57	51.97	33.72
D274	128.19	9.87	53.21	37.33
G275	101.03	8.12	42.88	
V276	121.14	7.78	56.43	32.30
P277				
L278				
P279				
L280				
P281				
P282				
S283				
P284			62.05	29.96
V285	115.02	7.53	57.39	33.19
L286	131.09	9.05	51.37	39.37
I287	125.93	8.13	57.69	37.26
L288	128.46	8.84	48.29	39.18
P289			59.89	30.16
E290	118.30	7.77	53.14	26.28
I291	119.79	8.47	60.51	35.99
G292	114.12	8.92	41.52	
P293			63.12	29.41
Q294	115.52	9.07	55.21	24.97
D295	116.83	8.15	52.91	40.51
Q296	117.75	7.23	54.93	28.59

Table D.1: continued

Res Num	N	NH	CA	CB
G297	108.65	8.16	42.63	
T298	116.06	8.85	58.84	67.34
Y299	127.02	9.52	54.63	40.56
S300	111.07	8.75	54.78	63.15
C301	114.10	8.58	50.53	45.37
V302	118.92	8.52	57.74	32.40
A303	128.73	8.90	47.59	19.22
T304	115.29	8.47	59.60	67.57
H305	121.48	8.47		
S306				
S307				
H308			52.71	27.41
G309	108.36	8.09	42.27	
P310				
Q311				
E312			52.07	29.98
S313	120.60	8.96	56.25	64.36
R314	116.74	8.51	53.85	27.50
A315	124.39	8.32	48.12	18.12
V316	119.01	9.13	57.63	32.55
S317	120.45	8.53	55.40	61.24
I318	127.23	9.05	55.79	35.02
S319	122.14	8.73	54.13	63.98
I320	119.33	8.08	56.56	36.62
I321	126.96	8.19	57.96	37.01
E322	126.67	8.37	51.61	27.27
P323			60.99	29.76
G324	109.11	8.40	42.60	
E325	120.10	8.24	54.03	27.93
E326	121.31	8.48	54.22	27.94
G327	115.64	7.93	43.71	

BIBLIOGRAPHY

1. Schmidt, A. M., Vianna, M., Gerlach, M., Brett, J., Ryan, J., Kao, J., Esposito, C., Hegarty, H., Hurley, W., Clauss, M., Wang, F., Pan, Y. C., Tsang, T. C., and Stern, D. (1992) Isolation and characterization of two binding proteins for advanced glycosylation end products from bovine lung which are present on the endothelial cell surface. *J Biol Chem* **267**, 14987-14997.

2. Neeper, M., Schmidt, A. M., Brett, J., Yan, S. D., Wang, F., Pan, Y. C., Elliston, K., Stern, D., and Shaw, A. (1992) Cloning and expression of a cell surface receptor for advanced glycosylation end products of proteins. *J Biol Chem* **267**, 14998-15004.

3. Stern, D. M., Yan, S. D., Yan, S. F., and Schmidt, A. M. (2002) Receptor for advanced glycation endproducts (RAGE) and the complications of diabetes. *Ageing Res Rev* **1**, 1-15.

4. Schmidt, A. M., Yan, S. D., Yan, S. F., and Stern, D. M. (2000) The biology of the receptor for advanced glycation end products and its ligands. *Biochim Biophys Acta* **1498**, 99-111.

5. Stern, D., Yan, S. D., Yan, S. F., and Schmidt, A. M. (2002) Receptor for advanced glycation endproducts: a multiligand receptor magnifying cell stress in diverse pathologic settings. *Adv Drug Deliv Rev* **54**, 1615-1625.

6. Kikuchi, S., Shinpo, K., Takeuchi, M., Yamagishi, S., Makita, Z., Sasaki, N., and Tashiro, K. (2003) Glycation--a sweet tempter for neuronal death. *Brain Res Brain Res Rev* **41**, 306-323.

7. Oimomi, M., Maeda, Y., Hata, F., Kitamura, Y., Matsumoto, S., Hatanaka, H., and Baba, S. (1988) A study of the age-related acceleration of glycation of tissue proteins in rats. *J Gerontol* **43**, B98-101.

8. Vishwanath, V., Frank, K. E., Elmets, C. A., Dauchot, P. J., and Monnier, V. M. (1986) Glycation of skin collagen in type I diabetes mellitus. Correlation with long-term complications. *Diabetes* **35**, 916-921.

9. Monnier, V. M., Vishwanath, V., Frank, K. E., Elmets, C. A., Dauchot, P., and Kohn, R. R. (1986) Relation between complications of type I diabetes mellitus and collagen-linked fluorescence. *N Engl J Med* **314**, 403-408.

10. Monnier, V. M. and Cerami, A. (1981) Nonenzymatic browning in vivo: possible process for aging of long-lived proteins. *Science* **211**, 491-493.

11. Schnider, S. L. and Kohn, R. R. (1980) Glucosylation of human collagen in aging and diabetes mellitus. *J Clin Invest* **66**, 1179-1181.

