8-2012

# Transmembrane Domain Structure and Function in the Erythropoietin Receptor

Ian Brett

# Stony Brook University

**Transmembrane Domain Structure and Function in the Erythropoietin Receptor**

A Dissertation Presented

by

**Ian Christopher Brett**

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

**Doctor of Philosophy**

in

**Biochemistry and Structural Biology**

Stony Brook University

**August 2012**

**Stony Brook University**

The Graduate School

**Ian Christopher Brett**

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

**Steven O. Smith, Ph.D. – Dissertation Advisor**
**Professor, Department of Biochemistry and Cell Biology**

**Huilin Li, Ph.D. - Chairperson of Defense**
**Professor, Department of Biochemistry and Cell Biology**

**Markus Seeliger, Ph.D.**
**Assistant Professor, Department of Pharmacological Sciences**

**Nancy Reich Marshall, Ph.D.**
**Professor, Department of Molecular Genetics & Microbiology**

This dissertation is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

**Transmembrane Domain Structure and Function in the Erythropoietin Receptor**

by

**Ian Christopher Brett**

**Doctor of Philosophy**

in

**Biochemistry and Structural Biology**

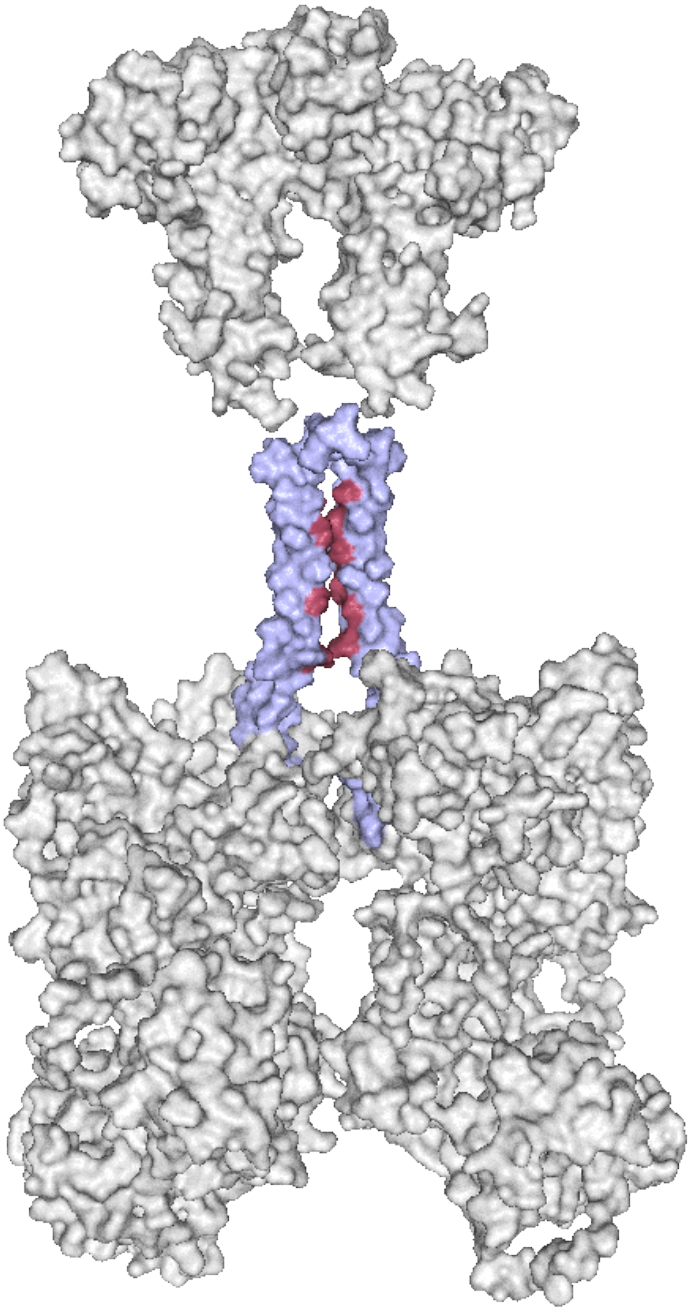
Stony Brook University

**2012**

Activation of the erythropoietin receptor (EpoR) by the soluble cytokine erythropoietin (Epo) is essential for the differentiation of erythrocyte progenitors and their development into red blood cells. The single transmembrane (TM) helix of the EpoR mediates dimerization of the receptor in the inactive state and is responsible for coupling ligand binding to activation of an intracellular Janus kinase. Neither the structure of the inactive dimer nor the structural changes in the TM region that occur upon ligand binding are known. This work presents the solution NMR structures of peptides corresponding to the TM and juxtamembrane (JM) sequences that bridge the extracellular and intracellular domains. The N-terminal end of the TM-JM peptides contains the transition point between the last β-strand of the extracellular D2 domain of the receptor and the TM α-helix. NMR measurements indicate that the TM helix extends to Pro225. This proline allows Asp224 to fold back and form side chain hydrogen bonds to the backbone NH of Leu226. Structural studies on the TM region of the EpoR alone reveal intermolecular contacts between polar residues (Ser231, Ser238 and Thr242). At the intracellular TM-JM boundary, the defined α-helical structure appears to break at Arg250-Arg251. However, Leu253-Lys256 exhibit downfield carbonyl chemical shifts consistent with helical structure for the JM switch region. To stabilize the TM-JM peptides in an active conformation, two approaches were undertaken. The first approach was to substitute Leu223 with cysteine; full length L223C EpoR is constitutively active. The second approach was to characterize the complex between the TM-JM dimer and the TM domain of an EpoR-activating viral membrane protein, gp55-P. In both cases, the largest

chemical shift changes were at the intracellular TM-JM boundary, particularly His249. Mechanisms of receptor activation that unite biophysical and biochemical data are discussed.

**Frontispiece**



## Table of Contents

## List of Figures

## List of Tables

## List of Abbreviations

TM- transmembrane

TMD- transmembrane domain

JM- juxtamembrane

Epo- erythropoietin

EpoR- erythropoietin receptor

Tpo- thrombopoietin

TpoR- thrombopoietin receptor

CHD- cytokine homology domain

ECD- extracellular domain

GH- growth hormone

GHR- growth hormone receptor

Prl- prolactin

PrlR- prolactin receptor

EGF- epidermal growth factor

Jak- Janus kinase

STAT- signal transducer and activator of transcription

His- 6x histidine tag

MBP- maltose binding protein tag

TEV- tobacco etch virus protease

IPTG- isopropyl β-D-1-thiogalactopyranoside

β-OG- octyl-β-D-maltoside

DDM- n-dodecyl-β-D-maltoside

CMC- critical micelle concentration

DHPC- 1,2-dihexanoyl-*sn*-glycerophosphocholine

DMPC- 1,2-dimyristyl-*sn*-glycerophosphocholine

PAGE- polyacrylamide gel electrophoresis

NMR- nuclear magnetic resonance

INEPT- insensitive nuclei enhancement using polarization transfer

HSQC- heteronuclear single quantum coherence

NOE- Nuclear Overhauser Effect

NOESY- Nuclear Overhauser Effect SpectroscopY

ARIA- ambiguous restraints by iterative assignment

CNS- Crystallography & NMR System

FERM- four point one, ezrin, radixin, moesin

PPIase- peptidyl-prolyl isomerase

## Acknowledgments

I gratefully acknowledge the use of the New York Structural Biology Center for NMR dataset collection, as well as the assistance of Shibani Bhattacharya, Michael Goger, and Kaushik Dutta during experiment setup and data collection.

I am also grateful for the assistance of Miki Itaya, who performed the analytical ultracentrifugation experiments and mass spectrometry on purified EpoR and TpoR peptides.

Lastly, I am indebted to Huilin Li, PhD; Stuart McLaughlin, PhD; Nancy Reich Marshall, PhD; and Markus Seeliger, PhD for their time, assistance, and invaluable advice during this process.

# Chapter 1-Introduction

## Cytokine Receptor Structure and Biology

The cytokine receptor family performs a wide variety of essential biological functions including the growth and development of multiple cell types (*e.g.*, blood cells) or regulation of the immune system (*1*). Members of this family of receptors commonly feature at least one cytokine receptor homology domain (CHD) composed of the fibronectin derived D1 and D2 domains; the D1 domain has 4 conserved cysteines and the D2 domain has a conserved WSXWS motif. The receptors that have functional intracellular domains (ICD) share homology as well, containing the conserved Janus kinase (Jak) binding sites Box 1 and Box 2. Each receptor shows high conservation across species (*2*). They are subdivided into 5 subgroups based on sequence homology, receptor structure, and type of functional oligomer (homo- or heterooligomer). Group 2 is the largest subgroup and has the most diverse extracellular domain (ECD), with a varying number of fibronectin and immunoglobulin domains in addition to the CHD. This group contains the gp130 common signaling chain receptor used by several of the Group 2 members. Group 3 also has a modified ECD containing an immunoglobulin domain in addition to the CHD; several of its members also use gp130 as a common signaling chain. Groups 4 and 5 are also heterocomplex receptors, but instead of using gp130 as the common signaling chain, either IL-2γc or IL-3Rβc are used.

Group 1 is made up of homodimeric receptors, including the receptor that is the focus of this research, the erythropoietin (Epo) receptor. Other Group 1 members include receptors for thrombopoietin (Tpo), growth hormone (GH) and prolactin (Prl) (Figure 1.1). Together, these four receptors are structurally similar with the exception of the TpoR. It has a duplicated ECD

(2x CHD), only one of which binds ligand, and a five-residue insert at the C-terminus of the transmembrane (TM) domain. Each of these receptors uses Jak2 and STAT5 (signal transducer and activator of transcription) for intracellular signaling.

The active form of Group 1 receptors is a dimer. The ligand is bound in a cleft between the two halves of the receptor dimer ECD in an asymmetric fashion, using a high-affinity site on one receptor monomer and a low-affinity site on the other (*3*). Ligand makes contact with several residues from both the D1 and the D2 domains. A short flexible linker region between the two domains allows changes in the relative position of the two domains in response to ligand. The conserved WSXWS motif in the second fibronectin fold of the CHD is necessary for proper receptor function and trafficking to the membrane. Solution nuclear magnetic resonance (NMR) structures of the D2 domain of the GH receptor and the IL-6 receptor (*4, 5*) and crystal structures of the GHR, PrlR and the EpoR extracellular domains (*6-9*) revealed that the conserved tryptophan residues of the WSXWS motif interact with conserved arginine residues in the CHD, likely stabilizing the structure of this domain. Mutagenesis of this motif performed in several receptors to gain insight into the potential structural and functional roles of the WSXWS motif (*10, 11*) usually resulted in decreased affinity for ligand. These results led researchers to hypothesize that WSXWS was involved in ligand binding. However, a more definitive WSXWS mutational study performed by Hilton *et al.* (*12*) determined that while the W232 and W235 residues had a very narrow mutational range (where functionality is preserved only if replaced by Phe or Tyr), S233 and S236 functioned with a wider range of mutations. In addition, the EpoR could still bind Epo when position A234 was mutated to almost any other residue. Despite the fact that mutation generally decreased receptor activity, Epo binding assays indicated that even mutant receptors that made it to the cell surface were capable of binding Epo and productive

signaling. This result is more consistent with the WSXWS motif performing a structural role in folding of the Epo receptor. Indeed, when the crystal structure of the Epo receptor ECD with ligand bound was solved, the WSXWS motif was located outside of the Epo binding site (*8*).

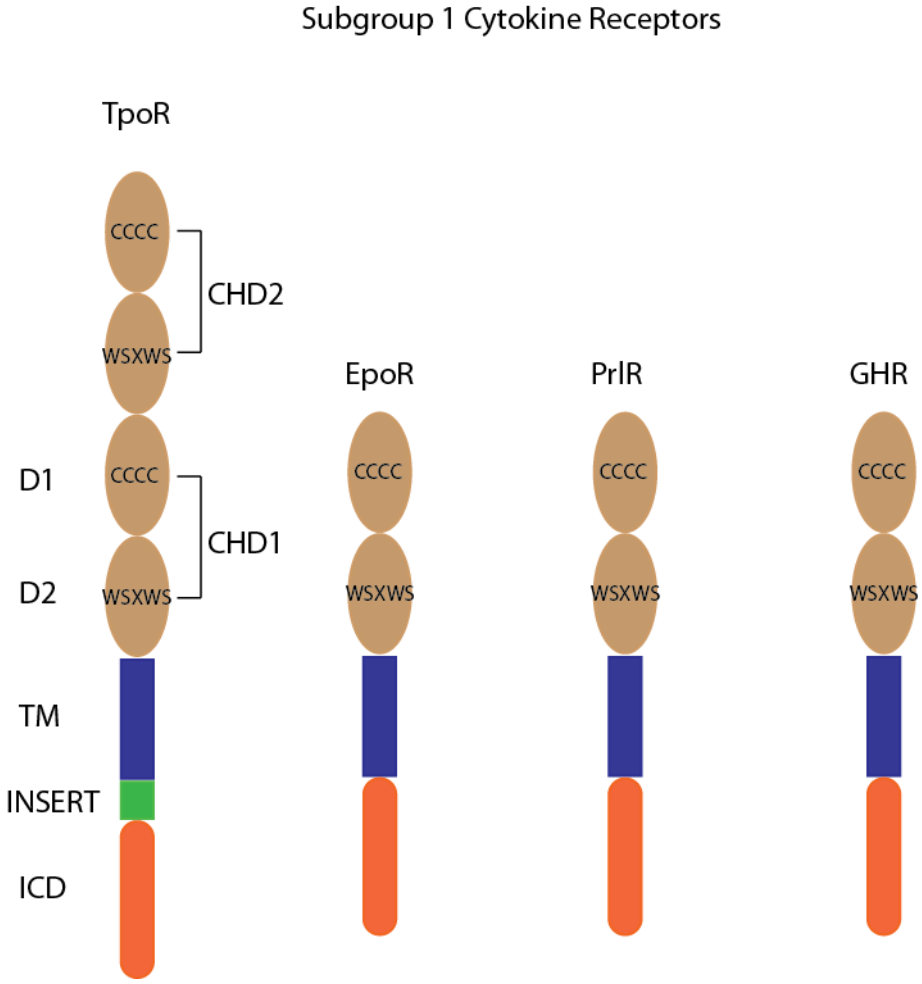



***Figure 1.1 Group 1 cytokine receptor family members.*** *Four members of the Group 1 cytokine receptors share general structural homology with each having an extracellular cytokine homology domain, a single transmembrane domain of varying length, and an ICD with a variable number of tyrosines, capable of non-covalent binding to a Janus kinase. The TpoR has two unique features: a duplicated extracellular domain and a five-residue insert (R/KWQFP) that modulates receptor activity. Epo:erythropoietin, Tpo:thrombopoietin, Prl:prolactin, GH:growth hormone*

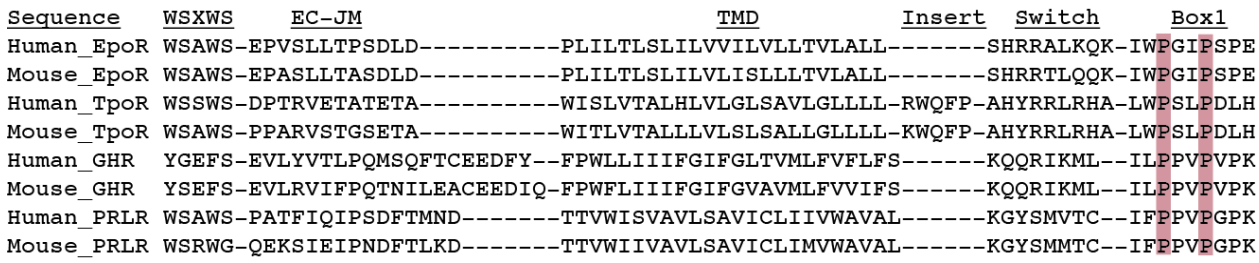



***Figure 1.2 Group 1 cytokine receptor family member TMD sequences, human and mouse.*** *The region including and around the TMD of the Group 1 receptor family members reveals interesting features. Beginning with the conserved WSXWS motif, each member has an acidic extracellular juxtamembrane domain of variable length, a hydrophobic TMD that shows interspecies conservation, an intracellular switch region with basic residues and a proline-rich Box 1 region that interacts with the Janus kinase. Conserved prolines are highlighted. Epo:erythropoietin, Tpo:thrombopoietin, Prl:prolactin, GH:growth hormone*

The TMD, in contrast, does not share sequence conservation across Group 1 cytokine receptor members, though sequence conservation of each receptor between species is generally seen (*2*). Alignments of the TMD sequences from Group 1 receptors produce no useful comparisons (Figure 1.2), and though these receptors are thought to function as dimers, no clear TM dimerization motifs are seen. Given the conflicting reports about the role of the TMD in receptor dimerization of this subgroup of receptors (*13-17*), this is not surprising. For instance, studies show that both wild type GHR and chimeric GHR with the TMD replaced with the low-density lipoprotein receptor (LDLR) TMD co-precipitate with a co-expressed, intracellular domain (ICD) truncated GHR mutant (*13*). This result not only indicates that the TMD is not necessary for dimerization of the GHR, but that the ECD is required. This observation is in contrast to experimental results from the Epo receptor. When EpoR-PrlR chimeras are co-expressed in cells with full-length WT Epo receptor, receptor co-patching only occurs when the EpoR TMD is present (*14*). A related study of truncated Prl receptors (ECD-TM or ICD-TM) indicated that when co-expressed in cells the ECD-TM inhibited signaling, while the ICD-TM construct augmented signaling (*16*). These results suggest that the TMD mediates PrlR

interaction. Using the TOXCAT expression-reporter system (*18*), further studies of the Epo and GH receptors indicate that the propensity for TM self-association is four times greater for the EpoR TMD than the GHR TMD (*19*).

While the EpoR and TpoR TMDs are leucine-rich, the GHR and PrlR TMDs are not. The presence of polar residues in the TMD is interesting. Studies show that when present they typically perform a critical function, such as mediate interhelical interaction (*20*). While the GHR is rich in aromatic residues, the other members are not. It is clear that the sequence of the TMD plays an important part in the function of each receptor, but if dimerization is mediated by the TMD a single dimerization motif that governs all of these receptors is unlikely to be found.

The ICDs of these receptors are arguably the most important and least understood part of the molecule. The ICD is responsible for both binding the tyrosine kinase molecule and harboring the tyrosine residues that, when phosphorylated, allow binding of adapter proteins for several downstream effector pathways. Each receptor has a variable length tail with a variable number of tyrosines, only some of which are essential for basic receptor function (*21*), the rest of which often function as docking sites for accessory pathway proteins (*22*). An important conserved region of the ICD is the Box 1 motif (Figure 1.2), necessary for binding the Janus kinase tyrosine kinase (*23, 24*). A second important Jak binding site is the smaller Box 2 motif (*24-26*).

Other interesting observations can be made regarding the receptor sequence, especially surrounding the TMD. First, the extracellular juxtamembrane region, like the TMD, is not highly conserved, but is certainly important as it mediates the transition between the β-sheet rich ECD and the α-helical TMD. It generally contains a number of acidic residues, the function of which,

if any, is unknown. In contrast the "switch" region, located in the membrane proximal intracellular juxtamembrane region, contains a number of basic residues. While these may be part of the "positive-inside" rule (*27, 28*), in at least the TpoR these basic residues have a defined role as part of a receptor-recycling motif (*29*).

**Cytokine Receptors and Disease**

The diverse array of biological processes regulated by the cytokine receptors make them targets for disease-causing mutation or other modulation resulting in pathology. Mutations can cause gain-of-function or loss-of-function, leading to physiological states that may be characterized as being within normal limits or pathologic, depending on severity. Interestingly, each receptor does not seem to have an equal chance of acquiring a pathologic mutation. For instance, the TpoR has several described mutations, but the EpoR does not. This may reflect a difference in the structural requirements for signaling, a difference in tolerance of variability in the ultimate physiological response, or differences in importance of the pathway either overall or at different stages of organism development.

The most common mutation associated with increased cytokine receptor function occurs in the associated Janus kinase, in EpoR/TpoR, Jak2 (JAKV617F) (*30*). It is unknown exactly how this mutation causes a gain of function, but the mutation occurs in the pseudokinase region of the kinase, which is thought to regulate C-terminal kinase activity (*31*). However, mutation in the receptors can modulate activity as well, presumably by inducing structural changes in the receptor. A mutation in the ECD of the EpoR (R129C) causes constitutive receptor activation in mice (*32*). Point mutations in the TpoR TMD (S505N), the EC-JM region (T487A), or the IC-JM region (W515K/L) cause constitutive TpoR activation (*33-35*). Alternatively, deletions in the intracellular C-terminus of the TpoR resulting in a truncated receptor can cause

thrombocytopenia (*36*), while organisms possessing an EpoR with a truncated intracellular C-terminus exhibit polycythemia (*22*).

The Friend spleen focus forming virus infects mice and causes erythroleukemia (*37*). It possesses a TM protein, gp55-P, which interacts with the TMD of the EpoR (*38*). This interaction is specific to the murine EpoR TMD, as it does not activate the human receptor. This specificity has been traced to the S238 residue in the murine EpoR, which is a leucine in the human receptor (*39*). Mutation of this residue to serine in the human TM sequence allows the EpoR to be activated by gp55-P (*39*). Expressed gp55-$P_{349\text{-}409}$ is used here in conjunction with expressed muEpo$R_{218\text{-}268}$ to probe possible effects on the EpoR TM structure.

**Erythropoietin and Erythrocyte Development**

Erythrocytes are the cells responsible for oxygen transport. In adult mammals these cells are produced in the bone marrow in a carefully regulated cascade involving several organs. Decreased oxygen delivery is sensed by cells in the kidney (*40*), which triggers a release of erythropoietin (Epo). Epo travels to the bone marrow where pluripotent stem cells (PSCs) give rise to all hematopoietic cell precursors. The particular precursor for erythrocytes, the erythrocyte blast forming unit (BFUe), is differentiated from PSCs by a combination of stem cell factor, IL-3, Tpo, and GM-CSF (granulocyte macrophage colony stimulating factor), but it is Epo that provides the definitive signal to begin differentiation to erythrocytes (*41, 42*).

The protein hormone Epo is a 30.4 kDa member of the four helix bundle cytokine family (*43*). Secreted by specialized cells in the outer cortex of the kidney it travels through the blood until it encounters the ligand-binding domain of the membrane-embedded Epo receptor on the cell surface of hematopoietic stem cells (HSCs) where it binds with high affinity. The Epo

receptor is a homodimer formed from two 66 kDa monomers which have sequence and structural homology to other members of the hematopoietin cytokine receptor subfamily (Figures 1.1, 1.2). The ECD portion of the Epo receptor has been well studied. Several crystal structures of the ECD exist at varying resolutions. In the absence (PDB ID:1ERN) (*9*) and presence of either the native ligand (PDB ID:1EER/1CN4) (*8*) or an Epo stimulatory peptide (PDB ID:1EBP) (*44*), the ECD is a dimer. The crystal structure of the Epo-bound ECD (PDB ID:1CN4) (*8*) confirms that the ratio of receptor to hormone in the active complex is 2:1, with the Epo molecule binding to the two halves of the ECD in an asymmetric fashion. Cellular studies confirm that the active state of the receptor is in fact a dimer (*45*).

The next well-defined domain is the hydrophobic TMD that consists of roughly 25 mostly hydrophobic amino acids. It allows tethering of the receptor in the membrane bilayer of the HSC and provides a route of communication between the EC and IC domains. The role of the ICD is perhaps the most poorly understood. This portion of the receptor has two conserved features important for proper receptor function, the Box 1 and Box 2 motifs that are important for Jak2 binding. There are 8 tyrosine residues present that are phosphorylated by Jak2 to allow binding of effector proteins, such as signal transducers and activators of transcription (STATs). These tyrosine residues are non-equivalent, Y343 is the only one shown to be absolutely necessary for activation (*21*). Most of the others can signal in a positive or negative regulatory fashion (*22*).

Receptor processing is an important part of the "life cycle" of the EpoR. After protein translation the receptor pre-associates with Jak2 in the endoplasmic reticulum (ER); most of the pool of Epo receptor does not leave the ER without Jak2 bound (*46*). This finding may be consistent with other members of the cytokine receptor family, as the TpoR behaves similarly

(*47*). After release from the ER the Epo receptor moves to the Golgi where oligosaccharide is added to the Jak2-bound fraction of the EpoR (*46*). This differentiates mature receptor (endo-H resistant oligosaccharide, 66 kDa) from the less mature form (endo-H sensitive, 64 kDa). Once the receptor is activated termination of signaling occurs by receptor internalization mediated by ubiquitination of cytoplasmic lysine residues, particularly K256, which mediates activation-induced internalization, and K428, which directs internalized receptor to the lysosome for degradation (*48*). These ubiquitination reactions are dependent on the activity of Jak2, and receptor that is not degraded by the lysosome can be degraded by the proteasome (*49*).

**Mechanism of Receptor Activation**

The activation mechanism of the EpoR is an area of active research. Previously it was thought that the receptor existed as a monomer, and two receptor monomers came together in the presence of ligand (*50*). More recently, it has become accepted that the Epo receptor exists as a pre-formed dimer (dimer in the absence of ligand) (*51*). The existence of a pre-formed, inactive dimer suggests that the receptor dimer resides in some low-energy conformation in the absence of ligand. Then, when ligand binds to the ECD, the global receptor structure is perturbed in a way that allows activation of the ICD-bound Jak2 molecules. This position is the result of two experimental studies. The first is that the crystal structure of the unliganded ECD is a dimer (*9*). The second comes from Förster resonance energy transfer (FRET) experiments using other receptors from this family. FRET studies using the GHR (*52*) indicate that these receptors form dimers in the absence of ligand. Furthermore, heterodimeric receptors associate as pre-formed dimers; as gp130 and LIF-R can be co-precipitated in the absence of ligand (*53*). Interestingly, similar FRET studies on the EGF receptor, a receptor tyrosine kinase (RTK) single TMD receptor shows that these receptors are also pre-formed dimers (*54*). However, this is

controversial as other reports have found inactive monomers (*55*), heterodimers (*56*), and activated higher-order oligomers (*57*). The importance of pre-association of receptor monomers is not known, but in theory having a pre-formed receptor complex awaiting only ligand should ultimately decrease activation latency. The following sections discuss receptor dimerization in the context of each domain of the receptor.

***Extracellular Domain***

The position or location of cytokine receptor monomer association is a point of contention. Studies of the ECD crystal structure of the ECD in the absence of Epo (PDB ID:1ERN) support the idea that the ECD mediates receptor dimerization (*9*). Several residues in the interface between the two halves of the ECD dimer are proposed to form near-symmetric interactions. When compared with the crystal structure of an non-Epo agonist-bound ECD (PDB ID:1EBP) (*44*) the domain movements involve a 13° rotation of D2 towards D1 and a change of the relative positions of the D2 domain C-termini from 73 Å to 39 Å (Figure 1.3). The Epo-bound structure shows similar domain movements (*8*). The relative positions of the two D1 domains in the ligand-bound structures are different, as the inter-D1 domain angle is 120° in the Epo-bound structure and 180° in the EMP-bound structure. This difference is likely due to more Epo-ECD contacts than the EMP-ECD structure, as Epo is a larger ligand. This observation suggests that the differences in biological activity of the two ligands (Epo > EMP) (*58*) originate from structural differences between ECD-ligand complexes.

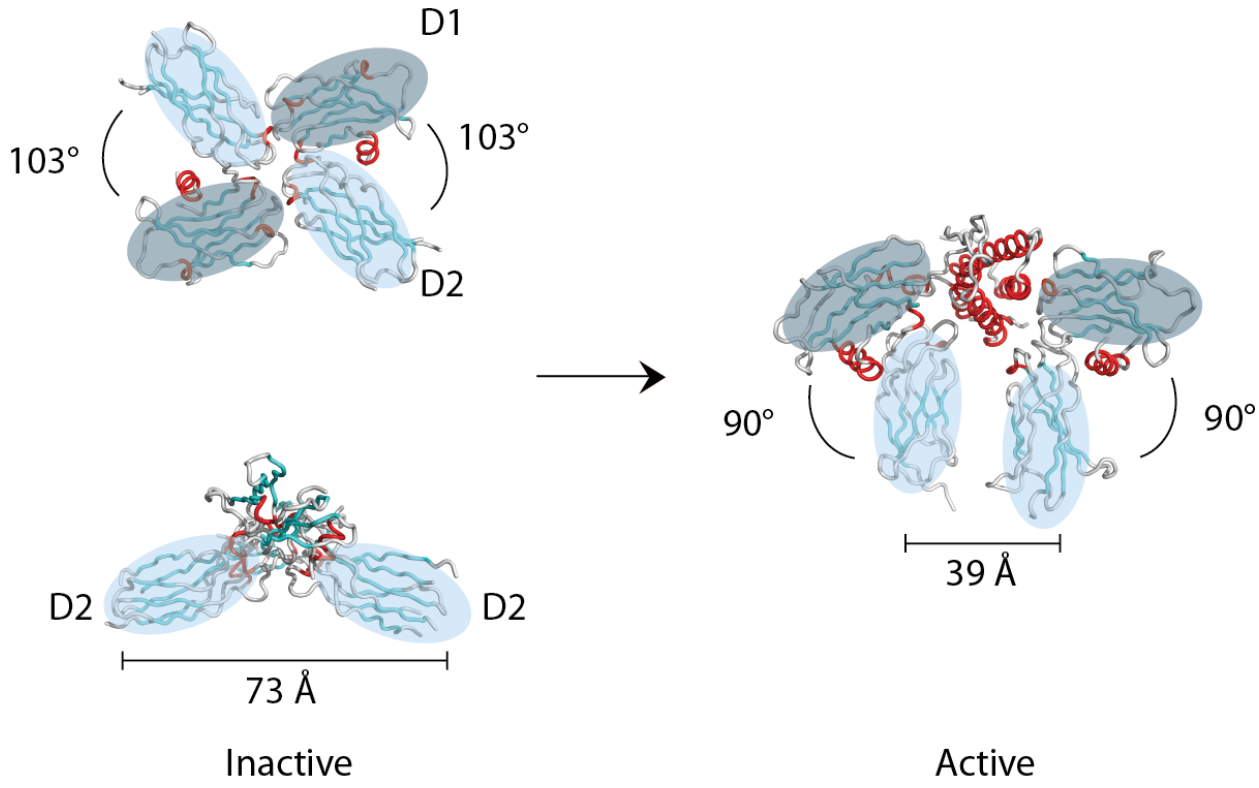



***Figure 1.3 EpoR ECD intra- and inter-domain motion upon ligand binding.*** *Crystal structures of the EpoR ECD in the absence (left, PDB ID:1ERN) and presence of Epo (right, PDB ID:1CN4) demonstrate intra-ECD domain motion (D1-D2) and inter-ECD motion upon ligand binding. Inter-ECD motion demonstrates that the D2 domains rotate to be in close proximity. D1 domains are highlighted with dark ellipses, D2 domains are highlighted with light ellipses.*

Points of inter-ECD contact are seen in both the unbound and Epo-bound structures. In the ligand-unbound structure, residues that normally interact with the ligand form contacts in the buried surface between the two ECD monomers (Phe93, Phe205, Leu33, Pro140, Met150). When either ligand is added the D2 domains move closer together, and residues in the D2 domain move within contact distance. Hydrogen bonds are formed between Glu134 and Ser135 on opposite D2 domains to stabilize the dimer (*8*). In contrast, a more extensive interface exists between the D2 domains of the GH receptor, which may contribute more to dimerization (*45*).

The result of studying crystal structures of the ECD without and with ligand suggests that there must be a significant rotation of the ECDs relative to each other as well as a change in D1-

D2 domain angle in order to activate (*8*). This hypothesis agrees with the computational studies on the GH receptor that describe a 45° rotation of the ECD upon GH binding (*59*). Pang and Zhou (*60*) also propose a similar rotation model based upon the crystal structures of the ECDs of the EpoR, PrlR, and GHR but with a twist. In their model each half of the ECD dimer performs a rotation and scissor motion, where they rotate toward each other about an axis perpendicular to the membrane, then the D2 domains pivot to bring together the C-terminal ends of the ECD (*60*). Consequently, the ICDs are brought closer together, allowing activation of Jak2.

***Transmembrane Domain***

Concurrent with the flurry of research on the Epo receptor ECD crystal structures, membrane protein groups investigated the role of the TMD on the function and activity of the Epo receptor. Gurezka *et al*. (*61*) used the TOXCAT assay to assess the propensity of the Epo receptor TMD to oligomerize and found that the self-association was the strongest of all TMDs studied that contained the leucine heptad repeat. This work led to a proposed dimerizing motif for the TMD of the Epo receptor, the leucine heptad repeat (LLxxLLx- LLxxLLxLL). This heptad repeat is a reference to the Leu zipper family of helical proteins that form sequence driven homodimers (*62, 63*). Originally found to exist in DNA-binding proteins, the dimerization of these proteins is driven by the interaction of Leu sidechains across the homodimer and this consensus sequence is found in a variety of TM and non-TM proteins. Subsequent experiments demonstrated that perturbation of this sequence by replacing Leu240 and Leu241 with Gly-Pro abrogated receptor activation as measured by colony formation, even when coupled with the activating R129C mutation (*19*). The leucine zipper dimerization motif in the EpoR was investigated further by examining the effect of asparagine mutagenesis on TMD oligomerization (*64*). L241N was found to increase TM oligomerization as measured by the TOXCAT assay.

This result led to a refinement of the Leu heptad repeat consensus to include serine residues in the interface (SxxLxxx-SxxLXxxxA).

Further work demonstrated a difference in the strength of association of the mouse and human Epo receptor TMDs. Ebie and Fleming (*65*) conducted analytical ultracentrifugation (AUC) on fusions of either the human or murine Epo receptor TMDs with staphylococcal nuclease. These studies demonstrated the intrinsic propensity of the Epo receptor TMD to dimerize in detergent micelles and found a difference in the free energy of association between the murine and human sequences of 1.3 kcal $mol^{-1}$ and 0.4 kcal $mol^{-1}$, respectively. Sequence alignments of the two receptors (Figure 1.4) point toward 3 differences between the two sequences that may be responsible for the change in association energy. These three changes surround Ser238(mouse)/Leu239(human), indicating the importance of this region to the murine Epo TM association. The fact that the murine TMD sequence dimerizes more strongly is hypothesized to be due to the mouse's higher turnover rate for erythrocytes (*65*) and implies that the basal rate of signaling activity is linked to the strength of the TM dimer. These results indicate that the oligomerization of the TMD is not indiscriminant clustering, but rather there exists a sequence-dependent association. These findings are consistent with sequence-specific oligomerization seen in other TM protein systems, such as glycophorin A (*66*), BNIP3 (*67*), and integrin TM helices (*68*).

***Intracellular Domain***

Research on the ICD sequence has identified several regions critical for Epo receptor function. The Box 1 (*69*) and Box 2 (*70*) regions are critical for binding Jak2, and the eight tyrosine residues are necessary for activation of downstream signaling pathways (*21, 71*). More detailed study of the residues connecting the C-terminus of the TMD and the Box 1 region have

identified 3 critical residues: Leu253, Ile257, and Trp258 (*72*). These residues are important for receptor activity, as mutation to alanine abrogates receptor function. Furthermore, their proper orientation is critical as insertion of 1 or 2 alanine residues between the TMD and Leu253 destroys receptor function. Interestingly, addition of three alanines (~1 helix turn) recovers activity, providing a hint that the region connecting the TMD and these "switch" residues is helical. Together, these data are suggestive of a rigid structural link between the TMD and the ICD.

**Discordance of Data from Different Regions**

Data collected from the study of the ECD suggest an inactive receptor model in which the TM domains are monomeric (*9*), while EpoR TMD studies in detergent (*65*) and in bacterial membranes (*64*) indicate that the EpoR TMD has a propensity to associate in a sequence-specific manner. The conflict between ECD dimerization data and TM dimerization data sets up a major question in cytokine receptor biology, "What is (are) the point(s) of receptor self-association?" This question cannot be answered by merely studying pieces of the receptor; it can only determine the relative strengths of self-association. Therefore, Constantinescu *et al.* (*73*) examined dimerization of the full-length receptor in cellular membranes. Using chimeras of the Epo and Prl receptors, it was determined that the TMD was both necessary and sufficient to cause receptor copatching in BOSC (human embryonic kidney) cells (*73*). Further studies were conducted to determine whether the orientation of the TMD is important in receptor activation. Cysteine mutagenesis studies on the TMD in the background of the full-length receptor in BaF3 cells indicated that when mutated, several residues in the N-terminus could either increase basal levels (Leu223, Leu226/Ile227) or severely decrease maximal levels of signaling (Asp224, Pro225) (*74, 75*). Furthermore, experiments studying Put3 coiled-coil-EpoR TMD fusions

demonstrated a dependence of activity on the rotational orientation of the TMD (*76*). Collectively, these results show that the TMD is critical for Epo receptor dimerization, and that receptor activity is dependent upon TMD orientation.

Complementing the cysteine mutagenesis studies of full-length EpoR is a report of asparagine mutagenesis of the EpoR TMD in cells (*77*), again correlating mutation with native function. This is an extension of the Ruan *et al.* (*64*) study that used asparagine mutagenesis in conjunction with the TOXCAT assay (*18*) to argue that the TM dimer interface of murine EpoR is mediated by a serine-leucine zipper. In Ruan *et al*. (*64*), L241N exhibited the strongest TOXCAT signal, indicating a strong dimer. However, in the background of the full-length EpoR in cells, the results were slightly different with L241N and A245N being hypersensitive to Epo (*77*). Interestingly, T242N responded at a lower level than WT to Epo, a result that was mimicked by T242Q as well. Oddly, similar to A245N, T242A demonstrated a hypersensitivity to Epo. The authors correlate the biological responses to TMD dimer models produced by molecular dynamics, which show that the Epo-hypersensitive mutants (T242A, A245N) have a smaller interhelical distance and packing volume than less active receptor mutants (T242N, Q) (*77*). Interestingly, different species utilize threonine or alanine at position 242 of the EpoR (Figure 1.4, (*77*)), but the reason for the difference is unclear. The AUC results mentioned above indicate that the three residue difference between the mouse and human sequences surrounding Ser238(mouse)/Leu239(human) indicate the importance of sequence specificity in TM association. Together, these results indicate that the role of the peptide sequence in TM dimerization is likely more complicated than a simple change in strength of TM association. They also hint that changes in the interhelical distance may be integral to the activation mechanism of the EpoR.

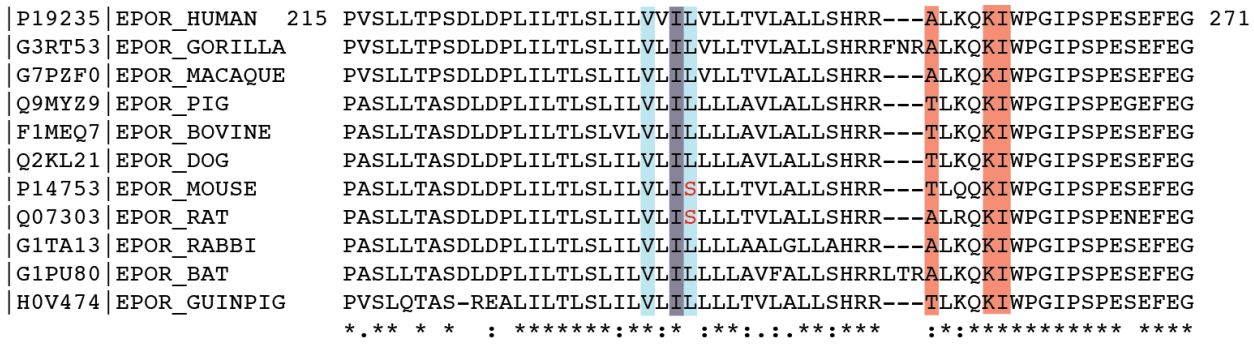



***Figure 1.4 Epo receptor TM region multiple species alignment.*** *Epo receptor protein sequences of the TM and surrounding region from 11 species show a high degree of conservation. Highlighted are residues in the middle of the TMD where the human sequence is different from the mouse (and others). Also indicated (orange) are the L-IW residues of the "switch" region that are shown to be important for receptor function. S238 in the mouse TMD is in red text. Sequence numbering provided is based on the human sequence minus the signal peptide.*

Putting all of these data in context is difficult. Given that multiple experiments show that the full-length Epo receptor is dimeric at the cell surface in the absence of ligand it seems reasonable that a pre-formed dimer exists. Yet the activation mechanisms proposed to date by study of the receptor piecewise are unsatisfying, as a single mechanism is not able to account for all of the biochemical and biophysical data. Studies of the ECD structure propose mechanisms of activation based upon data from what, in a functional sense, corresponds to only a third of the total receptor. Many of the TMD studies suffer because they are performed on severely truncated receptor TMDs. Also, conclusions from low-resolution TM association experiments are (perhaps erroneously) interpreted as being solely the result of studying TM dimers instead of being properly attributed to "oligomers of unknown constitution" (*e.g.*, TOXCAT assay results). The studies herein pick up where others leave off, studying the full TM sequence, at a sub-nanometer level of resolution, to determine the structure of the TMD and surrounding regions of the Epo receptor. From these studies, mechanisms of receptor function are proposed that are consistent with these and other data from the literature.

**Strategy for Structure-Function Studies on the EpoR**

Because the biochemical and biophysical evidence suggests that the TMD plays an integral role in the function of the EpoR, we decided to undertake 3D high-resolution structural studies of TMD-containing peptides in order to correlate structure and function. By starting with small constructs containing only the TMD we can initially capture the relevant TM structure and points of interaction between the receptor TM monomers. Then, studying longer constructs by adding portions of the receptor sequence gives us the ability to examine how changes in TMD structure are translated to the intracellular portions of the receptor.

This proposal assumes that the inactive and the active states can be "trapped," and that there will be some structural change that can be assessed by NMR. We assume that the WT peptide sequence approximates the "inactive state." We subsequently show using cysteine mutants that this assumption is correct. Comparisons between the inactive TMD structure and the active structures are accomplished in two ways. Because there are no clinically relevant mutations in the EpoR as in the TpoR (W515K/S505N/T487A), we rely on two types of activating mechanisms to approximate the active state, constitutive activity and allosteric modulation. The first involves the identification of a constitutively active mutant (L223C) from cysteine mutagenesis studies performed separately by two laboratories (*74, 75*). The second involves a viral receptor activating protein, gp55-P, from the spleen focus forming virus (SFFV) (*38*). This TM protein associates specifically with the TMD of the murine Epo receptor (*39, 78*), causing activation and cell proliferation that supports viral replication.

Starting with constructs containing the smallest independently folded domain, the TMD ($muEpoR_{220-248}$), structural studies were conducted to determine the structure of the wild type murine Epo receptor TM dimer. I found that the structure is a symmetric dimer, with

dimerization mediated by intermolecular hydrogen bonds formed between the sidechains of polar TM residues. My NMR studies of longer constructs that include the Box 1 region show a similar structure. In addition, I determined that residue His249 is a marker of the active state, and thus His249 is implicated in the activation mechanism.

