PURIFICATION AND MASS SPECTROMETRIC CHARACTERIZATION OF HUMAN CXCR4

by

JULIE PATRICIA WONG

B.Sc., The University of British Columbia, 2001

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Biochemistry and Molecular Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2008

Abstract

G protein- coupled receptors (GPCRs) are seven transmembrane receptors that comprise the largest superfamily of proteins in the body that are involved in a variety of fundamental processes including sight, smell, tactile as well as nervous responses. Tightly coordinated, control of these receptors is critical for normal physiology. Emphasis has been placed in pharmaceuticals to find ways to inhibit or accelerate these processes in GPCR implicated diseases. Although tertiary structural information would be beneficial for rational drug design, little is known about the structure of GPCRs despite extensive research in the field. Only two high resolution structures exist for any mammalian GPCR. Structural studies are challenged by the intrinsic difficulty associated with purifying these receptors in high quality and quantity. To overcome these challenges, we propose a mass spectrometric, sequence-based approach to characterize GPCRs allowing for the rapid characterization of an ectopically as well as endogenously expressed receptor employing a receptor tag. The sensitivity of the approach allows for the implementation of a mammalian expression system with the physiologically relevant post-translational modifications (PTMs). The sites of N-linked glycosylation has been mapped and receptor isoforms were identified for a model GPCR, CXCR4 and a related receptor, CCR5. This approach is versatile and applicable to other membrane proteins such as ATP-binding cassette (ABC) transporters and has been adapted into an antibody screening platform. As proof- of- principle, we have developed a monoclonal anti-human CXCR4 antibody with broad applications and potential for clinical diagnostics.

In view of the challenges preventing high resolution three-dimensional structure determination, photoaffinity crosslinking was combined with our sequence-based strategy for receptor binding site footprinting. CXCR4 ligand analogs containing a non-natural amino acid photocrosslinker, benzophenylalanine and a biotin tag for complex isolation were chemically synthesized and crosslinked to CXCR4 expressed on intact cells. Individual components of the receptor-ligand complex were identified by western blotting and tandem mass spectrometry. Regions on the bound receptor that were protected from protease treatment and consequently tandem MS/MS sequencing were identified as potential sites of ligand-receptor contact. These regions were identified as the receptor N-terminus and/or the first extracellular loop for ligand binding/docking.

Table of contents

List of Tables

List of Figures

List of symbols and abbreviations

ABC	ATP binding cassette
AIDS	Acquired immune deficiency syndrome
CAN	Acetonitrile
BPA	Benzoylphenylalanine
CID	Collision induced dissociation
conA	Concavalin A
DAPI	4`,6-diamidino-2-phenylindole
DMEM	Dubelcco`s modified eagle medium
EDTA	Ethylene diamine tetra acetic acid
ESI	Electrospray ionization
ELISA	Enzyme-linked immunosorbent assay
FA	Formic acid
Fmoc	9H-Fluoren-9-ylmethoxycarbonyl
FT-ICR	Fourier transform ion cyclotron resonance
GPCR	G protein-coupled receptor
GST	Glutathione S-transferase
GTP/GDP	Guanidine triphosphate/ guanidine diphosphate
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	Horse radish peroxidase
ICL, ECL	Intracellular loop, extracellular loop
IPTG	Isopropyl β–D-1-thiogalactopyranoside
LC	Liquid chromatography
LTQ	Linear ion trap
mAb	Monoclonal antibodies
MALDI	Matrix assisted laser desorption ionization
MMP	Matrix metalloproteinase
MRM	Multiple reaction monitoring
MS	Mass spectrometry
NCBI	National Centre for Biotechnology Information
NMR	Nuclear magnetic resonance
PBMC	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PTM	Post-translational modification
PVDF	Polyvinylidene fluoride
ROS	Rod outer segments
SDF-1	Stromal cell derived factor-1
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
svFc	Single-chain variable peptide fragments
t-BOC	Tert-butyloxycarbonyl
TBS	Tris buffered saline
TOF	Time-of-flight
UV	Ultraviolet

Acknowledgements and personal statement

I would like to thank all the past and present members of the Kast, Molday and former Clark-Lewis lab for support for if not, this work would not have been possible. I am really glad to have had the chance to work with such spectacular people. With the termination of the Clark-Lewis lab, all the personnel in Molday and the Kast labs really made me feel welcomed and a part of the group.

I feel very fortunate and grateful to have had the guidance of Drs. Robert S. Molday and Juergen Kast for the following years of my graduate studies. I really appreciate the time they took to help me refocus my studies and for supporting my decision to continue working on a risky project that was not a mainstay to either of their labs. Without either of their expertise I would not have been able to purify and characterize CXCR4.

I would like to extend my thanks to my advisory committee, Dr. Natalie Strynadka for providing me with other perspectives and suggestions to approaching my thesis aims and for all her support through the years.

A special dedication is made in memory of Dr. Ian Clark-Lewis my first supervisor who has instilled in me a curiosity for chemokine and peptide synthesis research. He was extremely generous with sharing his knowledge and reagents to the scientific community which helped advance research in the field. I admire his integrity and authenticity to his work and strive to emulate these qualities.

I would also like to acknowledge Dr. Jiang-Hong Gong, a former member of the Clark-Lewis lab for introducing and teaching me the basics of chemokine and immunology research. Her passion and dedication to the field is an inspiration.

Finally last but not least, I would like to thank all my family and friends (old, new and the ones I have gained during graduate work) for love and support. I would not have been able to do this without any of them.

There have been both happy and sad times but this has been a meaningful journey that I am very grateful to have had.

Statement of co-authorship:

Chapters II-IV are versions of manuscripts in preparation for submission and publication. I have designed and performed all of the experiments described, except for Figure 2-3 analysis of ATPase activity for ABCA4, which was performed and graphed by Dr. Emanuelle Reboul, post-doctoral fellow in the Molday Lab. I have written all three manuscripts presented in this dissertation.

Chapter I: Introduction

1.1 G protein-coupled receptors overview

G protein-coupled receptors (GPCRs) represent 3-4% of the human genome [2] and make up a significant proportion of cellular proteins. These receptors have seven transmembrane segments that are connected by extracellular and intracellular loops of varying lengths; the N-terminal segment extends into the extracellular space and the C-terminus is intracellular (Figure 1-1). GPCRs relay extracellular signals which may be in the form of photon or ligand to trigger intracellular heterotrimeric G-protein signalling. Upon receptor activation guanidinyl diphosphate (GDP) which is bound to the G-protein is exchanged for guanidinyl triphosphate (GTP) causing subunit dissociation [1]. The β and γ subunits are lipid anchored while the α subunit is released. Subsequent α subunit inactivation is caused by the hydrolysis of GTP to GDP causes the heterotrimetric G protein to reform (Figure 1-2) [1].

Mammalian G protein-coupled receptors are classified into 6 groups: A) rhodopsin – type, B) secretin/glucagon, C) metabotropic glutamate, D) fungal pheromone, E) cAMP and F) frizzled/smoothened receptors [3, 4]. Group A, the rhodopsin-type receptors represent the largest group in the superfamily (over 80% of all GPCRs) which account for many classical receptors for neurotransmitters, biogenic amines, nucleotides, prostaglandins and chemokines [3, 5]. Over 30 human diseases are associated with mutations in GPCR-coded genes which include retinis pigmentosa [5], nephrogenic diabetes insipidus [7, 8], hypo- and hyperthyroidism [5] as well as immune deficiencies [9]. Due to their importance in physiology, GPCRs are commonly targeted for drug discovery. GPCRs comprise approximately 60% of pharmaceutical targets [10] and 25% of the top selling drugs on the market are targeted to GPCRs [5]. The GPCR, CXCR4

is of particular interest due to implications in a variety of diseases and will be the primary focus in the following dissertation.

Drug development for GPCRs face challenges such as antagonist cross-reactivity with other GPCRs [11] and binding affinity discrepancies in mouse and human models [12]. To overcome these issues, high resolution structures for the GPCRs would be of great value for defining specificities for rational drug design; however, not much is known at this time due to inherent difficulties expressing and purifying these receptors. This challenge is reflected in the lack of information available for this very important group of proteins; only 2 out of almost 350 functional GPCRs (excluding olfactory receptors) have solved structures, 200 of these receptors have known ligands and 150 are orphan receptors [13]. Of the mammalian GPCRs, only bovine rhodopsin [14] and the human β_2-adrenrenergic receptor structure has been solved, since the discovery and characterization of the GPCR superfamily in the late 1980's [15, 16]. Bovine rhodopsin was solved by two collaborating research groups in 2000 and since then the structures of several related bacterial forms have also been established [14, 17, 18]. The structure of an engineered form of β_2-adrenergic receptor was recently determined which took almost 20 years to complete [15, 19, 20].

Although rhodopsin is virtually pure in bovine rod photoreceptor outer segments providing an abundant source for high resolution analysis, the dynamic nature of the receptor complicates uniform crystal formation [21]. Receptor stabilization for both rhodopsin and β_2-adrenergic receptor was the breakthrough that enabled structural determination. Palczewski's group added Zn^{2+} to rhodopsin [14] and the dynamic third intracellular loop in the β_2-adrenergic receptor was replaced with a more rigid lysozyme domain [19]. The rationale for the

stabilization was determined from previous work on the receptors employing other complementary biochemical methods such as mutagenesis and mapping studies. For example, the third intracellular loop in the β_2-adrenergic receptor was identified and characterized from earlier mutational studies by the same group that subsequently solved the structure [22-25].

Likewise, high resolution structures need to be consistent with the complementary structure-function information obtained from alternative methods. The rhodopsin structure was confirmed with available biochemical information collected from alternative approaches [26, 27] linking structure with function providing a more complete model template for other GPCRs.

1.1.1 Rhodopsin, a model GPCR

The first GPCR high resolution structure solved at 2.8 Å resolution was bovine rhodopsin[14]. This breakthrough greatly impacted the field as no other information was available at the time. Rhodopsin has become a structural template for understanding other GPCRs and several groups have extrapolated information from the rhodopsin prototype to validate other GPCR models. However, rhodopsin may not be the most suitable template due to functional differences as compared to ligand binding GPCRs.

Rhodopsin functions as a bimodal switch, and does not have a ligand binding site and the retinal cofactor binding site is buried unlike many other GPCRs. Moreover, rhodopsin shares low sequence identity with most GPCRs, for example 14% for serotonin receptor [28] and approximately 20% for CXCR4 as determined from protein blast 2 sequence alignment (www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). Despite these differences rhodopsin is still widely used for homology modelling due to the lack of other models, especially prior to the recent β_2-adrenergic receptor structure determination [20]. Homology modelling combined with other biochemical assays are common alternatives to understanding GPCR structure-function

relationships. Consequently this knowledge would facilitate high resolution analysis for this family of receptors.

1.1.2 GPCR activation inferences from rhodopsin as a model

Rhodopsin functions through the trimeric G protein, transducin which dissociates when a photon of light is absorbed. The high resolution structure solved for rhodopsin represents the dark adapted resting receptor in the absence of light. A light activated structure for the receptor has yet to be solved and what is currently known about receptor activation has been determined from mutational analysis [30, 31, 32, 33].

Activation upon ligand binding or photon absorption causes movements in the helical domains which lead to GTP binding and trimeric protein dissociation resulting in signal transduction. Electrostatic interactions and hydrogen bonding are the main factors that restrain as well as alter helical positions for rhodopsin [29]. In particular, electrostatic interactions between specific residues in the (E/D)RY motif and other amino acids on the extracellular surface dictate the straightening of helix VI and consequently the separation of helices III and VI which is referred to as the "Protonation Model"[30]. In this model, the (E/D)RY motif in transmembrane domain 3, bordering on intracellular loop 2, is constrained [30]. When the electrostatic interactions are released, as observed in mutational analysis which neutralizes the charge on the acidic residue in the motif, a constitutively active mutant is formed. Other GPCRs that follow this model include the opioid [31], alpha acetycholine [32, 33] and the beta adrenergic receptor [34, 35].

While this model holds true for many class A GPCRs, CXCR4 is an exception to the rule. Neutralization of the charge in the DRY motif results in partial loss of function but does not affect G protein coupling for CXCR4 [36].

1.1.3 Expression systems

In response to the limited expression levels of GPCRs from endogenous sources for elegant structural analysis, heterologous overexpression systems such as bacculovirus, *E. coli* and yeast have been implemented to study mammalian receptors. Portions of or the entire protein, are generally selected for expression. Heterologous expression systems offer several advantages, such as high protein yield and short generation times. Accurate folding, post-translational modification and solubility are complicating factors in these systems since they are dictated by the expression and trafficking machinery available in the host cell.

1.1.3.1 Accurate protein folding

Aggregated or misfolded proteins in inclusion bodies are sometimes observed in *E. coli* expression systems [37, 38]. However, misfolded proteins in inclusion bodies may be recovered and refolded under *in vitro* conditions; a technique which was first applied to overexpressed olfactory receptors [39] and more recently to the leukotriene B4 [40] and the serotonin HT_4-hydroxytryptamine receptors [41]. Refolding conditions were determined by systematic testing for each receptor. To improve solubility, a fusion protein was cloned onto the HT_4-hydroxytryptamine receptor which was subsequently cleaved off [41].

1.1.3.2 Post-translational modifications

Heterologous expression systems may not have the enzymes required to add on physiologically relevant post-translational modifications such as complex glycosylation, phosphorylation and lipidation which are required for normal function. Membranes in heterologous systems have different lipid environments for recombinant receptors, affecting expression and localization as well as extraction conditions. Unlike bacteria, yeast express endogenous GPCRs and thus have the enzymes required for eukaryotic post-translational modifications (PTMs) but there are some differences between yeast and mammalian PTMs. For

example yeast has ergosterol instead of cholesterol and different N-glycans from mammalian cells. Accurate post-translational modification is essential for proper expression, structure and function [42, 43].

In addition to PTMs, the host expression system needs to provide the physiologically relevant interacting proteins, such as chaperones or scaffold proteins for proper GPCR structure and function. Thus, heterologous systems may not provide a suitable environment for expression of mammalian GPCRs. Although mammalian expression systems offer lower protein yields, they are a more accurate source for the study of mammalian GPCRs.

1.1.3.3 Solubility

In addition to protein expression issues, protein extraction from the lipid membrane to more experimentally feasible context limit further analysis of this group of receptors. Selecting suitable receptor solubilising reagents such as detergent and chaotropes is non- trivial and differs from protein to protein. The type of membrane in which the GPCR is expressed also affects detergent choice. Moreover, it has also been reported that functional solubilisation of certain GPCRs requires ligand binding for stabilization [44].

1.1.4 Purification

GPCR detergent solubilisation may be incompatible with subsequent purification steps such as reverse phase chromatography and affinity purification that often requires three-dimensional shape recognition. Chromatographic methods targeted to protein-specific properties, immunoaffinity as well as affinity tags for the protein of interest are common enrichment strategies for GPCRs. Due to the challenging nature of GPCRs, accurate antigen expression as well as solubility issues influence antibody production and subsequent validation.

Many antibodies raised against GPCRs have been reported to have limited applicability to the protein of interest due to lack of reagent validation and unaccounted receptor heterogeneity [45].

Detection of CXCR4, an important GPCR will be discussed in more detail in the following. CXCR4 is often underestimated in samples with currently available antibodies. A combination of different anti-CXCR4 antibodies recognizing different epitopes on the receptor are necessary for complete and accurate detection due to antigenic heterogeneity [46]. Antibodies recognizing constant linear epitopes for GPCRs are highly desirable for receptor localization and quantification, one such antibody is the monoclonal antibody 1D4 directed to the C-terminus of bovine rhodopsin [47].

To overcome reagent and antibody limitations, general strategies such as affinity tags and immobilized ligands for the specific GPCR are often employed. However, trial and error testing is required to determine whether the tag affects receptor expression. If affinity tags are applicable, they are simple and economical alternatives for purification.

1.2 Chemokine receptors, an important group of rhodopsin family GPCRs and their cognate chemokine ligands

Chemokines are small structurally related secreted proteins that regulate homing and activation of leukocytes as well as non-immune cells such as stem cells through GPCRs. There are approximately 50 known human chemokines to date and approximately 20 cognate receptors (Table A-1). The disparity between the number of ligands and receptors reflects overlapping and promiscuous binding (Table A-1). Immunological functions that chemokines mediate include inflammatory responses, recruitment of various leukocyte subpopulations to sites of tissue damage, helper T cell responses as well as hematopoiesis [48, 49]. In addition, non-immune

chemokine activities involve angiogenesis, organogenesis, stem cell mobilization [50] and cell cycling [36, 51, 52].

1.2.1 Chemokine nomenclature and classification

At the Gordon Conference on chemotactic cytokines (June 23-28, 1996), systematic chemokine nomenclature was proposed. Prior to the meeting, chemokines were named by the investigators that discovered them. Some chemokines had several names and the acronyms chosen were usually random. The proposed systematic nomenclature was based on the current designation for chemokine receptors which is CC, CXC, XC, CX_3C followed by R for receptor and the number in which it was discovered. Similarly, chemokines under this classification are designated the same subfamily followed by L for ligand [53] (Table A-1). The numbering system follows the gene numbering encoded for each chemokine (www.rndsystems.com/chemokine_nomenclature.aspx). This standard is only applicable to human chemokines to date [54]. However, the non-systematic names are still used widely because they are familiar and descriptive. CXCL12, the ligand for CXCR4 will be referred to as stromal cell derived factor 1 (SDF-1) from here on in.

Chemokines are classified into 4 subfamilies based on the positioning of the first two of four conserved cysteine residues which are CXC, CC, CX_3C and C, where X represents any amino acid other than cysteine. Lymphotactin is an exception belonging to the C subfamily of chemokines which has only 2 cysteine residues and hence one disulfide bridge. Similarly, chemokine receptors are classified according to the subfamily of chemokine(s) that it binds to. For example CXCR4 is the corresponding receptor for CXC chemokine, SDF-1.

The CXC family can be further classified into the ELR and non-ELR containing motif chemokine. ELR chemokines such as interleukin-8 (IL-8) are involved in angiogenesis and bind to CXCR1 and CXCR2 receptors whereas non-ELR chemokines including interferon gamma (IFN-γ), inducible protein 10 (IP-10) and stromal cell derived factor 1 (SDF-1) bind to receptors CXCR3 and CXCR4, respectively (Table A-1).

1.2.2 Homeostatic and inflammatory chemokines

Chemokines are most commonly reported to be induced by pro-inflammatory cytokines such as IFN -γ and thus involved in inflammation. However, there is also a group of chemokines that have homeostatic functions and are constitutively expressed in organs and tissues in the absence of pro-inflammatory signals. Examples of homeostatic chemokines include CCL25, also known as TECK, which is involved in T cell development in the thymus, and SDF-1 which is involved in B cell development, organogenesis and stem cell homing. Yet another group of chemokines have both inflammatory and homeostatic functions, such as CX_3CL1 (fractalkine), which is highly expressed in the brain, an immunopriviledged organ, but is also induced by the pro-inflammatory cytokine TNF-α in endothelial cells [55].

1.2.3 SDF-1 and CXCR4

At present, only SDF-1 and CXCR4 gene deletion results in lethality in the chemokine and receptor families. Similar phenotypic abnormalities in both the receptor and chemokine knockouts as well as perinatal lethality suggested that these two binding partners were mutually exclusive. Other chemokines and chemokine receptors generally demonstrate promiscuous binding and functional redundancy compensates for the gene deletions resulting in unchanged

phenotypes. However, this only suggests that CXCR4 and SDF-1 are mutually exclusive binding partners in early life for the knock-out mice and there is a possibility that other ligands or receptors may bind later in life.

The exclusive binding between SDF-1 and CXCR4 was generally accepted up until CXCR7 was identified as an alternate receptor for SDF-1 [54]. CXCR7 binds to SDF-1 but does not induce Ca^{+2} mobilization or cell migration but is found to induce increased adhesive properties and promotes cell growth and survival [56, 57]

1.2.3.1 SDF-1

SDF-1 is highly conserved across species; mouse and human SDF-1 have one amino acid difference at position 17, from a valine in humans to an isoleucine in mice. SDF-1 was first coined the pre-B cell stimulating factor (PBSF) [58] and it is constitutively expressed in the stromal cells in the bone marrow [59]. SDF-1 retains pre-B cells in the appropriate niches for maturation prior to release into the periphery [60]. Subsequent *in vivo* work with SDF-1 knockout mice suggested that SDF-1 had roles beyond the immune system including embryonic development and homeostasis [61]. Defects in organ development indicated that SDF-1 is also involved in organogenesis. Major organs that were affected by SDF-1 or CXCR4 deletion were the hematopoietic, cardiovascular and central nervous system.

1.2.3.2 CXCR4

The NMR solution structure was solved for SDF-1; however, little yet is known at atomic resolution for CXCR4. Structural details on CXCR4 are mainly inferred from conserved motifs common with rhodopsin.

Similar to rhodopsin, CXCR4 has the DRY motif as well as a disulfide bridge in the second extracellular loop and a NPXXY motif in the C-terminus for receptor trafficking to the plasma membrane.

CXCR4 has ubiquitous expression in a variety of tissues such as various leukocyte populations, hematopoietic progenitor cells as well as non-hematopoietic cells such as endothelial and epithelial cells in the heart, central nervous system (CNS) and thymus [62]. Furthermore, different CXCR4 isoforms such as the splice variant CXCR4-lo, are not well characterized also contribute to receptor functional diversity [63]. Conditional knockouts were generated to clarify some of the SDF-1 and CXCR4 tissue specific functions post embryogenesis [64]. Conditional knockout mice, in which CXCR4 was selectively inactivated in B cell lineage precursors, revealed that CXCR4 has an important role in regulating homeostasis of B cell compartments and humoral immunity [64].

1.2.4 CXCR4 in cancer and tumour metastasis

Chemokine receptors are involved in the spread and metastatic growth of different tumours [65]. Under normal conditions CXCR4 is ubiquitously expressed at low levels in tissues including breast and ovary [54] but it is the most commonly overexpressed receptor in human cancer cells [66]. Expression levels are monitored for clinical diagnostics and targeted for anti-tumour therapies. However, receptor detection techniques are limited due to receptor heterogeneity which is not currently well understood; moreover, the reagents for detection have limited recognition, specificity and applicability [67]. We will describe a novel target probe validation for anti-CXCR4 and a high quality anti-CXCR4 antibody that we have developed to address these issues in this dissertation.

Small molecule drugs that have been developed against CXCR4 include AM5D3100, T140 and ALX20-4C. T140 and ALX20-4C are peptide antagonists to CXCR4 whereas AMD3100 is a bicyclam derivative. However, all of these inhibitors are of limited utility. AMD3100 and ALX20-4C have partial agonist activity and promote the mobilization of the different CXCR4 expressing cellular subsets [68, 69]. T140 is a short peptide derivative of an antiviral substance in horseshoe crab, which has secondary structure that mimics the strands of the SDF-1 antiparallel beta sheet structure and inhibits cancer metastasis [70] in pre-clinical models. However, T140 is small and fairly hydrophilic and is eliminated from the body after administration [71]. Consequently improved delivery and release systems are under investigation [71].

1.2.5 CXCR4 is a co-receptor for HIV-1 infection

Factors secreted from CD8+ cells that blocked HIV-1 infection of CD4+ T-cells were first leads to identifying chemokine receptors for HIV-1 infection [72-74]. The secreted factor was later identified as SDF-1 and the respective receptor, CXCR4 as a co-receptor for HIV-1 infection [75]. The envelope proteins on HIV-1 interact with CXCR4 to facilitate viral entry in macrophages and T-cells. Structure-function studies discovered that the CXCR4 HIV-1 binding determinants are in the second extracellular loop, particularly residues Asp 171 and Asp 262 [76, 77].

Small molecule antagonistic drugs have been developed to prevent HIV-1 infection through blocking viral interacting sites. AMD3100 blocks the CXCR4 receptor through electrostatic interactions with two asparagine residues on ends of the second extracellular loop

[78]. However, as previously mentioned, AMD3100 has residual agonist activity that may cause toxic side effects [68, 69].

1.2.6 Modified chemokine and peptide antagonists

A high resolution structure for CXCR4 would greatly advance rational drug design for chronic diseases such as cancer and AIDS. However, in the absence of high resolution information, modified chemokines as well as chemokine peptides with comparable binding affinity and high specificity to the receptor of interest are alternative approaches to the development of pharmaceutical antagonists. Some examples include modified chemokine Regulated upon Activation, Normal T cell Expressed (RANTES) or CCL5, MetRANTES [79] and amino-oxypentane (AOP)-RANTES [80] that are used to block CCR5, another HIV-1 coreceptor. MetRANTES has high binding affinity to CCR5 since it is the natural receptor chemokine that has an additional methionine attached on the N-terminus with partially muted activity and partial agonist activity. Antagonists to CXCR4 such as SDF-1 P2G which is SDF-1, in which proline in position 2 is replaced with glycine, is a high affinity antagonist. SDF-1 P2G was used to show that SDF-1 is involved in stem cell cycling which has ramifications for transplantation [48, 49].

1.3 Impact of chemically synthesized chemokines for advancing structure-function knowledge

The late Ian Clark-Lewis at the University of British Columbia was arguably the pioneer of chemical synthesis of functional chemokines. The Clark-Lewis group optimized solid phase t-Boc chemistry methods for the synthesis of 70-160 amino acid long full length bioactive

cytokines [81, 82] which were subsequently extended to include chemokines. His work was of great importance in advancing knowledge about motifs and structures that are involved in chemokine function through *in vitro* studies with synthetic hybrids and truncation variants. In particular, his seminal work with synthetic variants of SDF-1 proposed a model for chemokine docking and receptor activation [84] which has become a well accepted model for other members of the family.

Over the span of 15 years Dr. Clark-Lewis built an extensive peptide library of well over 1000 chemokines and related variants that were made readily available to other researchers in the field. The analogs and truncation variants that were generated for SDF-1 alone have been collated to show the breadth of the synthetic chemokine and corresponding analog repertoire that was generated in the Clark-Lewis laboratory (Table A-2). The high yield, purity and speed in which these chemokines could be generated greatly advanced high resolution structural analyses. Synthetic chemokines could be generated in less than a week and from a single synthesis, 1-100 milligrams of active homogenous chemokines were recovered [83], whereas purification from natural sources or recombinant approaches require more time.

These advances meant that for the first time functional chemokines could be rapidly generated by non-recombinant methods and in high purity and yield. They have been shown to be effective for *in vitro* [82, 84] as well as *in vivo* studies [85-87]. The first synthetic chemokine structurally characterized was CXCL8 or interleukin 8 (IL-8), which demonstrated to the scientific community that synthetic chemokines were equivalent to recombinantly produced chemokines [83, 88]. Taking advantage of the C-terminus to N-terminus synthesis directionality, several N-terminal truncation variants, analogs with common C-terminal sequence and varied N-

terminus as well as hybrids could be generated by splitting the solid-phase resin at intermediary steps prior to the N-terminus.

Although the initial IL-8 high resolution structure was solved with recombinant sources [89, 90], much of the structure-function relationships for IL-8 has been based on work with synthetic analogs and truncation variants. For instance, high resolution structural studies revealed a dimeric interface for IL-8 which was widely accepted to be active form of the chemokine [89, 90]. However, the Clark-Lewis and Sykes groups observed contradictory results using synthetic IL-8 analogs. They concluded that the dimerization observed in the NMR and crystallography may have been artifacts caused by the high protein concentrations used for those techniques. A synthetic analog that could not dimerize was shown to be fully bioactive in its monomeric form [84]. Synthetic approaches were crucial for the discovery of key motifs such as the ELR motif using other synthetic IL-8 variants [91].

1.3.1 Chemical synthesis of SDF-1

The high purity and yield in which chemokines could be generated by solid phase chemistry opened many new avenues for chemokine structure-function analysis. For example, this enabled the structural elucidation of a number of chemokines such as SDF-1. The SDF-1 solution structure was solved by NMR [84]; however, a high resolution structure is not available for the corresponding receptor CXCR4. What is currently known about the receptor has been largely inferred from motifs present on the ligand. There are two main sites on SDF-1 that are involved in the docking and activation of the receptor: the disordered N-terminus (residues 1- 8) and the RFFESH motif (residues 12-17) (Figure 1-3). In particular, the first two amino acids Lys-1 and Pro-2 on SDF-1 activate the receptor. When one or both of these amino acids are

removed, the truncated ligand binds with a dissociation constant (K_d) similar to the native form but is non-active [84]. Pro-2 is suggested to be important in positioning Lys-1 for binding to the extracellular loop(s) causing receptor conformational changes leading to downstream signal transduction. To further discern the role that Lys-1 plays in activation, an ornithine was substituted at the Lys-1 position [84]. Activity was retained with ornithine substitution at position 1 whereas arginine substitution drastically reduced activity; however, binding affinity was not affected suggesting that side chain size was important for activation. Addition of glycine on the absolute N-terminal of SDF-1 enhanced activity, suggesting that the N-terminal amine group is partially buried and can tolerate an additional amino acid on the N-terminus.

This work as well as other structure-function studies with N-terminal truncation variants identified the N-terminus as a key region involved in regulating chemokine activity suggesting that proteolysis may be involved in regulation. Matrix metalloproteinases (MMPs) have been implicated in clipping the N-terminus of the CC chemokine family of chemokines. Yeast two-hybrid screens offered the first indication that chemokines, particularly monocyte chemoattractant proteins, are substrates for MMP processing which was confirmed by *in vitro* functional and binding assays [92]. MMP chemokine processing produced high affinity antagonists that could also dampen inflammation *in vivo* [93].

1.3.2 Site-specific crosslinking analogs for receptor mapping

Site-specific crosslinkers have been extensively used to understand contact points between the receptor and ligand as a means of understanding receptor binding and activation. Photocrosslinkers are used for the study of many GPCRs such as the parathyroid hormone receptor [94], cholecystokinin [95] and the substance P (NK-1) receptor [96]. The crosslinkers

are activated by UV light to form radical intermediates that can react to form covalent linkages to the nearest neighbouring amino acid residues. The most common photoaffinity labels are aryl azides, aryl diazirines, α-diazocarbonyls and benzophenone derivatives [99].

Receptor crosslinking stabilizes protein interactions and near-neighbour relationships for known interacting partners. Receptor mapping of known binding proteins enable key regions on the receptor targeted for high resolution structural analyses. Mapping studies facilitate sectional analysis strategies that are reconstituted to understand the entire structure of the GPCR under study. This strategy allows for examination of the receptor in manageable segments.

Photocrosslinkers such as aryl diazirines and benzophenone derivatives associated with phenylalanine residues may be directly incorporated into the peptide of interest by solid phase synthesis. Benzoylphenylalanine (BPA) is by far the most popular photocrosslinkers because it offers several advantages: it can be directly incorporated into the ligand of interest by chemical synthesis and is commercially available, and it is stable in water. The reaction that it undergoes upon photoactivation is illustrated in Figure 1-4.

The crosslinked site is conventionally traced by radioisotopic labeling of the photoactive ligand, followed by enzymatic digestion, separation on polyacrylamide gels or reversed phase columns and then detection by autoradiography, UV absorbance or mass spectrometry. The interaction site is then deduced from mass changes between the crosslinked and non-crosslinked receptor species. When coupled to sensitive detection techniques such as mass spectrometry, the receptor may be probed in more physiologically relevant expression systems.

1.4 Mass spectrometry of membrane proteins and overview of membrane proteomics

Integral membrane proteins are not readily adaptable to experimental techniques that require abundant quantities of protein in solution. With the advent of soft ionization techniques [96-98], the increasing power of computational analysis and the sequencing of the human genome, sensitive mass spectrometry based approaches have been particularly useful for understanding the membrane protein complement of cells in isolated or global studies.

When the high data-collecting capacity, speed and sensitivity of mass spectrometry are applied to complex biological samples, the expression between proteins in two samples may be compared. Unique protein expression differences observed between healthy and diseased tissues are potential biomarkers. Biomarkers are employed for disease detection or to monitor disease progression. Accurate biomarkers for early disease diagnosis have implications for patient survival and/or identify therapies for immediate treatment.

The sequence of steps leading to protein MS fingerprinting, tandem mass spectrometry and characterization entails sample preparation, acquisition and data interpretation.

1.4.1 Sample preparation for mass spectrometric analysis

Both *in-solution* and *in-gel* based sample preparation have been applied to membrane protein mass spectrometric identification and mapping. Methods for membrane protein sample preparation has been commonly developed on readily available endogenous bovine rhodopsin [101] used as a model GPCR, as well as the related proton pump, bacteriorhodopsin [102].

1.4.1.1 *In-solution digestion*

A number of solution-based sample preparation methods for membrane extracted proteins for matrix-assisted laser desorption (MALDI) [103] and electrospray ionization (ESI) mass spectrometry [101] are reported despite the challenges involved with protein solubilisation. Detergents and chaotropes interfere with liquid chromatography (LC) interfaced ESI-MS instruments but weak detergents are tolerable at low concentrations on the MALDI-MS [104]. Strategies to overcome detergent and chaotrope interference include LC and MS compatible or cleavable detergents [105] and protein precipitation partition [106] but each of these methods may not be effective for every membrane protein and require trial and error testing.

Solubilisation may be avoided for some membrane proteins by digestion directly from the plasma membrane to release extracellular and intracellular peptides. However, in the membrane fluid mosaic many other proteins are present and thus produce a complex mixture of peptides that may mask lower abundant multipass membrane proteins or require special data interpretation software [107]. This approach does not provide information on hydrophobic regions on the receptor and is biased towards identifying proteins that are less functionally complex or hydrophobic. The majority of the membrane proteins reported with this method were single pass proteins, although some GPCRs were detected with low sequence coverage and consequently low confidence [107].

1.4.1.2 *In-gel digestion*

Alternatively, membrane proteins separated by sodium dodecyl sulphate gel polyacrylamide gel electrophoresis (SDS-PAGE) may be digested *in-situ* and the peptides extracted from detergents and other mass spectrometric incompatible reagents for analysis. Although this method enables one to concentrate the protein to one location for digestion,

peptide extraction from the gel is not always efficient. Extraction is dependent on the buffers used and the length and hydrophobicity of the peptides produced. Several groups investigated method and sample preparation improvements for *in-gel* digestions specifically for membrane proteins involving organic solvents [108] or small amounts of non-ionic detergents [102, 109] for digested peptide extraction from the gel.

1.4.2 Data acquisition and interpretation:

Mass spectra collected for in-solution or in-gel digested samples for complex samples can generate tens of thousands of unique peptides that are not all selected for further fragmentation or analysis. Although most modern mass spectrometers have the capacity to fragment peptide ions faster than one per second, the complexity of biological samples, either isolated or in mixtures can be overwhelming. Data acquisition is often information dependent and peptides that meet threshold intensities at a specific collision induced dissociation (CID) energies are selected for further downstream fragmentation. As a result of abundance or preferential ionization, redundant peptides are repeatedly sequenced whilst other segments of the protein or proteome are ignored. Thus, only a biased set of peptides are detected in a mass spectrum which may result in the absence or underestimation of particular protein(s) in a simple or complex sample.

A variety of algorithms have been developed to search MS as well as MS/MS data which include Mascot, Profinder, ProteinProphet, X!Tandem and SEQUEST for protein identification. These searches against protein databases such as NCBI, SwissProt, and MSDB are dependent on available sequence and PTM knowledge for the specific protein under study. Proteins containing post- translational modifications which are not accounted in existing databases influence protein

searches and confident identifications. Ironically, in this situation technology is driven as well as limited by the technology; there is a demand for database population with mass spectrometric data as well as biochemical post-translational modification data to improve global proteomics.

1.4.3 Proteotypic peptides and annotated peptide libraries

To improve efficiency of protein scoring and identification, platforms have been developed to integrate frequent experimentally observed peptides and ions for a particular protein or isoform in annotated databases for more efficient and ordered comparisons based on proteotypic peptides. Proteotypic peptides are defined as those peptides in a given sample that are consistently detected or fragmented from a protein in more than 50% of experiments in which the protein was identified [110]. These experimentally recurring characteristic peptides are targeted for analysis to improve identification. Proteotypic peptide scoring or searches provides an added level of confidence for protein identification and maybe extended for protein quantitative applications. However, for this platform to be effective these annotated libraries must be populated with experimental findings for each protein.

1.4.4 Targeted approaches

Targeted approaches are employed to specifically hone in on one particular protein such as disease biomarkers which encompass a variety of GPCRs. One such GPCR is CXCR4 which is implicated in cancer and AIDS. The ability to observe a given protein in a complex mixture by mass spectrometry is not controlled unless it is the most abundant protein in the mixture and thus acquisition parameters need to be refined for targeted detection.

Multiple reaction monitoring (MRM) is a targeted strategy which chooses a number of pre-specified precursor ions of a specific m/z value for fragmentation and pre-determined characteristic fragment ions are monitored for protein identification [111]. In MRM-based, LC-MS/MS experiments, the acquisition parameters are adjusted to collect information for specific precursor and fragment ion transitions to improve sensitivity, specificity and efficiency of the detection of important physiological targets in complex mixtures. In order for MRM to be viable for proteins that are less abundant or more difficult to detect, the specific precursor to fragment ion transitions need to be pre-determined. One way to collect MRM information is to employ reductive approaches which require established expression and enrichment methods for MS/MS sequencing and characterization of the given protein, especially if the protein is not readily available from natural sources.

