DISCOVERY OF CONSERVED MOTIFS IN MARCO

DISCOVERY OF CONSERVED MOTIFS IN MARCO THROUGH EVOLUTIONARY ANALYSES AND MOLECULAR BIOLOGY

BY

FIONA J. WHELAN, B.CompSc.

A THESIS

SUBMITTED TO THE DEPARTMENT OF MEDICAL SCIENCES

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Master of Science (2012) McMaster University

(Medical Sciences) Hamilton, Ontario, Canada

TITLE:	Discovery of conserved motifs in MARCO through evolutionary analyses and molecular biology
AUTHOR:	Fiona J. Whelan B.CompSc., (Bioinformatics) University of Waterloo, Waterloo, Canada
SUPERVISOR:	Dr. Dawn Bowdish
NUMBER OF PAGES:	xxv, 115

I dedicate this work to my parents, Bill Nye, and Jurassic Park, for if it were not for these 3 things, I would have no love of science.

Abstract

Of the pattern recognition receptors involved in innate immunity, the class A Scavenger receptors are involved in the recognition and clearance of bacteria, yeast, and senescent molecules. Previous research has implicated the intracellular region of these receptors as essential for the clearance of these substances via endocytosis. In this work, I used a bioinformatic approach to define the evolutionary history of the class A Scavenger receptor family while elucidating areas of conservation within the cytoplasmic domains of these proteins. With this information, in addition to further predictions of post translational modifications and potential docking motifs for interacting proteins, I conducted molecular biology experiments to study the *in vitro* functionality of the macrophage receptor with collagenous domain (MARCO), a member of the class A Scavenger receptors.

Evolutionary analyses of the 5 class A Scavenger receptors identified a shared ancestry between these proteins and allowed me to postulate that 4 distinct gene duplication events in addition to subsequent domain fusions, internal repeats, and deletions are responsible for the diverse protein structures and functions of this family. Despite some variation in domain structure, I found highly conserved regions across all 5 members, including a negatively charged region in the cytoplasmic domain. Further analyses of MARCO across organisms identified other conserved regions, including 2 residues predicted to be ubiquitinated, sumoylated, or phosphorylated by *in silico* predictive methods. However, molecular biology experiments demonstrated that these post translational modifications to not occur in the steady state. Additional *in vitro* experiments, including isolations of MARCO and an artifical construct containing only the intracellular regions of the protein, were unable to identify any candidate adaptor binding proteins. Further research is needed to determine whether modifications in this region occur in the presence of bound ligands and/or known co-receptors.

Acknowledgements

Even though the time I've spent in the Bowdish laboratory has been short, it has been 2 of my most rewarding years. This document is the result of the guidance, patience, and support of those who have surrounded me as I pursued this research. For this, I am humbly indebted.

Above all others, my supervisor, Dr. Dawn Bowdish, has treated me with her infinite guidance, encouragement, and patience. The care she takes with her students and peers, her ability to keep at the forefront of new research, and her dedication to promoting science to the general public, all while putting her family first, has dubbed her 'superwoman' among the female trainees. To me, she was, and will continue to be, an amazing role model.

The Bowdish laboratory has been an excellent environment in which to begin my scientific career. Each member showed incredible patience as I learnt molecular laboratory techniques at my own pace, and showed interest in the bioinformatic research I conducted. Thank you for the opportunity to work along aside you as friends.

I wish to thank all those who took on the role as my friends and mentors, specifically, Dr. Brendan McConkey, Dr. Brian Lichty, Dr. Brian Golding, Dr. Mark McDermott, Dr. Peter Pelka, Dr. Mathieu Morissette, Sarah Chauvin, and Mike Dorrington. Being able to comfortably approach each of you with questions, concerns, or expressions of frustration has been invaluable to my success.

I would like to acknowledge all others part of MIRC, the IIDR, and McMaster University as a whole. I feel exceptionally lucky to have developed such excellent friendships among my colleagues.

Lastly, thank you to friends and family for their support and understanding through my unpredictable research schedule.

Abbreviations and Symbols

acLDL	acetylated low-density lipoprotein
ATP	adenosine triphosphate
BLAST	basic local alignment search tool
BLAT	basic local alignment search tool-like alignment tool
BSA	bovine serum albumin
cA-SR	class A Scavenger receptor
CDD	Conserved Domain Database
CSR	cellular stress response
DMEM	Dulbecco's modified Eagle medium
DTT	dithiothreitol
EBI	European Bioinformatics Institute
EDTA	ethylenediaminetracetic acid
ELISA	enzyme-linked immunosorbent assay
ELM	Eukaryotic Linear Motif
FBS	fetal bovine serum
FPLC	fast protein liquid chromatography
GTR	Generalized Time-Reversible model
hMARCO	*Homo sapiens* MARCO
HRP	horse-radish peroxidise
IP	immunoprecipitation
LPS	lipopolysaccharide
malBSA	maleylated bovine serum albumin
MAP	mitogen-activating protein
MARCO	macrophage receptor with collagenous domain
MBL	mannose-binding lectin
MUSCLE	multiple sequence comparison by log-expectation
NCBI	National Center for Biotechnology Information
NF-κB	nuclear factor kappa B
oxLDL	oxidized low-density lipoprotein
PAMP	pathogen-associated molecular pattern

PBS	phosphate buffered saline
PCA	principal component analysis
PCR	polymerase chain reaction
PFA	paraformaldehyde
PPR	pattern recognition receptor
PSI-BLAST	position-specific iterated basic local alignment search tool
PTM	post-translational modification
PVDF	polyvinylidene difluoride
SCARA3	Scavenger receptor class A, member 3
SCARA4	Scavenger receptor class A, member 4
SCARA5	Scavenger receptor class A, member 5
SDS	sodium dodecyl sulfate
SR	Scavenger receptor
SRA	Scavenger receptor class A
SRCL	Scavenger receptor with C-type lectin domain
SRCR	Scavenger receptor cysteine rich domain
SUMO	small ubiquitin-related modifiers
TBST	tris(hydroxymethyl)aminomethane buffered saline with 0.1% Tween 20
TLR	Toll-like receptor
TMB	tetramethylbenzidine
TNF	tumor necrosis factor
TRAF6	tumor necrosis factor receptor-associated factor 6
Tris	tris(hydroxymethyl)aminomethane

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Pattern recognition receptors (PRRs)

It can be argued that one of the fundamental requirements for an organism's survival is a mechanism through which it is protected from infectious agents. As such, the development of immunity is likely universal across all kingdoms of life. However, the level of complexity of this system and, in general, how an organism defends itself from a pathogen differs depending on the type of animal. For example, only vertebrates are equipped with both an innate and adaptive immune response, whereas all other organisms rely solely on their innate defences (Schulenburg *et al.*, 2004). In order to mount this initial response, the innate immune system must be able to distinguish self from non-self using only a finite number of receptors present on the surface of immune cells such as macrophages and neutrophils. To accomplish this, these receptors, termed pattern recognition receptors (PRRs), have evolved to recognize highly conserved molecular structures present on the surface of pathogens that

are integral to their function, and thus not likely to be readily mutated. These motifs have been subsequently termed pathogen-associated molecular patterns (PAMPs) and include structures such as the lipopolysaccharides (LPS) and peptidoglycans that form bacterial cell walls and components of yeast cells such as mannans (Aderem and Ulevitch, 2000; Murphy *et al.*, 2011).

PRRs can be subclassified into groups based on their specific innate immune functions. For example, upon the recognition of specific PAMPs, a subset of PRRs will invoke a pro-inflammatory response that encourages additional neutrophils and macrophages to migrate to a site of infection (Murphy *et al.*, 2011). Further, other receptors are responsible for marking pathogenic molecules for destruction by other elements of the innate immune response, or readily phagocytose pathogens to rid them from circulation.

1.1.1 Proinflammatory receptors

In multicellular organisms, inflammation is essential for combating infection. Upon recognition of a pathogen, proinflammatory receptors signal the production of cytokines and chemokines. These soluble mediators affect resident cells, causing them to increase their own signalling and phagocytic capabilities in addition to encouraging the recruitment of effector cells to the site of infection (Murphy *et al.*, 2011). Perhaps the most well studied family of proinflammatory receptors are the Toll-like receptors (TLRs), so named due to their homology to the Toll protein originally discovered in *Drosophila melanogaster* (Anderson *et al.*, 1985). In vertebrate species, there have been at least 10 identified TLR proteins (Akira *et al.*, 2001), which fall into 6 distinct clades (Roach *et al.*, 2005). Each TLR recognizes a unique set of PAMPs;

for example, TLR4, the first identified human homolog (Aderem and Ulevitch, 2000), binds LPS of bacteria, TLR5 recognizes bacterial flagellin (Akira *et al.*, 2001), and TLR3 binds viral components (DeWitte-Orr *et al.*, 2010). Upon recognition of these molecules, TLRs induce a signalling cascade responsible for the activation of nuclear factor kappa B (NF-κB), a nuclear transcription factor which in turn incites pro-inflammatory responses and induces inflammation (Murphy *et al.*, 2011).