## Chapter 2-Material and Methods

### Cloning, Expression and Purification

Two fusion protein constructs were prepared using polymerase chain reaction in order to clone portions of the murine Epo receptor sequence into the ligation-independent cloning His-MBP vector (kindly supplied by Dr. Tim Cross, NHMFL, Tallahassee, FL). The two TMD-containing constructs correspond to residues 220-248 ($muEpoR_{220-248}$) or 218-268 ($muEpoR_{218-268}$) of the murine EpoR. The correct sequence was verified by DNA sequencing. Vectors containing the correct fusion were transformed into chemically-induced competent *Escherichia coli* BL21(DE3) cells. Expression of the fusion proteins was accomplished in M9 medium without isotopic labels or M9 medium containing either 1 g/L $^{15}$N-ammonium chloride or 1 g/L $^{15}$N-ammonium chloride and 3.6 g/L U-$^{13}$C-glucose (Cambridge Isotope Laboratories, Andover MA). A single colony from an isolation streak on a Luria-Bertani agar plate (100 ug/mL ampicillin) was expanded overnight in 25 mL Luria-Bertani broth containing 100 ug/mL ampicillin in a shaking incubator at 37 ˚C. These cells were pelleted by centrifugation at 3,300 x *g* for 20 minutes and then washed with 5 mL sterile M9 medium (*79*) and centrifuged again at 3,300 x *g* for 20 minutes. The resulting cell pellet was resuspended into 10 mL of M9 medium and the entire volume was used to inoculate 1 L of the appropriate M9 medium (as above). These cells were grown with shaking (200 rpm) at 37 ˚C until the $OD_{600}$ reached 0.5-0.6 (~6 hours) at which point the temperature was reduced to 23 ˚C and the culture was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM. The culture was allowed to continue incubating at 23 ˚C for 16-20 hours with shaking at 200 rpm. Cells were collected by centrifugation and cell pellets were resuspended in 10 mL of 'binding buffer' (50 mM Tris, pH 7.9, 500 mM NaCl, 5 mM imidazole) and frozen at -20 ˚C until future use. These

frozen cell aliquots were thawed and then the cells were lysed using a French press. Cell lysates were clarified by centrifugation at 25,000 x *g* for 25 minutes at 4 ˚C. 200 mg octyl-β-D-glucoside (β-OG) was added to the supernatant and dissolved by nutation at 23 ˚C, ~5 minutes. This mixture was then loaded onto a 10 mL $Ni^{+}$/NTA column equilibrated with 20 mL binding buffer (as above). Binding was accomplished by nutating the column for 2-4 hours at 4 ˚C. The column was then allowed to flow through and washed with 16 times the column volume with 'wash buffer' (50 mM Tris, pH 7.9, 500 mM NaCl, 20 mM imidazole). The protein was then eluted in 1 mL fractions with 'elution buffer' (50 mM Tris, pH 7.9, 500 mM NaCl, 500 mM imidazole) until the $A_{280}$ dropped to 0.05. Fractions above 0.05 were pooled (typically ~50 ml) and n-dodecyl-β-D-maltoside (DDM) was added to a final concentration of 1.7 mM, which is 10x the critical micelle concentration (CMC). The yields of the fusion proteins were typically about 100-120 mg/L.

To cleave the fusion tag (His-MBP) from the protein of interest, Tobacco Etch Virus (TEV) protease was used. His-tagged TEV protease (His-TEV) was produced in-house, using BL21(DE3) cells expressing His-TEV grown in Luria-Bertani broth and induced with 0.4 mM IPTG. Purification of His-TEV proceeded in a manner similar to the TM containing the TM fusion proteins described above, with the exception that no detergent was added to the cleared cell lysate before incubating on the $Ni^{+}$/NTA column. After elution, fractions with an $OD_{280}$ higher than 0.05 were combined and sterile glycerol was added to 50% of the final volume (*v:v*). Then dithiothrietol (DTT) was added to 5 mM final concentration and ethylenediamine tetraacetic acid (EDTA) was added to 1 mM final concentration. Typical yields of TEV were 17-22 mg/L. The fusion protein was combined with His-TEV in a 1:1 (*v:v)* ratio and the mixture was nutated at 23 ˚C for 36 hours. Cleavage was confirmed by SDS-PAGE. To separate the TM-

domain containing muEpoR$_{220-248}$ or muEpoR$_{218-268}$ from the His-MBP fusion tag, the His-TEV, and any uncleaved fusion protein, the entire mixture was precipitated by addition of trichloroacetic acid to a final concentration of 6%. The chalky white precipitate was collected by centrifugation (3,300 x *g*, 20 minutes) and washed twice with 10 mL distilled deionized water. The washed pellet was lyophilized for 16 hours and then the hydrophobic peptide content was extracted by nutating for 2 hours at 23 ˚C with 9 mL of methanol:chloroform (90:10, *v:v*). After two hours, the supernatant was removed by syringe and then filtered through a 0.22 μm PTFE syringe filter. Estimations of the final protein yield were made based on the $A_{280}$ (muEpoR$_{218-268}$) or $A_{230}$ (for the muEpoR$_{220-248}$ peptide), the molar extinction coefficient and Beer's Law ($A=\varepsilon \ell c$, where ε is the molar absorption coefficient, $\ell$ is the pathlength and c is the concentration), according to the method described by Aitken and Learmonth (*80*).

**NMR Sample Preparation**

The organic extract containing the peptide of interest was used for reconstitution based on the protocol of Sulistijo and MacKenzie (*81*). Briefly, the appropriate amount of peptide (for a particular final concentration in ~300 uL NMR sample, usually 1 mM) was aliquotted into a separate glass vial and evaporated under dry argon or nitrogen gas to approximately 2 mL. 10 mg *d*38-dodecylphosphocholine (DPC) was dissolved into this solution, and then water was added dropwise until the solution, when agitated, produced large bubbles that did not immediately dissipate. The sample was immediately frozen in a liquid nitrogen bath and then lyophilized in a low pressure (~10 mTor) and low temperature (-95 ˚C) lyophilizer. To maintain the sample in a frozen state, the entire lyophilization jar was placed in a Styrofoam container filled with ice. Lyophilization typically took place over a 16-hour period. The dried sample was rehydrated in 1 mL 10 mM sodium phosphate, pH 7.0, using sonication to dissolve any precipitates. The sample

was then dialyzed (1,000 Da MWCO) against 3 changes of 2 L 10 mM sodium phosphate, pH 7.0 to remove any residual glycerol or imidazole carried over from the TEV cleavage or TCA precipitation steps.

The sample was then concentrated to ~250 uL in a 4 mL Millipore ultra 3 kDa MWCO spin column (centrifuge speed, 3,000 x *g* for variable time). 30 uL of $D_2O$ was added for the solvent lock along with 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) to 0.2 mM for an internal reference.

Alternatively, a detergent exchange step is incorporated into the protocol. A non-negligible amount of DDM can be carried over from the elution of fusion protein through the TEV cleavage/TCA precipitation steps. While it does not seem to change the spectral data, the presence of residual amounts of DDM can adversely impact the collection of certain NMR datasets (*e.g.*, NOESY-$^{13}$C-HSQC). A protocol was developed to do an anion column detergent exchange for Epo receptor TM containing peptides, based on a method described elsewhere (*82*).

**Detergent Exchange**

After reconstitution the sample was dialyzed into 20 mM Tris, pH 8.5 to exchange it into a buffer that made the Epo receptor TMD negatively charged for anion exchange (and simultaneously remove residual glycerol and imidazole). Two buffers were prepared, START buffer (20 mM Tris, pH 8.5, 2x CMC *d38*-DPC) and ELUTE buffer (20 mM Tris, pH 8.5, 1 M NaCl, 2x CMC *d38*-DPC). All buffers were degassed under vacuum for at least 20 minutes before the addition of detergent. A 1 mL GE healthcare Q FF anion exchange column was warmed to room temperature and equilibrated as per the manufacturer's instructions (wash with 5 mL START buffer, 5 mL ELUTE buffer, then 5 mL START buffer). The sample was clarified

by centrifugation (30 minutes at 20,000 x *g*) and loaded onto the column. The column was washed with 10 column volumes of START buffer, collecting 1 mL fractions for gel analysis, then eluted with 10 column volumes of ELUTE buffer. All fractions were saved for SDS-PAGE analysis. The construct usually eluted in the first 5 fractions. These fractions are combined and dialyzed against 10 mM sodium phosphate, pH 7.0 in order to exchange the sample into a more "NMR friendly" buffer, then concentrated to ~270 uL using a 4 mL Millipore ultra 3 kDa MWCO spin column (centrifuge speed, 3,000 x *g* for variable time). 30 uL $D_2O$ was added to the sample and then it was loaded into a $D_2O$ susceptibility-matched Shigemi tube for NMR studies.

**Analytical Ultracentrifugation**

Sedimentation equilibrium experiments were performed on a Beckman XL-I analytical ultracentrifuge at 25 °C. Samples for AUC were prepared in a manner similar to solution NMR samples, co-dissolving peptide and detergent in organic solvent followed by lyophilization and rehydration. The rehydration solution (50 mM Tris-HCl, 0.1 M NaCl pH 7.5, and 15 mM DPC) was density matched to account for the buoyancy of DPC micelles by adding 52.5% $D_2O$ (*83*) upon rehydration. Absorbance ($A_{280}$) data points were collected in radial increments of 0.001 cm. Three different peptide concentrations (~60 μM, 70 μM and 180 μM) and two different speeds (40,000 rpm and 48,000 rpm) were used to ensure data quality. UltraScan II version 9.9 data analysis software (developed by B. Demeler, http://www.ultrascan.uthscsa.edu/) was used to process and analyze the data. Global curve fitting was accomplished using non-linear least-squares analysis.

**NMR Experiments**

Experiments were conducted on 3 different Bruker spectrometers; a 700 MHz equipped with a TXI probe, a 700 MHz equipped with a TCI cryoprobe, and an 800 MHz equipped with a TCI cryoprobe. To assign resonances to backbone atoms the following experiments were run: $^{1}$H-$^{15}$N-HSQC, HNCO, HNCACB, CBCACONH. For resonance assignment to sidechain atoms, HBHACONH, HCCCONH, CCCONH experiments were run. Through space NOESY experiments conducted to determine inter-residue distances were the $^{1}$H-$^{15}$N NOESY-HSQC and $^{1}$H-$^{13}$C NOESY-HSQC experiments. To examine intermonomer contacts, two different experiments were used. First, the X-filtered, edited NOESY experiment (*84*) was run using the 1:1 U-$^{13}$C, U-$^{15}$N:Unlabeled EpoR sample in 100% $D_2O$ (150 ms mixing). Second, the CN-NOESY experiment (*85*) was run using a fully $^{13}$C, $^{15}$N labeled sample in 10 mM sodium phosphate, pH 7.0, 200 mM DPC, 10% $D_2O$ (*v:v*).

Multidimensional experiments in solution NMR spectroscopy rely on two things, application of radiofrequency pulses of a defined length, power, and frequency and application of delays of a defined length of time. The radiofrequency pulses put energy into specific nuclei in the molecule and the delays allow that energy to evolve or dephase according to the particular set of nuclei that are being manipulated.

The values and particular combinations and precise application of these parameters are what defines a "pulse sequence," which is the computer code that controls the spectrometer's running of the experiment and allows magnetization transfer through nuclei in the molecule of interest. These parameters are set in the acquisition parameters window of the NMR software. Pulse sequence nomenclature for 3D experiments generally contains the names of atoms involved in the experiment (HNCO = amide proton/nitrogen, carbonyl carbon). The

magnetization transfer schemes presented in the figures below show how these multidimensional experiments are conducted. Atoms that are circled represent atoms through which magnetization is both transferred and evolve, that is, the frequency of the chemical shift of this atom is recorded experimentally. Bonds through which magnetization is transferred are colored red. Multiple bond transfers without chemical shift labeling of intervening atoms are represented by both colored bonds and indicated by curved red arrows. For experiments that begin with an INEPT transfer from proton to carbon to enhance carbon magnetization, such as CBCACONH, the proton resonances are indicated in red and the bonds between the proton and carbon atoms indicated in red, demonstrating that the proton resonance and the J-coupling between these atoms, respectively, are used in these experiments. For through-space experiments or multiple bond J-transfers, magnetization transfers during mixing times are indicated by curved arrows and all involved nuclei are circled in red. Potential interresidue contacts are indicated with a dashed green arrow. Where possible, the experiments are run using sensitivity enhancement and pulsed field gradients described by Muhandiram and Kay (*86*). Solvent suppression can be implemented using a variety of means (*e.g.*, flip-back pulses, WATERGATE sequences, gradient pulses). Where multiple pulses sequences are available, short experiments should be conducted to assess which version works best (*i.e.*, highest signal-to-noise) for the sample in question.

**Sample-Experiment Parameters**

It should be noted that it is not desirable to conduct all the experiments described below using a single prepared sample (*i.e.*, in the same buffer system). For instance, for NH-detected experiments the buffer system should contain no more than 10% $D_2O$, unless an H-D exchange experiment is being run. If the experiments being run are CH-detected, as in the $^{1}H$-$^{13}C$ HSQC, $^{1}H$-$^{13}C$ NOESY-HSQC or related experiments, then the buffer should be exchanged into buffer

prepared in 100% $D_2O$ in order to decrease the contribution of residual water protons to the spectrum. This reduction of water signal increases the likelihood that resonances near the water frequency will be detected (often Hα) and can increase the receiver gain (because water protons are not present) so that these experiments can have the maximum sensitivity possible. Another important consideration for these experiments is to minimize or eliminate any non-protein system components that are not deuterated. This includes using buffers that do not have protons (phosphate instead of Tris), using deuterated detergents or lipids instead of nondeuterated ones, and eliminating transfers of protonated molecules from other sources (*e.g.*, trace amounts of glycerol present in filter concentrators).

**Acquisition Parameters**

$^{1}H$ pulse lengths, as well as the O1 (resonance frequency for the $^{1}H$ channel) are measured on each sample. All $^{13}C$ and $^{15}N$ pulse lengths are calculated from the measured 90° pulse lengths determined either on a standard sample of urea or on the actual experimental sample. Typical experiment acquisition settings for the data matrix size would be 2048 x 48 x 128 (F3 x F2 x F1, where F3 is the direct dimension). Sweep widths in each dimension should be truncated as much as possible to increase resolution (Hz/ppm). The resonance frequency of each channel should be chosen to be in the center of the sweep width (for HNCO- $^{1}H$ ~4.7 ppm, $^{13}C$ ~176.0 ppm, $^{15}N$ ~118.0 ppm). Last, secondary acquisition parameters (*e.g.*, nd0/nd10, FnMODE) should be set correctly prior to acquisition using the pulse program as a guide.

**Processing Parameters**

Processing NMR spectra was performed using Bruker's Topspin software. However, because processing depends to an extent on the judgment of the operator, processing is more of an art than data acquisition. Despite the room for objectivity, there are general rules to follow

when doing this. It is acceptable to truncate the time domain of any spectrum that has decayed completely or where the end of the FID is noise. Datasets should be zero-filled no more than once (2 x TD). Proper FnMODE flags should be set depending upon acquisition parameters. Forward linear prediction is used in indirect dimensions to compensate for the lower number of points taken in the interest of saving experiment time. The number of coefficients (NCOEF) chosen for linear prediction is generally a number between 16-32, though some recommendations indicate that this should be roughly equal to the number of peaks expected in the spectrum. Post-acquisition referencing is done indirectly using DSS, either internally or externally, as described in (*87, 88*).

**2D Experiments**

***$^{1}$H-$^{15}$N-HSQC (Heteronuclear Single Quantum Coherence)***

The $^{1}$H-$^{15}$N-HSQC is arguably the most important experiment in protein solution NMR. While it is a simple experiment, providing only through bond correlations for 2 nuclei (*89*), the spectrum itself provides a means for assessing conformational heterogeneity or changes in sample conformation upon changes in sample conditions (pH, buffer, mutating residues).

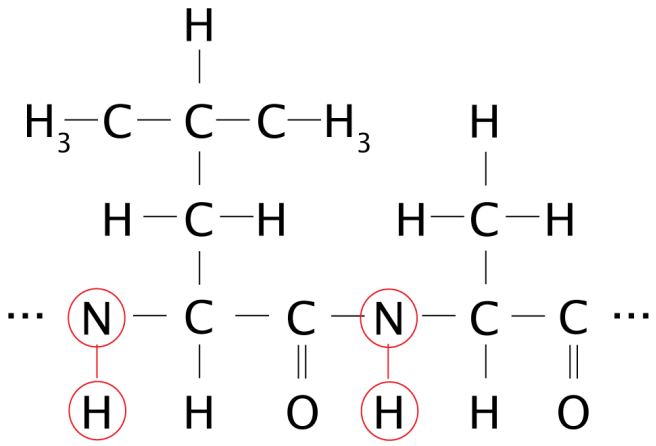


***Figure 2.1 $^{1}H$-$^{15}N$-HSQC expected resonances.*** *This experiment allows correlation of the chemical shift of nitrogen with that of its attached proton. Residues with sidechains containing nitrogens would generally show up in this experiment as well.*

### *$^{1}H$-$^{13}C$-HSQC (Heteronuclear Single Quantum Coherence)*

Essentially the same experiment as the $^{1}H$-$^{15}N$-HSQC above, the $^{1}H$-$^{13}C$-HSQC provides through bond correlations between $^{13}C$ and its directly attached $^{1}H$ atom(s). While the premise is essentially the same, there are many more variations of the $^{1}H$-$^{13}C$-HSQC than the $^{1}H$-$^{15}N$-HSQC. There are only two types of NH peaks in an $^{1}H$-$^{15}N$-HSQC (backbone and sidechain amide), but for the $^{1}H$-$^{13}C$-HSQC, the carbons and protons cover a wider frequency range in both dimensions. To make matters more complicated, aromatic C-H bonds have a different J-coupling value and show up at a different frequency (in C and H) than the aliphatic C-H do, so a second experiment with optimized values needs to be run to visualize the aromatic CH portion of the spectrum. The differences come mainly from the desire to look differently at the more numerous types of carbons in the spectrum, or at specific regions of the spectrum.

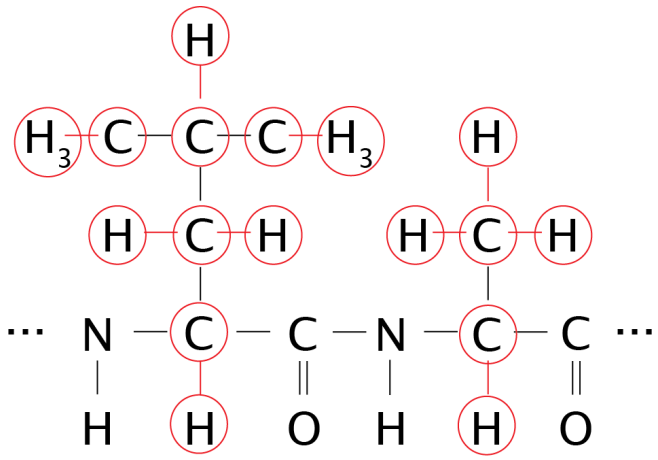



***Figure 2.2 $^{1}H$-$^{13}C$-HSQC expected resonances.*** *This experiment allows correlation of the chemical shift of carbon with that of its attached proton(s).*

**3D Backbone Experiments**

Through bond correlations of carbon resonances to backbone amide resonances are the basis of assigning particular NH resonances to residues in the protein sequence. This assignment approach requires the collection of multiple datasets that correlate $^{1}$H-$^{13}$C-$^{15}$N resonances in a through bond fashion. There are several approaches to sequential protein assignment; a discussion of the merits of each is beyond the scope of this text but is covered elsewhere (*90, 91*).

***HNCO***

This interresidue experiment allows correlation of the i-1 carbonyl carbon resonance with the i NH (*92*). It has the highest sensitivity of all the 3D backbone experiments, and the 2-3 plane (NH plane) should generally (but not always) replicate the $^{1}$H-$^{15}$N HSQC. The spectrum should yield one carbon peak per NH, so if two or more peaks are seen at a particular NH frequency, multiple conformations for that residue may exist (assuming there aren't two or more overlapping amide resonances at this frequency). Resonances missing would be any residue preceding a particular residue that does not have intensity in the $^{1}$H-$^{15}$N HSQC (*i.e.*, proline or flexible residues). This experiment has the highest sensitivity of the HN-detected 3D triple resonance experiments, so all expected peaks should be seen with good resolution after 8 scans before proceeding to the other 3D experiments. Magnetization begins on the amide proton, is transferred to the attached nitrogen using an INEPT transfer using the HN J-coupling, then to the carbonyl carbon using the NCO one-bond J-coupling. Then magnetization is transferred back to the amide nitrogen for detection. The 2-3 plane will recreate the $^{1}$H-$^{15}$N HSQC, the 1-3 plane shows the dispersion of the carbonyl carbon resonances.

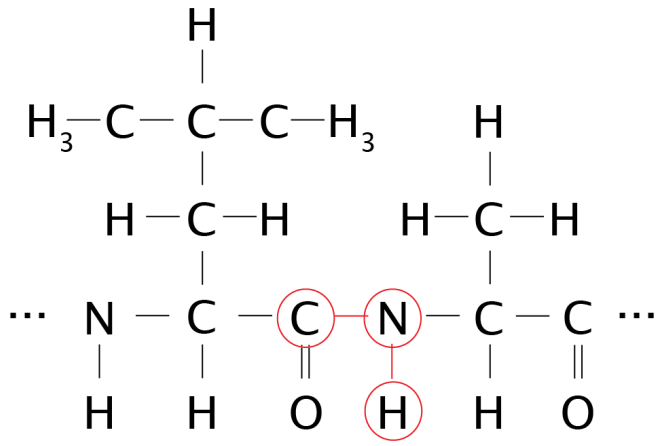


***Figure 2.3 HNCO magnetization transfer.*** *This experiment correlates the chemical shifts of the i nitrogen and proton with that of the i-1 carbonyl carbon.*

***CBCACONH***

Sidechain assignment begins with this through-bond interresidue experiment that correlates the i residue NH with the Cα and Cβ resonances of the i-1 residue. After the HNCO, it has the second highest sensitivity of the 3D backbone experiments described. Magnetization transfer begins with a $^{1}H$ to $^{13}C$ INEPT transfer to the Cα/Cβ carbons, then to the Cα, then a transfer through the carbonyl carbon to the amide nitrogen using the appropriate 1 bond J-coupling for each bond. An INEPT transfer back to the attached amide proton from the nitrogen allows detection. All experimental peaks are positive, and the 2-3 plane should recreate the $^{1}H$-$^{15}N$-HSQC. The 1-3 plane demonstrates two carbon resonances for each NH resonance, representing the Cα and Cβ peaks. Similar information can be obtained by the HNCOCACB experiment.

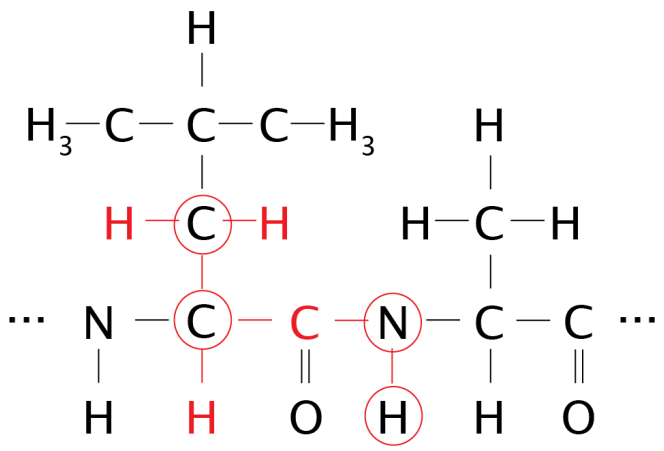


***Figure 2.4 The CBCACONH experiment.*** *This experiment correlates the chemical shifts of the i nitrogen and proton with that of the i-1 Cα/Cβ carbons.*

***HNCACB***

Each of the previous experiments are solely interresidue experiments, so an experiment that contains resonances from both the i and i-1 Cα and Cβ resonances is necessary for sequential reside assignment. Several possibilities are available to choose from when correlating inter- and intra-residual Cα/Cβ resonances to a single NH resonance in the same experiment, but the experiment used in this series is the HNCACB experiment (*93*). Magnetization begins on the amide proton, is transferred to the attached nitrogen, then to the Cα using an average of the J-couplings from the i and i-1 (2 bond) N-Cα couplings. Magnetization is then transferred to the Cβ using the Cα-Cβ coupling. Then everything happens in reverse to transfer magnetization back to the amide proton for detection. This experiment yields peaks in the carbon dimension that are 180˚ out of phase, which allows the phasing of the Cα resonances (~65ppm-45ppm) positive and the Cβ resonances (~42ppm-18ppm) negative. Another important consequence of this is that processing the 2-3 plane will generally sum to zero (for a single scan) because of the addition of positive and negative peak intensities, so recreation of the $^{1}$H-$^{15}$N HSQC is not possible. The 1-3

plane should overlap with the 1-3 plane of the CBCACONH above, albeit with twice as many peaks for each NH resonance.

Particular acquisition parameters to which attention should be paid are the shaped carbon pulses. These are generally calculated from the carbon 90˚ pulse, so if the 90˚ pulse is not correctly calibrated then the shaped pulses may be incorrect as well, causing phase errors in the carbon dimension and other spectral artifacts (*e.g.*, zero-quantum effects).

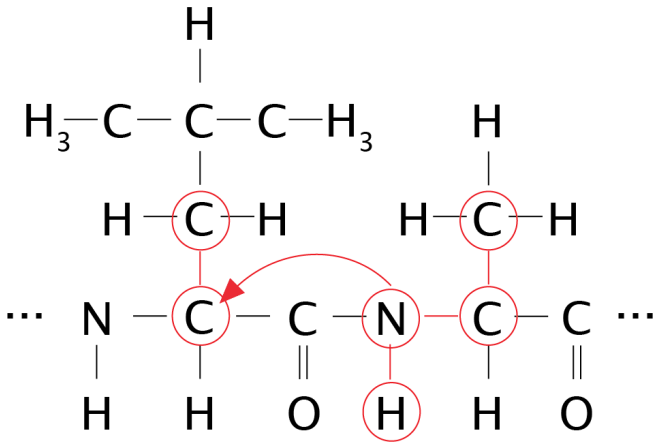


***Figure 2.5 The HNCACB experiment.*** *This experiment correlates the chemical shifts of the i nitrogen and proton with that of the i and i-1 Cα/Cβ carbons.*

**3D Sidechain Experiments**

While the above experiments are sufficient for assignment of the backbone atoms to particular resonances in the various spectra, because high-resolution structure calculation requires identification of nuclear Overhauser effect spectroscopy (NOESY) crosspeak components, assignment of as many protons in the protein as possible is essential. Therefore, using the already identified backbone resonances as a starting point, magnetization transfers are extended to atoms further out the sidechain for collection and identification of these resonances. Generally, these experiments have lower sensitivity than the HNCO backbone experiment.

***HBHACONH***

This experiment is essentially the same as the CBCACONH experiment described above, but instead of evolving and recording the i-1 carbon chemical shifts, the i-1 Hα/Hβ proton chemical shifts are recorded. Due to its relatively high sensitivity among the sidechain experiments, it should be run first.

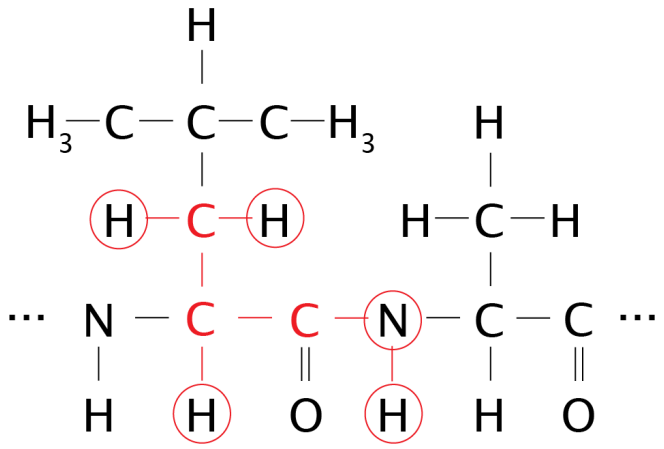


***Figure 2.6 The HBHACONH experiment.*** *This experiment correlates the chemical shifts of the i nitrogen and proton with that of the i-1 Hα/Hβ protons.*

***HCCCONH***

The HCCCONH experiment correlates all sidechain protons of aliphatic i-1 sidechains to the i NH resonances. It uses an isotropic $^{13}C$ mixing sequence to exchange magnetization between the sidechain carbon atoms (*94, 95*). Magnetization is then transferred through the carbonyl carbon to the i NH for detection. An important acquisition parameter to pay attention to is the length of the isotropic mixing sequence.

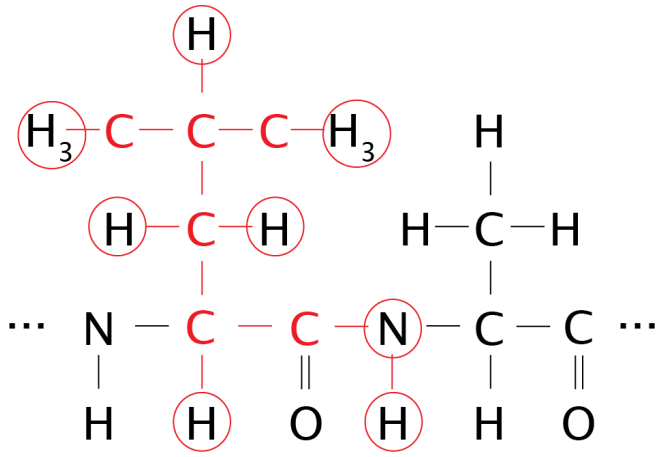


***Figure 2.7 The HCCCONH experiment.*** *This experiment correlates the chemical shifts of the i nitrogen and proton with that of the i-1 sidechain protons.*

***CCCONH***

This experiment is closely related the the HCCCONH above, but instead of recording the chemical shift of the aliphatic sidechain protons, the chemical shifts of the sidechain carbons are recorded instead. As above, the length of the isotropic mixing period should be paid particular attention.

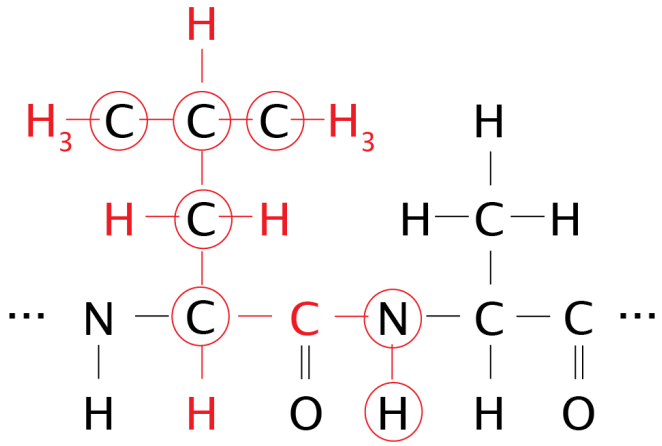


***Figure 2.8 The CCCONH experiment.*** *This experiment correlates the chemical shifts of the i nitrogen and proton with that of the i-1 sidechain carbons.*

**3D Through-Space Experiments**

High-resolution 3D structure calculation depends upon determining the relative positions of all of the molecule's atoms in space. In X-ray crystallography, the X-rays passing through the protein crystal cause a diffraction pattern that gives a picture of the relative position of these atoms in space. In NMR spectroscopy, this through-space picture is taken using correlations between different protons in the sample. These experiments utilize the Nuclear Overhauser Effect and are called "NOESYs" (Nuclear Overhauser Effect SpectroscopY). They use various magnetization transfer schemes to isolate different species of protons (amide, aliphatic, etc) but ultimately rely on a "mixing period" where different protons are allowed to exchange magnetization. For instance, the $^{1}$H-$^{15}$N-NOESY-HSQC is an experiment where all sample protons are correlated to NH protons. This is to say that it examines all protons that are within 6 Å (or so) of an amide proton. The mixing time controls how far the detection distance is; generally, shorter mixing times will generate less NOESY cross peaks (and cover less molecular distance) than longer ones will. Mixing times should be optimized on a per sample basis, generally larger molecules will have shorter optimal mixing times, but other factors such as buffer composition come into play here as well.

***NOESY-$^{1}$H-$^{15}$N-HSQC***

This experiment combines a $^{1}$H-$^{1}$H NOESY module with an $^{1}$H-$^{15}$N HSQC module. This means that first, magnetization is exchanged between all protons in the sample, then in the $^{1}$H-$^{15}$N-HSQC, magnetization is detected through the NH moiety using an INEPT transfer. In this way, we filter all of the NOE contacts that were made in the first module through the HSQC module so that we're only seeing the information exchange between protons that make contact

with the NH proton. Ideally, most of the i sidechain protons should make contacts with the i NH protons. In the figure below, these are denoted by solid red arrows. *Potential* contacts are contacts that may occur over longer distances from neighboring spin systems, and are indicated by dashed green arrows. Potential contacts are generally governed by secondary, tertiary, and possibly quaternary structural folds and intensity of these may be modulated by judicious choice of mixing time.

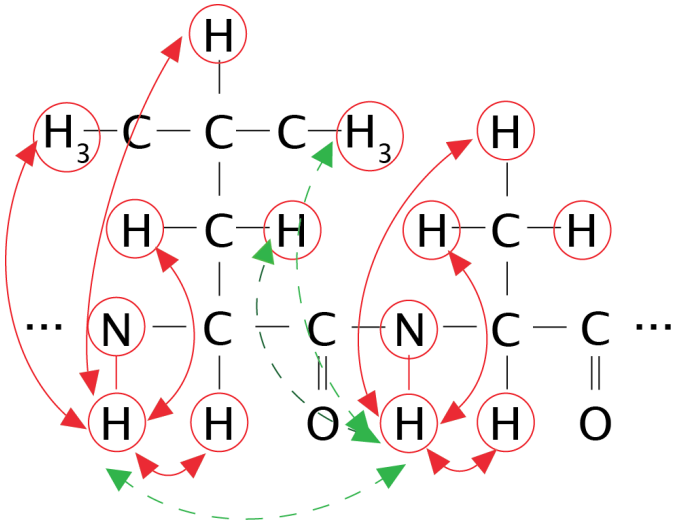


***Figure 2.9 The NOESY-$^{1}$H-$^{15}$N-HSQC experiment.*** *This experiment measures through-space contacts between amide protons and all other protons in the sample.*

### *NOESY-$^{1}$H-$^{13}$C -HSQC*

In a fashion similar to the $^{15}$N-NOESY-HSQC experiment above, this experiment combines a $^{1}$H-$^{1}$H NOESY module with an $^{1}$H-$^{13}$C-HSQC module. Correlations are detected between different protons attached to carbons, and nitrogens if the sample is not in 100% $D_2O$. *n.b.*, In the figure below (2.10), I have omitted contacts between the amide proton and nearby CH groups. Because this experiment is usually conducted in 100% $D_2O$, amide protons would be expected to exchange with deuterons and therefore be invisible. However, if for some reason the

experiment was not conducted in $D_2O$, or H-D exchange did not occur, these contacts would be expected in the spectrum.

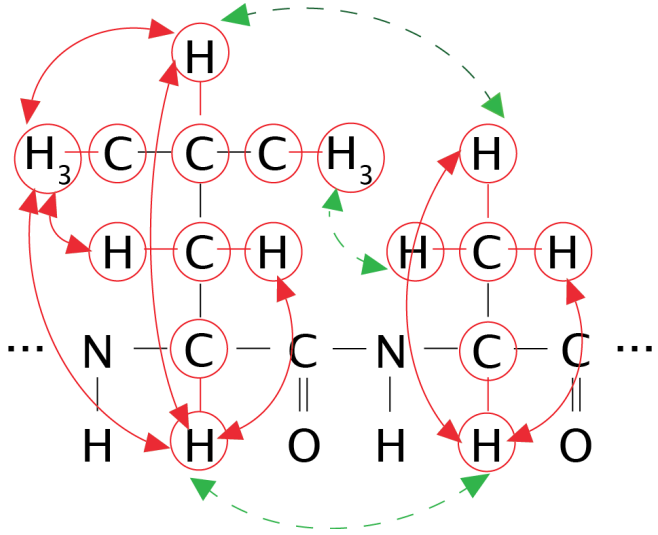



***Figure 2.10 The NOESY-$^{1}$H-$^{13}$C-HSQC experiment.*** *This experiment measures through-space contacts between aliphatic protons and all other protons in the sample.*

***CN-NOESY***

This experiment is also known as the HSQC-NOESY-HSQC because it sandwiches a NOESY experiment between an $^{1}$H-$^{13}$C-HSQC module and an $^{1}$H-$^{15}$N-HSQC module to allow detection through the amide NH resonances only those protons that are attached to a carbon (*85*). In this way, the flow of magnetization proceeds from carbon to its attached proton, then a NOESY mixing period occurs where protons are allowed to exchange magnetization, and finally detection occurs on the amide proton. This experiment deviates from a traditional NOESY in that the chemical shift axes recorded are carbon, nitrogen and the amide proton instead of two protons and one heteronucleus. In a traditional fully $^{13}$C$^{15}$N labeled protein sample, this experiment would not record any additional information when compared to a NOESY-$^{15}$N-HSQC. However, in a homodimeric system where one monomer is $^{13}$C labeled and the other is

$^{15}$N labeled, this experiment allows the transfer of NOEs in an intermonomer fashion, facilitating the identification of residues lining the dimer interface. One piece of information that is missing from this experiment is stereospecific information from the donating proton, assuming a particular carbon has multiple unique protons. This information can be regained by running the complementary NC-NOESY (see below). Important parameters are the J-couplings for the involved heteronuclei-proton pairs and the mixing time for the NOESY period.

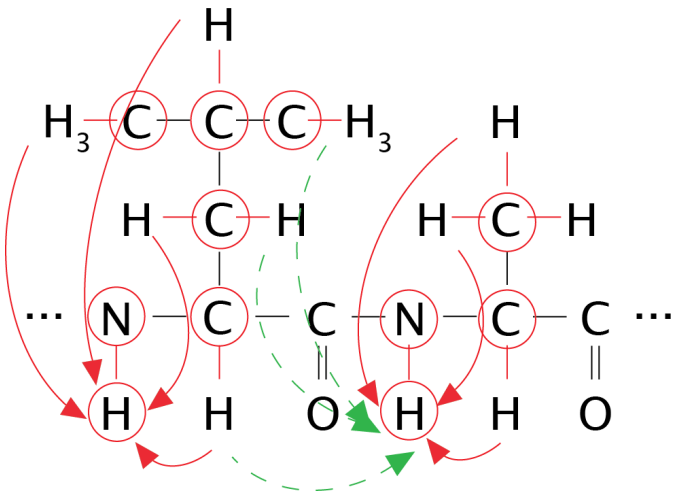



***Figure 2.11 The $^{13}C^{15}N$-NOESY experiment (aka. HSQC-NOESY-HSQC).*** *This experiment measures through-space contacts between the amide proton and protons attached to carbons, but records the chemical shift of the carbon atom instead of the proton.*

***NC-NOESY***

This is the complementary experiment to the CN-NOESY above (*85*), but it swaps the $^{1}$H-$^{15}$N-HSQC and the $^{1}$H-$^{13}$C-HSQC modules so that detection occurs through the CH instead of the NH. Important experimental considerations are as for the CN-NOESY above, but it is important to note that the signal to noise may be lower than the CN-NOESY because the NC-NOESY contains more stereospecific information.

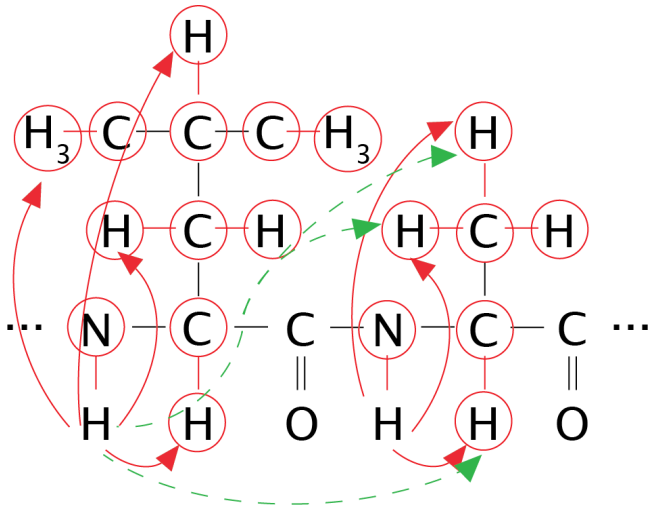



***Figure 2.12 The $^{15}N^{13}C$-NOESY experiment (aka. HSQC-NOESY-HSQC).*** *This experiment measures through-space contacts between aliphatic protons and protons attached to nitrogens, but records the chemical shift of the nitrogen atom instead of the proton.*

### *X-Filtered NOESY-HSQC (or HSQC-NOESY)*

The last, and arguably most complicated experiment is the X-filtered NOESY-HSQC. Originally developed to study points of interaction between parts of a 2-component system (*84*), it has more recently been used to determine which residues line the interface of TM helix homodimers (*81*). Intermonomer proton-proton NOEs are collected using the X-filtered NOESY-HSQC (*84*) and a sample that has half unlabeled and half $^{13}C^{15}N$-labeled monomers (see figure below-isotopically labeled atoms are green or blue). NOEs are transferred for one unlabeled molecule to the labeled one. Because this experiment depends upon detection of NOEs arising from unlabeled molecules, it is especially important to be certain that there are no components of the system that contain these types of molecules (residual unlabeled detergent, glycerol, buffer, or referencing agent). A related experiment switches the NOESY and the HSQC blocks of the experiment such that the experiment becomes an HSQC-NOESY. The information collected is

the same, however, because the NOESY crosspeaks are collected in the direct dimension (F3), greater resolution can be obtained.

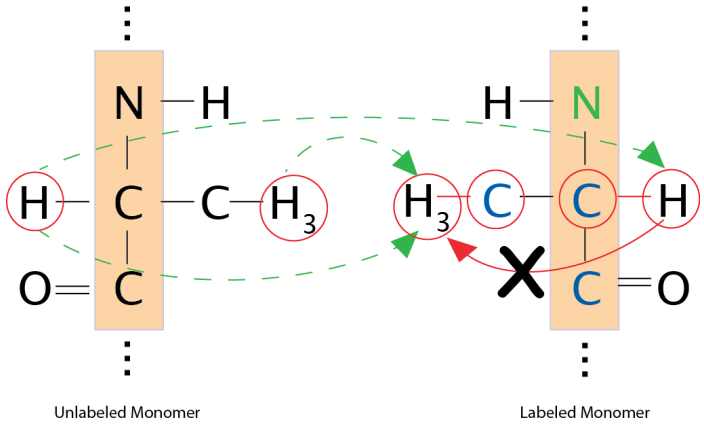



***Figure 2.13 The X-filtered NOESY experiment.*** *This experiment measures through-space contacts between of the protons of one unlabeled monomer and one isotopically labeled monomer.*

## Other Useful Experiments

### *HNCANNH*

Allows the correlation of the i-1 and i+1 nitrogen resonances with the i NH (*96*). It is particularly useful for samples where protein sequences or Cα/Cβ resonances are degenerate and correct assignment is precluded. The information produced allows one to walk through the $^{1}H^{15}N$-HSQC in a sequential fashion. While it is useful, for many samples it will have low sensitivity.

***HNCACO***

This experiment correlates the i NH with the i and i-1 carbonyl carbon resonances. It is useful in samples with proline residues, where the i-1 CO resonance for residues preceding proline are not visible in the HNCO experiment.

***TOCSY-HSQC***

This experiment correlates the i sidechain protons to the i NH, and can be useful if sidechain protons are missing in the HCCCONH experiment.

***HCC-TOCSY***

This experiment correlates sidechain carbons with one of the sidechain CH moieties. It can be useful for sidechain assignment, but in samples with multiple residues having similar chemical shifts, resonance overlap can preclude proper assignment.

***HCH-TOCSY***

This experiment correlates sidechain protons with one of the sidechain CH moieties. As with the HCC-TOCSY above, resonance overlap can prevent specific resonance assignment.

***Relaxation NMR Experiments***

Fluctuations about the NH bond can be used to determine the relative flexibility of the NH of each residue in the sample using the heteronuclear NOE experiment (*97*). T1 (spin-lattice) or T2 (spin-spin) relaxation parameters for each $^{15}N$ in the sample can be measured with the T1 relaxation and the T2 relaxation experiments (*97*).

***Solvent Accessibility***

NOEs of backbone or sidechain NH moieties to solvent (*e.g.*, water) can be measured using pulse sequences conaining the CLEANEX pulse sequence module (*98*).

**Peak Assignment**

All datasets were processed with Topspin 2.0 and either internally or externally referenced to the DSS methyl peak as described previously (*87, 88*). Processed data were then converted into a .ucsf file using the "bruk2ucsf" file conversion program in Sparky v3.113 (*99*). All spectra were loaded into Sparky, saved as a single project file and appropriate axes were synchronized using the "synchronize views" command window. Peaks in each experiment were picked and then assignments were made manually using the protein sequence and the $^{1}H$-$^{15}N$-HSQC, CBCACONH and HNCACB experiments. Where degeneracies or discrepancies arose, the NOESY-$^{15}N$-HSQC spectrum was used to determine sequential assignment of a particular NH resonance based on proximity. Also useful for this purpose was the HNCANNH experiment (*96*). Once assignment of backbone N, H, Cα and Cβ atoms was complete, the assignments of carbonyl atoms were made using the HNCO spectrum. Sidechain atom assignment began with the HBHACONH experiment, then were finished with the (H)CCCONH and the CCCONH experiments (*94*). Peaks for the various NOESY spectra collected were peak picked for each strip within the Sparky data model, but were not given resonance assignments.