12. Monnier, V. M., Kohn, R. R., and Cerami, A. (1984) Accelerated age-related browning of human collagen in diabetes mellitus. *Proc Natl Acad Sci U S A* **81**, 583-587.

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

14. Vitek, M. P., Bhattacharya, K., Glendening, J. M., Stopa, E., Vlassara, H., Bucala, R., Manogue, K., and Cerami, A. (1994) Advanced glycation end products contribute to amyloidosis in Alzheimer disease. *Proc Natl Acad Sci U S A* **91**, 4766-4770.

15. Smith, M. A., Taneda, S., Richey, P. L., Miyata, S., Yan, S. D., Stern, D., Sayre, L. M., Monnier, V. M., and Perry, G. (1994) Advanced Maillard reaction end products are associated with Alzheimer disease pathology. *Proc Natl Acad Sci U S A* **91**, 5710-5714.

16. Selkoe, D. J., Salazar, F. J., Abraham, C., and Kosik, K. S. (1982) Huntington's disease: changes in striatal proteins reflect astrocytic gliosis. *Brain Res* **245**, 117-125.

17. Radoff, S., Cerami, A., and Vlassara, H. (1990) Isolation of surface binding protein specific for advanced glycosylation end products from mouse macrophage-derived cell line RAW 264.7. *Diabetes* **39**, 1510-1518.

18. Radoff, S., Vlassara, H., and Cerami, A. (1988) Characterization of a solubilized cell surface binding protein on macrophages specific for proteins modified nonenzymatically by advanced glycosylated end products. *Arch Biochem Biophys* **263**, 418-423.

19. Esposito, C., Gerlach, H., Brett, J., Stern, D., and Vlassara, H. (1989) Endothelial receptor-mediated binding of glucose-modified albumin is associated with increased monolayer permeability and modulation of cell surface coagulant properties. *J Exp Med* **170**, 1387-1407.

20. Schmidt, A. M., Hori, O., Brett, J., Yan, S. D., Wautier, J. L., and Stern, D. (1994) Cellular receptors for advanced glycation end products. Implications for induction of oxidant stress and cellular dysfunction in the pathogenesis of vascular lesions. *Arteriosclerosis & Thrombosis* **14**, 1521-1528.

21. Baeuerle, P. A. and Henkel, T. (1994) Function and activation of NF-kappa B in the immune system. *Annu Rev Immunol* **12**, 141-179.

22. Schreck, R., Rieber, P., and Baeuerle, P. A. (1991) Reactive oxygen intermediates as apparently widely used messengers in the activation of the NF-kappa B transcription factor and HIV-1. *Embo J* **10**, 2247-2258.

23. Huttunen, H. J., Fages, C., and Rauvala, H. (1999) Receptor for advanced glycation end products (RAGE)-mediated neurite outgrowth and activation of NF-kappaB require the cytoplasmic domain of the receptor but different downstream signaling pathways. *J Biol Chem* **274**, 19919-19924.

24. Lander, H. M., Tauras, J. M., Ogiste, J. S., Hori, O., Moss, R. A., and Schmidt, A. M. (1997) Activation of the receptor for advanced glycation end products triggers a p21(ras)-dependent mitogen-activated protein kinase pathway regulated by oxidant stress. *Journal of Biological Chemistry* **272**, 17810-17814.

25. Yeh, C. H., Sturgis, L., Haidacher, J., Zhang, X. N., Sherwood, S. J., Bjercke, R. J., Juhasz, O., Crow, M. T., Tilton, R. G., and Denner, L. (2001) Requirement for p38 and p44/p42 mitogen-activated protein kinases in RAGE-mediated nuclear factor-kappaB transcriptional activation and cytokine secretion. *Diabetes* **50**, 1495-1504.

26. Brett, J., Schmidt, A. M., Yan, S. D., Zou, Y. S., Weidman, E., Pinsky, D., Nowygrod, R., Neeper, M., Przysiecki, C., and Shaw, A. (1993) Survey of the distribution of a newly characterized receptor for advanced glycation end products in tissues. *Am J Pathol* **143**, 1699-1712.

27. Hori, O., Brett, J., Slattery, T., Cao, R., Zhang, J., Chen, J. X., Nagashima, M., Lundh, E. R., Vijay, S., Nitecki, D., Morser, J., Stern, D., and Schmidt, A. M. (1995) The receptor for advanced glycation end products (RAGE) is a cellular binding site for amphoterin. Mediation of neurite outgrowth and co-expression of rage and amphoterin in the developing nervous system. *J Biol Chem* **270**, 25752-25761.

28. Collins, F. (1978) Induction of neurite outgrowth by a conditioned-medium factor bound to the culture substratum. *Proc Natl Acad Sci U S A* **75**, 5210-5213.

29. Rauvala, H. and Pihlaskari, R. (1987) Isolation and some characteristics of an adhesive factor of brain that enhances neurite outgrowth in central neurons. *J Biol Chem* **262**, 16625-16635.