1.1.2 Neutralization of pathogens via PRRs

Alternatively, pathogens can be 'marked' for destruction by the lectin complement pathway of the innate immune response (Murphy *et al.*, 2011). Similar to the classically activated complement pathway, the lectin pathway is responsible for the opsonization (coating) of pathogens with protein which subsequently induces their destruction by innate immune cells (Murphy *et al.*, 2011). Certain PRRs are responsible for activating the lectin complement pathway by recognizing conserved patterns on the surface of pathogens. The mannose-binding lectins (MBLs), for example, recognize mannose sugars on bacteria and other pathogens (Murphy *et al.*, 2011). Upon recognition, MBLs form a protein complex which induces the lectin complement pathway cascade, resulting in the opsonization of the pathogen (Murphy *et al.*, 2011).

1.1.3 Phagocytic receptors

Phagocytosis is the process used by both unicellular and eukaryotic cells to ingest particles (Allen and Aderem, 1996). This process has evolved from a simple feeding behaviour in protists to a method of engulfing pathogenic particles, aged

cells, and cellular debris in metazoan hosts (Alberts, 2008). In innate immunity, some cell types are more effective at phagocytosing particles than others; these so-called 'professional phagocytes' express high levels of phagocytic receptors on their cell surface, which advance the rate and sensitivity of phagocytosis. While there are many classes of phagocytic receptors, the laboratory of Dr. Dawn Bowdish is specifically interested in a subset of these, the Scavenger receptors (SRs), which reside on many cell types but most notably on the surface of macrophages.

1.1.4 The evolution of pattern recognition

The fact that all animals rely on innate defences suggests that the roots of innate host defence can be traced back to a common ancestor shared by both vertebrate and invertebrate species. It is interesting to postulate exactly where along this lineage the various components of innate immunity began. For example, many of the mammalian PRRs are part of larger, evolutionarily ancient recognition systems: the mammalian Toll-like receptors are homologous to the Toll proteins in *Drosophila melanogaster* (Murphy *et al.*, 2011) and orthologs of vertebrate MBLs have been identified in invertebrates (Fujita, 2002). Also, it has long been speculated that the classical complement pathway emerged from the innate lectin pathway, a hypothesis which recent research has supported (Matsushita *et al.*, 2004).

While the evolution of many families of PRRs, such as the TLRs and MBLs, have been extensively studied, the evolutionary history of the SRs has not yet been explored.

1.2 The Scavenger receptor family

The SRs, a protein family within the PRRs, are a structurally diverse group of receptors responsible for the recognition of foreign materials as well as modified host proteins (Bowdish and Gordon, 2009). SRs were originally identified by Goldstein *et al.* in 1979 based on their ability to bind and subsequently internalize acetylated low-density lipoprotein (acLDL) (Goldstein *et al.*, 1979). These receptors bind some (but not all) polyanions including oxidized low-density lipoprotein (oxLDL) (Brown *et al.*, 1980), maleylated bovine serum albumin (malBSA) (Goldstein *et al.*, 1979) and a variety of bacterial ligands (e.g. *Escherichia coli*, *Staphylococcus aureus* (Arredouani *et al.*, 2005), and *Mycobacterium tuberculosis* (Bowdish *et al.*, 2009)). Specifically, SRs have the ability to recognize ligands on both modified host proteins and apoptotic cells (Bowdish and Gordon, 2009).

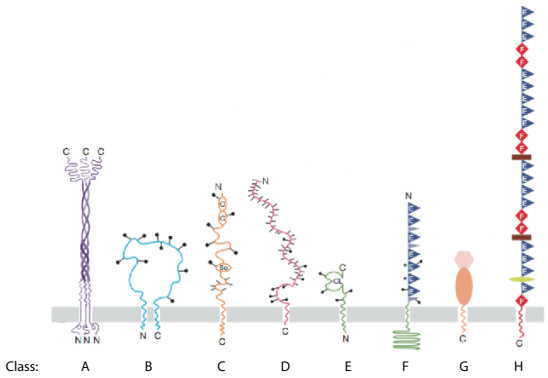


Figure 1.1: **Domain architecture of the 8 classes of Scavenger receptors.** Adapted from Pluddemann *et al.* (Plüddemann *et al.*, 2006). Because of the heterogeneity of the Scavenger receptors, the group was split into 8 classes, A through H, based on similarities in function and protein structure.

Since their initial discovery in 1979, a variety of proteins have been included in the SR family based on their ligand binding capabilities and/or similarities in their secondary structures, resulting in a diverse family of seemingly unrelated proteins (Krieger, 1997; Plüddemann *et al.*, 2007). Consequently, in 1997 Krieger suggested that the SRs be assigned to 8 distinct classes, termed A through H, on the basis of protein sequence comparisons and domain architecture (Krieger, 1997). Structural

representations of a typical member of each of these classes are displayed in *Figure 1.1.*

1.3 The class A Scavenger receptors

The class A Scavenger receptors (cA-SRs) are type II glycoproteins consisting of an intracellular N-terminal domain and an extracellular C-terminus (Murphy *et al.*, 2011). The cA-SRs consist of 2 original members, namely Scavenger receptor class A (SRAI) and macrophage receptor with collagenous domain (MARCO) (Krieger, 1997). Three additional members have been subsequently added to the cA-SRs: Scavenger receptor class A, member 3 (SCARA3)/CSR (cellular stress response) (Han *et al.*, 1998), SCARA4/SRCL (Scavenger receptor with C-type lectin domain) (Nakamura *et al.*, 2001), and SCARA5 (Jiang *et al.*, 2006).

All 5 members form homotrimers that are thought to be stabilized via α-helical coiled coil motifs in addition to their collagenous regions (*Fig. 1.1*) (Krieger, 1992; Doi *et al.*, 1993; Pearson, 1996). In general, the cA-SRs have similar domain structures with some obvious exceptions. For example, these proteins vary considerably in the length of their collagenous domains, ranging from approximately 75 residues in SCARA5 to 250 amino acids in MARCO (Jiang *et al.*, 2006; Elomaa *et al.*, 1995). Importantly, the members of this group also vary in which domain constitutes their C-terminus. SRAI, MARCO, and SCARA5 possess a terminating scavenger receptor cysteine rich (SRCR) domain (Bowdish and Gordon, 2009), whereas SCARA3 terminates at the collagenous domain (Han *et al.*, 1998) and SCARA4 possesses a C-type lectin domain (Nakamura *et al.*, 2001).

1.3.1 Scavenger receptor cysteine rich domain

Alongside the C-type lectin domain of the collectins (Holmskov *et al.*, 1994) and the leucine-rich repeat possessed by the TLRs (Roach *et al.*, 2005), the SRCR domain is one of the most ancient pattern recognition domains associated with innate immunity, having been found in organisms as diverse as humans to sea urchins (Gordon, 2002). This domain possesses 6 highly conserved cysteine residues that form a distinguishable pattern of disulfide bonding (*Fig. 1.2*) (Martínez *et al.*, 2011).

Other proteins possessing this domain have been implicated in a wide variety of functions, including pathogen recognition, endocytosis, and immune response homeostasis (reviewed in Martinez *et al.*, 2011); however, the role of the SRCR domain in the cA-SRs remains unclear. Studies of MARCO and SRAI implicate a region at the junction of the collagenous and SRCR domains as a potential ligand binding motif (Brännström *et al.*, 2002; Goh *et al.*, 2010). Both SRAI and its isoform, SRAII, which lacks the SRCR domain, bind acLDL with equal affinity, indicating that the ligand binding domain, at least for this ligand, is within the collagenous region (Doi *et al.*, 1993). Whether this discrepancy is due to the particular ligands examined and/or multiple binding motifs is unknown.

Figure 1.2: **The protein structure of the SRCR domain.** The domain was crystalized from the *Homo sapiens* Mac-2 binding protein and deposited in the Protein Data Bank (Identifier 1BY2). Cysteine residues involved in disulfide bonding are highlighted in yellow.

1.3.2 Function and expression profile

SRAI is primarily implicated in homeostatic functions such as the uptake of modified lipids and proteins, in addition to having a role in pathogen clearance (Krieger, 1992; Pearson, 1996). In contrast, MARCO has been primarily implicated in host defence via the direct recognition and subsequent endocytosis of pathogens, and the modulation of cytokine production (Kraal *et al.*, 2000; Arredouani *et al.*, 2004). Both SCARA4 and SCARA5 have been documented *in vitro* to bind bacteria, although this ability has not been established *in vivo* (Nakamura *et al.*, 2001; Jiang *et al.*, 2006). In constrast, SCARA3 has been associated with the protection of cells from reactive oxygen species during oxidative stress (Han *et al.*, 1998).

While SRAI and MARCO are primarily expressed on macrophages (Elomaa *et al.*, 1995; Kodama *et al.*, 1990), SCARA3, SCARA4, and SCARA5 are expressed on a variety of other cell types, including epithelial cells (Jiang *et al.*, 2006), and cells of the placenta, lungs, heart, and small intestine (Nakamura *et al.*, 2001). This combination of diverse patterns of expression and function raise questions concerning whether these proteins are related to one another and, if so, what evolutionary relationships exist between them.