**Structure Calculation**

Sparky peaklists, resonance assignments and ucsf spectra were transferred into the Collaborative Computing Project for NMR (CCPN) data model (*100*) to facilitate structure calculation. CCPN exports data directly into Aria2.3 (*101*), the program used to assign NOE spectra during the structure calculation process. Aria2.3 was used to automatically assign peak-picked (but unassigned) NOESY-$^{13}C$-HSQC spectra and Crystallography and NMR System (CNS) 1.21 (*102, 103*) was used as the structure calculation engine. Details of restraints and

structure parameters are discussed in Chapter 3, but briefly, constraint lists for structure prediction consisted of dihedral angle prediction and hydrogen bond lists for helical TM structure (i C=O/i+4 H-N bonding), derived from the chemical shift analysis (*104*). Initial rounds of structure calculation included these constraints and 1 or more of the NOESY spectra described above. Processing and structure calculations were performed on an Apple MacBook Pro with OS X 10.6.8, a 2.4 GHz Intel Core 2 Duo processor and 4 GB of RAM.

***CCPN***

The Collaborative Computing Project for NMR (CCPN) has produced a multifunctional software package for processing of NMR data (*100*). In addition to spectrum organization and peak assignment, there are a variety of tools incorporated into the software package that aid in structure calculation. DANGLE (*105*) is a dihedral angle prediction program that works in a manner similar to TALOS (*106*). Using the resonance list and the protein sequence, dihedral angles are predicted from the chemical shifts of the backbone atoms H, N, CO, Cα, Cβ, Hα and stored within the project directory for later export to Aria2.3. Tools for creating other constraints used in structure calculation, such as hydrogen bonding lists or other distance restraints are also incorporated into the software package.

***ARIA2.3/CNS 1.21***

Ambiguous Restraints for Iterative Assignment (Aria2.3) is a program that is used to analyze NOESY spectra in the context of a fully assigned resonance list to determine NOE peak contributions, make assignments to these NOEs and then use this information iteratively to calculate protein structure (*101*). Through successive iterations of structure calculation, correct NOE assignments would be enforced and carried through to the next rounds of calculation and incorrect ones would be discarded, ultimately producing a list of correct NOE assignments and

an accurate protein structure. Aria2.3 works with CNS (*102, 103*) where Aria2.3 calibrates and assigns NOEs at the beginning of each structure calculation iteration and CNS uses this information in the structure calculation.

To do this, Aria2.3 source code is compiled with CNS source code to make an executable program. The Aria2.3 gui window allows easy incorporation of project data from a saved CCPN project or independently from user modified xml files. Constraint lists can be read into Aria2.3, along with the molecule information, and CNS iteration parameters can be set. These data can be saved-Aria2.3's output is an xml file with links to all of the constraints and all the information that CNS needs to run structure calculations. Once saved, the script is setup and then activated from an X11 terminal window. At this point, CNS takes over and runs iterations, returning structures and data after each step. These results can be read back into CCPN to refine the data model in order to perform more precise structure calculations.

***NOE Assignment with Aria2.3***

Crosspeaks in NOESY spectra contain two pieces of structural information. The frequencies of the crosspeak in the relevant spectral dimensions reveal the identity of the two interacting spins, and the crosspeak intensity generally provides information about the distance separating those spins. The combination of these two pieces of information allows structure calculation. In theory, for a given structure, a "correct" set of NOE assignments exists where the identity and distance information for all peaks are self-consistent, that is, they are derived from and support calculation of a single structure. A typical NOESY spectrum from even a small protein can contain hundreds of NOE crosspeaks, making manual assignment arduous. Aria2.3 increases efficiency and accuracy of NOE spectrum assignment by automatically calibrating NOE crosspeaks and assigning them using the resonance list (*101*). Assignments that cannot be

unambiguously assigned are given an *ambiguous* assignment, where several possible assignments are treated as a single restraint. Then, these assignments are evaluated by a round of structure calculation. Through successive iterations of assignment and structure calculation, structural and NOE peaklist ambiguity are reduced, with correct assignments saved and incorrect ones discarded, resulting in structural refinement.

Input data for the structure calculation begins with the molecular system as defined by CCPN. This consists of the protein sequence and number of peptide chains in the system. The chemical shift list for each chain is read in, and the symmetry is set for "C2" symmetry, representing a symmetric homodimer system. Constraint lists begin with dihedral angle predictions for each chain, as well as hydrogen bond lists for the α-helical region of each TMD. Peak-picked NOESY spectra are added for each through space experiment being used for the calculation. An unambiguous constraint list consists of NOE contacts from manually calibrated NOE spectra such as the CN-NOESY and any NOE crosspeaks that can be unambiguously assigned to specific, non-overlapping resonances. All manually calibrated NOEs are sorted into 3 categories, weak, medium, and strong based upon the intensity of the crosspeak. Where possible, peak intensities are calibrated using an internal standard, such as crosspeaks within the leucine sidechain (*e.g.*, Hα -> Hγ). All lower bounds for manually calibrated crosspeaks were set to 1.8 Å. Upper bounds added 0.6 Å to the calibrated value. Other structural constraints used are the dihedral angle prediction for the molecular system and the hydrogen bond constraint list.

***Structure Calculation***

Structures were calculated using Aria2.3 default settings (*107*). Unambiguous constraints enter the calculation in the first round of calculation. Default iteration parameters consist of 8 rounds of structure calculations, with 20 structures generated each round. Refinement of the

resulting structures in explicit solvent occurred using a shell of water molecules. Analysis of the 20 structures generated uses the seven lowest energy structures from each round to create an ensemble average. Procheck2 was used to evaluate the dihedral angles of each average structure as well as the root mean square deviation (RMSD) of backbone atoms and angles (*108, 109*).

## Chapter 3-Structure of the Erythropoietin Receptor Transmembrane Domain Dimer

While there is compelling evidence that the Epo receptor exists as a pre-formed dimer, and the TMD mediates dimerization, several questions still remain unanswered. What is the TM dimerization interface? Does the murine Epo receptor TMD follow the proposed interface with Leu and Ser residues mediating dimerization (*61*)? How is it that the three amino acid difference between the mouse and human receptors influence 1) gp55-P binding and activation, and 2) the dimerization propensity seen by Ebie and Fleming (*65*)? Do the mouse and human receptors share the same interface? Most important of all, how is it that ligand binding is transmitted through the TM dimer to the ICD in order to activate Jak2?

To begin to define the role of the TMD in the structure and function of the Epo receptor, structural studies were performed on the smallest independently folding region, the TMD sequence from residues 220-248, referred to as muEpoR$_{220\text{-}248}$ (Figure 3.1). My studies began with the murine sequence. The murine sequence was selected for structural studies on the basis of more available biochemical and biophysical data. The construct is expressed as a fusion protein and contains both the putative TMD and several residues from the region that connects the extracellular D2 domain with the TMD (Ala220-Asp224). The cleaved, purified peptide is solubilized in a membrane mimetic system and used for solution NMR and other biophysical experiments.

SNASDLDPLILTLSLILVLISLLLTVLALLS

***Figure 3.1 Sequence of the muEpoR$_{220\text{-}248}$ peptide.*** *muEpoR$_{220\text{-}248}$ captures the entire TMD (P225-L247) and a small segment of the extracellular JM domain sequence (in red) that overlaps with the C-terminus of the ECD crystal structure (PDB ID:1CN4). Proteolytic cleavage of the fusion protein leaves two non-native amino acids (Ser-Asn) attached to the N-terminus.*

**Choice of Membrane Mimetic System for Solution NMR Studies on Monotopic Membrane Proteins**

Many decisions need to be made when undertaking biophysical studies on a particular protein. Solvent, buffer, pH, salt concentration, minor buffer components, and cofactors are several of the parameters that need to be considered for collection of biologically relevant data. Additionally, system components may not be compatible with certain techniques at the desired concentration, if at all. In an ideal situation, a preparation would resemble the protein *in vivo*, possessing the same structural features and any native biological activity. Finding the appropriate milieu can be difficult, even for soluble proteins. For membrane proteins, the natural hydrophobic character of TMDs presents an additional complexity as they are not soluble in water. Organic solvents, lipids, detergents (*110*) and amphipols (*111*) are typically used to solubilize membrane proteins.

Solution experiments can be run on peptides in organic solvent (*112, 113*), but any hydrophilic portions of the peptide may not adopt native structural features and TM oligomerization may be adversely affected (*114*). Another method for solubilizing hydrophobic TMDs involves simulating the hydrophobic environment of the lipid bilayer. This is possible by including molecules that have a hydrophobic and a hydrophilic domain. Such molecules would include lipids, detergents (*110*) and amphipols (*111*). Lipids normally self-assemble into multilamellar vesicles (MLV) consisting of many stacked bilayers; these MLVs are too large to be useful in solution NMR. Small unilamellar vesicles (SUVs) are smaller, but still too large and unstable to be used for solution NMR of integral membrane proteins; they are more commonly used to study soluble proteins that interact with lipid bilayers (*110*). Isotropic bicelles contain long and short chain lipids, typically 1,2-dimyristyl-*sn*-glycerophosphocholine (DMPC) and 1,2-

dihexanoyl-*sn*-glycerophosphocholine (DHPC) at a ratio of 1:4. These are becoming more popular for solution NMR studies for integral membrane proteins, as they may lessen α-helical structure distortion or allow the protein to retain some native activity (*115*). However, bicelles are still about twice as large as a micelle, which may necessitate use of deuteration and/or TROSY type pulse sequences (*116*). Additionally, lipids are more expensive than detergents, and deuterated lipids may be much more costly if they are available at all.

Detergents possess a hydrophilic head group and a hydrophobic "lipid-like" tail; these molecules spontaneously form micelles in aqueous solutions that have a hydrophobic interior (tails) and a hydrophilic surface (head groups). Many different detergents are available; they differ in the constituent groups. Typically, the molecular weight of the micelle plus the protein of interest is small enough to be studied by solution NMR (*110*).

Solution NMR studies of TM proteins in detergents are becoming increasingly common. Structures or structural features of entire TM proteins (*117*) or portions of larger polytopic TM proteins (*116, 118*) are being studied with increasing frequency due to technological advancements in spectrometer/probe design (*e.g.*, cryoprobes) or pulse sequence development (TROSY) (*119*). Advances in both fields allow the study of larger, more quickly relaxing systems. These methods are particularly suited to studying TM dimers. Judicious sequence and detergent choice allow the molecular weight of the complex to remain under 40 kDa, small enough to be studied by traditional solution NMR methods. This size is ideal, as crystallization of these systems has not progressed as quickly as for multitopic membrane proteins (*e.g.*, GPCRs) (*120*).

Regardless of what membrane mimetic is chosen, the complex under study must approximate the structure (and if possible function) of the biologically native sample. Many single TM receptors signal productively through a dimeric intermediate. Receptor interactions in single-pass membrane proteins are often mediated by sequence specific contacts in the TMD (*82, 121*). TM dimerization has been shown for the ErbB receptors, which are known to dimerize in lipid bilayers (*122*), the ephrin receptors (*123, 124*), and glycophorin A (*125*), among others. The solution structures of TM homodimers for Eph2A (*126*), the CD3 TCR ζζ domain (*127*), and BNIP3 (*81, 128*) have been solved.

However, important differences exist between lipid bilayers and detergent micelles that may bias structural studies. First, micelles have a greater degree of surface curvature that may adversely influence the structure of proteins solubilized by detergent micelles (*129*). Second, the width of the hydrophobic portion of the micelle may not match the native lipid environment, allowing or inducing changes in helix tilt or changes in packing of helices. Third, the packing between the hydrophobic detergent tails and the hydrophobic exposed amino acid sidechains may be suboptimal (*130*). In the end, if possible, it is best to correlate biophysical and structural data from studies performed in lipid bilayers and studies performed in detergent micelles to minimize or eliminate bias induced by the system chosen.

These studies use dodecylphosphocholine (DPC) as the membrane mimetic system. DPC is the most commonly used mimetic for solving membrane protein structures (*110*). It also has been widely used for AUC studies on TM proteins (*131*). DPC mimics several characteristics of the lipid bilayer. The zwitterionic phosphocholine headgroup is the most common headgroup of all phospholipids in the typical eukaryotic cell membrane (*132*). The width of the hydrophobic

portion of a DPC micelle (25.5-27.5 Å) (*133*) as measured by small-angle X-ray scattering is roughly the same as the hydrophobic portion of a lipid bilayer (~25-30 Å) (*132*).

**AUC Measurements Confirm Dimer Formation of muEpoR TM Peptides in DPC**

The existing cellular, biochemical, and biophysical data regarding dimerization of the EpoR TMD has been discussed above. Once reconstituted into DPC detergent micelles, our TMD-containing peptide was studied by analytical ultracentrifugation (AUC). Because of difficulty studying very small molecular weight proteins in detergent micelles by AUC, we used the slightly larger muEpoR$_{218\text{-}268}$ peptide. AUC results (Figure 3.2) demonstrate that in DPC micelles, the molecular weight of the complex is ~12 kDa, approximately twice the weight of a muEpoR$_{218\text{-}268}$ monomer (5.9 kDa).

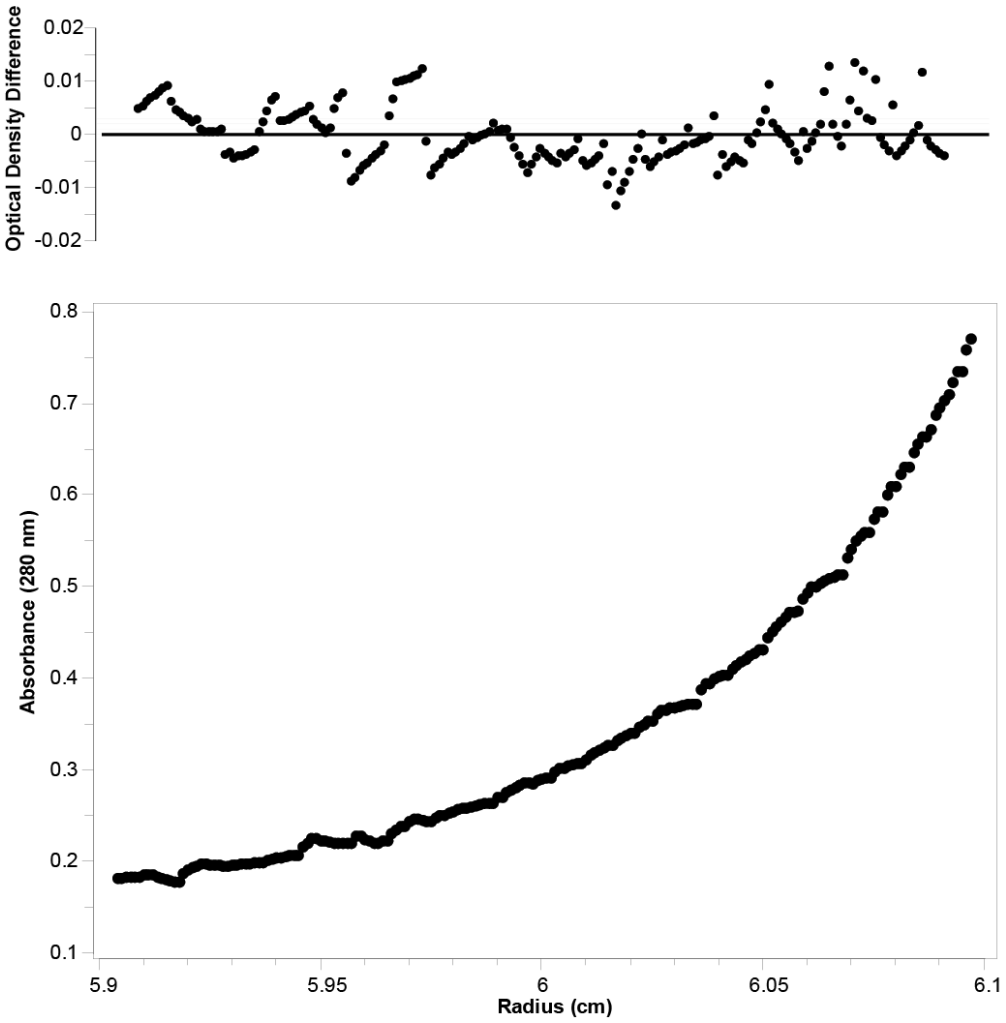



***Figure 3.2 AUC of muEpoR$_{218-268}$ in DPC micelles.*** *The muEpoR$_{218-268}$ peptide was solubilized in DPC detergent micelles, 50 mM Tris, 100 mM NaCl, 15 mM DPC, pH 7.5 at peptide concentrations of 60, 70 and 180 μM. Centrifuge speeds used were 40,000 rpm and 48,000 rpm. Global curve fit analysis reveals that the MW of the complex is ~12 kDa, consistent with a dimer of muEpoR$_{218-268}$. Shown is a representative curve and residuals.*

**muEpoR$_{220-248}$ NMR Measurements**

For these experiments, the sequence of the mouse EpoR TMD (muEpoR$_{220-248}$, Figure 3.1) along with a few EC-JM residues were cloned into the His-MBP vector (Chapter 2), expressed and purified. muEpoR$_{220-248}$ represents the smallest independently folding region of the receptor. The $^{1}$H-$^{15}$N-HSQC of the muEpoR$_{220-248}$ is presented in Figure 3.3. The spectrum

exhibits well-resolved, uniform peaks indicative of a homogeneous sample preparation. The spectrum shows the expected number of peaks, an indication that the sample under study is either monomeric or a symmetric oligomer. If the system under study was not structurally homogeneous, or an asymmetric oligomer, multiple peaks would be expected for each amino acid. Three-dimensional NMR measurements allowed the assignment of each peak in the spectrum to a particular amino acid (with the exception of Pro225). Chemical shift analysis carried out on the fully-backbone assigned sample indicates that the muEpoR$_{220\text{-}248}$ peptide is α-helical from residues Pro225-Leu246 (Figure 3.4). Circular dichroism spectroscopy of the muEpoR$_{220\text{-}248}$ peptide solubilized in DPC micelles confirms these measurements.

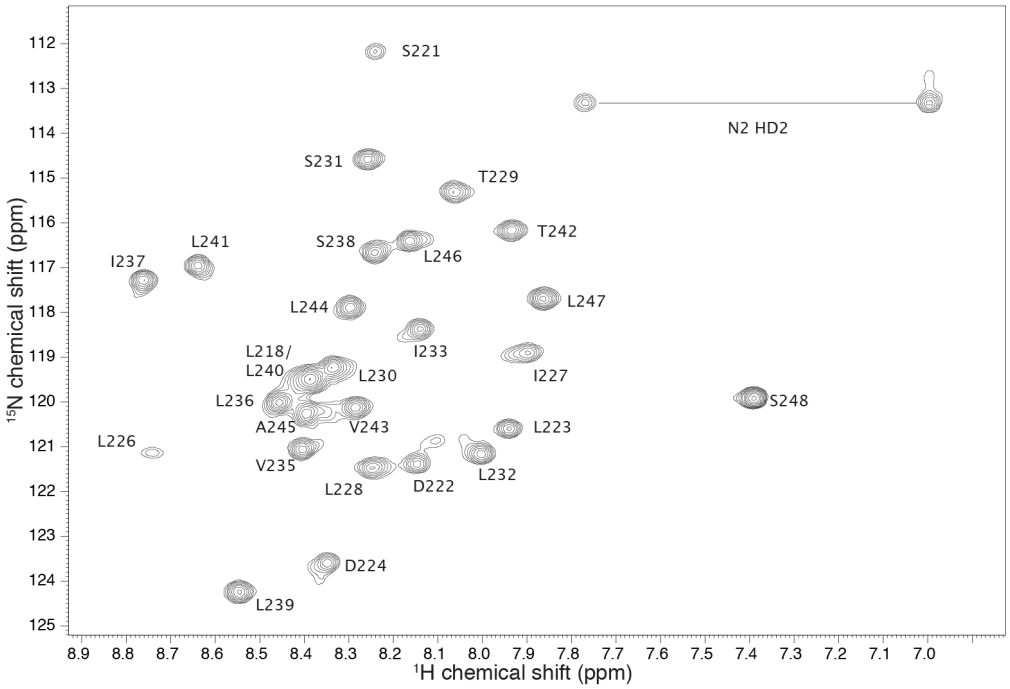



***Figure 3.3 $^{1}$H-$^{15}$N-HSQC of muEpoR$_{220\text{-}248}$.*** *The $^{1}$H-$^{15}$N-HSQC demonstrates good peak resolution and sensitivity, indicating sample homogeneity. The spectrum was collected on a 700 MHz spectrometer at 313 K with 32 scans. Sample is ~1 mM protein, 10 mM sodium phosphate, pH 7.0, 200 mM d38-DPC, 10% $D_2O$ (v:v).*


***Figure 3.4 Secondary structure of muEpoR$_{220-248}$ in DPC micelles from NMR measurements.*** *Chemical shift analysis of NMR measurements predicts dihedral angle values consistent with alpha-helical secondary structure between P225-L246. The chemical shift analysis was made using the program DANGLE (105).*

**Unusual Chemical Shifts of muEpoR$_{220-248}$ TM Polar Residues**

One curiosity that arose from analysis of the chemical shifts of the TM residues in muEpoR$_{220-248}$ is the observation of downfield shifted $^{13}$Cα resonances for Ser and Thr residues within the TMD (listed in Table 3.1). While the Cβ chemical shifts seem to be within one standard deviation of the mean value, the Cα resonances for Ser/Thr residues in the muEpoR$_{220-248}$ peptide in aqueous solution (10 mM sodium phosphate, pH 7.0) are shifted downfield by more than 2 standard deviations from their mean values. Hypothesizing these downfield shifts were directly related to dimerization, a muEpoR$_{220-248}$ sample was prepared in organic solvent (90% trifluoroethanol:10% $CDCl_3$, *v:v*), backbone resonances were measured by NMR and assigned to specific amino acids. Trifluoroethanol (TFE) has been used for NMR data acquisition on TM peptides previously (*134*); it is expected to retain, but not induce α-helical structure for peptides, yet disrupts intermolecular interactions (*114*). Interestingly, this TFE preparation of the muEpoR$_{220-248}$ peptide exhibits chemical shifts of the Ser/Thr Cα atoms in more upfield positions. Secondary structure influences chemical shift such that Cα resonances shift downfield and Cβ resonances shift upfield. However, because both peptides are α-helical and differ only in oligomeric status, we suggest that these residues may be involved in or affected by dimerization. The unusual downfield shift of the Cα resonances could be due to several factors. The BMRB

keeps detailed records of all chemical shifts (http://www.bmrb.wisc.edu/ref_info/statsel.htm), and outliers (> 3 SD from the mean) are flagged separately. Interestingly, of the first six NMR structures in the BMRB with outlying Ser Cα chemical shifts downfield of the mean value, all are in α-helical structure. Three (PDB IDs:1QQY, 1R7J, 1ZGG) are within hydrogen bonding distance to an acidic residue side chain (Asp, Glu), and two others (PDB IDs:1HA8, 2KZT) are within 6 Å of several polar sidechains (Gln, Arg, Lys). This observation would suggest that a combination of secondary structure and polar environment can cause such a chemical shift deviation. If the polar residues mediate an interface between two α-helical monomers, the proximity of the sidechain OH groups may create a polar environment able to perturb the chemical shifts in the manner seen.

| Residue | Cα, DPC | Cα, TFE | Cβ, DPC | Cβ, TFE |
|---|---|---|---|---|
| S221 | 59.054 | 57.875 | 63.661 | 61.313 |
| T229 | 68.434 | 65.532 | 68.083 | 67.009 |
| S231 | 63.708 | 60.367 | 62.891 | 60.367 |
| S238 | 63.753 | 60.686 | 62.688 | 60.686 |
| T242 | 68.33 | 65.2 | 67.747 | 66.698 |
| S248 | 61.068 | 56.125 | 65.056 | 61.926 |
| BMRB Avg Ser/Thr (sd) | Ser: 58.73 (2.08), Thr: 62.22 (2.59) | | Ser: 63.8 (1.48), Thr: 69.74 (1.65) | |

***Table 3.1 Unusual chemical shifts of TM polar residues suggests dimerization.*** *The Cα resonances of TM polar residues are downfield shifted from non-TM polar residues (shaded region). In particular, the Ser231/Ser238 Cα resonances are shifted more than 2 standard deviations from the mean. The EC-JM (Ser221) and TM (Ser248) are presented for comparison, these remain unshifted. Average values were obtained from the BMRB (http://www.bmrb.wisc.edu/ref_info/statsel.htm).*

**Structural Model of the EpoR TM Dimer Based Upon Intermonomer NOE Contacts**

***CN-NOESY***

Perhaps the most compelling evidence for a dimer interface driven by polar sidechain interactions comes from specific NMR measurements using a sample that contained an equimolar ratio of $^{15}N^{12}C$-labeled muEpoR$_{220\text{-}248}$ and $^{14}N^{13}C$-labeled muEpoR$_{220\text{-}248}$. The CN-NOESY experiment (*85*) transfers magnetization from an H-C group to an H-N group. Because one monomer of the dimer is labeled with $^{13}C$ and the other monomer is only labeled with $^{15}N$, NOEs are transferred *across the dimer interface*. Contacts made to either $^{12}C$- or $^{14}N$-containing peptide are effectively filtered out. Representative strips of the spectrum are presented in Figure 3.5. In these measurements, contacts are observed between the $^{13}C$-attached sidechain protons and amide nitrogens of Ser231, Val235, Ser238, Thr242, and Ala245, indicating these residues line the dimer interface. It is important to note that in general, stronger NOEs for a given residue are made for carbon atoms further out on the sidechain (*e.g.*, Val235 Cγ1, Cβ > Cα), which would be expected as those atoms are closer to the opposite amide in space. All strips show variable intensity peaks to unspecified Leu Cβ resonances. This is not surprising, as there are 12 Leu residues in the TMD and chemical shift degeneracy prevents accurate assignment.

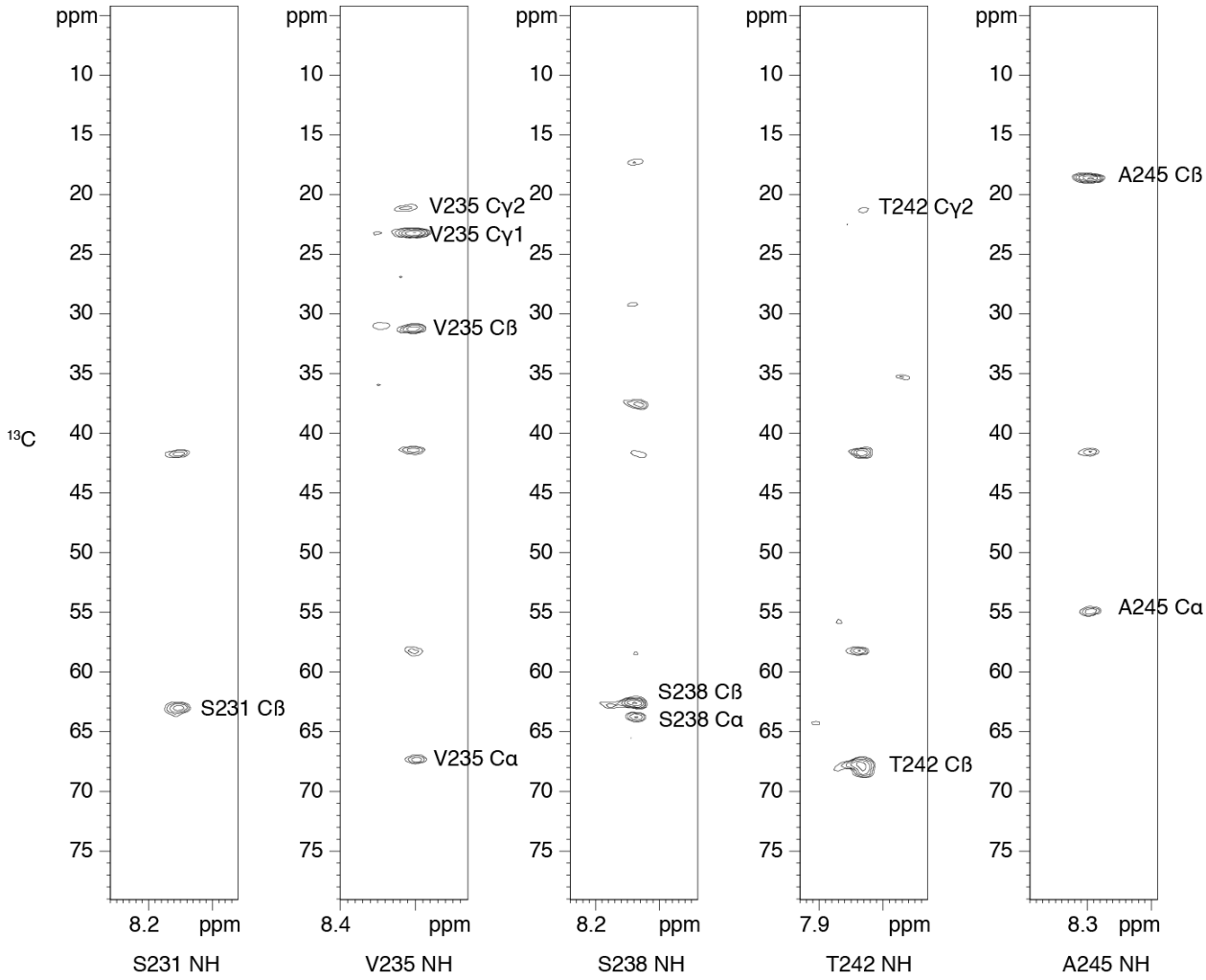



***Figure 3.5 CN-NOESY NMR Spectra of muEpoR$_{220-248}$ define TM dimer interfacial residues.*** *CN-NOESY spectra of $^{15}N^{12}C$ muEpoR$_{220-248}$ mixed with $^{14}N^{13}C$ muEpoR$_{220-248}$ (1:1 molar) demonstrate that S231, V235, S238, T242 and A245 line the TM dimer interface. The spectrum was collected on a 700 MHz spectrometer at 313 K with 32 scans. Sample was ~2 mM total protein concentration in 10 mM sodium phosphate, pH 7.0, 200 mM d38-DPC, 10% $D_2O$ (v:v).*

***CHI***

Prior to conducting the NMR experiments, computational searches for low-energy symmetric homodimers using the CHI computational modeling program (*135*) were conducted on the murine EpoR TM sequence. The interhelical axis distance was specified at 9.5 Å and the dielectric was set to 1, reflecting the relatively low permittivity of a membrane bilayer. More left-handed symmetric interfaces were found than right-handed ones, though generally the

energies were similar. The interface of the resulting symmetric structures generally contained at least one polar residue making an interhelical hydrogen bond. Interestingly, one of the lowest-energy left-handed structures (Figure 3.6) has Ser231, Val235, Ser238, Thr242, and Ala245 in the dimer interface. This low-energy interface was also found in computational searches that specified the interhelical axis distance at 10 Å.

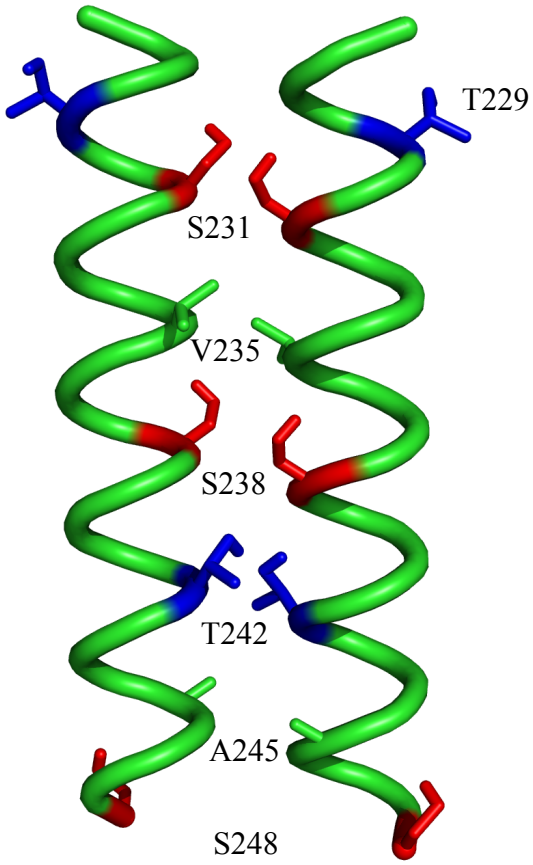



***Figure 3.6 CHI computational modeling predicts a low-energy dimer interface.*** *Computational searches for symmetric, low-energy structures using the muEpoR$_{220-248}$ sequence yield a left-handed coiled coil dimer with S231, V235, S238, T242, and A245 lining the dimer interface. The separation between the axes of the TM helices in the dimer (input parameter) is fixed at 9.5 Å.*

***X-Filtered NOESY***

The protein structure calculation depends upon collecting through-space NOE datasets. This requires proper and complete assignment of backbone and sidechain atom resonances. While the results of the CN-NOESY described above provide a low-resolution picture of the dimer interface, it suffers from the lack of stereospecific information regarding proton contacts. Two experiments were used to gain a higher resolution picture of the muEpoR$_{220\text{-}248}$ homodimer. First, the $^{1}$H-$^{13}$C NOESY-HSQC provides distance information for pairs of protons in the sample. These can be ambiguous as to whether the NOE contacts are inter- or intramonomer, due to the proximity of the residues in the interface. They can also be ambiguous if several atoms have the same chemical shift. Intermonomer NOEs were collected for a 1:1 $^{13}$C$^{15}$N labeled:unlabeled muEpoR$_{220\text{-}248}$ sample in $D_2O$ sodium phosphate (10 mM), 200 mM *d38*-DPC, using an NMR experiment that detects only NOEs transferred from an unlabeled to a labeled monomer (*84*). A third NOE dataset, a $^{1}$H-$^{15}$N NOESY-HSQC was collected with a sample in sodium phosphate, pH 7.0, 200 mM *d38*-DPC, 10% $D_2O$ (*v:v*). Constraints from the CN-NOESY experiment were manually sorted into strong/medium/weak crosspeaks by intensity and included in the unambiguous constraint list. The other NOE spectra were assigned using the program Aria2.3 and the assigned peaklist through successive iterations of structure calculations. Lists of ambiguous and unambiguous NOE constraints were generated after each round of calculation. The ambiguous NOE list was analyzed and converted to unambiguous NOEs, which are automatically included in the structure calculation. Representative slices from the X-filtered NOESY are presented in Figure 3.8. These data demonstrate that small and polar residues line the interface of the EpoR TM helix dimer, confirming the results from the CN-NOESY experiment (Figure 3.5). A final list of distance constraints was prepared from these assigned NOE spectra, a partial list is available in Table 3.2.

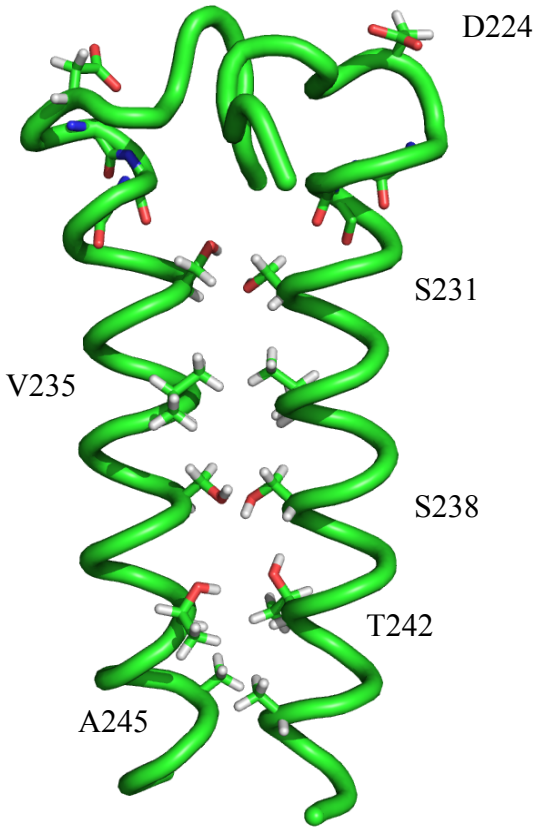



***Figure 3.7 muEpoR$_{220-248}$ dimer structure determined from solution NMR measurements.*** *Structure determination from NMR data of the muEpoR$_{220-248}$ protein in DPC micelles reveals a symmetric left-handed coiled coil dimer with S231, V235, S238, T242, and A245 lining the dimer interface.*

| Spectrum | Residues |
|---|---|
| CN-NOESY | Ser231Cβ - Ser231HN |
| | Val235Cβ – V235HN |
| | Val235Cα – V235HN |
| | Val235Cγ1 – V235HN |
| | Val235Cγ2 – V235HN |
| | Ser238Cβ - Ser238HN |
| | Ser238Cα - Ser238HN |
| | Thr242Cβ - Thr238HN |
| | Thr242Cα – Thr242HN |
| | Ala245Cβ – Ala245HN |

| | |
|---|---|
| | Ala245Cα – Ala245HN |
| | |
| X-Filtered NOESY | |
| | Ile227Hγ-Ile227Hδ |
| | Ile227Hα-Ile227Hδ |
| | Val235Hα-Val235Hγ |
| | Val235Hβ-Val235Hγ |
| | Ser238Hβ-Val235Hβ |
| | Ser238Hβ-Ser238Hα |
| | Thr242Hα-Thr242Hγ |
| | Thr242Hα-Ala245Hβ |
| | Ala245Hβ-Val243Hγ |

***Table 3.2 muEpoR$_{220-248}$ interhelical dimer contacts.*** *Intermonomer constraints were generated from the CN-NOESY experiment and the X-filtered NOESY. Each constraint was counted twice (From monomer A to monomer B and vice versa).*

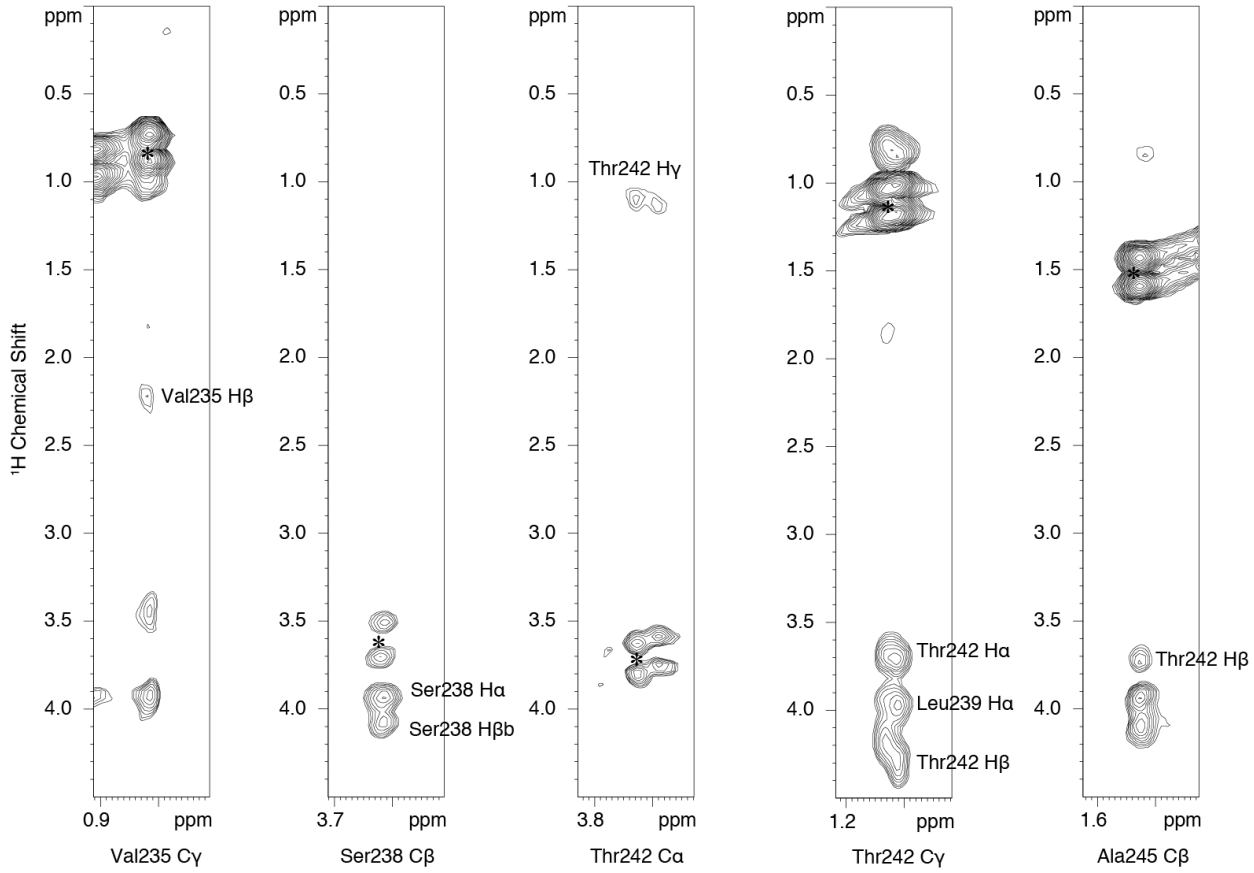


***Figure 3.8 X-filtered NOESY of muEpoR$_{220-248}$ defines TM dimer interfacial residues.*** *$^{15}N^{13}C$ muEpoR$_{220-248}$ mixed with $^{14}N^{12}C$ muEpoR$_{220-248}$ (1:1 molar) demonstrates that V235, S238, T242, and A245 line the TM helix dimer interface. Asterisks mark diagonal peaks. The spectrum was collected on an 800 MHz spectrometer at 313 K with 24 scans. Sample was ~2 mM total protein concentration in 10 mM sodium phosphate, pH 7.0, 200 mM d38-DPC, 100% $D_2O$.*

**TM Dimer Interface**

The structure of the muEpoR$_{220-248}$ in DPC micelles calculated from NMR measurements reveals that polar residues line the muEpoR$_{220-248}$ peptide TM dimer interface (Figure 3.7), with hydrophobic Leu sidechains facing the hydrophobic detergent tails. This interface correlates with low energy structures calculated from computational searches for symmetric TMD dimers. TMD dimerization mediated by polar residues has been demonstrated by asparagine mutagenesis studies of model polyleucine helices (*136*). Eilers *et al*. (*20*) have found that Ser and Thr residues mediate tight packing of helices in left-handed coiled-coil dimers. Furthermore, a clinically relevant mutation of the TpoR TMD (S505N) causes TM self-assembly in the TOXCAT system (*15*). The S505N mutation has been found in patients with thrombocythemia (*33*). These observations indicate that the TMD plays a critical role in receptor activation, and mutation of S505 to asparagine causes receptor hyperactivity via TM-mediated oligomerization.

The TM dimer interface presented for the muEpoR$_{220-248}$ TM dimer can be compared with interfaces proposed in the literature. Gurezka *et al*. (*61*) first proposed that muEpoR TM self assembly was mediated by a leucine heptad repeat (Leu zipper) based on sequence similarity of the muEpoR TMD with self-assembling model peptides containing a Leu heptad repeat (*61*). Ruan *et al.* (*64*) took this work a step further and demonstrated by asparagine scanning mutagenesis that several Asn mutations of the TMD caused greater activity in the TOXCAT system; these mutations followed a different heptad repeat pattern that included Ser231 and Ser238 at the “a” positions and Leu234 and Leu241 at the “d” positions (SxxLxxx-SxxLxxx-A).

However, characterization of the L241N mutation in the context of the full Epo receptor in BaF3 cells demonstrated that this receptor was not constitutively active, but could be activated by Epo. Further work investigated mutation to asparagine of the other residues in the interface (Leu234, Ser238, Leu241, and Ala245) in the context of the full receptor (*77*). Curiously, and without explanation, Ser231 is omitted from this study. However, Thr242 is added, a residue not previously thought to be in the interface. None of the asparagine mutants were constitutively active and only T242N exhibited cell surface expression comparable to wild type. However, even though Epo could activate T242N, the receptor's cellular staining pattern differed from the wild type EpoR. While the authors ultimately make comparisons between helix packing density and biological activity, the results are interpreted as if asparagine were driving specific TM dimerization, which is not clear as the results could also be due to receptor dimer clustering or trimer formation. Asparagine can drive trimer formation, as seen in mutants of αIIbβ3 integrin (*137*), and asparagine containing TM peptides (*138*). Further examination of the TOXCAT method demonstrates that asparagine insertions in the background of a leucine heptad repeat can cause stronger than normal TOXCAT readings due to higher-order oligomer formation that may mislead interpretation of results (*139*). This may be the case in the L241N mutant, which showed TOXCAT assay results ~4 times higher than other asparagine mutants (*64*).