30. Bianchi, M. E., Beltrame, M., and Paonessa, G. (1989) Specific recognition of cruciform DNA by nuclear protein HMG1. *Science* **243**, 1056-1059.

31. Einck, L. and Bustin, M. (1985) The intracellular distribution and function of the high mobility group chromosomal proteins. *Exp Cell Res* **156**, 295-310.

32. Ge, H. and Roeder, R. G. (1994) The high mobility group protein HMG1 can reversibly inhibit class II gene transcription by interaction with the TATA-binding protein. *J Biol Chem* **269**, 17136-17140.

33. Onate, S. A., Prendergast, P., Wagner, J. P., Nissen, M., Reeves, R., Pettijohn, D. E., and Edwards, D. P. (1994) The DNA-bending protein HMG-1 enhances progesterone receptor binding to its target DNA sequences. *Mol Cell Biol* **14**, 3376-3391.

34. Reeck, G. R., Isackson, P. J., and Teller, D. C. (1982) Domain structure in high molecular weight high mobility group nonhistone chromatin proteins. *Nature* **300**, 76-78.

35. Cary, P. D., Turner, C. H., Mayes, E., and Crane-Robinson, C. (1983) Conformation and domain structure of the non-histone chromosomal proteins, HMG 1 and 2. Isolation of two folded fragments from HMG 1 and 2. *Eur J Biochem* **131**, 367-374.

36. Hardman, C. H., Broadhurst, R. W., Raine, A. R., Grasser, K. D., Thomas, J. O., and Laue, E. D. (1995) Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry* **34**, 16596-16607.

37. Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R., Laue, E. D., and Thomas, J. O. (1993) Structure of the HMG box motif in the B-domain of HMG1. *Embo J* **12**, 1311-1319.

38. Huttunen, H. J., Fages, C., Kuja-Panula, J., Ridley, A. J., and Rauvala, H. (2002) Receptor for advanced glycation end products-binding COOH-terminal motif of amphoterin inhibits invasive migration and metastasis. *Cancer Res* **62**, 4805-4811.

39. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612.

40. Yan, S. D., Chen, X., Fu, J., Chen, M., Zhu, H., Roher, A., Slattery, T., Zhao, L., Nagashima, M., Morser, J., Migheli, A., Nawroth, P., Stern, D., and Schmidt, A. M. (1996) RAGE and amyloid-beta peptide neurotoxicity in Alzheimer's disease.[see comment]. *Nature* **382**, 685-691.

41. Haass, C. and Selkoe, D. J. (1993) Cellular processing of beta-amyloid precursor protein and the genesis of amyloid beta-peptide. *Cell* **75**, 1039-1042.

42. Arancio, O., Zhang, H. P., Chen, X., Lin, C., Trinchese, F., Puzzo, D., Liu, S., Hegde, A., Yan, S. F., Stern, A., Luddy, J. S., Lue, L. F., Walker, D. G., Roher, A., Buttini, M., Mucke, L., Li, W., Schmidt, A. M., Kindy, M., Hyslop, P. A., Stern, D. M., and Du Yan, S. S. (2004) RAGE potentiates Abeta-induced perturbation of neuronal function in transgenic mice. *Embo J* **23**, 4096-4105.

43. Hofmann, M. A., Drury, S., Fu, C., Qu, W., Taguchi, A., Lu, Y., Avila, C., Kambham, N., Bierhaus, A., Nawroth, P., Neurath, M. F., Slattery, T., Beach, D., McClary, J., Nagashima, M., Morser, J., Stern, D., and Schmidt, A. M. (1999) RAGE mediates a novel proinflammatory axis: a central cell surface receptor for S100/calgranulin polypeptides. *Cell* **97**, 889-901.

44. Huttunen, H. J., Kuja-Panula, J., Sorci, G., Agneletti, A. L., Donato, R., and Rauvala, H. (2000) Coregulation of neurite outgrowth and cell survival by amphoterin and S100 proteins through receptor for advanced glycation end products (RAGE) activation. *J Biol Chem* **275**, 40096-40105.

45. Kligman, D. and Marshak, D. R. (1985) Purification and characterization of a neurite extension factor from bovine brain. *Proc Natl Acad Sci U S A* **82**, 7136-7139.

46. Winningham-Major, F., Staecker, J. L., Barger, S. W., Coats, S., and Van Eldik, L. J. (1989) Neurite extension and neuronal survival activities of recombinant S100 beta proteins that differ in the content and position of cysteine residues. *J Cell Biol* **109**, 3063-3071.

47. Arumugam, T., Ramachandran, V., and Logsdon, C. D. (2006) Effect of cromolyn on S100P interactions with RAGE and pancreatic cancer growth and invasion in mouse models. *J Natl Cancer Inst* **98**, 1806-1818.

48. Arumugam, T., Simeone, D. M., Schmidt, A. M., and Logsdon, C. D. (2004) S100P stimulates cell proliferation and survival via receptor for activated glycation end products (RAGE). *J Biol Chem* **279**, 5059-5065.