Specific data obtained by hypothesis-driven reductive approaches provide a basis for targeted global studies for improved biomarker detection of disease conditions.

1.4.5 GPCRs as biomarkers in global proteomics

Unusual protein expression profiles typically signify disease and the deregulated protein involved is often selected as a biomarker traced for diagnostics and treatment. As discussed in the previous sections, GPCRs mediate many cellular processes and thus tight functional control of these receptors is essential for normal physiology as well as therapeutic impact in GPCR-related diseases.

GPCRs are implicated in a broad spectrum of diseases and are important biomarkers which potentially can be traced by antibody-based or mass spectrometric approaches. However, antibody-based methods are contingent on the reagents available for the protein under study.

Global mass spectrometric approaches are powerful for tracking diagnostic protein markers in human tissues. As separation methods, detection sensitivity and bioinformatics for global proteomic analyses improve, mass spectrometry will assume a more central role in clinical research.

1.4.5.1 Antibody-based biomarker identification

It is reported that only 35 % of antibodies on the market are effective for the studied target [45] and consequently result in underestimations of proteins in expression profiling. This is evident for anti-GPCR antibodies such as CXCR4 which is antigenically heterogenous [67]. Thus, antibodies that are generated to non-variable antigens are well sought after. In response to this, there have been strategies such as the protein atlas (www.proteinatlas.org) established by Matthias Uhlen's group to generate a consolidated repository of universal validated antibodies for antibody based proteomics.

1.4.5.2 Label free biomarker identification

Shotgun approaches are one way of discovering novel biomarkers. This approach attempts to identify as many proteins as possible which involve trypsin digestion or other proteases [110], reversed phase separation and mass spectrometry. Prospective biomarkers are identified by comparative analysis of proteins identified in different physiological states. Expression levels and degrees of post-translational modification are compared between different sample conditions.

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)-based methods examining expression changes and PTM have also been described which involve densitometry and specialized detection reagents. Differences visualized on the gel are then confirmed by

mass spectrometry to identify the potential biomarker. This method is particularly useful for soluble proteins but not well-suited for membrane proteins [113].

1.4.5.3 Biomarker validation

Biomarkers selected from global proteomic analyses are conditional on physiological verification. Determining the significance of expression change is a challenge for biomarker discovery. Biological validation often follows a trial and error approach. The candidate protein may be selected for subsequent validation based on the reagents available for further analysis (ie: antibodies) and the quality of the tandem mass spectra.

Conversely, potential biomarkers may be identified by alternate biochemical approaches or be based on prior knowledge for global proteomic analysis, in which case mass spectrometric detection parameters for complex mixture can be refined for improved detection and analysis. For the latter, clinical proteomics is dependent on existing information in database to guide more targeted mass spectrometric approaches.

1.4.6 Quantification

Mass spectrometry is conventionally a sequence-based tool for qualitative analysis of proteins; however, with the introduction of covalent stable isotopic labeling for proteins referred to as isotope-coded affinity tag (ICAT), relative protein quantification was made possible [114]. Typical protein quantification methods such as quantitative reverse transcriptase –polymerase chain reaction (QRT-PCR) indirectly measure mRNA transcripts present in cells and infer protein concentration whereas mass spectrometric approaches directly examine expressed proteins.

Another method for quantification that does not covalently modify the endogenous biomarker is referred to as absolute quantification (AQUA) [115]. Stable isotopic forms of known proteotypic peptides for a given protein are spiked into digested protein mixtures for absolute quantification based on spectral intensity computations. Different stable isotopes are incorporated in to synthetic biomarker proteotypic peptides are spiked into the complex mixtures that are under study as an internal standard. Relative intensity differences between the isotopically labeled peptide and the corresponding peptide derived from the endogenous protein is used in the absolute quantification. Similarly, challenges observed with accurate protein scoring with regards to incomplete or inaccurate databases also affect quantitative approaches.

Comprehensive methods to populate databases and annotated repositories would greatly improve global proteomic studies.

1.5 Thesis objectives

GPCRs are well studied but tertiary structural information for this family of proteins is under-represented due to intrinsic challenges with expression and purification. In addition to intrinsic challenges, the lack in consistent reagents and validation, as well as the lack of PTM site information introduces bottlenecks to the study of GPCRs. We aim to improve the understanding of GPCRs expressed in less heterologous systems employing front of the line technologies. We have mainly focussed on human CXCR4, an important GPCR for this dissertation.

We intend to achieve these goals by:

1) Devising a general, comprehensive and versatile enrichment approach to characterize GPCRs expressed in physiological contexts.

2) Improving reagents available for the study of GPCRs for specific protein analysis

    a) Generate an effective anti-human CXCR4 monoclonal antibody

    b) Develop an antibody screening platform for accurate target probe validation

3) Applying the findings from the previous points to study the structure-function relationships for SDF-1 and CXCR4 employing novel site-specific crosslinkers and mass spectrometry.

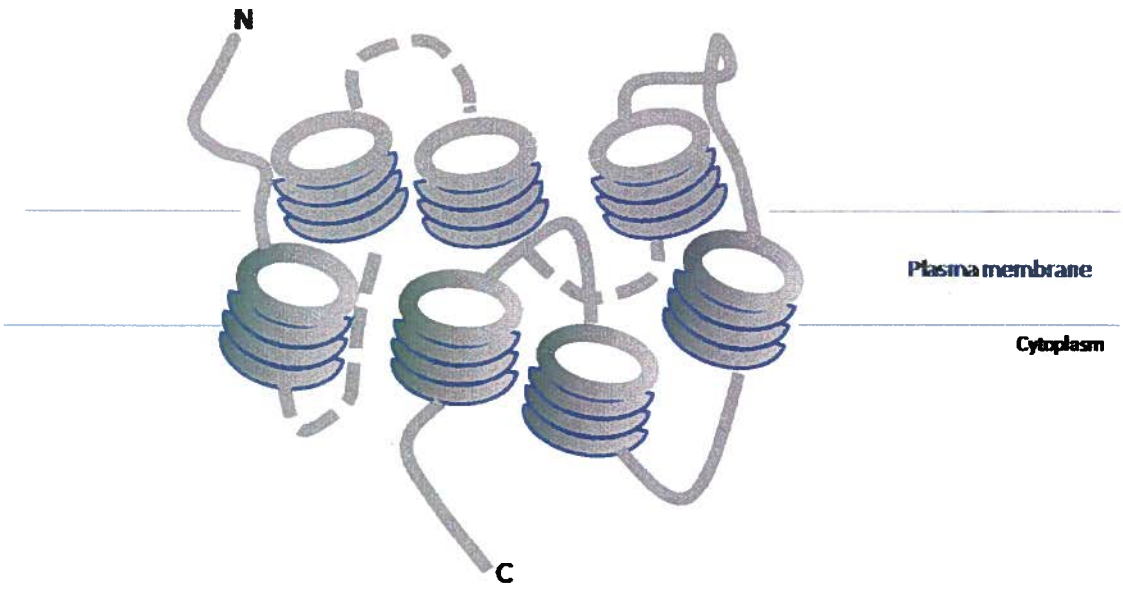


Figure 1-1: Diagrammatic representation of a termini orientation of a typical G-protein coupled receptor, adapted from the structure of rhodopsin. © Julie Patricia Wong, 2008

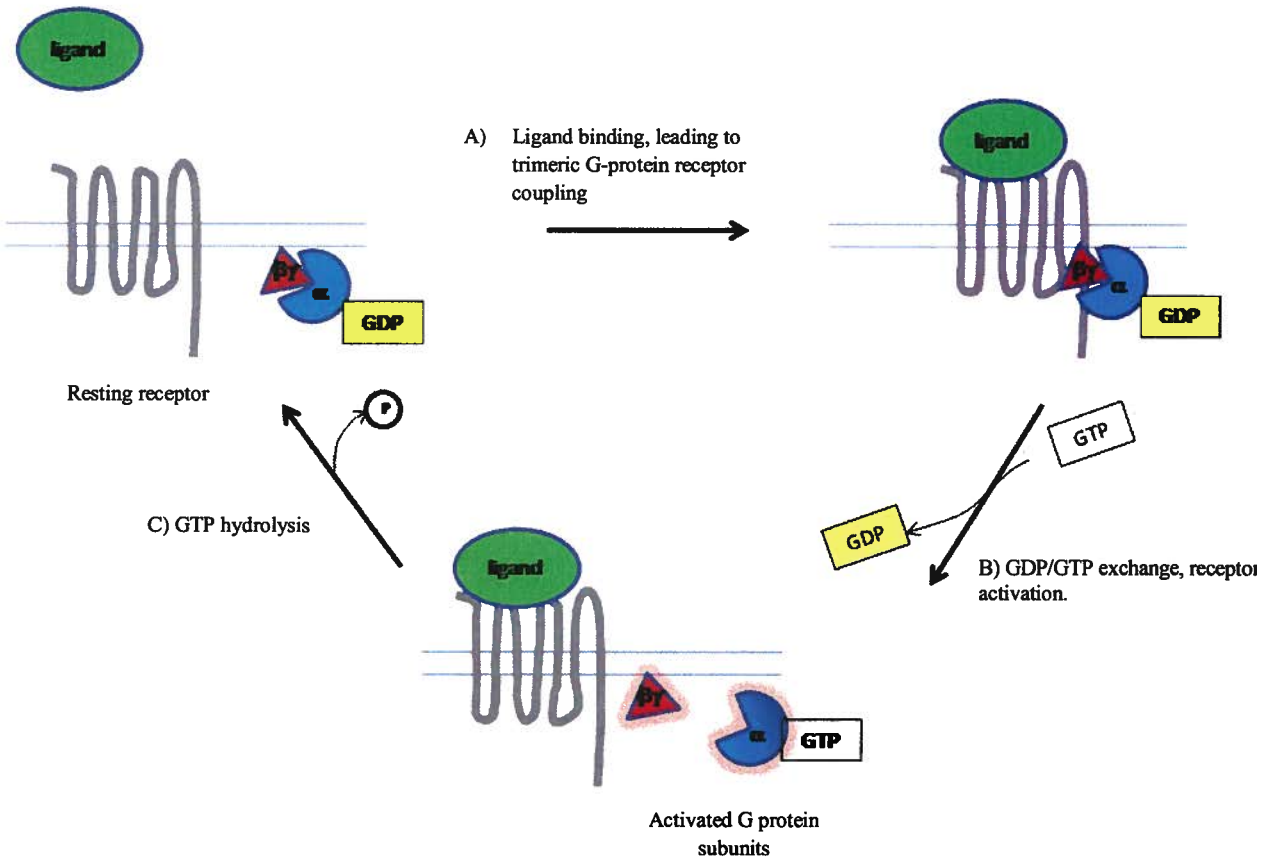


Figure 1-2: General diagrammatic representation of GPCR A) ligand binding, B) activation and C) GTP hydrolysis. © Julie Patricia Wong, 2008.

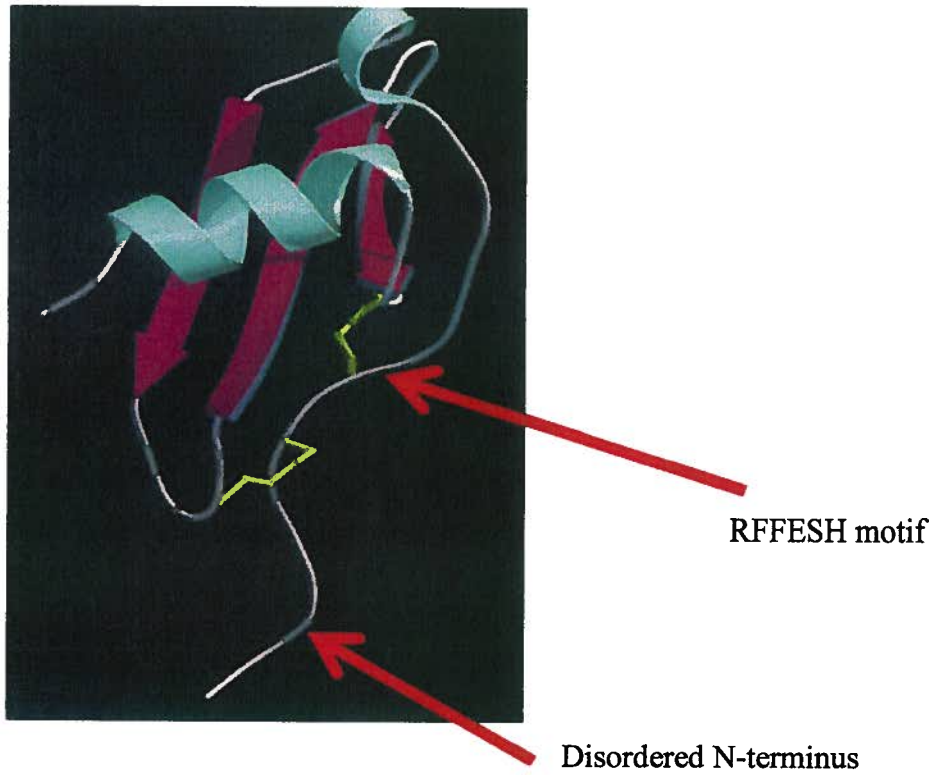


Figure 1-3: Ribbon diagram of the NMR solution structure for synthetic SDF-1, key motifs are indicated with red arrows and yellow bonds represent disulfide bridges. Figure adapted by permission from Macmillan Publishers Ltd: [EMBO] (Crump et al. 1997 [84]), copyright 1997.

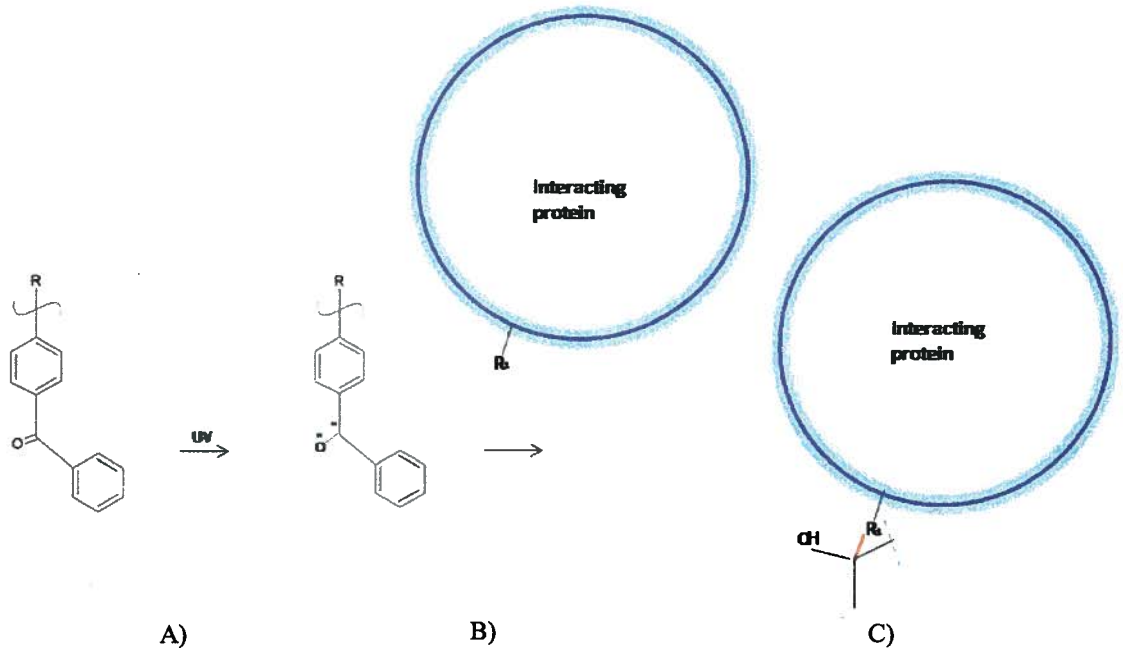


Figure 1-4: Diagrammatic representation of the steps involved in photocrosslinking with A) BPA incorporated into a protein, which upon B) UV irradiation forms a reactive radical intermediate that C) forms a covalent linkage to neighbouring residues on an interacting protein. The covalent linkage is indicated in red and the rest of the crosslinker and remaining portions of the ligand analog has been deleted for clarity. R represents phenylalanine and R_1 represents a neighbouring carbon atom on an amino acid side chain. Figure is not drawn to scale. © Julie Patricia Wong, 2008

1.6 References

1. Alberts B, et al., *Essential Cell Biology: An Introduction to the Molecular Biology of the Cell.* 1998, New York: Garland Publishing.

2. Schoneberg T, Schultz, and Gudermann, *The structural basis of G-protein-coupled receptor function and dysfunction in human diseases.* Rev. Phys. Biochem. Pharm., 2002 **144**: p. 145-227.

3. Kolakowski LF Jr, *GCRDb: a G-protein-coupled receptor database.* Receptors Channels, 1994. **2**(1): p. 1-7.

4. Vriend G, *WHAT IF: a molecular modeling and drug design program.* J Mol Graph, 1990. **8**(1): p. 52-56.

5. Davies MN, et al., *Proteomic applications of automated GPCR classification.* Proteomics, 2007. **7**: p. 2800-2814.

6. Dryja TP, et al., *A point mutation of the rhodopsin gene in one form of retinis pigmentosa.* Nature, 1990. **343**: p. 364-366.

7. Rosenthal W, S.A., Antaramian A, Lonergan M, Arthus M-F, Hendy GN, Birnbaumer M and Hichet DG, *Molecular identification of the gene responsible for congentital nephrogenic diabetes insipidus.* Nature, 1992. **359**: p. 233-235.

8. Pan Y, M.A., Das S, Jing B, Gitschier J, *Mutations in the V2 vasopressin receptor gene are associated with X-linked nephrogenic diabetes insipidus.* Nature Genetics, 1992. **2**(2): p. 103-106.

9. Hernandez PA, et al., *Mutations in the chemokine receptor gene CXCR4 are associated with WHIM syndrome, a combined immunodeficiency disease.* Nature Genetics, 2003. **34**: p. 70-74.

10. Schoneberg T, et al., *Mutant G-protein-coupled receptors as a cause of human diseases.* Pharmacology & Therapeutics, 2004. **104**(3): p. 173-206.

11. Tagat JR, et al., *Piperazine-based CCR5 antagonists as HIV-1 inhibitors II. Discovery of 1-[(2,4-Dimethyl-3-pyridinyl)carbonyl]-4-methyl-4-[3(S)-methyl -4-[1(S)-[4-(trifluoro-methyl)phenyl]ethyl]-1-piperazinyl]-piperidine N1-oxide (Sch-350634), an orally bioavailable, potent CCR5 antagonist.* J. Med. Chem., 2001 **44**(21): p. 3343-3346.

12. Horuk R, *Development and evaluation of pharmacological agents targeting chemokine receptor* Methods, 2003. **29**: p. 369-375.

13. Civelli O, et al., *Novel neurotransmitters as natural ligands of orphan G-protein-coupled receptors.* Trends in Neuroscience, 2001. **24**: p. 230-237.

14. Palczewski K, K.T., Hori T, Behnke C, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M., *Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor* Science, 2000. **289**(5480): p. 739-745.

15. Dixon RAF, et al., *Cloning of the gene and cDNA for mammalian β-adrenergic receptor and homology with rhodopsin.* Nature, 1986. **321**(May 1): p. 75-79.

16. Buck L and Axel R, *A novel multigene family may encode odorantreceptors: A molecular basis for odor recognition.* Cell Mol. Life Sci., 1991. **65**(1): p. 175-187.

17. Royant A, Nollert P, Edman K, Neutze R, Landau EM, Pebay-Peyroula E and Navarro J, *X-ray structure of sensory rhodopsin II at 2.1 Å resolution.* PNAS, 2001. **98**(10131-10136).

18. Belrhali H, *Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å.* Structure, 1999. **7**: p. 909-917.

19. Cherezov V, et al., *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science, 2007. **318**(5854): p. 1258-1265.

20. Rasmussen SG, C.H., Rosenbaum DM, Kobilka TS, Thian FS, Edwards PC, Burghammer M, Ratnala VR, Sanishvili R, Fischetti RF, Schertler GF, Weis WI and Kobilka BK, *Crystal structure of the human beta2 adrenergic G-protein-coupled receptor.* Nature, 2007. **450**(7168): p. 383-387.

21. Rosenbusche JP, *Stability of membrane proteins: relevance for the selection of appropriate methods for high-resolution structure determinations.* J. Struct. Biol. , 2000. **136**(144-157).

22. Day PW, et al., *A monoclonal antibody for G protein coupled receptor crystallography.* Nature Methods, 2007. **4**(11): p. 927-929.

23. Granier S, et al., *Structure and conformational changes in the C-terminal domain of the beta 2-adrenoceptor: insights from fluorescence resonance energy transfer studies.* Journal of Biological Chemistry, 2007. **282**(18): p. 13895-13905.

24. Kobilka BK, K.T., Daniel K, Regan JW, Caron MG and Lefkowitz R, *Chimeric alpha 2-, beta 2-adrenergic receptors: delineation of domains involved in effector coupling and ligand binding specificity.* Science, 1988. **240**(4857): p. 1310-1316.

25. Chakir K, et al., *The third intracellular loop and the carboxyl terminus of β_2 adrenergic receptor confer spontaneous activity of the receptor.* Molecular Pharmacology, 2003. **64**: p. 1048-1058.

26. Cai K, Klein-Seetharaman J, Farrens D, Zhang C, Altenback C, Hubbell WL, and Khorana G, *Single-cysteine substitution mutants at amino acid positions 306-321 in ry, rhodopsin, the sequence between the cytoplasmic end of helix VII and the palmitoylation sites: sulfhydrl reactivity and transducin activation reveal a tertiary structure.* Biochemistry, 1999. **38**: p. 7925-7930.

27. Yang K, Farrens DL, Altenbach C, Farahbakhsh ZT, Hubbell WL, and KhoranaHG, *Structure and function in rhodopsin. Cysteines 65 and 316 are in proximity in a rhodopsin mutant as indicated by disulfide formation and interactions between attached spin labels.* Biochemistry, 1996. **35**: p. 14040-14046.

28. Archer E, et al., *Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors?* . Trends in Pharmacology Science, 2003. **24**: p. 36-40.

29. Arnis S and Hofmann KP, *Two different forms of metarhodopsin II: schiff base deprotonation precedes proton uptake and signaling state.* PNAS, 1993. **90**(16): p. 7849-7853.

30. Arnis S and Hofmann KP, *Photogeneration of bovine rhodopsin from its signaling state.* Biochemistry, 1995. **34**(29): p. 9333-9340.

31. Li J, H.P., Chen C, de Reil JK, Weinstein H, and Liu-Chen LY, *Constitutive activaton of the μ–opioid receptor by mutation of D3.49(164), but not D3.32(147): D3.49(164) is critical for stabilization of the inactive form of the receptor and for its expression* Biochemistry, 2001. **40**(40): p. 12039-12050.

32. Scheer A, et al., *Mutational analysis of the highly conserved arginine within the Glu/Asp-Arg-Tyr motif of the alpha1b-adrenergic receptor: effects on receptor isomerization and activation.* Molecular Pharmacology, 2000. **57**: p. 219-231.

33. Scheer A, et al., *Constituitively active mutants of the alpha 1B-adrenergic receptor: roleof highly conserved polar amino acids in receptor activation.* EMBO, 1996. **15**: p. 3566-3578.

34. Jensen AD, G.F., Rasmussen SG, Asmar F, Ballesteros JA, and Gether U, *Agonist-induced conformational changes at the cytoplasmic side of transmembrane segment 6 in the beta 2 adrenergic receptor mapped by site-selective fluorescent labeling.* Journal of Biological Chemistry, 2001. **276**(12): p. 9279-9290.

35. Gether U, et al., *Agonists induce conformational changes in transmembrane domains III and VI of the beta2 adrenoceptor.* EMBO, 1997. **16**(22): p. 6737-6747.

36. Roland J, et al., *Role of the intracellular domains of CXCR4 in SDF-1-mediated signaling.* Blood, 2003. **101**(2): p. 399-406.

37. Carrio MM and Villaverde A, *Construction and deconstruction of bacterial inclusion bodies* Journal of Biotechnology 2002 **96**: p. 3-12.

38. Sarramegna V, T.F., Demange P and Milon A. , *Heterologous expression of G-protein-coupled receptors: comparison of expression systems from the standpoint of large-scale production and purification.* Cell Mol. Life Sci. , 2003. **60**(8): p. 1529-1546.

39. Kiefer H, M.K., and Vogel R, *Refolding of G protein-coupled receptors from inclusion bodies produced in E. coli.* . Biochem. Soc. Trans., 1999. **27**: p. 908-912.

40. Baneres JL and Parello J, *Structure-based analysis of GPCR function: evidence for a novel pentameric assembly between the dimeric leukotriene B4 receptor BLT1 and the G-protein.* Journal of Molecular Biology, 2003. **329**(4): p. 815-829.

41. Baneres JL, et al., *Molecular characterization of a purified 5-HT4 receptor: a structural basis for drug efficacy.* Journal of Biological Chemistry, 2005. **280**: p. 20253-20260.

42. Lanctot PM, Leclerc PC, Clement M, Auger-Messier M, Escher E, Leduc R and Guillemette G, *Importance of N-glycosylation position for cell-surface expression, targeting, affinity and quality control of the human AT_1 receptor.* Biochem J., 2005. **390**: p. 367-376.

43. Lanctot PM, Leclerc PC, Escher E, Guillemette G and Leduc R, *Role of N-glycan-dependent quality control in the cell-surface expression of the AT_1 receptor.* Biochemical and Biophysical Research Communications 2006. **340**: p. 395-402.

44. Liitti S, Matikainen MT, Scheinin M, Glumoff T and Goldman A, *Immunoaffinity purification and reconstitution of human alpha(2)-adrenergic receptor subtype C2 into phospholipid vesicles.* Protein Expression Purification 2001. **22**: p. 1-10.

45. Blow N, *Antibodies: The generation game.* Nature, 2007. **447**: p. 741-744.

46. Baribaud F, Edwards TG, Sharron M, Brelot A, Heveker N, Price K, Mortari F, Alizon M, Tsang M, and Doms RW, *Antigenically distinct conformations of CXCR4.* Journal of Virology, 2001. **75**(19): p. 8957-8967.

47. Hodges RS, Heaton RJ, Parker JM, Molday L and MoldayRS, *Antigen-antibody interaction.* The Journal of Biological Chemistry, 1988. **263**: p. 11768-11775.

48. Misslitz A, Pabst O, Hintzen G, Ohl L, Kremmer E, Petrie HT, and Forster R, *Thymic T cell development and progenitor localization depend on CCR7.* Journal of Experimental Medicine, 2004. **200**(4): p. 481-491.

49. Nagasawa T, Kikutani H and Kishimoto T, *Molecular cloning and structure of a pre-B-cell growth-stimulating factor.* PNAS, 1994. **91**: p. 2305-2309.

50. Cashman J, et al., *Changes in the proliferative activity of human hematopoietic stem cells in NOD/SCID mice and enhancement of their transplantability after in vivo treatment with cell cycle inhibitors* J. Exp. Med., 2002. **196**(9): p. 1141-1149.

51. Cashman J, et al., *Stromal-derived factor 1 inhibits the cycling of very primative human hematopoietic cells in vitro and in NOD/SCID mice.* Blood, 2002. **99**(3): p. 792-799.

52. Lataillade, J-J, Clay D, Bourin P, Herodin F, Dupuy C, Jasmin C and Le Bousse-Kerdiles M-C, *Stromal cell-derived factor 1 regulates primative hematopoiesis by suppressing apoptosis and by promoting G_0/G_1 transition in $CD34^+$ cells: evidence for an autocrine/paracrine mechanism.* Blood, 2002. **99**: p. 1117-1129.

53. Bacon K, et al., *Chemokine/Chemokine receptor nomenclature.* Journal of Leukocyte Biology, 2001. **70**(September): p. 465-466.

54. Muller A, Homey B, Soto H, Ge N, Catron D, Buchannan ME, McClanaha T, Murphy E, Yuan W, Wagner SN, Barrera JL, Mohar A, Verastegui E and Zlotnik A, *Involvement of chemokine receptors in breast cancer metastasis.* Nature, 2001. **410**: p. 50-56.

55. Barzan JF, et al., *A new class of membrane-bound chemokine with a CX_3C motif.* Nature, 1997. **385**: p. 640-6.

56. Balabanian K, et al., *The chemokine SDF-1/CXCL12 binds to and signals through the orphan receptor RDC1 in T lymphocytes.* Journal of Biological Chemistry, 2005. **280**(42): p. 35760-35766.

57. Burns JM, SB, Wang Y, Melikian A, Berahovich R, Miao Z, Penfold MET, Sunshine MJ, Littman DR, Kuo CJ, Wei K, McMaster BE, Wright K, Howard MC and Schall TJ. , *A novel chemokine receptor for SDF-1and I-TAC involved in cell survival, cell adhesion, and tumor development.* Journal of Experimental Medicine, 2006. **203**(9): p. 2201-2213.

58. Nagasawa T, Tachibana K and Kishimoto T, *A novel CXC chemokine PBSF/SDF-1 and its receptor CXCR4: their functions in development, hematopoiesis and HIV infection.* Immunology, 1998. **10**: p. 179-185.

58. Shirozu M, et al., *Structure and chromosomal localization of the human stromal cell-derived factor 1 (SDF-1) gene.* Genomics, 1995. **28**: p. 495-500.

60. Ma Q, Jones D, and Springer T, *The chemokine receptor CXCR4 is required for the retention of B lineage granulocytic precursors within the bone marrow microenvironment.* Immunity, 1999. **10**: p. 463-471.

61. Peled A, Petit I, Kollet O, Magid M, Ponomaryov T, Byk T, Nagler A, Ben-Hur H, Many A, Shultz Lider, O Alon, R, Zipori D and Lapidot T, *Dependence of human stem cell engraftment and repopulation of NOD/SCID mice on CXCR4.* Science, 1999. **283**: p. 845-848.

62. Zou Y-R, Kottmann AH, Kuroda M, Taniuchi I and Littman D, *Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development.* Nature, 1998. **393**: p. 595-599.

63. Gupta SK and Pillarisetti K, *Cutting Edge: CXCR4-Lo: Molecular Cloning and Functional Expression of a Novel Human CXCR4 Splice Variant.* Journal of Immunology, 1999. **163**(5): p. 2368-2372.

64. Nie Y, Waite J, Brewer F, Sunshine M-J, Littman DR and Zou Y-R, *The role of CXCR4 in maintaining peripheral B cell compartments and humoral immunity* Journal of Experimental Medicine, 2004. **200**(9): p. 1145-1156.

65. Balkwill F, *Cancer and the chemokine network* Nature Reviews Cancer, 2004. **4**: p. 540-550.

66. Ruffini PA, M., Cabioglu N, Altundag K, and Cristofanilli M, *Manipulating the chemokine-chemokine receptor network to treat cancer.* Cancer, 2007. **109**(12): p. 2392-2404.

67. Carnec X, Quan L, Olson WC, Hazan U and Dragic T, *Anti-CXCR4 monoclonal antibodies recognizing overlapping epitopes differ significantly in their ability to inhibit entry of human immunodeficiency virus type 1.* Journal of Virology, 2005. **79**(3): p. 1930-19-33.

68. Zhang WB, Naveno JM, Haribabu B, Tamamura H, Hiramatu K, Omagari A, Pei G, Manfredi JP, Fuji N, Broach JR and Peiper SC, *A point mutation that confers constituitive activity to CXCR4 reveals that T140 is an inverse agaonist and that AMD3100 and ALX40-4C are wearch partial agonists.* . The Journal of Biological Chemistry, 2002. **277**(27): p. 24515-24521.

69. Liles WC, Broxymeyer HE, Rodger E, Wood B, Jubel K, Cooper S, Hangoc G, Bridger GJ, Henson GW, Calandra G, and Dale DC, *Mobilization of hematopoietic progenitor cells in healthy volunteers by AMD3100, a CXCR4 antagonist.* Blood, 2003. **102**(8): p. 2728-2730.

70. Hirokazu T, et al., *T140 analogs as CXCR4 antagonists identified as anti-metastatic agents in the treatment of breast cancer.* FEBS Letters, 2003. **550**(1-3): p. 79-83.

71. Takenaga M, et al., *A single treatment with microcapsules containing a CXCR4 antagonist suppresses pulmonary metastasis of murine melanoma.* Biochemical and Biophysical Research Communications, 2004. **320**(1): p. 226-232.

72. Cocchi F, DeVico AL, Garzino-Demo A, Cara A, Gallo RC and Lusso P, *The V3 domain of the HIV-1 gp120 envelope glycoprotein is critical for chemokine-mediated blockade of infection.* Nature Medicine, 1996. **2**(11): p. 1244-1247.

73. Blackbourn DJ, et al., *Human CD8+ cell non-cytolytic anti-HIV activity mediated by a novel cytokine.* Res. Immunol., 1994. **145**(8-9): p. 658-659.

74. Balter M, *Elusive HIV-suppressor factors found.* Science, 1995. **270**(5242): p. 1560-1561.

75. Mackewicz CE, Barker E, and Levy JA, *Role of beta-chemokines in suppressing HIV replication.* Science, 1996. **274**(5291): p. 1393-1395.

76. Hatse S, Princen K, Lars-Ole G, Bridger G, Henson G, De Clercq E, Swartz TW, and Schols D, *Mutation of Asp^{171} and Asp^{262} of the chemokine receptor CXCR4 impairs its coreceptor function for human immunodeficiency virus-1 entry and abrogates the antagonistic activity of AMD 3100.* Molecular Pharmacology, 2001. **60**(1): p. 164-173.

77. Knuesel M, Wan Y, Xiao Z, Holinger E, Lowe N, Wang W and Liu X, *Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system.* Molecular and cellular proteomics, 2003. **2**(11): p. 1225-1233.

78. Hatse S, Princen K, Gerlach LO, Bridger G, Henson G, De Clercq E, Schwartz TW, and Schols D, *Mutation of Asp(171) and Asp(262) of the chemokine receptor CXCR4 impairs its coreceptor function for human immunodeficiency virus-1 entry and abrogates the antagonistic activity of AMD3100.* Mol Pharmacol, 2001. **60**(1): p. 164-173.

79. Proudfoot AE, Power CA, Hoogewerf AJ, Montjovent M-O, Borlat F, Offord RE and Wells TNC, *Extension of recombinant human RANTES by the retention of the initiating methionine produces a potent antagonist.* Journal of Biological Chemistry, 1996. **271**(5): p. 2599-2603.

80. Mack M, Luckow B, Nelson PJ, Cihak J, Simmons G, Clapham PR, Signoret N, Marsh M, Stangassinger M, Borlat F, Wells TNC, Schlondorff D and Proudfoot AEI, *Aminooxypentane-RANTES induces CCR5 internalization but inhibits recycling: a novel inhibitory mechanism of HIV infectivity.* Journal of Experimental Medicine, 1998. **187**(8): p. 1215-1224.

81. Clark-Lewis I, Hood LE, and Kent SB, *Role of disulfide bridges in determining the biological activity of interleukin 3.* PNAS, 1988. **85**(November): p. 7897-7901.

82. Clark-Lewis I, Aebersold, Ziltner H, Schrader JW, Hood LE and Kent SB., *Automated chemical synthesis of a protein growth factor for hemopoietic cells, interleukin-3.* Science, 1986. **231**(4734): 134-139.

83. Clark-Lewis I, *Synthesis of Chemokines*, in *Chemokine Protocols*, W.T. Proudfoot AEI, Power CA, Editor. 2000, Humana Press: Ottawa. p. 47-63.

84. Crump MP, G.J., Loetscher P, Rajarathnam K, Mara A, Arenzana-Seisdedos F, Virelizier JL, Baggiolini M, Sykes BD and Clark-Lewis I., *Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 from binding and inhibition of HIV-1.* EMBO, 1997. **16**(23): p. 6996-7007.

85. Kohler RE, Caon A, Willenborg DO, Clark-Lewis I, and McColl SR, *A role for macrophage inflammatory protein-3 alpha/CC chemokine ligand 20 in immune priming during T cell-mediate inflammation of the central nervous system.* Journal of Immunology, 2003. **170**(12): p. 6298-6306.

86. Gong JH, Ratkay LG, Waterfield JD, and Clark-Lewis, *An antoagonist of monocyte chemoattractant protein 1 (MCP-1) inhibits arthritis in the MRL-lpr mouse model.* Journal of Experimental Medicine, 1997. **186**(1): p. 131-137.

87. Gong JH, Yan R, Waterfield JD and Clark-Lewis I, *Post-onset inhibition of murine arthritis using combined chemokine antagonist therapy* Rheumatology 2004. **43**(1): p. 39-42.

88. Clark-Lewis I, et al., *Chemical synthesis, purification and characterization of two inflammatory proteins, neutrophil activating peptide 1 (interleukin-8) and neutrophil activating peptide 2.* Biochemistry, 1991. **30**(12): p. 3128-3135.

89. Clore GM, et al., *Three-dimensional structure of interleukin 8 in solution.* Biochemistry, 1990. **29**(7): p. 1689-1696.

90. Baldwin ET, et al., *Crystal structure of interleukin 8: symbiosis of NMR and crystallography.* PNAS, 1991. **88**: p. 502-506.

91. Clark-Lewis I, et al., *Structure-Activity Relationships of Interleukin-8 Determined Using Chemically Synthesized Analogs.* Journal of Biological Chemistry, 1991. **266**(December 5): p. 23128-23134.