1.4 Macrophage receptor with collagenous domain

MARCO, a cA-SR, was originally identified on specific populations of macrophages in *Mus musculus* for its ability to bind bacteria and acLDL, but not yeast (*Fig. 1.3*) (Elomaa *et al.*, 1995). Since its discovery, MARCO has been shown to be important

in the removal of inhaled particles and pathogens from the lungs due in part to its increased expression on alveolar macrophages (Arredouani *et al.*, 2004, 2005). Whereas the role of the SRCR domain across the cA-SRs has been debated, this domain has been shown to be critical for bacterial binding by MARCO (Brännström *et al.*, 2002). Specifically, mutagenesis studies identified the RGRAEVYY motif as essential for high-affinity bacterial binding (Brännström *et al.*, 2002).

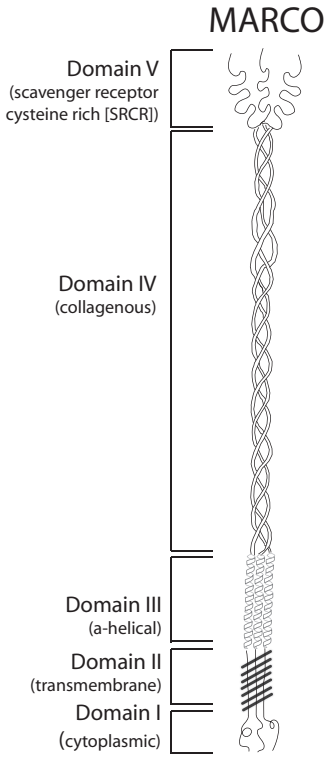


Figure 1.3: **A representation of the domain architecture of MARCO.** MARCO, a class A Scavenger receptor, consists of 5 domains spanning approximately 520 amino acids.

In contrast to what is known regarding the functional role of the C-terminus, the function of the cytoplasmic region of MARCO has not been well described. Much of what is known stems from the research completed by a prior student of Dr. Dawn Bowdish, Mariliis Kroos. Kroos studied MARCO-mediated binding and uptake in addition to the involvement of MARCO in cell surface adhesion and motility (Kroos, 2008). Kroos constructed a mutant of MARCO lacking the cytoplasmic domain, and used this mutant to determine that the intracellular region was essential for endocytosis and adhesion, and additionally affected the cell surface expression of the protein (Kroos, 2008). These

findings indicate that there may be specific functional regions in the cytoplasmic domain of MARCO that mediate these functions.

Even though little is currently known with regards to the intracellular region of MARCO, another cA-SR, SRAI, has been more extensively studied.

1.5 The function of the cytoplasmic region of SRAI

To date, the cytoplasmic domain of SRAI has been studied in more detail than that of MARCO. Early work by Fong *et al.* (1996) suggests that SRAI's intracellular domain is phosphorylated and shows that disrupting this process alters SRAI's cell-surface expression patterns and ability to internalize bound ligands. Further work identified serine residues at positions 21 and 49 as potential sites of phosphorylation since mutating either of these residues effected internalization of acLDL and/or the surface expression of SRAI (Fong and Le, 1999). Additionally, studies of the kinetics of ligand internalization indicated that serine phosphorylation of the cytoplasmic domain increases upon internalization of a bound ligand (Fong and Le, 1999). Further, the disruption of a serine residue at position 48 in the cytoplasmic domain of SRAI reduced surface expression of the receptor (Heider and Wintergerst, 2001), indicating that this residue, and perhaps its phosphorylation, may be critical for this process.

In addition to predictions of post-translational modifications, it has recently been demonstrated that there are proteins that interact with the intracellular region of SRAI. In particular, this region directly interact with tumor necrosis factor (TNF) receptor-associated factor 6 (TRAF6) to inhibit TLR4-induced activation of NF-κB signalling (Yu *et al.*, 2011). Findings such as these suggest that the cytoplasmic domains of other cA-SRs may be fundamental to their functional role and thus deserve

further investigation.

1.6 Central paradigm

Previous work with both SRAI and MARCO has demonstrated the importance of the intracellular regions of the cA-SRs. Specifically, experiments conducted by Kroos using cytoplasmic mutants of MARCO suggest that this region is important for a variety of the protein's functions. It is well established that through the course of evolution, some regions of a gene are more susceptible to the accumulation of mutational change than others, and that this bias is due to the biological importance of such a region (Alberts, 2008). Highly conserved regions are hypothesized to encode for essential motifs within a protein that cannot support any changes to their function brought on by mutation (Alberts, 2008). Therefore, I hypothesized that comparing the sequences of MARCO proteins across a range of organisms would identify specific areas of conservation within the cytoplasmic domain and that this conservation would be indicative of areas essential for function. Identified motifs in the cytoplasmic region may be conserved sites of post-translational modifications or docking sites for interacting proteins. Thus, I hypothesize that evolutionarily conserved motifs in the cytoplasmic domain of MARCO are essential for the function of the protein. Based on this central paradigm, I propose the following specific hypotheses.

1.6.1 Specific hypotheses

1. Areas of conservation shared amongst the 5 members of the cA-SR protein family may indicate a shared evolutionary history, which would explain some similarities in their function and expression patterns.

2. Motifs in MARCO and possibly other cA-SRs, specifically in the intracellular region, are conserved throughout the evolution of these proteins.

3. Areas of conservation in the cytoplasmic domain of MARCO are important for the function of this protein. Post-translational modifications and putative adaptor binding proteins are associated with this region in a manner that is essential to its function.

1.6.2 Aims

Based on these specific hypotheses, my research has been conducted under 3 aims as follows.

1. Determine the putative evolutionary history of the cA-SRs and use this to determine areas of conservation in the cytoplasmic domain of MARCO.

2. Use *in silico* techniques to predict the presence of post-translational modifications and putative adaptor binding proteins in the cytoplasmic domain of the *Homo sapiens* MARCO protein.

3. Assess the validity of the *in silico* predictions of post-translational modifications and putative docking sites for interacting proteins in MARCO via *in vitro* methods.

Chapter 2

Materials and Methods

2.1 Cell culture

2.1.1 General conditions for mammalian cell culture

HEK293T (ATCC, Number CRL-11268), HeLa (Sir William Dunn School of Pathology, Oxford), and HT1080 (a kind gift from Dr. Peter Pelka) cell lines were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS), 2mM L-glutamine, and 100 units/ml penicillin (herein referred to as complete DMEM media). Cells were incubated at 37°C in a 5% CO_2, 95% humidity incubator. Cells were sub-cultured between 1:5 and 1:10 twice weekly by detaching cells with the addition of 2ml of 0.05% Trypsin (Gibco, Invitrogen) at 37°C for approximately 5 minutes.

2.2 Molecular biological methods

2.2.1 Antibody production

Monoclonal anti-Myc antibody

The mouse anti-human Myc antibody-producing hybridoma, 9E10, was grown in house. Cell stocks stored in liquid nitrogen were rapidly thawed in a 37°C water bath then incubated in 20ml of complete DMEM media in a T50 flask (Falcon). Upon reaching 70-80% confluency, the cells were split into 3 separate T75 flasks (Falcon). These were grown until complete cell death (approximately 3 weeks) at which time the solution was transfered to a 50ml Falcon tube which was supplemented with 20mM of 1M tris(hydroxymethyl)aminomethane (Tris) (pH 8.0) and 0.01% sodium azide. The solution was centrifuged at 4°C at a speed of 1000rpm for 5 minutes. The resultant supernatant containing anti-Myc monoclonal antibody was used in further experiments at a dilution of 1:1000.

Polyclonal anti-hMARCO antibodies

The rabbit anti-human MARCO antibodies were produced in house by colleagues in the Bowdish laboratory. This reagent was specific for hMARCO except for an additional unidentified protein at approximately 37kDa in size, and is used in all subsequent experiments at a dilution of 1:2500.

2.2.2 Transfection of HEK293T, HeLa, and HT1080 cells

Cells were seeded on day 0 at approximately 3 x 10^5 cells per 1 ml of complete DMEM media, which translates to 4 x 10^4 cells per well in a Lab-Tek II microscopy

chamber slide (Thermo Scientific), 9 x 10^5 cells per well in a 6-well plate (Falcon), or 3 x 10^6 cells per 20 cm tissue culture dish (Nunc). On day 1, the cells were transfected using 8μl of branched polyethylenimine (Sigma-Aldrich, C.408727) diluted 1:1000 in phosphate buffered saline (PBS) in combination with 2μg of DNA. Cells were used for subsequent analysis on day 4.

2.2.3 Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide separating gels of 10% and 13.5% (pH 8.8) and stacking gels of 5% (pH 6.8) were used to separate proteins based on molecular size. Cell lysates and column fractions were mixed with SDS Sample Buffer (50mM Tris, 2% SDS, 0.1% bromophenol blue, 10% glycerol) at a ratio of 3:1 then boiled for 10 minutes, unless otherwise stated. All gels were run using the Mini-PROTEAN III gel system (Biorad). Gels were electrophoresed at 200V per 1.5mm-thick gel in running buffer (25mM Tris, 250mM glycine (pH 8.3), 0.1% SDS) until the dye in the loaded material had reached the end of the separating gel (approximately 40 minutes). Molecular sizes of resulting protein bands were estimated by comparison to standards of known size ranging from 10 to 250kDa (Kaleidoscope, Biorad). A subset of these gels were stained using Coommassie Blue Silver (0.12% Coommassie G-250, 10% ammonium sulfate, 10% phosphoric acid, 20% methanol in ultrapure water) overnight.