49. Yammani, R. R., Carlson, C. S., Bresnick, A. R., and Loeser, R. F. (2006) Increase in production of matrix metalloproteinase 13 by human articular chondrocytes due to stimulation with S100A4: Role of the receptor for advanced glycation end products. *Arthritis Rheum* **54**, 2901-2911.

50. Cecil, D. L., Johnson, K., Rediske, J., Lotz, M., Schmidt, A. M., and Terkeltaub, R. (2005) Inflammation-induced chondrocyte hypertrophy is driven by receptor for advanced glycation end products. *J Immunol* **175**, 8296-8302.

51. Ishihara, K., Tsutsumi, K., Kawane, S., Nakajima, M., and Kasaoka, T. (2003) The receptor for advanced glycation end-products (RAGE) directly binds to ERK by a D-domain-like docking site. *FEBS Lett* **550**, 107-113.

52. Li, J. and Schmidt, A. M. (1997) Characterization and functional analysis of the promoter of RAGE, the receptor for advanced glycation end products. *Journal of Biological Chemistry* **272**, 16498-16506.

53. Schmidt, A. M., Yan, S. D., and Stern, D. M. (1995) The dark side of glucose. *Nat Med* **1**, 1002-1004.

54. Schmidt, A. M., Yan, S. D., Yan, S. F., and Stern, D. M. (2001) The multiligand receptor RAGE as a progression factor amplifying immune and inflammatory responses. *Journal of Clinical Investigation* **108**, 949-955.

55. Hudson, B. I., Bucciarelli, L. G., Wendt, T., Sakaguchi, T., Lalla, E., Qu, W., Lu, Y., Lee, L., Stern, D. M., Naka, Y., Ramasamy, R., Yan, S. D., Yan, S. F., D'Agati, V., and Schmidt, A. M. (2003) Blockade of receptor for advanced glycation endproducts: a new target for therapeutic intervention in diabetic complications and inflammatory disorders. *Archives of Biochemistry & Biophysics* **419**, 80-88.

56. Plump, A. S., Smith, J. D., Hayek, T., Aalto-Setala, K., Walsh, A., Verstuyft, J. G., Rubin, E. M., and Breslow, J. L. (1992) Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by homologous recombination in ES cells. *Cell* **71**, 343-353.

57. Park, L., Raman, K. G., Lee, K. J., Lu, Y., Ferran, L. J., Jr., Chow, W. S., Stern, D., and Schmidt, A. M. (1998) Suppression of accelerated diabetic atherosclerosis by the soluble receptor for advanced glycation endproducts. *Nat Med* **4**, 1025-1031.

58. Taguchi, A., Blood, D. C., del Toro, G., Canet, A., Lee, D. C., Qu, W., Tanji, N., Lu, Y., Lalla, E., Fu, C., Hofmann, M. A., Kislinger, T., Ingram, M., Lu, A., Tanaka, H., Hori, O., Ogawa, S., Stern, D. M., and Schmidt, A. M. (2000) Blockade of RAGE-amphoterin signalling suppresses tumour growth and metastases. *Nature* **405**, 354-360.

59. Zlokovic, B. (1997) Can blood-brain barrier play a role in the development of cerebral amyloidosis and Alzheimer's disease pathology. *Neurobiol Dis* **4**, 23-26.

60. Mackic, J. B., Stins, M., McComb, J. G., Calero, M., Ghiso, J., Kim, K. S., Yan, S. D., Stern, D., Schmidt, A. M., Frangione, B., and Zlokovic, B. V. (1998) Human blood-brain barrier receptors for Alzheimer's amyloid-beta 1- 40. Asymmetrical binding, endocytosis, and transcytosis at the apical side of brain microvascular endothelial cell monolayer. *J Clin Invest* **102**, 734-743.

61. Hsiao, K., Chapman, P., Nilsen, S., Eckman, C., Harigaya, Y., Younkin, S., Yang, F., and Cole, G. (1996) Correlative memory deficits, Abeta elevation, and amyloid plaques in transgenic mice. *Science* **274**, 99-102.

62. Deane, R., Du Yan, S., Submamaryan, R. K., LaRue, B., Jovanovic, S., Hogg, E., Welch, D., Manness, L., Lin, C., Yu, J., Zhu, H., Ghiso, J., Frangione, B., Stern, A., Schmidt, A. M., Armstrong, D. L., Arnold, B., Liliensiek, B., Nawroth, P., Hofman, F., Kindy, M., Stern, D., and Zlokovic, B. (2003) RAGE mediates amyloid-beta peptide transport across the blood-brain barrier and accumulation in brain. *Nat Med* **9**, 907-913.

63. Kuhn, R., Lohler, J., Rennick, D., Rajewsky, K., and Muller, W. (1993) Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75**, 263-274.