The interfaces defined by the Ruan *et al*. (*64*) and Becker *et al*. (*77*) studies differ from the interface defined here by ~52°, but share the Ser231, Ser238, and Ala245 residues. However, this would leave Thr242 outside the interface and exposed to the hydrophobic lipid environment. This is in direct contradiction with the T242N results obtained by Becker *et al*. (*77*). While the two Thr sidechains (one for each monomer) could back hydrogen bond to the helix backbone (*140*), this would otherwise lead to an increase in energy equivalent to 0.5 to 2.0 kcal/bond (*141*).

The interface presented here from NMR measurements satisfies as many of those hydrogen bonds as possible by burying the polar residues in the dimer interface. In support of this model the interface defined here is the only interface identified by Seubert *et al*. (*76*) that could be activated by the viral TM protein gp55-P, which has been shown to activate EpoR through TM-TM interactions (*14*).

**Active Interfaces**

The results from the muEpoR$_{220-248}$ structure determination are presumed to be from the inactive TM dimer interface. However, asparagine mutants determined by Becker *et al.* (*77*) that are shared by this interface (Ser238, Val235, Thr242, and Ala245) are able to be activated by Epo. If the active dimers have a different interface than the inactive dimers (as in TM rotation), then the presence of asparagine should make it more difficult to activate, which does not appear to be the case based on Epo response curves. These results suggest three possibilities: 1) the activation is able to overcome presumed asparagine-induced dimerization, 2) the inserted asparagine does not affect the inactive dimer interface, or 3) the inactive and active interfaces are the same. None of the TM residues, when mutated to cysteine, caused constitutive activation, though L241C, L244C, and A245C could be crosslinked to varying degrees (*74*). However, all responded normally to Epo. The combination of these results suggests that some aspect of EpoR activation with respect to the TMD dimerization is misunderstood, and more work remains to be done.

**Human Receptor TM Domain**

It is unclear if the human EpoR TMD uses the same interface. NMR measurements ($^{1}$H-$^{15}$N-HSQC) conducted on the same segment of the human receptor huEpoR$_{221-249}$ (numbering is slightly different because of an single amino acid insertion in the ECD) demonstrates a very

different spectrum (Figure 3.9). Differences of this magnitude would not be expected from a sequence that differs in only 3 amino acids in the center of the helix but uses the same dimerization interface. This is especially true for residues not in the interface or at the ends of the helix; if the dimerization interface remained the same, then those residues should be in a relatively similar environment in both peptides. This result suggests that either the dimerization interface has changed, or the huEpoR$_{221\text{-}249}$ is no longer a dimer. However, AUC studies have demonstrated that the huEpoR TM dimerizes in detergent micelles (*65*), and other studies have determined a propensity of the huEpoR TM sequence to self-associate (*19*). If the human receptor has a different TM dimer interface than the mouse does due to the loss of Ser238, then this would suggest that Ser238 is important for dimerization, and explain the inability of the viral TM protein gp55-P to activate the human EpoR (*39*).

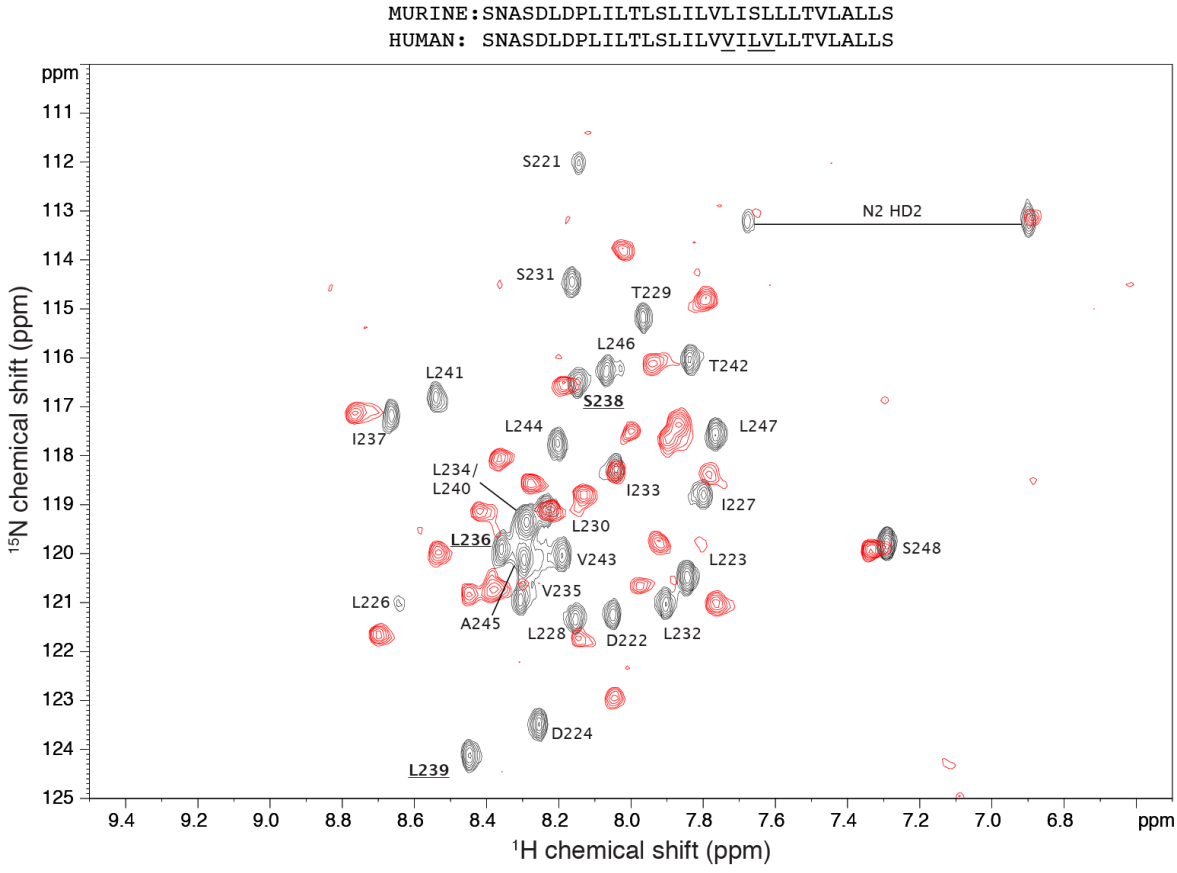
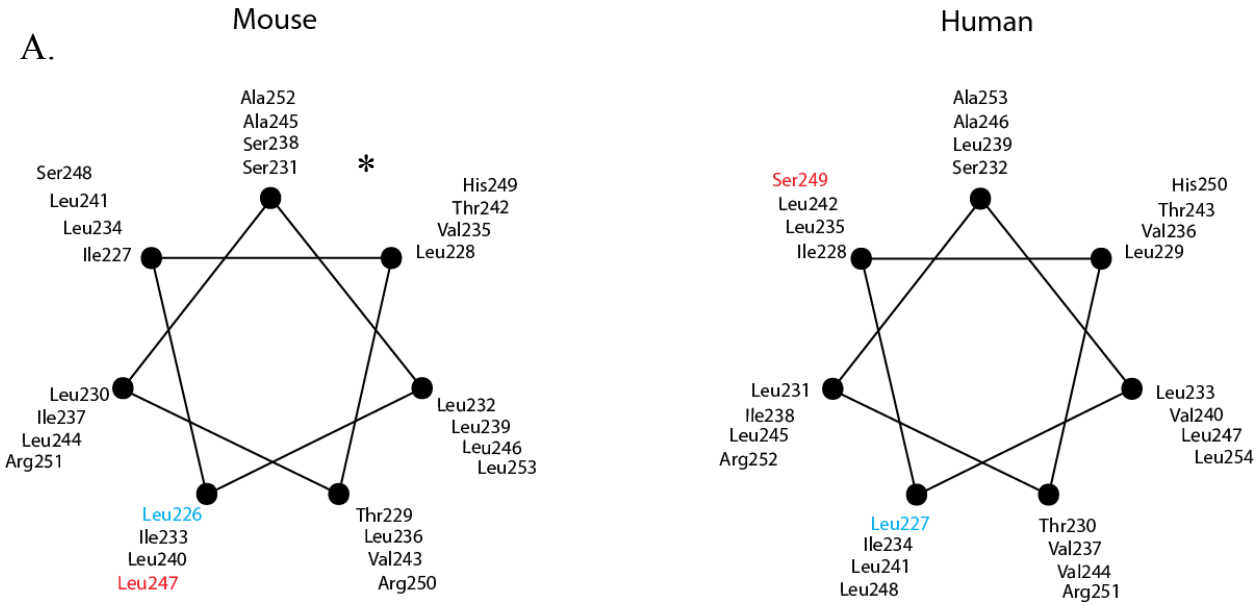


***Figure 3.9 Comparison of muEpoR$_{220-248}$ and huEpoR$_{221-249}$ $^{1}$H-$^{15}$N HSQC spectra.*** *Overlay of huEpoR$_{221-249}$ (red) and muEpoR$_{220-248}$ spectra demonstrates a completely different HSQC spectrum for the huEpoR$_{221-249}$ as compared to the muEpoR$_{220-248}$ spectrum. Underlined residues represent sequence differences between mouse and human (all peak labels refer to mouse amino acid assignments). Both peptides are at ~1 mM in 10 mM sodium phosphate, pH 7.0, 200 mM DPC, 10% $D_2O$ (v:v). muEpoR$_{220-248}$ was collected with ns=2, huEpoR$_{221-249}$ with ns=2 on a 700 MHz spectrometer at 313 K.*

What would the human dimerization interface look like if it were different than the murine one? Figure 3.10 (panel A) shows helical wheel diagrams for the two EpoR TM sequences. The muEpoR$_{220-248}$ structure presented uses the dimer interface lined with Ser231, Val235, Ser238, Thr242, and Ala245 (marked with asterisk). AUC experiments (*65*) and TOXCAT experiments (*19*) indicate that the human sequence associates less strongly than the mouse sequence. If the same helix interface is used in the human receptor, the replacement of Ser238 with leucine in the human sequence may destabilize the helix dimer, assuming the polar Ser238 sidechain contributes to dimerization. Alternatively, mutation of Ser238 to leucine may force the use of a different interface. Leucine residues are not uncommon in TM helix dimer interfaces, and are almost twice as common in the "d" position than the "a" position (*142*), so why a leucine substitution at position 238 would be so destabilizing is unclear. If another dimerization interface were to be used by the huEpoR, it is unclear what interface that would be. However, maximizing the number of small or polar residues in the interface and restricting leucine residues to the "d" position suggests that an interface ~52° counterclockwise (Leu235, Leu242, Ser249; Ser232, Leu239, Ala246) might be used. This is the same interface described for the murine TM dimer by Ruan *et al*. (*64*). Computational searches for low-energy symmetric TM dimer interfaces suggest another possibility (Figure 3.10, Panel B). Comparison between the symmetric dimers of the mouse and human sequences show that for both sequences, sidechains of Thr242/243 form interhelical hydrogen bonds. However, the mutation of Ser238 in the

muEpoR to leucine in the huEpoR forces a larger crossing angle in the human TMD (6° vs. 16°, respectively). Interestingly, mutation of Leu239 to serine in the human TM sequence returns the structure to a narrow crossing angle (7.5°) as seen in the mouse WT sequence model. The mouse WT TM dimer energy is lower than that of the human WT TM dimer (-60.58 kcal/mol, -58.8 kcal/mol, respectively), which agrees with the difference in association seen using AUC experiments (*65*). These data may also agree with other computational studies correlating the packing density of different murine EpoR TMD-based sequences with cellular activity (*77*). In that study, Thr242 was mutated to asparagine, glutamine and alanine, and Epo-induced receptor activity increased with decreasing sidechain length (increased packing density, Ala>WT>Asn>Gln). If decreased packing density of the EpoR TMD negatively affects receptor activity, then the lower activity of the human receptor is explained by the larger crossing angle.

A.




B.

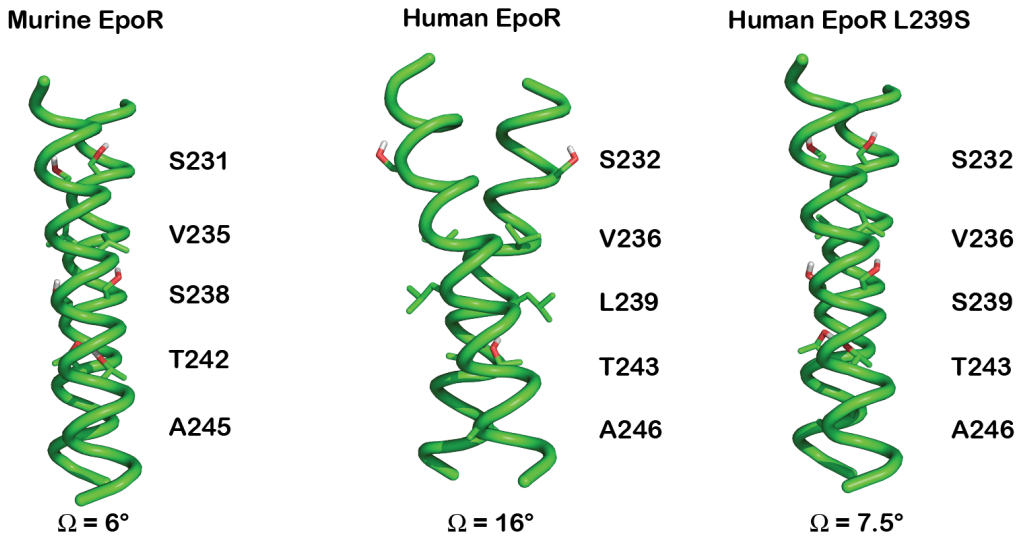



***Figure 3.10 muEpoR and huEpoR helix interfaces. A. muEpoR$_{220-253}$ and huEpoR$_{221-254}$ helical wheel diagrams.*** *Left-handed coiled-coil (LHCC) helical wheel diagrams demonstrate relative positions of the residues in the helix if the EpoR TM dimer is a LHCC. The murine and human are positioned to be in the same register based upon the putative start of the TM helix (in blue). The putative end is marked in red. An asterisk marks the TM dimer interface in the muEpoR$_{220-248}$ dimer structure. B. Computational modeling of mouse and human TM dimer interfaces. Computational searches for low-energy symmetric TM dimer structures for the mouse and human sequences result in interfaces with the same general interface, but with different crossing angles (Ω). Mutation of Leu239 to serine in the human sequence recovers the narrow crossing angle. The separation between the axes of the TM helices in the dimer (input parameter) is fixed at 9.5 Å.*

How could a change in the EpoR TM dimer crossing angle result in a difference in its ability to be activated by gp55-P? While it has been shown that the ability of the gp55-P TMD to interact is dependent upon a single serine residue in the TMD, the same study demonstrates that gp55-P is able to activate the mouse TMD with S238 mutated to alanine (*39*), so S238 is unlikely to be intimately involved in the interactions between the EpoR and gp55-P. This conclusion is despite presentation in the same paper of an EpoR-gp55-P TM heterodimer model with Met390 from gp55-P and EpoR Ser238 in the interface of the heterodimer. The change in crossing angle (Ω) in the EpoR TM homodimer (Figure 3.10) demonstrates a second, subtle consequence of leucine at position 239 of the huEpoR TMD, namely a slight rotation of the TM helices with respect to each other and a change in the interface of the EpoR TM dimer. Ser232 is clearly

rotated to a position outside the interface; this rotation is reversed in the huEpoR L239S mutation. This rotation in the huEpoR WT TMD may result in an obfuscation of the TM helix dimer 'outerface' that would be available for interaction with gp55-P in the native EpoR TM dimer structure. Therefore, if specific interactions between the EpoR and gp55-P TMDs are disrupted by changes in the sequence of the TMDs, the disruption is likely the result of two different changes. The first is a change in the relative homodimer's (EpoR:EpoR or gp55-P:gp55-P) ability to retain their native interfaces. Secondary to that is the inability to form specific EpoR:gp55-P TMD interactions that are critical to TMD repositioning related to the EpoR activation.

**Helix Cap**

The i+4 backbone hydrogen bonding network of α-helices contribute to the stability of this particular type of secondary structure. An important and particularly interesting characteristic of the α-helix has to do with the structural transition at the ends of the helix. At the ends of the helix, the characteristic i carbonyl oxygen to i+4 amide proton hydrogen bonding pattern is disrupted, leaving several unsatisfied hydrogen bonds. These bonds could be satisfied by water, but a water NOESY experiment (*98*) demonstrates that these amide protons are not in exchange with water (data not shown). If a hydrogen bond contributes -0.5 to -2.0 kcal/mol of energy, then leaving 3 unsatisfied hydrogen bonds means an increase in the free energy of the system of 1.5-6 kcal/mol per monomer (*141*). In order to minimize this energy increase, a sequence dependent cap structure forms to stabilize the end of the helix (*143*). Typically, these involve both polar and non-polar interactions. In an N-terminal cap, residues with a polar sidechain forms a hydrogen bond with an amide nitrogen further C-terminal to the capping

residue. Further stability comes from the packing of non-polar sidechains. This cap structure successfully reduces the amount of free energy in the system, stabilizes the helix termini and prevents helix unraveling.

Helix cap structures have been well studied and consensus sequences for both the C-termini and N-termini caps have been determined (*143*). Unfortunately, a survey of the currently available NMR solution structures of TM dimers (glycophorin A, T-cell receptor ζζ, EphA2, and BNIP3), demonstrates none with a definitive helix cap structure (Table 3.3). This is in spite of each of the sequences possessing a N-terminal capping motif sequence, as predicted by the consensus sequences of Aurora and Rose (*143*). This begs the question as to whether the lack of a cap structure is the result of a post-NMR data collection error (bias in structure calculation) or a problem with everything leading up to the structure calculation input (NMR data collection, data analysis or a problem with the system being studied). One possibility is that structure calculation algorithms could be to blame; the necessary forcefields may not be 'mature' enough to correctly handle correct structure calculation in the odd chemical environment at the polar/nonpolar interface (membrane boundary). However, the consensus sequences for the current cap motifs were in part determined through PDB structure searches. Given that these structures were calculated with similar or identical forcefields, an error at this step is unlikely, and we must search for other ways to explain the lack of defined helix caps in the structures listed below.

| Protein (PDB ID) | Sequence | Potential Cap Motif |
|---|---|---|
| Glycophorin A (1AFO) | VQLAHHFSEPEITLIIFGVMAGVIGTILLISYGI | IIIb (F) |

| | | |
|---|---|---|
| EphA2 (2K9Y) | EFQTLSPEGSGNLAVIGGVAVGVVLLLVLAGVGFFIHRRRK | Ia (V), Ib (I) |
| BNIP3 (2KA1) | GGIFSAEFLKVFLPSLLLSHLLAIGLGIYIGRRLT | Ia (L) |
| T-cell receptor ζζ (2HAC) | DSKLCYLLDGILFIYGVILTALFLRVKFSRSAD | Ia (L) |

***Table 3.3 Potential helix caps in currently available PDB TM dimers.*** *Several TM homodimer structures are available for comparison. Each of these has a potential N-terminal cap sequence motif, yet none of the structures demonstrate such a feature.*

Interestingly, in examining our spectral data and the resulting calculated structure, an N-terminal helix cap is formed involving the sidechain of Asp224 and the amide proton of Leu226 (Figure 3.10). Furthermore, this cap is predicted from the protein sequence based on the Aurora and Rose Ia (h-xpxhx) consensus sequence, where the first residue of the capping motif is Leu223 and the N1 residue is Pro225. Statistically, proline is the most common residue at the N1 position for most of the N-terminus capping motifs (*143*). At the N3 position is Ile227, which makes contacts with the N' hydrophobic residue (Leu223). This structure is directed by our through-space NMR data, Table 3.4 shows these contacts in the CN-NOESY spectrum. Specific contacts are made between Asp224 and Ile227, Pro225 and Ile227/Leu228. The sidechain of Asp224 seems to be within hydrogen bonding distance of the backbone amide NH groups of Leu226. There are also contacts between Leu223 and Ile227, indicating that hydrophobic packing contacts between the sidechains of these two residues contribute to the structure of the cap. Interestingly, when Asp224 is mutated to cysteine, receptor processing and trafficking to the membrane is affected, leading to about half the wild type receptor at the cell surface (*75*). The receptor that does make it to the membrane is capable of binding Epo, but has a decreased level

of activity. It is plausible that this decreased trafficking and activity is due to a disruption of the local cap structure.

| Spectrum | Contact |
|---|---|
| CN-NOESY | D222HN-S221Cβ/Cα |
| | D222HN-L226Cα |
| | L223HN-S221Cα |
| | L223HN-D222Cα |
| | L223HN-P225Cβ/Cd |
| | L223HN-L226Cβ |
| | D224HN-I227Cβ |
| | D224HN-I227Cγ1/Cγ2/Cd |
| | D224HN-L223Cγ |
| | I227HN-L232Cδ |
| | L228HN-P225Cα |
| | L228HN-I227Cβ/Cγ1/Cγ2 |
| | L228HN-S231Cβ |
| | T229HN-P225Cα/Cβ |
| | T229HN-L223Cγ |
| | T229HN-L226Cα |
| | T229HN-L230Cγ |
| | T229HN-L232Cδ |

***Table 3.4 Helix cap contacts in NOESY datasets.*** *Contacts defining the helix cap structure in the muEpoR$_{220-248}$ are listed. Contacts between residues S221-L232 are available to define the structure.*

Previous computational work by our group on the murine EpoR TMD indicated that a cap structure forms between the sidechain OH of Thr229 and the amide proton of Leu232, one helical turn lower than we see in our current NMR-based structure (*74*). The differences in the two murine EpoR TM cap structures (Figures 3.11, 3.12) could have several different explanations. First, the Kubatzky structure is the result of computational studies based upon only two carbonyl chemical shifts (Leu228, Leu230). These chemical shifts were collected using solid-state NMR methods from a peptide reconstituted into multilamellar lipid vesicles (DMPC). Though counterintuitive, the DMPC bilayer has a shorter hydrophobic cross section (24 Å) than

a DPC micelle (~30 Å) (*144, 145*). Therefore, a longer hydrophobic helix would be expected in the DPC sample. It could be explained by the differences in surface curvature of DPC detergent versus DMPC lipid. Detergent micelles have an inherent surface curvature that is greater than that of the multilamellar lipid vesicles formed by the solid-state NMR preparation, which may induce structural variations, especially at the helix termini. Perhaps the greatest drawback of the solid-state based structure is that this structure was calculated based upon only two carbonyl chemical shift constraints resulting from solid-state NMR studies in lipid bilayers. In contrast, the solution NMR cap structure is the result of multiple through space contacts and measurement of chemical shifts of all relevant atoms.

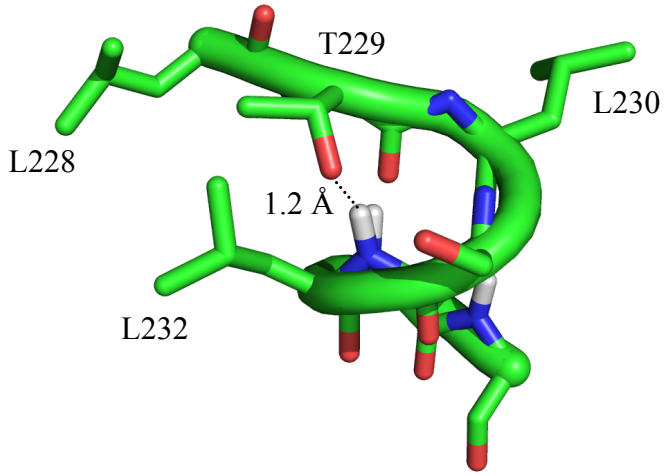



***Figure 3.11 Previous muEpoR dimer helix cap structure.*** *Helix cap structure from Kubatzky et al. (74) indicating that the sidechain OH of T229 forms a hydrogen bond with the HN of L232. The sidechains of L228 and L232 form hydrophobic contacts.*

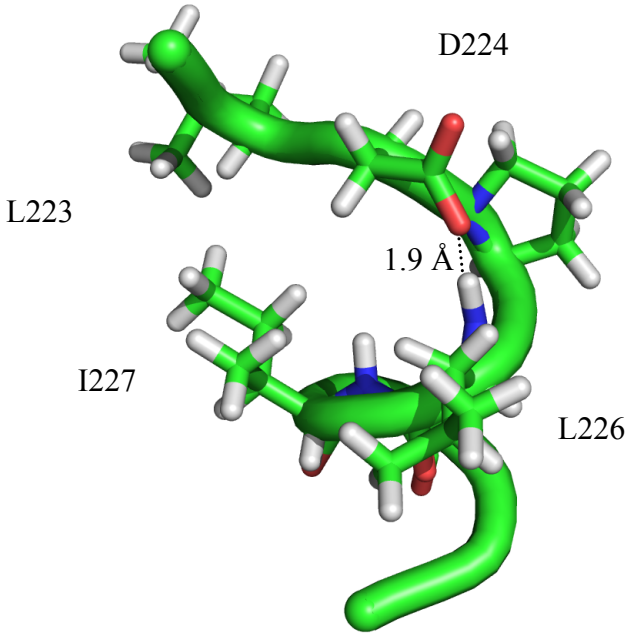



***Figure 3.12 Current muEpoR dimer helix cap structure.*** *Helix cap structure demonstrates D224 sidechain capping the helix by forming hydrogen bonds with the NH of L226. The sidechain of L223 makes contacts with the sidechain of I227.*

## Chapter 4-Structural Changes in the Erythropoietin Receptor Upon Activation

To begin to answer the question of how the TMD is involved in the function of the EpoR, we extend the sequence of the construct to include the conserved intracellular Box 1 region. Also added are several amino acid at the N-terminus to capture a larger portion of the EC-JM region, a part of the receptor which captures a number of positions that modulate receptor activity when mutated to cysteine. Figure 4.1 depicts the sequence of the muEpoR$_{218\text{-}268}$ peptide and the regions of the receptor added (red text) for these studies. The underlined are those residues that, when mutated to cysteine (Leu223, Leu228) modulate receptor activity (*74, 75*).


***Figure 4.1 Amino acid sequence of the recombinant muEpoR$_{218\text{-}268}$ peptide.*** *The murine EpoR TM Box 1 construct includes 20 residues of the ICD corresponding to the region including the Jak2 binding Box 1 region (PGIPSP). Underlined are residues that, when mutated to cysteine, modulate receptor activity. Also included are the residues of the switch region (L253, I257, W258) demonstrated to be important for Jak2 activation.*

This construct retains the hydrophobic TMD and adds three residues to the N-terminus (LTA). This N-terminal extension represents most of the remainder of the sequence between the WSXWS motif in the D2 domain and the TMD. Added to the C-terminus are 20 residues that represent the "switch" region and the Jak2 binding site, Box 1. The addition of this intracellular region serves two purposes. First, it would allow interaction studies between the Jak2 FERM domain and this longer EpoR TM containing peptide, as the Box 1 region is demonstrated to be necessary for association of these Jak2 with the EpoR (*46, 146*). More importantly, it allows the assessment of intracellular structural changes upon the simulation of the active state. Two

methods of simulating the active state are specific cysteine mutagenesis or the addition of the gp55-P TM peptide, which activates the EpoR (*38*). Therefore, the addition of more of the protein sequence surrounding the TMD adds functional relevance and context for examining the propagation of structural changes to the ICD upon activation.

**Cysteine Mutants**

Two important cysteine mutagenesis studies have been performed on the EpoR TMD and surrounding regions (*74, 75*). The difference between the two studies is the method of crosslinking, passive in Lu *et al*. (*75*) versus active with the lipid soluble linker ortho-phenylenedimaleimide (o-PDM) in Kubatzky *et al*. (*74*). Both studies identified L223C as a constitutively active mutant. L226C and I227C were also identified to be constitutively active, but at a lower level than L223C. D222C, P225C and L228C are not constitutively active, but can respond to Epo (wild-type like), though P225C responds at lower levels. D224C is an interesting case, as it is expressed at the cell surface at reduced levels (~50%), and is not processed through the Golgi (endo-H negative). It shows a very small amount of constitutive activity, but cannot respond to WT levels when exposed to Epo.

The Kubatzky study (*74*) differs from the Lu *et al.* (*75*) study because of the advantage of being able to actively crosslink cysteines inside the TMD, using the membrane-permeable crosslinker o-PDM, where disulfides normally do not form. Interestingly, crosslinked receptor is found with L241C, L244C, and A245C. However, these receptors are not constitutively active, though oddly both Leu241 and Leu244 are in the "active" interface described in Seubert *et al*. (*76*). Recombinant L223C and L228C muEpoR$_{218\text{-}268}$ are used in this study to approximate the active and the WT-like receptor, respectively.

**Wild-Type muEpoR$_{218\text{-}268}$**

Figure 4.2 shows the $^{1}H$-$^{15}N$-HSQC spectrum of the muEpoR$_{218\text{-}268}$ construct in DPC detergent micelles. This spectrum has more peaks than the muEpoR$_{220\text{-}248}$ spectrum (Figure 3.2), reflecting the extra residues added to the construct sequence. Peaks for residues in the TMD are generally of lower intensity, reflecting an overall lower mobility of this region than the N- and C-termini. Despite the added molecular weight, the spectrum peaks are generally well resolved, indicating that the preparation is homogenous. Significant peak overlap exists between the muEpoR$_{218\text{-}268}$ and the muEpoR$_{220\text{-}248}$, indicating that the TM conformation is preserved in the larger construct (Figure 4.3). Chemical shift deviation (Δδ) is calculated for each peak using the formula:

$$\Delta\delta = \sqrt{\frac{(CS_{H1} - CS_{H2})^2 + (CS_{N1} - CS_{N2})^2 / 5}{2}}$$

where $CS_{H1}$/$CS_{N1}$ and $CS_{H2}$/$CS_{N2}$ are the chemical shifts of the relevant hydrogen and nitrogen in the first spectrum minus the second spectrum (*147*). Most of the intermonomer residues identified by the TMD structure (Ser231, Val235, Ser238, Thr242) are in the same position in the muEpoR$_{218\text{-}268}$ spectrum. The only intermonomer peak that is not in the same position is Ala245. It is likely shifted because the entire C-terminus is shifted due to the additional C-terminus residues. Full 3D assignment reveals that all amino acids can be assigned with the exception of Leu226 and Arg251. Thr252 has very low intensity, and Arg250 is buried in the more intense Leu223 resonance, but can be assigned by examining the 3D experiment strips. The low intensity of these peaks is likely due to flexibility at the intracellular membrane boundary. At this site, where the TM helix emerges from the detergent micelle and transitions from α-helical to random

coil, structural heterogeneity is not unsurprising. It is unclear if the structural heterogeneity is an integral part of the system or a result of the study methodology.

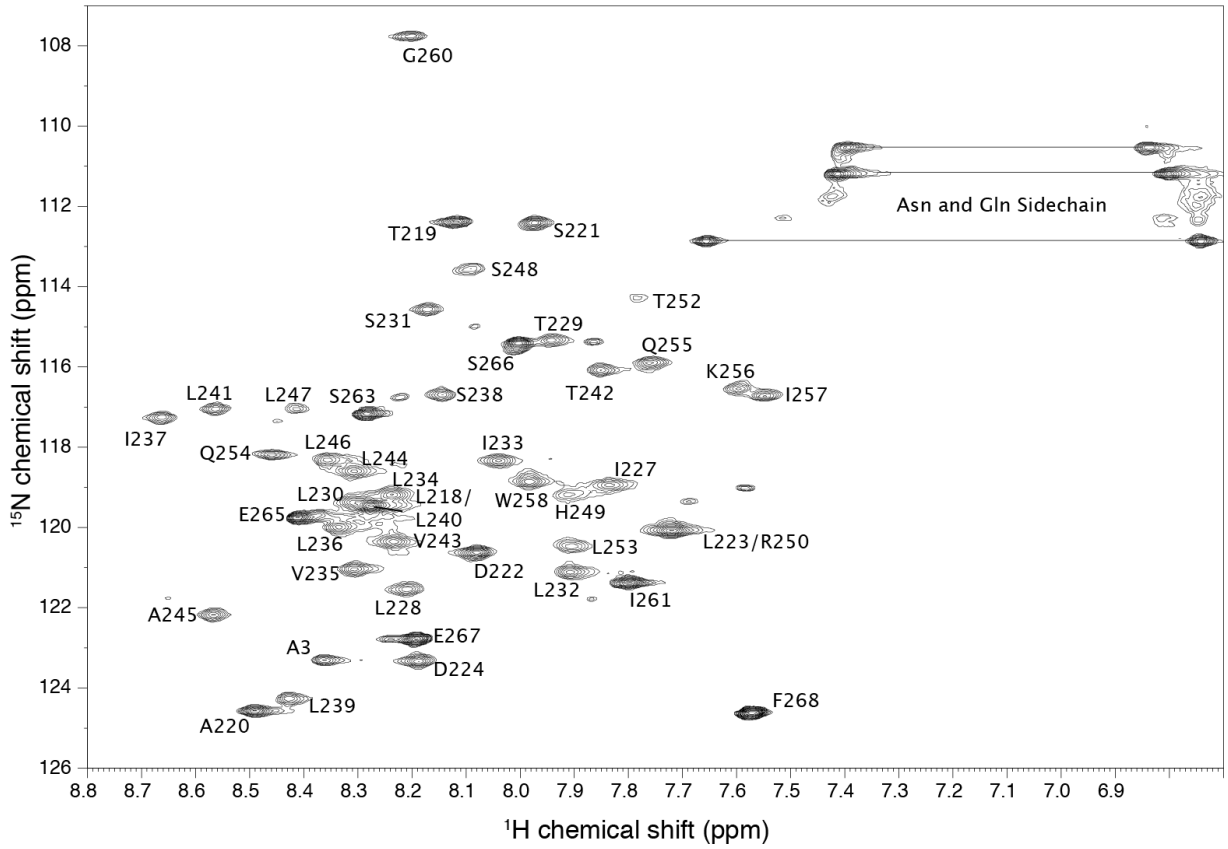



***Figure 4.2 The murine EpoR WT TM Box 1 $^{1}$H-$^{15}$N-HSQC spectrum.*** *The spectrum for muEpoR$_{218-268}$ shows good resolution given the size of the detergent-protein complex. Peak assignments are shown on the spectrum; only 2 amino acids remain unassigned (L226, R251). Spectrum was collected on a Bruker 700 MHz spectrometer at 313 K, with 32 scans. Sample is ~1 mM protein, 10 mM sodium phosphate, pH 7.0, 200 mM d38-DPC, 10% $D_2O$ (v:v).*

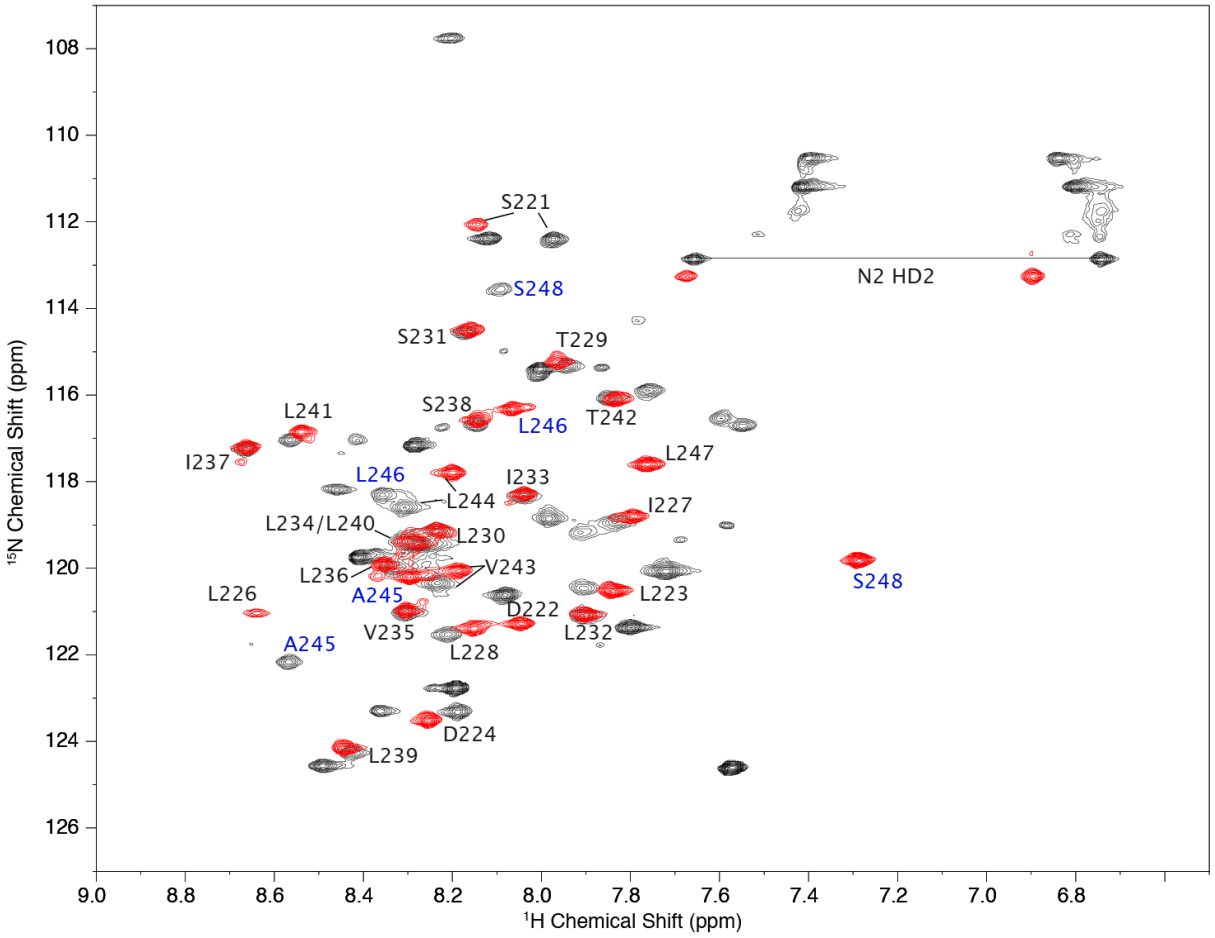



***Figure 4.3 Overlay of muEpoR$_{220-248}$ $^{1}$H-$^{15}$N-HSQC on muEpoR$_{218-268}$ $^{1}$H-$^{15}$N-HSQC.*** *Only five of the resonances in the transmembrane region (L244-S248) shift when the peptide is lengthened to muEpoR$_{218-268}$ (only S221-S248 are labeled). Residues in the TMD are numbered, numbers in blue represent a significant peak shift (and peaks for each construct are labeled). Generally these shifts occur in the last turn of the helix (A245-S248). Sample is ~1 mM protein, 10 mM sodium phosphate, pH 7.0, 200 mM d38-DPC, 10% $D_2O$ (v:v).*

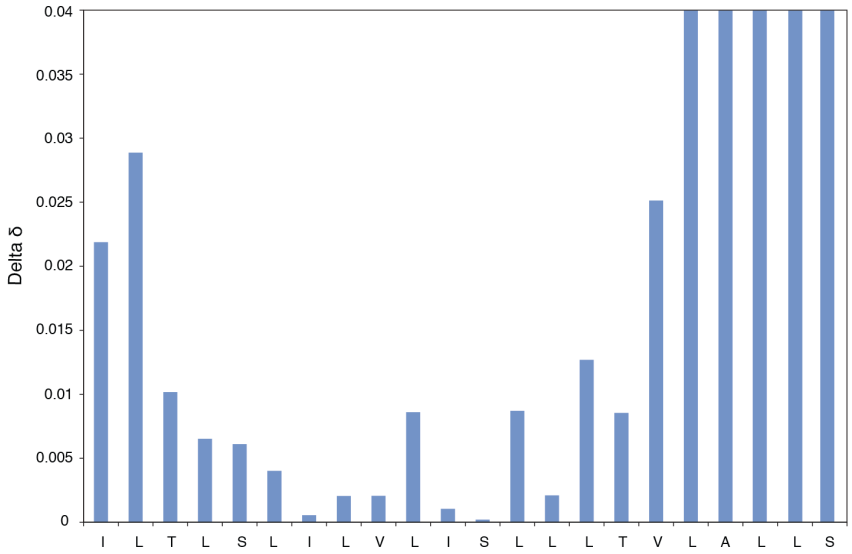



***Figure 4.4 Chemical shift differences of the TM domain between muEpoR$_{220-248}$ and muEpoR$_{218-268}$.*** *The core residues of the TM domain, including the dimer interface residues S231, V235, S238, and T242 remain in the same position, an indication of conformation preservation.*

**Simulating the Active State: L223C and gp55-P**

In order to begin probing the functional mechanism for the EpoR, spectra of the WT (inactive) and the active state were compared. Peptides of the muEpoR$_{218-268}$ with the L223C mutation were expressed, purified, and reconstituted into DPC micelles. Some optimization was necessary to obtain high quality spectra. First, DTT (20 mM final concentration) was added to a dilute (~2 ml) sample, and then the sample was dialyzed versus 3 changes of 10 mM sodium phosphate, pH 7.0 (1 kDa MWCO). This prevents TM dimers from two micelles from crosslinking, which degrades spectral quality because of slower molecular tumbling. Then the sample was concentrated to 250 uL and loaded into the Shigemi NMR tube. Other than these deviations, the NMR samples were prepared in the same manner as the WT TM peptides.

The L223C muEpoR$_{218-268}$ spectrum-WT spectrum comparison is shown in Figure 4.5. It is clear from the spectral comparison that there are few shifted peaks from the WT spectrum, an indication that the structures are comparable. As expected, peaks in the vicinity of L223C were shifted quite dramatically (Asp222, Asp224), likely due to perturbation to the local structure.

Chemical shift deviation calculation of the region from Ile227-Phe268 (Figure 4.6) shows that outside of the residues immediately surrounding L223C, His249 shifts the most of all residues, followed by Gly260 and Leu253. These results indicate that a mutation in the extracellular region can effect a change in the intracellular region of the protein. The significance of the large shift in Ile227 is thought to be due to local disruption of the helix cap structure, where the side chain of Asp224 hydrogen bonds to the NH of Ile227.