2.2.4 Western blotting

Proteins were transfered from SDS-PAGE to a polyvinylidene difluoride (PVDF) membrane using the Mini-PROTEAN III cassette (Biorad) at 100V, 4°C for 1 hour in transfer buffer (25mM Tris, 250mM glycine (pH 8.3)) mixed with methanol in a 9:1 ratio. The membrane was then blocked for 1 hour at room temperature in blocking solution (5% skim milk powder dissolved in Tris-buffered saline with 0.1% Tween 20 (TBST)) followed by incubation overnight in primary antibody diluted in blocking solution. The membrane was then washed 3 times for 10 minutes each in TBST followed by incubation with horse radish peroxidise (HRP)-conjugated secondary antibody (used at a dilution of 1:10000, as per the manufacturer's instructions) for 1 hour at room temperature. The membrane was again washed 3 times for 10 minutes in TBST and developed using the ECL Western blotting detection system (GE Healthcare, as per the manufacturer's instructions), with Kodak BioMax XAR film.

2.2.5 Immunofluorescence and microscopy

HEK293T and HeLa cells were seeded into Lab-Tek II microscopy chamber slides (Thermo Scientific) at 4 x 10^4 cells per well in complete DMEM and transfected as described in section *2.2.2*. For immunofluorescence staining, cells were fixed in 300μl of 4% paraformaldehyde (PFA) for 10 minutes at room temperature followed by permeabilization of the cells using 300μl of 0.1% Triton X-100 in PBS for 10 minutes at room temperature. Further, cells were blocked in 300μl of blocking buffer (1% goat serum, 1% bovine serum albumin (BSA) in PBS) overnight at 4 °C. To immunolabel, the cells were incubated with the primary antibody, diluted in blocking buffer, for 1 hour at room temperature. Following washes with PBS, the cells were exposed to a

secondary antibody and phalloidin mixture diluted in blocking buffer for 1 hour at room temperature. The microscopy slide was then removed from the chamber, and mounted with ProLong Gold (Invitrogen) and cover slips (22 x 50, VWR). Once the mounting media had dried (approximately 20 minutes), nailpolish was applied to the edges of the cover slip in order to seal the specimen.

In the case of HEK293T cells, the protocol was altered to lift cells before fixing with 4% PFA. Fixing, permeabilization, blocking, and immunolabelling were all done in suspension. The cells were then deposited onto microscopy slides using a Cytospin (Thermo Scientific). The protocol was altered for these cells because of their low adherence to tissue culture plastic which made obtaining large numbers of them difficult.

2.2.6 His-tagged protein isolation

Fast protein liquid chromatography (FPLC) was used to isolate the hMARCO-T4-construct from a protein mixture. First, the hMARCO-T4-construct was transfected into 30 150mm dishes of HEK293T cells as per section *2.2.2*. The cells were collected and lysed in 1% NP40 lysis buffer (20mM Tris, 137mM NaCl, 10% glycerol, 1% NP40 diluted in ultrapure water) then passed over a nickel affinity chromatography column to which the polyhistidine-tag of the hMARCO-T4-construct bound. Binding buffer (20mM Tris, 150mM NaCl, 5mM imidazole) was replaced with elution buffer (20mM Tris, 500mM NaCl, 500mM imidazole) in a linear fashion to promote the elution of proteins bound to the column over 200 100μl fractions.

2.2.7 Enzyme-linked immunosorbent assay

The enzyme-linked immunosorbent assay (ELISA) was used to determine which fractions from the His-tagged protein isolation protocol (*2.2.6*) contained the hMARCO-T4-construct. On day 1, samples from each fraction (diluted 1:25 in PBS) were placed into a 96-well high affinity binding ELISA plate (Nunc) to a final volume of 100μl. Following overnight incubation, the plate was washed with PBS supplemented with 0.05% Tween 20 and blocked for 2 hours with 300μl of assay diluent (10% FBS in PBS). Further, the primary antibody, 9E10, was diluted in assay diluent to a final concentration of 1:50, applied to the plate, and allowed to incubate overnight at 4°C. On day 3, the plate was washed with PBS supplemented with 0.05% Tween 20 and incubated with HRP-conjugated goat anti-mouse secondary antibody diluted 1:5000 in assay diluent. The plate was again washed and exposed to 100μl of tetramethylbenzidine (TMB) substrate in order to detect the HRP activity. Once the wells had turned a vibrant shade of blue (approximately 20 minutes), 50μl of 2N H_2SO_4 was applied to each well and the plate was read on a EnVision plate reader.

2.2.8 Bradford assay

A Bradford assay was used to determine the amount of protein present in fractions from the His-tagged protein isolation method (*2.2.6*). A DC Bradford Assay kit (Biorad) was used as per manufacturer instructions on 5μl of non-diluted fractions and compared to a BSA standard by determining the absorbance of the plate at a wavelength of 595nm.

2.2.9 Protein immunoprecipitation

Proteins were transfected into HEK293T cells (as per *2.2.2*), collected, and lysed with either 0.1% Triton X-100, 1% RIPA buffer (150mM NaCl, 1.0% NP40, 0.5% sodium deoxycholate, 0.1% SDS, 50mM Tris (pH 8.0) in ultrapure water), or 0.5% NP40 lysis buffer supplemented with protease inhibitor and N-ethylmaleimide. Extracts were than incubated with the 9E10 primary antibody for 1 hour at 4°C while in constant motion. Protein A sepharose beads were then added and the solution was further incubated overnight. The solution was centrifuged at 13,000rpm for 5 minutes at 4°C, the supernatant removed and saved for later use, and the beads washed in the lysis buffer solution. After 3 wash steps, the samples were analysed using a SDS-PAGE gel as per *2.2.3*.

2.3 Bioinformatic methods

2.3.1 Mining and annotation of class A Scavenger receptor mRNA and amino acid sequences

Vertebrate genomes from the NCBI's GenBank (http://www.ncbi.nlm.nih.gov /genbank/) and EBI's Ensembl (http://www.ensembl.org) databases were analysed for novel cA-SR amino acid sequences. Known cA-SRs were used as queries to the basic local alignment search tool (BLAST) (Altschul *et al.*, 1990), position-specific iterated BLAST (PSI-BLAST) (Altschul *et al.*, 1997), BLAST-like alignment tool (BLAT) (Kent, 2002), and HMMER (Eddy, 1998). Class A Scavenger receptors were identified as consisting of a C-terminal SRCR domain in the case of MARCO, SRAI, and SCARA5, or a C-type lectin domain in the case of SCARA4, connected to a

collagenous region, consisting of at least 70 amino acids in length, that additionally shared significant sequence similarity with known cA-SR sequences. Novel SCARA3 proteins were annotated based on full-length sequence similarity to known SCARA3 sequences. Additional gene synteny analyses were conducted with the aid of the UCSC Genome Browser (Kent *et al.*, 2002) when only partial sequences were available. When appropriate, publicly available predicted transcript data were manually edited to reflect known cA-SR exon structure. In the case where only partial sequences were available, the sequences were omitted from further analyses.

2.3.2 Multiple sequence alignments

Multiple sequence alignments of the cA-SR mRNA and amino acid sequences were generated using multiple sequence comparison by log-expectation (MUSCLE) (Edgar, 2004) and viewed using JalView 6.7.1 (Waterhouse *et al.*, 2009). Known and newly annotated cA-SR sequences are presented in *Table 3.2*.

2.3.3 Construction of phylogenetic trees

Molecular phylogenies of the cA-SR mRNA and amino acid sequences were created using both maximum likelihood and Bayesian probabilistic methods of evolution. These methods were implemented using the RAxML-VI-HPC v7.2.8 (Stamatakis, 2006) and MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) software packages, respectively. The appropriate substitution models for each phylogeny were determined by JModelTest (Posada, 2008) and ProtTest (Abascal *et al.*, 2005). The MARCO mRNA data were estimated to fit most appropriately with the Generalized Time-Reversible (GTR) model including both invariable sites

(I) and a discrete gamma (G) distribution. All other mRNA data were estimated to be best represented by the GTR+G model. To create the phylogenies for gene trees based on full-length mRNA sequences, MrBayes analyses were run for 3 million generations; for comparisons of SRCR domains across SRAI, MARCO, and SCARA5, MrBayes was run for 10 million generations. All Bayesian phylogenies were sampled every 1000 generations and a 25% burn-in period was used. Convergence was confirmed by use of the AWTY (Nylander *et al.*, 2008) software package and variation in likelihood values were visualized using Tracer v1.5 (Rambaut and Drummond, 2009b). Maximum likelihood phylogenies were also created using the appropriate substitution models and were subject to 100 bootstrap replicates. All trees were mid-point rooted using FigTree v1.3.1 (Rambaut and Drummond, 2009a).