64. Bhattacharya, S., Bunick, C. G., and Chazin, W. J. (2004) Target selectivity in EF-hand calcium binding proteins. *Biochim Biophys Acta* **1742**, 69-79.

65. Kretsinger, R. H. and Nockolds, C. E. (1973) Carp muscle calcium-binding protein. II. Structure determination and general description. *J Biol Chem* **248**, 3313-3326.

66. Van Eldik, L. J. and Watterson, D. M. (1998) *Calmodulin and Signal Transduction*, Academic Press, Boston, MA.

67. Linse, S., Helmersson, A., and Forsen, S. (1991) Calcium binding to calmodulin and its globular domains. *J Biol Chem* **266**, 8050-8054.

68. Potts, B. C., Smith, J., Akke, M., Macke, T. J., Okazaki, K., Hidaka, H., Case, D. A., and Chazin, W. J. (1995) The structure of calcyclin reveals a novel homodimeric fold for S100 Ca(2+)-binding proteins. *Nat Struct Biol* **2**, 790-796.

69. Szebenyi, D. M. and Moffat, K. (1986) The refined structure of vitamin D-dependent calcium-binding protein from bovine intestine. Molecular details, ion binding, and implications for the structure of other calcium-binding proteins. *J Biol Chem* **261**, 8761-8777.

70. Maler, L., Sastry, M., and Chazin, W. J. (2002) A structural basis for S100 protein specificity derived from comparative analysis of apo and Ca(2+)-calcyclin. *J Mol Biol* **317**, 279-290.

71. Sastry, M., Ketchem, R. R., Crescenzi, O., Weber, C., Lubienski, M. J., Hidaka, H., and Chazin, W. J. (1998) The three-dimensional structure of Ca(2+)-bound calcyclin: implications for Ca(2+)-signal transduction by S100 proteins. *Structure* **6**, 223-231.

72. Fritz, G., Mittl, P. R., Vasak, M., Grutter, M. G., and Heizmann, C. W. (2002) The crystal structure of metal-free human EF-hand protein S100A3 at 1.7-A resolution. *J Biol Chem* **277**, 33092-33098.

73. Heizmann, C. W., Fritz, G., and Schafer, B. W. (2002) S100 proteins: structure, functions and pathology. *Front Biosci* **7**, d1356-1368.

74. Marenholz, I., Heizmann, C. W., and Fritz, G. (2004) S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). *Biochem Biophys Res Commun* **322**, 1111-1122.

75. Rety, S., Sopkova, J., Renouard, M., Osterloh, D., Gerke, V., Tabaries, S., Russo-Marie, F., and Lewit-Bentley, A. (1999) The crystal structure of a complex of p11 with the annexin II N-terminal peptide. *Nat Struct Biol* **6**, 89-95.

76. Koch, M., Bhattacharya, S., Kehl, T., Gimona, M., Vasak, M., Chazin, W., Heizmann, C. W., Kroneck, P. M., and Fritz, G. (2007) Implications on zinc binding to S100A2. *Biochim Biophys Acta* **1773**, 457-470.

77. Wilder, P. T., Varney, K. M., Weiss, M. B., Gitti, R. K., and Weber, D. J. (2005) Solution structure of zinc- and calcium-bound rat S100B as determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **44**, 5690-5702.

78. Loomans, H. J., Hahn, B. L., Li, Q. Q., Phadnis, S. H., and Sohnle, P. G. (1998) Histidine-based zinc-binding sequences and the antimicrobial activity of calprotectin. *J Infect Dis* **177**, 812-814.

79. Steinbakk, M., Naess-Andresen, C. F., Lingaas, E., Dale, I., Brandtzaeg, P., and Fagerhol, M. K. (1990) Antimicrobial actions of calcium binding leucocyte L1 protein, calprotectin. *Lancet* **336**, 763-765.

80. Sohnle, P. G., Hunter, M. J., Hahn, B., and Chazin, W. J. (2000) Zinc-reversible antimicrobial activity of recombinant calprotectin (migration inhibitory factor-related proteins 8 and 14). *J Infect Dis* **182**, 1272-1275.

81. Sampson, B., Fagerhol, M. K., Sunderkotter, C., Golden, B. E., Richmond, P., Klein, N., Kovar, I. Z., Beattie, J. H., Wolska-Kusnierz, B., Saito, Y., and Roth, J. (2002) Hyperzincaemia and hypercalprotectinaemia: a new disorder of zinc metabolism. *Lancet* **360**, 1742-1745.

82. Barger, S. W., Wolchok, S. R., and Van Eldik, L. J. (1992) Disulfide-linked S100 beta dimers and signal transduction. *Biochim Biophys Acta* **1160**, 105-112.

83. Moroz, O. V., Antson, A. A., Murshudov, G. N., Maitland, N. J., Dodson, G. G., Wilson, K. S., Skibshoj, I., Lukanidin, E. M., and Bronstein, I. B. (2001) The

three-dimensional structure of human S100A12. *Acta Crystallogr D Biol Crystallogr* **57**, 20-29.