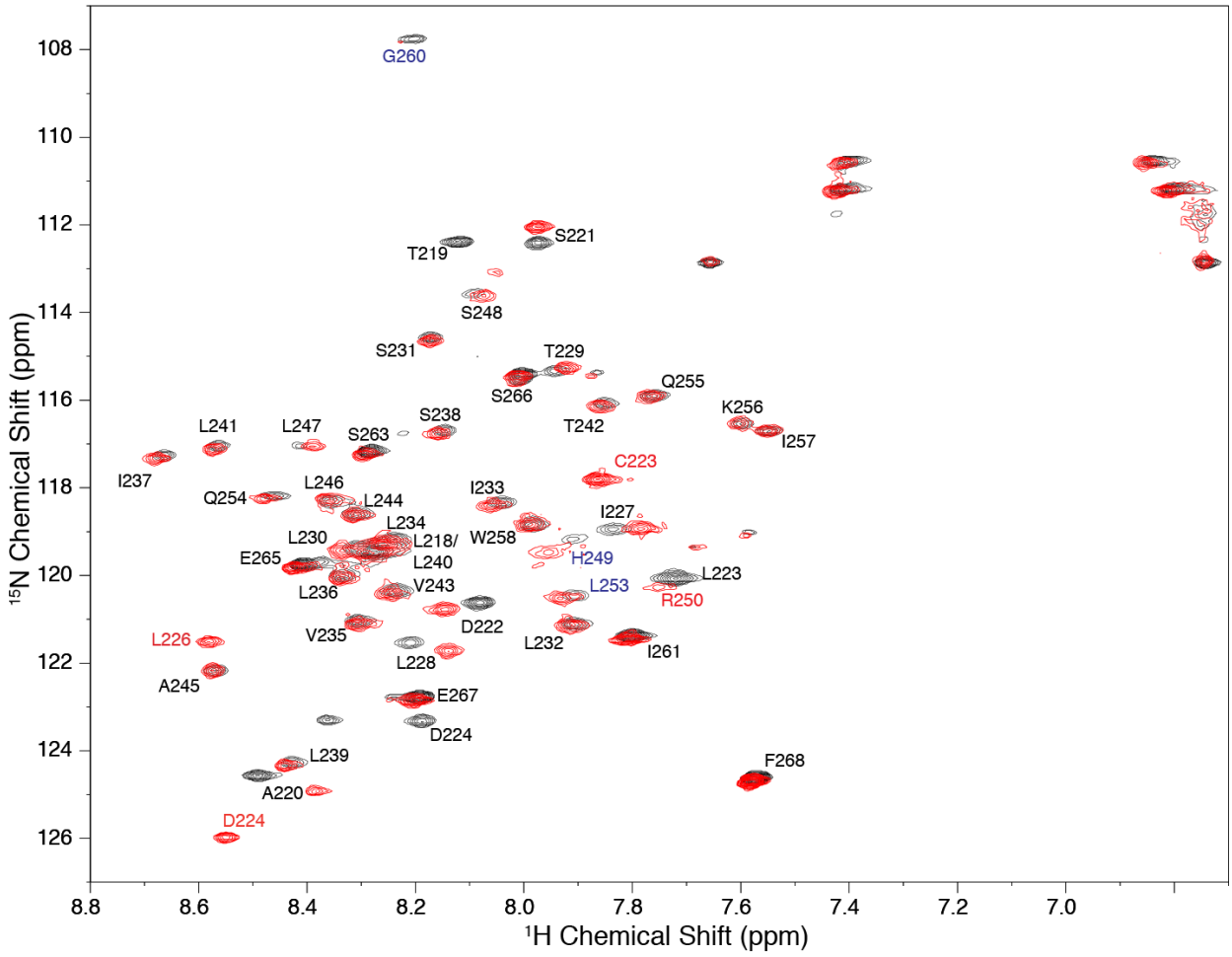



***Figure 4.5 $^{1}H$-$^{15}N$-HSQC comparison, L223C (red) compared to WT (black).*** *The $^{1}H$-$^{15}N$-HSQC of the L223C muEpoR$_{218-268}$ was prepared as described and a spectrum was collected at 313 K with 32 scans. The overlay demonstrates that while many peaks overlap, several do not. Red text denotes new peaks/peak positions for the L223C mutant. H249, L253, and G260 are in blue. Sample is ~1 mM protein, 10 mM sodium phosphate, pH 7.0, 200 mM d38-DPC, 10% $D_2O$ (v:v).*

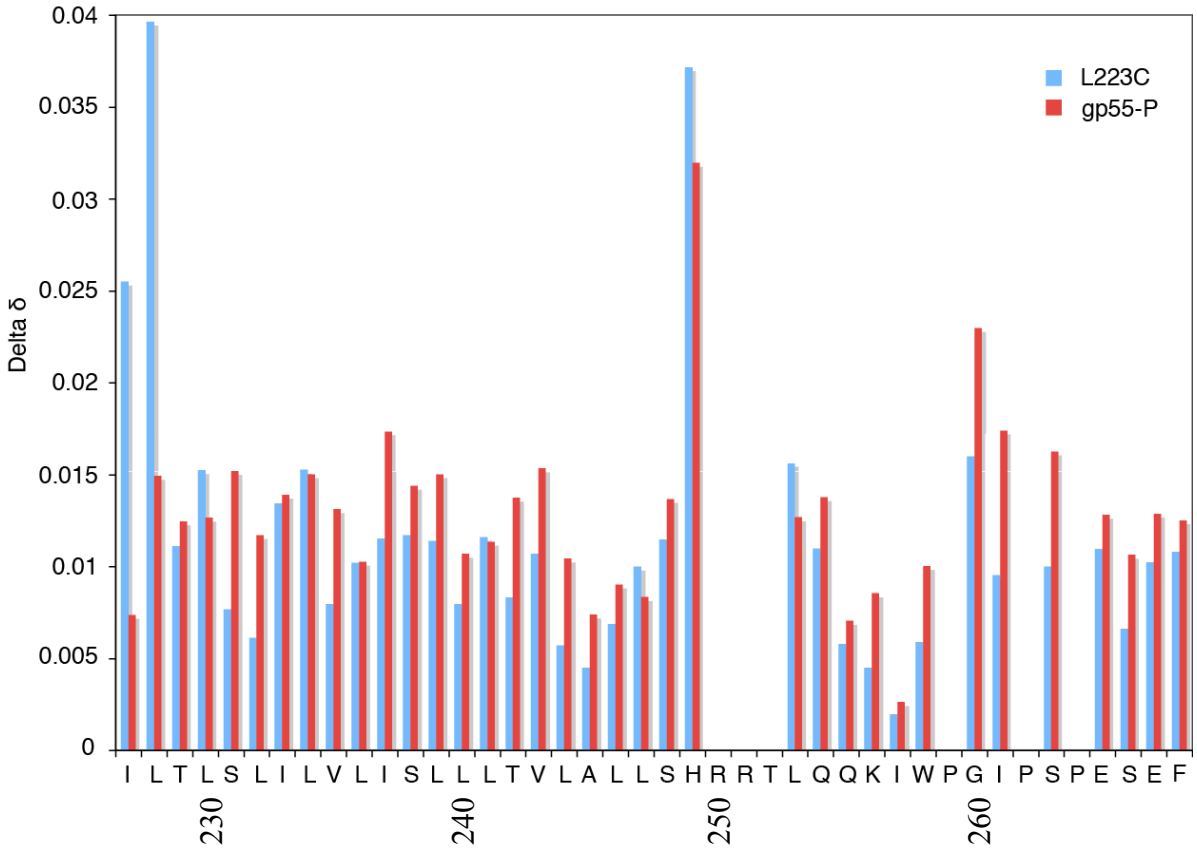



***Figure 4.6 $^{1}$H-$^{15}$N HSQC chemical shift differences, active compared to wild type.*** *Chemical shift differences from the WT spectrum are calculated for the peaks in $^{1}$H-$^{15}$N-HSQC for each active sample, using the formula in the text. Differences from WT are blue:L223C, red:gp55-P.*

**gp55-P**

Recombinant gp55-P$_{349\text{-}409}$ was expressed as a His-MBP fusion protein without isotopic labels and purified using the same method as the His-MBP EpoR TM fusion proteins (Chapter 2). Purified TM peptide was mixed 1:1 (molar ratio) with $^{15}$N labeled muEpoR$_{218\text{-}268}$ and reconstituted into DPC micelles. The resulting $^{1}$H-$^{15}$N-HSQC spectrum, when overlaid with the EpoR WT spectrum, demonstrates that many peaks are in the same position indicating that the

overall structure is the same (Figure 4.7). However, shifts of the same magnitude and direction are seen in His249, Leu253, and Gly260 (Figure 4.6) as were seen in the L223C mutant.

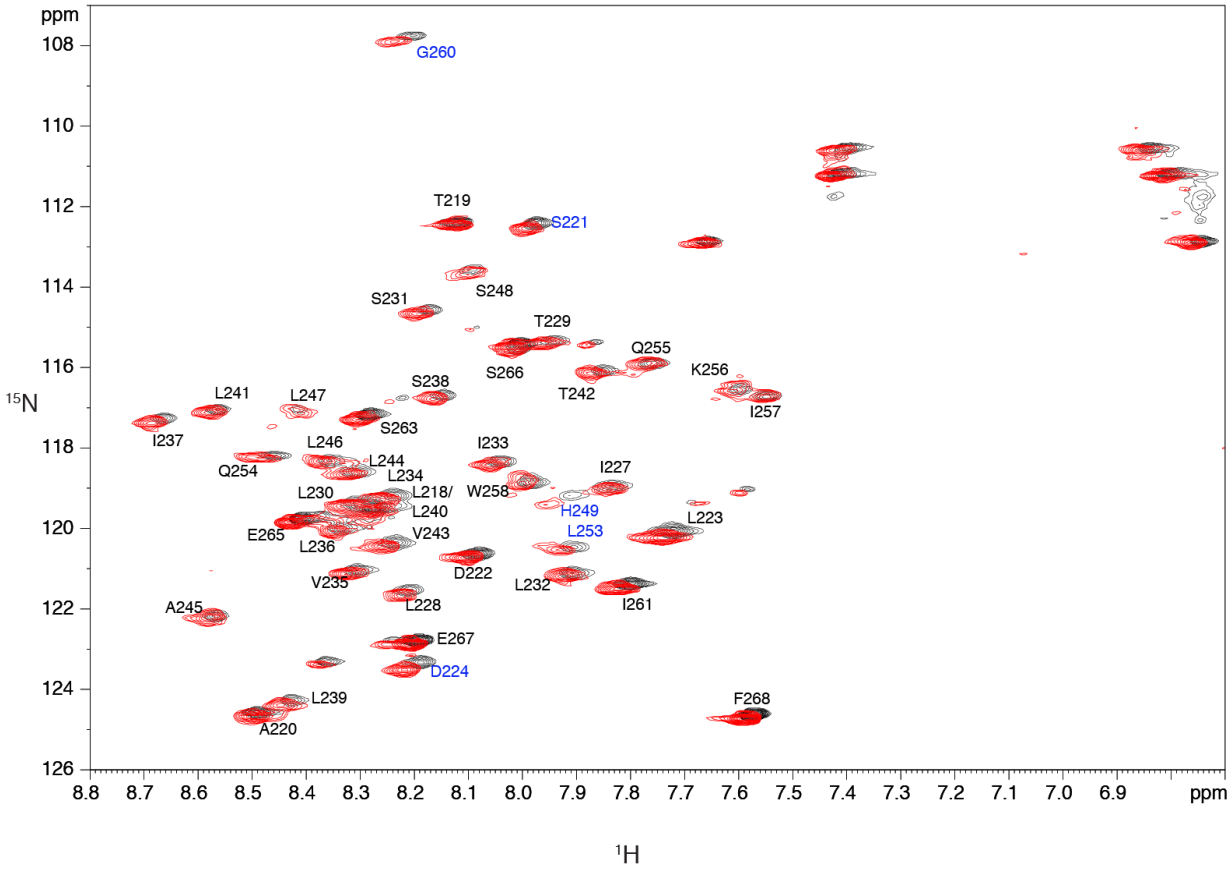



***Figure 4.7 $^{1}H$-$^{15}N$-HSQC comparison, WT compared to WT + gp55-P.*** *Wild type muEpoR$_{218-268}$ + gp55-P was prepared with an EpoR:gp55-P molar ratio of 1:1. The spectra were collected on a 700 MHz spectrometer at 313 K with 32 scans. The overlay demonstrates that while many peaks overlap, there are several that do not, including H249. black=WT, red=WT+gp55-P. Sample is ~1 mM, 10 mM sodium phosphate, pH 7.0, 200 mM d38-DPC, 10% $D_2O$ (v:v). Residues that shift significantly are indicated with blue text.*

**The Inactive State: L228C = Wild Type**

As mentioned in previous sections, the assumption is made that in this system, the WT sequence represents the inactive state of the EpoR. This assumption is necessary in order to establish a baseline for our structural studies, but must be confirmed experimentally if WT spectrum comparisons are to be made with active structures. This is accomplished using the L228C mutant. The mutant is demonstrated not to be constitutively active, but responds to Epo at the same level as wild type (*74, 75*), therefore it resembles the wild-type receptor in the absence of ligand. When this construct is prepared in the same manner as the L223C sample, peak shifts are seen in the residues surrounding the inserted cysteine as expected (Figure 4.8). However, the residues that shifted in both the L223C mutant and the WT + gp55-P spectra, notably His249 and Leu253 remain unshifted, and while Gly260 is the peak that shifts most in the spectrum, the magnitude of the shift is less than both active spectra (Figure 4.9). This result indicates that His249 and Leu253 are markers of the active state.

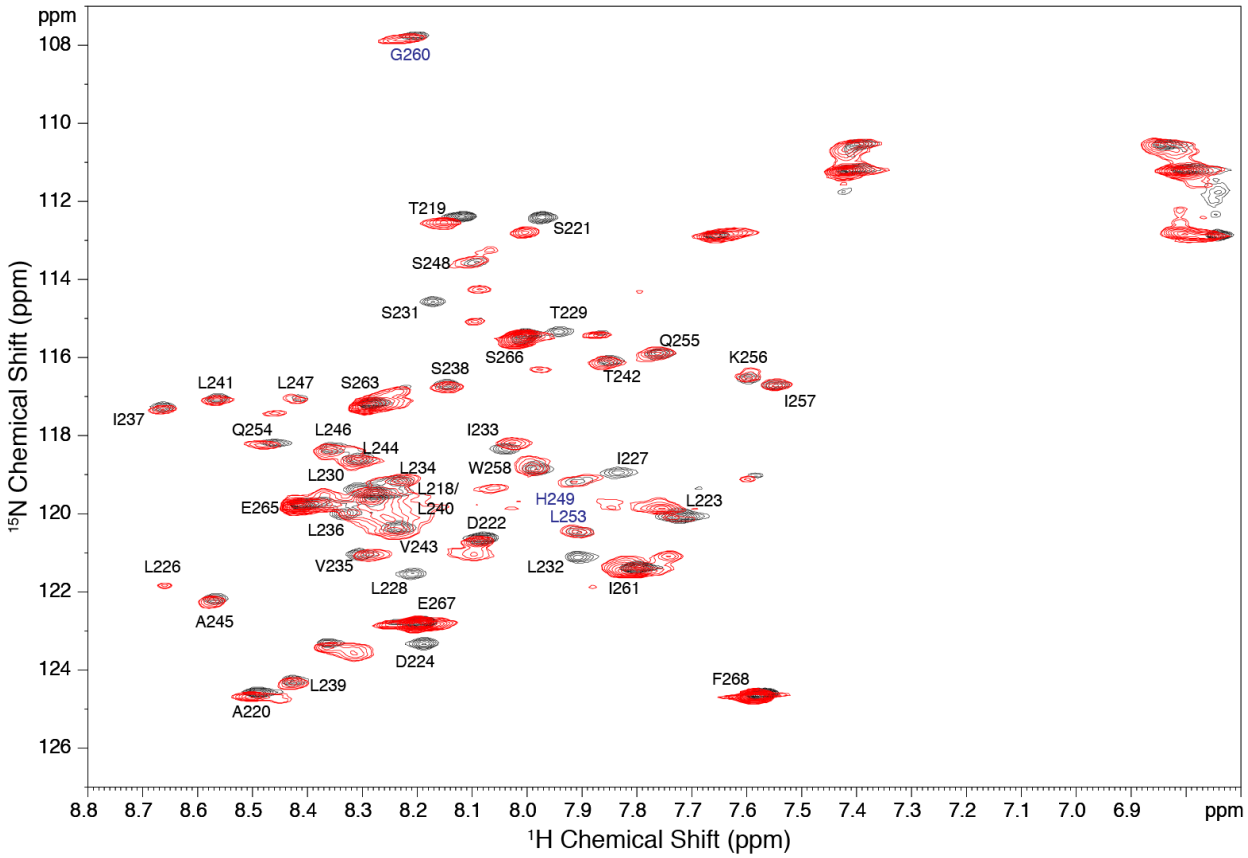



***Figure 4.8 $^{1}H$-$^{15}N$ HSQC comparison, WT compared to L228C.*** *L228C muEpoR$_{218-268}$ was prepared as described and a spectrum was collected on a 700 MHz spectrometer at 313 K with 32 scans. The overlay demonstrates that while many peaks overlap, a few do not. H249, L253, and G260 are shown in blue. Black=WT, red=L228C. Sample is ~1 mM, 10 mM sodium phosphate, pH 7.0, 200 mM d38-DPC, 10% $D_2O$ (v:v).*

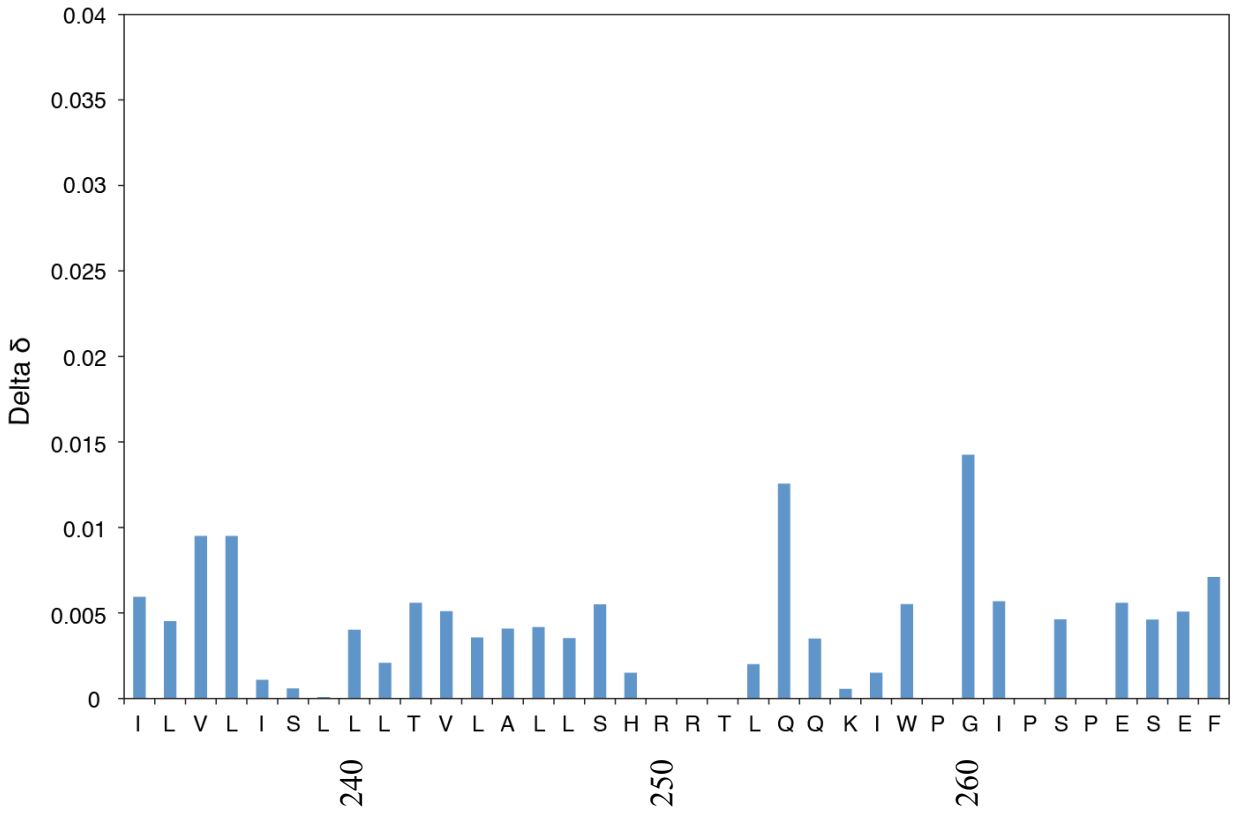



***Figure 4.9 $^{1}$H-$^{15}$N-HSQC chemical shift differences, inactive compared to wild type.*** *Chemical shift differences from the WT spectrum are calculated for the peaks in the L228C $^{1}$H-$^{15}$N HSQC, using the formula in the text.*

**Structural Changes Between Inactive and Active Constructs**

Engineered cysteine mutations are expected to cause local changes in protein structure and therefore the $^{1}$H-$^{15}$N-HSQC spectrum due to changes in the local environment of surrounding nuclei, as we see with the residues surrounding L223C and L228C. However, spectral shifts of more distant amino acids (*e.g.*, the opposite end of the helix) are interpreted as a structural change associated with activation.

***His249***

His249 is seen to shift in the same direction in the same magnitude in both the L223C and the WT+gp55-P spectra (Figure 4.10, B/C). This result, coupled with the lack of movement in the L228C spectrum (Figure 4.10, A), suggests that His249 is a marker of the active state. There are two pieces of data in the literature regarding His249. First, in an EpoR-PrlR chimera study (*73*), one of the receptor constructs used was a double mutant (L223V, H249G), presumably to insert restriction cloning sites for connection of chimeric receptor pieces. This L223V/H249G mutant appeared to be wild type like; without constitutive activity, but could respond to Epo at the same level as the WT receptor, as measured by GFP fluorescence, a measure of cell proliferation. The second piece of data also comes from Constantinescu, an alanine mutagenesis study of the intracellular juxtamembrane region (*72*). In one mutant, His249, Arg250, and Arg251 are mutated to alanine; this construct retains much of its wild-type level of activity, as measured by cell proliferation assays.

These points argue against His249 being intimately involved in the activation mechanism. However, this is only one possible extrapolation of our data to a potential mechanism. While chemical shift deviation would be expected if His249 were important for activation, shifts would also be expected if His249 were influenced by a change in local structure that depended upon transition to the active state. For example, if His249 were pulled into the membrane bilayer a change in chemical shift would be expected, as would one if a change in secondary structure occurred. In this way, the NMR data and the cell biological data are not incongruent.

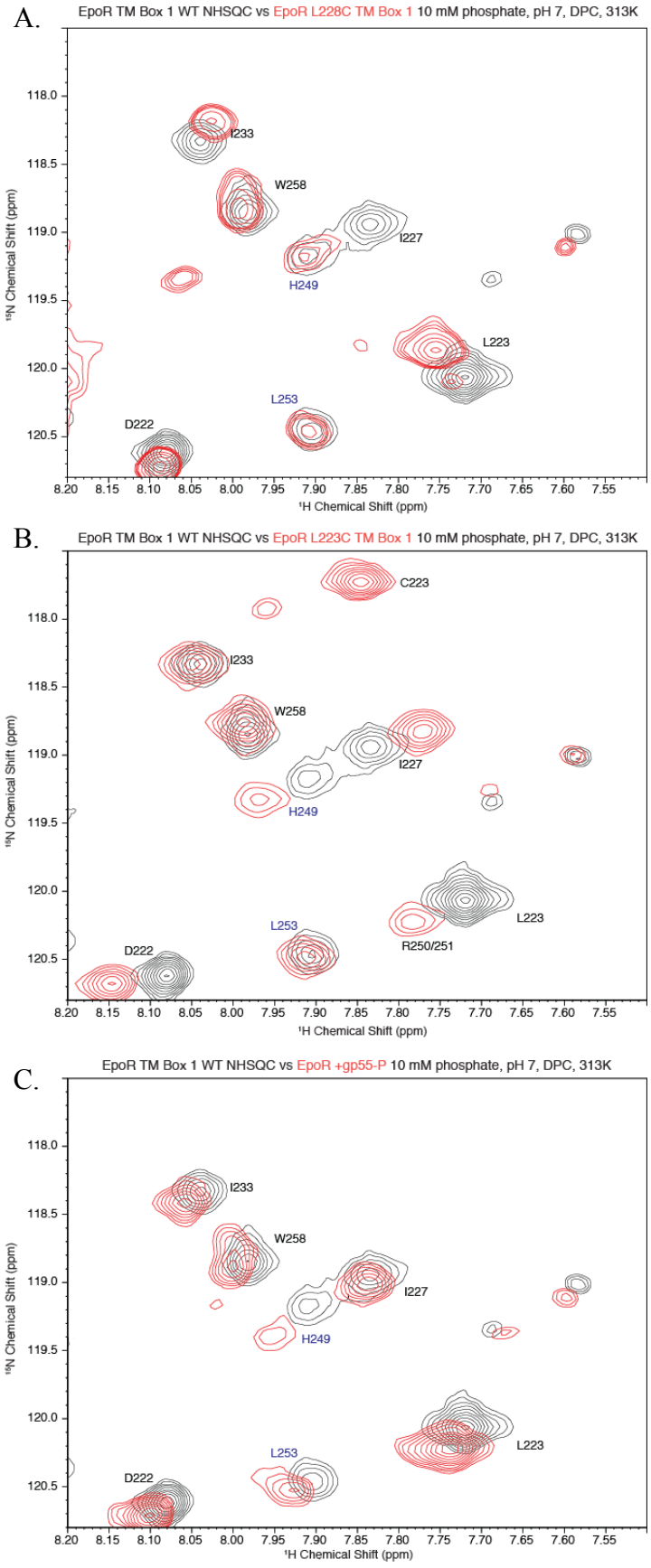



***Figure 4.10 Position of H249 is affected by active/inactive structures.*** *Overlays of the inset regions of inactive and active spectra showing the area surrounding H249 demonstrate that H249 shifts in the active L223C and EpoR+gp55-P spectra, but not the L228C wild type-like spectrum. H249 and L253 are shown in blue. A) WT (black) vs. L228C (red) B) WT (black) vs. L223C (red) C) WT (black) vs. EpoR+gp55-P (red). Sample is ~1 mM, 10 mM sodium phosphate, pH 7.0, 200 mM d38-DPC, 10% $D_2O$ (v:v).*

***Leu253***

In the active structures, Leu253 NH shows chemical shift deviation from the WT spectrum. Three conserved residues (Leu253, Ile257, Trp258) have been shown, through mutational studies, to be absolutely required for receptor activity (*72, 76*). These residues constitute a hydrophobic motif that in addition to being highly conserved across species for the EpoR (Figure 1.4), they are relatively conserved among the Group 1 cytokine receptors (Figure 1.2, and (*72*)). Despite conservation, the importance of the positioning of these residues is unclear. The position of these residues relative to the Box 1 motif is preserved across cytokine receptors, but not with respect to the end of the TMD (Figure 1.2). This would seem to indicate that whatever function they perform, they act in concert with the Box 1 motif and independently of the TMD. However, this is not necessarily true, because the function of the TMD in the rest of the cytokine receptors is unclear. If the TMD in the GHR and the PrlR drive dimerization as they do in the EpoR (*73*) and the TpoR (*15*), then two pieces of information are still missing. First, the putative dimerization interface in these receptors is unknown-the relative position of the three-residue motif could indeed be consistent with that of the EpoR. Second, the conformation of the residues linking the TMD and the L, I, W residues is unknown. Much research still needs to be performed on this interesting region before cytokine receptor activation is understood.

Chemical shift index analysis (*148*) and dihedral angle predictions for the wild type and L223C muEpoR$_{218\text{-}268}$ samples indicate that both are α-helical from Pro225-Ile257 (Figure 4.11). However, it is important to note that intensity is lost for Arg250-Arg251 in the $^{1}$H-$^{15}$N-HSQC, which would indicate flexibility or structural heterogeneity of this region. Interestingly, if the IC-JM is a rigid α-helix, then the hydrogen bonding partner for Leu253 would be the carbonyl of His249, the chemical shift of which is unclear at the moment in the wild type sample-there are

two possible chemical shift assignments, 175.8 ppm or 173.3 ppm. The L223C active carbonyl chemical shift of His249 is 176.05 ppm, which is predicted to be helical. Because the chemical shift of the wild type sample is upfield of the L223C sample, it is possible that there is a difference of secondary structure in this region between the two samples. However, because of the uncertainty of the exact chemical shift of the wild type sample, further investigation is required.

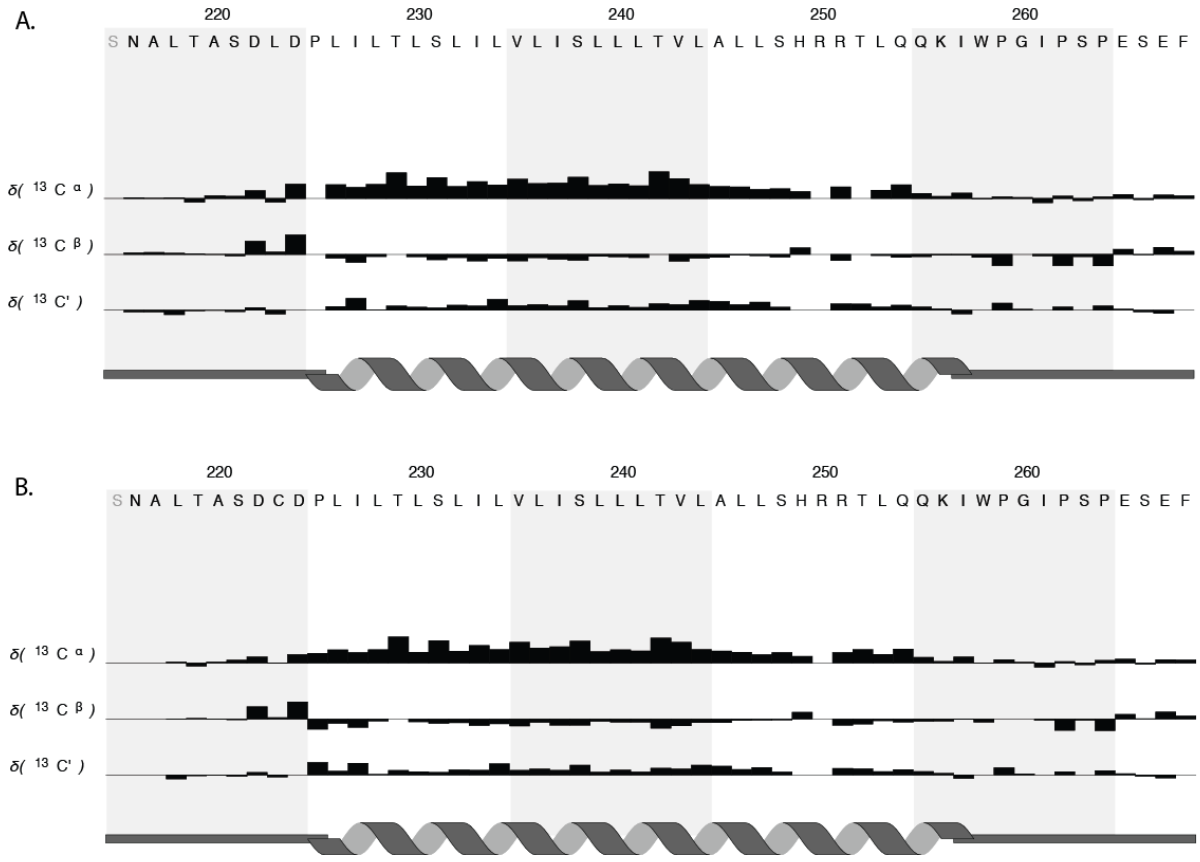



***Figure 4.11 Chemical shift index and secondary structure, muEpoR$_{218-268}$ wild type and L223C.*** *Chemical shift index is calculated on Cα, Cβ and C=O chemical shifts between the WT and L223C samples. Secondary structure prediction is a result of these calculations and dihedral angle predictions. A) Wild Type B) L223C.*

***Box 1***

Little is known about this region with the exception that it is necessary for Jak binding. This nine residue, proline-rich motif is highly conserved across species (for the EpoR) and somewhat conserved across receptor classes (Figure 1.2). With the possible exception of Gly260, this study was unable to detect any changes in the structure of the Box 1 region upon activation (L223C or gp55-P). The likely explanation is that because this region normally binds Jak2, whatever the structural conformation adopted in the TM-JM constructs in the absence of Jak2 may be irrelevant.

**Cysteine Mutants and Protein Structure**

Cysteine mutagenesis is an important tool in protein biochemistry, used to modulate a protein's state of activity, but is a tool of somewhat dubious value. Cysteine residues do not normally oxidize to form disulfides in the hydrophobic membrane environment, so structural effects of cysteine mutagenesis are unknown. While the data herein are presented without explicitly stating that the L223C construct is a disulfide-linked dimer, it is important to disclose that chemical shift analysis of the Cβ of Cys223 indicates that it is unlikely to be oxidized. $^{1}$H-$^{15}$N-HSQC studies were performed on a copper-phenanthroline crosslinked L223C and crosslinking was confirmed by SDS-PAGE. However, with the exception of the residues in the region of the copper-phenanthroline (lost due to paramagnetic effects), the non-actively crosslinked spectrum was the same as the copper-phenanthroline crosslinked spectrum (data not shown). Therefore, whatever structural changes are induced by the L223C mutation is independent of disulfide crosslinking. Indeed, in Lu *et al.'s* cysteine mutagenesis study (*75*), only a fraction of the L223C receptor exists as a crosslinked dimer. The strong effect of cysteine

without disulfide linking, while surprising, is not without explanation, as this region is part of a helix cap motif that is sequence dependent (Chapter 3).

**Helix Cap**

While we lack the datasets necessary for constructing a high resolution model of the N-terminal helix cap region for the muEpoR$_{218-268}$ WT and L223C proteins, the dihedral angles predicted from the chemical shifts of backbone atoms can be used to create a model. While the same helix cap contacts are made in both the WT and the L223C structures as in the muEpoR$_{220-248}$ structure (Chapter 3), the replacement of Leu with Cys at position 223 causes a coiling of the remainder of the N-terminus (Figure 4.12 A, B) compared with WT, which seems to be fully extended. The significance of this is at the moment unknown. However, considering the rigid coupling that is required for an ECD structural change to be transmitted through the TMD to the ICD, even small changes can have a significant impact on structure, and therefore activity. This is even more important given that the region between the WSXWS motif and the helix cap contains only 10 amino acids.

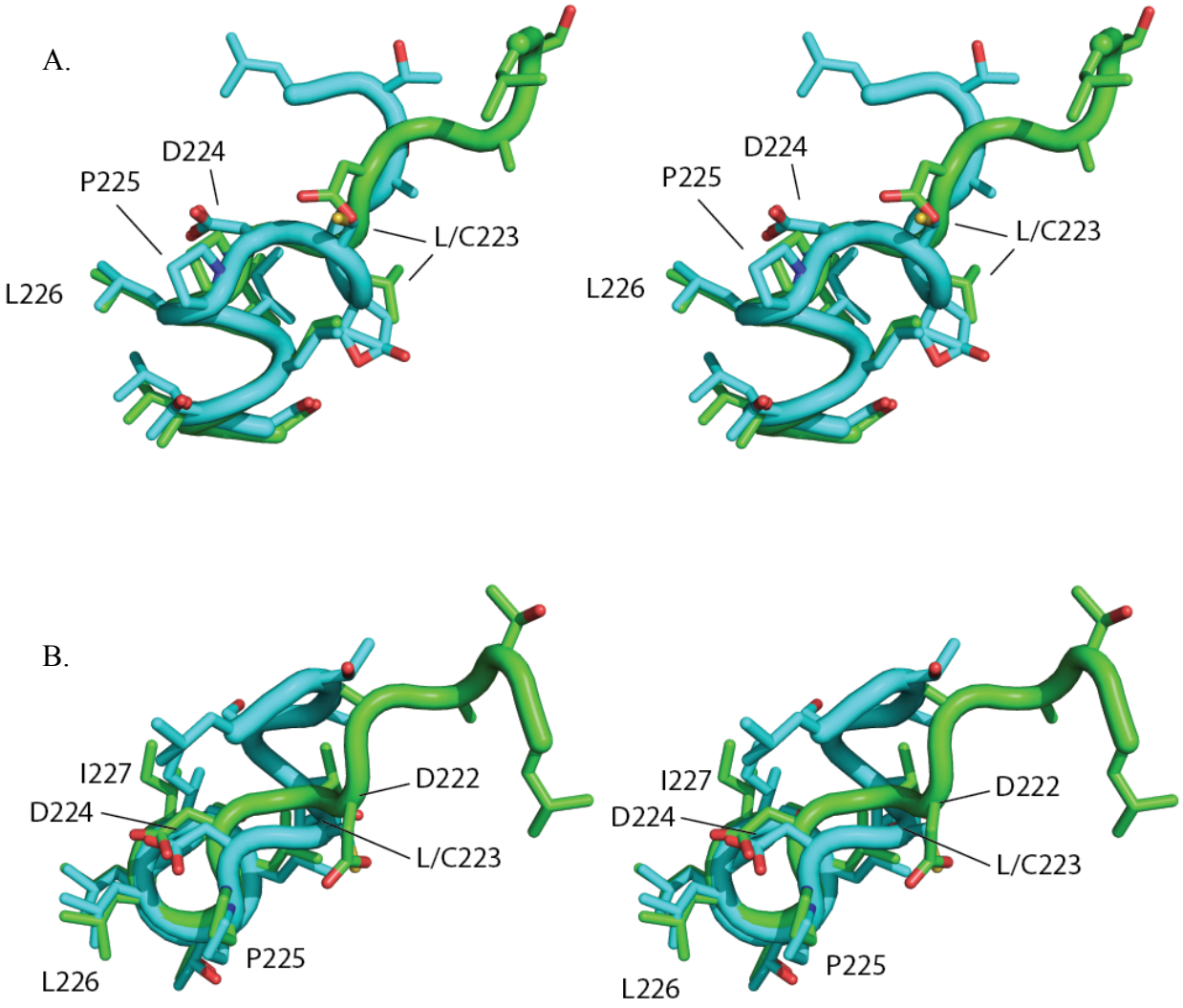



***Figure 4.12 Helix cap models, WT & L223C.*** *Helix cap models were constructed based upon dihedral angles calculated from the chemical shift values of backbone atoms. Green=WT, Blue=L223C. Cysteine at position 223 causes a significant change in structure of the N-terminus, as the A) side view (stereo), L218-S231 B) view from N-terminus (stereo), L218-S231.*

**Activation Mechanism**

The NMR measurements made on the wild type, L223C and the wild-type + gp55-P reveal several details about receptor activation. First, the fact that the spectral shifts between the inactive and the active structures are unlikely to be due to a rotation of the two helices and formation of a new interface, as proposed by Seubert *et al.* (*76*). In fact, the residues demonstrated to be lining the interface by intermonomer NMR measurements show very

minimal shifts. Second, several key juxtamembrane residues are perturbed during activation. Third, while the results of activation with L223C or gp55-P are essentially the same (His249 movement *et al.*), structural changes in the TMD between L223C and gp55-P are different. L223C induces structural changes at the N-terminus that cause or allow TMD movement to an activating position. Gp55-P causes the same activating changes in the C-terminus as L223C, but does not perturb the structure of the N-terminus. This is an indication that the interaction between gp55-P and the EpoR forcibly repositions the TMD into the active conformation. It is strange that greater shifts are not seen in the TMD residues of the EpoR upon gp55-P binding, but the average shift over the length of the TMD is greater for gp55-P than for L223C. This disparity indicates that while the L223C mutation works by changing the N-terminus structure to reposition the interfacial residues in order to move the C-terminus, gp55-P may work from the outside (helix 'outerface') to accomplish the same task.

## Chapter 5-Models for Erythropoietin Receptor Activation

Chapters 3 and 4 present the first atomic resolution studies of the EpoR TMD. An important aspect of these studies is that the TM domain is first shown to be dimeric under the conditions used and then measurements are made using a high resolution approach, namely NMR spectroscopy. Previous reports studying the TMD of the EpoR are unable to definitively say that the EpoR TM system under study is dimeric. This shortcoming leads to conclusions that overstate the data. In other studies, if a dimer could be demonstrated (*e.g.,* with AUC), then the resolution to determine an actual dimer interface was missing. What makes the studies described here all the more important is that the dimeric nature of the system under study is demonstrated, and that the system is probed with angstrom resolution.

### TM Helix Association/Interactions

Several sequence motifs for helix association have been identified. Perhaps the best studied is the GxxxG motif. This motif is the basis for TM dimerization of glycophorin A (*125*) and BNIP3 (*149*), among other TM dimers. The two glycine residues in this motif allow the close approach of the helices within the dimer, while amino acids distal but on the same face modulate the strength of the TM interaction (*150*). The leucine zipper is another dimerization motif (*62*). Present in the DNA binding transcription factor GCN4, the hydrophobic residues (Leu/Val) form the hydrophobic dimer interface (*63*), while charged or polar residues form the surface that is exposed to water. However, it is unclear if homodimeric TM proteins make use of the strict leucine zipper as well. Despite TOXCAT studies on TM proteins possessing a similar sequence motif as GCN4 indicating variable propensities for self-assembly (*61*), studies by the developer of the TOXCAT assay indicated that model leucine zipper peptides containing Val at the “a” position did not self-associate, and a polar residue was required for the peptide to do so

(*139*). Further work on the EpoR TMD defined a slightly different interface with serines at the "a" positions (*64*). Such a motif may more appropriately be characterized as a serine zipper, which has been defined in model TM peptides (*151*). However, the asparagine mutagenesis studies by Ruan *et al.* (*64*) may be misleading, as the interface results may be due to trimer or oligomer formation. Studies by Zhou *et al*. (*139*) and Choma *et al.* (*138*) have determined that Asn-containing peptides can form trimers that cause abnormally intense results in the TOXCAT assay. These strong results can mislead researchers into believing that a particular Asn mutation causes "strong" dimerization. Therefore, while TOXCAT studies can be a useful tool to define whether TM segments self-associate, caution must be exercised when interpreting results from such studies as dimers, especially when it is the only method used to assess association. The implication for EpoR signaling is that the dimerization interface defined by Ruan *et al*. (*64*) may not be the actual dimer interface. NMR measurements made on muEpoR$_{220\text{-}248}$ indicate that Ser231, Ser238 and Thr242 line the dimer interface, with Val235 and Ala245 allowing close approximation of the two helices. This interface has support in the literature from Put3-EpoR fusion studies in which only constructs with this interface are activated by gp55-P (*76*).

**EpoR TM Helix Movement During Activation**

Several general models have been proposed for the function of proteins that associate via their TMDs. The three possibilities include a rigid turning of the TMDs relative to each other (*15*), a change in the helix crossing angles relative to each other (scissoring) (*152*), or a longitudinal motion of one or both of the TMDs in the lipid bilayer (pistoning) (*153*). It is important to note that regardless of mechanism, there appears to be rigid coupling of the TMD and the ICD. Put3-EpoR TM fusion experiments demonstrate that receptor activity has a dependence on the rotational orientation of the TMD (*76*). Insertion of alanine residues between

the TMD and the intracellular "switch" region of the EpoR also modulates receptor activity (*72*), with 1-2 inserted alanines reducing activity and 3 inserted alanines (~1 helix turn) recovering wild type activity. This result, coupled with the Put3-EpoR fusion data, suggested that the helix continues after His249 and that the position of the ICDs, and by extension receptor activity, depends upon TMD rotational orientation. This hypothesis fit nicely with the ECD crystal structures that demonstrated rotation of the two ECD halves upon ligand binding (*8*).

The TM rotation model for activation also has support from a related receptor family, the ErbB receptors. For example, cysteine mutagenesis studies have demonstrated that EGF receptor kinase activity is coupled to the rotational orientation of the TMD (*154*). The kinase domain of the Neu receptor, a related RTK, was also shown to have activity that is coupled to the rotational orientation of the TM region, albeit with a model TM sequence instead of the actual Neu receptor sequence (*155*). The receptor for atrial natriuretic peptide (NPRA) also seems to have a similar rotational mechanism of activation (*156*).

The bacterial aspartate chemoreceptor TAR is a good example of the piston model of signaling. This TM receptor is a homodimer, but the functional unit is thought to be a trimer of dimers. Tryptophan residues flank the TMD sequence, but activity seems to be controlled by the Trp at the C-terminus of the TMD, which is followed by a tyrosine. When the Trp-Tyr motif is moved closer to the N-terminus of the helix, the kinase is inactive (*153*). The opposite occurs when the Trp-Tyr motif is moved further C-terminal. The transitions along the bilayer normal are small, ~1.5 Å, which is accompanied by a small change in helix tilt (~5°) (*153*).

While the NMR data presented here for the muEpoR$_{218\text{-}268}$ do not necessarily disagree with a rigid coupling of the TMD and the ICD, it is incompatible with a TM rotation model. The

NMR spectra of the muEpoR$_{220-248}$ and muEpoR$_{218-268}$ are those of a symmetric molecule where the TMDs do not change conformation greatly when switching from the inactive (WT) to active (L223C, gp55-P) conformations. A change in the TM dimer interface would be expected to cause shifts in the residues that were part of the old (inactive) interface, as well as shifts in residues in the new interface. Therefore, the NMR measurements presented here appear to be more consistent with a helix tilting model or a piston model.

**Implications for Other Group 1 Cytokine Receptors**

It was hoped that the specific interaction of the TMDs of EpoR would provide a model for the function of other cytokine receptors, especially the members of the Group 1 receptors (GHR, PrlR, TpoR). However, it seems that even between members of this subgroup there may be divergent function. The GHR has the fewest polar residues in the putative TMD of all Group 1 cytokine receptors (Figure 1.2), and no other immediately discernable dimerization motif. TOXCAT assays performed as controls for self-assembly of the EpoR TM domain indicate that the GHR TMD has no propensity to self assemble (*19*). This correlates well with the receptor monomer association studies performed on chimeric GH receptors harboring the LDLR TMD (*13*). The results of these studies indicate that the EC domain, not the TMD, mediates receptor self-assembly. Furthermore, comparisons between the ECD crystal structures of the EpoR and GHR reveal that the GHR has more inter-D2 domain contacts than does the EpoR (*45*). Together, these observations indicate that although the inactive and active forms of the cytokine receptor are dimeric, the mechanism of dimerization may differ between receptors.