2.3.4 Domain characterization similarity measures, and principal component analyses

Full-length cA-SR sequences were analysed via phylogenetics and principal component analysis (PCA). PCA plots were based on pairwise distances calculated between the cA-SRs across all species and were generated using STAMPv2.0.0 (Parks and Beiko, 2010). Additionally, each domain of these receptors was analysed independently. In order to determine the domain architecture of each cA-SR, the boundaries of each domain were calculated using bioinformatic software. The cytoplasmic and transmembrane domains were determined with TMHMM2.0 (Krogh *et al.*, 2001). The α-helical regions were identified with the JUFO Server (http://www.meilerlab.org /index.php/servers/show?s_id=5) and PSIPred (Jones, 1999). The collagenous, SRCR, and C-type lectin domain boundaries were determined via NCBI's Conserved Domain

Database (CDD) (Marchler-Bauer *et al.*, 2010). Additionally, permutation tests to compare each of the *Homo sapiens* cA-SR amino acid sequences were generated using PRSS with 1000 iterations (Smith *et al.*, 1981; Pearson and Lipman, 1988). Permutation tests such as these allow one to determine the homology between 2 sequences by calculating the likelihood that the 2 sequences are related followed by randomized tests that determine a probability score that explains the likelihood that such a score could indicate homology by chance. Percent identity measures calculated for the same sequences were based on pairwise distance scores calculated using EBI's EMBOSS Needle global alignment algorithm using default settings (Rice *et al.*, 2000).

2.3.5 Prediction of post-translational modifications in the cytoplasmic domain of MARCO

In an effort to use *in silico* techniques to inform *in vitro* analyses, a variety of bioinformatic tools were utilized to hypothesize what types of post-translational modifications may be taking place in the cytoplasmic domain of the MARCO protein. For simplicity, the amino acid sequence of *Homo sapiens* MARCO (hMARCO) protein was used for all analyses. Multiple tools were utilized, including the NetCGlyc 1.0 Server to predict C-mannosylation sites in mammalian proteins (Julenius, 2007), NMT - The MYR Predictor for prediction of N-terminal N-myristolyation of proteins (http://mendel.imp.univie.ac.at/myristate/SUPLpredictor.htm), PrePS - Prenylation Prediction Suite to examine whether hMARCO is a prenylated protein (Maurer-Stroh and Eisenhaber, 2005), the NetPhos 2.0 Server for the predictions of phosphorylation sites on serine, theonine and tyrosine residues (Blom *et al.*, 1999), ExPASy's Sulfinator tool for the prediction of tyrosine sulfation sites (Monigatti

et al., 2002), the SUMOplot Analysis tool to predict the probability of sumoylation sites within a protein sequence (http://www.abgent.com/tools/), the ProP 1.0 server which predicts arginine and lysine propeptide cleavage sites (Duckert *et al.*, 2004), and UbPred which predicts protein ubiquitination sites (Radivojac *et al.*, 2010). The full-length hMARCO amino acid sequence was used as input for these tools, with emphasis being placed on those results showing post translational modifications in the cytoplasmic domain of the protein.

2.3.6 Prediction of adaptor binding proteins in the cytoplasmic domain of MARCO

In order to make *in silico* predictions of the occurrence of short linear motifs in the cytoplasmic domain of hMARCO, the following tools were used:

- ELM - Eukaryotic Linear Motifs, (Puntervoll, 2003)
- PROSITE, (Bairoch, 1991)
- MotifScan, (Obenauer, 2003)
- AutoMotif Server 2.0, (Plewczynski *et al.*, 2005)

Chapter 3

The Evolution of the class A Scavenger receptors

3.1 Introduction

3.1.1 Rationale

MARCO was originally identified in the *Mus musculus* (Elomaa *et al.*, 1995), and shortly after, the *Homo sapiens* (Elomaa *et al.*, 1998) genomes. Sequence identity comparisons identify 51% similarity between the cytoplasmic regions of these 2 proteins (Elomaa *et al.*, 1998). Since it is known that conserved regions are more important for function than areas that mutate over evolutionary time, it can be deduced that motifs of functional importance lie within the homologous regions shared between the human and mouse. However, there are many other identified MARCO proteins that have not yet been analysed in this fashion. By investigating these areas of conservation, I can further identify conserved motifs that may, as of yet, have

undescribed function. Additionally, by identifying how MARCO is related to the other members of the cA-SRs, I can use comparisons of these sequences to identify conserved regions that may be essential to the function of all cA-SRs.

3.2 Results

3.2.1 The cA-SRs share similar domain architectures

Sixteen SRAI, 21 MARCO, 21 SCARA3, 25 SCARA4, and 22 SCARA5 full-length mRNA and protein sequences were identified and analysed in this study (*Table 3.2*). An exhaustive bioinformatic search was undertaken in order to identify these receptors, including searches of all SRCR-containing proteins for transmembrane, α-helical, and collagenous domains using various bioinformatic tools. These extensive methods were used in order to best identify any ancient homologs, pseudogenes, or related proteins that had undergone various domain swap or fusion events. Many of the cA-SRs examined have not been previously annotated and therefore represent novel cA-SR sequences. While a subset of these are partial sequences, no pseudogenes were identified.

Previous analyses of the domain structures of the cA-SRs have been inconsistent; therefore, I re-examined these predictions using current bioinformatic tools. Amino acid sequences of cA-SR proteins from the *Homo sapiens* genome were used as representative members to determine domain architectures. The relative lengths and composition of these domains are visualized in *Figure 3.1* and explained in detail in *Table 3.3*. Cytoplasmic and transmembrane domains were established using the TMHMM software (Krogh *et al.*, 2001) and were determined to be approximately

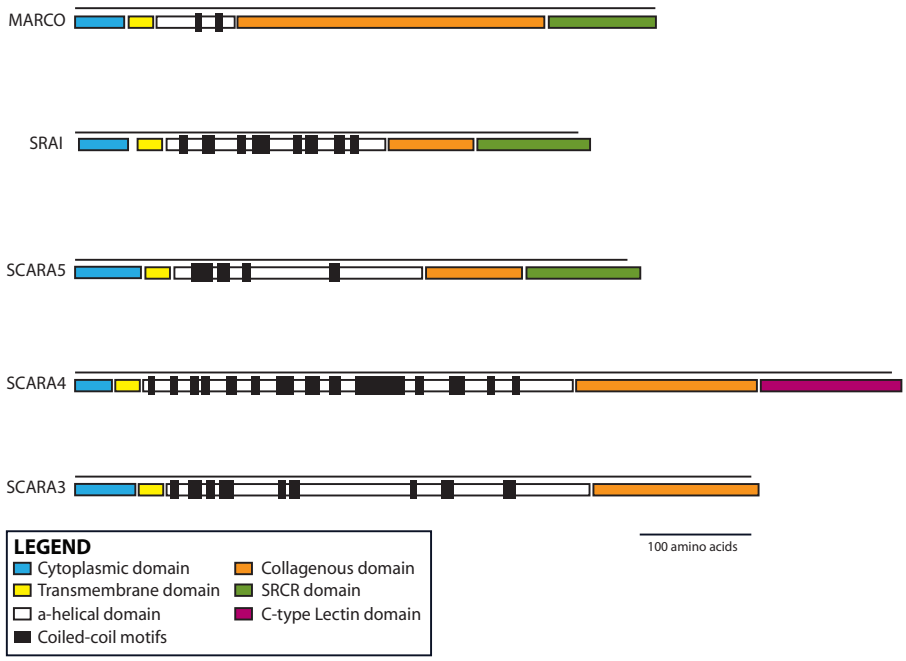


Figure 3.1: **The protein domain architecture of the class A Scavenger receptors.** Structures are scaled based on the length of each domain. The cytoplasmic and transmembrane domains were determined using TMHMM software; α-helical domains were determined using the JUFO Server and PSIPred. The boundaries of the collagenous, SRCR, and C-type lectin domains were determined using NCBI's CDD. Heptad coiled-coil motifs of the α-helical domains were identified based on the definition put forth in McAlinden *et al.* (2003) and Parry *et al.* (2008). Domain boundaries are supported by *Table 3.3*.

30-55 and 20 amino acids long, respectively, in each receptor.

Previous work indicated the region between the transmembrane and collagenous domains to be a combination of a spacer and α-coiled-coil region dependent on the receptor in question (reviewed in Bowdish and Gordon, 2009). My analyses using the JUFO Server (www.jens-meiler.de/jufo.html) and PSIPred (Jones, 1999) indicate that this region is primarily α-helical in all 5 receptors and includes multiple coiled-coil motifs (*Fig. 3.1, black boxes*). The coiled-coil motifs are based on heptad

motifs of the form HxxHccc (McAlinden, 2003; Parry *et al.*, 2008), where hydrophobic residues (H) appear at the first and fourth positions of a 7 amino acid sequence, with positions 5 to 7 tending to be charged (c). Variations on this 3-4 separation pattern of hydrophobic residues include 4-4, 3-3, and 3-1 repeats (Parry *et al.*, 2008). These motifs have been shown to be necessary for oligomerization in other proteins (McAlinden, 2003) and, thus, are likely to contribute to the trimerization of the cA-SRs.