84. Moroz, O. V., Dodson, G. G., Wilson, K. S., Lukanidin, E., and Bronstein, I. B. (2003) Multiple structural states of S100A12: A key to its functional diversity. *Microsc Res Tech* **60**, 581-592.

85. Moroz, O. V., Antson, A. A., Dodson, E. J., Burrell, H. J., Grist, S. J., Lloyd, R. M., Maitland, N. J., Dodson, G. G., Wilson, K. S., Lukanidin, E., and Bronstein, I. B. (2002) The structure of S100A12 in a hexameric form and its proposed role in receptor signalling. *Acta Crystallogr D Biol Crystallogr* **58**, 407-413.

86. Novitskaya, V., Grigorian, M., Kriajevska, M., Tarabykina, S., Bronstein, I., Berezin, V., Bock, E., and Lukanidin, E. (2000) Oligomeric forms of the metastasis-related Mts1 (S100A4) protein stimulate neuronal differentiation in cultures of rat hippocampal neurons. *J Biol Chem* **275**, 41278-41286.

87. Strupat, K., Rogniaux, H., Van Dorsselaer, A., Roth, J., and Vogl, T. (2000) Calcium-induced noncovalently linked tetramers of MRP8 and MRP14 are confirmed by electrospray ionization-mass analysis. *J Am Soc Mass Spectrom* **11**, 780-788.

88. Teigelkamp, S., Bhardwaj, R. S., Roth, J., Meinardus-Hager, G., Karas, M., and Sorg, C. (1991) Calcium-dependent complex assembly of the myeloic differentiation proteins MRP-8 and MRP-14. *J Biol Chem* **266**, 13462-13467.

89. Ostendorp, T., Heizmann, C. W., Kroneck, P. M., and Fritz, G. (2005) Purification, crystallization and preliminary X-ray diffraction studies on human Ca2+-binding protein S100B. *Acta Crystallograph Sect F Struct Biol Cryst Commun* **61**, 673-675.

90. Rety, S., Osterloh, D., Arie, J. P., Tabaries, S., Seeman, J., Russo-Marie, F., Gerke, V., and Lewit-Bentley, A. (2000) Structural basis of the Ca(2+)-dependent association between S100C (S100A11) and its target, the N-terminal part of annexin I. *Structure* **8**, 175-184.

91. Bhattacharya, S., Large, E., Heizmann, C. W., Hemmings, B., and Chazin, W. J. (2003) Structure of the Ca2+/S100B/NDR kinase peptide complex: insights into S100 target specificity and activation of the kinase. *Biochemistry* **42**, 14416-14426.

92. Rustandi, R. R., Baldisseri, D. M., and Weber, D. J. (2000) Structure of the negative regulatory domain of p53 bound to S100B(betabeta). *Nat Struct Biol* **7**, 570-574.

93. Grotzinger, J. (2002) Molecular mechanisms of cytokine receptor activation. *Biochim Biophys Acta* **1592**, 215-223.

94. Watowich, S. S., Hilton, D. J., and Lodish, H. F. (1994) Activation and inhibition of erythropoietin receptor function: role of receptor dimerization. *Mol Cell Biol* **14**, 3535-3549.

95. Wrighton, N. C., Farrell, F. X., Chang, R., Kashyap, A. K., Barbone, F. P., Mulcahy, L. S., Johnson, D. L., Barrett, R. W., Jolliffe, L. K., and Dower, W. J. (1996) Small peptides as potent mimetics of the protein hormone erythropoietin. *Science* **273**, 458-464.

96. Livnah, O., Stura, E. A., Johnson, D. L., Middleton, S. A., Mulcahy, L. S., Wrighton, N. C., Dower, W. J., Jolliffe, L. K., and Wilson, I. A. (1996) Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 A. *Science* **273**, 464-471.

97. Livnah, O., Johnson, D. L., Stura, E. A., Farrell, F. X., Barbone, F. P., You, Y., Liu, K. D., Goldsmith, M. A., He, W., Krause, C. D., Pestka, S., Jolliffe, L. K., and Wilson, I. A. (1998) An antagonist peptide-EPO receptor complex suggests that receptor dimerization is not sufficient for activation. *Nat Struct Biol* **5**, 993-1004.

98. Livnah, O., Stura, E. A., Middleton, S. A., Johnson, D. L., Jolliffe, L. K., and Wilson, I. A. (1999) Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation. *Science* **283**, 987-990.

99. Syed, R. S., Reid, S. W., Li, C., Cheetham, J. C., Aoki, K. H., Liu, B., Zhan, H., Osslund, T. D., Chirino, A. J., Zhang, J., Finer-Moore, J., Elliott, S., Sitney, K., Katz, B. A., Matthews, D. J., Wendoloski, J. J., Egrie, J., and Stroud, R. M. (1998) Efficiency of signalling through cytokine receptors depends critically on receptor orientation. *Nature* **395**, 511-516.