The TpoR is divergent as well. It possesses a duplicated ECD, not uncommon in the cytokine receptor family. However, when this extra ECD is deleted, it causes constitutive receptor activation (*157*). There is also an additional five-residue insert (R/KWQFP) that

regulates activity of the TpoR (*158*). Deletion of this insert causes constitutive receptor activation. There are also several clinically relevant mutations of the TMD (S505N) or surrounding regions (T487A, W515K/L) that cause receptor hyperactivity, while for the EpoR none are known. Given these structural and functional differences between the Epo and the other Group 1 receptors, while they seem to generally function similarly, it seems likely that each receptor has evolved receptor-specific mechanisms for precise control of activation and cell signaling. These are likely related to pathway importance, cellular receptor density, and concentration and affinity for ligand.

**Implications for the Activation Mechanism of EpoR**

While there is evidence for a rigid coupling of the TM-ICD functions, there is almost no data on the relationship between the ECD and the TMD and how information is transmitted between the two. The crystal structures of the ECD with and without ligand include several of the residues between these two regions, but the structure is undefined. The only constitutive mutant is the R129C mutation, which crosslinks the two D2 domains of the ECD (*32*). In the crystal structure of the Epo-bound ECD (PDB ID:1CN4) these residues are more than 30 Å apart. Positioning the D2 domains of receptor monomers so the R129C residues are within disulfide bonding distance would also reposition the EC-JM region to be in close proximity. If the constitutive activity of the L223C mutant depends upon disulfide linkage, then it would seem to accomplish the same thing. A recently described activating mutation of the TpoR (T487A) is located in this same region of the receptor (*34*). Recent studies on the Prl receptor where one to four alanine residues were inserted between the ECD and the TMD indicated that ligand-induced activity did not suffer (*17*). It is clear that ligand binding to the ECD of cytokine receptors causes structural changes that are propagated through the TMD to the ICD, but how remains a mystery.

**Box 1, Jak2 and Prolines**

Chemical shift analysis and dynamic NMR studies indicate that the rigid helical nature of the EpoR TM and IC-JM region is encoded by the sequence of the receptor (up to Ile257), but C-terminal to that the structure is unknown. The conformations of the WT and the L223C Box 1 regions seem to be generally very similar as judged by the $^{1}H$-$^{15}N$-HSQC (Figure 4.5). These studies lack the ability to determine the conformation of the Box 1 region that binds Jak2. It may be that the conformation of the Box 1 region in the absence of Jak2 is irrelevant, as there is evidence that the EpoR (*46*) and TpoR (*47*) require Jak2 to be properly processed through the Golgi and transported to the membrane.

It is interesting to speculate about the role of the proline-rich Box 1 region and activation of the cytokine receptors. The proline residues in Box 1 throughout the cytokine receptor family are highly conserved, but their function is unknown. Given proline's ability to be in the *cis* or *trans* conformation, it seems like a natural point of pathway regulation. Indeed, others have investigated peptidyl-prolyl isomerase (PPIase) activity in regulating other signaling pathways (*159*), and the processing of amyloid precursor protein (*160*). Pin1 is a PPIase that recognizes Ser/Thr-Pro motifs (*161*). Given the presence of the Ser263-Pro264 sequence, this is a potential target for regulation. A biophysical study of a peptide homologous to the sequence of the PrlR Box 1 region (IFPPVPGP) indicates that it is very flexible, but has a definite preference for *cis*- or *trans*-conformations in water (*trans*-*trans*-*cis*-*trans*) (*162*). The first, third and fourth Pro are conserved between the EpoR and PrlR (Figure 1.2). The chemical shift of the Cβ resonance is sensitive to *cis*/*trans* isomerization, the chemical shifts of Cβ for Pro259, Pro262, and Pro264 for both wild type and L223C EpoR are all more consistent with a *cis* conformation. This is unusual, as the *cis* conformation is expected to be found only ~10% of the time (*163*). The significance of

this is unclear, but given that our peptides were expressed in bacteria, it is possible that processing is different from mammalian expression systems. It is also possible that the Cβ chemical shifts observed are unrelated to the isomer present and are otherwise affected by the environment of the peptide.

**Mechanism of Activation**

To postulate a mechanism for receptor activation, I begin with two premises: 1) that the EpoR dimers are pre-formed, and 2) that the TMD mediates receptor dimerization. The structure of the unliganded ECD dimer (*9*) leaves the C-terminal residues too far apart (~78 Å) to connect to a TM dimer, the TMDs would have to be monomeric. Because these two data points are incompatible, and several lines of evidence suggest that the TMD mediates dimerization of the EpoR in cells (*73*), the unliganded ECD structure is unlikely to be correct. Interestingly, the structure of the TM dimer presented here and the ligand-bound ECD structure (PDB ID:1CN4) are compatible, the C-terminus of the ECD and the N-terminus of muEpoR$_{220\text{-}248}$ are separated by about the same distance and can be linked without difficulty (Figure 5.1). This exercise does not immediately suggest any clues as to how Epo binding changes in the ECD are propagated through the EC-JM domain to the TMD.

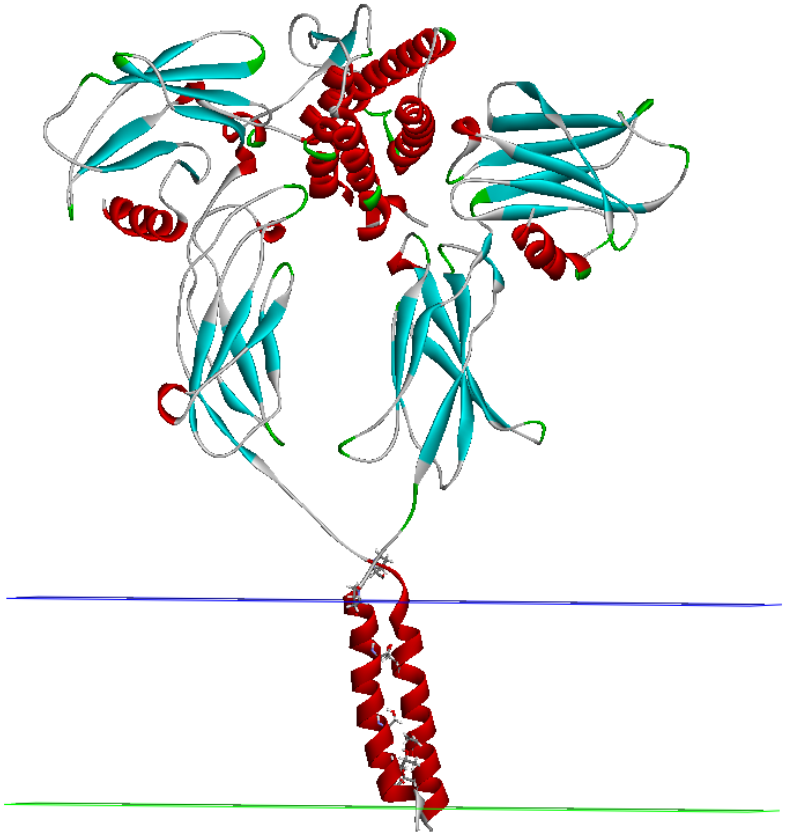


***Figure 5.1 Human EpoR ECD linked to mouse TMD.*** *The human EpoR ECD dimer with bound-ligand (Pro7-Thr220, PDB ID:1CN4) is joined to the mouse EpoR TM dimer (Ala221-Ser248). Pro225, Ser231, Ser238, and Thr242 sidechains are shown for each monomer. Membrane boundaries are shown as horizontal lines.*

Studies on the region that connects the TMD and the Box 1 motif indicate that Ser248-Ile257 is a continuous helix (*72*). The NMR measurements on the wild type and the L223C sequence (Chapter 4) suggest that this may be the case in the muEpoR$_{218\text{-}268}$ structure. Some controversy exists with regard to how activation is transmitted through the TM dimer of cytokine receptors. Studies on the EpoR indicate that receptor activity may be linked to the rotational orientation of the TMD (*72*). However, similar studies on the PrlR indicate that its activation is not consistent with rotation or a piston motion of the TMDs (*17*). The NMR measurements

presented here on the muEpoR$_{218-268}$ peptide are inconsistent with a rotation model. The $^{1}$H-$^{15}$N HSQC spectra of the L223C and the wild type EpoR + gp55-P show considerable overlap with the wild type spectrum, an indication that the overall structure does not differ greatly between the wild type and the active conformations. If the L223C and the wild type EpoR + gp55-P spectra are truly the result of studying the active conformations, then the only models that would be consistent with the structural data are a helix tilting model or a piston model where both helices move in or out of the membrane simultaneously. With respect to the helix tilting model, two modes of tilting are possible, one where the helix crossing angle changes (Figure 5.1, A), or one where the interhelical distance becomes unequal (*i.e.*, greater at the bottom than the top, Figure 5.2, B). For these two models, small changes in relative helix tilt should result in a significant change in intracellular distance. The residues from His249-Ile257 would constitute an additional length of ~13.5 Å, if it is a continuous alpha helix. If the crossing angle were centered in the middle of the TM segment (~17 Å), then the horizontal displacement of the two helix C-termini would be 5.4 Å, assuming a 5° tilt with the bilayer normal (sin 5° = X/30.5 Å). This is additive to the interhelical separation that already exists. The other possibility is that the entire peptide dimer piston model where the entire complex moves in or out of the membrane, parallel to the bilayer normal (Figure 5.2, C). How this would activate the intracellular kinases is not clear, unless the act of physically changing the distance of the kinase from the bilayer causes an activating (or inactivating) conformational change. Because distance measurements have not been made for the larger muEpoR$_{218-268}$ peptide, the precise interface and thus the activation mechanism remains unknown.

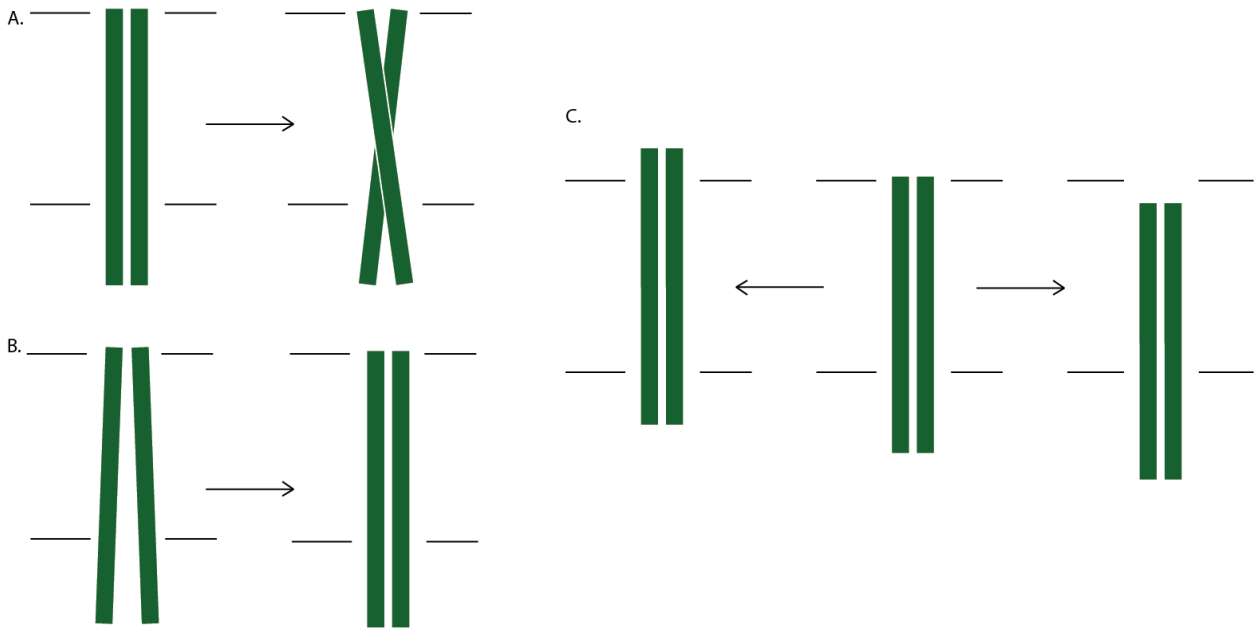



***Figure 5.2 Possible activation mechanisms.*** *Potential activation mechanisms consistent with the NMR measurements presented here. A) Helix tilting repositions the C-terminal ends of the helices with respect to each other. B) Second helix tilt mechanism allows relative repositioning of the C-termini in a different manner. C) A piston model where the interface remains the same in the inactive and the active structures.*

For the intracellular side, changes that occur upon whatever conformational change the TMD undergoes are also unknown. This is even more of a mystery because no structure of the ICD exists. It is known that the FERM domain of Jak2 binds to the proline-rich Box 1 sequence, but the relative positions of the FERM and the kinase from the two receptor chains are unknown. The NMR measurements made here of the Box 1 region are consistent with a random coil geometry. Therefore, it is not clear if the Box 1 region has a defined structure in the absence of Jak2. However, the structure of the ICD may not be as important as the state of Jak2 residence on the receptor ICD.

Funakoshi-Tago *et al*. (*164*) have demonstrated with mutants of Jak2 that phosphorylation of Tyr119 after receptor activation triggers Jak2 dissociation from Box 1. This

fits with another observation about the EpoR that suggests a dissociation of Jak2 from Box 1, the ubiquitination of Lys256 (*48*) in response to Epo-induced activation and subsequent degradation of the EpoR (*49*). Lys256 is a residue in the switch region, only one amino acid away from the beginning of the Box 1 sequence, where Jak2 binds. If Jak2 remained bound to Box 1 after Epo-induced activation, then it seems unlikely that a rapid ubiquitination event could occur, as this would require space for the binding of ubiquitin and the ubiquitination machinery. More recent experiments on the GHR demonstrate a similar detachment of Jak2 upon receptor activation (*165*). Therefore, the activation mechanism of cytokine receptors upon ligand binding may involve dissociation of the bound Jak.

In light of the structural information regarding the EpoR TMD and ICD ascertained by these NMR studies, it is possible to begin connecting the various structural pieces of the receptor (Figure 5.3). I have mentioned above that it is possible to connect the ligand-bound ECD crystal structure (PDB ID:1CN4) with the EpoR TM dimer structure determined here by solution NMR (Chapter 3). Chapter 4 describes the extension of structural studies past the TMD dimer to include the Box 1 region that interacts with the Jak2 FERM domain. Using the dihedral angles predicted from chemical shifts of backbone atoms in the ICD region, I extended the TM dimer structure calculated in Chapter 3 to include ICD residues His249-Phe268. The conformation of the Box 1 residues is governed by the *cis* conformations of the proline residues in this region.

No crystal structure exists for the Jak2 FERM domain, but through homology modeling and computational studies, a model of the complete Jak2 structure has been proposed (*166*). The FERM domain structure for the focal adhesion kinase (FAK) has been solved (*167*), it shows the tri-lobed structure seen in other FERM domains (PDB ID:2AL6). One feature of this structure is that a region of the FAK C-terminal to the FERM domain that corresponds to the SH3 binding

site (RxxPxxP) binds in a groove between the F1 and F3 lobes of the FERM domain. This peptide sequence has suspicious homology to the sequence immediately preceding and including the Box 1 region of the EpoR (KIWPGIP). Hypothesizing that the Box 1 region of muEpoR$_{218-268}$ also binds between the F1-F3 lobes of the Jak2 FERM domain, the Jak2 structure is positioned such that the peptide aligns in this groove.

This exercise does not immediately suggest any keys for unlocking the mystery of receptor activation. Rather, it hints that with the proper perspective, studying the receptor piecewise and reassembling the pieces later may allow construction of a rational model of the receptor. For instance, as mentioned in earlier sections, while the unliganded EpoR ECD is incompatible with a TM dimer, the liganded ECD D2 domains connect easily with a TM dimer (Figure 5.1). Combining the knowledge of Box 1's role in Jak2 binding with the structural knowledge of the FERM domain of other kinases provides clues about how to connect the two proteins. Interestingly, when the Jak2 molecules are positioned on the EpoR TM dimer such that the F1-F3 lobes of the FERM domain interact with the Box 1 region, there is no steric hindrance between the Jak2 molecules on opposite monomers. It seems that with Jak2 positioned as such, the FERM domain would just clear the membrane. The important Y1007/Y1008 residues near the active site are positioned at the bottom of the Jak2 molecule, near the interface of the two kinase molecules.

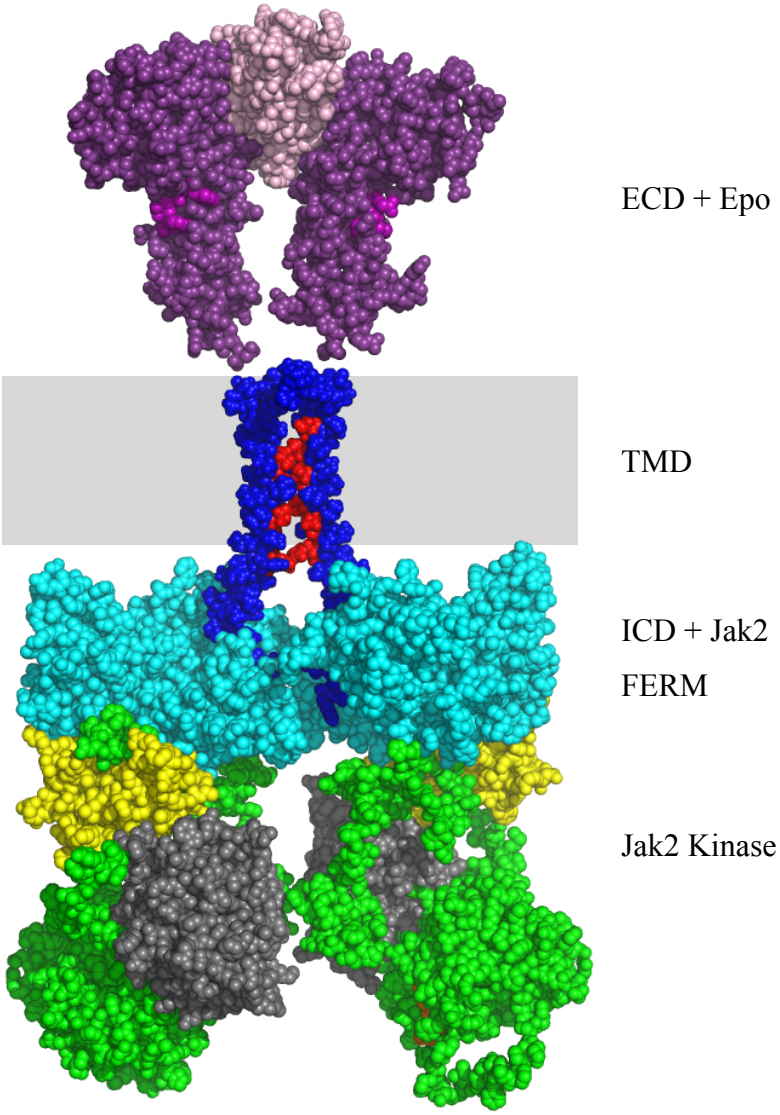



***Figure 5.3 EpoR-Jak2 structure composite.*** *The muEpoR$_{218-268}$ structure based upon dihedral angles is colored blue, with the interfacial residues of the dimer (S231, V235, S238, T242, A245, and H249) are colored red. The Epo-bound EpoR ECD dimer (PDB ID:1CN4) is colored purple, the WSXWS motif in magenta. Epo is light pink. The different domains of Jak2 (Jak2.pdb, (166)) are color-coded; the FERM domain is cyan, the SH2 domain is yellow, the pseudokinase domain is grey, the kinase domain is green. Y1007/Y1008 are colored orange.*

While the model presented above provides a compelling look into the possible structure of the full-length EpoR with Jak2, it unfortunately does not give any indication of how small changes in the EpoR TM structure upon activation would induce Jak2 activity. Clearly, these studies are merely the beginning of the road with respect to determination of the full receptor

structure and ultimately, mechanism of activation. What is clear is that the TMD provides a connection between the ECD and Jak2, and the activation signal is transduced through this region. Building out from the TMD and successively studying larger and therefore more biologically relevant receptor constructs, in an manner similar to cryo-electron microscopy studies of the IL-6/IL-6R/gp130 complex (*168*), and the IL-6/IL-6R/gp130/Jak1 complex (*169*), it should be possible to determine structures of EpoR in the inactive and active state using solution NMR. The EpoR peptide/micelle system, as defined here, will allow the addition of the FERM domain of Jak2 to the muEpoR$_{218\text{-}268}$ sample in order to ascertain changes in Box 1 structure, as well as specific interactions between the two proteins. A range of experiment possibilities is explored further in Chapter 7.

## Chapter 6-The S505N Mutation in the Thrombopoietin Receptor Drives Receptor Dimerization

The thrombopoietin receptor (TpoR), along with its cognate ligand, thrombopoietin (Tpo) governs megakaryocyte development and ultimately platelet production (*170, 171*). The sequence of the receptor is related to the rest of the Group 1 cytokine receptors (EpoR, PrlR, GHR), with two notable exceptions. First, the TpoR contains a duplicated ECD, resulting in an ECD that contains two D1 domains and two D2 domains. Deletion of the combined membrane distal D1-D2 domain results in constitutive receptor activity (*157*). Second, the TpoR contains a five-residue insert directly C-terminal to the TMD, RWQFP (KWQFP in mice), that has been shown to modulate receptor activity (*158*).

Much less is known about the structure of the TpoR than other Group 1 cytokine receptors. A crystal structure of the ECD has not been solved. However, several point mutations of the TpoR that cause disease in humans have been described; these have been helpful in correlation of receptor structure with function. The S505N (*172*), W515K (*35*) and T487A (*34*) mutations cause a gain of function of the TpoR, resulting in disease from thrombocythemia to myelofibrosis. The positioning of these mutations is interesting. The S505N mutation resides in the middle of the TMD, while the W515K and T487A mutations are in the intracellular and extracellular juxtamembrane regions. The clustering of these mutations in and around the TMD seems to underscore the importance of this region in receptor function. What is not completely clear is how these mutations perturb receptor structure in order to cause activation.

### TpoR Dimerization

Because the other Group 1 cytokine receptors have been shown to function in the active state as dimers (*45*), it is reasonable to assume that the TpoR does as well. However, while

experiments to determine preassociation of receptor chains in the absence of ligand has been shown for the EpoR (*73*), GHR (*52*), and PrlR (*16*), the same experiments have not been conducted for the TpoR. Association studies on the TMD sequence of the human TpoR using cysteine mutagenesis and the TOXCAT expression reporter assay indicate that the wild type TMD sequence of the huTpoR$_{491-515}$ associates as a dimer (*15*). In this model, the interfacial residues are G503, A506, L510, and L513. Interestingly, W515, when mutated to cysteine is also able to crosslink receptors, despite being on the helix 'outerface' with respect to the heptad interface defined above. There are two possible explanations for this result. The first is that the W515C induces oligomers composed of more than just two monomers. The second is less obvious. Alignments of the TMD-ICD of the EpoR and TpoR (Figure 1.4) indicate that the R/KWQFP insert has no overlap with the rest of the EpoR sequence and is a true insert of five residues (*158*). This would indicate that the five-residue insert may function as a unit, and the lack of the QFP residues affects the structure of the TMD, and therefore, the results of the Matthews *et al*. (*15*) series of experiments. For instance, typical α-helical backbone hydrogen bonding proceeds in an i C=O to i+4 NH pattern. If we stipulate that the secondary structure is helical through Pro518, then the NH groups of the missing Q516, F517 residues would be unable to hydrogen bond with L512 and L513 carbonyls, disrupting the secondary structure of this important region. This is especially relevant because activity studies on the full-length receptor with the RWQFP insert deleted is constitutively active (*158*).

In order to determine the contribution of the TMD and the insert region to the function of the TpoR, we decided probe the structural characteristics of the entire human TpoR TMD including the insert region using solution NMR. The construct sequence chosen (huTpoR$_{481-520}$) stretches from the end of the β-sheet rich D2 domain to two residues after Pro518, so it should be

possible to assess the native structure and function of the complex. Because there are several clinically important disease-causing mutations in and around the TMD of the TpoR (T487A, S505N, W515K), peptides corresponding to these mutations can be studied alongside the WT peptide in order to determine the biological relevance of the results. We find that the WT peptide contains a kink in the TM helix such that there are two separate helices, an observation that corresponds to previous NMR studies of this peptide (*147*). We also find that the S505N mutation increases the oligomerization state of the $TpoR_{481-520}$ peptide. Furthermore, there is also a change in secondary structure of the peptide associated with this active state, a straightening of the helix through the kinked region observed in the WT peptide. Together, these results indicate that the S505N mutation functions by changing the oligomerization state of the TpoR TMD. Furthermore, the change in secondary structure may mimic the change in TMD structure induced by the binding of the natural ligand, thrombopoietin.

**Structural Data on the TpoR WT and S505N Peptides**

The $huTpoR_{481-520}$ WT and S505N peptides were expressed as His-MBP fusion proteins in *E. coli* BL21(DE3) cells and purified as previously described (*173*). Reconstitution of the peptides into DPC micelles as was performed for the EpoR TM peptides revealed that the $huTpoR_{481-520}$ S505N peptide yielded poor quality spectra (data not shown). Repeated purification and different buffer conditions (changing pH, buffer and salt concentrations) did not improve the spectrum. However, reconstitution of each peptide in *d25*-sodium dodecyl sulfate (SDS) micelles yielded good quality spectra (Figure 6.1) that were used for 3D backbone studies. Interestingly, despite the sequence differing by a single amino acid, the two spectra were very different, an indication that the two conformations are very different.

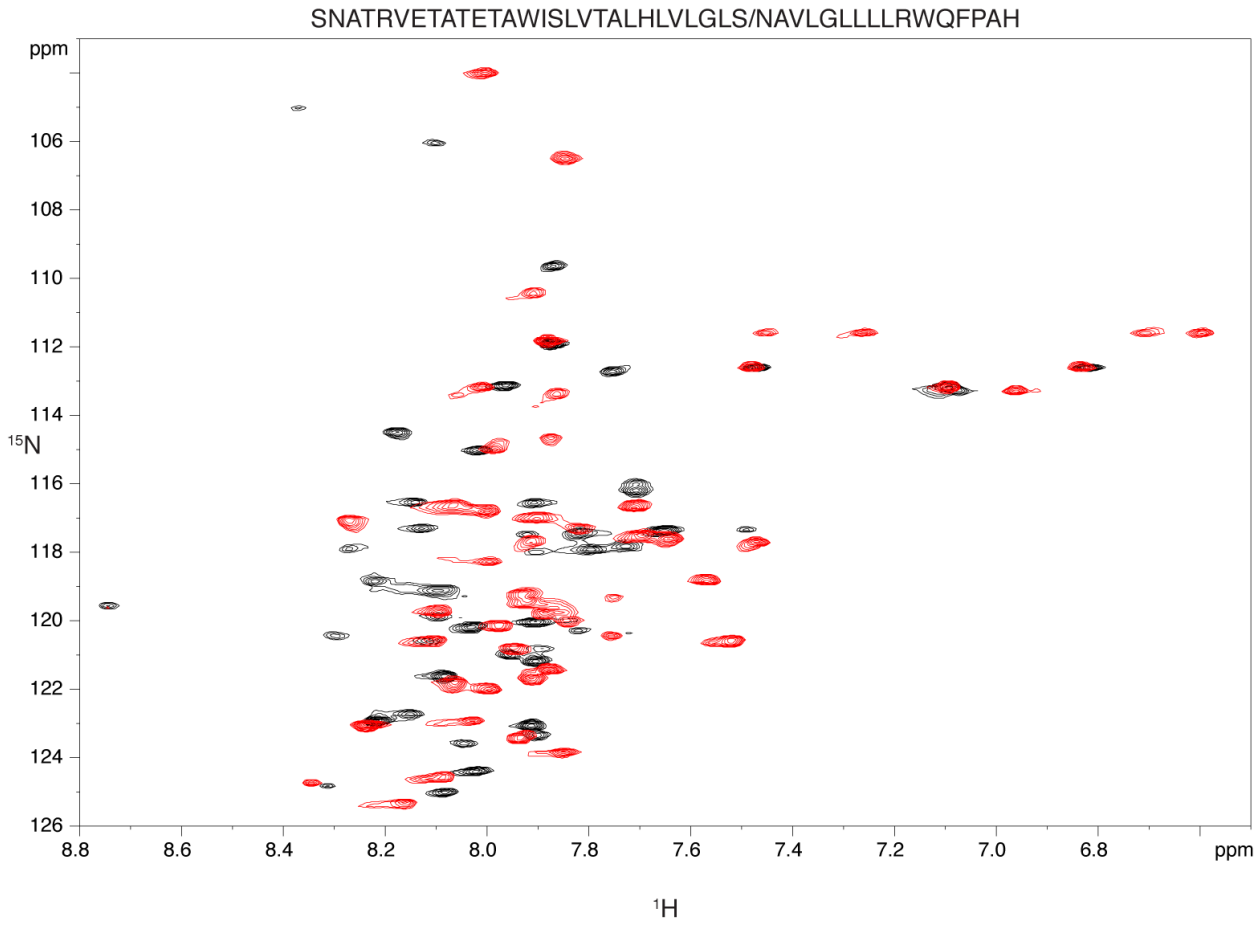
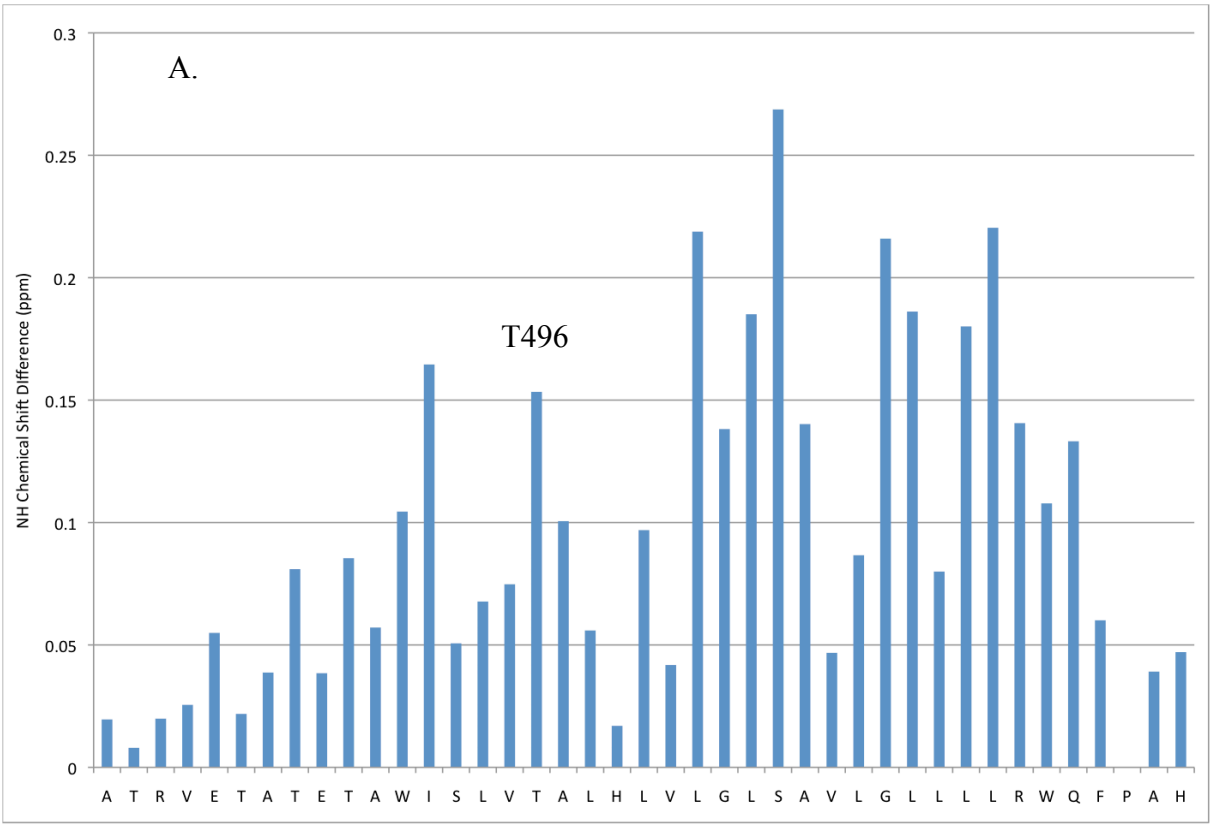



***Figure 6.1 Comparison between $^{1}H$-$^{15}N$ HSQC from huTpoR$_{481-520}$ WT and S505N demonstrate large chemical shift changes.*** *huTpoR$_{481-520}$ WT (black) and S505N (red) reconstituted into SDS detergent micelles yielded high quality HSQC spectra. Samples were 1 mM peptide in 10 mM sodium phosphate, pH 4.7, 85x CMC d25-SDS, 10% $D_2O$ (v:v). Spectra were collected at on a 700 MHz spectrometer at 320 K, 32 scans each.*

Backbone chemical shift assignments for the WT and the S505N peptides (Figure 6.2 A, B) reveal secondary structure changes from WT to S505N peptides reconstituted into SDS micelles. Carbonyl chemical shift changes in the N-terminus of the TM helix occur between WT and S505N, particularly at residue I492. Changes in α-helical geometry for the TM helix would be expected to result in paired shift of the carbonyl and its i+4 hydrogen-bonding partner's amide proton. This is seen for the I492 carbonyl-T496 amide pair, and the downfield shift of the I492

carbonyl is indicative of greater α-helical geometry for this region of the S505N peptide. The conclusion from these observations is that there seems to be an increase in α-helical character of the TMD N-terminus in the S505N peptide in the region surrounding I492-T496.



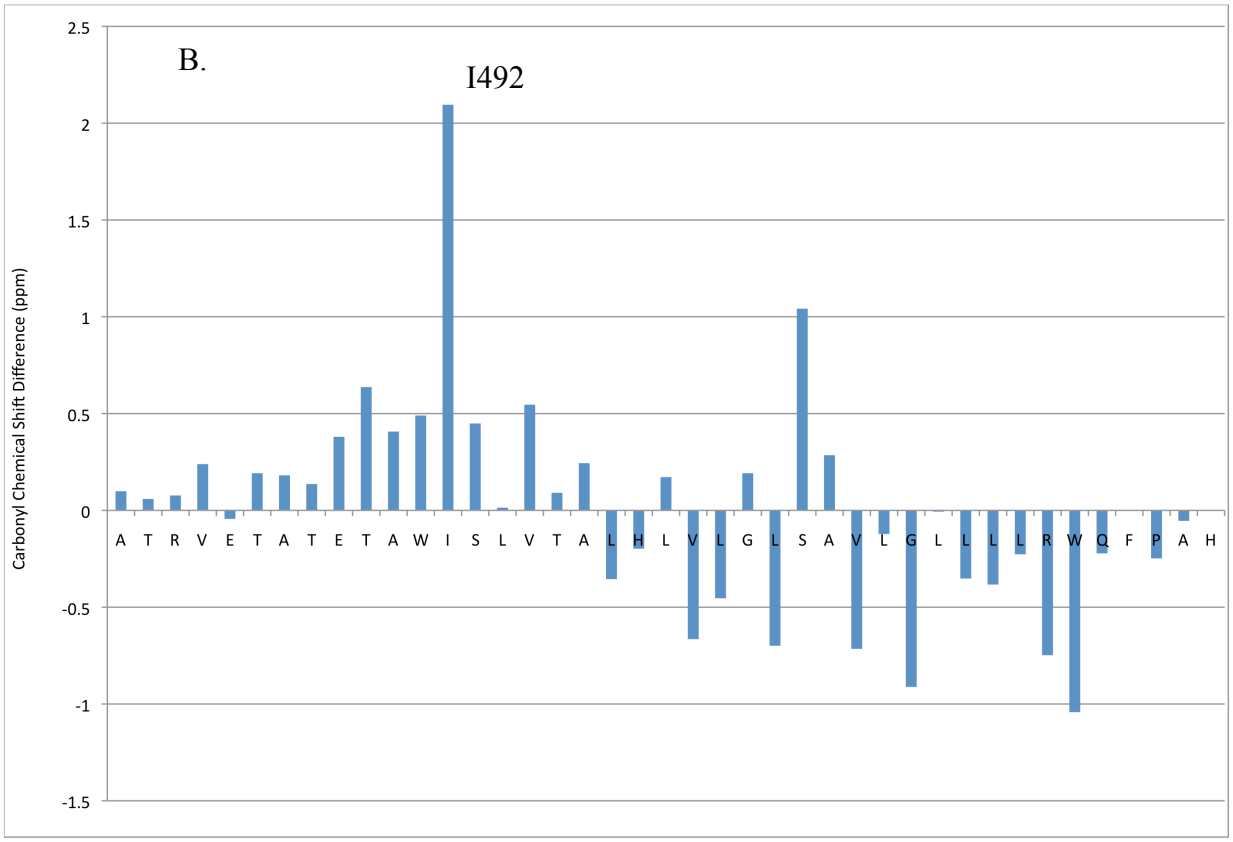



***Figure 6.2 Backbone chemical shift changes between WT and S505N peptides suggest secondary structure changes from WT to S505N.*** *A) Backbone NH chemical shift changes from huTpoR$_{481-520}$ WT and S505N demonstrate large changes in I492 and T496. B) Carbonyl chemical shift changes indicate that the carbonyl of I492 shifts downfield in the S505N peptide.*

The NH bond vector is particularly sensitive to changes is secondary structure and local environment. Flexibility of the peptide backbone can be assessed by probing the $^{1}$H-$^{15}$N heteronuclear NOE relaxation NMR experiments (*97*). The heteronuclear NOE values of huTpoR$_{481-520}$ WT and S505N in SDS micelles are presented in Figure 6.3. One of the most rigid residues in the WT peptide is Thr496, previously identified as a residue that has a large NH chemical shift change from WT to S505N. In the S505N peptide, the rigidity of Thr496 decreases, while Thr487 becomes more rigid. The changes in rigidity of Thr487 and Thr496 seem to be coupled to the structural change indicated by the chemical shift deviations seen in the Ile492 carbonyl and the Thr496 NH. Together, these changes in flexibility are consistent with a

change in secondary structure of Ile492-Thr96 to α-helix. Why Thr487 is perturbed by a mutation at S505N is not clear. A Thr487 mutation to alanine has previously been identified as a clinically important TpoR activating mutant in humans (*34*). Perhaps a shift in the conformation of Thr487 is critical for achieving the active state. Interestingly, in a continuous helix, Thr487 would be on the same face as S505N. This would place it in the interface of a helix dimer that is mediated by the S505N mutation, potentially explaining the increase in rigidity seen in the S505N peptide.

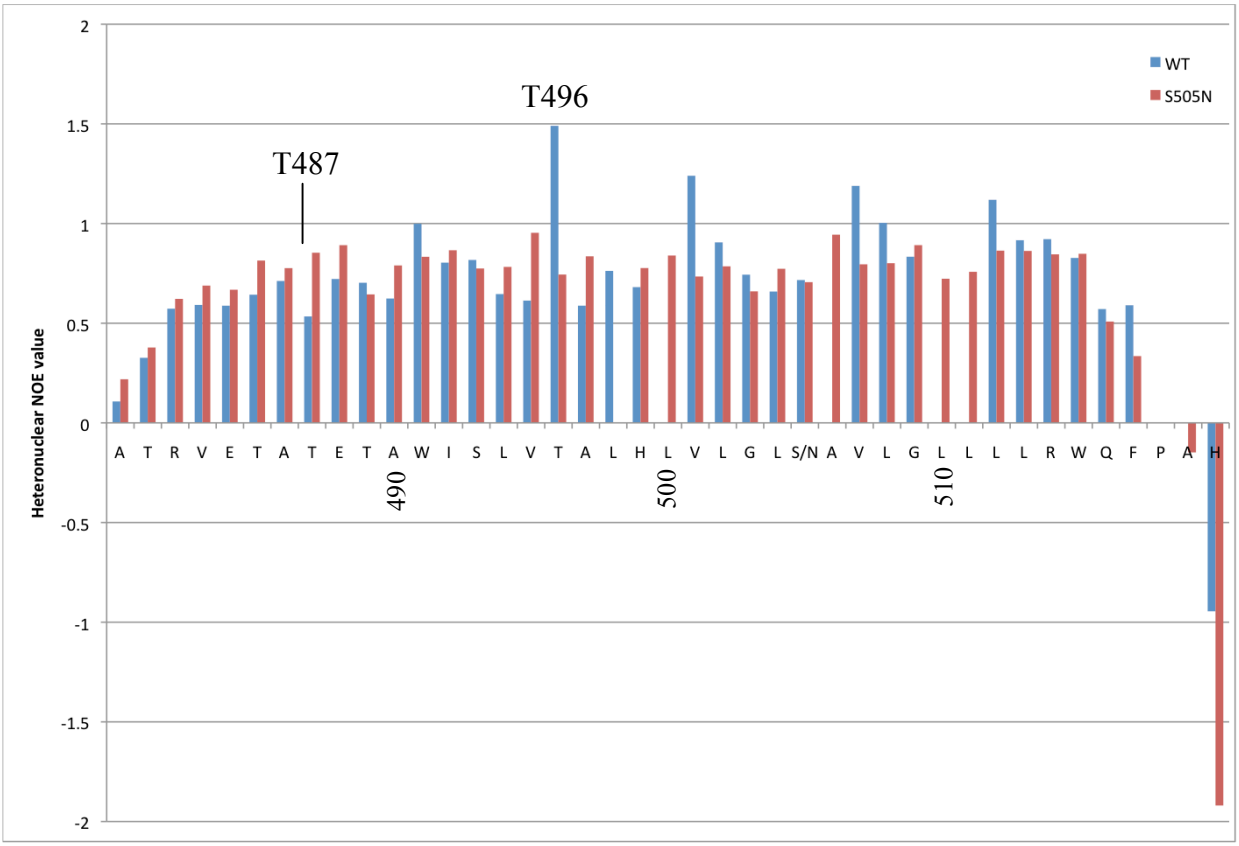



***Figure 6.3 Backbone flexibility of the WT and S505N huTpoR$_{481-520}$ peptides in SDS micelles.*** *$^{1}$H-$^{15}$N Heteronuclear HSQC experiments conducted on the WT and S505N peptides indicate a change in flexibility of two threonine residues at the N-terminus of the peptide. Interscan delay = 3s. blue=WT, red=S505N*

The structural information encoded in the chemical shift measurements allows prediction of dihedral angles using the program DANGLE (*105*). The dihedral angles were used to create

structural models using CNS (*102, 103*). A single round of structure calculation (representing structures that are the result of dihedral angle input and a single round of minimization) produced the structures in Figure 6.4. The WT model (in green) demonstrates a kink in the helix around T496, which is consistent with a structural model based upon solid-state NMR measurements of the WT peptide in lipid (*147*). Conversely, the S505N model (in cyan) shows a continuous helix in this region of T496, consistent with the downfield shift of the carbonyl of I492. Together these data demonstrate that the mutation at position 505, which is more than 2 helix turns away from T496, induces a dramatic structural change in the TMD.