92. McQuibban GA, Gong JH, Tam EM, McCulloch CA, Clark-Lewis I, and Overall CM, *Inflammation dampened by gelatinase A cleavage of monocyte chemoattractant protein-3.* Science, 2000. **289**(5482): p. 1202-1206.

93. McQuibban GA, G.J., Wong JP, Wallace JL, Clark-Lewis I, and Overall CM, *Matrix metalloproteinase processing of monocyte chemoattractant proteins generates CC chemokine receptor antagonists with anti-inflammatory properties in vivo.* Blood, 2002. **100**(4): p. 1160-1167.

94. Nakamoto C, Behar V, Chin KR, Adams AE, Suva LJ, Rosenblatt M and Chorev M, *Probing the bimolecular interactions of parathyroid hormone with the human parathyroid hormone/parathyroid hormone-related protein receptor. 1. Design, synthesis and characterization of photoreactive benzophenone-containing analogs of parathyroid hormone.*. Biochemistry, 1995. **34**: p. 10546-10552.

95. Gimpl G, Anders J, Thiele C and Fahrenholz F, *Photoaffinity labeling of the human brain cholecystokinin receptor overexpressed in insect cells. Solubilization, deglycosylation and purification.* Eur. J. Biochem, 1996. **237**: p. 768-777.

96. Li, YM, Marnerakis M, Stimson ER, and Maggio JE, *Mapping peptide-binding domains of the substance P (NK-1) receptor from P388D$_1$ cells with photolabile agonists.* . Journal of Biological Chemistry, 1995. **270**(3): p. 1213-1220.

97. Weber PJ, Beck-Sickinger AG, *Comparision of the photochemical behavior of four different photoactivable probes.* Journal of Peptide Research, 1997. **49**(5): p. 375-383.

98. Fenn JB, Mann M, Meng CK, Wong SF, and Whitehouse CM, *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

99. Tanaka K, et al., *Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry.* Rapid Communications in Mass Spectrometry, 1988. **2**(8): p. 151-153.

100. Karas M and Hillenkamp F, *Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons.* Analytical Chemistry, 1988. **57**(20): p. 2299-2301.

101. Ball LE, et al., *Mass spectrometric analysis of integral membrane proteins: Application to complete mapping of bacteriorhodopsins and rhodopsin.* Protein Science, 1998. **7**(3): p. 758-764.

102. Quach TTT, LN, Richards DP, Zheng J, Keller B and Li L, *Development and applications of in-gel CNBr/tryptic digestion combined with mass spectrometry for the analysis of membrane proteins.* Journal of Proteome Research, 2003. **2**(5): p. 543-552.

103. Zhong H, Marcus SL, and Li L, *Microwave-assisted acid hydrolysis of proteins combined with liquid chromatography MALDI MS/MS for protein identification.* American Society for Mass Spectrometry, 2005. **16**: p. 471-481.

104. Zhang N and Li L, *Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation.* Rapid Communications in Mass Spectrometry, 2004. **18**(8): p. 889-896.

105. Norris, JL, Porter NA, and Caprioli RM, *Mass Spectrometry of Intracellular and Membrane Proteins Using Cleavable Detergents.* Anal. Chem., 2003. **75**(23): p. 6642-6647.

106. Barnidge DR, Dratz EA, Jesaitis AJ and Sunner J, *Extraction method for analysis of detergent*

-solubilized bacteriorhodopsin and hydrophobic peptides by electrospray ionization mass spectrometry. Analytical Biochemistry, 1999. **269**: p. 1-9.

107. Wu CC, MacCross MJ, Howell KE, and Yates JR 3rd, *A method for the comprehensive proteomic analysis of membrane proteins.* Nature Biotechnology, 2003. **21**(5): p. 532-538.

108. Soskic V and Godovac-Zimmermann J, *Improvement of an in-gel tryptic digestion method for matrix-assisted laser desorption/ionization-time of flight mass spectrometry peptide mapping by use of volatile solubilizing agents.* Proteomics, 2001. **1**(11): p. 1364-1367.

109. van Montfort BA, Canas B, Duurkens R, Godovac-Zimmermann J and Robillard GT, *Improved in-gel approaches to generate peptide maps of integral membrane proteins with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.* Journal of Mass Spectrometry, 2002. **37**: p. 322-330.

110. Kuster B, Schirle M, Mallick P, and Aebersold R, *Scoring proteomes with proteotypic peptide probes.* Nature Reviews, 2005. **6**(July): p. 577-583.

111. Domon B and Aebersold R, *Mass spectrometry and protein analysis* Science, 2006. **312**: p. 212-217.

112. Blow N, *Antibodies: The generation game.* Nature, 2007. **447**: p. 741-744.

113. Pisitkun T, et al., *Tandem mass spectrometry in physiology.* Physiology, 2007. **22**: p. 390-400.

114. Santoni V, et al., *Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels.* Electrophoresis, 1999. **20**: p. 705-711.

115. Gygi SP, et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nature Biotechnology, 1999. **17**(10): p. 994-999.

116. Kirkpatrick DS, Gerber SA, and Gygi S, *The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications.* Methods, 2005. **35**: p. 265-273.

Chapter II: An efficient C-terminal affinity tag for the purification and comprehensive analysis of membrane proteins-
Application to G protein-coupled receptors and ATP binding cassette transporters

2.1 Introduction

G-protein coupled receptors (GPCRs) represent the largest class of membrane proteins in the human genome and are important in the control of fundamental processes such as vision, olfaction, stress responses as well as nervous system functions. Controlled, coordinated expression, localization and signaling are essential for normal physiology as malfunction leads to disease. Due to these critical implications, GPCRs are popular targets for drug discovery. Although extensive research has been invested in the field, still little is known about the three dimensional structure of these receptors due to inherent difficulties in protein expression and purification [1].

GPCRs are difficult to recover in sufficient quantities from natural sources, with the exception of bovine rhodopsin. To overcome this difficulty, overexpression systems are often implemented. *E. coli* and yeast systems are commonly employed to improve GPCR quantity, but generally do not produce proteins with normal post-translational modifications or folding [2, 3]. Alternatively, low expression levels may be addressed with enrichment approaches such as affinity purification to concentrate protein to suitable levels for experimental analysis.

A general enrichment strategy for a more comprehensive survey of GPCRs expressed in physiologically relevant contexts would be beneficial to understanding the receptor of interest as well as resolve experimental artifacts produced by heterologous expression systems. Affinity

A version of this chapter will be submitted for publication. Co-authors in no particular order: Wong JP, Kast J, Molday RS, and Reboulle E.

tags are one such general purification strategy for proteins that cannot be isolated conveniently via specific intrinsic properties. However, selecting an appropriate affinity purification system for a membrane protein of interest is not trivial and may require extensive detergent screening and optimization [4].

Protein enrichment may be combined with sensitive detection techniques such as mass spectrometry to further improve analysis of low GPCR expression in more physiologically relevant cellular contexts [5]. Targeted approaches such as Multiple Reaction Monitoring (MRM) and proteotypic peptide searches [6-9] improve sensitivity through selective detection of pre-determined signature peptides and fragment ions characteristic for the protein of interest. However, the applicability of targeted analysis is contingent on the experimental information available in existing databases and repositories.

Targeted analysis focuses on experimentally established peptide ions, rather than entire predicted theoretical protein sequences which are often irrelevant and inefficient in the analyses. For targeted approaches to work at full potential for improved sensitivity and specificity, repositories must be populated with detailed and accurate peptide as well as associated fragmentation patterns which are often underrepresented for GPCRs due to challenges associated with protein hydrophobicity, solubility and enzyme accessibility.

A versatile GPCR enrichment strategy that involves a membrane protein compatible 1D4 affinity tag derived from bovine rhodopsin will be described in this investigation. The 1D4 tag specific monoclonal antibody (mAb) was mapped to the absolute bovine rhodopsin C-terminal sequence TETSQVAPA. The 1D4 enrichment strategy is applicable to GPCRs and other membrane proteins such as ATP binding cassette (ABC) transporters employing one common

detergent and wash conditions. In this study, CXCR4-1D4, CCR5-1D4, ABCA1-1D4 and ABCA4-1D4 were enriched for comprehensive mass spectrometric characterization. Enzymatic combinations for each of the membrane proteins that produce high sequence coverage were determined. High sequence coverage is critical for confident protein identification in both hypothesis-driven as well as global analyses.

2.2 Material and methods

Construction of 1D4 tagged receptor constructs and site directed mutagenesis:

Restriction endonuclease recognition sites and the C-terminal 1D4 tag (TETSQVAPA) were introduced into the primers used in the polymerase chain reaction (PCR) with wild type CXCR4 in PCDNA 3.1+ as the template. The PCR reactions were carried out for 35 cycles with 60°C annealing temperature with *Pfu* polymerase (Fermentas, Glen Burnie, MD), forward and reverse primer sequences are as follows: *cggaattcgccgccatggaggggatcagtatatac* and *cgctcgagttaggcaggcgccacttggctggtctctgtgctggagtgaaaacttga*, respectively. The 1D4 sequence is underlined. The PCR product was ligated into EcoRI and XhoI in the pcDNA 3.0+ multiple cloning site. The CXCR4 and CCR5 templates used were generous gifts from Dr. Bernhard Moser (Theodor-Kocher Institute; Bern, Switzerland) and the ABCA1-1D4 and ABCA4-1D4 constructs were generated as previously described [10].

N-linked glycosylation deletion mutants were generated by site directed mutagenesis following the QuikChange kit (Stratagene; La Jolla, CA) with the CXCR4-1D4 pcDNA 3.0+ construct as the template. The following primers were used to mutate each of the asparagine residues to alanine: N11A (forward primer *agtatatacacttcagatgcctacaccgaggaaatgggc*) and N176A (forward primer *cccgacttcatctttgccgccgtcagtgaggcagatgac*). The sites of NΔA are underlined. The double mutant N11/176A was created by using the N11A construct as a template and mutating site N176 with the N176A primers with QuikChange kit.

Cell culture and transfection:

Human Embryonic Kidney (HEK) 293T cells (ATCC, Manassas, VA) were cultured in Dulbecco's modified eagle media (DMEM) (Invitrogen, Carlsbad CA) and 10% fetal bovine serum (FBS). A confluent 20mm dish of 293T cells were seeded 1 in 10 dilution 12 h prior to calcium phosphate mediated transfection. CXCR4-1D4/pcDNA 3.0 (20μg), CCR5-1D4/pcDNA 3.0 (20μg), ABCA1-1D4/pCEP4 (20μg) or ABCA4-1D4/pCEP4 (20μg) were transfected into one 20mm dish of cells. Twelve hours post-transfection the cells were gently rinsed with phosphate buffered saline (PBS), fresh media was replaced and incubated at 37°C in CO_2 overnight before harvest.

Cell lysis:

One 100mm x 20mm dish of transiently tranfected 293T cells was solubilized in 1ml of 2% Triton X-100 in TBS for 1h at 4°C with stirring and 1X complete protease inhibitors (Roche, Mannheim, Germany). Insoluble material and nuclei were pelleted at 200, 000 x g in a TLA.3 Beckman rotor for 15min. The resultant solubilized cells were subjected to affinity purification.

Membrane fragment preparation:

ABCA1 and ABCA1-1D4 transfected cells were resuspended and washed twice in a 10 mM Tris Buffer pH 7.5 supplemented with 0.5 mM EDTA. Pellets were then resuspended in the same buffer with protease inhibitors, lysed for 1h on ice vortexing occasionally, and homogenized with a glass homogenizer. This mixture was loaded on top of a gradient made with

5% sucrose and 60% sucrose in a 20 mM Tris buffer pH 7.4 supplemented in 150 mM NaCl, 1mM $MgCl_2$, 1mM$CaCl_2$ and 0.1 mM EDTA, and centrifuged for 30 min at 24,000 rpm in a SW27.1 swinging bucket rotor. Membranes which banded on top of the 60% sucrose fraction were harvested, washed once and centrifuged for 15 min at 40,000 rpm. Membranes were then stored in the same buffer supplemented with protease inhibitors and 10% glycerol at -30°C prior to solublization.

Membrane fragments prepared from 10 plates, approximately $5x10^8$ transiently transfected cells, were solubilized in 1ml of 2% Triton X-100/TBS with protease inhibitors for 1h with stirring at 4°C. The insoluble material was removed by centrifugation at 100,000 x g for 10 min.

Immunoprecipitation and SDS-polyacrylamide gel electrophoresis:

The cell lysate or solubilized sample was added to 50µl of 1D4 monoclonal antibody (mAb) conjugated Sepharose prepared as previously described [11] with complete protease inhibitors (Roche, Mannheim, Germany) and rotated for 1h at 4°C. Beads were washed 10 times with 0.1% Triton X-100/TBS. Tagged receptors were eluted 2 times with 100µl 0.5mg/ml 1D4 peptide in 0.1% Triton X-100 and one final time with 1% SDS/TBS at room temperature with shaking. SDS sample loading buffer (1% SDS, 10% glycerol, 0.1M TrisHCl pH 6.8, bromophenol blue tracking dye) was added to samples and loaded on a 10% SDS polyacrylamide gel with a 5% stacking gel. Corresponding western blots were performed with these samples on polyvinylidene fluoride membranes (Millipore, Bedford, MA) for antibody staining with 1D4

monoclonal antibody and goat anti-mouse IgG IR Dye 680 secondary (LI-COR, Lincoln, NB) followed by LI-COR Odyssey infrared imaging (Lincoln, NB).

Concavalin A western blot probing:

N11A CXCR4-1D4, N176A CXCR4-1D4, N11/176A CXCR4-1D4 and CXCR4-1D4 were transferred onto PVDF and analysed by lectin binding with concavalin A (ConA). The membrane was blocked at 37°C in 5% bovine serum albumin (BSA) in PBS with 0.01% Tween-20 (PBST) and then bound with biotinyl succinylated ConA (Vector Laboratories, Burlingame, CA) in blocking buffer. After 3 PBST washes, the blot was labeled with streptavidin 680 IR dye in (LI-COR, Lincoln, NB) in 5% BSA, 0.02% SDS in TBS and visualized on the LI-COR infrared imager.

In-gel digestion:

Coomassie Blue stained gel bands were washed three times 15min each with 25mM ammonium bicarbonate/50% acetonitrile followed by reduction with 1.5 mg/ml dithiothreitol (DTT) at 50°C for 30min and alkylation with 10mg/ml iodoacetamide (IAA) at room temperature for 20min. Gel bands were rehydrated with 37.5ng/ml sequencing grade trypsin and/or chymotrypsin (Roche, Mannheim, Germany) with 5-10 mM $CaCl_2$ in 50mM ammonium bicarbonate (ABC) on ice for 30min. Excess enzyme was removed and 15µl of additional digestion buffer was added to prevent gel drying during digestion. The gel bands were incubated at 37°C with shaking at 450 rpm for 18 h. Digests were extracted with 1:1 50mM ammonium

bicarbonate/acetonitrile and the final extraction was carried out with 1:1 5% formic acid/acetonitrile. Extracts were pooled and lyophilized to dryness and stored at -80°C until analysed. For double digests, the initial enzymes or chemicals employed were removed. The remaining gel band was dried with one volume of acetonitrile and speed vacuum centrifugation prior to rehydration with the second enzyme and subsequently extracted as described in the previous.

Cyanogen bromide and trypsin double digestion:

Gel bands were washed, reduced and alkylated as indicated above. The cyanogen bromide (CNBr) solution was prepared in a chemical fumehood, one crystal of CNBr (Aldrich, Milwaukee, Wisconsin) was dissolved in 100µl of 50% TFA. CNBr (20µl) was used to rehydrate gel bands which were kept in the dark overnight

LC-MS/ MS conditions:

Trypsin-digested peptides were base and then acid extracted from the gel pieces, pooled and lyophilized. Samples were reconstituted in 6µl of formic acid (FA) and injected onto the Qstar XL LC/MS/MS (Applied Biosystems, Foster City, CA). A PepMap C18, 3µm particle size and 100A pore size column from LC packings (Amsterdam, Netherlands) was used for peptide separation. Solvents B and A contained, 20% acetonitrile (ACN) and 5% ACN, respectively. LC conditions started at 2% solvent B with a gradient to 60% B over 60 min, to 95% B at 63 min and held for three minutes before returning to 2%B.

Confocal microscopy:

Transfected monkey COS-7 cells were grown onto polylysine coated cover slips and then fixed with 4% paraformaldehyde (Fisher; Fair Lawn, New Jersey) for 20 min at room temperature. Cells were blocked with normal goat serum and labeled with 1/50 dilution of 1D4 antibody for 1h and subsequently labeled with anti-mouse Alexi 488 secondary antibody. The sections were counter stained for nuclei with 4', 6-diamidino-2-phenylindole (DAPI).

Calcium Flux assays:

CXCR4 and CXCR4-1D4 ectopically expressed on 300-19 pre-B cells ($2x10^6$cells) were harvested and loaded with Fluor-3AM fluorescent dye and kept in the dark at 37°C until assay. Cells ($1 x10^6$) were resuspended in 2.5 ml 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 140 mM NaCl, 10 mM glucose, 1.8 mM $CaCl_2$, 1 mM $MgCl_2$ and 3 mM KCl, pH 7.3. Cells were stirred in a Horiba Jobin FL3-22 TAU (Edison, New Jersey) spectrofluorimeter at 506 nM excitation and 526 nM emission wavelength. After approximately 20 s baseline stabilization 100nM SDF-1 was added to the cells and resultant fluorescence emission and calcium flux was recorded for 60 s time-based acquisition.

ABCA4 and ABCA4-1D4 reconstitution of proteins in liposomes:

The lipids used for liposome formation were reconstituted in 25 mM HEPES pH 7.5, 140 mM NaCl, 3 mM $MgCl_2$ and 10% glycerol. The mixture was sonicated for 15min at room temperature. The lipid stock (9 µl) was mixed with 15% n-octylglucoside (6 µl) and purified

ABCA4 or ABCA4-1D4 (100 µl) by slow addition. The resultant mixture was incubated 30 min on ice. The detergent was removed from this mixture by passing through equilibrated Extracti-gel D resin (Pierce, Rockford, IL) and the flow through was collected. The reconstitution buffer consisted of 25 mM HEPES pH 7.5, 140 mM NaCl, 10% glycerol, 3 mM $MgCl_2$, 1 mM EDTA and 1mM DTT.

ATPase assays:

ATPase assays were performed in triplicate for each point determination using purified protein reconstituted into liposomes. The reaction was initiated with the addition of 1 µl of a mixture of 0.5 mM ATP, 5-10 µCi [α^{32}P]-ATP in reconstitution buffer to the prepared liposomes and incubated at 37°C for 30 min. An aliquot of each reaction mixture was collected after 0, 5, 10, 20 and 30 min and spotted on polyethyleneimine cellulose (Aldrich, Milwaukee, WI). The hydrolysed ATP was separated from the unhydrolysed ATP by thin layer chromatography (TLC) in 0.5 M LiCl/ 1 M formic acid. The TLC plate was air dried and exposed to a phosphor screen for 3h. The ABCA4 or ABCA4-1D4 content of the liposomes was assessed by Coomassie Blue gel and densitometry.

2.3 Results

C-terminally tagged receptors express and demonstrate accurate cellular localization:

COS-7 cells were transiently transfected with the CXCR4-1D4, CCR5-1D4, ABCA1-1D4 and ABCA4-1D4 constructs and labeled for confocal microscopy to visualize receptor expression and localization. CXCR4-1D4, CCR5-1D4 and ABCA1-1D4 expression was

observed along the plasma membrane (Figures 2-1A, 2-1B and 2-1C respectively). In contrast ABCA4-1D4 was observed in vesicles within the cytosolic compartment of the transfected COS-7 cells (Figure 2-1D). Ectopically expressed bovine rhodopsin was used as a positive control for the 1D4 monoclonal antibody (Figure 2-1E). The identity of each purified receptor was confirmed in independent experiments described in the following sections.

The 1D4 tag does not alter receptor function:

To determine the effect of the 1D4 tag on CXCR4 function, murine B300-19 pre-B cells which are commonly used to study GPCR activity and signal transduction, were transfected with either the CXCR4 or CXCR4-1D4 construct. The cells were loaded with Fluo-3AM calcium binding fluorescent dye indicator to monitor phospholipase gamma signal transduction and intracellular calcium mobilization. Receptor activation upon chemokine binding resulted in increased flourescence signal which is indicative of increased intracellular calcium levels detected by Fluo-3AM binding. Both CXCR4-1D4 and CXCR4 showed similar fluorescence signal response times. Maximum fluorescence was observed at 40 s with the addition of 100nM stromal cell derived factor 1 (SDF-1) at 18 s for both CXCR4 and CXCR4-1D4 (Figure 2-2A, 2-2B). Addition of SDF-1 at 18 s to a final concentration of 100nM resulted in 450,000 and 350,000 cps change in fluorescence for CXCR4 and CXCR4-1D4 transfected cells, respectively (Figure 2-2A, 2- 2B). When the same volume of PBS was introduced at 18 s, baseline fluorescence was observed at 40 s and beyond (Figure 2-2C, 2-2D).

The activity of ABCA4-1D4 was measured with ATP hydrolysis assays. Both ABCA4 and ABCA41D4 showed a similar increase in the amount of ATP hydrolysed over the 30min

period. The activity after 30 min was 265 ± 3.65 nmol and 248 ± 47.35 nmol ATP/mg purified protein for ABCA4 and ABCA4-1D4, respectively (Figure 2-3).

The 1D4 affinity achieves high recovery for membrane proteins:

The 1D4 tag was used to enrich for CXCR4-1D4 expressed transiently in a human cell line. The purification was separated on polyacrylamide gel for western blotting, densitometry and protein recovery determination. Bovine rhodopsin from rod outer segments (ROS) was employed as a standard. Using one constant buffer and detergent in a standard 1D4 enrichment protocol applied to different membrane proteins, 50%-60% of receptor was recovered. To determine the peptide concentration necessary for maximum receptor recovery, the 1D4 peptide epitope was titrated for CXCR4-1D4 and ABCA1-1D4 elution from the immunoaffinity column. Optimum peptide concentration for receptor elution determined from densitometry was 0.25 mg/ml for both CXCR4-1D4 and ABCA1-1D4 (Figure 2-4). At concentrations beyond 0.25 mg/ml, the dose response curve started to plateau while lower concentrations resulted in significantly reduced receptor recovery (Figure 2-4). Similarly, complete removal of Triton X-100 from the peptide elution also reduced receptor recovery whereas improved receptor yield was obtained with 0.1% and 0.2% Triton X-100 (Figure 2-5A, 2-5B).

The remaining receptor on the immunoaffinity beads was recovered from a final non-specific elution with 1% (v/v) SDS. From a single purification of $5x10^7$cells transiently expressing CXCR4-1D4, approximately 500 ng of protein from peptide elution and another 500 ng from the final SDS elution was recovered, corresponding to approximately 90% total protein recovery. On average 1000 ± 20 ng and 300 ± 50ng of CXCR4-1D4 and CCR5-1D4 was

recovered from 1D4 purification of 10^8 transiently transfected 293T cells, calculated from 3 independent experiments.

The 1D4 tag is specific:

The 1D4 tagged receptors were effectively separated from other cellular proteins as shown in the reduction of bands in the peptide eluate relative to the pre-column lane of the whole cell lysate lane 1 (Figures 2-6(A-E). CXCR4-1D4 was observed as a broad band that extended from ~40-50kDa and a lower molecular weight band at ~30kDa. The identity of these bands was confirmed by *in situ* digestion and mass spectrometry as discussed below. Two bands were isolated for CCR5-1D4, 3 bands for ABCA1-1D4 and 4 bands for ABCA4-1D4 and MS/MS sequenced.

Mass spectrometry and sequence coverage

Due to the natural abundance of rhodopsin from rod outer segments in bovine eyes, electrophoresis alone was adequate for separation and subsequent mass spectrometric analysis. Enriched CXCR4-1D4, CCR5-1D4, ABCA1-1D4 and ABCA4-1D4 were cleaved with different enzymatic combinations for MS/MS sequencing.

The gel bands indicated by the red arrows were excised for digestion and mass spectrometry (Figure 2-6). Trypsin digestion of CXCR4, CCR5 and rhodopsin yielded few peptides: 1, 2 and 3 respectively (Table 2-1). Additional peptides for these GPCRs were identified with other enzymatic combinations. The greatest number of unique peptides detected

was from a chymotrypsin/trypsin double digestion for CXCR4 and rhodopsin which yielded 4 and 6 unique peptides, respectively (Table 2-1). CNBr/trypsin digestion yielded 6 unique peptides for CCR5 (Table 2-1). Unlike the GPCRs, a single tryptic digestion of the ABC transporters identified 24 and 6 unique peptides for ABCA1 and ABCA4, respectively.

With these enzymatic combinations we were able to achieve the top sequence coverage and match score in the Global Proteome Machine database (www.thgpm.org) from 101 submitted entries for human CXCR4 with a top score of -142, followed by -42 for the next entry (data not shown). The scoring represents the logarithm of the probability in which the identification is a false positive. Thus, a peptide score of -2 has a 1/100 chance of being a false positive [12]. No entries were available for human CCR5 in the database (data not shown).

CXCR4 post-translational modifications

A peptide of CXCR4 that contained the NX(S/T) consensus sequence for N-linked glycoylation as determined from computational predictions performed by NetNGlyc 1.0 server (linked from http://www.expasy.org) was sequenced by tandem mass spectrometry. According to the prediction software, there are two putative N-linked glycosylation sites at positions 11 and 176. While mass spectrometric sequencing of peptide $AN_{176}VSEADDRY$ obtained in a chymotrypsin and trypsin sequential double digestion (Table 2-1) identified the unmodified asparagine at position 176, no peptide containing N11 could be observed by mass spectrometry. To further confirm that failed detection was not due to the absence of the unmodified peptide and that residue N11 was in fact glycosylated as indicated in previous reports [13, 14], we examined

these sites of post-translational modification with site-directed mutagenesis as well as western blotting.

N11A, N176A and N11/176A CXCR4-1D4 mutant constructs were transiently transfected into human 293T cells and analyzed on a western blot probed with Rho1D4 mAb and LICOR goat anti-mouse 680 secondary antibody. When position 11 was mutated, the protein migrated at a lower molecular weight (Figure 2-7A lane 2). Mutation at position 176 did not change the migration relative to the tagged wild type receptor. The double mutant N11/176A migrated at the same molecular weight as the N11A mutant (Figure 2-7A, lane 4).

In addition, duplicate lanes of the N11A, N176A and N11/176A CXCR4-1D4 mutant samples were run on the same polyacrylamide gel and transferred onto the same polyvinylidene fluoride (PVDF), probed with succinyl-biotinylated ConA and subsequently streptavidin-680IRDye (LI-COR) to detect mannose containing sites on CXCR4 (Figure 2-7B). A sharp Con A positive band migrated at approximately 40kDa was observed for CXCR4-1D4 and the N176A mutant corresponding to a similar smaller species band for CXCR4-1D4 detected in the 1D4 probed blot (Fig 2-7B, lanes 1 and 3). However, this band was absent for the N11A and the N11/176A double mutants (Fig 2-7B, lanes 2 and 4).

2.4 Discussion

1D4 enrichment is a general approach for membrane protein characterization and protein database population:

1D4 affinity enriched CXCR4-1D4, CCR5-1D4, ABCA1-1D4 and ABCA4-1D4 separated on polyacrylamide gel, Coomassie Blue stained gel bands were subjected to *in-gel* digestion and mass spectrometric sequencing.

The mass spectrometer is capable of collecting an overwhelming amount of data and only a biased fraction of the data is actually selected for subsequent fragmentation in complex samples. Peptides with specific biophysical properties and peptides of high abundance [7, 15] are often chosen ignoring other important peptides. These preferential peptides are redundantly sequenced reducing the acquisition space available for other unique peptides which affects the accuracy and efficiency of protein identification. However, implementation of targeted approach such as MRM for biomarker identification will refine acquisition parameters to select specific peptides for a given biomarker for MS/MS analysis improving detection sensitivity in complex mixtures. Targeted analysis relies on the previous knowledge of signature fragment ion information for peptides detected which may be available in annotated peptide libraries. For these libraries to be effective and efficient for protein searches, they must be populated with accurate biochemical and mass spectral data. Existing protein libraries are often limited with membrane protein information due to intrinsic challenges related to protein purification. However, we have demonstrated that the 1D4 membrane protein enrichment strategy is versatile, comprehensive and efficient for membrane protein mass spectrometric characterization for library population.

The short 1D4 sequence is functionally compatible with GPCRs and other membrane proteins:

The 1D4 tag represents the C-terminal 9 amino acids on bovine rhodopsin and is a suitable choice for studying other GPCRs and other membrane proteins due to structural and functional similarities which would not likely alter protein localization and function. We have shown that this approach is applicable to several GPCRs as well as other multi-pass membrane proteins such as ABC transporters. When the 1D4 sequence is attached onto the C-terminal end

of CXCR4, CCR5, ABCA1 and ABCA4, localization and functional properties (Figures 2-1, 2-2 and 2-3 respectively) were not perturbed.

CXCR4-1D4, CCR5-1D4 and ABCA1-1D4 demonstrate accurate cellular localization to the plasma membrane; ABCA4-1D4 localized to intracellular vesicles mimicking localization of ABCA4 to the intracellular discs in the rod outer segments [16] (Figure 2-1C). Moreover, CXCR4-1D4 stimulation with 100nM SDF-1 induced calcium flux levels that are comparable to wild type CXCR4 (Figure 2-2A, 2-2B). Both receptors show similar response times to chemokine addition and the absolute change in fluorescence were both on the order of 400, 000 cps. The variation in fluorescence intensity between CXCR4 and CXCR4-1D4 upon SDF-1 addition observed is semi-quantitative and most likely reflects the quantity of receptors expressed in the cell line which is subject to differences in expression levels and transfection efficiency. To ensure that the fluorescence response observed upon SDF-1 introduction were the result of SDF-1 binding and CXCR4 signal transduction, equivalent volumes of PBS were added to the cells and no change in fluorescence and thus activity was observed. Similarly the 1D4 tag does not affect the ATPase activity of a representative ABC transporter, ABCA4.

Enrichment specificity and sensitivity of approach enables analysis of receptors expressed in physiologically relevant contexts:

Functional expression of the membrane receptors is important for accurate downstream mass spectrometric analyses. GPCRs display low endogenous expression and solubility often depends on the functional complexity of the protein of interest. Receptor isolation is also complicated by heterogeneous glycosylation on the cell surface. The specificity of the 1D4

approach effectively enriches for the tagged receptor of interest for sensitive detection by mass spectrometry.

We have optimized the enrichment procedure for one common detergent including the wash conditions and have acquired Coomassie Blue stained levels for all of the receptors purified from as little as one dish or approximately $5x10^7$ human 293T transiently transfected cells. The high level of protein enrichment achieved with our purification is attributed to the specificity and mild elution conditions. Mild elution conditions minimize receptor loss through protein precipitation and aggregation which may occur with heating or lowering pH. Peptide competitive elution is specific, gentle and non-denaturing, reducing non-specific protein and antibody leaching from the affinity column.

MS data acquisition for low quantities of the protein of interest is compromised by abundant proteins such as cytokeratins and external keratins introduced during sample preparation. In addition, other contaminating proteins suppress peptide signals for the protein of interest and affect identification. The specificity of our purification strategy has enabled us to take advantage of the full potential of mass spectrometry. CXCR4-1D4, CCR5-1D4, ABCA1-1D4 and ABCA4-1D4 were MS/MS sequenced with similar sequence coverage to highly abundant bovine rhodopsin which is virtually pure in bovine rod outer segments (Table 2-1) [17]. Peptide elution is specific and efficient. From a single bed volume of peptide epitope solution approximately 50% of the tagged receptor is recovered at room temperature, repeated elution recovers an additional 40% of the remaining receptor.

Mapping receptor post-translational modifications:

Specific enrichment of diluted receptors expressed in physiologically relevant contexts for GPCRs or natural sources to experimentally feasible concentrations is advantageous for further biochemical analysis. Heterologous expression systems are commonly employed to improve protein levels but accurate protein folding, post-translational modifications and function may not be possible in foreign host systems. Comprehensive analysis of post-translational modifications for mammalian receptors expressed in a mammalian system are important for resolving artifacts observed from heterologous expression as well as for subsequent MS mapping and targeted disease profiling in clinical proteomic applications.

Thorough mass spectrometric sequence coverage and efficient protein identification for rhodopsin was only made possible from prior knowledge of the PTM sites [18-20] and has been since included in existing databases. We have selected human CXCR4 as our model GPCR for detailed post-translational modification characterization because of disease implications such as AIDS and cancer metastasis. The specific isoforms that are involved in these various diseases are not well understood. CXCR4 as a biomarker for disease diagnosis or profiling is not possible unless the corresponding PTMs on the receptor are characterized in the healthy and disease conditions for accurate protein identification and relative quantification.

We have observed from tandem mass spectrometry, site-directed mutagenesis and lectin binding that N11 was the only N-linked glycosylated site on CXCR4. Mutation on N176 does not affect the migration size as detected on a 1D4 western blot and ConA binding on the N176A/CXCR4-1D4 mutant indicating that only position 11 is N-linked glycosylated (Figure 2-

7). This is also consistent with tandem mass spectra observed for a peptide containing unmodified N176.

CXCR4 heterogeneity relates to differing degrees of glycosylation on N11 and not N176 as N11 is likely the only site for N-linked glycosylation. Most GPCRs are predominantly glycosylated at the second extracellular loop and less often on the N-terminus according to the GPCR database (http://www.gpcr.org). It is reported that approximately 66% of receptors with at least one glycosylation occur at extracellular loop (ECL) 2, 14% ECL1 and 20% ECL3 not on the N-terminal [21]. We observed unmodified N176 in a sequenced peptide which was not theoretically predicted for CXCR4 as this ECL2 site contains a glycosylation consensus sequence.

Improving biomarker profiling and targeted approaches:

In silico digestion of the primary sequence of CXCR4 and CCR5 receptor indicates that several peptides are possible from trypsin digestion for CXCR4 and CCR5; however, a small proportion of these peptides are experimentally detected. From MS/MS sequencing of the receptors, we have determined enzymatic combinations that provided several unique peptides for the enriched receptors. This information was subsequently entered into thegpm annotated database. We have observed that trypsin, which is commonly used in proteomic studies alone, did not produce a suitable number of peptides for mass spectrometric identification, especially for the GPCRs we have studied.

The 1D4 tag has enabled us to comprehensively study the best enzyme combinations for GPCR MS detection. We have determined enzyme combinations that provide the greatest

number of peptides that can be sequenced for each of the receptors. Applying this knowledge to global analyses would improve GPCR biomarker identification by increasing the number of possible peptides for detection as well as identify proteotypic peptides to target.

This information in turn was used to populate protein information databases to improve protein identification. To stress this point, we have achieved the top sequence coverage for CXCR4 from 101 entries submitted to thegpm annotated protein database, whereas no entries are currently available for CCR5. This data was collected on the LC-Qstar instrument, we have also achieved further sequence coverage on a more sensitive instrument the Fourier Transform ion cyclotron resonance mass spectrometer.

1D4 enrichment for classical proteomic studies:

The 1D4 enrichment method outlined opens new avenues not only for the study of PTMs and filling gaps for membrane proteins in databases for improved protein identification but may also help identify constant, soluble regions on the receptor for high resolution analysis. As an affinity tag, the specificity and mild elution conditions for the 1D4 tag may also be applied to classical proteomic analyses such as novel interacting protein identification for mammalian receptors expressed in mammalian cells. In addition, understanding the enzyme accessibility for each of the receptors investigated would also facilitate ligand-receptor or small molecule-receptor MS fingerprinting or footprinting studies.

Table 2-1: Sequence coverage for membrane proteins employing different enzymatic combinations for *in-gel* digestion.