100. Remy, I., Wilson, I. A., and Michnick, S. W. (1999) Erythropoietin receptor activation by a ligand-induced conformation change. *Science* **283**, 990-993.

101. Cunningham, B. C., Ultsch, M., De Vos, A. M., Mulkerrin, M. G., Clauser, K. R., and Wells, J. A. (1991) Dimerization of the extracellular domain of the human growth hormone receptor by a single hormone molecule. *Science* **254**, 821-825.

102. de Vos, A. M., Ultsch, M., and Kossiakoff, A. A. (1992) Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306-312.

103. Rowlinson, S. W., Behncken, S. N., Rowland, J. E., Clarkson, R. W., Strasburger, C. J., Wu, Z., Baumbach, W., and Waters, M. J. (1998) Activation of chimeric

and full-length growth hormone receptors by growth hormone receptor monoclonal antibodies. A specific conformational change may be required for full-length receptor signaling. *J Biol Chem* **273**, 5307-5314.

104. Brown, R. J., Adams, J. J., Pelekanos, R. A., Wan, Y., McKinstry, W. J., Palethorpe, K., Seeber, R. M., Monks, T. A., Eidne, K. A., Parker, M. W., and Waters, M. J. (2005) Model for growth hormone receptor activation based on subunit rotation within a receptor dimer. *Nat Struct Mol Biol* **12**, 814-821.

105. Fontana, A., de Laureto, P. P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004) Probing protein structure by limited proteolysis. *Acta Biochimica Polonica* **51**, 299-321.

106. Soroka, V., Kolkova, K., Kastrup, J. S., Diederichs, K., Breed, J., Kiselyov, V. V., Poulsen, F. M., Larsen, I. K., Welte, W., Berezin, V., Bock, E., and Kasper, C. (2003) Structure and interactions of NCAM Ig1-2-3 suggest a novel zipper mechanism for homophilic adhesion. *Structure* **11**, 1291-1301.

107. Wilton, R., Yousef, M. A., Saxena, P., Szpunar, M., and Stevens, F. J. (2006) Expression and purification of recombinant human receptor for advanced glycation endproducts in Escherichia coli. *Protein Expr Purif* **47**, 25-35.

108. Renard, C., Chappey, O., Wautier, M. P., Nagashima, M., Lundh, E., Morser, J., Zhao, L., Schmidt, A. M., Scherrmann, J. M., and Wautier, J. L. (1997) Recombinant advanced glycation end product receptor pharmacokinetics in normal and diabetic rats. *Molecular Pharmacology* **52**, 54-62.

109. Ostendorp, T., Weibel, M., Leclerc, E., Kleinert, P., Kroneck, P. M., Heizmann, C. W., and Fritz, G. (2006) Expression and purification of the soluble isoform of human receptor for advanced glycation end products (sRAGE) from Pichia pastoris. *Biochemical & Biophysical Research Communications* **347**, 4-11.

110. Chung, C. M., Chiu, J. D., Connors, L. H., Gursky, O., Lim, A., Dykstra, A. B., Liepnieks, J., Benson, M. D., Costello, C. E., Skinner, M., and Walsh, M. T. (2005) Thermodynamic stability of a kappaI immunoglobulin light chain: relevance to multiple myeloma. *Biophys J* **88**, 4232-4242.

111. Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12366-12371.

112. Eletsky, A., Kienhofer, A., and Pervushin, K. (2001) TROSY NMR with partially deuterated proteins. *Journal of Biomolecular NMR* **20**, 177-180.

113. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* **26**, 795-800.

114. Schneider, T. R. and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58**, 1772-1779.

115. Fortelle, E. d. l. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods in Enzymology* **276**, 472-494.

116. Terwilliger, T. C. (2000) Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* **56**, 965-972.

117. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.

118. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255.

119. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* **57**, 122-133.

120. Bork, P., Holm, L., and Sander, C. (1994) The immunoglobulin fold. Structural classification, sequence patterns and common core. *Journal of Molecular Biology* **242**, 309-320.

121. Donato, R. (2001) S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *Int J Biochem Cell Biol* **33**, 637-668.

122. Kislinger, T., Fu, C., Huber, B., Qu, W., Taguchi, A., Du Yan, S., Hofmann, M., Yan, S. F., Pischetsrieder, M., Stern, D., and Schmidt, A. M. (1999) N(epsilon)-(carboxymethyl)lysine adducts of proteins are ligands for receptor for advanced glycation end products that activate cell signaling pathways and modulate gene expression. *J Biol Chem* **274**, 31740-31749.

123. Ikeda, K., Higashi, T., Sano, H., Jinnouchi, Y., Yoshida, M., Araki, T., Ueda, S., and Horiuchi, S. (1996) N (epsilon)-(carboxymethyl)lysine protein adduct is a major immunological epitope in proteins modified with advanced glycation end products of the Maillard reaction. *Biochemistry* **35**, 8075-8083.