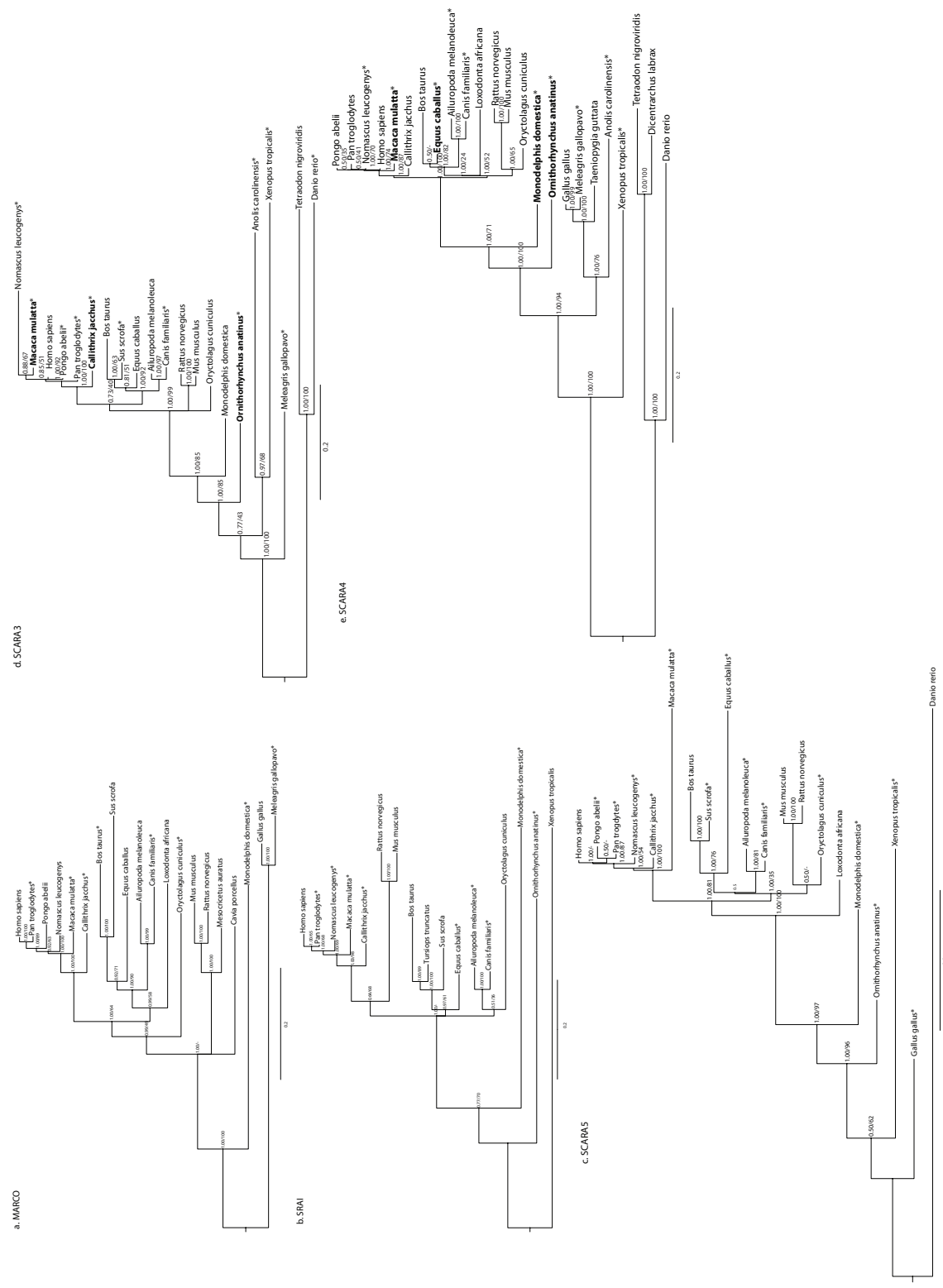
The boundary between this α-helical domain and the collagenous region was determined using the characteristic Gly-Xxx-Yyy repeat (reviewed in Bowdish and Gordon, 2009), which appears over the full-length of the collagenous domain. The C-terminal domains have been previously annotated in NCBI and were confirmed using NCBI's CDD. The resulting domain architecture shows strong similarities across the cA-SR protein family.

3.2.2 Classification of known and novel cA-SRs

Bayesian and maximum likelihood phylogenies were constructed for each of the 5 protein family members using full-length mRNA sequences of the known and novel cA-SRs gathered from available genomes present in the NCBI and Ensembl databases. Novel cA-SRs were identified based on domain structure, synteny analyses, and pairwise sequence identity scores as compared to known cA-SRs. Phylogenies of these sequences were created to examine and confirm the within group relatedness of these proteins across vertebrate species.

The molecular phylogeny of full-length MARCO mRNA sequences (*Fig. 3.2a*) details the conservation of MARCO across mammalian and avian species. A partial

Figure 3.2 *(following page)*: **Phylogenetic trees of known and novel class A Scavenger receptors indicate conservation of these receptors in a subset of vertebrate genomes.** Phylogenies were created based on full-length cA-SR mRNA sequences. Novel sequences are indicated with bold font. Those sequences marked as only predicted in GenBank and Ensembl databases are labelled with an asterisk. MARCO (*a*) was discovered in avian and mammalian genomes. SRAI (*b*) was found in organisms including *Xenopus tropicalis* and mammals; no SRAI sequences were found in publicly-available avian genomes. SCARA5 (*c*) was more phylogenetically diverse then MARCO and SRAI as 3 additional instances of this receptor were found in fish genomes. SCARA3 (*d*) was found in two Teleost fish genomes, *Danio rerio* and *Tetraodon nigroviridis*. SCARA4 (*e*) is the most phylogenetically widespread of the cA-SRs, present in 4 distinct fish species including the early bony fish of the superorder *Acanthopterygii*. Tree topologies were determined using both Bayesian and maximum likelihood methods and are supported by posterior probabilities and bootstrap values as indicated on node labels [BY/ML]. All phylogenies are midpoint rooted; scale bars indicate the number of substitutions per site.


a. MARCO
Homo sapiens
Pan troglodytes*
Pongo abelii
Nomascus leucogenys
Macaca mulatta*
Callithrix jacchus*
Bos taurus*
Sus scrofa
Equus caballus
Alluropoda melanoleuca
Canis familiaris*
Loxodonta africana
Oryctolagus cuniculus*
Mus musculus
Rattus norvegicus
Mesocricetus auratus
Cavia porcellus
Monodelphis domestica*
Gallus gallus
Meleagris gallopavo*
b. 5RAI
Homo sapiens
Pan troglodytes*
Nomascus leucogenys*
Macaca mulatta*
Callithrix jacchus*
Rattus norvegicus
Mus musculus
Bos taurus
Tursiops truncatus
Sus scrofa
Equus caballus*
Alluropoda melanoleuca*
Canis familiaris*
Oryctolagus cuniculus
Monodelphis domestica*
Ornithorhynchus anatinus*
Xenopus tropicalis
c. SCARA5
Homo sapiens
Pongo abelii*
Pan troglodytes*
Nomascus leucogenys*
Callithrix jacchus*
Macaca mulatta*
Bos taurus
Sus scrofa*
Equus caballus*
Alluropoda melanoleuca*
Canis familiaris*
Mus musculus
Rattus norvegicus
Oryctolagus cuniculus*
Loxodonta africana
Monodelphis domestica*
Ornithorhynchus anatinus*
Xenopus tropicalis*
Gallus gallus*
Danio rerio
d. SCARA3
Nomascus leucogenys*
Macaca mulatta*
Homo sapiens
Pongo abelii*
Pan troglodytes*
Callithrix jacchus*
Bos taurus
Sus scrofa*
Equus caballus
Alluropoda melanoleuca
Canis familiaris*
Rattus norvegicus
Mus musculus
Oryctolagus cuniculus
Monodelphis domestica
Ornithorhynchus anatinus*
Anolis carolinensis*
Xenopus tropicalis*
Meleagris gallopavo*
Tetraodon nigroviridis
Danio rerio*
e. SCARA4
Pongo abelii
Pan troglodytes
Nomascus leucogenys*
Homo sapiens
Macaca mulatta*
Callithrix jacchus
Bos taurus
Equus caballus*
Alluropoda melanoleuca*
Canis familiaris*
Loxodonta africana
Rattus norvegicus
Mus musculus
Oryctolagus cuniculus
Monodelphis domestica*
Ornithorhynchus anatinus*
Gallus gallus
Meleagris gallopavo*
Taeniopygia guttata
Anolis carolinensis*
Xenopus tropicalis*
Tetraodon nigroviridis
Dicentrarchus labrax
Danio rerio

transcript of a MARCO-like gene covering the SRCR and a piece of the collagenous domain was found in *Xenopus tropicalis* (*Table 3.2*), indicating that a functional MARCO gene might also be present in amphibians. However, the sequence was excluded from further analyses since the full-length protein sequence spans multiple contigs and could not be reliably constructed. Similarly, SRAI is present in mammalian and amphibian genomes (*Fig. 3.2b*), yet there appears to be a secondary loss of SRAI in avian species as it is absent from the fully sequenced *Gallus gallus* and *Meleagris gallopavo* genomes. SCARA5 appears to be the most abundant of the SRCR-containing cA-SRs, as the gene is conserved in mammals, birds, amphibians, reptiles, and fish (*Fig. 3.2c*).