Receptor	Enzyme(s)	Number of unique peptides	% Sequence coverage (proportion of residues)	Sequenced peptides
CXCR4	Trypsin	2	7.4% (26/352)	^{134}YLAIVHATNSQRPR^{148} ^{309}TSAQHALTSVSR^{322}
	Chymotrypsin/trypsin	4	16% (55/352)	^{134}YLAIVHATNSQRPR^{148} ^{171}TIPDFIFANVSEADDR^{183} ^{271}QGCEFENTVHK^{282} ^{307}FKTSAQHALTSVSR^{322}
CCR5	Trypsin	1	3% (12/352)	^{322}CCSIFQQEAPER^{334}
	CNBr/trypsin	6	21% (75/352)	^{49}LVILILINCKR^{60} ^{126}YLAVVHAVFALK^{138} ^{210}VICYSGILKTLLR^{223} ^{287}THCCINPIIYAFVGEK^{303} ^{303}FRNYLLVFFQK^{314} ^{322}CCSIFQQEAPER^{334}
Rhodopsin	Trypsin	3	11.5% (40/348)	^{135}YVVCKPMSNFR^{147} ^{231}EAAAQQQESATTQK^{245} ^{296}TSAVYNPVIYIMMNK^{311}
	CNBr/trypsin	3	11% (38/348)	^{126}SLVVLAIER^{135} ^{296}TSAVYNPVIYIMMNK^{311} ^{325}NPLGDDEASTTVSK^{339}
	Chymotrypsin/trypsin	6	21% (72/348)	^{21}SPFEAPQYY^{30} ^{146}RFGENHAIMGVAF^{159} ^{228}TVKEAAAQQQESATTQ^{244} ^{276}IFTHQGSDFGPIFMTIPAFF^{294} ^{325}NPLGDDEASTTVSK^{339}

Receptor	Enzyme(s)	Number of unique peptides	% Sequence coverage (proportion of residues)	Sequenced peptides
ABCA1	Trypsin	24	13% (299/2261)	^{230}SNMDILKPILR^{241} ^{320}SLNWYEDNNYK^{331} ^{382}ILYTPDTPATR^{393} ^{500}FMECVNLNK^{509} ^{862}SYWFGEESDEK^{873} ^{896}LGVSIQNLVK^{906} ^{1039}KLSVALAFVGGSK^{1052} ^{1052}VVILDEPTAGVDPYSR^{1068} ^{1082}TIILSTHHMDEADVLGDR^{1100} ^{1108}LCCVGSSLFLK^{1119} ^{1119}NQLGTGYYLTLVK^{1132} ^{1133}DVESSLSSCR^{1143} ^{1250}VAEESGVDAETSDGTLPAR^{1269} ^{1303}ETDLLSGMDGK^{1314} ^{1323}LTQQQFVALLWK^{1335} ^{1490}QNTADILQDLTGR^{1503} ^{1839}FVSPLSWDLVGR^{1851} ^{1884}LSPLNDEDEDVR^{1896} ^{1901}ILDGGGQNDILEIK^{1915} ^{2023}VGEWAIR^{2030} ^{2081}FLWNCALSVVK^{2092} ^{2134}FGDGYTIVVR^{2144} ^{2145}IAGSNPDLKPVQDFFGLAFPG SVLK^{2169} ^{2189}IFSILSQSK^{2198}
ABCA4	Trypsin	6	3% (70/2273)	^{992}DIETSLDAVR^{1002} ^{1556}YGGISIGGK^{1565} ^{1911}EPIVDEDDDVAEER^{1925} ^{1945}IYPGTSSPAVDR^{1957} ^{1983}MLTGDTTVTSGDATVAGK^{2001} ^{2049}VANWSIK^{2056}

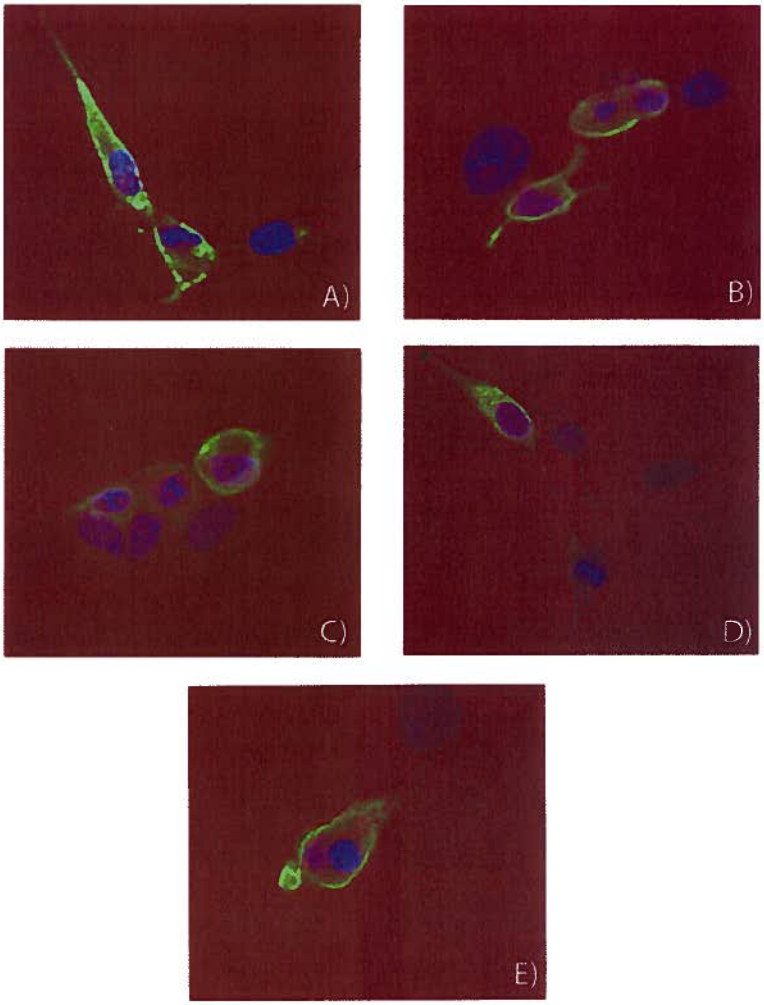


Figure 2-1: 1D4 C-teminally tagged receptors express and demonstrate accurate localization. Confocal microscopy slices of COS-7 cells transiently transfected with A) CXCR4-1D4, B) CCR5-1D4, C) ABCA1-1D4, D) ABCA4-1D4 and E) rhodopsin. Cells were fixed with paraformaldehyde and stained with mouse 1D4 mAb, anti-mouse IgG Alexi 488 (green) and DAPI nuclear staining (blue).

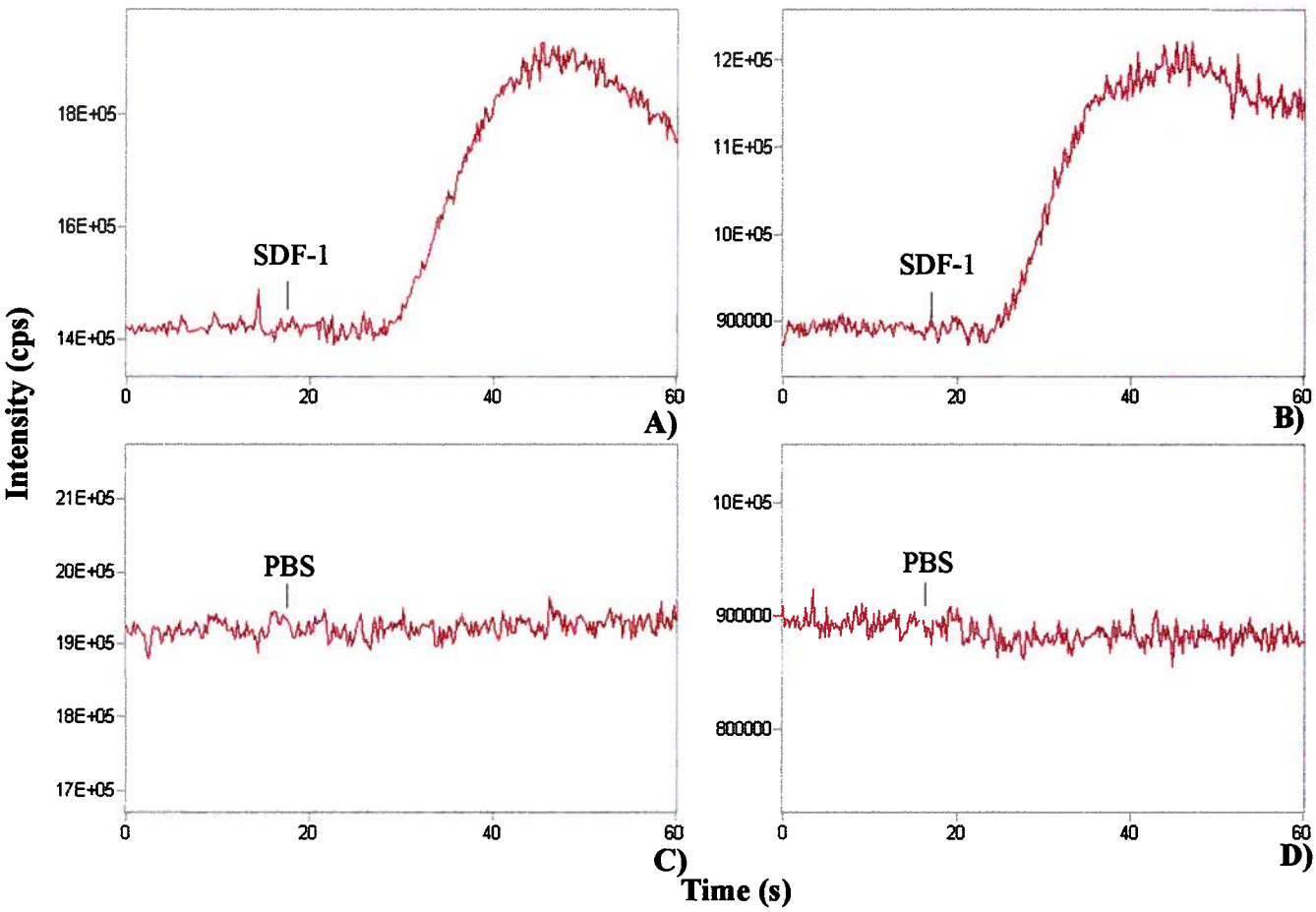


Figure 2-2: CXCR4-1D4 is functional and induces calcium mobilization upon SDF-1 binding. Calcium mobilization and PLC-γ activation was detected in A) CXCR4 and B) CXCR4-1D4 expressing B300-19 cells loaded with Fluo-3AM calcium indicator fluorescent dye. Corresponding controls with equivalent volumes of PBS alone were added to C) CXCR4 and D) CXCR4-1D4 cells as a negative control.

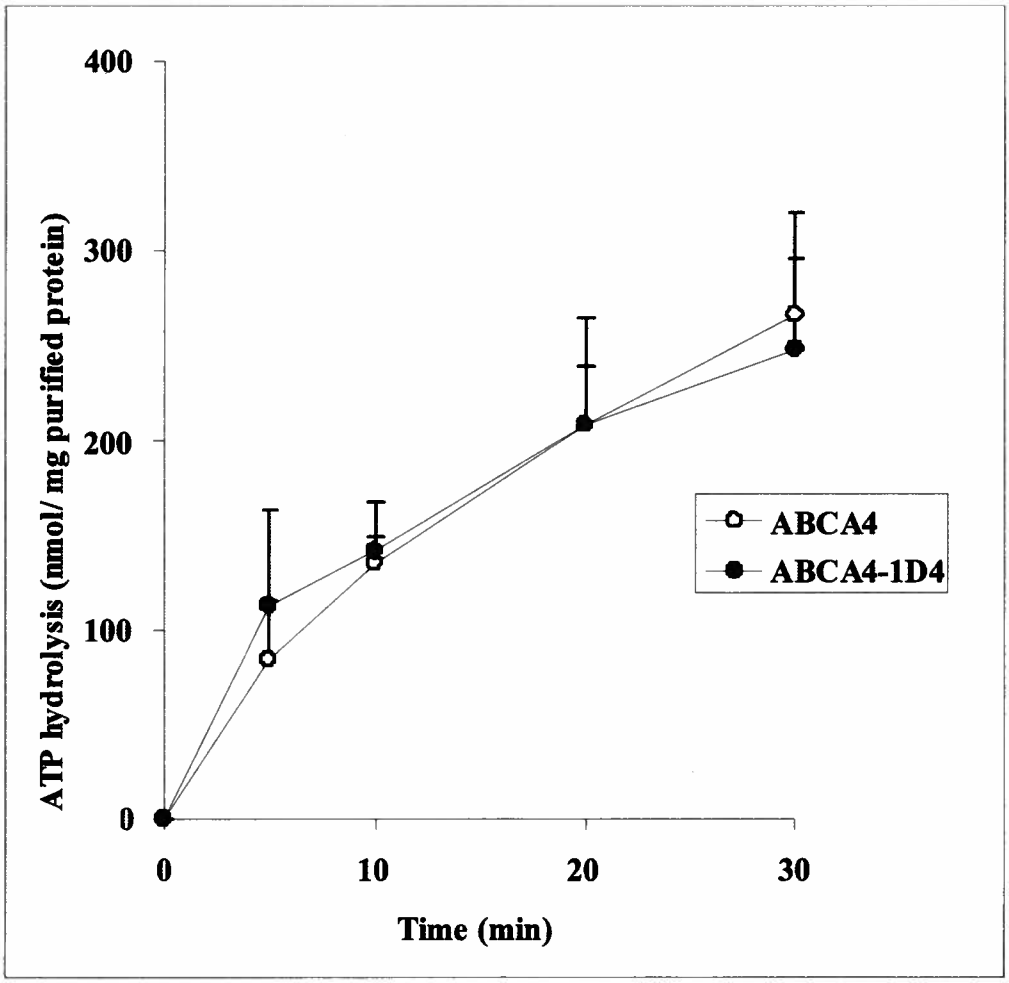


Figure 2-3: Purified ABCA4 (○) and ABCA4-1D4 (●) were reconstituted in brain polar lipid liposomes for ATPase activity assay. ATPase activity was monitored with the release of radiolabelled ADP from [α32P]-ATP over the course of 30 min at 37oC. The average and standard deviation from 3 experiments are shown.

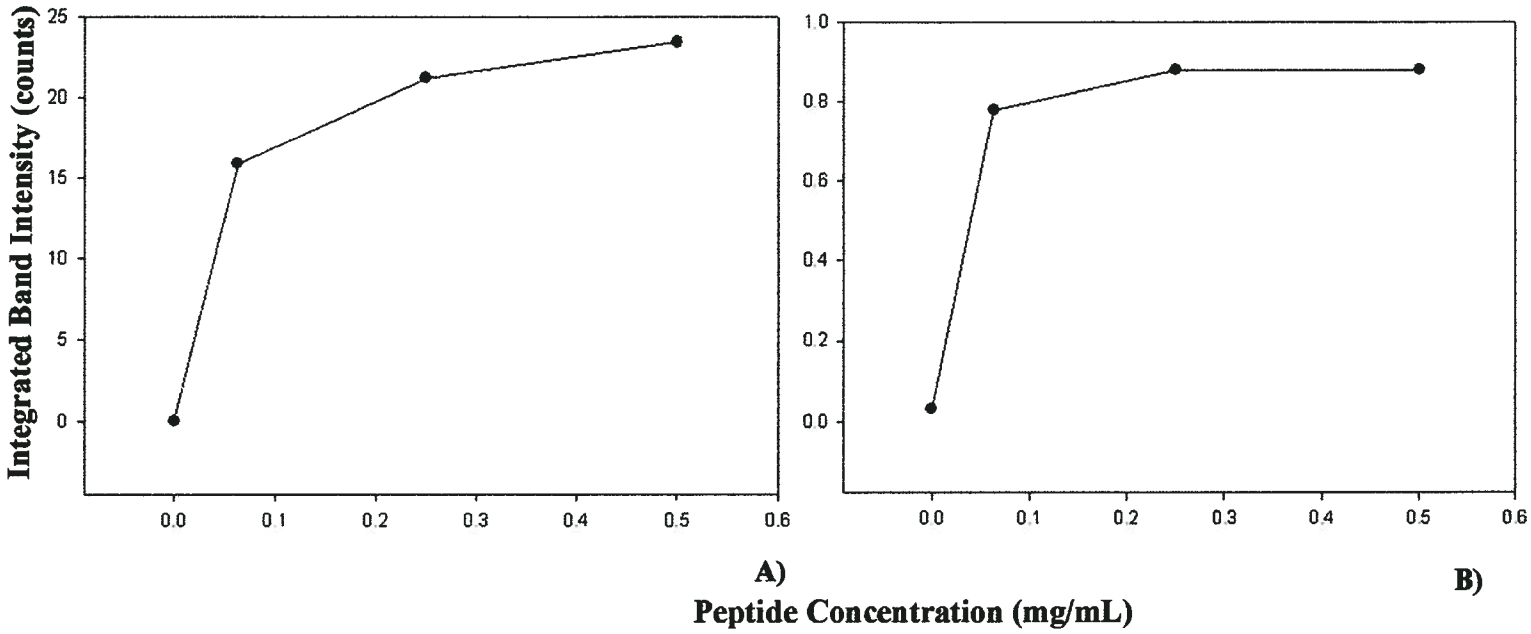


Figure 2-4: Determining the optimum 1D4 peptide concentration for 1D4-tagged receptor recovery. Integrated intensities were determined from densitometry measurements of 1D4 probed western blots of purified A) CXCR4-1D4 and B) ABCA1-1D4 plotted against peptide concentration used for competitive elution.

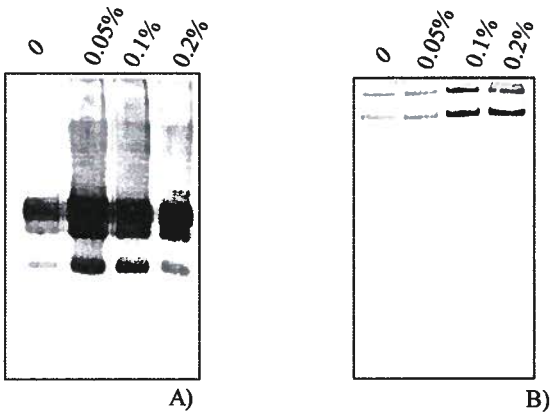


Figure 2-5: Detergent effects on receptor elution recovery. Low concentration of Triton X-100 was titrated in 0.5mg/ml 1D4 peptide in TBS for competitive elution of A) CXCR4-1D4 and B) ABCA1-1D4 from 1D4 mAb conjugated Sepharose.

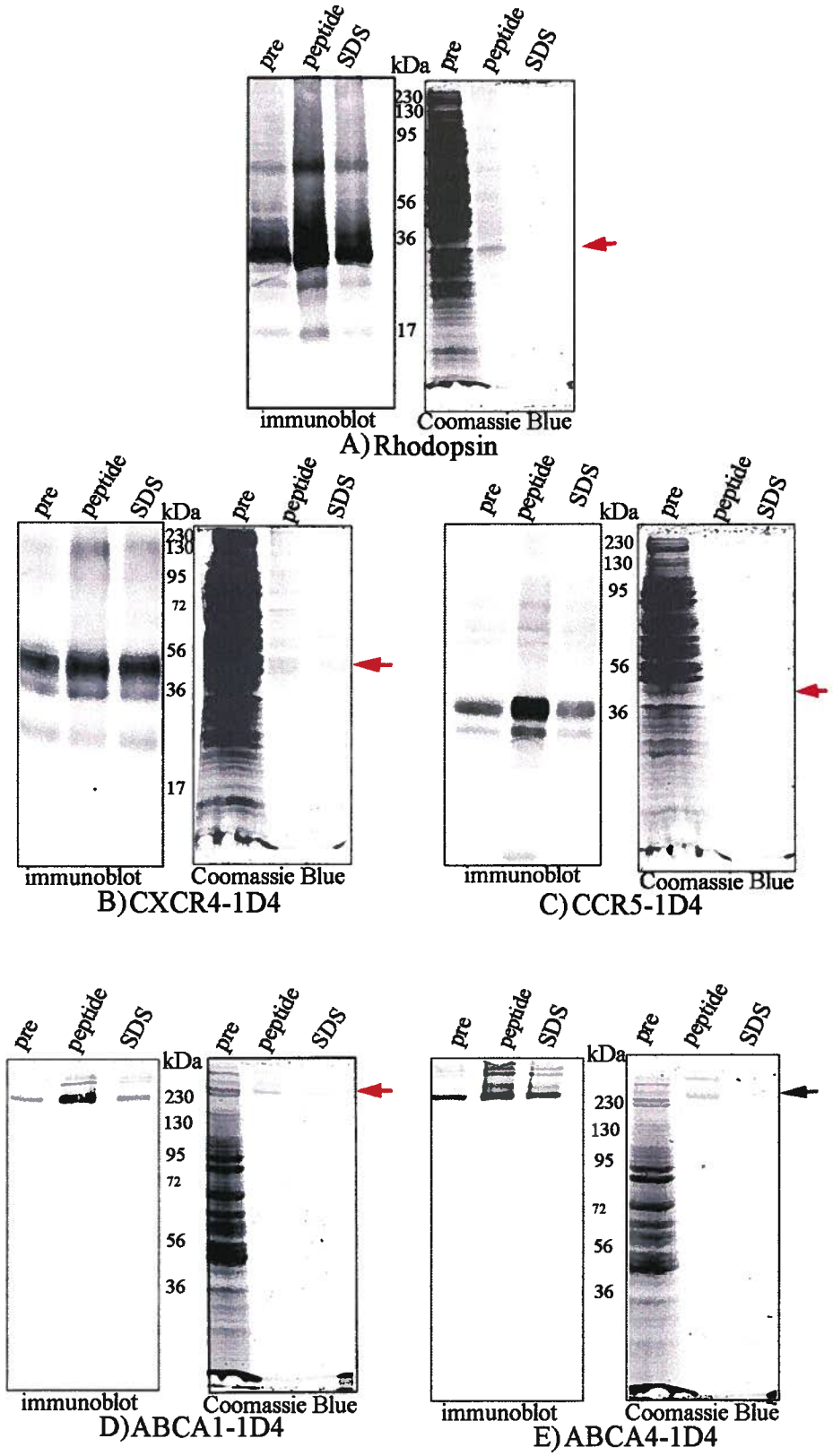


Figure 2-6: C-terminally tagged1D4 receptors expressed in 293T cells affinity purified and separated on SDS-PAGE. Immunoblots and coomassie stained gels show receptor enrichment post immunoaffinity capture and peptide specific elution and a final elution with 1 % SDS to recover remaining receptor. Ectopically expressed A) bovine rhodopsin was used as a control for the purification since rhodopsin was the natural antigen in which the antibody was raised against. Receptor recovery and expression varies amongst the different receptors B) CXCR4-1D4, C) CCR5-1D4, ABCA1-1D4 and ABCA4-1D4.

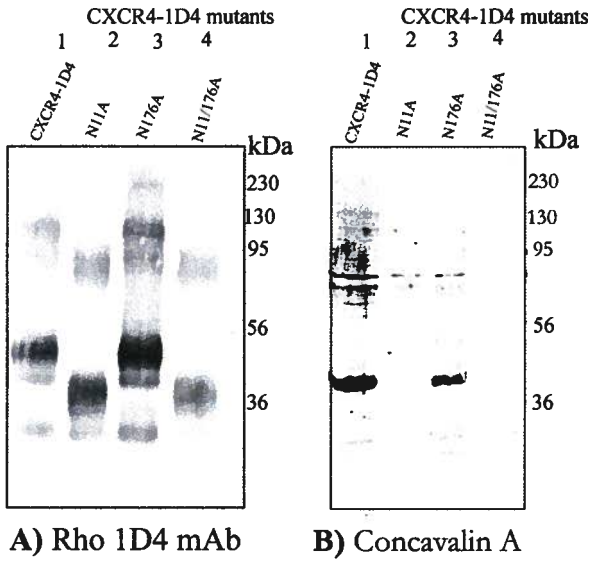


Figure 2-7: Site of CXCR4 post-translational modification were examined by site-directed mutagenesis, western blotting and concavalin A labeling. Western blot of N11A, N176A and N11/176A CXCR4-1D4 mutants were labeled with A) anti-1D4 and B) Concavalin.

2.5 References

1. Luche S, Santoni V, and R. T, *Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis.* Proteomics, 2003. **3**: p. 249-253.

2. Logsdon CD, *The Influence of the cellular context on receptor function: a necessary consideration for physiological interpretations of receptor expression studies.* Life Sciences, 1999. **64**(6/7): p. 369-374.

3. Molloy MP, et al., *Extraction of Escerichia coli proteins with organic solvents prior to two-dimensional electrophoresis.* Electrophoresis, 1999. **20**: p. 701-704.

4. Terpe K, *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.* Applied Microbiol Biotechnology, 2003. **60**: p. 523-533.

5. Wu CC, Yates III Jr, *The application of mass spectrometry to membrane protomics.* Nature Biotechnology, 2003. **21**: p. 262-267.

6. Mirzaei H, et al., *Comparative evaluation of current peptide production platforms used in absolute quantification in proteomics.* Molecular and Cellular Proteomics, 2008. **7**(4): p. 813-823.

7. Mallick P, et al., *Computational prediction of proteotypic peptides for quantitative proteomics.* Nature Biotechnology, 2007. **25**: p. 125-131.

8. Kuster, B., Schirle, M, Mallick, P, and Aebersold, R., *Scoring proteome with proteotypic peptide probes.* Nature Reviews/Molecular Cell Biology, 2005. **6**: p. 577583.

9. Craig R, Cortens JC, and Beavis RC, *The use of proteotypic peptide libraries for protein identification.* Rapid Communications in Mass Spectrometry, 2005. **19**: p. 1844-1850.

10. Ahn J, et al., *Functional Interactions between the two halves of the photoreceptor-specific ATP binding cassette protein ABCR (ABCA4).* The Journal of Biological Chemistry, 2003. **278**: p. 39600-39607.

11. Weber BH, et al., *Inactivation of the murine X-linked juvenile retinoschisis gene, Rs1h, suggests a role of retinoschisin in retinal cell layer organization and synaptic structure.* \. Proc. Natl, Acad. Sci. USA, 2002. **99**: p. 6222-6227.

12. Craig R, Cortens JP, and Beavis RC., *Open source system for analyzing, validating, and storing protein identification data.* Journal of Proteome Research., 2004. **3**: p. 1234-1242.

12. Chabot DJ, et al., *N-linked glycosylation of CXCR4 masks coreceptor function for CCR5-dependent human immunodeficiency virus type 1 isolates.* J. Virology, 2000. **74**(9): p. 4404-4413.

14. Wang J, et al., *N-linked glycosylation in the CXCR4-N-terminus inhibits binding to HIV-1 envelope glycoproteins.* Virology, 2004. **324**(1): p. 140-150.

15. Kuster B, et al., *Scoring proteomes with proteotypic peptide probes.* Nature Reviews/Molecular Cell Biology, 2005. **6**: p. 577-583.

16. Ahn J and Molday RS, *Purification and characterization of ABCR from bovine rod outer segments.* Methods in Enzymology, 2000. **315**: p. 864-79.

17. Ball LE, et al., *Mass spectrometric analysis of integral membrane proteins application to complete mapping of bacteriorhodopsin and rhodopsin.* Protein Science, 1998. **7**: p. 758-764.

18. McDowell JH and Kuhn H, *Light-induced phosphorylation of rhodopsin in cattle photoreceptor membranes: substrate activation and inactivation.* Biochemistry, 1977. **16**: p. 4054-4060.

19. Duffin KL, et al., *Identification and oligosaccharide structure analysis of rhodopsin glycoforms containing galactose and sialic acid.* Glycobiololgy, 1993. **3**: p. 365-380.

20. Hargrave PA, et al., *Rhodopsin's protein and carbohydrate structure: selected aspects.* Vision Research, 1984. **24**: p. 1487-1499.

21. Lanctot PM, et al., *Importance of N-glycosylation positioning for cell-surface expression, targeting, affinity and quality control of the human AT1 receptor.* Biochem J., 2005. **390**(1): p. 367-376.

Chapter III: A novel rapid screening approach for the generation of effective anti-G protein-coupled receptor monoclonal antibodies: target-probe validation

3.1 Introduction

Antibodies are versatile tools for the detection and capture of specific proteins for global as well as targeted protein analyses. In this post-genome era, there is a growing demand for a broad scope of high quality antibodies for exploring all proteins in the human proteome. However, well studied proteins generally have extensive libraries of antibodies recognizing various epitopes, whereas less well studied proteins lack the tools for further elucidation.

Monoclonal antibodies are a valuable source of consistent and renewable affinity reagents with high specificity and affinity. The bottleneck for monoclonal antibody production is not at the animal injection and the hybridoma fusion stage, but rather at the screening or validation stage.

Initiatives such as the human protein atlas (www.proteinatlas.org) set forth by Mathias Uhlen's group is dedicated to collecting and creating a repository of useful antibodies for global human proteomics. However, universal validation does not exist for most available antibodies and quality depends on how the investigators generate and assess the probe. It has been reported that only about 35% of commercial antibodies are effective for the target of interest according to the protein atlas initiative assessment standards [1]. Human proteome analysis is hampered by the reagents and validation platforms available. In response to the growing demand for antibodies, alternative reagents have been developed such as recombinant antibodies, single-

A version of this chapter will be submitted for publication. Co-authors in no particular order Wong JP, Kast J, and Molday RS.

chain variable peptide fragments (svFc) [2], affibodies [3] and aptamers [4]. However, these too are not producible in a high-throughput manner and are subject to validation.

Relatively high throughput platforms have been under development for antibody screening which are mainly enzyme linked immunosorbant assays (ELISA) and flow cytometry-based [5, 6]. Although ELISA and flow cytometric applications may be interfaced with relatively high throughput platforms, these assays are one dimensional and select antibodies based on these specific applications and a multiplexed platform would be desirable to screen for other applications. Antibodies are generated in response to the lack of accurate affinity reagents available for detecting or isolating a protein of interest and thus defining an accurate positive control for the target becomes tricky, especially for proteins such as poorly expressed G protein-coupled receptors (GPCRs). To address this problem, we have employed a well characterized membrane protein compatible 1D4 tag [7] (refer to Chapter II), as a means of detecting a tagged fusion of the ectopically expressed target protein for subsequent antibody screening.

We are interested in drug discovery targets such as G-protein coupled receptors (GPCRs) which are a challenging family of proteins to study due to challenges involved in purification. This is further hampered by the lack of high quality affinity reagents. In particular, we have investigated the GPCR, CXCR4 which has disease relevance in acquired immune deficiency syndrome (AIDS) and cancer [8-10]. Several antibodies directed to human CXCR4 are commercially available but we have noticed limited applicability and specificity. Monoclonal antibodies directed to CXCR4 that are currently available target the N- terminus and extracellular loops on the receptor which have been reported to display antigenic and conformational heterogeneity [11-13]. The epitopes are variable and depending on the cellular context and activity of the receptor, most available antibodies recognize only a subset. Thus,

CXCR4 concentration is generally underestimated in studies that utilize these available antibodies. In fact, a survey of available antibodies indicates that many available antibodies recognize conformationally dependent epitopes [14-16]. Antibodies recognizing linear epitopes are more versatile for proteomic analyses which often require partial or complete protein denaturation [16].

A standardized method assessing application, specificity and affinity would advance antibody based proteomics. Here, we propose a simple, rapid validation platform for monoclonal antibodies raised to membrane proteins involving the 1D4 affinity tag and *in-cell* western detection. As proof-of-principle we have produced a specific monoclonal antibody directed to human CXCR4 C-terminal region with broad ranging applications.

3.2 Material and methods

PCR and cloning:

CXCR4-1D4 constructs were cloned as previously described (refer to Chapter II methods). The C-terminal 45 amino acids and the N-terminus 43 amino acids of human CXCR4 were cloned in frame with the C- terminus of glutathione *S*-transferase (GST) in the pGEX-4T-1 vector (Amersham, Little Chalfont, Buckinghamshire). The restriction endonuclease sites were introduced into the primers used for the polymerase chain reaction (PCR) with the CXCR4-1D4/pcDNA3.0+ as the template. The PCR was carried out for 30cycles at 60°C annealing temperature with *Pfu* polymerase (Fermentas, Glen Burnie, MD). The forward and reverse primers were as follows for the C-terminal construct: *cggaattctttaaaacctctgcccag* and *cgctcgagttagctggagtgaaaac* respectively. The forward and reverse primers employed for the N-terminal fusion construct are as follows: *cggaattcgccgccatggaggggatcagtatatac* and *cgctcgagttagggcaggaagattttattg*, respectively. The PCR products were ligated into EcoRI and XhoI sites, in the pGEX-4T-1 vector. Mice were immunized with fusion protein that had been purified on glutathione-agarose beads (Sigma, Louis, MO)

Expressing and purifying GST-fusion proteins:

A culture (250ml) of N-terminal and C-terminal CXCR4 GST fusion protein constructs transformed into BL21 cells was grown to OD_{600} ~0.9. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was also added to a final concentration of 0.3mM and the cells were grown for a further 2-3 h. Cells were centrifuged and sonicated for 5 times 30 s pulses.

Unbroken cells were pelleted and the resultant lysate supernatant was bound onto glutathione agarose for 2h at 4°C. Beads were washed 5 times in cold phosphate buffered saline (PBS) and fusion proteins were eluted 6 times with 1ml aliquots of elution buffer (3mM EDTA, 10mM glutathione in PBS) at 4°C. The absorbance at 280 nM was measured and the samples were analyzed on sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE).

Injecting Mice:

GST-C-terminal or GST-N-terminal fusion proteins, 10μg each were diluted in PBS and mixed 1:1 with TitreMax Gold adjuvant (Sigma, St. Louis, MO). The antigen adjuvant mixture was injected into 10-12 week old Swiss Webster mice. After 3 weeks, the blood was obtained and the serum was tested for an immune response by *in-cell* or western blot screening. Spleens from selected positive mice were harvested for fusion.

Myeloma and mouse spleen fusions: Hybridoma cell lines were generated as previously described [17] and screened for specificity to CXCR4 using western blots of purified CXCR4-1D4 as will be described in the following.

Cell Culture:

Human embryonic kidney (HEK) 293T (American Type Culture Collection, Manassas, VA) and monkey COS-7 cells (American Type Culture Collection, Manassas, VA) were cultured

in Dulbecco's modified eagle medium (DMEM) (Invitrogen, Auckland, NZ) and 10% fetal bovine serum (FBS) at 37°C and 5% CO_2. A confluent 20mm dish of 293T cells were seeded 1 in 10 dilution 12 h prior to calcium phosphate mediate transfection.

In-cell westerns for hybridoma screening applying 1D4 tagged protein positive control:

Approximately $5x10^7$ HEK 293T cells were transfected with 20μg of CXCR4-1D4/pcDNA3.0+ construct following the calcium phosphate mediated procedure. For one 100 x 20mm dish 100μl of a 1M calcium chloride solution was added to a 400μl solution of DNA, prior to the addition of 500 μl of *N,N*-bis(2-hydroxy-ethyl)-2-aminoethane-buffered saline at pH 6.96 and incubated at room temperature for 20min. The DNA mixture was then added to the cells and incubated at 37°C in a 5% CO_2. Media was changed 8 h post transfection and incubated overnight. The transiently transfected cells were resuspended in fresh media and evenly split between two 96well poly-lysine (Sigma, St. Louis, MO) coated plates and incubated overnight. The cells were fixed with 4% paraformaldehyde for 20min and washed with PBS and blocked with normal goat serum prior to antibody binding. The fixed cells were bound with hybridoma supernatants or Rho 1D4 primary antibodies and then subsequently with goat anti-mouse IgG conjugated Alexa 680 IRdye secondary antibody (LI-COR, Lincoln, NB). *In-cell* western visualization and densitometry were performed on the LI-COR Odyssey Infrared imager (Lincoln, NB).

Secondary Screen, Western blot strips:

Approximately $5x10^7$ HEK 293T cells, transiently transfected with 20μg of CXCR4-1D4/pcDNA 3.0 construct via calcium phosphate mediated transfection were harvested and solubilized in 1ml 2% Triton X-100 detergent in Tris buffer saline (TBS) with complete protease inhibitors (Roche, Mannheim, Germany) 24 h post transfection. Insoluble material was spun down with ultracentrifuge at 100,000 x g force. The lysate was run on 10% SDS-PAGE in wide wells and transferred onto polyvinylidene fluoride (PVDF) membrane. Lanes containing the lysate were cut into 2-5 mm strips and probed for CXCR4-1D4 with 1D4 monoclonal antibody and anti-mouse horse radish peroxidase (HRP) antibody and enhanced chemiluminscence (ECL) detection.

Determining Antibody Isotype:

Mouse monoclonal antibody isotope was determined with Isostrip kit (Roche, Mannheim, Germany) as per manufacturer's protocol.

Immunofluorescence: Transfected Human 293T cells were fixed onto polylysine coated cover slips with 4% paraformaldehyde (Fisher; Fair Lawn, New Jersey) for 20min at room temperature. Cells were blocked with normal goat serum and stained with 1/50 dilution of 1F2 antibody. Anti-mouse Alexi 488 secondary antibody was subsequently used and the nuclei was stained with 4', 6-diamidino-2-phenylindole (DAPI).

Preparing Immunoaffinity matrix:

1F2 monoclonal antibody conjugated beads were made with cyanogen bromide activation of Sepharose-2B beads (Amersham, Little Chalfont, Buckinghamshire) and 2mg 1F2 antibody per 1ml of activated Sepharose concentration. The reaction was stopped with 0.1M glycine (Sigma, St. Louis, MO) pH 2.5 and the beads were washed with TBS and stored in PBS containing 0.01% azide at 4°C.

Immunoprecipitation:

Approximately 10^8 Jurkat E6-1 cells (American type culture collection, ATCC) were washed with PBS and solubilized in 2% Triton X-100 in Tris buffered saline (TBS). The solubilized cells were added to 1F2 monoclonal antibody conjugated Sepharose and allowed to mix and bind for 1h at 4°C in the presence of complete protease inhibitors (Roche, Mannheim, Germany). The beads were washed 10 times with 0.2% Triton X-100 in TBS and CXCR4 was competitively eluted with one bed volume of 0.5mg/ml 1F2 peptide in 0.2% Triton X-100 in TBS for 20min at room temperature and repeated with an additional bed volume of 1F2 peptide. SDS (1%) in TBS was used to remove remaining receptor bound to the beads.

Isolating white blood cells from whole blood:

Mononuclear cells were isolated from 20ml of peripheral blood by Ficoll density gradient centrifugation. The recovered cells were resuspended in 90% fetal bovine serum (FBS) (v/v), 10% dimethyl sulfoxide (DMSO) (v/v) in Roswell Park Memorial Institute (RPMI) media

and stored in liquid nitrogen, prior to use. The purified PBMCs were provided by Dr. Lenka Allan, Child and Family Research Institute, UBC.