124. Smith, S. P. and Shaw, G. S. (1997) Assignment and secondary structure of calcium-bound human S100B. *J Biomol NMR* **10**, 77-88.

125. Koppal, T., Lam, A. G., Guo, L., and Van Eldik, L. J. (2001) S100B proteins that lack one or both cysteine residues can induce inflammatory responses in astrocytes and microglia. *Neurochemistry International* **39**, 401-407.

126. Haspel, J. and Grumet, M. (2003) The L1CAM extracellular region: a multi-domain protein with modular and cooperative binding modes. *Frontiers in Bioscience* **8**, s1210-1225.

127. Xie, J., Burz, D. S., He, W., Bronstein, I. B., Lednev, I., and Shekhtman, A. (2007) Hexameric calgranulin C (S100A12) binds to the receptor for advanced glycated end products (RAGE) using symmetric hydrophobic target-binding patches. *J Biol Chem* **282**, 4218-4231.

128. Wilder, P. T., Lin, J., Bair, C. L., Charpentier, T. H., Yang, D., Liriano, M., Varney, K. M., Lee, A., Oppenheim, A. B., Adhya, S., Carrier, F., and Weber, D. J. (2006) Recognition of the tumor suppressor protein p53 and other protein targets by the calcium-binding protein S100B. *Biochim Biophys Acta* **1763**, 1284-1297.

129. Srikrishna, G., Huttunen, H. J., Johansson, L., Weigle, B., Yamaguchi, Y., Rauvala, H., and Freeze, H. H. (2002) N -Glycans on the receptor for advanced glycation end products influence amphoterin binding and neurite outgrowth. *J Neurochem* **80**, 998-1008.

130. Fages, C., Nolo, R., Huttunen, H. J., Eskelinen, E., and Rauvala, H. (2000) Regulation of cell migration by amphoterin. *Journal of Cell Science* **113**, 611-620.

131. Yonekura, H., Yamamoto, Y., Sakurai, S., Petrova, R. G., Abedin, M. J., Li, H., Yasui, K., Takeuchi, M., Makita, Z., Takasawa, S., Okamoto, H., Watanabe, T., and Yamamoto, H. (2003) Novel splice variants of the receptor for advanced glycation end-products expressed in human vascular endothelial cells and pericytes, and their putative roles in diabetes-induced vascular injury. *Biochem J* **370**, 1097-1109.

132. Hanford, L. E., Enghild, J. J., Valnickova, Z., Petersen, S. V., Schaefer, L. M., Schaefer, T. M., Reinhart, T. A., and Oury, T. D. (2004) Purification and characterization of mouse soluble receptor for advanced glycation end products (sRAGE). *Journal of Biological Chemistry* **279**, 50019-50024.

133. Russ, W. P. and Engelman, D. M. (2000) The GxxxG motif: a framework for transmembrane helix-helix association. *Journal of Molecular Biology* **296**, 911-919.

134. McClintock, K. A. and Shaw, G. S. (2000) A logical sequence search for S100B target proteins. *Protein Sci* **9**, 2043-2046.

135. Weidemann, T., Hofinger, S., Muller, K., and Auer, M. (2007) Beyond dimerization: a membrane-dependent activation model for interleukin-4 receptor-mediated signalling. *J Mol Biol* **366**, 1365-1373.

136. Kurth, I., Horsten, U., Pflanz, S., Timmermann, A., Kuster, A., Dahmen, H., Tacken, I., Heinrich, P. C., and Muller-Newen, G. (2000) Importance of the membrane-proximal extracellular domains for activation of the signal transducer glycoprotein 130. *J Immunol* **164**, 273-282.

137. Malherbe, P., Richards, J. G., Gaillard, H., Thompson, A., Diener, C., Schuler, A., and Huber, G. (1999) cDNA cloning of a novel secreted isoform of the human receptor for advanced glycation end products and characterization of cells co-expressing cell-surface scavenger receptors and Swedish mutant amyloid precursor protein. *Brain Res Mol Brain Res* **71**, 159-170.

138. Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng* **6**, 383-390.

139. Johnsson, B., Lofas, S., and Lindquist, G. (1991) Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Analytical Biochemistry* **198**, 268-277.

140. Myszka, D. G. (1999) Improving biosensor analysis. *Journal of Molecular Recognition* **12**, 279-284.

141. Roden, L. D. and Myszka, D. G. (1996) Global analysis of a macromolecular interaction measured on BIAcore. *Biochemical & Biophysical Research Communications* **225**, 1073-1077.

142. Bartels, C., Xia, T., Billeter, M., Guntert, P., and Wuthrich, K. (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *Journal of Biomolecular NMR* **6**, 1-10.

143. Lambert, C., Leonard, N., De Bolle, X., and Depiereux, E. (2002) ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics* **18**, 1250-1256.

144. Canutescu, A. A., Shelenkov, A. A., and Dunbrack, R. L., Jr. (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci* **12**, 2001-2014.

145. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) Stereochemical quality of protein structure coordinates. *Proteins* **12**, 345-364.