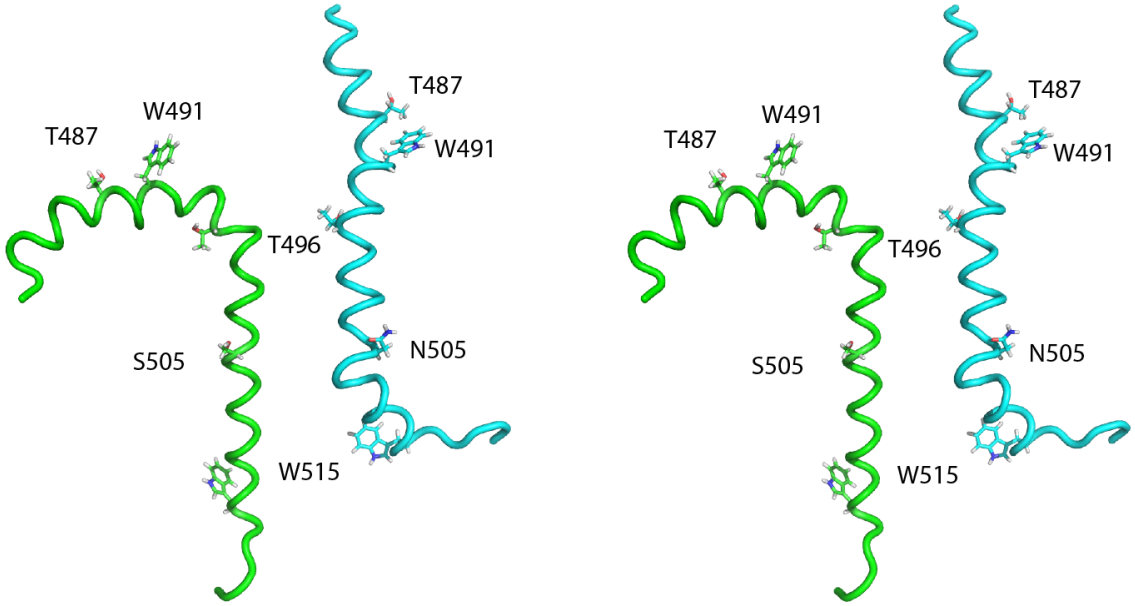



***Figure 6.4 Structural models of the WT and S505N huTpoR$_{481-520}$ peptides in SDS micelles.*** *Dihedral angles computed from backbone chemical shift measurements were used to create stereo structural models of the WT and S505N peptides. The sidechains of T487, W491, T496, S/N505, and W515 are visible. Green=WT, Cyan=S505N.*

That the $^{1}$H-$^{15}$N-HSQC for the TpoR WT and S505N peptides demonstrate such a dramatic structural change upon the mutation of a single amino acid is unexpected. The $^{1}$H-$^{15}$N-HSQC of muEpoR$_{218-268}$ peptides with the L223C mutation (Chapter 4) demonstrates mostly

local changes in the peptide structure, along with a few residues distal to the site (His249, Leu253) associated with a shift to the activating state. One potential model that explains how such a large change in structure occurs upon the mutation of a single amino acid is TMD oligomerization. In model TMD peptides, polar residues such as asparagine can drive helix association (*174*). Changes in the oligomerization status of the huTpoR$_{481\text{-}520}$ peptide in SDS detergent micelles upon mutation of serine 505 to asparagine can be assessed by relaxation NMR techniques (*97*). T1 (spin-lattice) relaxation and T2 (spin-spin) relaxation of the amide NH are sensitive to the rate of molecular tumbling and can be easily assessed with a series of $^{1}H$-$^{15}N$-HSQC based experiments. Histograms of these measurements are presented in Figure 6.5 and 6.6, demonstrating that while there is no difference in the T1 values for the WT peptide, the T2 values for the S505N peptide are roughly half that of the WT peptide. Faster relaxation for the S505N peptide is consistent with a larger complex, suggesting TM oligomerization.

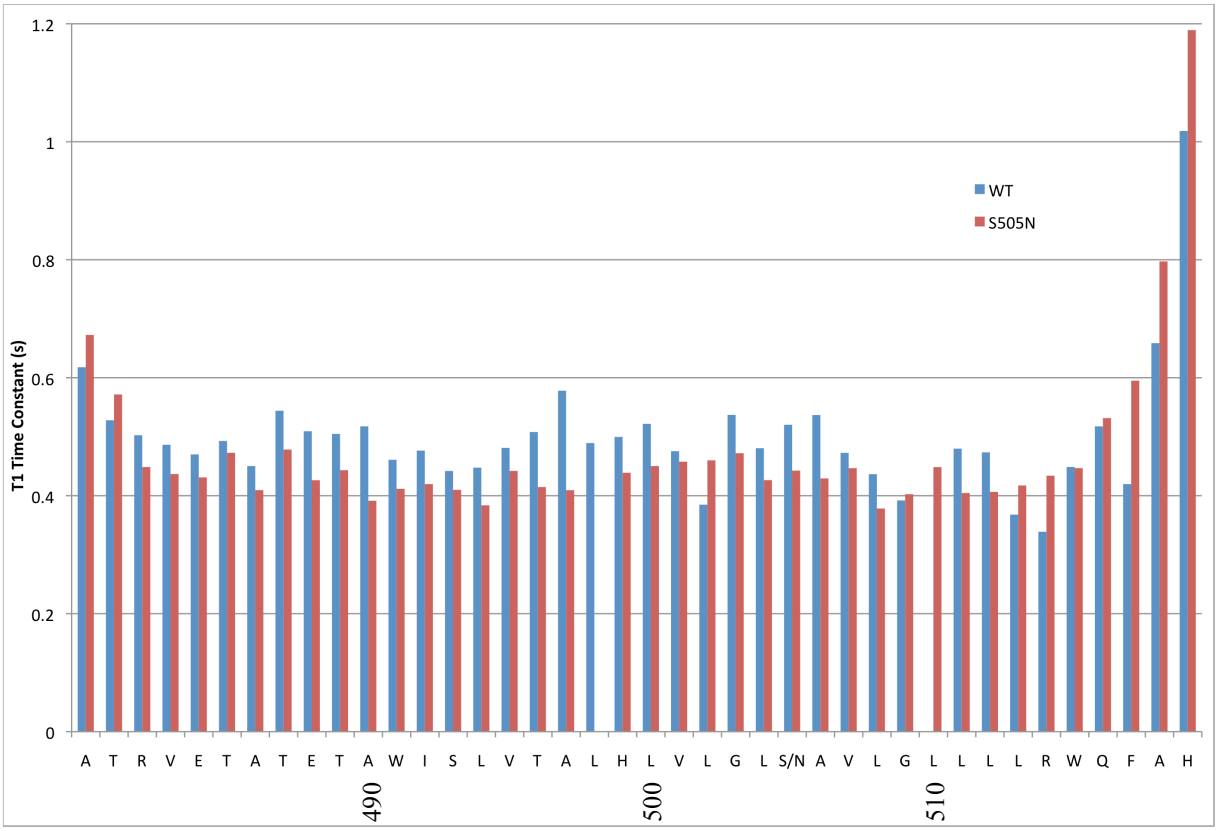



***Figure 6.5 Determination of T1 values for the WT and S505N huTpoR$_{481-520}$ peptides in SDS micelles.*** *T1 relaxation values for each NH in the WT and the S505N samples were determined using solution NMR. Pro518 has been omitted from the histogram. Interscan delay = 3s. blue=WT, red=S505N.*

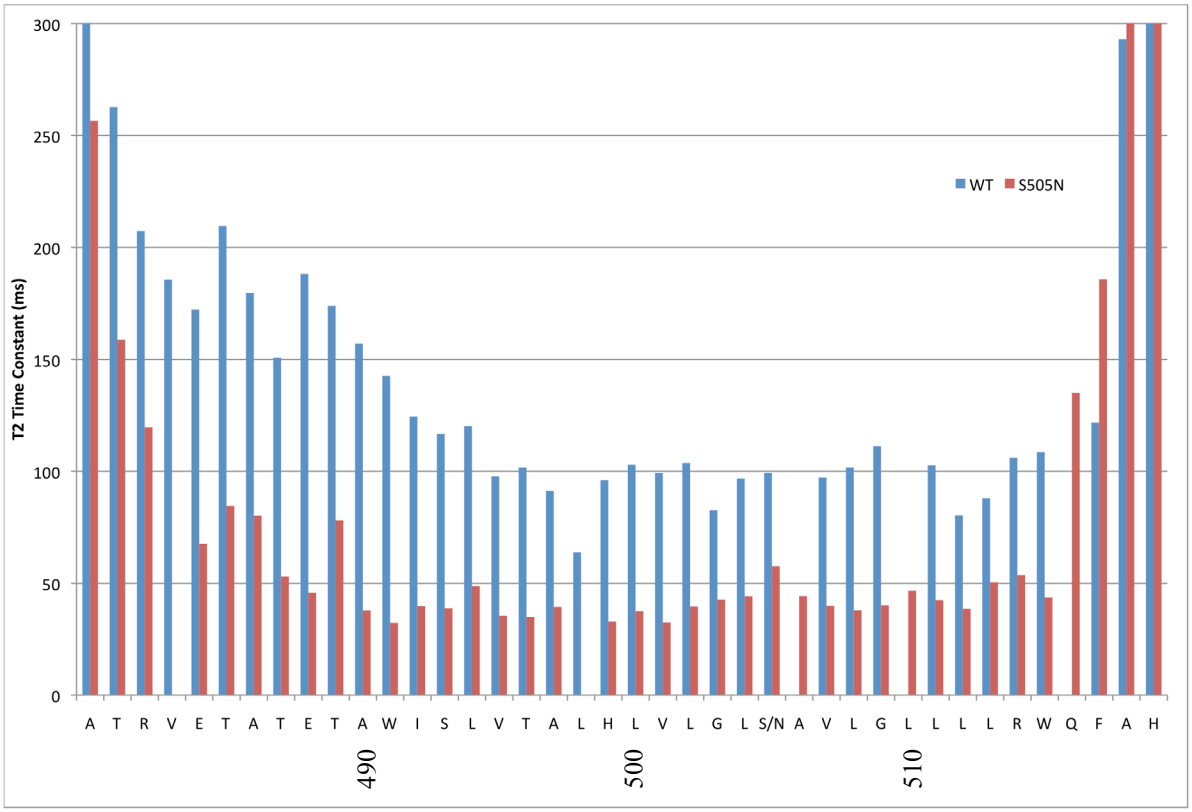



***Figure 6.6 Determination of T2 values for the WT and S505N huTpoR$_{481-520}$ peptides in SDS micelles.*** *T2 relaxation values for each NH in the WT and the S505N samples were determined using solution NMR. Pro518 has been omitted from the histogram. Interscan delay = 3s. blue=WT, red=S505N.*

The correlation time ($\tau_c$) of each residue in the complex can be calculated using the formula (*97*):

$$\tau_c = \left( \sqrt{\frac{6T_1}{T_2} - 7} \right) / 4\pi\upsilon_N$$

Typically, these values are averaged over the amino acids that demonstrate a heteronuclear NOE value greater than 0.6 (relatively rigid) to measure the correlation time of the molecule (*128*). The $\tau_c$ calculations were performed for the WT and S505N peptides solubilized in SDS micelles and the results are presented in the histogram below (Figure 6.7). Because the calculation of $\tau_c$ is

dependent on the ratio of T1/T2, as expected the correlation time of S505N is roughly twice the rate of the WT peptide. This is another indication that the oligomerization state has changed from the WT to S505N peptides.

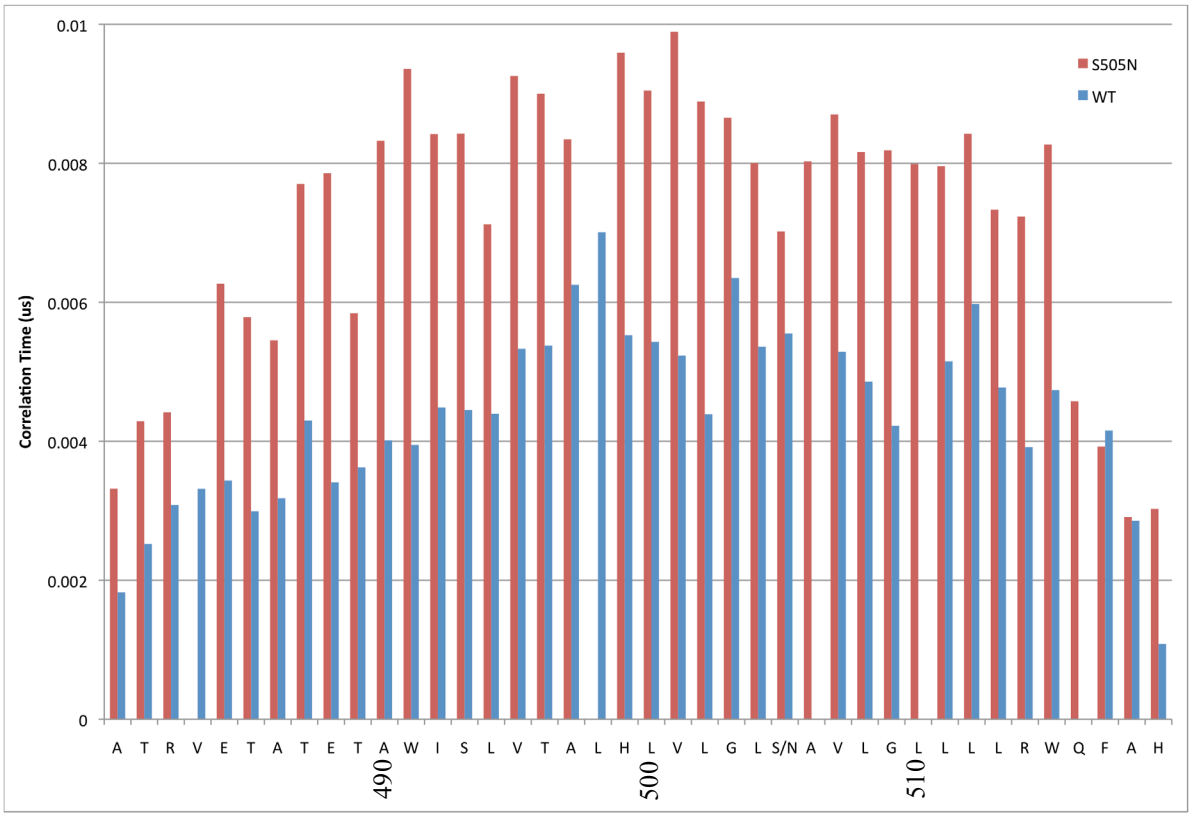



***Figure 6.7 Calculated values of $\tau_c$ values for the WT and S505N huTpoR$_{481-520}$ peptides in SDS micelles.*** *Correlation time of each NH in the WT and the S505N samples were determined from the T1/T2 relaxation values and the formula in the text. Pro518 has been omitted from the histogram. blue=WT, red=S505N.*

## Conclusions

The peak shifts in the $^{1}$H-$^{15}$N-HSQC between WT and S505N peptides suggests that the single amino acid difference causes a dramatic change in the structure of the TM peptide. The relaxation NMR data suggests that this dramatic structural change is related to a change in

oligomerization state of the peptide. The huTpoR$_{481-520}$ TM peptide monomer is ~4.7 kDa, and the average SDS micelle is ~17 kDa (288.3 g/mol x aggregation number 62), for a total peptide:detergent complex size of ~26.4 kDa. A study of the BNIP3 TM peptide dimer prepared in isotropic DHPC/DMPC bicelles with a complex size of ~50 kDa measured a correlation time of ~18 ns (*128*). The S505N correlation time is ~8 ns, about half the BNIP3 $\tau_c$, which fits well with the molecular weight of a huTpoR$_{481-520}$ peptide dimer in an SDS micelle. Interestingly, other data in our lab suggest that S505N induces TM dimerization. AUC experiments in DPC on peptides corresponding to the TMD of the human TpoR demonstrate that the WT sequence is monomeric and the S505N peptide forms TM dimers (*175*). Similar results were seen using deuterium NMR experiments with the same peptides reconstituted in DMPC bilayers (*175*).

***Interface***

If the S505N TM peptide forms a dimer in SDS micelles, then what is the peptide dimer interface? While there is not enough data to determine precisely what the interface of such a dimer would be, it is reasonable to hypothesize that the Asn505 residue would be in a helix dimer interface. Several studies on Asn-containing model TM peptides (*136, 138, 139*) and Asn mutagenesis of actual TMD peptides (*64, 176*) indicate that asparagine can mediate oligomerization of TM helices. If Asn505 is in the interface of the TM dimer, then two interfaces are possible, one where Asn505 is in the "a" position, one where it is in the "d"position (labeled "1" and "2" on Figure 6.8 below). If interface "1" is the actual interface, then this would place Thr487 and Trp515 in the interface of the active dimer. This is an interesting observation, as mutation of each of those residues causes disease associated with receptor hyperactivity in humans (*34, 35*). Interface "2" represents a model supported by TOXCAT assay data (*15*).

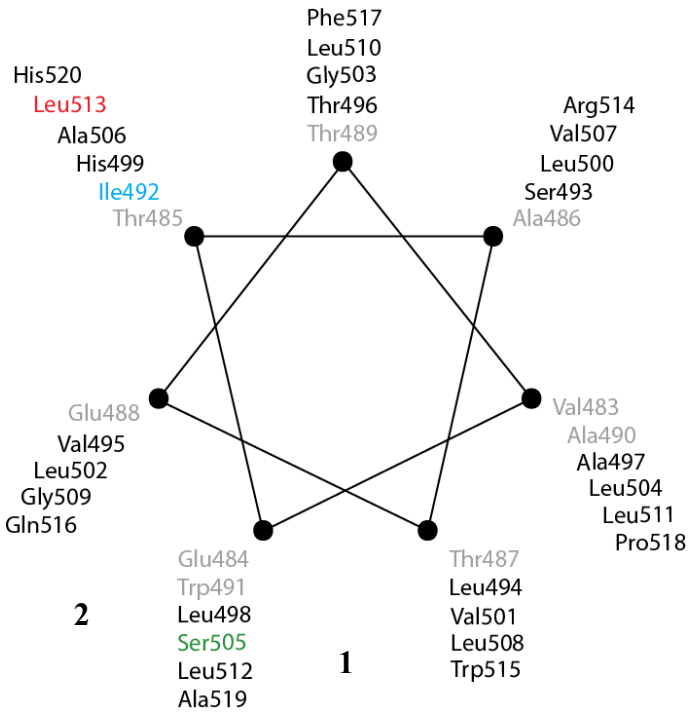



***Figure 6.8 Helical wheel diagram of the huTpoR$_{481-520}$ peptide TMD sequence.*** *Left-handed coiled coil diagram for the WT huTpoR$_{481-520}$ peptide sequence. Potential active interfaces "1" and "2" are labeled. Blue=TM helix start, red=TM helix end, gray=N-terminal helical region.*

The results of these experiments suggest that the clinically relevant S505N mutation induces receptor activation by driving dimerization of the TMD. This conclusion is supported by the change in correlation time of the peptide containing the S505N mutation, which suggests a change in oligomerization state. What is less clear is how the change in oligomerization state is associated with structural changes in the N-terminus of the receptor TMD. Using a combination of solid and solution NMR, Kim *et al*., (*147*) demonstrate that the WT TpoR TMD is broken into two helices, separated by a kink around I492-T496. This helical break is reproduced by our solution NMR data as well. It is clear from the downfield shift in carbonyl resonance of I492 and the amide chemical shift of T496 that in the presence of the S505N mutation, this region becomes more helical. The reason behind this structural change is unclear. One possibility is that

the straightening of the helix could occur to accommodate formation of interhelical interactions between N-terminal residues of opposite receptor monomers, a "zippering-up" of this region.

Assuming the Kim *et al*. (*147*) model is correct, and the N-terminal helix resides along the membrane surface, helix straightening would serve to lift it off the membrane. This could lift the entire EC-JM region off of the membrane. Interestingly, such a model has some support. Weidemann *et al*. (*177*) have hypothesized that the IL-4 receptor WSXWS motif functions as a molecular switch, modulating receptor activity. Using free energy perturbation calculations, they demonstrate that the tryptophan residues of the WSXWS motif could reside in two conformations. In the 'off,' or inactive state, the Trp residues partition into the membrane, tethering the ECD to the cell membrane. In the 'on,' or active state, the ECD is lifted off the membrane and the Trp residues rotate inward to interact in the Trp-Arg zipper motif seen in the various crystal and NMR structures of the ECD D2 domain (*4, 10*).

More recent research gives support to a role for WSXWS as a molecular switch. NMR studies on the PrlR ECD in the absence and presence of Prl demonstrate a change in the conformation of the tryptophan residues in the WSXWS motif; in the ligand-bound state, the Trp sidechains form the Trp-Arg zipper (*178*). However, in the unbound state, the two Trp sidechains form a T-stack conformation, with the edge of one Trp ring stacking on the face of the other Trp. The consequence of this conformational change is a structural rearrangement of the D2 domain β-sheets. These changes do not mirror the structural changes hypothesized by the free energy perturbation studies of Weidemann *et al*., (*177*), since lipid was clearly unavailable for Trp sidechain partitioning. Given that mutational studies have clearly implicated the WSXWS motif in receptor activity (*12*), and now structural studies have demonstrated a conformational difference of the WSXWS motif between the ligand bound and free ECD structures, it seems

reasonable that this motif plays a greater role in receptor function than originally thought. Clearly more research on larger receptor constructs in a biologically relevant system (lipid) is necessary to probe the structure and function of the hematopoietin receptor family. However, the experiments described here are the necessary precursor to these more difficult studies and establish the structural changes induced by the S505N mutation that are associated with receptor activation.

## Chapter 7-Conclusions and Future Directions

The data contained herein describe the structure of cytokine receptor TMD dimers and provide insights into the mechanism(s) of how structural changes induced by ligand binding may be transmitted through the TMD to the ICD. However, the broader picture that emerges from the study of these peptides includes three important observations. First, the structures of cytokine receptor TMD dimers can be elucidated by solution NMR. While the ability to derive structural information from TMD dimers has been well established (*81, 125, 179*), none of the cytokine receptor TMD structures have been reported. The second observation from these studies is that mutations in the TM or JM regions can be added to the WT peptides to assess structural changes either locally or distally. Assessment of structural changes upon the addition of the S505N mutation in the background of the huTpoR TMD peptide demonstrates that not only can differences in oligomerization state be examined, but also the structural changes associated with oligomerization. The third point is the observation that the sequence of the TMD encodes secondary structure information (α-helix) that directs the proper folding and assembly of this region. To date, solution NMR studies on single TM receptors have focused on separate regions of the receptor such as the D2 domain (*4*) or the TMD dimer (*122, 179*). These studies are the necessary first steps for investigation of relationships between receptor structure and function. The muEpoR$_{220\text{-}248}$ peptide spontaneously folds into an α-helix and dimerizes in DPC micelles (Chapter 3). When additional residues corresponding to the switch and Box 1 regions are added (muEpoR$_{218\text{-}268}$), the secondary structure and TMD dimer interactions are preserved (Chapter 4). The addition of these regions provides context for assessing the role of activating mutations with respect to receptor structural changes.

Finally, the recapitulation of interactions that occur *in vivo* can be made. While solution NMR has long been used to assess structural changes of soluble interacting proteins, it has not been widely used for TMD-containing proteins. The TMD of gp55-P interacts with the TMD of muEpoR, thereby activating the receptor. This work demonstrates that those TMD interactions can be recreated in a membrane mimetic system and associated structural perturbations can be assessed using solution NMR. Together, these experimental observations set the stage for collecting more biologically relevant data to elucidate the relationship between the structure and function of cytokine receptors.

**EpoR**

Several potential avenues for extending the studies of the EpoR are available. Two subsets of experiments in particular can be considered based on experiments already performed. These fall into two separate categories, the addition of secondary interacting molecules or proteins (*e.g.* gp55-P) or extension of the sequence to study different regions of the receptor. For the first subset of experiments, it would be helpful to characterize the molecular interactions of the EpoR and gp55-P TMDs. This would not just define the interface between the EpoR and gp55-P, but also help define what changes in the EpoR TMD are associated with activation. An additional experiment that would be helpful here is to study the human EpoR TMD sequence with the L239S mutation that allows gp55-P activation of the receptor with the human sequence (*39*).

Interestingly, there is a second, albeit synthetic, TMD-containing protein that has been demonstrated to activate the EpoR through TMD interactions (*180*). This peptide has no sequence homology to gp55-P. It would be instructive to perform the same structural studies

with the muEpoR$_{218-268}$ peptide in DPC micelles to assess whether the same structural changes are seen as with the active constructs studied here (L223C, +gp55-P).

Extension of the sequence construct from the core TMD region that we have studied here could help shed light on regions of the receptor that to date have not been studied. For instance, the JM regions of the receptor are areas of structural transition between the ECD β-sheet rich region and the α-helical TMD. In TpoR, mutations in these regions cause constitutive receptor activity and disease (*34, 35*). However, the changes in structure of these regions associated with the active state are unknown. By adding the sequence of the ECD D2 domain to the TMD-containing peptide, the structure of the JM region between the WSXWS motif and the N-terminus of the TMD in a more native conformation can be studied.

Furthermore, two modalities in addition to the JM cysteine mutations (*74*) can be used to simulate the active state. The R129C mutation causes EpoR hyperactivity by inducing receptor dimer crosslinking (*32*). This mutation can be introduced and NMR can ascertain structural changes from the WT peptide. A second method of activation is to add an activating peptide. Naranda *et al*. (*181*) have determined that a sequence corresponding to residues 194-216 of the EpoR activates the full length EpoR. This peptide could be added to WT D2-TM peptides to simulate an active state, and again structural changes caused by the addition of the activating peptide can be determined through NMR experiments. The EpoR 194-216 peptide has been shown to bind to its identical sequence on the full receptor (*181*). Interestingly, the peptide includes an arginine residue (R171) demonstrated by X-ray crystallography to interact with a tryptophan residue (W212) of the WSXWS motif (*8*), suggesting this region has a role in receptor activation. Moreover, given the concentration of activating mutations in the EC-JM region in both the EpoR and TpoR, perhaps the WSXWS motif and the EC-JM region work in

concert to regulate receptor activity. The structural importance of this region is an area of active study, as a recent study of PrlR demonstrated that the WSXWS motif acts as a molecular switch (*178*). Clearly, study of this region of the receptor is critically important for determining how changes in ECD orientation upon ligand binding are transmitted to the TMD, and would ultimately shed light on an activation mechanism of these receptors.

Extension of the TMD-containing construct further into the ICD would be helpful as well. We have cloned and expressed a construct (muEpoR$_{218-392}$) that extends the TM+Box 1 construct through the Box 2 region and includes the first tyrosine phosphorylated by Jak2 upon Epo binding (Y343). We were able to purify this peptide, reconstitute it into DPC micelles and perform NMR studies. Unfortunately, the spectrum exhibited severe overlap, either because of chemical shift broadening due to slow molecular tumbling (approximate complex size is ~55 kDa) or conformational heterogeneity (data not shown). However, use of technological advances for study of larger complexes can be helpful here. Deuteration of peptide sidechains, combined with TROSY NMR pulse sequences (*119*), can narrow linewidths of resonances in larger molecular weight complexes and make them amenable to study by solution NMR methods.

Once the structural conformations of the WT muEpoR$_{218-268}$ and muEpoR$_{218-392}$ complexes have been determined, the FERM domain from Jak2, expressed separately, can be added to each peptide complex. This would provide information on the binding site of the FERM domain on the EpoR and how that interaction changes the local and the global structure of the complex. Difficulties due to the size of this complex may arise, but can be overcome with the TROSY method mentioned above (*119*). Structural details of near-megadalton protein complexes have successfully been studied in this manner (*182*).

**TpoR**

The avenues for continuing the research on the TpoR TMD are similar to the EpoR project. Given that the data described here clearly implicate the S505N mutation in receptor TMD association, it seems natural to continue experiments to determine the receptor interface. This would be more than merely an academic exercise. The rotational orientation of the murine TpoR TMD is linked to differential signaling of the receptor (*183*). The interface represented by the active S505N TM dimer induces proliferation to near WT-like levels, activating the Jak2/STATs, the MAPK and PI3K pathways, but not Tyk2. Further structural study of these interfaces will elucidate how the changes in TMD rotational orientation is coupled to changes in ICD orientation, and thus allows differential activation of intracellular pathways to effect specific physiological responses.

Another interesting question to be answered involves the role of the K/RWQFP insert in TpoR. This insert modulates the activity of the receptor in a sequence specific fashion (*158*). However, the actual mechanism of function is unknown. Mutation of Trp515 to lysine, leucine (*35*) or alanine (*158*) results in a constitutively active receptor, suggesting a role for Trp514 of this motif in the preventing activation. Even more mysterious is how activation by native ligand or S505N mutation is able to overcome the inhibition provided by this tryptophan. By studying TMD-containing peptides in the same manner as described here for the EpoR TM constructs, it is possible to determine how Trp515 and various mutants act in these cases. In a similar fashion, the structure of the active T487A mutant can be studied as well. It would be interesting to see if these mutants share a common structure and therefore method of activation, or if different structures lead to the same result.

Finally, in the same vein as the EpoR, longer TpoR TMD-containing constructs can be expressed that include the Box 1 region in order to determine the molecular contacts between the TpoR ICD and the FERM domain of Jak2. This experiment has particular importance, as the TpoR is known to use two different kinases, Jak2 and Tyk2 (*184*). Such experiments could help determine the mechanism of kinase specificity in cytokine receptors.

Of course, while the potential experiments described here primarily utilize solution NMR, other approaches to assess structure should not be ruled out. X-ray crystallography may be useful in determining the structure of the ICD region if expressed by itself, or in complex with the FERM domain of Jak2. Solid-state NMR experiments are generally unable to obtain the resolution of solution NMR experiments, but have the advantage of being able to examine peptide structure in a native lipid environment, which may be desirable.

These structural characterizations or TMD containing peptides all lead to increasing the understanding of how these receptors function. Once Epo is bound to the ECD, the TMD shifts to the active conformation, and the ICD structure is reorganized to allow activation of Jak2. But how does this occur? If Jak2 disengages from Box 1 when the EpoR is activated, what structural changes would cause this to happen? What methods of regulation would exist to keep a free, soluble Jak2 from phosphorylating without limits? Much still remains to be discovered regarding the function of these receptors. It will be even longer until therapeutics can be developed based on the structural insights gleaned from these studies. However, the studies described here begin to define the role of the TMD in these receptors both structurally and functionally and represent the necessary first steps for understanding how these proteins function.

## References


1. Boulay, J. L., O'Shea, J. J., and Paul, W. E. (2003) Molecular phylogeny within type I cytokines and their cognate receptors, *Immunity 19*, 159-163.

2. Liongue, C., and Ward, A. C. (2007) Evolution of class I cytokine receptors, *BMC Evol. Biol. 7*.

3. Philo, J. S., Aoki, K. H., Arakawa, T., Narhi, L. O., and Wen, J. (1996) Dimerization of the extracellular domain of the erythropoietin (EPO) receptor by EPO: One high-affinity and one low-affinity interaction, *Biochemistry 35*, 1681-1691.

4. Kernebeck, T., Pflanz, S., Müller-Newen, G., Kurapkat, G., Scheek, R. M., Dijkstra, K., Heinrich, P. C., Wollmer, A., Grzesiek, S., and Grötzinger, J. (1999) The signal transducer gp130: Solution structure of the carboxy-terminal domain of the cytokine receptor homology region, *Protein Sci. 8*, 5-12.

5. Hecht, O., Dingley, A. J., Schwanter, A., Özbek, S., Rose-John, S., and Grötzinger, J. (2006) The solution structure of the membrane-proximal cytokine receptor domain of the human interleukin-6 receptor, *Biol. Chem. 387*, 1255-1259.

6. Devos, 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.

7. van Agthoven, J., Zhang, C., Tallet, E., Raynal, B., Hoos, S., Baron, B., England, P., Goffin, V., and Broutin, I. (2010) Structural characterization of the stem-stem dimerization interface between prolactin receptor chains complexed with the natural hormone, *J. Mol. Biol. 404*, 112-126.

8. Syed, R. S., Reid, S. W., Li, C. W., Cheetham, J. C., Aoki, K. H., Liu, B. S., Zhan, H. J., Osslund, T. D., Chirino, A. J., Zhang, J. D., 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.

9. 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.

10. Baumgartner, J. W., Wells, C. A., Chen, C. M., and Waters, M. J. (1994) The role of the WSXWS equivalent motif in growth-hormone receptor function, *J. Biol. Chem. 269*, 29094-29101.

11. Quelle, D. E., Quelle, F. W., and Wojchowski, D. M. (1992) Mutations in the WSAWSE and cytosolic domains of the erythropoietin receptor affect signal transduction and ligand-binding and internalization, *Mol. Cell. Biol. 12*, 4553-4561.

12. Hilton, D. J., Watowich, S. S., Katz, L., and Lodish, H. F. (1996) Saturation mutagenesis of the WSXWS motif of the erythropoietin receptor, *J. Biol. Chem. 271*, 4699-4708.

13. Yang, N., Wang, X., Jiang, J., and Frank, S. J. (2007) Role of the growth hormone (GH) receptor transmembrane domain in receptor predimerization and GH-induced activation, *Mol. Endocrinol. 21*, 1642-1655.

14. Constantinescu, S. N., Keren, T., Russ, W. P., Ubarretxena-Belandia, I., Malka, Y., Kubatzky, K. F., Engelman, D. M., Lodish, H. F., and Henis, Y. I. (2003) The erythropoietin receptor transmembrane domain mediates complex formation with viral anemic and polycythemic gp55 proteins, *J. Biol. Chem. 278*, 43755-43763.

15. Matthews, E. E., Thevenin, D., Rogers, J. M., Gotow, L., Lira, P. D., Reiter, L. A., Brissette, W. H., and Engelman, D. M. (2011) Thrombopoietin receptor activation: transmembrane helix dimerization, rotation, and allosteric modulation, *FASEB J. 25*, 2234-2244.

16. Gadd, S. L., and Clevenger, C. V. (2006) Ligand-independent dimerization of the human prolactin receptor isoforms: Functional implications, *Mol. Endocrinol. 20*, 2734-2746.

17. Liu, W., and Brooks, C. L. (2011) Functional impact of manipulation on the relative orientation of human prolactin receptor domains, *Biochemistry 50*, 5333-5344.

18. Russ, W. P., and Engelman, D. M. (1999) TOXCAT: A measure of transmembrane helix association in a biological membrane, *Proc. Natl. Acad. Sci. U. S. A. 96*, 863-868.

19. Kubatzky, K. F., Ruan, W. M., Gurezka, R., Cohen, J., Ketteler, R., Watowich, S. S., Neumann, D., Langosch, D., and Klingmüller, U. (2001) Self assembly of the transmembrane domain promotes signal transduction through the erythropoietin receptor, *Curr. Biol. 11*, 110-115.

20. Eilers, M., Patel, A. B., Liu, W., and Smith, S. O. (2002) Comparison of helix interactions in membrane and soluble α-bundle proteins, *Biophys. J. 82*, 2720-2736.

21. Zang, H. S., Sato, K., Nakajima, H., McKay, C., Ney, P. A., and Ihle, J. N. (2001) The distal region and receptor tyrosines of the Epo receptor are non-essential for *in vivo* erythropoiesis, *EMBO J. 20*, 3156-3166.

22. Huang, L. J., Shen, Y. M., and Bulut, G. B. (2010) Advances in understanding the pathogenesis of primary familial and congenital polycythaemia, *Br. J. Haematol. 148*, 844-852.

23. Pelletier, S., Gingras, S., Funakoshi-Tago, M., Howell, S., and Ihle, J. N. (2006) Two domains of the erythropoietin receptor are sufficient for Jak2 binding/activation and function, *Mol. Cell. Biol. 26*, 8527-8538.

24. Tong, W., Sulahian, R., Gross, A. W., Hendon, N., Lodish, H. F., and Huang, L. J. S. (2006) The membrane-proximal region of the thrombopoietin receptor confers its high surface expression by JAK2-dependent and -independent mechanisms, *J. Biol. Chem. 281*, 38930-38940.

25. Hintzen, C., Evers, C., Lippok, B. E., Volkmer, R., Heinrich, P. C., Radtke, S., and Hermanns, H. M. (2008) Box 2 region of the oncostatin M receptor determines specificity for recruitment of Janus kinases and STAT5 activation, *J. Biol. Chem. 283*, 19465-19477.

26. Miller, C. P., Liu, Z. Y., Noguchi, C. T., and Wojchowski, D. M. (1999) A minimal cytoplasmic subdomain of the erythropoietin receptor mediates erythroid and megakaryocytic cell development, *Blood 94*, 3381-3387.

27. von Heijne, G. (1986) Mitochondrial targeting sequences may form amphiphilic helices, *EMBO J. 5*, 1335-1342.

28. von Heijne, G. (1986) The distribution of positively charged residues in bacterial inner membrane-proteins correlates with the trans-membrane topology, *EMBO J. 5*, 3021-3027.

29. Hitchcock, I. S., Chen, M. M., King, J. R., and Kaushansky, K. (2008) YRRL motifs in the cytoplasmic domain of the thrombopoietin receptor regulate receptor internalization and degradation, *Blood 112*, 2222-2231.

30. Kralovics, R., Passamonti, F., Buser, A. S., Teo, S., Tiedt, R., Passweg, J. R., Tichelli, A., Cazzola, M., and Skoda, R. C. (2005) A gain-of-function mutation of JAK2 in myeloproliferative disorders, *New Engl. J. Med. 352*, 1779-1790.

31. Ungureanu, D., Wu, J. H., Pekkala, T., Niranjan, Y., Young, C., Jensen, O. N., Xu, C. F., Neubert, T. A., Skoda, R. C., Hubbard, S. R., and Silvennoinen, O. (2011) The pseudokinase domain of JAK2 is a dual-specificity protein kinase that negatively regulates cytokine signaling, *Nat. Struct. Mol. Biol. 18*, 971-U921.

32. Watowich, S. S., Yoshimura, A., Longmore, G. D., Hilton, D. J., Yoshimura, Y., and Lodish, H. F. (1992) Homodimerization and constitutive activation of the erythropoietin receptor, *Proc. Natl. Acad. Sci. U. S. A. 89*, 2140-2144.

33. Ding, F., Komatsu, H., Wakita, A., Kato-Uranishi, M., Ito, M., Satoh, A., Tsuboi, K., Nitta, M., Miyazaki, H., Lida, S., and Ueda, R. (2004) Familial essential thrombocythemia associated with a dominant-positive activating mutation of the c-MPL gene, which encodes for the receptor for thrombopoietin, *Blood 103*, 4198-4200.

34. Malinge, S., Ragu, C., Della-Valle, V., Pisani, D., Constantinescu, S. N., Perez, C., Villeval, J. L., Reinhardt, D., Landman-Parker, J., Michaux, L., Dastugue, N., Baruchel, A., Vainchenker, W., Bourquin, J. P., Penard-Lacronique, V., and Bernard, O. A. (2008) Activating mutations in human acute megakaryoblastic leukemia, *Blood 112*, 4220-4226.

35. Pardanani, A. D., Levine, R. L., Lasho, T., Pikman, Y., Mesa, R. A., Wadleigh, M., Steensma, D. P., Elliott, M. A., Wolanskyj, A. R., Hogan, W. J., McClure, R. F., Litzow, M. R., Gilliland, D. G., and Tefferi, A. (2006) MPL515 mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients, *Blood 108*, 3472-3476.

36. Ihara, K., Ishii, E., Eguchi, M., Takada, H., Suminoe, A., Good, R. A., and Hara, T. (1999) Identification of mutations in the c-mpl gene in congenital amegakaryocytic thrombocytopenia, *Proc. Natl. Acad. Sci. U. S. A. 96*, 3132-3136.

37. Ruscetti, S. K. (1995) Erythroleukemia Induction by the Friend Spleen Focus-Forming Virus, *Baillieres Clinical Haematology 8*, 225-247.

38. Zon, L. I., Moreau, J.-F., Koo, J.-W., Mathey-Prevot, B., and D'Andrea, A. D. (1992) The erythropoietin receptor transmembrane region is necessary for activation by the Friend spleen focus-forming virus gp55 glycoprotein, *Mol. Cell. Biol. 12*, 2949-2957.

39. Constantinescu, S. N., Liu, X. D., Beyer, W., Fallon, A., Shekar, S. C., Henis, Y. I., Smith, S. O., and Lodish, H. F. (1999) Activation of the erythropoietin receptor by the gp55-P viral envelope protein is determined by a single amino acid in its transmembrane domain, *EMBO J. 18*, 3334-3347.

40. Jacobson, L. O., Goldwasser, E., Fried, W., and Plzak, L. (1957) Role of the kidney in erythropoiesis, *Nature 179*, 633-634.

41. Hattangadi, S. M., Wong, P., Zhang, L. B., Flygare, J., and Lodish, H. F. (2011) From stem cell to red cell: Regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications, *Blood 118*, 6258-6268.

42. Chateauvieux, S., Grigorakaki, C., Morceau, F., Dicato, M., and Diederich, M. (2011) Erythropoietin, erythropoiesis and beyond, *Biochem. Pharmacol. 82*, 1291-1303.

43. Miyake, T., Kung, C. K. H., and Goldwasser, E. (1977) Purification of human erythropoietin, *J. Biol. Chem. 252*, 5558-5564.

44. 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 angstrom, *Science 273*, 464-471.

45. Frank, S. J. (2002) Minireview: Receptor dimerization in GH and erythropoietin action - It takes two to tango, but how?, *Endocrinology 143*, 2-10.

46. Huang, L. J. S., Constantinescu, S. N., and Lodish, H. F. (2001) The N-terminal domain of Janus kinase 2 is required for Golgi processing and cell surface expression of erythropoietin receptor, *Mol. Cell 8*, 1327-1338.

47. Royer, Y., Staerk, J., Costuleanu, M., Courtoy, P. J., and Constantinescu, S. N. (2005) Janus kinases affect thrombopoietin receptor cell surface localization and stability, *J. Biol. Chem. 280*, 27251-27261.

48. Bulut, G. B., Sulahian, R., Ma, Y., Chi, N. W., and Huang, L. J. S. (2011) Ubiquitination regulates the internalization, endolysosomal sorting, and signaling of the erythropoietin receptor, *J. Biol. Chem. 286*, 6449-6457.

49. Walrafen, P., Verdier, F., Kadri, Z., Chrétien, S., Lacombe, C., and Mayeux, P. (2005) Both proteasomes and lysosomes degrade the activated erythropoietin receptor, *Blood 105*, 600-608.

50. 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.

51. Richmond, T. D., Chohan, M., and Barber, D. L. (2005) Turning cells red: Signal transduction mediated by erythropoietin, *Trends Cell Biol. 15*, 146-155.

52. 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.

53. Tenhumberg, S., Schuster, B., Zhu, L. X., Kovaleva, M., Scheller, J., Kallen, K. J., and Rose-John, S. (2006) gp130 dimerization in the absence of ligand: Preformed cytokine receptor complexes, *Biochem. Biophys. Res. Commun. 346*, 649-657.

54. Yu, X. C., Sharma, K. D., Takahashi, T., Iwamoto, R., and Mekada, E. (2002) Ligand-independent dimer formation of epidermal growth factor receptor (EGFR) is a step separable from ligand-induced EGFR signaling, *Molecular Biology of the Cell 13*, 2547-2557.

55. Chung, I., Akita, R., Vandlen, R., Toomre, D., Schlessinger, J., and Mellman, I. (2010) Spatial control of EGF receptor activation by reversible dimerization on living cells, *Nature 464*, 783-787.

56. Tao, R. H., and Maruyama, I. N. (2008) All EGF(ErbB) receptors have preformed homo- and heterodimeric structures in living cells, *J. Cell Sci. 121*, 3207-3217.

57. Clayton, A. H. A., Orchard, S. G., Nice, E. C., Posner, R. G., and Burgess, A. W. (2008) Predominance of activated EGFR higher-order oligomers on the cell surface, *Growth Factors 26*, 316-324.

58. Johnson, D. L., Farrell, F. X., Barbone, F. P., McMahon, F. J., Tullai, J., Hoey, K., Livnah, O., Wrighton, N. C., Middleton, S. A., Loughney, D. A., Stura, E. A., Dower, W. J., Mulcahy, L. S., Wilson, I. A., and Jolliffe, L. K. (1998) Identification of a 13 amino

acid peptide mimetic of erythropoietin and description of amino acids critical for the mimetic activity of EMP1, *Biochemistry 37*, 3699-3710.

59. Poger, D., and Mark, A. E. (2010) Turning the growth hormone receptor on: Evidence that hormone binding induces subunit rotation, *Proteins 78*, 1163-1174.

60. Pang, X., and Zhou, H.-X. (2012) A common model for cytokine receptor activation: combined scissor-like rotation and self-rotation of receptor dimer induced by class I cytokine, *PLoS Comp. Biol. 8*, e1002427.

61. Gurezka, R., Laage, R., Brosig, B., and Langosch, D. (1999) A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments, *J. Biol. Chem. 274*, 9265-9270.

62. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins, *Science 240*, 1759-1764.

63. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil, *Science 254*, 539-544.

64. Ruan, W. M., Becker, V., Klingmüller, U., and Langosch, D. (2004) The interface between self-assembling erythropoietin receptor transmembrane segments corresponds to a membrane-spanning leucine zipper, *J. Biol. Chem. 279*, 3273-3279.

65. Ebie, A. Z., and Fleming, K. G. (2007) Dimerization of the erythropoietin receptor transmembrane domain in micelles, *J. Mol. Biol. 366*, 517-524.