Mapping epitope:

The 1F2 epitope was mapped using the SPOTS kit (Sigma Genosys, Woodlands, TX) according to the manufacturer's protocol. Synthetic peptides (9 amino acids long, 2 amino acid overlap) spanning the C-terminal 45 amino acids of CXCR4 were coupled to a cellulose membrane, and their immunoreactivity was probed with subcloned 1F2 primary antibody and anti-mouse IgG-horse radish peroxidase (HRP) for detection by enhanced chemiluminescence (ECL).

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and western blotting:

Samples were run on 10% separating gel with a 5% stacking polyacrylamide gel following the Laemmli method. The gels were either stained with coommassie brilliant blue or transferred onto polyvinylidene fluoride (PVDF) (Millipore, Bedford, MA) membranes for western blot detection and densitometry measurements with a LI-COR Odyssey (Lincoln, NB) infrared scanner.

In-gel digestion:

Coomassie Blue stained gel bands were washed three times 15min each with 25mM ammonium bicarbonate/50% acetonitrile followed by reduction with 1.5 mg/ml dithiothreitol (DTT) at 50°C for 30min and alkylation with 10mg/ml iodoacetamide (IAA) at room temperature for 20min. Gel bands were rehydrated with 37.5ng/ml sequencing grade trypsin and/or chymotrypsin (Roche, Mannheim, Germany) with 5-10 mM $CaCl_2$ in 50mM ammonium bicarbonate (ABC) on ice for 30min. Excess enzyme was removed and 15µl of additional digestion buffer was added to prevent gel drying during digestion. The gel bands were incubated at 37°C with shaking at 450 rpm for 18 h. Digests were extracted with 1:1 50mM ammonium bicarbonate/acetonitrile and the final extraction was carried out with 1:1 5% formic acid/acetonitrile. Extracts were pooled and lyophilized to dryness and stored at -80°C until analysed.

LC-MS/ MS conditions:

Tryptically digested peptides were base and then acid extracted from the gel pieces, pooled and lyophilized. Samples were reconstituted in 6µl of formic acid (FA) and injected onto the Qstar XL LC/MS/MS (Applied Biosystems, Foster City, CA). A PepMap C18, 3µm particle size and 100Å pore size column from LC packings (Amsterdam, Netherlands) was used for peptide separation. Solvents B and A were 20% acetonitrile (ACN) and 5% ACN, respectively. Chromatography conditions started at 2% solvent B with a gradient to 60% B over 60min, to 95% B at 63min and held for three minutes before returning to 2%B.

3.3 Results

GST-fusion protein antigen expression test:

A small scale induction was performed for both the CXCR4-N-terminual and C-terminal GST fusion protein constructs and an empty vector control to determine whether both fusion proteins are expressed (Figure 3-1A). A dominant band was observed at approximately 30kDa for both the +IPTG lanes for both the C-terminal and N-terminal fusion preparations but absent in the non-induced lanes (Figure 3-1A, +IPTG lanes). These 30kDa bands were not observed in the –IPTG samples. A GST protein was observed at 25kDa in the empty vector control (VC) lane. A similar band was observed in the +IPTG C-terminal fusion lanes slightly below the fusion protein band at 25kDa.

GST-fusion protein antigen purification:

A large scale overnight culture was induced with IPTG to generate antigen for purification and mouse injection. The GST C-terminal protein was observed as a single 30kDa band (Figure 3-1B, eluate lanes 1-6) separated from the rest of the cellular proteins observed in the precolumn and flowthrough lanes (Figure 3-1B). A smaller band at 25kDa was observed in Figure 3-1A in all 6 elution lanes. Similarly, a unique band was observed for the GST N-terminal fusion protein at 30kDa (Figure 3-1C, eluate lanes 1-6).

In-cell western detection for assessing mouse response and large scale hybridoma screening:

Mouse antibody response and hybridoma screening was tested on the prepared microtitre

plates employing LI-COR infrared *in-cell* western detection. The microtitre plates were prepared in advance for large scale screening. The plates were titrated with 1D4 monoclonal antibody as a positive control (Figure 3-2A, wells A1-A3), prior to storage at 4°C with 0.01% sodium azide before use. Wells A1-A3(Figure 3-2A) inclusive were bound with 1/10, 1/100 and 1/1000 dilution of 1D4 antibody respectively displayed a gradual decrease in integrated intensity from densitometry measurements in response to decreasing antibody concentration.

The *in-cell* westerns were also used to assess mouse response to the antigen introduced. A small sample of mouse serum was collected for in-cell screening three weeks post injection. Mouse serum that displayed higher than background integrated intensity signals on the *in-cell* screen was selected for hybridoma fusion (Figure 3-2B). The mouse serum also tested on western blot strips for additional verification as will be discussed in the following section. Three of the six injected mice showed response; however, one mouse injected with the C-terminal GST fusion protein displayed the highest titre by *in-cell* western and western blot strip screening was selected for hybridoma production.

Over 800 hybridomas were produced from the injected mouse and subsequently screened for specific antibody production by *in-cell* western. Well H12 on the screening plates was reserved for a 1D4 monoclonal antibody positive control. Figure 2C is one representative in-cell western screening plate out of 9 bound with supernatant transferred directly from hybridomas grown in a corresponding microtitre plate coordinates. Well F2 which had integrated intensity of 8.4 which was well above background signal was selected for further screening (Figure 3- 2C).

Secondary screen- western blotting on PVDF membrane strips:

Sixty positive clones were selected from the preliminary screen. These clones were expanded in 24 well plates and the respective supernatants were harvested for testing on narrow strips of PVDF membrane transferred from CXCR4-1D4 lysate after SDS-PAGE separation. For each western blot strip screen, a positive control with 1D4 monoclonal antibody was run in parallel (Figure 3-3B, last ++ labeled strips). CXCR4-1D4 positive control strips detected 2 protein bands near the bottom half of the strip and a larger molecular weight band representing a dimer. One clone, 1F2 recognized two protein bands in a region similar to the ones observed in the positive control with similar intensity and location (Figure 3- 3B). Several clones were non-specific and stained the entire strip (Figure 3-3A; strips 1E11, 1D6, 1D4, 1B12) or were blank on film.

In-cell western detection for 1F2 subcloning:

Clone 1F2 was selected for further subcloning and characterization. In the first round of subcloning approximately 90% (86/96 clones were detected) (Figure 3-4A) of the clones were positive as compared to positive 1D4 control (Figure 3-4A, well H12). One well was reserved as a negative control which was only bound with secondary antibody (Figure 3-4A, well G12). The most intense subclone was subjected to further dilution and subcloning until all wells or subclones were positive as assessed by in-cell western detection.

The titre of the subcloned 1F2 clone was tested by serial dilution on the *in-cell* western (Figure 3-4B, wells C12, D12, E12) and compared against Rho1D4 mab positive control wells (Figure 3-4B, F12, G12, H12).

1F2 mAb specificity:

To confirm that the selected hybridoma specifically recognizes wild type CXCR4 without the 1D4 C-terminal tag, cell lysate from Jurkat E6-1, CXCR4 and CXCR4-1D4 transient transfectants in HEK 293T cells were separated on polyacrylamide gel and transferred onto PVDF membrane (Figure 3- 5). 1F2 mAb recognized both the ectopically expressed wild type CXCR4 and CXCR4-1D4 (Figure 3-5A). A series of bands extending from 40-50kDa and a thinner band at 30kDa was observed for CXCR4 and CXCR4-1D4 on the western blot strip (Figure 3-5A, lanes 2 and 3). CXCR4-1D4 migrates at a slightly higher molecular weight than CXCR4 (Figure 3- 5A). IF2mAb also recognizes a band at 50kDa in Jurkat E6-1 lane (Figure 5A). A parallel blot run with the same samples in figure 3-5A was labeled with the 1D4 antibody. Rho1D4 mAb only recognized CXCR4-1D4 which is observed as a series of 3-4 bands at 40-50kDa and a smaller band at 30kDa and nothing in the Jurkat or CXCR4 lysate.

Mapping the 1F2mAb epitope:

Nine amino acid long stretches of the CXCR4 C-terminus with 2 amino acid overlap were synthesized onto a membrane support for antibody epitope mapping (Figure 3-6). The 1F2 mAb detects an internal sequence on the C-terminus corresponding to the overlapping sequences

QHALTSVSR, ALTSVSRGS, TSVSRGSSL and VSRGSSLKI as these peptides were visible on the film (Figure 3-6A). The two end spots recognized were weaker in intensity and thus were closer to the regions beyond the epitope. We finalized the epitope sequence as ALTSVSRGS. The peptide epitope was synthesized with a C-terminus amide cap and tested for specific elution in immunoaffinity purification.

Testing 1F2mAb for immunoprecipitation application:

The 1F2 mAb was tested for immunoprecipitation application on cells transfected with CXCR4-1D4 or CXCR4. The purifications were run on SDS-PAGE, transferred for western blot analysis and densitometry. Total receptor concentration was determined from densitometry measurements of the pre-column sample. The eluate concentrations were also determined by densitometry and then normalized against total receptor present in the precolumn sample to calculate percent recovery. Approximately 20% of the receptor is released with peptide competition as determined from densitometry and 40% is recovered from a final non-specific elution with 1% SDS (Figure 3-7).

1F2mAb for endogenous CXCR4 immunoprecipitation and detection:

Endogenous CXCR4 from peripheral blood mononuclear cells (PBMC) and Jurkat E6-1 cells was purified with the 1F2 conjugated Sepharose (Figure 3-8). A single 40kDa band was purified from the PBMC (Figure 3-8A) and three bands were recognized from the Jurkat cell line (Figure 3-8B). Peptide elution releases one band in the PBMC sample but in the Jurkat sample

only the smallest band in the purification was released with peptide competition. Additional bands were eluted non-specifically with SDS in the Jurkat cell sample (Figure 3-8B).

Applications-Immunofluorescence:

As determined from the initial *in-cell* screen, the 1F2 mAb clone recognizes CXCR4 in paraformaldehyde fixed cells. We further tested the 1F2 monoclonal antibody for immunofluorescence applicability. 1F2mAb recognizes transfected CXCR4 in HEK 293T cells as observed by the green staining (Figure 3-9).

Assessing anti-CXCR4 antibody specificity, 1D4 tagged receptors for antibody validation:

We have compared the effectiveness of some commercially available polyclonal and monoclonal antibodies directed against human CXCR4 by western blot analysis. CXCR4-1D4 and an empty vector control cell lysates were run on gels and transferred onto PVDF membranes for western blotting with various anti-CXCR4 antibodies (Figure 3-10). A positive control blot probed with 1D4 monoclonal antibody was used to detect CXCR4-1D4 (Figure 3-10A). Commercial polyclonal anti-huCXCR4 (1-14) IgG antibody fraction generated similar banding patterns in both vector control and receptor lane, and no band was evident at the theoretical molecular weight for CXCR4 (Figure 3-10B). The polyclonal rabbit anti-serum raised against a chemically synthesized tetox fusion protein of the entire CXCR4 N-terminus showed a ~50kDa band in the receptor transfected lane and not the vector control lane. However, several non-specific bands at higher molecular weight were also detected in the vector control lane (Figure 3-

10C). The 1F2 mAb antibody recognized a broad band ~50kDa that resembles the positive control blot and extraneous bands were not detected in the vector control lane (Fig 3-10D).

3.4 Discussion

A rapid, general large scale screening approach for effective and efficient monoclonal antibody production for GPCRs was developed. The 1F2 mAb produced and validated by our rapid *in-cell* screen, recognizes human CXCR4 C-terminal residues 315-324 and has broad applications. The linear epitope may be employed for competitive elution from 1F2 mAb conjugated Sepharose for immunoaffinity purification. Currently available CXCR4 monoclonal antibodies are mainly directed to the N-terminal residues and extracellular loops of the receptor with limited application. These antibodies were characterized and are reported to sample only a subset of the receptor due to receptor extracellular site conformational and antigenic heterogeneity [11, 13]. To accurately estimate CXCR4 expression in a particular cell type or condition, a combination of the available antibodies should be used or a constant region on the receptor for antibody development needs to be identified and targeted.

Generating GST-fusion proteins for immunostimulation:

Short segments of the receptor were selected for expression and antigen target for antibody production. The entire CXCR4 C- and the N-terminus were selected for mouse immunostimulation for the generation of an antibody targeted to a linear CXCR4 epitope. To ensure that the termini of the receptor are expressed as a GST fusion protein, we performed a small scale crude analysis on SDS-PAGE (Figure 3-1). The C and N-terminals were observed at

approximately 30kDa separated on 10% SDS-PAGE which corresponds to 25kDa for the GST portion plus approximately 4500kDa for the 45 amino acid long terminal CXCR4 peptides. The smaller band observed at 25kDa for the GST-C-terminal fusion proteins observed in the elution lanes for Figure 3-1B may be degradation products. Degradation was not observed for the GST-N-terminal and the fusion protein purified as a single band (Figure 3-1C).

1D4 C-terminally tagged human CXCR4 for antigen target validation:

From previous work, we have shown that the C-terminal 1D4 tag is a suitable enrichment method for membrane proteins expressed in mammalian systems (refer to Chapter II) [18]. Addition of the 1D4 tag does not affect receptor expression, localization or function (refer to Chapter II). We have applied this enrichment method for target antigen validation for the production of monoclonal antibodies directed against our model GPCR, human CXCR4.

The 1D4 tag is specific and consistent in the detection of CXCR4-1D4 the positive control in the antibody *in-cell* screens described. Furthermore the tag is also applicable on other membrane proteins without affecting functional activity as previously shown (refer to Chapter II, pg 50).

In-cell westerns for rapid large scale hybridoma screening

Employing our 1D4 tagged receptor approach coupled to *in-cell* western detection, we were able to select and characterize an effective antibody with broad applications in 3 months. This approach offers several advantages. Screening is fast, simple, versatile and robust. There is

a small window of time in which freshly generated hybridoma cell lines are selected and subcloned. Screening and selection after hybridoma production has to be fairly quick, as delayed subcloning may result in positive clone loss from dilution of other non-antibody producing or spurious clones. Although the clones may be maintained or expanded in culture to extend the time required for hybridoma validation this becomes cumbersome and uneconomical when hundreds of clones are produced. The 1D4 tag is applicable to several types of membrane proteins (refer to Chapter II) and when combined in a microtitre plate format allows for quick hybridoma screening and subcloning. Hybridoma supernatants from the resultant fusions are directly transferred into microtitre plates with the same plate coordinates for simplicity. Moreover, detecting the tag on the receptor serves as a control for ensuring that the prepared screening plates are consistent from screen to screen. The *in-cell* screening plates can be stored with sodium azide at 4°C for as long as one week and thus can be prepared in advance. Sodium azide preservative in the storage buffer does not affect subsequent detection with the LI-COR secondary antibody as it would for HRP and ECL detection in traditional western blot visualization.

In-cell screening follows similar steps as a conventional western blot and only requires approximately 3-4 h to complete from blocking, primary antibody binding, secondary binding and washing to complete. Final detection does not require a substrate chemical reaction. The tested plates are visualized on the LI-COR Odyssey instrument which can scan multiple microtitre plates at one time.

The Rho 1D4 monoclonal antibody is a constant renewable source for testing the screening plates. Intensity measurements determined for the *in-cell* western may be applied to approximate antibody titre relative to the 1D4 positive control which provides an additional

parameter for hybridoma selection in the screen due to the large dynamic range capable on the LI-COR instrument. We have effectively tested 9 microtitre plates, approximately 800 clones in one evening and we were able to retest a second time to confirm and select positive hybridomas.

We have for the first time generated an anti-human CXCR4 mAb raised against a recombinant fragment of CXCR4 that recognizes a linear epitope. We have applied the 1D4 tag and the *in-cell* western technology as a platform to screen ensuring that the antibody recognizes the human receptor as it is expressed in the relevant cellular context.

1F2 mAb for target validation in various cell types and disease profiling:

1F2 recognizes ectopically and endogenously expressed CXCR4. We have shown that the antibody is able to isolate the receptor expressed in primary human PBMC, the Jurkat E6-1 T-cell line as well as CXCR4 transfected in HEK 293T cells which was also confirmed by mass spectrometry. The 1D4 tagged CXCR4 serves as an initial positive control to select and validate antibodies based on recognition of physiologically relevant forms of CXCR4.

Our validation platform is effective for the rapid production of versatile and specific monoclonal antibodies for the study of GPCRs. As proof of principle we have generated the 1F2 mAb that we have shown recognizes endogenous CXCR4 in primary cells. The 1F2 mAb provides a useful tool for receptor profiling in various CXCR4 implicated diseases such as AIDS and cancer metastasis.

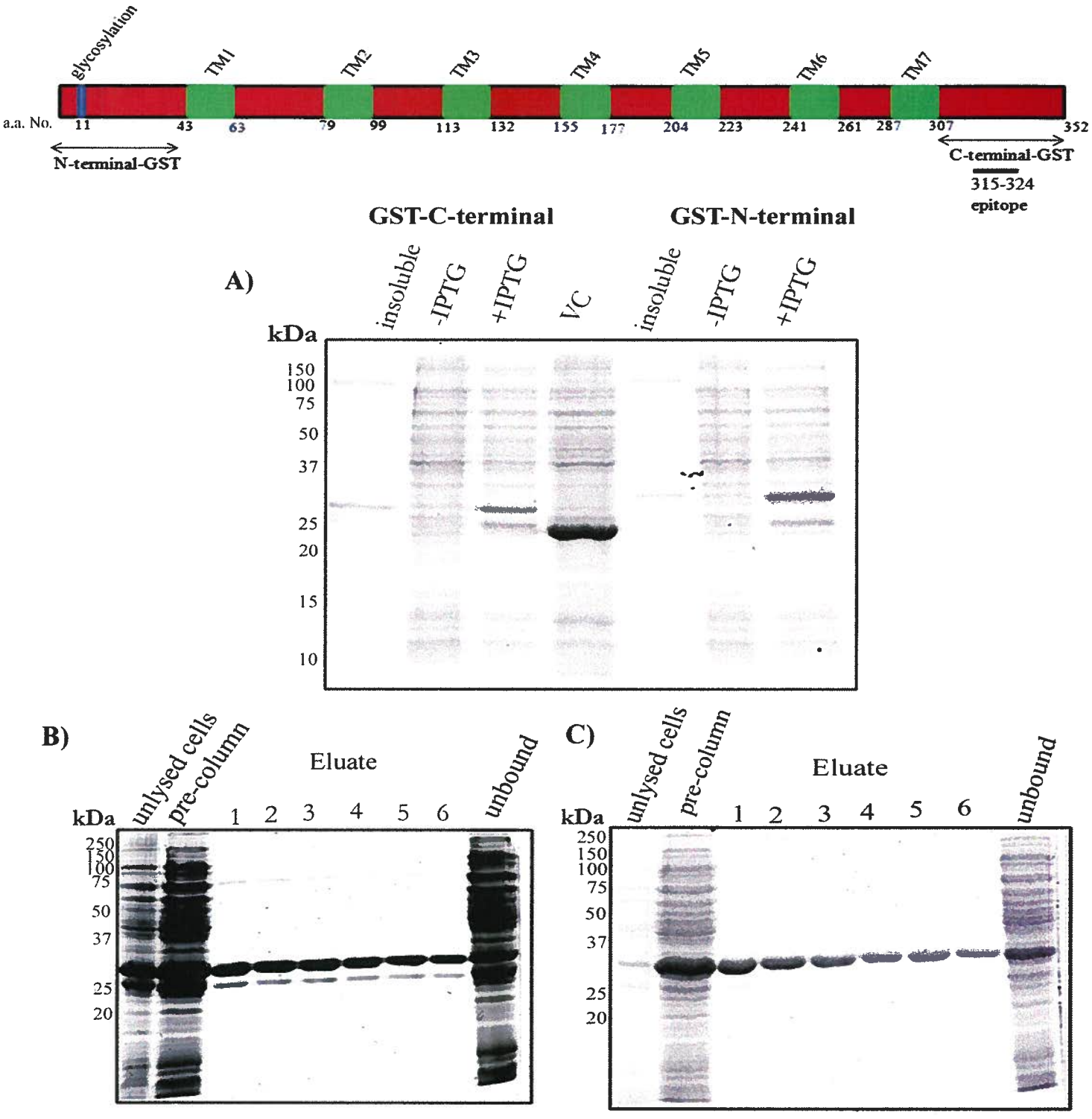


Figure 3-1: A N- and C-terminal fusion protein expression test, small scale induction. Samples were sonicated and separated on SDS-PAGE and Coomassie Blue stained A). Expressed GST-fusion proteins B) C-terminal and C) N-terminal proteins were purified on glutathione agarose and eluates were evaluated on SDS-PAGE. A linear amino acid (a.a.) map showing the position of the transmembrane (TM) segments and termini selected for CXCR4 termini fusion protein construction is displayed at the top of the figure.

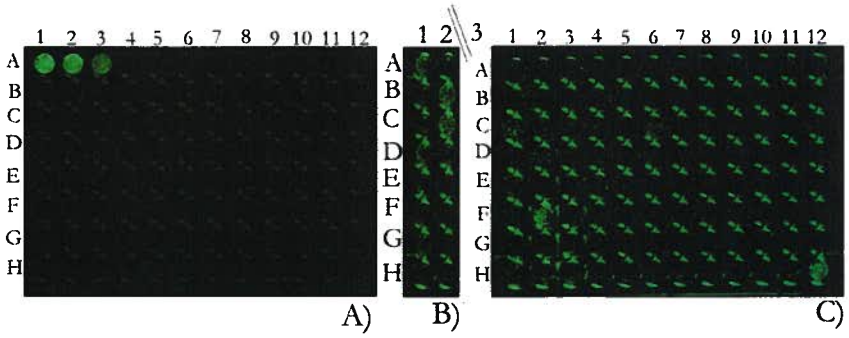


Figure 3-2: *In-cell* western detection for hybridoma screening. Human 293T cells transfected with CXCR4-1D4/pcDNA3.1+ constructs were grown onto poly-lysine coated 96 well microtitre plates. A positive control plate A) was used to titrate Rho 1D4 monoclonal antibody at 1/10, 1/100, and 1/1000 dilution (wells A1-A3 inclusive and respectively). The *in-cell* westerns were also used to assess mouse response 3 weeks post injection in plate B) and hybridoma supernatant was transferred directly from the 96 wel microtitre plate for hybridoma screening in plate C). Well H12 was reserved for the 1D4 mAb positive control.

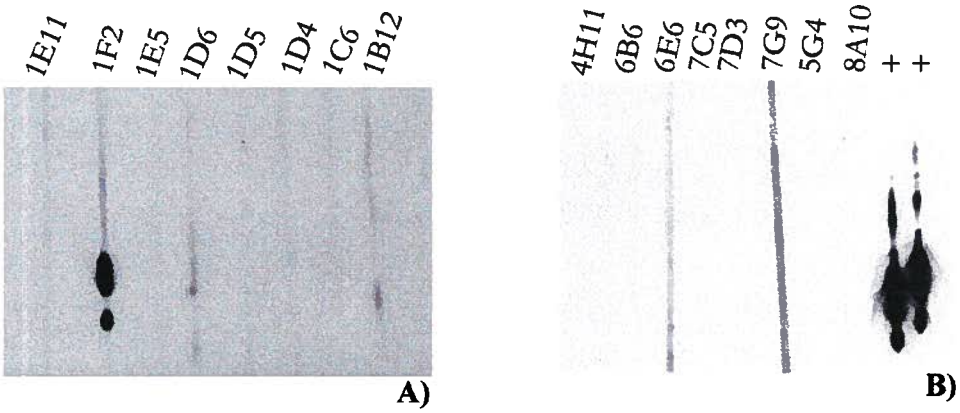


Figure 3-3: *In-cell* western positive hybridomas were selected and further screened by western blot analysis. CXCR4-1D4 transfected 293T cell lysate was separated on SDS-PAGE and transferred onto PVDF. Lanes with lysate were divided into 3-5mm strips for hybridoma screening with an anti-mouse IgG HRP conjugated secondary antibody and ECL detection. Western blot strips were labelled with supernatant from positive hybridomas and the Rho 1D4 monoclonal antibody as a positive control, ++.

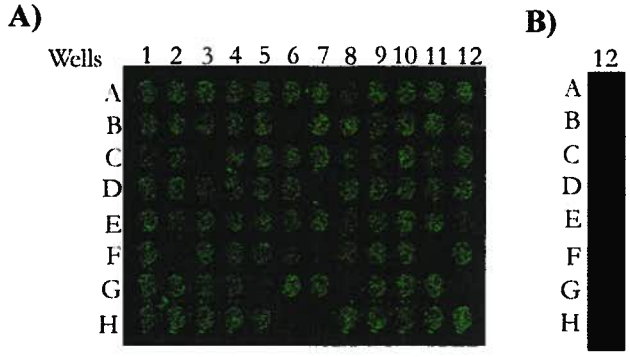


Figure 3-4: *In-cell* western screening for 1F2 mAb subcloning. One representative screen for A) the first round of subcloning and B) final titre for the supernatant harvest from doubly subcloned 1F2 mAb compared to standard Rho 1D4 mAb. Wells C12, D12, E12 correspond to 1F2 concentrations: neat, 1/10 and 1/100 dilution. Whereas wells F12, G12, H12 correspond to: Rho 1D4 mAb 1/10, 1/100, and 1/1000 dilutions, respectively.

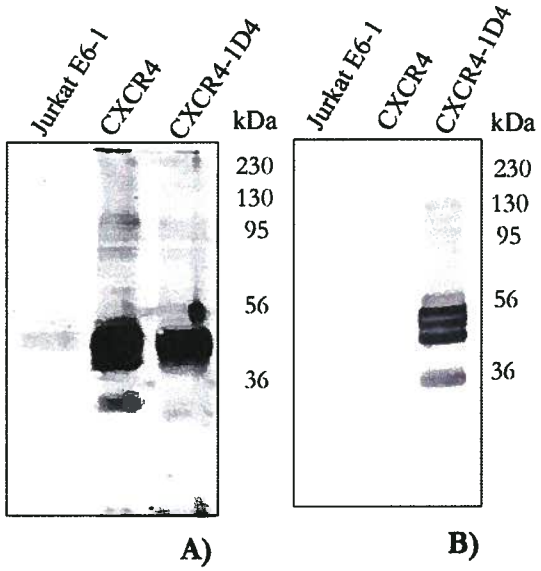


Figure 3-5: 1F2 and 1D4mAb specificity. 1F2 mAb recognizes endogenous CXCR4 in Jurkat E6-1 cells and transfected CXCR4 and CXCR4-1D4 in HEK 293T cells. Cell lysates were separated on SDS PAGE and transferred onto PVDF for western blottting with A) 1F2 mAb and B) Rho 1D4 mAb.

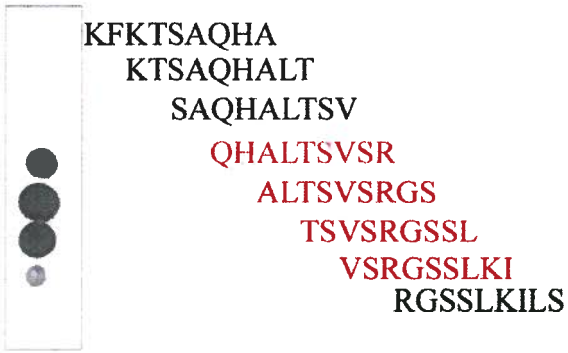


Figure 3-6: 1F2 mAb epitope mapping. Nine amino acid segments of the C-terminus with 2 amino acid overlap were synthesized following Fmoc chemistry on a filter paper support. The entire support upon synthesis completion was blocked, probed with 1F2 mAb and anti-mouse IgG HRP secondary for ECL detection.

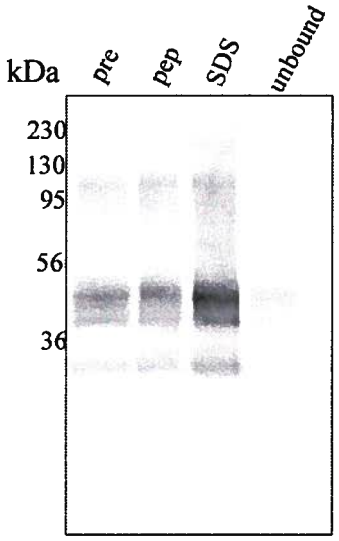


Figure 3-7: 1F2 mAb for immunoprecipitation. CXCR4 was immunoaffinity purified with 1F2 mAb conjugated Sepharose and analyzed on SDS-PAGE and western blot. Receptor recovery was determined by densitometry.

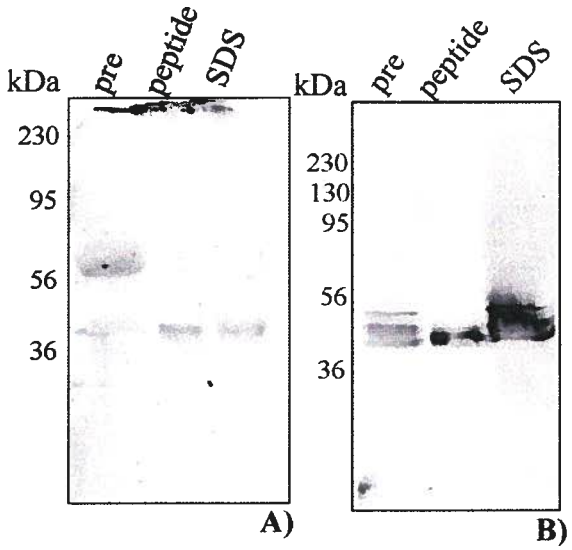


Figure 3-8: 1F2 mAb immunoprecipitation of endogenous CXCR4 from A) human peripheral blood mononuclear cells and B) human Jurkat E6-1 cells.

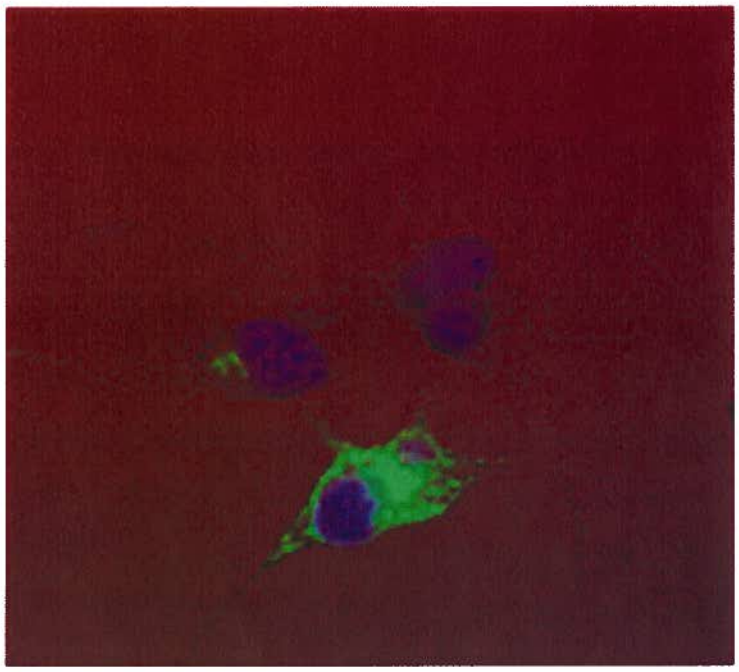


Figure 3-9: 1F2 mAb for immunofluorescence. HEK 293T cells transiently transfected with CXCR4 were fixed onto poly-lysine coated cloverslips and fixed with paraformaldehyde. The coverslips were blocked with normal goat serum, labelled with 1F2 mAb and anti-mouse Alexi 488 secondary antibody for immunofluorescence visualization.

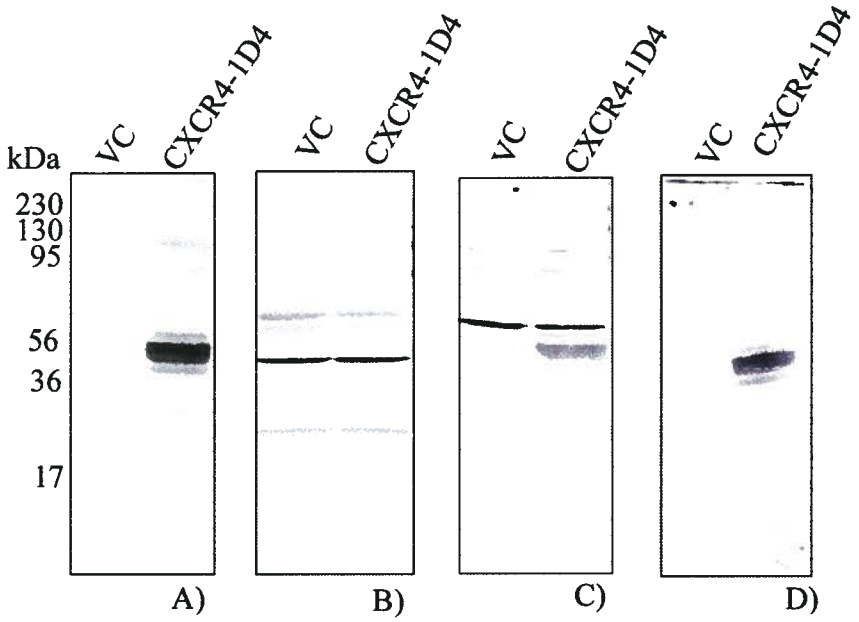


Figure 3-10: Assessing anti-CXCR4 antibody specificity. Western blots of CXCR4-1D4/pcDNA3.0 and pcDNA3.0 vector control were transiently transfected into HEK 293T cells and solubilized in Triton X-100 detergent. The blots were labelled with primary antibodies A) Rho 1D4 mAb, B) commercial anti-human CXCR4 (1-14) IgG fraction of rabbit serum, C) polyclonal rabbit serum raised against synthetic tetox N-terminal peptide and D) in-house anti-CXCR4 C-terminal mAb.

3.5 References

1. Blow N, *Antibodies: The generation game.* Nature, 2007. **447**: p. 741-744.

2. Bonnin E, et al., *Generation of functional scFv intrabodies for triggering anti-tumor.* Methods, 2004. **34**: p. 225-232.

3. Nord K, et al., *Binding proteins selected from combinatorial libraries of an α-helical bacterial receptor domain.* Nature Biotechnology, 1997. **15**: p. 772-777.

4. Lee JF, Stovall GM, and Ellington AD, *Aptamer therapeutics advance.* Current Opinion in Chemical Biology, 2006. **10**: p. 282-289.

5. Ayriss J, et al., *High-throughput screening of single-chain antibodies using multiplexed flow cytometry.* J. Proteome Research, 2007. **6**(3): p. 1072-1082.

6. Marks JD, et al., *By-passing immunization. Human antibodies from V-gene libraries displayed on phage.* J. Mol. Biol., 1991. **222**: p. 581-597.

7. Hodges RS, et al., *Antigen-antibody interaction. Synthetic peptides define linear antigenic determinants recognized by monoclonal antibodies directed to the cytoplasmic carboxy terminus of rhodopsin.* Journal of Biological Chemistry, 1988. **263**(24): p. 11768-11775.

8. McKnight A, et al., *Inhibition of human immunodeficiency virus fusion by a monoclonal antibody to a coreceptor (CXCR4) is both cell type and virus strain dependent.* Journal of Virology, 1997. **71**(2): p. 1692-1696.

9. Endres MJ, et al., *CD4-independent infection by HIV-2 is mediated by fusin/CXCR4.* Cell, 1996. **87**: p. 745-756.

10. Darash-Yahana M, et al., *Role of high expression levels of CXCR4 in tumor growth, vascularization, and metastasis.* FASEB journal, 2004. **18**: p. 1240-1242.

11. Baribaud F, et al., *Antigenically distinct conformations of CXCR4.* Journal of Virology, 2001. **75**(19): p. 8957-8967.

12. Strizki JM, et al., *A monoclonal antibody (12G5) directed against CXCR4 inhibits infection with the dual-tropic human immunodeficiency virus type 1 isolate HIV-$1_{89.6}$ but not the T-tropic isolate HIV-1_{HxB}.* Journal of Virology, 1997: p. 5678-5683.

13. Xu C, et al., *Human anti-CXCR4 antibodies undergo V_H replacement, exhibit functional V-region sulfation, and define CXCR4 antigenic heterogeneity.* Journal of immunology, 2007. **179**: p. 2408-2418.

14. Van Regenmortel MHV, *Mapping epitope structure and activity: from one-dimensional prediction to four-dimensional description of antigenic specificity.* Methods, 1996. **9**: p. 465-472.

15. Uhlen M and Ponten F, *Antibody-based proteomics for human tissue profiling.* Mol. Cell. Proteomics, 2005. **4**: p. 384-393.

16. Berglund L, et al., *The epitope space of the human proteome.* Protein Science, 2008. **17**: p. 606-613.

17. Mackenzie D and Molday RS, *Organization of rhodopsin and a high molecular weight glycoprotein in rod photoreceptor disc membranes using monoclonal antibodies.* Journal of Biological Chemistry, 1982. **257**(12): p. 7100-7105.