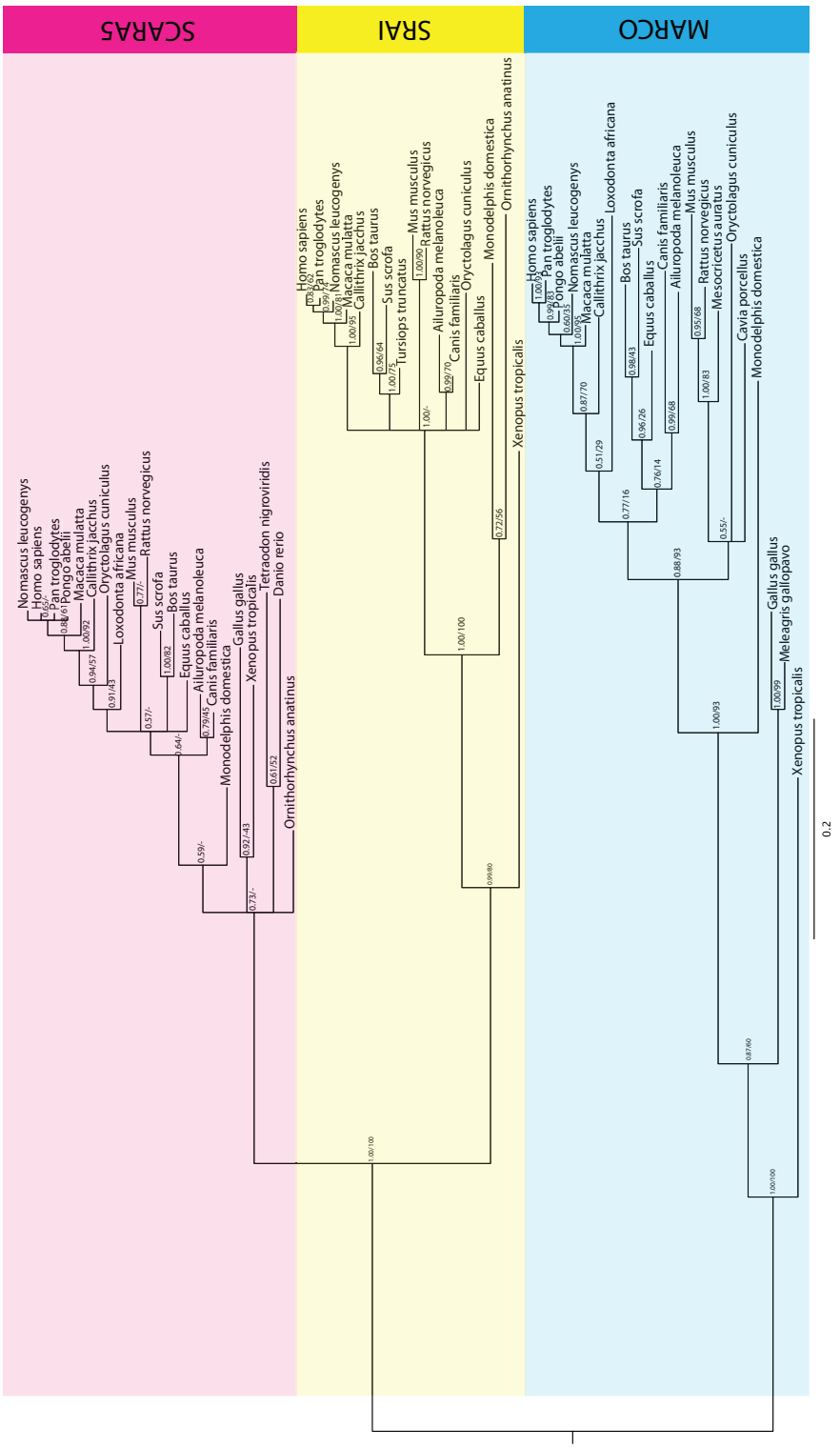
Both of the non-SRCR-containing cA-SRs, SCARA3 and SCARA4, are also present in mammalian, avian, amphibian, reptilian, and fish genomes. Of the 2 proteins, SCARA3 (*Fig. 3.2d*) was found in Ostariophysian and Salmonidae fish species, while SCARA4 (*Fig. 3.2e*) is present in these genomes as well as the bony Acanthopterygii fishes.

Additional phylogenies based on the full-length protein sequences of these receptors were created and found to have tree topologies matching those displayed in *Figure 3.2*.

3.2.3 MARCO, SRAI, and SCARA5 share a highly conserved SRCR domain

Three of the cA-SRs (MARCO, SRAI, and SCARA5) possess an evolutionarily conserved SRCR domain. The SRCR domain is present in many proteins and is highly conserved across both vertebrate and invertebrate species (Martínez *et al.*, 2011).

Figure 3.3 *(following page)*: **Analysis of the SRCR domain across the SRCR-containing class A Scavenger receptor mRNA sequences.** A phylogeny built using both Bayesian and maximum likelihood methods demonstrates the relatedness of the mRNA SRCR domain sequences across MARCO (blue), SRAI (yellow), and SCARA5 (red). Node labels indicate posterior probabilities generated from the Bayesian analysis and bootstrap values of the maximum likelihood tree [BY/ML]. The phylogeny is midpoint rooted; scale bar indicates the number of substitutions per site.


SCARA5
SRAI
MARCO
Nomascus leucogenys
Homo sapiens
Pan troglodytes
Pongo abelii
Macaca mulatta
Callithrix jacchus
Oryctolagus cuniculus
Loxodonta africana
Mus musculus
Rattus norvegicus
Sus scrofa
Bos taurus
Equus caballus
Ailuropoda melanoleuca
Canis familiaris
Monodelphis domestica
Gallus gallus
Xenopus tropicalis
Tetraodon nigroviridis
Danio rerio
Ornithorhynchus anatinus
Tursiops truncatus
Meleagris gallopavo
Cavia porcellus
Mesocricetus auratus
0.2

Phylogenetic analysis of the SRCR domains from these 3 cA-SRs were conducted in order to determine the evolutionary relations between them. By both Bayesian and maximum likelihood methods, the SRCR domains of each receptor cluster together, with those domains from SRAI grouping closer to those of SCARA5 when compared to MARCO (*Fig. 3.3*), indicating that the SRCR domains of SRAI and SCARA5 are more similar to each other than to those of MARCO, and are likely to have diverged from a more recent common ancestor.

3.2.4 The non-SRCR containing cA-SRs - SCARA3 and SCARA4 - are evolutionarily related to each other

Because SCARA3 and SCARA4 do not possess the canonical SRCR domain, it was unclear whether they fit into this family. Before examining this question, I first determined whether SCARA3 and SCARA4 were related to each other. Instead of an SRCR domain, SCARA4 has a C-type lectin domain and SCARA3 terminates after its collagenous region. Permutation tests of *Homo sapiens* SCARA3 and SCARA4 confirmed that their full-length amino acid sequences are statistically similar to each other (*Table 3.1*). Further phylogenetic analyses of the domains shared between these two cA-SRs determined the clustering of SCARA3 and SCARA4 sequences across vertebrate species (*Fig. 3.4*).

3.2.5 Pairwise distances between shared domains demonstrate evolutionary relationships between SCARA3, SCARA4, and the SRCR-containing proteins

In order to examine a possible evolutionary relationship between SCARA3, SCARA4, and the other family members, I employed principal component analysis (PCA) based on pairwise distance scores between cA-SR amino acid sequences.