66. Russ, W. P., and Engelman, D. M. (2000) The GxxxG motif: A framework for transmembrane helix-helix association, *J. Mol. Biol. 296*, 911-919.

67. Sulistijo, E. S., and MacKenzie, K. R. (2006) Sequence dependence of BNIP3 transmembrane domain dimerization implicates side-chain hydrogen bonding and a tandem GxxxG motif in specific helix-helix interactions, *J. Mol. Biol. 364*, 974-990.

68. Berger, B. W., Kulp, D. W., Span, L. M., DeGrado, J. L., Billings, P. C., Senes, A., Bennett, J. S., and DeGrado, W. F. (2010) Consensus motif for integrin transmembrane helix association, *Proc. Natl. Acad. Sci. U. S. A. 107*, 703-708.

69. Miura, O., Cleveland, J. L., and Ihle, J. N. (1993) Inactivation of erythropoietin receptor function by point mutations in a region having homology with other cytokine receptors, *Mol. Cell. Biol. 13*, 1788-1795.

70. He, T. C., Jiang, N., Zhuang, H. M., Quelle, D. E., and Wojchowski, D. M. (1994) The extended Box-2 subdomain of erythropoietin receptor is nonessential for JAK2 activation yet critical for efficient mitogenesis in FDC-ER cells, *J. Biol. Chem. 269*, 18291-18294.

71. Quelle, F. W., Wang, D., Nosaka, T., Thierfelder, W. E., Stravopodis, D., Weinstein, Y., and Ihle, J. N. (1996) Erythropoietin induces activation of Stat5 through association with specific tyrosines on the receptor that are not required for a mitogenic response, *Mol. Cell. Biol. 16*, 1622-1631.

72. Constantinescu, S. N., Huang, L. J. S., Nam, H. S., and Lodish, H. F. (2001) The erythropoietin receptor cytosolic juxtamembrane domain contains an essential, precisely oriented, hydrophobic motif, *Mol. Cell 7*, 377-385.

73. Constantinescu, S. N., Keren, T., Socolovsky, M., Nam, H. S., Henis, Y. I., and Lodish, H. F. (2001) Ligand-independent oligomerization of cell-surface erythropoietin receptor is mediated by the transmembrane domain, *Proc. Natl. Acad. Sci. U. S. A. 98*, 4379-4384.

74. Kubatzky, K. F., Liu, W., Goldgraben, K., Simmerling, C., Smith, S. O., and Constantinescu, S. N. (2005) Structural requirements of the extracellular to transmembrane domain junction for erythropoietin receptor function, *J. Biol. Chem. 280*, 14844-14854.

75. Lu, X. H., Gross, A. W., and Lodish, H. F. (2006) Active conformation of the erythropoietin receptor: Random and cysteine-scanning mutagenesis of the extracellular juxtamembrane and transmembrane domains, *J. Biol. Chem. 281*, 7002-7011.

76. Seubert, N., Royer, Y., Staerk, J., Kubatzky, K. F., Moucadel, V., Krishnakumar, S., Smith, S. O., and Constantinescu, S. N. (2003) Active and inactive orientations of the transmembrane and cytosolic domains of the erythropoietin receptor dimer, *Mol. Cell 12*, 1239-1250.

77. Becker, V., Sengupta, D., Ketteler, R., Ullmann, G. M., Smith, J. C., and Klingmüller, U. (2008) Packing density of the erythropoietin receptor transmembrane domain correlates with amplification of biological responses, *Biochemistry 47*, 11771-11782.

78. Fang, C., Choi, E., Nie, L. G., and Li, J. P. (1998) Role of the transmembrane sequence of spleen focus-forming virus gp55 in erythroleukemogenesis, *Virology 252*, 46-53.

79. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, Plainview, NY.

80. Aitken, A., and Learmonth, M. (1996) Protein Determination by UV Absorption, In *The Protein Protocols Handbook* (Walker, J. M., Ed.), pp 3-6, Springer Science+Business Media, LLC, Philadelphia, PA.

81. Sulistijo, E. S., and MacKenzie, K. R. (2009) Structural basis for dimerization of the BNIP3 transmembrane domain, *Biochemistry 48*, 5106-5120.

82. Fleming, K. G., Ackerman, A. L., and Engelman, D. M. (1997) The effect of point mutations on the free energy of transmembrane alpha-helix dimerization, *J. Mol. Biol. 272*, 266-275.

83. Kochendoerfer, G. G., Salom, D., Lear, J. D., Wilk-Orescan, R., Kent, S. B. H., and DeGrado, W. F. (1999) Total chemical synthesis of the integral membrane protein influenza A virus M2: Role of its C-terminal domain in tetramer, *Biochemistry 38*, 11905-11913.

84. Zwahlen, C., Legault, P., Vincent, S. J. F., Greenblatt, J., Konrat, R., and Kay, L. E. (1997) Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex, *J. Am. Chem. Soc. 119*, 6711-6721.

85. Diercks, T., Coles, M., and Kessler, H. (1999) An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments, *J. Biomol. NMR 15*, 177-180.

86. Muhandiram, D. R., and Kay, L. E. (1994) Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity, *J. Magn. Reson. Ser. B 103*, 203-216.

87. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) $^{1}H$, $^{13}C$ and $^{15}N$ chemical shift referencing in biomolecular NMR, *J. Biomol. NMR 6*, 135-140.

88. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., and Wüthrich, K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids - IUPAC-IUBMB-IUPAB Inter-Union Task Group on the Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy, *J. Biomol. NMR 12*, 1-23.

89. Kay, L. E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, *J. Am. Chem. Soc. 114*, 10663-10665.

90. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (1996) *Protein NMR Spectroscopy*, Academic Press, San Diego.

91. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.

92. Grzesiek, S., and Bax, A. (1992) Improved 3D triple-resonance NMR techniques applied to a 31-kDa protein, *J. Magn. Reson. 96*, 432-440.

93. Wittekind, M., and Mueller, L. (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins, *J. Magn. Reson. Ser. B 101*, 201-205.

94. Grzesiek, S., Anglister, J., and Bax, A. (1993) Correlation of backbone amide and aliphatic side-chain resonances in $^{13}C/^{15}N$-enriched proteins by isotropic mixing of $^{13}C$ magnetization, *J. Magn. Reson. Ser. B 101*, 114-119.

95. Montelione, G. T., Lyons, B. A., Emerson, S. D., and Tashiro, M. (1992) An efficient triple resonance experiment using $^{13}C$ isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins, *J. Am. Chem. Soc. 114*, 10974-10975.

96. Weisemann, R., Ruterjans, H., and Bermel, W. (1993) 3D Triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in 15N- and 13C-labelled proteins *J. Biomol. NMR 3*, 113-120.

97. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by $^{15}N$ inverse detected heteronuclear NMR-spectroscopy: Application to staphylococcal nuclease, *Biochemistry 28*, 8972-8979.

98. Hwang, T. L., van Zijl, P. C. M., and Mori, S. (1998) Accurate quantitation of water-amide proton exchange rates using the Phase-Modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme, *J. Biomol. NMR 11*, 221-226.

99. Kneller, D. G., and Kuntz, I. D. (1993) UCSF Sparky - An NMR display, annotation and assignment tool, *J. Cell. Biochem.*, 254-254.

100. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, P., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline, *Proteins 59*, 687-696.

101. Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E., and Nilges, M. (2007) ARIA2: Automated NOE assignment and data integration in NMR structure calculation, *Bioinformatics 23*, 381-382.

102. Brunger, A. T. (2007) Version 1.2 of the Crystallography and NMR system, *Nat. Protoc. 2*, 2728-2733.

103. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallographica Section D-Biological Crystallography 54*, 905-921.

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

105. Cheung, M. S., Maguire, M. L., Stevens, T. J., and Broadhurst, R. W. (2010) DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure, *J. Magn. Reson. 202*, 223-233.

106. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J. Biomol. NMR 13*, 289-302.

107. Habeck, M., Rieping, W., Linge, J. P., and Nilges, M. (2004) NOE assignment with ARIA 2.0: the nuts and bolts, *Methods Mol. Biol. 278*, 379-402.

108. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR 8*, 477-486.

109. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK - A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283-291.

110. Warschawski, D. E., Arnold, A. A., Beaugrand, M., Gravel, A., Chartrand, E., and Marcotte, I. (2011) Choosing membrane mimetics for NMR structural studies of transmembrane proteins, *Biochim. Biophys. Acta 1808*, 1957-1974.

111. Tribet, C., Audebert, R., and Popot, J. L. (1996) Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions, *Proc. Natl. Acad. Sci. U. S. A. 93*, 15047-15050.

112. Houliston, R. S., Hodges, R. S., Sharom, F. J., and Davis, J. H. (2003) Comparison of proto-oncogenic and mutant forms of the transmembrane region of the Neu receptor in TFE, *FEBS Lett. 535*, 39-43.

113. Libich, D. S., and Harauz, G. (2008) Solution NMR and CD spectroscopy of an intrinsically disordered, peripheral membrane protein: evaluation of aqueous and membrane-mimetic solvent conditions for studying the conformational adaptability of the 18.5 kDa isoform of myelin basic protein (MBP), *Eur. Biophys. J. Biophys. Lett. 37*, 1015-1029.

114. Sönnichsen, F. D., Vaneyk, J. E., Hodges, R. S., and Sykes, B. D. (1992) Effect of trifluoroethanol on protein secondary structure - an NMR and CD study using a synthetic actin peptide, *Biochemistry 31*, 8790-8798.

115. Mi, L. Z., Grey, M. J., Nishida, N., Walz, T., Lu, C. F., and Springer, T. A. (2008) Functional and structural stability of the epidermal growth factor receptor in detergent micelles and phospholipid nanodiscs, *Biochemistry 47*, 10314-10323.

116. Pervushin, K. V., and Arseniev, A. S. (1992) Three-dimensional structure of (1-36)bacterioopsin in methanol-chloroform mixture and SDS micelles determined by 2D 1H-NMR spectroscopy, *FEBS Lett. 308*, 190-196.

117. Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R., and Choe, S. (2005) NMR structure of Mistic, a membrane-integrating protein for membrane protein expression, *Science 307*, 1317-1321.

118. Lomize, A. L., Pervushin, K. V., and Arseniev, A. S. (1992) Spatial structure of (34-65)bacterioopsin polypeptide in SDS micelles determined from nuclear magnetic resonance data, *J. Biomol. NMR 2*, 361-372.

119. Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. (1997) Attenuated T-2 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, *Proc. Natl. Acad. Sci. U. S. A. 94*, 12366-12371.

120. Steyaert, J., and Kobilka, B. K. (2011) Nanobody stabilization of G protein-coupled receptor conformational states, *Curr. Opin. Struct. Biol. 21*, 567-572.

121. Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J., and Engelman, D. M. (1992) Sequence specificity in the dimerization of transmembrane α-helices, *Biochemistry 31*, 12719-12725.

122. Bocharov, E. V., Mineev, K. S., Volynsky, P. E., Ermolyuk, Y. S., Tkach, E. N., Sobol, A. G., Chupin, V. V., Kirpichnikov, M. P., Efremov, R. G., and Arseniev, A. S. (2008) Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state, *J. Biol. Chem. 283*, 6950-6956.

123. Bocharov, E. V., Mayzel, M. L., Volynsky, P. E., Goncharuk, M. V., Ermolyuk, Y. S., Schulga, A. A., Artemenko, E. O., Efremov, R. G., and Arseniev, A. S. (2008) Spatial structure and pH-dependent conformational diversity of dimeric transmembrane domain of the receptor tyrosine kinase EphA1, *J. Biol. Chem. 283*, 29385-29395.

124. Artemenko, E. O., Egorova, N. S., Arseniev, A. S., and Feofanov, A. V. (2008) Transmembrane domain of EphA1 receptor forms dimers in membrane-like environment, *Biochim. Biophys. Acta-Biomembr. 1778*, 2361-2367.

125. MacKenzie, K. R., Prestegard, J. H., and Engelman, D. M. (1997) A transmembrane helix dimer: Structure and implications, *Science 276*, 131-133.

126. Bocharov, E. V., Mayzel, M. L., Volynsky, P. E., Mineev, K. S., Tkach, E. N., Ermolyuk, Y. S., Schulga, A. A., Efremov, R. G., and Arseniev, A. S. (2010) Left-handed dimer of

EphA2 transmembrane domain: Helix Packing diversity among receptor tyrosine kinases, *Biophys. J. 98*, 881-889.

127. Call, M. E., Schnell, J. R., Xu, C., Lutz, R. A., Chou, J. J., and Wucherpfennig, K. W. (2006) The structure of the ζ ζ transmembrane dimer reveals features essential for its assembly with the T cell receptor, *Cell 127*, 355-368.

128. Bocharov, E. V., Pustovalova, Y. E., Pavlov, K. V., Volynsky, P. E., Goncharuk, M. V., Ermolyuk, Y. S., Karpunin, D. V., Schulga, A. A., Kirpichnikov, M. P., Efremov, R. G., Maslennikov, I. V., and Arseniev, A. S. (2007) Unique dimeric structure of BNip3 transmembrane domain suggests membrane permeabilization as a cell death trigger, *J. Biol. Chem. 282*, 16256-16266.

129. Chou, J. J., Kaufman, J. D., Stahl, S. J., Wingfield, P. T., and Bax, A. (2002) Micelle-induced curvature in a water-insoluble HIV-1 Env peptide revealed by NMR dipolar coupling measurement in stretched polyacrylamide gel, *J. Am. Chem. Soc. 124*, 2450-2451.

130. Khao, J., Arce-Lopera, J., Sturgis, J. N., and Duneau, J. P. (2011) Structure of a protein-detergent complex: The balance between detergent cohesion and binding, *Eur. Biophys. J. Biophys. Lett. 40*, 1143-1155.

131. Zhang, Y., Kulp, D. W., Lear, J. D., and DeGrado, W. F. (2009) Experimental and computational evaluation of forces directing the association of transmembrane helices, *J. Am. Chem. Soc. 131*, 11341-11343.

132. Gennis, R. B. (1989) Membrane dynamics and protein-lipid Interactions, In *Biomembranes*, pp 166-198, Springer-Verlag, New York.

133. Lipfert, J., Columbus, L., Chu, V. B., Lesley, S. A., and Doniach, S. (2007) Size and shape of detergent micelles determined by small-angle x-ray scattering, *J. Phys. Chem. B 111*, 12427-12438.

134. Rigby, A. C., Grant, C. W. M., and Shaw, G. S. (1998) Solution and solid state conformation of the human EGF receptor transmembrane region, *Biochim. Biophys. Acta 1371*, 241-253.

135. Adams, P. D., Engelman, D. M., and Brünger, A. T. (1996) Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching, *Proteins: Structure, Function, and Genetics 26*, 257-261.

136. Ruan, W. M., Lindner, E., and Langosch, D. (2004) The interface of a membrane-spanning leucine zipper mapped by asparagine-scanning mutagenesis, *Protein Sci. 13*, 555-559.

137. Li, R. H., Mitra, N., Gratkowski, H., Vilaire, G., Litvinov, R., Nagasami, C., Weisel, J. W., Lear, J. D., DeGrado, W. F., and Bennett, J. S. (2003) Activation of integrin αIIbβ3 by modulation of transmembrane helix associations, *Science 300*, 795-798.

138. Choma, C., Gratkowski, H., Lear, J. D., and DeGrado, W. F. (2000) Asparagine-mediated self-association of a model transmembrane helix, *Nat. Struct. Biol. 7*, 161-166.

139. Zhou, F. X., Cocco, M. J., Russ, W. P., Brünger, A. T., and Engelman, D. M. (2000) Interhelical hydrogen bonding drives strong interactions in membrane proteins, *Nat. Struct. Biol. 7*, 154-160.

140. Gray, T. M., and Matthews, B. W. (1984) Intrahelical hydrogen bonding of serine, threonine and cysteine residues within α-helices and its relevance to membrane-bound proteins, *J. Mol. Biol. 175*, 75-81.

141. Presta, L. G., and Rose, G. D. (1988) Helix signals in proteins, *Science 240*, 1632-1641.

142. Langosch, D., and Heringa, J. (1998) Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils, *Proteins: Structure, Function, and Genetics 31*, 150-159.

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

144. Morein, S., Killian, J. A., and Sperotto, M. M. (2002) Characterization of the thermotropic behavior and lateral organization of lipid-peptide mixtures by a combined experimental and theoretical approach: Effects of hydrophobic mismatch and role of flanking residues, *Biophys. J. 82*, 1405-1417.

145. Columbus, L., Lipfert, J., Jambunathan, K., Fox, D. A., Sim, A. Y. L., Doniach, S., and Lesley, S. A. (2009) Mixing and matching detergents for membrane protein NMR structure determination, *J. Am. Chem. Soc. 131*, 7320-7326.

146. Frank, S. J., Yi, W. S., Zhao, Y. M., Goldsmith, J. F., Gilliland, G., and Jiang, J. (1995) Regions of the JAK2 tyrosine kinase required for coupling to the growth hormone receptor, *J. Biol. Chem. 270*, 14776-14785.

147. Kim, M. J., Park, S. H., Opella, S. J., Marsilje, T. H., Michellys, P. Y., Seidel, H. M., and Tian, S. S. (2007) NMR structural studies of interactions of a small, nonpeptidyl Tpo mimic with the thrombopoietin receptor extracellular juxtamembrane and transmembrane domains, *J. Biol. Chem. 282*, 14253-14261.

148. Wishart, D. S., and Sykes, B. D. (1994) The C-13 Chemical-Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data, *J. Biomol. NMR 4*, 171-180.

149. Sulistijo, E. S., Jaszewski, T. M., and MacKenzie, K. R. (2003) Sequence-specific dimerization of the transmembrane domain of the "BH3-only" protein BNIP3 in membranes and detergent, *J. Biol. Chem. 278*, 51950-51956.

150. Melnyk, R. A., Kim, S., Curran, A. R., Engelman, D. M., Bowie, J. U., and Deber, C. M. (2004) The affinity of GXXXG motifs in transmembrane helix-helix interactions is modulated by long-range communication, *J. Biol. Chem. 279*, 16591-16597.

151. North, B., Cristian, L., Stowell, X. F., Lear, J. D., Saven, J. G., and DeGrado, W. F. (2006) Characterization of a membrane protein folding motif the ser zipper, using designed peptides, *J. Mol. Biol. 359*, 930-939.

152. Kalli, A. C., Campbell, I. D., and Sansom, M. S. P. (2011) Multiscale simulations suggest a mechanism for integrin inside-out activation, *Proc. Natl. Acad. Sci. U. S. A. 108*, 11890-11895.

153. Hall, B. A., Armitage, J. P., and Sansom, M. S. P. (2011) Transmembrane helix dynamics of bacterial chemoreceptors supports a piston model of signalling, *PLoS Comp. Biol. 7*, e1002204.

154. Moriki, T., Maruyama, H., and Maruyama, I. N. (2001) Activation of preformed EGF receptor dimers by ligand-induced rotation of the transmembrane domain, *J. Mol. Biol. 311*, 1011-1026.

155. Bell, C. A., Tynan, J. A., Hart, K. C., Meyer, A. N., Robertson, S. C., and Donoghue, D. J. (2000) Rotational coupling of the transmembrane and kinase domains of the Neu receptor tyrosine kinase, *Molecular Biology of the Cell 11*, 3589-3599.

156. Parat, M., Blanchet, J., and De Léan, A. (2010) Role of juxtamembrane and transmembrane domains in the mechanism of natriuretic peptide receptor A activation, *Biochemistry 49*, 4601-4610.

157. Sabath, D. F., Kaushansky, K., and Broudy, V. C. (1999) Deletion of the extracellular membrane-distal cytokine receptor homology module of Mpl results in constitutive cell growth and loss of thrombopoietin binding, *Blood 94*, 365-367.

158. Staerk, J., Lacout, C., Sato, T., Smith, S. O., Vainchenker, W., and Constantinescu, S. N. (2006) An amphipathic motif at the transmembrane-cytoplasmic junction prevents autonomous activation of the thrombopoietin receptor, *Blood 107*, 1864-1871.

159. Lopez-Ilasaca, M., Schiene, C., Kullertz, G., Tradler, T., Fischer, G., and Wetzker, R. (1998) Effects of FK506-binding protein 12 and FK506 on autophosphorylation of epidermal growth factor receptor, *J. Biol. Chem. 273*, 9430-9434.

160. Pastorino, L., Sun, A., Lu, P. J., Zhou, X. Z., Balastik, M., Finn, G., Wulf, G., Lim, J., Li, S. H., Li, X. J., Xia, W. M., Nicholson, L. K., and Lu, K. P. (2006) The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid-β production, *Nature 440*, 528-534.

161. Namanja, A. T., Peng, T., Zintsmaster, J. S., Elson, A. C., Shakour, M. G., and Peng, J. W. (2007) Substrate recognition reduces side-chain flexibility for conserved hydrophobic residues in human Pin1, *Structure 15*, 313-327.

162. O'Neal, K. D., Chari, M. V., McDonald, C. H., Cook, R. G., YuLee, L. Y., Morrisett, J. D., and Shearer, W. T. (1996) Multiple cis-trans conformers of the prolactin receptor proline-rich motif (PRM) peptide detected by reverse-phase HPLC, CD and NMR spectroscopy, *Biochem. J. 315*, 833-844.

163. Gothel, S. F., and Marahiel, M. A. (1999) Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts, *Cell. Mol. Life Sci. 55*, 423-436.

164. Funakoshi-Tago, M., Pelletier, S., Matsuda, T., Parganas, E., and Ihle, J. N. (2006) Receptor specific downregulation of cytokine signaling by autophosphorylation in the FERM domain of Jak2, *EMBO J. 25*, 4763-4772.

165. Putters, J., Almeida, A. C. D., van Kerkhof, P., van Rossum, A., Gracanin, A., and Strous, G. J. (2011) Jak2 Is a Negative Regulator of Ubiquitin-Dependent Endocytosis of the Growth Hormone Receptor, *PLoS One 6*, e14676.

166. Giordanetto, F., and Kroemer, R. T. (2002) Prediction of the structure of human Janus kinase 2 (JAK2) comprising JAK homology domains 1 through 7, *Protein Engineering 15*, 727-737.

167. Ceccarelli, D. F. J., Song, H. K., Poy, F., Schaller, M. D., and Eck, M. J. (2006) Crystal structure of the FERM domain of focal adhesion kinase, *J. Biol. Chem. 281*, 252-259.

168. Skiniotis, G., Boulanger, M. J., Garcia, K. C., and Walz, T. (2005) Signaling conformations of the tall cytokine receptor gp130 when in complex with IL-6 and IL-6 receptor, *Nat. Struct. Mol. Biol. 12*, 545-551.

169. Lupardus, P. J., Skiniotis, G., Rice, A. J., Thomas, C., Fischer, S., Walz, T., and Garcia, K. C. (2011) Structural snapshots of full-length Jak1, a transmembrane gp130/IL-6/IL-6Rα cytokine receptor complex, and the receptor-Jak1 holocomplex, *Structure 19*, 45-55.

170. Alexander, W. S., and Dunn, A. R. (1995) Structure and transcription of the genomic locus encoding murine c-mpl, a receptor for thrombopoietin *Oncogene 10*, 795-803.

171. Desauvage, F. J., Hass, P. E., Spencer, S. D., Malloy, B. E., Gurney, A. L., Spencer, S. A., Darbonne, W. C., Henzel, W. J., Wong, S. C., Kuang, W. J., Oles, K. J., Hultgren, B., Solberg, L. A., Goeddel, D. V., and Eaton, D. L. (1994) Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-mpl ligand, *Nature 369*, 533-538.

172. Onishi, M., Mui, A. L., Morikawa, Y., Cho, L., Kinoshita, S., Nolan, G. P., Gorman, D. M., Miyajima, A., and Kitamura, T. (1996) Identification of an oncogenic form of the thrombopoietin receptor MPL using retrovirus-mediated gene transfer, *Blood 88*, 1399-1406.

173. Itaya, M., Brett, I. C., and Smith, S. O. (2012) Synthesis, purification, and characterization of single helix membrane peptides and proteins for NMR spectroscopy, *Methods Mol. Biol. 831*, 333-357.

174. Zhou, F. X., Merianos, H. J., Brünger, A. T., and Engelman, D. M. (2001) Polar residues drive association of polyleucine transmembrane helices, *Proc. Natl. Acad. Sci. U. S. A. 98*, 2250-2255.

175. Itaya, M. (2012) Regulation of Dimerization and Activation of the Thrombopoietin Receptor, In *Department of Biochemistry and Cell Biology*, Stony Brook University, Stony Brook, NY.

176. Parthasarathy, K., Lin, X., Tan, S. M., Law, S. K. A., and Torres, J. (2008) Transmembrane helices that form two opposite homodimeric interactions: An asparagine scan study of αM and β2 integrins, *Protein Sci. 17*, 930-938.

177. 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.

178. Dagil, R., Knudsen, M. J., Olsen, J. G., O'Shea, C., Franzmann, M., Goffin, V., Teilum, K., Breinholt, J., and Kragelund, B. B. (2012) The WSXWS Motif in Cytokine Receptors Is a Molecular Switch Involved in Receptor Activation: Insight from Structures of the Prolactin Receptor, *Structure 20*, 270-282.

179. Mineev, K. S., Khabibullina, N. F., Lyukmanova, E. N., Dolgikh, D. A., Kirpichnikov, M. P., and Arseniev, A. S. (2011) Spatial structure and dimer-monomer equilibrium of the ErbB3 transmembrane domain in DPC micelles, *Biochim. Biophys. Acta 1808*, 2081-2088.

180. Cammett, T. J., Jun, S. J., Cohen, E. B., Barrera, F. N., Engelman, D. M., and DiMaio, D. (2010) Construction and genetic selection of small transmembrane proteins that activate the human erythropoietin receptor, *Proc. Natl. Acad. Sci. U. S. A. 107*, 3447-3452.

181. Naranda, T., Kaufman, R. I., Li, J., Wong, K., Boge, A., Hallen, D., Fung, K. Y. C., Duncan, M. W., Andersen, N., Goldstein, A., and Olsson, L. (2002) Activation of erythropoietin receptor through a novel extracellular binding site, *Endocrinology 143*, 2293-2302.

182. Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wuthrich, K. (2002) NMR analysis of a 900K GroEL-GroES complex, *Nature 418*, 207-211.

183. Staerk, J., Defour, J.-P., Pecquet, C., Leroy, E., Antoine-Poirel, H., Brett, I., Itaya, M., Smith, S. O., Vainchenker, W., and Constantinescu, S. N. (2011) Orientation-specific signalling by thrombopoietin receptor dimers, *EMBO J. 30*, 4398-4413.

184. Morita, H., Tahara, T., Matsumoto, A., Kato, T., Miyazaki, H., and Ohashi, H. (1996) Functional analysis of the cytoplasmic domain of the human Mpl receptor for tyrosine-phosphorylation of the signaling molecules, proliferation and differentiation, *FEBS Lett. 395*, 228-234.

# Appendix

## Chemical Shift Tables

### *A.1 muEpoR$_{220-248}$ chemical shifts in 10 mM sodium phosphate, pH 7.0, 200 mM d38-DPC, 10% $D_2O$ (v:v)*

| | Seq | C=O | Ca | Cb | Ha | N |
|---|---|---|---|---|---|---|
| | S | | | | | |
| | N | | 53.163 | 39.314 | 4.793 | |
| 220 | A | 177.651 | 53.586 | 18.923 | 4.202 | |
| 221 | S | 174.347 | 59.054 | 63.661 | 4.317 | 112.08236 |
| 222 | D | 176 | 54.648 | 41.39 | 4.623 | 121.31871 |
| 223 | L | 176.053 | 54.522 | 42.967 | 4.297 | 120.5382 |
| 224 | D | | 53.491 | 42.3 | | 123.5524 |
| 225 | P | 178.879 | 65.405 | 32.287 | 4.231 | |
| 226 | L | 178.314 | 58.395 | 41.467 | 4.117 | 121.08517 |
| 227 | I | 179.21 | 64.227 | 36.932 | 3.675 | 118.829 |
| 228 | L | 177.787 | 58.613 | 41.963 | 3.978 | 121.42733 |
| 229 | T | 175.918 | 68.434 | 68.083 | 3.699 | 115.23447 |
| 230 | L | 178.261 | 58.249 | 41.798 | 3.942 | 119.39 |
| 231 | S | 175.117 | 63.708 | 62.891 | 3.996 | 114.49025 |
| 232 | L | 178.507 | 58.205 | 41.605 | 3.942 | 121.10517 |
| 233 | I | 177.269 | 65.608 | 37.282 | 3.554 | 118.28963 |
| 234 | L | 179.929 | 58.436 | 41.393 | 3.974 | 119.15964 |
| 235 | V | 177.227 | 67.332 | 31.212 | 3.468 | 120.9879 |
| 236 | L | 178.717 | 58.895 | 41.688 | 3.943 | 119.94409 |
| 237 | I | 177.451 | 65.476 | 37.565 | 3.584 | 117.21661 |
| 238 | S | 176.856 | 63.753 | 62.688 | 4.019 | 116.59416 |
| 239 | L | 178.073 | 58.297 | 41.932 | 3.989 | 124.19183 |
| 240 | L | 178.503 | 58.531 | 41.729 | 3.924 | 119.42783 |
| 241 | L | 178.448 | 58.18 | 41.77 | 3.96 | 116.863 |
| 242 | T | 176.177 | 68.33 | 67.747 | 3.727 | 116.1027 |
| 243 | V | 177.602 | 67.362 | 31.25 | 3.503 | 120.07525 |
| 244 | L | 179.241 | 58.233 | 41.674 | 3.942 | 117.81175 |
| 245 | A | 179.747 | 54.891 | 18.6 | 4.045 | 120.20555 |
| 246 | L | 178.162 | 56.602 | 42.762 | 4.238 | 116.33357 |
| 247 | L | 176.206 | 55.018 | 42.795 | 4.4 | 117.61337 |
| 248 | S | | 61.068 | 65.056 | 4.187 | 119.84088 |

### *A.2 muEpoR$_{220-248}$ chemical shifts in 90% trifluoroethanol, 10% deuterated chloroform (v:v)*

| | C | Ca | Cb | H | N |
|---|---|---|---|---|---|
| N219 | 172.185 | 50.546 | 36.334 | 8.658 | 116.43 |
| A220 | 175.321 | 51.914 | 15.324 | 8.075 | 122.801 |
| S221 | | 57.875 | 61.313 | 8.393 | 111.785 |
| D222 | 172.61 | 51.784 | 35.041 | 8.233 | 118.254 |

| | | | | | |
|---|---|---|---|---|---|
| L223 | 174.829 | 53.174 | 40.241 | 8.049 | 119.041 |
| D224 | | 53.117 | 33.927 | 8.146 | 118.323 |
| P225 | 176.219 | 63.836 | 28.797 | | |
| L226 | 175.574 | 56.137 | 39.406 | 7.761 | 119.359 |
| I227 | 177.117 | 62.278 | 35.276 | 8.036 | 118.958 |
| L228 | 176.453 | 56.348 | 39.59 | 8.711 | 121.774 |
| T229 | 173.234 | 65.532 | 67.009 | 8.199 | 114.9 |
| L230 | 176.541 | 56.1 | 39.476 | 8.826 | 120.732 |
| S231 | 172.499 | 60.367 | | 8.491 | 113.326 |
| L232 | | 55.874 | 39.501 | 8.21 | 122.352 |
| I233 | 175.147 | 63.018 | 35.41 | 8.453 | 119.769 |
| L234 | 178.532 | 56.172 | 39.139 | 8.512 | 118.811 |
| V235 | 175.641 | 65.15 | 29.186 | 8.469 | 122.012 |
| L236 | 177.528 | 56.512 | 39.255 | 8.69 | 121.5 |
| I237 | 175.513 | 62.904 | 35.423 | 9.298 | 118.944 |
| S238 | 174.383 | 60.686 | | 8.477 | 115.886 |
| L239 | 176.21 | 56.339 | 39.54 | 8.853 | 126.305 |
| L240 | 176.991 | 56.392 | 39.376 | 8.722 | 121.366 |
| L241 | 177.438 | 55.989 | 39.386 | 9.121 | 117.1 |
| T242 | 173.249 | 65.2 | 66.698 | 8.254 | 116.163 |
| V243 | 175.725 | 65.321 | 29.247 | 8.78 | 122.551 |
| L244 | 178.038 | 56.134 | 39.339 | 8.78 | 118.664 |
| A245 | 178.136 | 53.163 | 14.965 | 8.594 | 123.071 |
| L246 | 177.699 | 55.788 | 39.854 | 8.72 | 119.949 |
| L247 | 176.107 | 54.243 | 40.153 | 8.873 | 118.144 |
| S248 | | 56.125 | 61.926 | 8.199 | 113.901 |

### *A.3 Wild type muEpoR$_{218-268}$ chemical shifts in 10 mM sodium phosphate, pH 7.0, 200 mM d38-DPC, 10% D$_2$O (v:v)*

Wild Type

| | | C | Ca | Cb | H | N |
|---|---|---|---|---|---|---|
| | S | | | | | |
| | N | 174.64904 | 53.51418 | 39.50878 | | |
| | A | 177.39068 | 53.00189 | 19.75242 | 8.36448 | 123.36188 |
| 218 | L | 176.76146 | 55.65966 | 42.75389 | 8.27975 | 119.51994 |
| 219 | T | 174.49302 | 61.00128 | 70.21197 | 8.12354 | 112.39972 |
| 220 | A | 177.8703 | 53.55676 | 19.2455 | 8.49114 | 124.62308 |
| 221 | S | 174.27282 | 59.26771 | 63.72775 | 7.97884 | 112.44594 |
| 222 | D | 175.82701 | 54.8776 | 41.52073 | 8.08274 | 120.65205 |
| 223 | L | 176.07547 | 54.58111 | 43.08118 | 7.72845 | 120.09381 |
| 224 | D | | 54.42775 | 43.03436 | 8.189 | 123.34001 |
| 225 | P | | | | | |
| 226 | L | 178.29115 | 58.58417 | 41.52057 | | |
| 227 | I | 179.25282 | 64.25483 | 36.98435 | 7.84571 | 119.01423 |
| 228 | L | 177.80552 | 58.84924 | 41.93387 | 8.21059 | 121.60099 |
| 229 | T | 175.95305 | 68.0918 | | 7.94914 | 115.36954 |
| 230 | L | 178.29815 | 58.51818 | 41.88574 | 8.31024 | 119.41597 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 231 | S | 175.15035 | 63.53921 | 62.76026 | 8.17696 | 114.61253 |
| 232 | L | 178.55808 | 58.30516 | 41.67067 | 7.91081 | 121.13175 |
| 233 | I | 177.30515 | 65.517 | 37.35079 | 8.04518 | 118.3965 |
| 234 | L | 179.97502 | 58.49818 | 41.49307 | 8.24393 | 119.21893 |
| 235 | V | 177.25092 | 67.10518 | 31.29531 | 8.31149 | 121.11908 |
| 236 | L | 178.7409 | 58.82468 | 41.67313 | 8.34538 | 119.97857 |
| 237 | I | 177.48313 | 65.28394 | 37.73762 | 8.66729 | 117.3096 |
| 238 | S | 176.86862 | 63.63482 | 62.63005 | 8.14779 | 116.71553 |
| 239 | L | 178.08519 | 58.60595 | 41.99451 | 8.43321 | 124.34448 |
| 240 | L | 178.57424 | 58.8893 | 41.88437 | 8.2851 | 119.48266 |
| 241 | L | 178.36977 | 58.43049 | 41.71903 | 8.56275 | 117.0672 |
| 242 | T | 176.30861 | 68.21235 | | 7.85895 | 116.02906 |
| 243 | V | 177.63102 | 67.39633 | 31.215 | 8.23644 | 120.37464 |
| 244 | L | 179.37623 | 58.57203 | 41.53489 | 8.31669 | 118.70518 |
| 245 | A | 179.82444 | 55.64047 | 18.53608 | 8.56764 | 122.22491 |
| 246 | L | 178.89799 | 58.23828 | 42.27027 | 8.36416 | 118.32277 |
| 247 | L | 179.52534 | 57.5618 | 42.15666 | 8.41682 | 117.03651 |
| 248 | S | 175.31898 | 60.95305 | 63.40101 | 8.09369 | 113.59026 |
| 249 | H | | 57.14116 | 30.74674 | 7.91674 | 119.27719 |
| 250 | R | | | | 7.72036 | 120.13266 |
| 251 | R | 178.09474 | 59.20649 | 29.44758 | | |
| 252 | T | 176.35464 | 65.4327 | 68.5985 | 7.78609 | 114.32036 |
| 253 | L | 178.3113 | 57.51019 | 42.18068 | 7.91977 | 120.51861 |
| 254 | Q | 177.31 | 59.46324 | 28.80213 | 8.46463 | 118.15121 |
| 255 | Q | 176.71167 | 57.34407 | 28.98848 | 7.76434 | 115.92335 |
| 256 | K | 176.75494 | 57.0598 | 33.02405 | 7.60337 | 116.54817 |
| 257 | I | 174.73794 | 62.52286 | 38.75947 | 7.55505 | 116.74118 |
| 258 | W | | 55.49808 | 29.03703 | 7.98595 | 118.84649 |
| 259 | P | 177.92646 | 63.41782 | 32.10553 | | |
| 260 | G | 173.84061 | 45.17598 | | 8.21028 | 107.78397 |
| 261 | I | | 58.64655 | 38.58278 | 7.80287 | 121.40432 |
| 262 | P | 176.46355 | 63.24791 | 32.12214 | | |
| 263 | S | | 56.18249 | 63.54117 | 8.29015 | 117.18823 |
| 264 | P | 177.03438 | 63.52479 | 32.0578 | | |
| 265 | E | 176.45209 | 56.99261 | 30.1038 | 8.41722 | 119.77598 |
| 266 | S | 174.06548 | 58.32796 | 63.97491 | 8.00553 | 115.44215 |
| 267 | E | 175.0503 | 56.85047 | 30.60347 | 8.19624 | 122.80193 |
| 268 | F | | 58.83643 | 40.52133 | 7.5768 | 124.65485 |

***A.4 L223C muEpoR$_{218-268}$ chemical shifts in 10 mM Sodium phosphate, pH 7.0, 200 mM d38-DPC, 10% D$_2$O (v:v)***

| L223C | | | | | | |
|---|---|---|---|---|---|---|
| | | C | Ca | Cb | H | N |
| | SNA | | | | | |
| 218 | L | 176.821 | 55.732 | 42.386 | 8.277 | 119.142 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 219 | T | 174.384 | 61.141 | 70.1395 | 7.961 | 111.8895 |
| 220 | A | 177.838 | 53.08433 | 19.211 | 8.38925 | 124.86525 |
| 221 | S | 174.237 | 59.39933 | 63.88067 | 8.05525 | 113.06175 |
| 222 | D | 175.7 | 54.51467 | 41.23033 | 8.14813 | 120.69475 |
| 223 | C | 173.546 | 58.552 | 28.66067 | 7.854 | 117.7355 |
| 224 | D | | 53.121 | 42.399 | 8.55413 | 125.88562 |
| 225 | P | 178.82 | 65.383 | 32.35 | | |
| 226 | L | 178.272 | 58.40767 | 41.38433 | 8.584 | 121.447 |
| 227 | I | 179.136 | 64.13933 | 36.904 | 7.77787 | 118.83925 |
| 228 | L | 177.829 | 58.59967 | 41.95367 | 8.1305 | 121.71637 |
| 229 | T | 175.934 | 68.23767 | | 7.91743 | 115.20071 |
| 230 | L | 178.294 | 58.097 | 41.77667 | 8.34033 | 119.38467 |
| 231 | S | 175.132 | 63.894 | 62.99533 | 8.17475 | 114.58988 |
| 232 | L | 178.552 | 58.244 | 41.555 | 7.914 | 121.06043 |
| 233 | I | 177.328 | 65.668 | 37.32067 | 8.05513 | 118.35988 |
| 234 | L | 179.954 | 58.45467 | 41.2975 | 8.25713 | 119.238 |
| 235 | V | 177.227 | 67.44333 | 31.158 | 8.3085 | 121.05838 |
| 236 | L | 178.728 | 58.86033 | 41.65033 | 8.34237 | 120.03037 |
| 237 | I | 177.474 | 65.42567 | 37.42867 | 8.68112 | 117.27037 |
| 238 | S | 176.87 | 63.805 | 62.571 | 8.15888 | 116.71625 |
| 239 | L | 178.08 | 58.241 | 41.9075 | 8.44688 | 124.30113 |
| 240 | L | 178.587 | 58.536 | 41.678 | 8.299 | 119.4355 |
| 241 | L | 178.376 | 58.2865 | 41.7 | 8.57817 | 117.08717 |
| 242 | T | 176.326 | 67.791 | 67.818 | 7.867 | 116.12637 |
| 243 | V | 177.636 | 67.576 | 31.167 | 8.2554 | 120.3908 |
| 244 | L | 179.388 | 58.47233 | 41.56333 | 8.32175 | 118.6065 |
| 245 | A | 179.798 | 55.603 | 18.38233 | 8.57387 | 122.08013 |
| 246 | L | 178.843 | 58.041 | 42.091 | 8.36925 | 118.25538 |
| 247 | L | 179.33 | 57.41333 | 42.0985 | 8.38917 | 117.09367 |
| 248 | S | 175.255 | 61.134 | 63.771 | 8.06267 | 113.39817 |
| 249 | H | 176.05 | 57.081 | 30.734 | 7.973 | 119.339 |
| 250 | R | | | | 7.792 | 120.218 |
| 251 | R | 178.058 | 58.986 | 29.38193 | | |
| 252 | T | 176.33 | 65.131 | 68.77733 | 7.78555 | 114.31694 |
| 253 | L | 178.313 | 57.5965 | 42.023 | 7.91443 | 120.52314 |
| 254 | Q | 177.311 | 59.517 | 28.741 | 8.46525 | 118.20675 |
| 255 | Q | 176.694 | 57.446 | 28.93833 | 7.78586 | 115.93529 |
| 256 | K | 176.715 | 56.912 | 32.663 | 7.61057 | 116.58271 |
| 257 | I | 174.714 | 62.66167 | 38.799 | 7.54887 | 116.6725 |
| 258 | W | | 55.546 | 28.988 | 7.99063 | 118.80862 |
| 259 | P | 177.904 | 63.788 | 32.013 | | |
| 260 | G | 173.843 | 45.1335 | | 8.208 | 107.74725 |
| 261 | I | | 58.672 | 38.576 | 7.803 | 121.37267 |
| 262 | P | 176.45 | 63.0305 | 32.0975 | | 117.15088 |
| 263 | S | | 56.293 | 63.728 | 8.29262 | |
| 264 | P | 177.024 | 63.7025 | 32.017 | | |
| 265 | E | 176.451 | 56.96433 | 29.99633 | 8.42513 | 119.76538 |
| 266 | S | 174.063 | 58.30467 | 64.121 | 8.00975 | 115.43213 |
| 267 | E | 175.054 | 56.802 | 30.54633 | 8.201 | 122.782 |

| | | | | | |
|---|---|---|---|---|---|
| 268 | F | 58.927 | 40.463 | 7.579 | 124.65387 |

**Structure Parameter Table**

***A.5 Murine EpoR TM dimer structure parameters***

| **Experimental Restraints per Monomer** | |
|---|---|
| # of NOE restraints | |
| Intraresidue | 241 |
| Inter-residue | |
| sequential | 105 |
| medium-range | 92 |
| long-range | 7 |
| unambiguous intermonomer | 66 |
| # of Hydrogen Bond Restraints | 15 |
| # of Torsion Angle Restraints | 52 |
| Backbone Φ | 26 |
| Backbone Ψ | 26 |

| **Structure Statistics** | |
|---|---|
| Restraint Violations | |
| Distance (>0.5 Å) | 2.53 |
| Distance (>0.3 Å) | 2.42 |
| Dihedrals (>5°) | 13 |
| Deviation from idealized geometry | |
| Bonds (Å) | 0.0085 ±0.0005 |
| Angles (deg) | 1.141 ±0.096 |
| Impropers (deg) | 4.23 ±0.569 |
| Deviation from experimental restraints | |
| NOEs | 24 |
| Dihedrals (deg) | 7.39 |
| RMSD | |
| Backbone atoms | 2.05 |
| All Heavy Atoms | 2.64 |
| Ramachandran Analysis | |
| % Most Favored | 72.1 |
| % Allowed | 24.8 |
| % Generously Allowed | 2.34 |
| % Disallowed | 0.72 |