18. Ahn J, et al., *Functional interaction between two halves of the photoreceptor-specific ATP binding cassette protein ABCR (ABCA4). Evidence for a non-exchangeable ADP in the first nucleotide binding domain.* Journal of Biological Chemistry, 2003. **278**(41): p. 39600-39608.

Chapter IV: Mapping stromal cell derived factor 1 and CXCR4 interactions by photoaffinity crosslinking and tandem mass spectrometry

4.1 Introduction

G protein-coupled receptors (GPCRs) comprise 60% of drug targets; however, only 2 mammalian GPCRs have solved high resolution structures to date, bovine rhodopsin [1] and the human beta adrenergic receptor [2, 3]. Intrinsic problems involved with low protein expression, solubility and structural heterogeneity limit structural analysis.

Unlike other GPCRs, bovine rhodopsin is readily available from natural sources in higher quantities. However, its structural determination still required two collaborating research groups to trouble shoot purification and crystallization [1]. Similarly, the recent crystal structure for human adrenergic receptor was quoted in a Science editorial 2007 breakthroughs as a "tour de force" [4] taking nearly 20 years to solve and involved protein engineering [2] as well as the binding of a specific monoclonal antibody [5] to stabilize the dynamic protein at the third intracellular loop.

As a result of limited three-dimensional structure information, rhodopsin is often employed as a structural template for modelling other GPCRs in the family even though it is not always a perfect replacement for structural analysis of the protein of interest. A range of alternative "low resolution" techniques are necessary for finding complementary information for GPCR structure-function relationships; including site-directed mutagenesis, spectroscopy, protein engineering and crosslinking.

A version of this chapter will be submitted for publication. Co-authors in no particular order: Wong JP, Kast J, and Molday RS.

Site-specific photocrosslinkers are employed to capture a snapshot of transient protein interactions and when incorporated at a specific site on the probe may be employed for binding site modelling. Photocrosslinkers such as benzoylphenylalanine (BPA), a modified amino acid, can be directly incorporated into the protein of interest during solid phase synthesis allowing it to be used to study a number of GPCRs which include the cholecystokinin [6], parathyroid hormone (PTH) [7] and substance P (NK-1) receptors [8, 9]. Using photoactive ligand probes it is possible to map receptor interactions expressed in the natural plasma membrane context, allowing accurate binding site modelling. The sites of contact and covalent linkage are commonly traced indirectly with digestion and radioisotopic labels; however, we will describe in this present investigation a method to footprint the contact site by *in-gel* proteolysis and mass spectrometry.

CXCR4, an important GPCR implicated in a variety of diseases such as Acquired Immune Deficiency Syndrome (AIDS) [10], cancerous tumour growth [11] and metastasis [12] will be the focus of this present work. Previous studies have shown that CXCR4 binds to Stromal cell Derived Factor 1 (SDF-1), through a proposed two site binding model common to some other GPCRs [13,14]. The proposed two site binding model was inferred from structure-function studies with *in vitro* folded, full length chemokines [14]. A nuclear magnetic resonance (NMR) solution structure was solved for synthetic SDF-1, identifying two key sites: the disordered N-terminal region (residues 1-8) which is responsible for triggering signal transduction and the RFFESH motif (residues 12-17) docking site [14]. Furthermore, the N-terminus has been shown to be functional as an independent entity on SDF-1, but has reduced binding affinity when the rest of the protein scaffold is absent or interchanged with other chemokine scaffolds [15]. The C-terminal alpha helical domain on SDF-1 is not directly

involved in receptor binding as determined from hybrid forms of SDF-1 [14, 16]. N-terminal residues when combined with other CXC chemokine scaffolds modestly affect receptor binding and activity [14, 16] suggesting that the SDF-1 alpha helical domain is available for manipulation. Residue Leu-66 at the penultimate position in the C-terminal alpha helix has been chosen for affinity tag substitution in the SDF-1 photoaffinity analogs. Here we study a non-active analog that binds to resting receptor. The rationale and design of the following work is illustrated in Figure 4-1 and the sequences for the ligands are outlined in Table 4-1. The modified ligand is bound to the receptor and photo-activated for covalent linkage to stabilize the site-specific interaction (Figure 4-1). The contact site is then identified by proteolytically mapping the interpeptide crosslink site and sequencing the surrounding area to generate a binding footprint of the interaction. The transmembrane, extracellular and intracellular region designations referred to in this current work was predicted with the HMMTOP algorithm for CXCR4 [21,22] (Table 4-2) for this manuscript.

The sits of interaction on SDF-1 have been mapped for receptor binding and activation; however, the site(s) on the corresponding receptor has not yet been determined. Whether two distinct sites exist for docking and activation of the receptor also remains unconfirmed. In addition to defining a preliminary footprint of a non-active, photosensitive SDF-1 analog bound receptor, we have optimized purification and mass spectrometric detection methods that will then be applied to other site-specific probes for activation site(s) identification.

4.2 Methods

Ligand synthesis:

SDF-1, SDF-1 BPA3 and SDF-1 BPA5 were synthesized by solid phase synthesis employing tert-butoxycarbonyl (t-BOC) chemistry on an ABI 430A automated peptide synthesizer (Applied Biosystems, Foster City, CA), purified by reverse phase high performance liquid chromatography (HPLC) and subsequently folded. To this end, purified peptides were dissolved to a final concentration of 1 mg/ml in 1M guanidine, 0.1M TRIS and 10% methyl sulfoxide (Aldrich, Milwaukee, WI) in a sealed container at room temperature overnight in the dark. Folded peptides were subsequently separated from the non-folded by a second run through the HPLC.

^{125}I radiolabeling:

SDF-1 was selectively labeled on tyrosine residues employing a sodium iodide protocol as previously described [14]. SDF-1 (5μg) was combined with 1mCi of $Na^{125}I$ (MP Biomedicals, Canada) and lactoperoxidase (1μg) (Sigma, St. Louis, MO) in 0.25M sodium acetate pH6.5 in a final volume of 113μl and incubated at room temperature for 3min. The reaction was stopped with an equal volume of saturated tyrosine (Sigma, St. Louis, MO) solution. The resultant mixture was passed through a pre-packed G-25 Sephadex column (Pharmacia, Uppsala, Sweden) with gelatin (Biorad, Richmond, CA) as a carrier molecule for separation of the reacted and non-reacted iodine-125. The labeled fractions were pooled and stored at 4°C before use.

Cell Culture:

Murine B300-19 cells expressing CXCR4 (generous gift from Dr. Bernhard Moser, Theodore-Kocher Institute, Switzerland) were grown and maintained in RPMI 1640 (Gibco, Auckland, NZ), supplemented with non-essential amino acids, sodium pyruvate (1mM), β–mercaptoethanol (0.5μM), 10% fetal bovine serum, penicillin/streptomycin and selective agent puromycin (1.5μg/ml) in 5% CO_2 at 37°C.

Binding Competition assay:

CXCR4 expressing murine B300-19 ($2x10^6$ cells) were bound with 4 nM of ^{125}I-SDF-1 in the presence of unlabeled ligand at 0-1μM range in concentration, in 200μl Roswell Park Memorial Institute (RPMI) cell culture media (Gibco, Auckland, NZ), 10mg/ml BSA (Sigma, St. Louis, MO) and 0.1% sodium azide for 20min. The cells were separated through a 2:3 mixture of diacetylphthalate (Sigma, Milwaukee, WI) and dibutylphthalate (Fisher, Fairlawn, NJ) and the resultant cell pellet was measured on the gamma counter. ^{125}I-SDF-1 specifically bound to the cells in the absence of unlabeled competitor was a reference for 100% binding, and all subsequent measurements were normalized against this value for a percentage value for the competitive binding curve (Figure 4-2).

Chemotaxis Assays:

Migration of CXCR4 expressing murine B300-19 cells were assessed with disposable 12 well transwell plates (Costar, Corning, NY) with 5μm pore size polycarbonate membranes. The

chemoattractants, SDF-1 and the analogs were dissolved in HEPES buffered RPMI 1649 (Gibco, Auckland, NZ) supplemented with 10mg/ml bovine serum albumin (BSA) and added to the bottom chamber of the transwell system. CXCR4/B300-19 cells were added to the upper chamber at10^6 cells/well and the system was incubated for 3 h at 37°C in 5% CO_2 atmosphere. Cells migrating to the lower chamber were counted with a hemocytometer and light microscopy.

Calcium flux assays:

CXCR4 transfected murine B300-19 cells ($2x10^6$ cells) were loaded with Fluo-3AM (Molecular Probes, Eugene, OR) and kept in the dark at 37°C until ready to assay. Cells ($1x10^6$) were resuspended in 2.5ml of 25 mM HEPES, 140 mM NaCl, 10mM glucose, 1.8 mM $CaCl_2$, 1 mM $MgCl_2$ and 3 mM KCl, pH 7 and stirred in a Horiba Jobin FL3-22 TAU (Edison, NJ) spectrofluorimeter at 506 nM excitation and 526 nM emission wavelength. After approximately 20 s baseline stabilization, 100nM SDF-1, SDF-1 BPA3 or SDF-1 BPA5 was added to the cells and resultant fluorescence emission and calcium flux was recorded for 60 s time-based acquisition. PBS (25µl) was added at 20 s as a negative control.

Binding and crosslinking:

Transiently transfected HEK 293T cells ($5x10^8$ cells) were washed two times with cold PBS, resuspended in 20ml cold PBS with 1µM SDF-1 BPA 3 or SDF-1 BPA 5 and mixed for 15min at 4°C. The cells were subsequently pelleted and resuspended in fresh cold PBS (20ml) in a 10 cm glass petri dish on ice for UV irradiation (UV Stratalinker 1800, Stratagene, La Jolla,

CA) at 365 nm. The cells were irradiated for 60 min at 10cm from the UV source on ice. After irradiation, the cells were washed twice with cold PBS and lysed in 1ml 2% Triton X-100 with complete protease inhibitors (Roche, Mannheim, Germany) for 1h at 4°C.

Solubilization and 1D4 affinity purification:

The insoluble material and nuclei from the cell solublization was removed by centrifugation at 200,000x g in a TLA.3 (Beckman, Mississauga, ON) rotor for 15min. The resulting supernatant was mixed with 50µl mouse Rho 1D4 monoclonal antibody (mAb) conjugated Sepharose for 1h at 4°C with complete protease inhibitors (Roche, Mannheim, Germany). The flow through was reserved for sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis. The bound Sepharose was washed 10 times with cold 0.2% Triton X-100 in tris buffered saline (TBS) and eluted at room temperature with two consecutive aliquots of one bed volume of 0.5mg/ml 1D4 peptide in 0.2% Triton X-100/TBS for 20 min each. The eluates were pooled and subjected to monomeric avidin (Pierce, Rockford, IL) affinity purification.

Monomeric avidin affinity purification:

Monomeric avidin beads were blocked and primed following manufacturer's protocol (Pierce, Rockford, IL). The 1D4 eluate was diluted 5 fold with TBS and layered onto 50 µl of primed monomeric avidin beads with 1x complete protease inhibitors (Roche, Mannheim, Germany) and mixed for 1h at 4°C. The beads were washed 10X with 0.2% Triton X-100 in

TBS and eluted with two bed volumes of 1% sodium dodecyl sulphate (SDS) in TBS for 10min at room temperature. The final eluate was immediately loaded onto 10% SDS-PAGE for further separation.

SDS-polyacrylamide gel electrophoresis and western blot analysis:

The final eluates were separated on 10% SDS-PAGE with a 5% stacking gel following the Laemmli method. Gel bands were stained with Coomassie Brilliant blue and a corresponding western blot was run in parallel to confirm the presence of ligand and receptor in the complex, prior to identifying a region on the gel for excision and *in gel* digestion. To ensure electrophoretic migration was consistent, duplicate lanes were run on the same polyacrylamide gel and transferred onto the same polyvinylidene fluoride (PVDF) membrane. The membrane was divided into two blots, one for 1D4 labeling and the second for biotin detection. CXCR4-1D4 was detected with mouse anti-1D4 IgG monoclonal antibody and LI-COR goat anti-mouse 1gG conjugated Alexa 680 infrared (IR) secondary antibody (LI-COR, Lincoln, NB). SDF-1 BPA 5 was detected with streptavidin-IR 680 (LI-COR, Lincoln, NB) or streptavidin conjugated horse radish peroxidase (HRP) (Sigma, St. Louis, MO) for enhanced chemiluminescence detection (Amersham, Little Chalfont, Buckinghamshire). Blot visualization and protein densitometry was performed on the LI-COR Odyssey Infrared imager (Lincoln, NB).

In-gel digestion:

Excised Coomassie Blue stained protein gel bands were washed with at least 3 volumes of 25 mM ammonium bicarbonate (ABC) in 50% acetonitrile, shaking until the gel bands were completely destained. The bands were subsequently reduced with 1.5mg/ml dithiothreitol (DTT) at 56°C for 30 min and alkylated with iodoacetamide (10mg/ml) for 20 min at room temperature. The gel bands were washed, dried and rehydrated with 0.1μg sequencing grade bovine chymotrypsin (Roche, Mannheim, Germany) in 50 mM ABC and10mM $CaCl_2$ and incubated at room temperature. The reaction was allowed to proceed for 4 h and subsequently quenched with acetonitrile to dehydrate and remove chymotrypsin from the gel pieces. The acetonitrile supernatant was pooled with final digested extract for analysis. The gel pieces were rehydrated with 0.1μg of porcine trypsin with 5 mM $CaCl_2$ in 50 mM ABC and incubated at 37°C overnight for final digestion. Digests were extracted with 1:1 25mM ABC/ acetonitrile with water bath sonication at room temperature for 15 min and a final extraction with 1:1 5% formic acid/acetonitrile. All extracts were aliquoted into two silicon coated mirocentrifuge tubes (Axygen, Union City, CA), snap frozen and lyophilized to dryness (3 h).

In-solution protein digestion:

SDF-1 (0.25pmol/μL) and SDF-1 BPA 5 (0.25pmol/μL) in 50mM ABC were thermally denatured, reduced with 1mM dithiothreitol at 56°C for 1 h and alkylated with 1mM iodoacetamide at room temperature in the dark for 45min. Sequencing grade bovine chymotrypsin (0.5μg) (Roche, Mannheim, Germany) was added to each sample and the mixture was incubated at room temperature for 2 h with shaking followed by heat inactivation at 90°C

and cooling on ice. Sequencing grade porcine trypsin (0.5μg) (Roche, Mannheim, Germany) was added to the denatured, reduced and alkylated protein for incubation at 37°C overnight with shaking. The reaction was quenched with the addition of trifluoroacetic acid (TFA) to a final concentration of 0.1% v/v.

Mass Spectrometry:

Peptide samples were desalted with C-18 stage tips [18-20]. The desalted samples were analyzed on a linear ion trap – Fourier Transform Ion Cyclotron Resonance Mass spectrometer (LTQ-FTICRMS) (ThermoFisher Scientific, Medina, OH) on-line coupled to an Agilent 1100 Series nanoflow HPLC.

Reversed phase chromatography was run on in-house made 150 mm x75μm fused silica emitters using 3- μm-diameter Reprosil-Pur C-18-AQ beads (Dr. Maisch, www.Dr-Maisch.com). Liquid chromatography gradients were formed with mobile phase A consisting of 0.5% acetic acid and mobile phase B consisting of 0.5% acetic acid and 80% acetonitrile. Gradients were run from 6% B to 30% B over 45 min (flow 0.2 μL/min), then from 30% B to 80% B in the next 20 min, held at 80% B for 10 min (while increasing the flow rate from 0.2 μL/min to 0.6 μL/min), and then dropped to 6% B for another 10 min to recondition the column.

The LTQ-FT (Linear Ion Trap-FTICR) was set to acquire a MS scan at 25,000 resolution from 350 to 1500 Th in the FT-ICR cell, followed by selected ion monitoring (SIM) of the top three peptide ions in each cycle at resolution 50,000 for accurate mass measurement and MS/MS of these 3 ions in the LTQ (Linear Ion Trap) (minimum intensity 500,000 counts).

In-solution digested SDF-1 and SDF-1 BPA 5 sample was mixed 1:1 with a saturated α-cyano-4-hydroxycinnaminic acid (Sigma, St. Louis, MO) in 1:1 0.1%TFA and acetonitrile (v/v) solution and 1μl was spotted onto a target for analysis on a matrix assisted laser desorption ionization time-of-flight/time-of-flight (MALDI TOF/TOF) mass spectrometer, the 4700 Proteomics Analyzer (Applied Biosystems, Foster City, CA).

FTICRMS data extraction and database search:

Monoisotopic peak masses and charge states acquired from the FTICR-MS data were extracted using DTA SuperCharge, part of the MSQuant suite of software (http://msquant.sourceforge.net). Corrected data was submitted for identification employing the Mascot (www.matrix science.com) or X! Tandem search engines against the National Center for Biotechnology Information (NCBIr_20020711) or the Global proteome machine (www.thegpm.org) protein databases. To maintain data interpretation consistency we used one version of the NCBIr database for all replicates. MS peptide and MS/MS error tolerance values were 10ppm and 0.5Da, respectively. Up to three missed cleavages were selected in the database search for the best sequence coverage. The quality of spectra collected for all identified peptides were also manually inspected.

4.3 Results

Binding competition assays:

To ensure that BPA group substitution into SDF-1 does not perturb receptor binding, SDF-1 was iodine -125 labeled on tyrosine residues for binding competition assays. ^{125}I-SDF-1 was competed against increasing concentrations of SDF-1, SDF-1 BPA 3 and SDF-1 BPA 5 (Figure 4-2). Both SDF-1 BPA 3 and 5 compete with ^{125}I-SDF-1 and have similar binding dissociation constant (K_d) values approximated at half saturation of 8 nM and 12 nM, respectively (Figure 4-2, yellow and red trace). The values correspond to approximately 10 nM which is consistent with the dissociation constant for wild type SDF-1 [15] (Figure 4-2).

Chemotaxis and calcium flux functional assays:

The SDF-1 BPA analogs were further tested for activity by *in vitro* chemotaxis assays. Cells expressing CXCR4 were titrated for response to increasing concentrations of SDF-1 as a positive control (Figure 4-3, blue bars). Maximum cell migration was observed at 10 nM concentration for SDF-1 with a chemotactic index of 40 and at 30 nM SDF-1 BPA 3 with a chemotactic index of 37 (Figure 4-3, yellow bars). SDF-1 BPA 5 did not induce cell migration on CXCR4 over the range of concentrations (Figure 4-3, red bars).

To further confirm this observation, an alternative functional assay to measure CXCR4 calcium mobilization or signal transduction in response to chemokine analog was also performed. Calcium mobilization was detected with an intracellular fluorescent calcium indicator that increases in fluorescence intensity when bound to Ca^{2+}. Maximum fluorescence intensity was observed at 500,000 cps when 100 nM SDF-1BPA 3 was introduced at 20s (Figure

4-4A). This was consistent with the positive control SDF-1 response which also reached a maximum of 500,000cps (Figure 4-4C). When SDF-1 BPA 5 was added after 20s, the overall fluorescence intensity did not change over the entire 60s scan (Figure 4-4B) and was similar to the flat trace observed for the negative control (Figure 4-4D).

Nearest neighbour accessibility for crosslinking:

To determine whether the crosslinking site introduced onto the ligand could find a nearby residue in the receptor binding site, a preliminary screen was performed with active and non-active SDF-1 BPA analogs bound to transiently overexpressed CXCR4-1D4 whole cells in the presence and absence of UV irradiation. A single band was recognized for both SDF-1BPA 3 and 5 with streptavidin conjugated horse radish peroxidase (HRP) at approximately 54 kDa for both the peptide and SDS eluates (Figure 4-5A, +UV lanes). In the absence of UV, no biotin containing protein band could be detected (Figure 4-5B, -UV lanes).

Identifying a crosslinked band for digestion and mass spectrometric analysis:

The purified crosslinked protein band was located by western blot detection of both the tagged ligand and receptor. Several receptor isoforms were observed in the pre-column sample on the 1D4 probed blot (Figure 4-6A) representing all forms of CXCR4 in the photoactivated cell lysate. Bands representing the monomer of the receptor were observed as four distinct bands at 54kDa, 50kDa, 45kDa and 44kDa. The 50kDa and 45kDa bands were the most intense and dominant receptor species in the sample. A larger band at 130kDa band was also observed in the pre-column lane (Figure 4-6A). The 130kDa band was also enriched in the final eluate lane

(Figure 4-6A) and may be a multimeric form of the receptor. Two 1D4 positive bands corresponding to the receptor monomer at 50kDa and 44kDa were observed in the final eluate. The 50kDa band was more intense (Figure 4-6A).

When the duplicate set of sample lanes were probed with streptavidin-680 IR dye, a predominant band was observed at 54kDa in the blot (Figure 4-6B). Additional bands at 90kDa (doublet), 130kDa and at the stacking/separating gel interface as well as at the loading dye front and 16kDa were observed (Figure 4-6B). The same banding pattern was observed in the 1D4 affinity purification flow through; however the band at 54kDa was fainter as compared to the pre-column lane suggesting that this streptavidin positive species was depleted by the 1D4 affinity purification. The other higher molecular weight steptavidin positive bands (Figure 4-6B) were at the same intensity as in the corresponding bands in the pre-column sample (Figure 4-6B) indicating that these proteins were not depleted by 1D4 affinity purification and thus were not co-purified with CXCR4-1D4. No bands were observed in the streptavidin Sepharose flow through lane (Figure 4-6B), but the corresponding lane in the 1D4 blot was positive for unbound receptor (Figure 4-6A, streptavidin flow through lane). A single band corresponding to the migration of the dominant band at 54kDa in the pre-column lane was present in the final eluate lane and was more intense after purification and enrichment (Figure 4-6B).

The single streptavidin positive band in the final eluate (Figure 4-6B) aligned virtually exactly with the migration of the most intense receptor band observed in the 1D4 blot at 54kDa and are similar in size in thickness; however, the band in the 1D4 blot is more intense than the band observed in the streptavidin blot (Figure 4-6A and 4-6B, final eluate lanes). Densitometry measurements of the streptavidin probed blots determined that approximately 10% of the receptor was recovered after crosslinking and two rounds of purification.

The final crosslinked and doubly purified sample was separated on a preparative SDS-polyacrylamide gel and stained with Coomassie Blue. A gel band that corresponded to the migration position of the Rho1D4 mAb and streptavidin positive bands at 54kDa observed in Figures 4-6A and 4-6B was excised for digestion and mass spectrometry (preparative gel not shown).

The quality of the mass spectra for each of the peptides was evaluated for adequate signal to noise ratios and isotopic distributions corresponding to the correct charge states and mass accuracy of at least 10ppm for MS data collected. Monoisotopic masses for the peptides identified and mass to charge values for the doubly and triply charged species are shown in Table 4-3.

SDF-1 and CXCR4-1D4 uncrosslinked control digests:

Control digests of the receptor and ligand alone were performed to determine enzyme accessibility and peptides that are detectable for the uncrosslinked proteins involved in the complex. Comparison of the peptides observed between the uncrosslinked (Table 4-4) and crosslinked (Table 4-5) will be employed to identify a potential receptor footprint.

In-solution chymotrypsin/trypsin double digestion of SDF-1 BPA 5 was analysed on the MALDI-MS. The N-terminal peptide containing BPA released from the digestion was used to estimate the mass of the ligand peptide that could possibly crosslink with the corresponding segment of the receptor released from the complex, assuming that the cleavage site(s) on the ligand is not blocked after the crosslinking procedure. This assumption is based on previous

studies in which the SDF-1 N-terminus was extended by an additional glycine residue enhanced rather than abolish receptor activity suggesting that this region can tolerate this modification as it is not completely buried within the receptor when bound [11]. The mass of the ligand N-terminal peptide containing the BPA group was determined and used to predict and calculate the mass of the crosslinked interpeptide that would be released from the digested complex. The ligand N-terminal peptide $K^{1}PVSBSYR^{9}$ with a mass of 1087.6Da was observed in the spectra and the sequence was confirmed by tandem mass spectrometry (Figure 4-8). The resultant ligand-receptor interpeptide crosslink would be composed of a 1087.6 Da mass and the mass of the corresponding CXCR4 extracellular peptide released from the chymotrypsin/trypsin digestion. Two other internal peptides: $^{12}FFESHVAR^{20}$ and $^{28}LNTPNCALQIVARL^{42}$ were also detected for SDF-1 BPA 5 confirming the sequence and identity of the ligand.

The peptides from each of the CXCR4 extracellular loops 1, 2, and 3 were sequenced on the FTICR-MS (Figure 4-9A). This peptides sequenced represent 100% of ECL1, 70% of ECL2, and 70% of ECL3 were sequenced for the unbound CXCR4 digested with chymotrypsin/trypsin (Table 4-5). Other receptor peptides apart from the extracellular loops included a segment from the N-terminus that extends into TM1, TM2, TM4, ICL2 and a peptide from the C-terminus (Figure 4-9A). With this enzymatic combination, we have achieved nearly complete sequence coverage of the extracellular loops on CXCR4, including peptides that anchor the loops that extend into the transmembrane regions.

Receptor peptides recovered from the crosslinked complex:

SDF-1 BPA 5 was bound to CXCR4-1D4 overexpressed transiently in HEK 293T cells and photoactivated for crosslinking. The resulting crosslinked complexes were isolated by affinity purification and further separation on SDS-PAGE. The complex was excised from the gel for digestion and mass spectrometric fingerprinting. Receptor sequence coverage recovered for the extracellular regions were primarily from ECL2, ECL3 and a small N-terminal peptide (Figure 4-9A, red highlighted segments). The proportion of the extracellular loops that were sequenced was 20% of ECL1, 50% of ECL2 and 70% of ECL3 (Table 4-5).

The C-terminal peptides ^{310}TSAQHALTSVSR^{322} was most frequently identified from the crosslinked complex, followed by ^{110}AVHVIYTVN^{119} from ECL1, ^{174}ANVSEADDRY^{184} from ECL2 and then ^{271}QGCEFENTVKHK^{282} from ECL3 (Figure 4-7). The peptide ^{36}NKIFLPTIY^{45} representing a segment in the N-terminal just proceeding TM1 was identified in 2 of the 15 replicates. Peptides of the transmembrane helices: TM1, TM4 and TM7 were also sequenced in 10-30% of the experiments (Figure 4-7).

Comparing crosslinked and unbound receptor sequence coverage:

An analysis of the peptides sequenced from the crosslinked and unbound receptor reveal overlapping and unique sequence coverage (Table 4-6). Extracellular stretches observed uniquely for the unbound receptor include: ^{36}IFLPTIY^{45}(C-terminus), ^{87}VITLPFWAVDAVANW^{102} (TM4/ECL1), ^{102}YFGNFLCK^{110} (ECL1), ^{184}ICDRFYPNDLW^{195} (ECL2) and ^{264}ILLEIIK^{272} (ECL3) (Table 4-6). Only a few stretches unique to the crosslinked receptor were observed, ^{28}FREENANF^{36} (N-terminus) and ^{283}ISITEALAFF^{293} (ECL3/TM7)

(Table 4-6). Peptides sequenced that were common to the unbound and crosslinked receptor were: ^{34}NK^{36}(C-terminal), ^{110}AVHIYTVN^{119} (ECL1), ^{166}LTIPDFIFANVSEADDR^{183} (TM4/ECL2) and ^{271}QGCEFENTVHK^{282} (ECL3) (Table 4-6). These peptides are indicative ot regions on the receptor that are accessible for proteolysis before and after photoactivation.

4.4 Discussion

Previous structure-function work on SDF-1 [14] has provided the basis for the design and synthesis of site-specific crosslinkers for CXCR4 binding site mapping. An engineered site-specific SDF-1 crosslinking analog was synthesized and thoroughly characterized. We have optimized and defined binding, crosslinking and purification conditions for the complex. Tandem mass spectrometry was applied to footprint the ligand- bound crosslinked complex. A large proportion of the extracellular loops were sequenced, narrowing down the possible regions available for interaction with the SDF-1 N-terminus (Figure 4-9). Elimination of these regions as possible sites for interaction will allow us to focus on other regions that have not been sequenced in greater detail. Sequencing of the exact site of crosslinking would provide direct evidence for interaction but are challenged by intrinsic and technical limitations. Identification and/or data interpretation of the peptide residue(s) involved in the peptide-peptide crosslink is complicated, as product ions will arise from different peptides derived from the same crosslink [23].

Despite these obstacles, we were able to generate a preliminary footprint of CXCR4 which was made possible due to the high purity of our purification methods and superior performance of instrumentation employed.

SDF-1 N-terminus flexibility allows for photocrosslinker substitution:

The strategy was successful due to the flexibility of the SDF-1 N-terminal which has enabled BPA substitution without affecting ligand folding and receptor binding. Amino acid residues with properties similar to BPA were chosen for substitution (Table 4-1). Both active and non-active SDF-1 BPA analogs with near native binding affinities were generated for these studies to distinguish between "activating" versus "non-activating" interaction sites on CXCR4. The inactive analog SDF-1BPA5 was initially employed to identify binding or docking sites.

Binding site residue accessibility and proximity for photocrosslinking:

We screened the analogs for crosslinking feasibility to ensure that the analogs have nearby residues available for covalent linkage. The formed complexes were purified by 1D4 affinity chromatography. Receptors that were complexed to the biotin containing ligand were detected by western blotting and streptavidin-HRP probing. A single band at approximately 54 kDa, was recognized by streptavidin-HRP and enhanced chemiluminescence detection for the 1D4 peptide and SDS eluates for both SDF-1 BPA3 and BPA5 UV activated samples (Figure 4-5A). This suggests that there is a population of CXCR4-1D4 that co-purifies with SDF-1 BPA3 and BPA5 after 1D4 affinity purification, denaturation and disulfide bond reduction and SDS-PAGE. The corresponding negative control shows that in the absence of UV, following the same purification procedures did not recover a protein band at 54kDa (Figure 4-5B, -UV lanes). Under these purification conditions the SDF-1 ligand did not purify with CXCR4-1D4, unless there was UV activation and thus stable covalent linkage (Figure 4-5 A, +UV). This strongly

suggests that covalent linkage occurred resulting in a band containing both the receptor and ligand affinity tags at the expected crosslinked complex molecular weight.

High sample purity is required for MS mapping of the crosslinked complex:

The stringency of two complementary purifications was necessary for identifying both CXCR4 and SDF-1 BPA5. We were unable to identify both proteins in the complex in one affinity purification via either tag (data not shown).

A single band migrating at 54kDa was recognized by Rho 1D4 mAb and streptavidin after both affinity purification steps (Figure 4-6). Although the 54kDa band in the 1D4mAb blot and the streptavidin blot were of different intensity, the bands displayed similar shape, thickness and migration suggesting that they are migrating as a single entity rather than two co-migrating proteins. Differences in intensity between the staining of the final eluate lane in the1D4 and streptavidin blots may be attributed to epitope masking or anomalous migration patterns due to non-linear nature of the denatured crosslinked complex. Moreover, differences between antibody affinity and detection reagents employed for each of the blots also contribute to intensity variations of the protein bands. Thus, the intensity of the bands is not a reflection of the binding stoichiometry of the two proteins in complex.

High mass accuracy and duty cycle for protein identification:

The LTQ-FT-MS was employed to analyse extracted peptides from the *in-gel* digested crosslinked complex for interaction site mapping. Intrinsic heterogeneity of the receptor, mixed

products from receptor photocrosslinking, as well as the promiscuity of chymotrypsin in *in-gel* double digestions may produce an overwhelming number of different peptides for detection. Fragment mass spectral data acquisition is information dependent and only a certain number of peptides that meet threshold requirements are selected for sequencing. Complex mixtures of peptides in a sample may bury lower abundant peptides of interest preventing analysis and accurate identification of both receptor and ligand in the complex.

We have taken advantage of the high duty cycle and sensitivity of the LTQ-FT-MS to detect both respective proteins in the crosslinked complex. Both the receptor and ligand were identified in a single band on the LTQ-FT-MS in 15 independent experiments. The high duty cycle and the high mass accuracy of the FTICR-MS also increased unique peptide detection. Twelve unique peptides from CXCR4 in the crosslinked complex were identified.

Inferences made from the peptides sequenced

Peptides identified from ECL2 and ECL3 were observed in approximately 50% of the replicates. The masses observed did not correspond to crosslinking derivatized masses and were consistently detected in the digested complex. This suggests that these sequences are not directly involved in covalent bonding to SDF-1 BPA5 upon photoactivation. The high frequency observed and quality of spectra collected for these peptides are an indication that these peptides were available in reasonable quantity and quality for sequencing. The presence of crosslink-modified species would dilute out the concentration of the non-modified peptides from the receptor complex preventing data acquisition for these regions, which were selected for MS/MS analysis.

From comparison of the sequence coverage in the unbound and crosslinked receptor indicate that much of the ECL1 and the N-terminus were not recovered from the crosslinked complex, while they were present in the unbound receptor analysis (Figure 4-9, Table 4-6). Interestingly, the transmembrane segments at TM3 that anchors ECL1 was sequenced. Moreover, less of ECL3 extending from TM6 was sequenced but more was sequenced at TM7 for the crosslinked complex. This suggests that these sites are accessible for enzymatic cleavage after crosslinking despite reduced external loop sequence coverage which may be blocked by ligand binding.

Sequence coverage for ECL3 was the same for both crosslinked and unbound receptor suggesting that enzymatic accessibility in this region upon crosslinking is not affected. The decrease in sequence coverage in ECL1 and at a lesser extent ECL2 is an indication that these sites may be blocked or partially blocked by ligand crosslinking. Peptides were not observed from ECL1 for the crosslinked receptor but were detectable in the unbound form. These regions may contain potential interaction sites for SDF-1 BPA5. Possible mass combinations of the N-terminal peptide of SDF-1 BPA5 were combined with the masses of the peptides suggested in the above and searched against experimental data. However, we were unable to find any match at this time.

Only a small proportion of CXCR4 N-terminus was sequenced in both the crosslinked (19%) and unbound receptor (16%) which may be attributed to heterogeneity in this region. Heterogeneous glycosylation at CXCR4 N-terminus complicates identification since the resulting set of heterogeneous peptide masses are difficult to account for in our calculations as well as dilution of signal caused by signal splitting. Thus, the sensitivity of detection of the N-terminus is reduced, as peptides containing differing degrees of PTM will split signals reducing

overall signal intensity for that particular region on the receptor. Therefore, the possible involvement of the N-terminus region requires further investigation with glycosidases to remove the heterogeneous modification to consolidate the peptides excised from the region. Peptides detected and sequenced after carbohydrate removal from the complex will eliminate regions on the receptor N-terminus that are not modified by the crosslinking procedure to further refine the ligand bound receptor footprint.

Footprinting with the photosensitive probe SDF-1 BPA 5 suggest that position 5 potentially may interact with a site in the receptor N-terminus or ECL1 which is possibly obstructed upon crosslinking preventing mass spectrometric detection. These regions need to be verified with more specific targeted approaches such multiple reaction monitoring (MRM) and heavy isotopic and light label at a site that will be excised with the crosslinked interpeptide and traced by correlation analysis (discussed further in Chapter V).

Table 4-1: Site-specific photoaffinity crosslinking SDF-1 analog and SDF-1 sequence alignment. Sites of photocrosslinker, benzoylphenylalanine substitution are represented as B in the above. Lysine-ε-biotin substitution at position 66 is denoted K_b.

residue no.	1	5	10	15	20	30	35	40	45	50	55	60	65	70
SDF-1α	KPVSL	SYRCP	CRFFE	SHVAR	ANVKH	TPNCA	LQIVA	RLKNN	NRQVC	IDPKL	KWIQE	YLEKA	LN	
B3L66K$_b$	. . B . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	K$_b$	
B5L66K$_b$	. . . . B	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	K$_b$.	

Table 4-2: Transmembrane protein topology for human CXCR4 predicted with HMMTOP version 2.0 online software (http://www.enzim.hu/hmmtop/)[19, 20]. "H" indicates amino acids predicted in the transmembrane domains, "i" for intracellular domains (upper case "I" represents the intracellular C-terminal) and "o" for extracellular domains (upper case "O" denotes the extracellular N-terminal). The top row denoted "seq" represents the primary sequence for human CXCR4 and "pred" represents the predicted topology for the receptor. Extracellular sequences from the crosslinked receptor, indicated in blue text were sequenced by tandem mass spectrometry.

seq	MEGISSIPLP LLQIYTSDNY TEEMGSGDYD SMKEPCFREE NANFNKIFLP	50
pred	OOOOOOOOOO OOOOOOOOOO OOOOOOOOOO OOOOOOoooo oooooooooo	
seq	TIYSIIFLTG IVGNGLVILV MGYQKKLRSM TDKYRLHLSV ADLLFVITLP	100
pred	oHHHHHHHHHH HHHHHHHHHH Hiiiiiiiii iiiiiiiHHH HHHHHHHHHH	
seq	FWAVDAVANW YFGNFLCKAV HVIYTVNLYS SVLILAFISL DRYLAIVHAT	150
pred	HHHHHHHooo oooooooooo oHHHHHHHHH HHHHHHHHHH iiiiiiiiii	
seq	NSQRPRKLLA EKVVYVGVWI PALLLTIPDF IFANVSEADD RYICDRFYPN	200
pred	iiiiiiiiii iiHHHHHHHH HHHHHHHHHH HHHHHoooo oooooooooo	
seq	DLWVVVFQFQ HIMVGLILPG IVILSCYCII ISKLSHSKGH QKRKALKTTV	250
pred	oooooooooo oHHHHHHHHH HHHHHHHHHH Hiiiiiiiii iiiiiiiiiH	
seq	ILILAFFACW LPYYIGISID SFILLEIIKQ GCEFENTVHK WISITEALAF	300
pred	HHHHHHHHHH HHHHHHHHHo oooooooooo oooooooooo oooooHHHHH	
seq	FHCCLNPILY AFLGAKFKTS AQHALTSVSR GSSLKILSKG KRGGHSSVST	350
pred	HHHHHHHHHH HHHHHiiiii iiiiiiiiii IIIIIIIIII IIIIIIIIII	
seq	ESESSSFHSS 360	
pred	IIIIIIIIII	

Table 4-3: Receptor peptides sequenced from SDF-1 BPA5 crosslinked CXCR4-1D4 complex band *in-gel* digested with chymotrypsin and trypsin sequentially. Peptides listed were identified 3 or more times from 15 independent experiments in which both proteins were identified in the same protein band.

Location	Sequence	Monoisotopic mass (experimental) (Da)	Mass observed (m/z)
TM4/ECL 2	^{166}LTIPDFIFANVSEADDR^{183}	1921.95	1043.51, 961.98
ECL 2	^{174}ANVSEADDRY^{184}	1138.48	570.25
ECL 3	^{264}ILLEIIK^{272}	840.57	421.28
ECL 3	^{271}QCEFENTVHK^{282}	1347.59	674.80, 450.20
N-terminal	^{26}FREENANFNK^{36}	1267.59	423.53
C-terminal	^{309}TSAQHALTSVSR^{322}	1256.65	629.33, 419.89

Table 4-4: Receptor peptides sequenced of unbound CXCR4-1D4 digested *in-gel* with chymotrypsin and trypsin sequentially. Peptides compiled in this table were obtained from both the LC-Qstar and FTICR mass spectrometers.

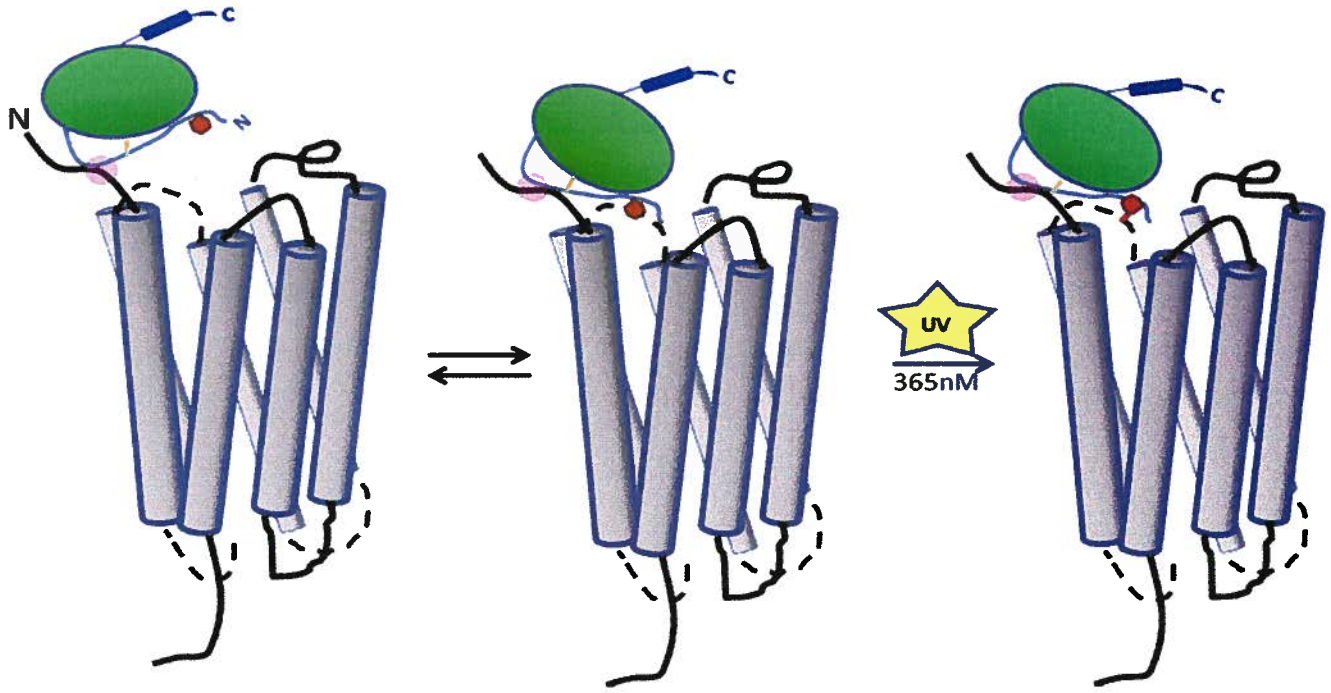
Location	Sequence	Monoisotopic mass (experimental) (Da)	Mass observed (m/z)
TM2/ECL1	^{87}VITLPFWAVDAVANW^{102}	1701.89	845.19
ECL1	^{102}YFGNFLCK^{110}	1047.48	520.58
ECL1	^{110}AVHIYTVN^{119}	351.20	700.39
TM4/ECL 2	^{166}LTIPDFIFANVSEADDR^{183}	1921.95	1043.51, 961.98
ECL 2	^{174}ANVSEADDRY^{184}	1138.48	570.25
ECL2	^{184}ICDRFYPNDLW^{195}	1497.67	749.84
ECL 3	^{264}ILLEIIK^{272}	840.57	421.28
ECL 3	^{271}QCEFENTVHK^{282}	1347.59	674.80, 450.20
N-terminal	^{34}NKIFLPTIY^{45}	1107.63	550.42
C-terminal	^{309}TSAQHALTSVSR^{322}	1256.65	629.33, 419.89

Table 4-5: Proportion of CXCR4 each extracellular loop sequenced before and after photocrosslinking

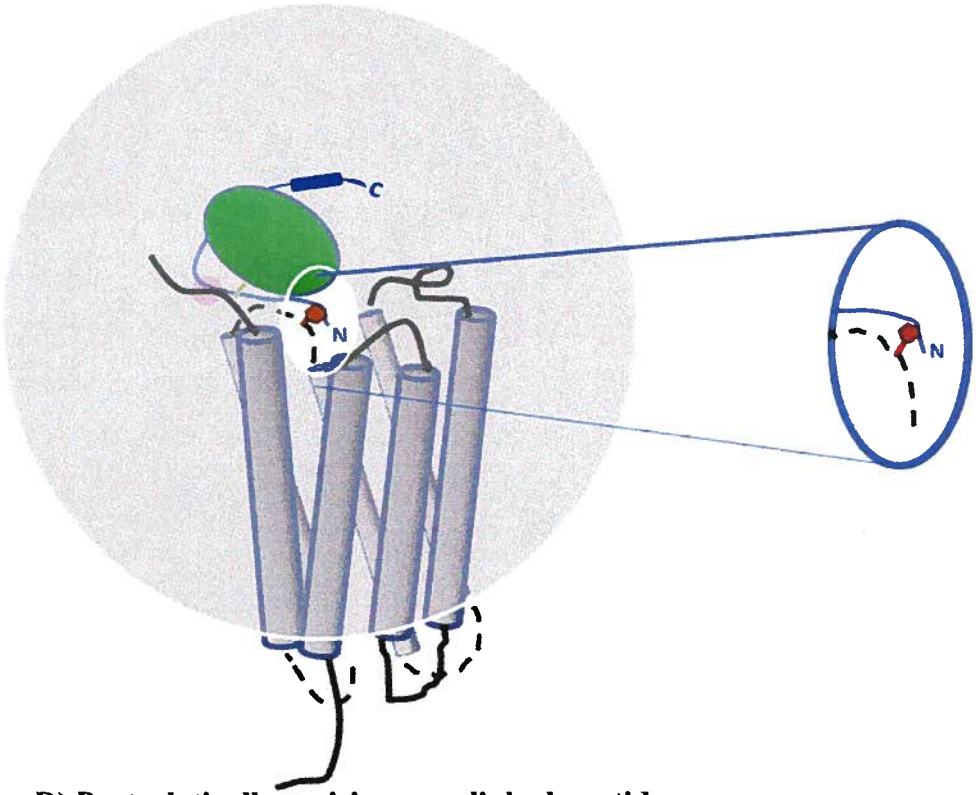
Unbound		**Crosslinked**		**Location**
%	**Residues**	**%**	**residues**	
16	7/43	19	8/43	N-terminus
100	14/14	20	3/14	ECL1
70	18/26	50	11/23	ECL2
70	18/26	70	18/26	ECL3

Table 4-6: Unique and overlapping peptide coverage for the unbound and crosslinked CXCR4-1D4.

Unbound		**Crosslinked**		**Overlapping**	
Sequence	**location**	**Sequence**	**Location**	**Sequence**	**Location**
^{36}1FLPTIY^{45}	N-terminus	^{28}FREENANF^{36}	N-terminus	^{34}NK^{36}	N-terminus
^{87}VITLPFWAVDAVANW^{102}	TM2/ECL1	^{283}ISITEALAFF^{293}	ELC3/TM7	^{110}AVHIYTVN^{119}	ECL1
^{102}YFGNFLCK^{110}	ECL1			^{166}LTIPDFIFANVSEADDR^{183}	TM4/ECL2



A) Docking B) Binding C) Photocrosslinking



D) Proteolytically excising crosslinked peptides

Figure 4-1: Proposed schematic for crosslinking studies to determine docking and activation site(s) A) represents an SDF-1 analog (green) and receptor (grey) docking. Then enabling B) SDF-1 N-terminal binding to the extracellular loops (exact site is unknown, a random site was chosen for this theoretical illustration), C) crosslinker photactivation by UV, and covalent linkage to nearest neighbouring residues and D) punching out the crosslinked peptides with proteases for subsequent mass spectrometric analysis.

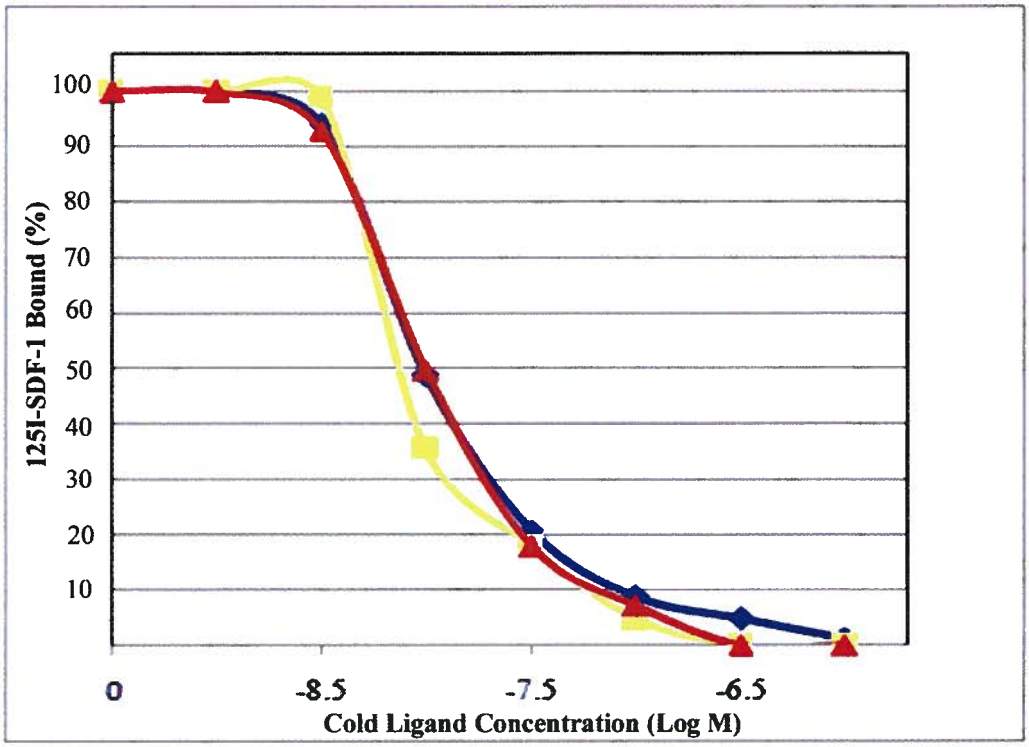


Figure 4-2: Binding competition curves for binding affinity determination for SDF-1 BPA3 (yellow), BPA5 (red) and SDF-1 (blue) against 125I-SDF-1on CXCR4 transfected B300-19 cell line. Each data point is an average calculated from 3 replicates.

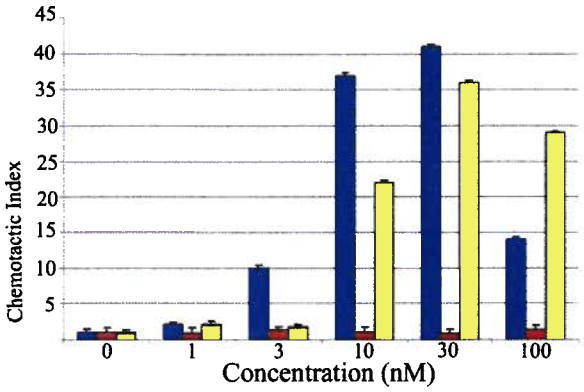


Figure 4-3: Chemotaxis activity assays for SDF-1 and SDF-1 BPA analogs. Disposable transwell dual-chambered plates were employed to form an artificial chemokine or analog gradient for cell migration. Cells migrated to the chemokine/analog chambers were innumerated by light microscopy and a hemocytometer. The chemotactic index was calculated as the ratio of cells in the lower chamber in the presence over the absence of SDF-1 (blue), SDF-1 BPA 3 (yellow) and SDF-1 BPA 5 (red) over a concentration range. Standard error was calculated from 4 experiments.

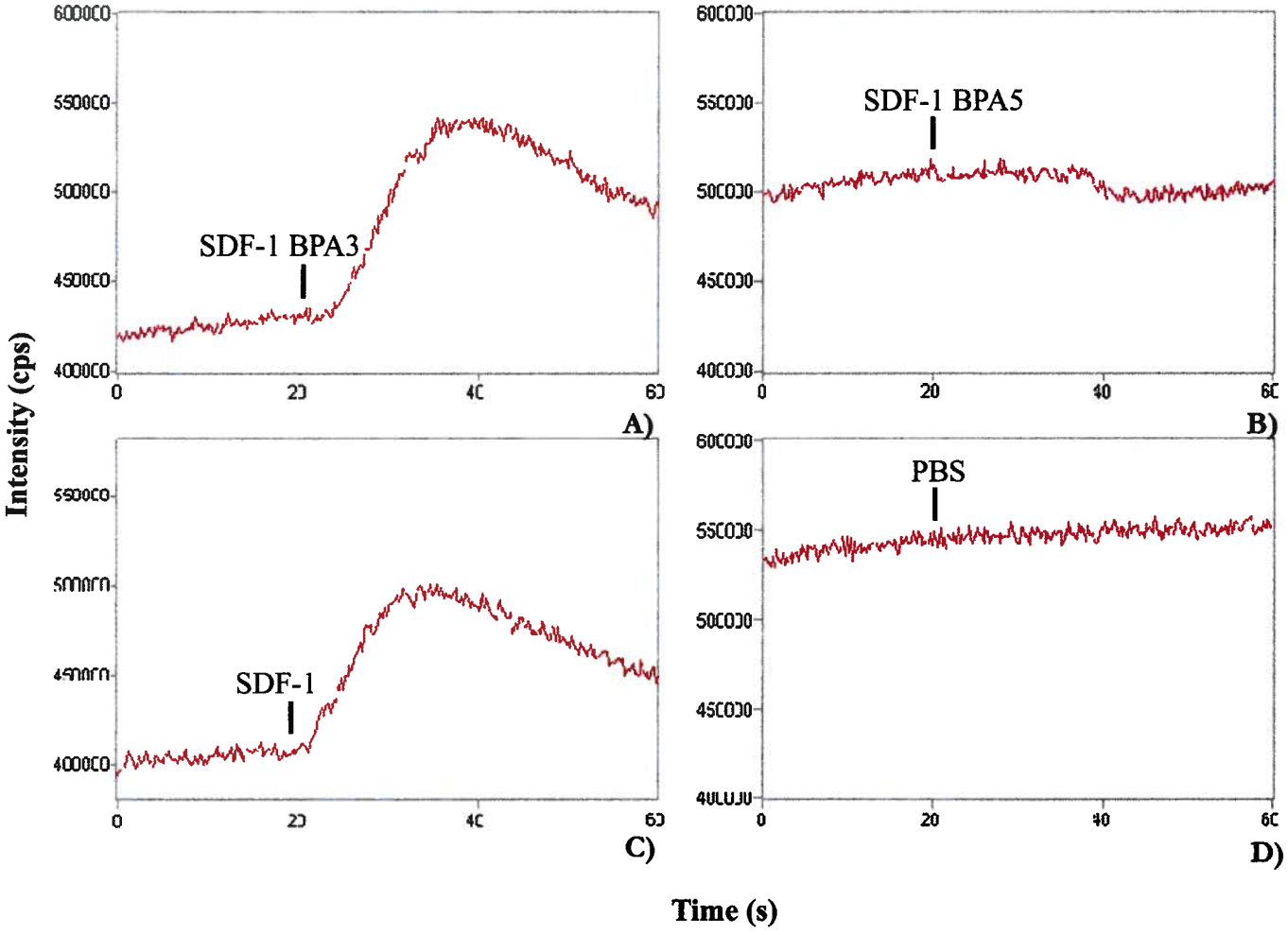


Figure 4-4: Intracellular calcium mobilization functional assays for SDF-1 and analog stimulated CXCR4 expressing B300-19 cells. Fluorescence changes were monitored for cells loaded with calcium indicator Fluor-3AM with the addition of A) SDF-1 BPA3, B) SDF-1 BPA5, C) SDF-1 and D) PBS after 18s baseline over 60 scan. This is one representative experiment of 3 replicates

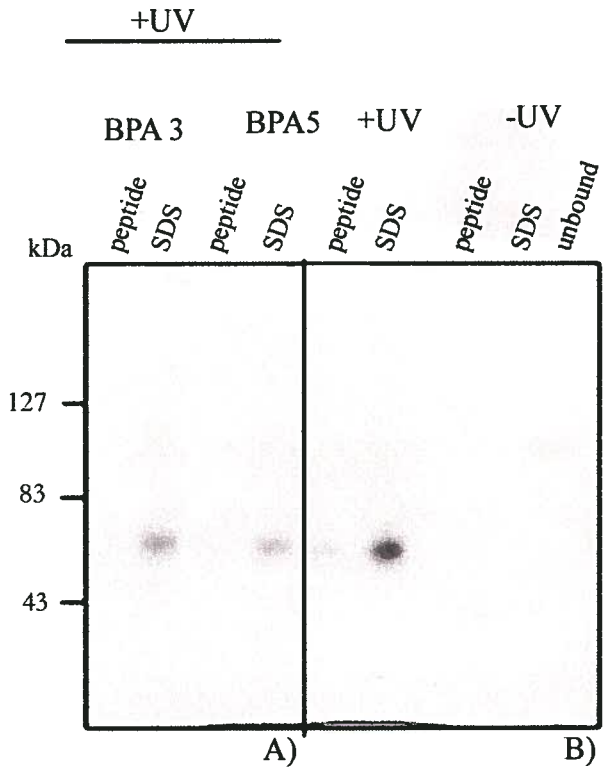


Figure 4-5: Preliminary screen for nearest neighbouring CXCR4 residues available for SDF-1 BPA3 and BPA 5 for photolabelling. After UV activation the complexes conjugated to A) SDF-1 BPA 3 and BPA5 were isolated via the receptor 1D4 tag and separated on SDS PAGE and then transferred for western blot detection of the ligand tag. Similar purification and gel separation and membrane detection was performed in B) for the corresponding +/- UV source controls for SDF-1 BPA5

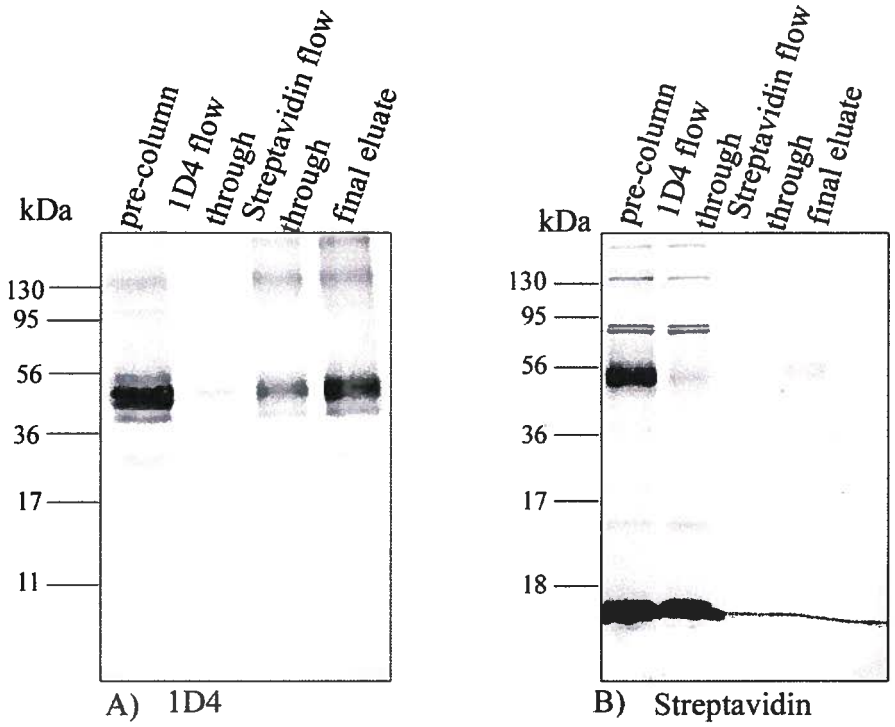


Figure 4-6: Crosslinking and crosslinked complex purification and verification by western blot detection. The steps leading to the purification of the photocrosslinked receptors were separated on SDS-PAGE transferred onto membranes for receptor and ligand detection. For consistency duplicate lanes of the same purified samples were run on the same gel and separated into two for A) 1D4mAbs and B) streptavidin-680 IR dye probing and LICOR Odyssey detection.

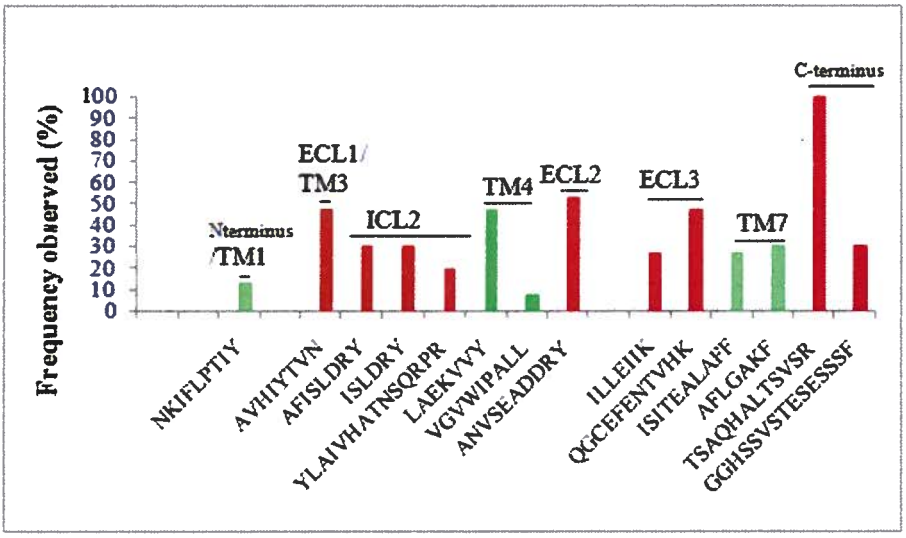


Figure 4-7: Frequency of observed peptides from chymotrypsin, trypsin double digestion of the photocrosslinked complex in 15 replicate experiments in which both SDF-1 and CXCR4 were identified. The red and green columns represent extracellular or intracellular and transmembrane segments of the receptor, respectively.

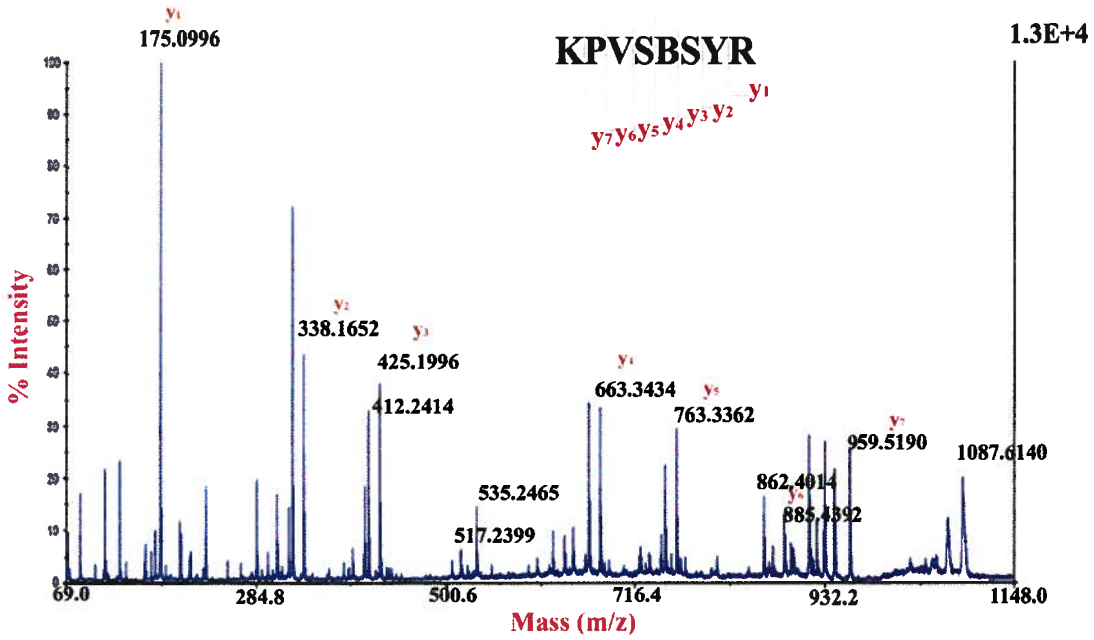


Figure 4-8: MALDI MS/MS spectrum for chymotrypsin, trypsin digested SDF-1 BPA5 N-terminal peptide 1087.6Da. The y-ion series is indicated in red.

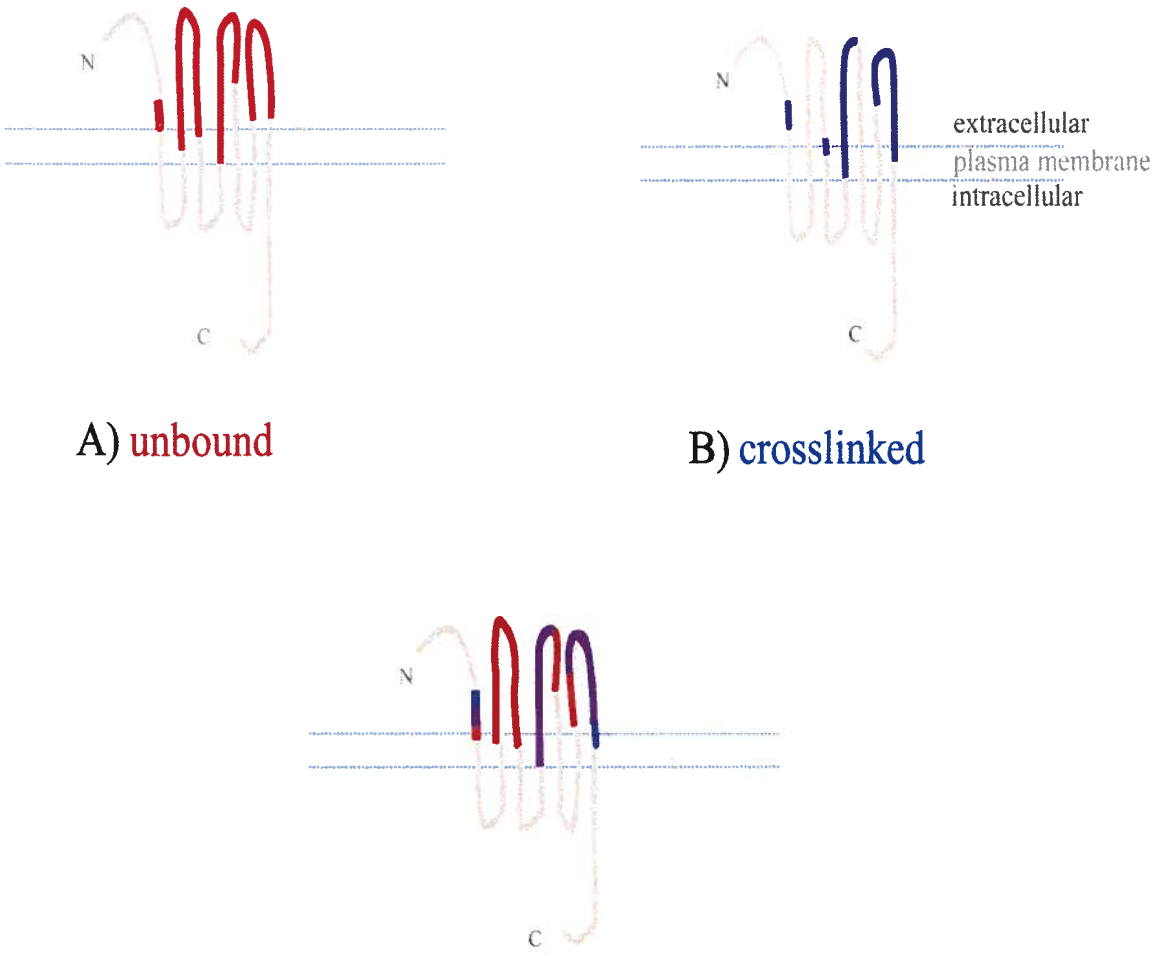


Figure 4-9: Diagrammatic representation of receptor sequence coverage of crosslinked and unbound receptor. Receptor segments obtained from the A) unbound receptor is indicated in red, B) photocrosslinked complex in blue, and C) the overlapping sequence coverage in purple.

4.5 References

1. Palczewski, K., Kumasaka, T, Hori, T, Behnke, CA, Motoshima, H, Fox, BA, Le Trong, I, Teller, DC, Okada, T, Stenkamp, RE, Yamamoto, M and Miyano, M., *Crystal structure of rhodopsin a G protein-coupled receptor.* Science, 2000. **289**(August 4): p. 739-745.

2. Cherezov V, et al., *High-resolution crystal structure of an engineered human beta2-adrenergic G-protein coupled receptor.* Science, 2007. **318**(5854): p. 1258-1265.

3. Rasmussen SG, et al., *Crystal structure of the human beta2 adrenergic G-protein-coupled receptor.* Nature, 2007. **450**(7168): p. 383-387.

4. Kennedy, D., *Breakthrough of the year.* Science, 2007. **318**: p. 1846.

5. Day PW, R.S., Parnot C, Fung JJ, Masood A, Kobilka TS, Yao XJ, Choi HJ, Weis WI, Rohrer, and Kobilka BK, *A monoclonal antibody for G protein-coupled receptor crystallography.* Nature Methods, 2007. **4**(11): p. 927-929.

6. Gimpl, G., Anders, J, Thiele, C and Fahrenholz, F., *Photoaffinity labeling of the human brain cholecystokinin receptor overexpressed in insect cells. Solubilization, deglycosylation and purification.* Eur. J. Biochem, 1996. **237**: p. 768-777.

7. Li, Y.-M., Marnerakis, M, Stimson, ER, and Maggio, JE., *Mapping peptide-binding domains of the substance P (NK-1) receptor from P388D$_1$ cells with photolabile agonists* Journal of Biological Chemistry, 1995. **270**(3): p. 1213-1220.

8. Boyd, N., Kage, R, Dumas, JJ, Krause, JE and Leeman, SE., *The peptide binding site of the substance P (NK-1) receptor localized by a photoreactive analogue of substance P: Presence of a disulfide bond.* PNAS, 1996. **93**: p. 433-437.

9. Girault S, Sagan S, Bolbach G, La Veille S, and Chassaing G, *The use of photolabeled peptides to localize the substance-P-binding site in the human neurokinin-1 tachykinin receptor.* Eur. J. Biochem, 1996. **240**: p. 215-222.

10. Connor RI, et al., *Changes in coreceptor use correlates with disease progression in HIV-1 infected individuals.* J. Exp. Med., 1997. **185**: p. 621-628.

11. Orimo A, Gupta PB, Sgroi DC, et al., *Stromal fibroblasts present in invasive human breast carcinomas promotes tumour growth and angiogenesis through elevated SDF-1/ CXCL12 secretion.* Cell, 2005. **121**: 335-348.

12. Muller A, Homey B, Soto H, et al., *Involvement of chemokine receptors in breast cancer metastasis.* Nature, 2001. **410**: 50-56.

13. Siciliano SJ, Rollins TE, DeMartino J, Konteatis Z, Malkowitz L, Van Riper G, Bondy S, Rosen H, Springer MS, *Two-site binding of C5a by its receptor: an alternative binding paradigm for G protein-coupled receptors.* PNAS, 1994. **91**(4): 1214-1218.

14. Crump, M., et al., *Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1.* EMBO, 1997. **16**(23): p. 6996-7007.

15. Loetscher P, et al., *N-terminal peptides of stromal cell-derived factor-1 with CXC chemokine receptor 4 agonist and antagonist activities.* Journal of Biological Chemistry, 1998. **273**(35): p. 22279-22283.

16. Tudan C, et al., *C-terminal cyclization of an SDF-1 small peptide analogue dramatically increases receptor affinity and activation of the CXCR4 receptor.* J. Med. Chem., 2002. **45**: p. 2024-2031.

17. Rappsilber J, Mann M, and Ishihama Y, *Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.* Nature Protocols, 2007. **2**(8): p. 1896-1906.

18. Rappsilber J and Mann M, *Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics.* Anal. Chem., 2003. **75**(663-670).

19. Ishihama Y, Rappsilber J, and Mann M, *Modular stop and go extraction tips with stacked disks for parallel and multidimensional peptide fractionation in proteomics.* Journal of Proteome Research, 2006. **5**: p. 988-994.

20. Crump MP, G.J., Loetscher P, Rajarathnam K, Amara A, Arenzana-Seisdedos F, Virelizier JL, Baggiolini M, Sykes BD and Clark-Lewis I, *Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1.* EMBO, 1997. **16**(23): p. 6996-7007.

21. Tusnády GE, S.I., *Principles governing amino acid composition of integral membrane proteins: applications to topology prediction.* J. Mol. Biol., 1998. **283**: p. 489-506.

22. Tusnády GE and, S.I., *The HMMTOP transmembrane topology prediction server.* Bioinfomatics, 2001. **17**(9): p. 849-850.

23. Zhou, A., Bessalle R, Bisello, A, Nakamoto, C, Rosenblatt, M, Suva, LJ and Chorev, M., *Direct mapping of an agonist-binding domain within the parathyroid hormone/parathyroid hormone-related protein receptor by photoaffinity crosslinking.* PNAS, 1997. **94**: p. 3644-3649.

24. Ruhmann, A., Kopke, AKE, Dautzenberg, FM, and Spiess J., *Synthesis and characterization of a photoactivatable analog of corticotropin-releasing factor for specific receptor labeling.* PNAS, 1996. **93**: p. 10609-10613.

25. Williams K and Schoelson SE., *A photoaffinity scan maps regions of the p85 SH2 domain involved in phosphoprotein binding.* Journal of Biological Chemistry, 1993. **268**(8): p. 5361-5364.

26. Soderblom EJ, et al., *Tandem mass spectrometry acquisition approaches to enhance identification of protein-protein interactions using low-energy collision-induced dissociative chemical crosslinking reagents.* Rapid Communications in Mass Spectrometry 2007. **21**: p. 3395-3408.

Chapter V: Conclusions and future directions

Inherent and technical challenges undermine the structural analysis of the GPCRs despite the relevance of this family of receptors in pharmaceuticals. As such, bovine rhodopsin and human β_2 adrenergic receptor are the only two high resolution structures available for mammalian GPCRs. Classical approaches to improving low expression and solubility for GPCRs involve painstakingly time consuming, trial-and-error approaches to find suitable expression and purification systems. In particular, heterologous expression systems that are adapted to large-scale production have seemingly become a requirement for structural studies; however, these systems may not be equipped with the transcription, translation and protein folding machinery for physiologically relevant expression of the receptor studied. Given the paucity of high resolution structures that are available, we have developed a general, versatile and robust enrichment method that is applicable to transiently expressed GPCRs as well as ABC transporters in a mammalian expression system. Enrichment has enabled characterization of a model GPCR, CXCR4 by sensitive mass spectrometric techniques.

Taking into consideration of artifactual expression possible in heterologous systems, mammalian GPCRs investigated in this dissertation were maintained in contextually relevant mammalian expression systems.

5.1 Expression systems tested

A number of mammalian expression systems were investigated to improve receptor quantities for characterization which include fluorescence activated cell sorting of high expressing stably transfected cells, a tetracycline inducible CXCR4 cell line and transient

transfectants (Appendix B). The highest expression was observed for the transient HEK 293T cell transfectants which were employed for the purification and mass spectrometric characterization of CXCR4.

5.2 Purification

Purification strategies that involve three-dimensional shape recognition are detergent and chaotroph sensitive and tend to be applicable to soluble proteins as opposed to hydrophobic proteins. Purification strategies that can tolerate low levels of detergent such as immunoaffinity purifications that recognize small linear tags are employed to study membrane proteins. However, the feasibility of an affinity tag on a receptor requires individual expression and functional testing. We have examined the 1D4 affinity tag for the study of membrane proteins. The 1D4 tag which was derived from a membrane protein, bovine rhodopsin, is suitable for GPCRs as well as ABC transporters. The specificity of this enrichment has allowed for protein at as little as subpicomolar quantities of enriched protein to be analysed by mass spectrometry.

Although affinity tags are powerful tools for the purification of proteins that do not have suitable purification regents, they are not suitable for detection of endogenous protein. GPCRs are notoriously antigenic and conformationally heterogeneous and thus antibodies raised against these regions would not detect all receptor present in a sample. Combinations of complementing antibody are required to ensure that all receptor present is detected.

We have optimized and developed a general purification method for membrane proteins which was tested on the receptors: CXCR4, CCR5, ABCA1 and ABCA4 described in Chapter II. The strategy involved a small membrane protein related tag, the 1D4 tag that is versatile and specific. With this approach, we determined one constant detergent, wash conditions and elution

conditions for the purification of all 4 receptors. Receptor purification enabled human receptors expressed in a human cell line to be characterized including PTM site mapping employing mass spectrometry. From this analysis of our model GPCR, CXCR4, we have identified a peptide that contains a consensus sequence for N-linked glycosylation that is not modified. This finding was atypical to other GPCRs as predicted from homology comparisons.

5.2.1 Anti-CXCR4 antibodies

The most commonly used anti-human CXCR4 monoclonal antibody12G5, is one such antibody that recognizes a heterogeneous epitope on the receptor. 12G5 is a neutralizing antibody that recognizes the second extracellular loop on CXCR4 and is commonly employed in flow cytometry and immunofluorescence applications [1] and when used alone may underestimate CXCR4 concentrations at the cell surface.

Similarly, antibodies raised to the N-terminus of CXCR4 have limited applicability as they overlook PTMs that are present. We have observed that antibodies raised against the N-terminus of CXCR4 were either synthetically generated or bacterially expressed without the relevant PTMs present on the N-terminus (Chapter III). From our mass spectrometric characterization of CXCR4, we have observed that there is an N-linked glycosylation site at asparagine 11, which was confirmed with site-directed mutagenesis, Con A labeling (Chapter II) and PNGase F treatment (data not shown).

In response to the lack of effective affinity reagents for human CXCR4, we have developed an antibody validation platform that incorporates the 1D4 tag for anti-GPCR monoclonal antibody screening. Specific isolation of C-terminally tagged CXCR4 enabled not only comprehensive mass spectrometric fingerprinting but also provided an accurate positive

control for monoclonal antibody production and screening. We successfully applied the 1D4 affinity purification approach to develop a target-probe validation platform to screen for a specific broad functioning monoclonal antibody, 1F2 mAb (Chapter III).

5.2.2 Anti-human CXCR4 monoclonal antibody 1F2

We have for the first time developed a specific monoclonal antibody, 1F2 for human CXCR4 mapped to a linear epitope. The 1F2 mAb has broad applications including, immunoprecipitation, immunofluorescence and western blotting. When employed in biomarker profiling for various CXCR4 implicated diseases or conditions, the 1F2 provides a renewable reagent for disease diagnostics. We have shown that the 1F2 mAb can recover endogenous CXCR4 from the Jurkat E6-1 cell line and from primary human PBMCs.

5.2.2.1 1F2 monoclonal antibody future perspectives

We will employ the 1F2 mAb to profile and examine CXCR4 expressed in PBMCs obtained from HIV-1 infected, cancer and warts, hypogammaglobulinemia, immunodeficiency syndrome patients. The ethical approval certificate for the use of human samples for our future work is attached in Appendix E.

5.3 Receptor footprinting

We have applied the expression, purification and mass spectrometric detection strategies that we have optimized to footprint the CXCR4 binding site. SDF-1-CXCR4 intermolecular interactions were examined with synthetic photoaffinity crosslinking analogs of the cognate ligand as a probe (Chapter IV). Through a comparison of the sequence coverage obtained

between the complexed and unbound receptor we have narrowed down the possible CXCR4 extracellular sites that are involved in direct interaction with the SDF-1 N-terminus.

Classical genetic footprinting approaches search for missing nucleotide sequences in a gene inferred from the absence of PCR products to identify gene function by insertional mutagenesis under different selection conditions [2]. Similarly footprinting approaches have been applied to map protein-protein interaction sites. In particular, site-specific BPA crosslinker is incorporated into the corresponding ligand to probe the receptor being studied. Crosslinking has proven to be effective for GPCR binding surface mapping [3]. The contact regions between the ligand and receptor are stabilized by covalent linkage and crosslinked site may be then determined by digestive mapping [4], radioiolabeling [5] and autoradiography [6], liquid chromatography [7], UV detection [7] or mass spectrometry [8]. The sites of contact in BPA crosslinked complexes are generally inferred from mass changes detected by radioisotopic labeling and autoradiography of digested pieces of the crosslinked complex. There have been reports of mass spectrometric detection without radioactive labels in some substance P receptor photocrosslinking studies; however these sites were not sequenced and similarly mass changes were used to infer sites of contact [9].

Photocrosslinking with BPA is an effective method to develop a preliminary footprint of GPCRs because of its specificity, stability in water and ambient light, photokinetics, availability and ease of incorporation into ligands in solid-phase synthesis [3]. We also recognize that there are limitations to BPA photocrosslinking technology such as low resolution and reactive preference for methionine [10]. However, the advantages of photocrosslinking mass spectrometric mapping outweigh the shortcomings. Reactive preference for methionine was not an issue for CXCR4, as methionine was not present in the extracellular regions of the receptor

for non-specific crosslinking. Photocrosslinking enables contact region identification between ligand receptor binding that may be refined with other high resolution techniques. Contact regions on the parathyroid hormone receptor were identified initially by photocrosslinking for example and then followed up with further refinement by NMR [11, 12]. Photocrosslinking is a powerful technology that can be employed as an initial investigation into ligand-receptor interactions to identify key contact regions that are targeted for further examination in more molecular detail. Site-specific photocrosslinking provides the groundwork for rational drug design.

Although identifying or visualizing the exact crosslinked amino acid residues from photolabeled complex would be the most direct way of identifying contact regions but remains a challenge and the strategy for detection depends on the specific question examined. However, even if the crosslinked peptide was detectable, data interpretation would be complicated because of the complex product ions produced. Data interpretation is also limited by the quality, accuracy and completeness of existing protein databases. Two proteomic search engines and protein databases were required to accurately identify as well as confirm receptor sequence coverage of the SDF-1BPA5-CXCR4-1D4 complex.

5.3.2 CXCR4 footprinting future perspectives

We are currently developing more sensitive, specific targeted MS approaches to hone in on the exact region of contact such as MRM to focus in on the peptides that have not been sequenced in the receptor, the receptor N-terminus and ECL1 as described in Chapter IV. Sample heterogeneity will be address by enzymatic removal of the glycosylation on CXCR4 residue N11 after crosslinking to distinguish whether the N-terminus or ECL is involved in

contacting amino acid 5 on SDF-1 BPA5. Glycosidase treatment conditions of Triton X-100 solubilized CXCR4 has been optimized for subsequent complex analysis (data not shown).

Novel strategies for detection of the actual crosslinked site by mass spectrometry to provide direct contact point determination between CXCR4 and SDF-1 has also been examined. A stable octa-deuterated isotopic label was incorporated at position 3 for the SDF-1 BPA analog for subsequent crosslinked site detection. A 1:1 mixture of the heavy and light photoactive analogs with equal binding affinities as determined from binding competition experiments were crosslinked to CXCR4-1D4 for correlational analysis. A theoretical correlation of 8.04Da is used to trace the heavy and light crosslinked species. The identity of the species and corresponding regions on the receptor and ligand will be determined by manual inspection of tandem mass spectra if available. Since stable isotopic forms of BPA were not commercially available and a custom synthesis was not economically feasible, a separate site close to the BPA site was chosen for stable isotopic label incorporation which was at valine-3. Preliminary data for the isotopically labelled crosslinked site for correlation analysis is presented in Appendix D.

The rationale for isotopic label placement close to the BPA site in the ligand so that it is excised together with crosslinked site for mass spectrometric detection. This was confirmed with control digestions of the ligand alone, both the isotopic label and crosslinker were in the same peptide with *in-solution* tryptic as well as sequential chymotrypsin, trypsin digestion. Control digestions for the ligand and receptor alone, without crosslinking were sequenced by tandem mass spectrometry to verify primary sequence as well determine expected peptide mass contributions from each moiety in the final crosslinked complex for the particular digestion condition.

5.4 References

1. Strizki JM, et al., *A monoclonal antibody (12G5) directed against CXCR4 inhibits infection with the dual-tropic human immunodeficiency virus type 1 isolate HIV-189.6 but not the T-tropic isolate HIV-1HxB.* Journal of Virology, 1997. **71**(7): p. 5678-5683.

2. *Smith V, Botstein D, and Brown PO, Genetic footprinting: a genomic strategy for determining a gene's function given it's sequence.* PNAS, 1995. **3**(92): p. 6479-6483.

3. Wittelsberger A, Mierke DF, and Rosenblatt M, *Mapping Ligand-receptor interfaces: approaching the resolution limit of benzophenone-based photoaffinity scanning.* Chemical Biological Drug Design, 2008. **71**: p. 380-383.

4. Wittelsberger A, et al., *The mid-region of parathyroid hormone (1-34) serves as a functional docking domain in the receptor activation.* Biochemistry, 2006. **45**: p. 2027-2034.

5. Perodin, J., et al., *Residues 293 and 294 Are Ligand Contact Points of the Human Angiotensin Type 1 Receptor.* Biochemistry, 2002. **41**(48): p. 14348-14356.

6. Gensure, R.C., et al., *Identification of Determinants of Inverse Agonism in a Constitutively Active Parathyroid Hormone/Parathyroid Hormone-related Peptide Receptor by Photoaffinity Cross-linking and Mutational Analysis.* J. Biol. Chem., 2001. **276**(46): p. 42692-42699.

7. Li Y-M, et al., *Mapping Peptide-binding Domains of the Substance P (NK-1) Receptor from P388D1 Cells with Photolabile Agonists.* Journal of Biological Chemistry, 1995. **270**: p. 1213-1220.

8. Mills, J.S., et al., *Identification of a Ligand Binding Site in the Human Neutrophil Formyl Peptide Receptor Using a Site-specific Fluorescent Photoaffinity Label and Mass Spectrometry.* J. Biol. Chem., 1998. **273**(17): p. 10428-10435.

9. Girault S, Sagan S, Bolbach G, La Veille S, and Chassaing G. *The use of photolabeled peptides to localize the substance –p binding site in the human neurokinin-1 tachykinin receptor.* Eur. J. Biochem., 1996. **240:** p. 215-222.

10. Wittelsberger A, et al., *Methionine acts as a "magnet" in photoaffinity crosslinking experiments.* FEBS Lett, 2006. **580**: p. 1872-1876.

11. Pellegrini M, et al., *Addressing the tertiary structure of human parathyroid hormone-(1-34).* Journal of Biological Chemistry, 1998. **273**: p. 10420-10427.

12. Pellegrini M, et al., *Binding domain of human parathyroid hormone receptor: from conformation to function.* Biochemistry, 1998. **37**: p. 12737-12743.

Appendix A: Introduction, chemokine nomenclature and synthetic SDF-1 variants

Table A-1: Adapted from Zlotnik and Yosie, Immunity 2000, R&D systems Chemokine Nomenclature and the International Union of Immunological Societies/World Health Organization Subcommittee on Chemokine Nomenclature (Journal of Leukocyte Biology, 70:465-466). CXCR7, Balabanian et al. 2005.

Human aliases	Systematic name	small secreted cytokine (SCY) gene designation	Receptors
Lptn, SCM-1, ATAC	XCL1	SCYC1/2	XCR1
SCM-1β	XCL2		XCR1
Fractalkine, ABCD-3	CX_3CL1		CX_3CR1
I-309	CCL1	SCYA1	CCR8
MCP-1, MCAF, LDFG, GDCF, TDCF, SMC-CF, HC11, TSG8	CCL2	SCYA2	CCR2
MIP-1α, LD78α, LD78β, GOS19, Pat464	CCL3	SCYA3	CCR1, CCR5
MIP-1β, PAT744, ACT-2, G-26, HC21, H400, MAD-5, LAG-1	CCL4	SCYA4	CCR5
RANTES	CCL5	SCYA5	CCR1, CCR3, CCR5
?	CCL6	SCYA6	?
MCP-3, NC28, FIC, MARC	CCL7	SCYA7	CCR1, CCR2, CCR3
MCP-2, HC-14	CCL8	SCYA8	CCR3
?	CCL9/10	SCYA9	?
Eotaxin	CCL11	SCYA11	CCR3
?	CCL12	SCYA12	CCR2
MCP-4, Ckβ10, NCC-1	CCL13	SCYA13	CCR2, CCR3
HCC-1, MCIF, Ckβ1, NCC-2, HCC-3	CCL14	SCYA14	CCR1
MIP-1δ, Lkn-1, CC-2, MIP-5, HCC-2, CCF-18, NCC-3	CCL15	SCYA15	CCR1, CCR3
HCC-4, LEC, ILINK, NCC-4, LEC, LMC, Ckβ12	CCL16	SCYA16	CCR1
TARC, Dendrokine, ABCD-2	CCL17	SCYA17	CCR4, CCR8
PARC, DC-CK1, AMAC-1, MIP-4, Dctactin	CCL18	SCYA18	?
MIP-3β, ELC, Exodus-3, Ckβ11	CCL19	SCYA19	CCR7
MIP-3α, LARC, Exodus-1, Mexikine, ST38, Ckβ4	CCL20	SCYA20	CCR6
6-Ckine, Exodus-2, SLC, TCA4, Ckβ9	CCL21	SCYA21	CCR7
MDC, STCP-1, DCtactin β, ABCD-1, DC/BCK	CCL22	SCYA22	CCR4

Human aliases	Systematic name	small secreted cytokine (SCY) gene designation	Receptors
MPIF-1, Ckβ8-1	CCL23	SCYA23	CCR1
Eotaxin-2, MPIF-2, Ckβ6	CCL24	SCYA24	CCR3
TECK	CCL25	SCYA25	CCR9
Eotaxin-3	CCL26	SCYA26	CCR3
CTACK	CCL27	SCYA27	CCR10
	CCL28	SCYA28	CCR10?
GROα, MGSA–α, GRO–1, NAP–3	CXCL1	SCYB1	CXCR2>CXCR1
GRO-β, MGSA-β, MIP-2α, GRO-2	CXCL2	SCYB2	CXCR2>CXCR1
MGSA-β, MIP-2β, GRO-3	CXCL3	SCYB3	CXCR2>CXCR1
GRO-γ, MGSA-γ, MIP-2β, GRO-3	CXCL3	SCYB3	?
PFA	CXCL4	SCYB4	CXCR2, CXCR1
ENA-78, AMCF-11	CXCL5	SCYB5	CXCR2, CXCR1
GCP-2, Ckα3	CXCL6	SCYB6	CXCR1, CXCR2
NAP-2, MDGF	CXCL7	SCYB7	CXCR2
IL-8, NCF, NAP-1, MDNCF, LUCT, AMCF-1, MONAP	CXCL8	SCYB8	CXCR1,CXCR2
MIG	CXCL9	SCYB9	CXCR3
IP-10	CXCL10	SCYB10	CXCR3
I-TAC, b-R1, H174, IP-9	CXCL11	SCYB11	CXCR3, CXCR7
SDF-1α/β, PBSF, hiRH, TLSR0α/β, TPAR1	CXCL12	SCYB12	CXCR4, CXCR7
BLC/BCA-1, CXC-X, BLR1L, Angie	CXCL13	SCYB13	CXCR5
BRAK, CXC-X3, Bolekine, NJAC	CXCL14	SCYB14	?
CINC-2β-like	CXCL15	SCYB15	
	CXCL16	SYB16	CXCR6

Table A-2: Synthetic SDF-1 analogs, adapted from Crump M et al, EMBO, 1997.**16**(23):p.6996-7007; Loetscher P et al, JBC, 1998. **273**(35): 22279-83; and Wong et al. (unpublished)

Chemokine or variant	K_d fold increase relative to SDF-1	Sequence
SDF-1α	1	1 67 KPVSLSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
SDF, 1-8	Na	KPVSLSYR
SDF, 1-9	1,500	KPVSLSYRC
SDF, 1-9 [aba-9]	Na	KPVSLSYRAba
SDF, 1-9, dimer	82	KPVSLSYRC KPVSLSYRC
SDF, 1-9 [P2G] dimer	290	KGVSLSYRC KGVSLSYRC
SDF, 1-17	94	KPVSLSYRCPCRFFESH
		N-terminal truncation variants
SDF1 2-67	6	2 67 PVSLSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
SDF1 3-67	13	VSLSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
SDF1 4-67	94	SLSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
SDF1 5-67	110	LSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
SDF1 6-67	110	SYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
SDF1 7-67	130	YRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
SDF1 8-67	140	RCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
SDF19-67	Na	CPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
		Hybrids
IP10H1	250	KPVSLSYRCPCISISNQPVNPRSLEKLEIIPASQFCPRVEIIATMKKKGEKACLNPESKAIKNLLKAVSKEMSKRSP
IP10H2	16	KPVSLSYRCPCRFFESHVNPRSLEKLEIIPASQFCPRVEIIATMKKKGEKACLNPESKAIKNLLKAVSKEMSKRSP
IP10	Na	VPLSRTVRCTCISISNQPVNPRSLEKLEIIPASQFCPRVEIIATMKKKGEKACLNPESKAIKNLLKAVSKEMSKRSP
GROH1	14	KPVSLSYRCPCLQTLQGIHPKNIQHLKILNTPNCAQTEVIATLKNGRKACLNPASPIVKKIIEKMLNSDKSN
GROH2	3	KPVSLSYRCPCRFFESHIHPKNIQHLKILNTPNCAQTEVIATLKNGRKACLNPASPIVKKIIEKMLNSDKSN
GRO	Na	ASVATELRCQCLQTLQGIHPKNIQSVNVKSPGPHCAQTEVIATLKNGRKACLNPASPIVKKIIEKMLNSDKSN
IL-8H1	Na	KPVSLSYRCPCIKTYSKPFHPKFIKELRVIETPNCANTEIIVKLSDGRELCLDPKENWVQRVVEKFLKRAENS
IL-8H2	Na	KPVSLSYRCPCRFFESHFHPKFIKELRVIETPNCANTEIIVKLSDGRELCLDPKENWVQRVVEKFLKRAENS
IL-8	Na	SAKELRCQCIKTYSKPFHPKFIKELRVIESGPHCANTEIIVKLSDGRELCLDPKENWVQRVVEKFLKRAENS
		Single addition or substitution variants
SDF-Gly	<1	GKPVSLSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
K1R	4	RPVSLSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
K1Orn	2	OrnPVSLSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
P2G	3	KGVSLSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
V3I	4	KPISLSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
AQA	3	KPVAQALSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
Y7A	4	KPVSLSARCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
Y7H	<1	KPVSLSHRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
R8K	5	KPVSLSYKCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKALN
		Novel crosslinking analogs
SDF1Bpa3	3	KPBSLSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKAK_bN
SDF1Bpa5	1	KPVSBSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKAK_bN
SDF1Bpa7	1	KPVSLSBRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKAK_bN
SDF1Bpa 13	3	KPVSLSYRCPCRBFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKAK_bN
SDF1Bpa5, $V3_{ps}$	1	KPV_{DX}SBSYRCPCRFFESHVARANVKHLKILNTPNCALQIVARLKNNNRQVCIDPKLKWIQEYLEKAK_bN

Appendix B: Expression systems

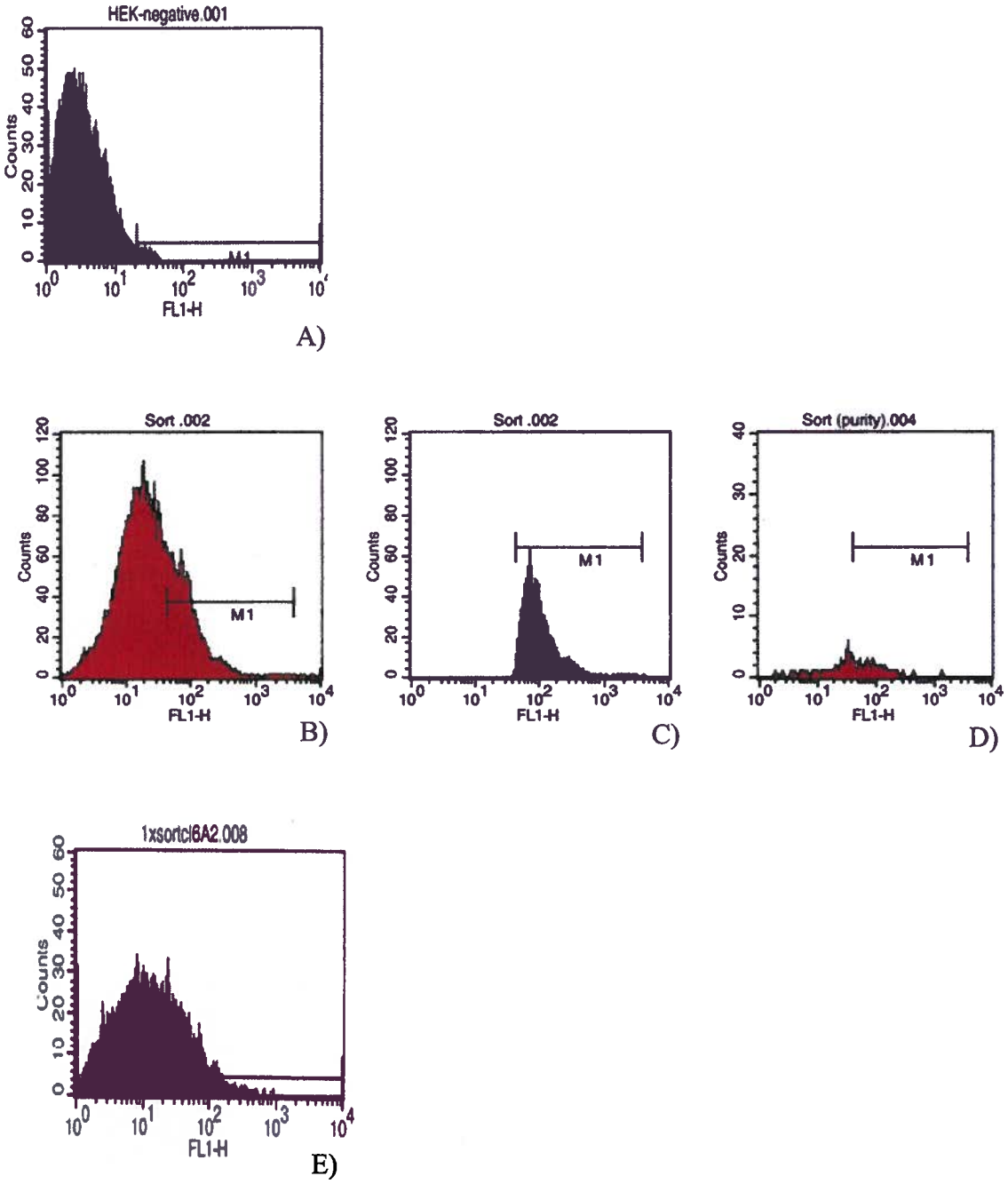


Figure B-1: Fluorescence activated cell sorting for high CXCR4 expressing stably transfected Human Embryonic Kidney (HEK) cell line. The cells were stained with neutralizing antibody 12G5 and anti-mouse conjugated fluorescein isothiocyanate (FITC). Flow cytometric analysis of the A) vector control, B) stable CXCR4 transfection before sorting, C) from sorting, D) after sorting, and a week of culturing E) show the level of receptor expressed.

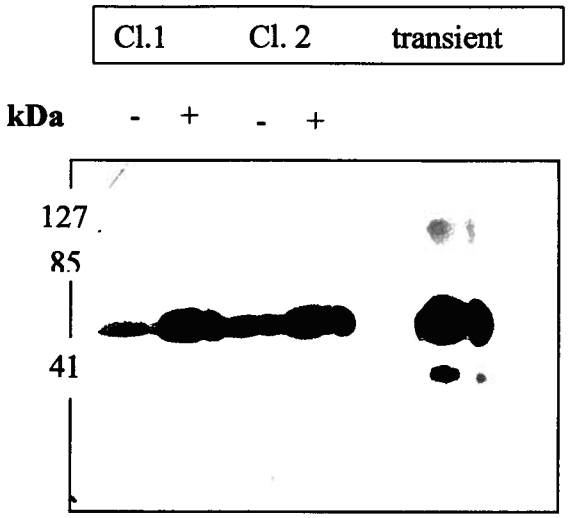


Figure B-2: Tetracycline inducible HEK cell line for CXCR4 expression. Clones with low background and high inducible expression were screened and expression levels were compared against a transient expression. The western blot was probed with Rho1D4 mAb for the C-terminal tag, an anti-mouse horse radish peroxidase secondary antibody and enhanced chemiluminescence detection.

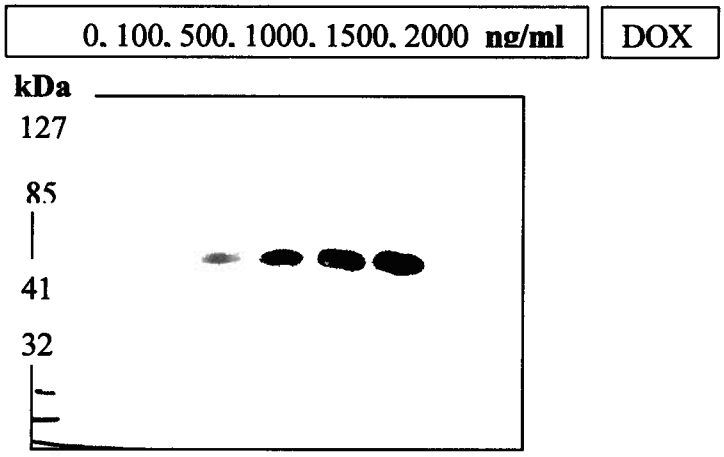


Figure B-3: Titrating tetracycline concentrations on selected clone for maximal CXCR4 induced expression. The western blot was probed with Rho1D4 mAb for the C-terminal tag, an anti-mouse horse radish peroxidase secondary antibody and enhanced chemiluminescence detection.

Appendix C: Two more active analogs, SDF-1 BPA 7 and SDF-1BPA13 were synthesized and characterized

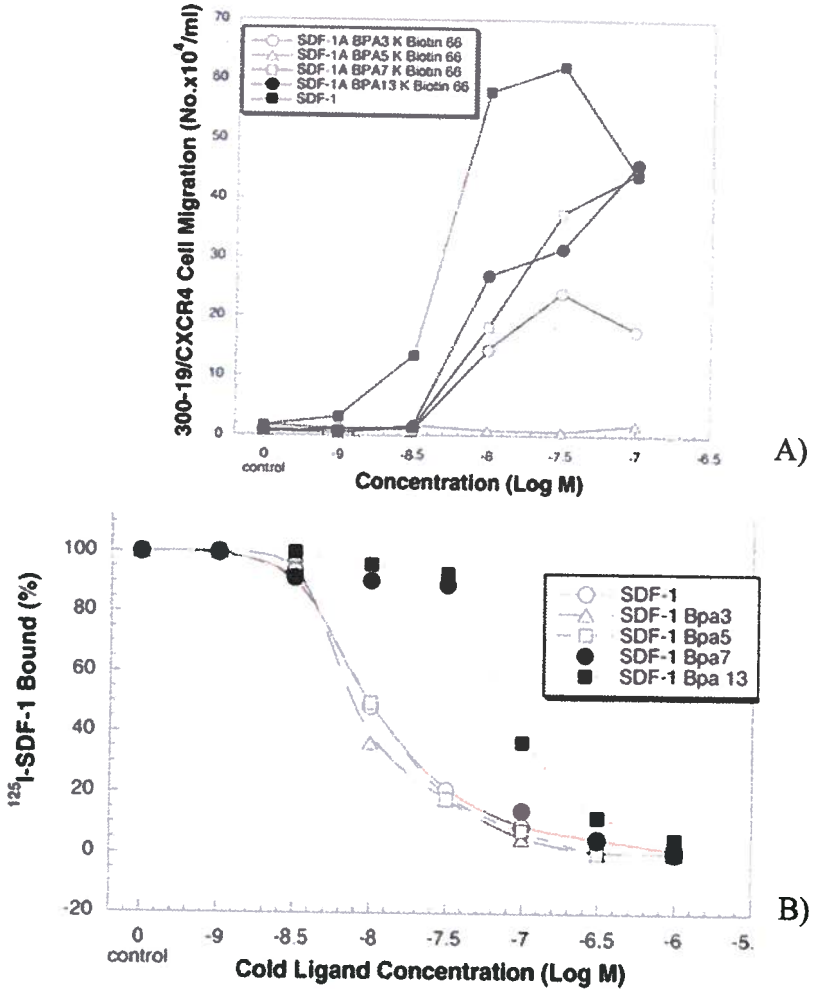


Figure C-1: Two additional analogs SDF-1 BPA 7 and SDF-1 BPA 13 were synthesized, characterized for A) activity and B) binding. For comparison of all synthesized ligands, SDF-1 BPA 3 and SDF-1 BPA5 (the analogs that were described in the dissertation) were also included in the above graphs along with the SDF-1 positive control.

Appendix D: Stable isotopic labels for locating crosslinking site. A 1:1 mixture of an octadeuterated heavy SDF-1 BPA 5 and SDF-1 BPA 5 for photocrosslinking and crosslink site tracing by computational correlation analysis.

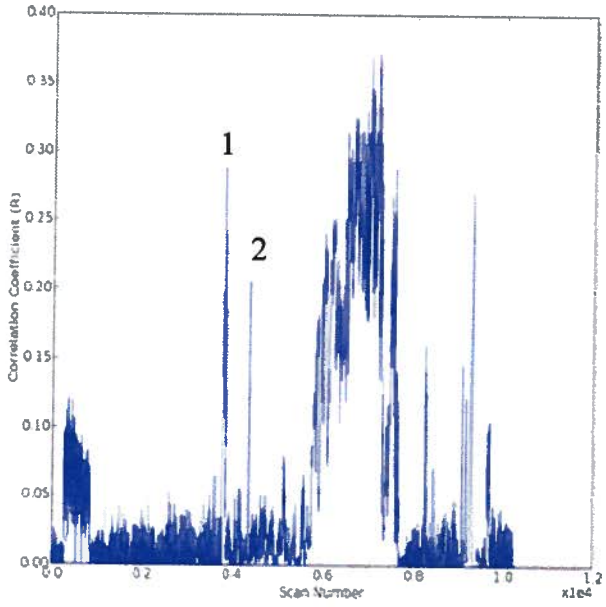


Figure D-1: Delta 8 Th correlation analysis on the precursor ion spectra from the FT-ICR mass spectrometer was converted to text format using an Xcalibur script and pertinent information was parsed from the file with in-house software for correlation analysis. Positive correlation peaks selected for further investigation are numbered.

Peak 1 is a delta 16Da difference for a doubly charged species and thus is a false positive. Peak 2 represents an 8 Th correlation for a singly charged species. The extracted ion chromatogram for the corresponding scan number for the run is shown in Figure D1. The broad peak spanning from approximately scan number 0.58 to 0.79 represents random correlations at the end of the chromatography run when the gradient reaches high organic solvent concentration.

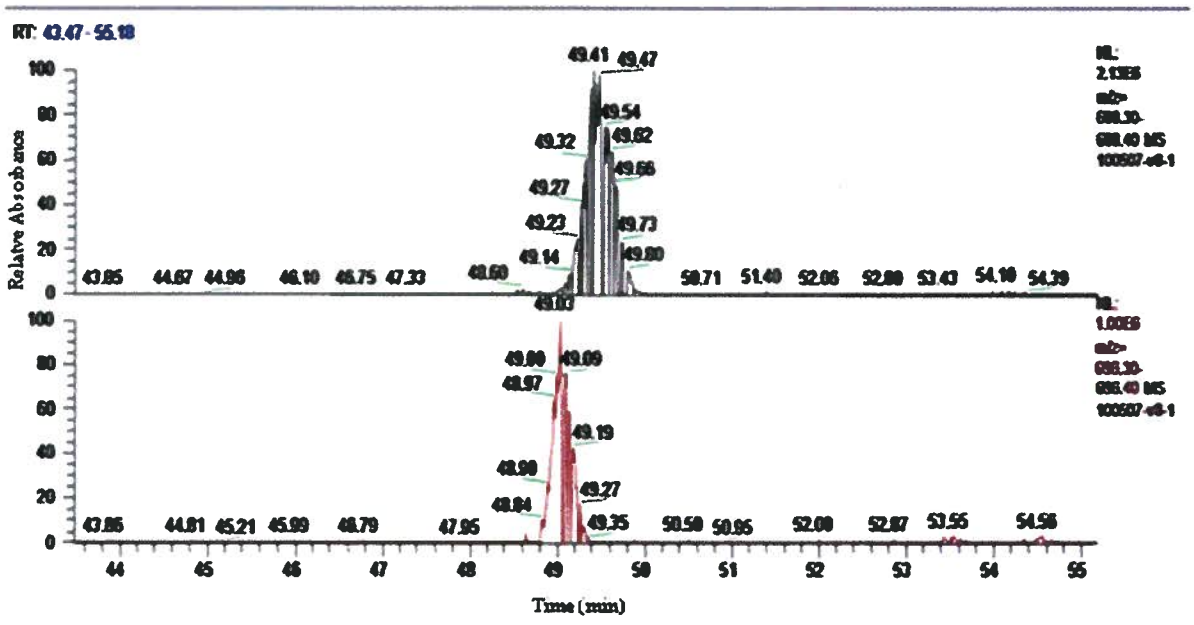


Figure D-2: Extracted ion chromatogram for the A) 688.39 and B) 696.35 Da singly charged species identified in the peak 2 of the 8 Th correlation analysis.

The heavier species has a shorter retention time than the lighter species as would be expected for a deuterated species. Both species elute at relatively similar times and display similar peak shapes. Closer examination of these peaks, we found that the correlation is only a 7.96 Th difference and not the expected 8.04 Th for an octadeuterated species.

Appendix E: Ethical approval for the use of human samples

The University of British Columbia
Office of Research Services
Clinical Research Ethics Board – Room 210, 828 West 10th Avenue, Vancouver, BC V5Z 1L8

ETHICS CERTIFICATE OF EXPEDITED APPROVAL

PRINCIPAL INVESTIGATOR:	INSTITUTION / DEPARTMENT:	UBC CREB NUMBER:
Anne K. Junker	UBC/Medicine, Faculty of/Paediatrics	H07-01837

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Vancouver (excludes UBC Hospital)
Children's and Women's Health Centre of BC (incl. Sunny Hill)	Children's and Women's Health Centre of BC (incl. Sunny Hill)
Other locations where the research will be conducted:	
N/A	

CO-INVESTIGATOR(S):

David R. Burdge
Evelyn Maan
Juergen Kast
Julie Patricia Wong
John C. Forbes
Ariane Alimenti
Robert S. Molday

SPONSORING AGENCIES:

Canadian Institutes of Health Research (CIHR)
Michael Smith Foundation for Health Research

PROJECT TITLE:

CXCR4 Expression in Human Leukocyte Populations

THE CURRENT UBC CREB APPROVAL FOR THIS STUDY EXPIRES: February 11, 2009
The UBC Clinical Research Ethics Board Chair or Associate Chair, has reviewed the above described research project, including associated documentation noted below, and finds the research project acceptable on ethical grounds for research involving human subjects and hereby grants approval.

DOCUMENTS INCLUDED IN THIS APPROVAL:			**APPROVAL DATE:**
Document Name	**Version**	**Date**	**February 11, 2008**
Protocol:			
Protocol	1.2	December 12, 2007	
Consent Forms:			
Study Consent	Version 1	November 13, 2006	
Advertisements:			
Recruitment Notice	1	November 13, 2007	

CERTIFICATION:
In respect of clinical trials:
1. The membership of this Research Ethics Board complies with the membership requirements for Research Ethics Boards defined in Division 5 of the Food and Drug Regulations.
2. The Research Ethics Board carries out its functions in a manner consistent with Good Clinical Practices.
3. This Research Ethics Board has reviewed and approved the clinical trial protocol and informed consent form for the trial which is to be conducted by the qualified investigator named above at the specified clinical trial site. This approval and the views of this Research Ethics Board have been documented in writing.

The documentation included for the above-named project has been reviewed by the UBC CREB, and the research study, as presented in the documentation, was found to be acceptable on ethical grounds for research involving human subjects and was approved by the UBC CREB.

Approval of the Clinical Research Ethics Board by one of:

Dr. Caron Strahlendorf, Associate Chair