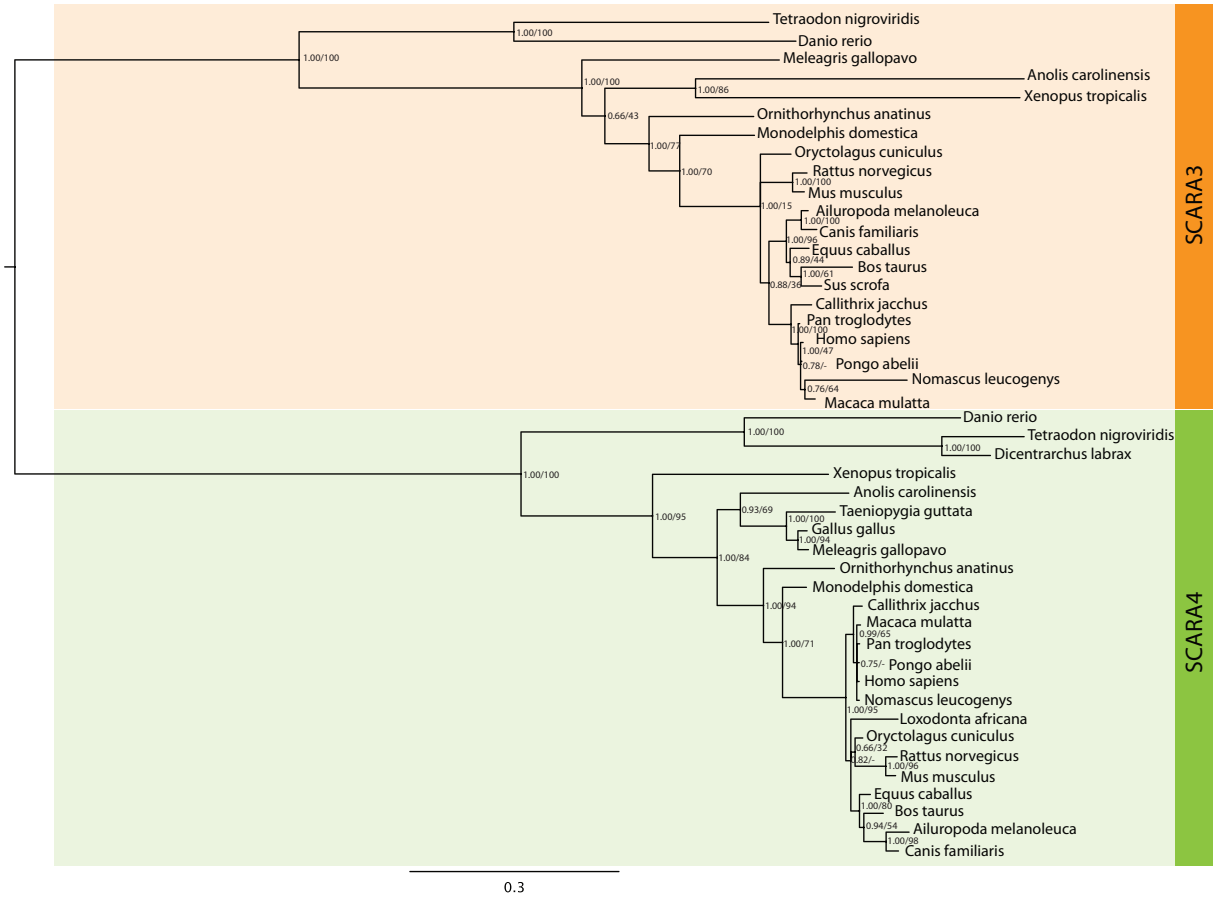


Figure 3.4: **Phylogenetic analysis of the domains shared by SCARA3 and SCARA4 mRNA sequences.** A phylogeny built using both Bayesian and maximum likelihood methods demonstrates the clustering of SCARA3 (orange) and SCARA4 (green) proteins across vertebrate species. Node labels indicate posterior probabilities generated from the Bayesian analysis and bootstrap values of the maximum likelihood tree [BY/ML]. Phylogeny is midpoint rooted; scale bar indicates the number of substitutions per site.

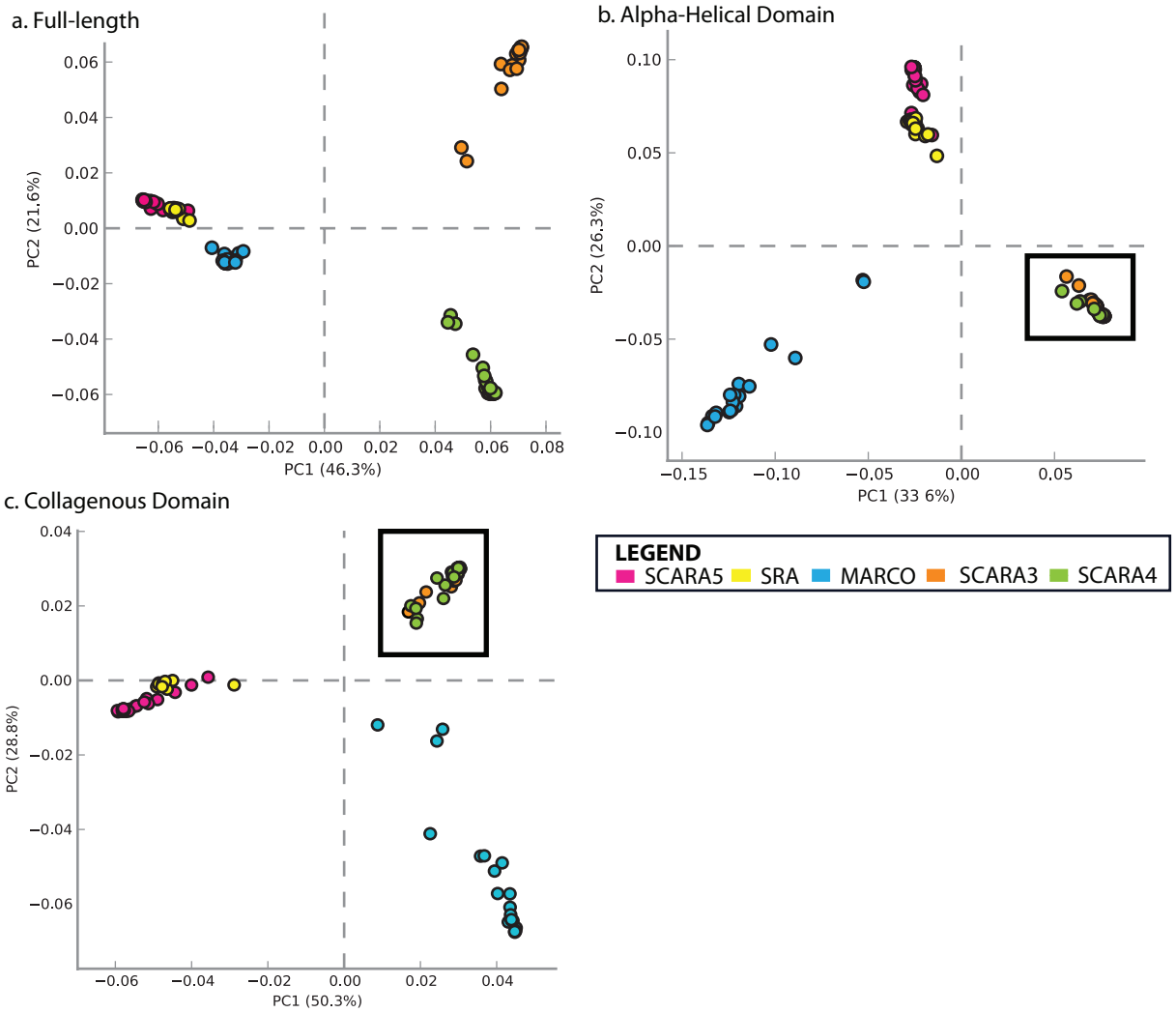


Figure 3.5: **Principal component analysis (PCA) of the pairwise distance scores between the class A Scavenger receptors. a.** Analysis of distance scores between full-length cA-SR amino acid sequences indicates similarity between the SRCR-containing receptors, MARCO, SRAI, and SCARA5. The α-helical (**b**) and collagenous (**c**) domains of SCARA3 and SCARA4 (boxed) cluster closely together, consistent with a shared ancestry between these 2 proteins.

Analyses of the full-length cA-SRs (*Fig. 3.5a*) resulted in a clustering of the SRCR-containing cA-SRs, but not of SCARA3 and SCARA4. This is not surprising given the differences at the C-termini of these 2 receptors when compared to those possessing

the SRCR domain. However, when domains of interest were analysed individually, SCARA3 and SCARA4 often form tight clusters (*Fig. 3.5b-c*), demonstrating that these 2 receptors are closely related, and that a combination of domain deletion and/or swapping events has resulted in the differences in their C-terminal domain structures.

3.2.6 A common ancestry is shared between all 5 members of the cA-SRs

Permutation tests performed using the PRSS software established that each *Homo sapiens* cA-SR amino acid sequence is statistically similar to each other, suggesting a strong evolutionary relationship connecting all members of this family (*Table 3.1*). Additional analyses of similarities across the cA-SR *Homo sapiens* amino acid sequences confirmed significant sequence similarity amongst these proteins. Analyses identified 4 conserved motifs including a cluster of negatively charged amino acids in the cytoplasmic domain of the 5 cA-SRs (*Fig. 3.6, orange boxes*). Furthermore, in addition to the plethora of coiled-coil heptad motifs, a conserved motif in the α-helical domains of each receptor, excluding MARCO, was established (*Fig. 3.6, teal boxes*). A previously predicted ligand-binding motif in MARCO (Brännström *et al.*, 2002) was not found in the SRCR domains of SRAI and SCARA5 (*Fig. 3.6, yellow box*); however, the lysine-rich region in the collagenous domain of SRAI hypothesized to be necessary for ligand binding was found in all other cA-SRs (*Fig. 3.6, pink boxes*). These similarities in domain structure, conserved motifs, and full-length sequence identity support a common evolutionary relationship between all 5 cA-SRs.

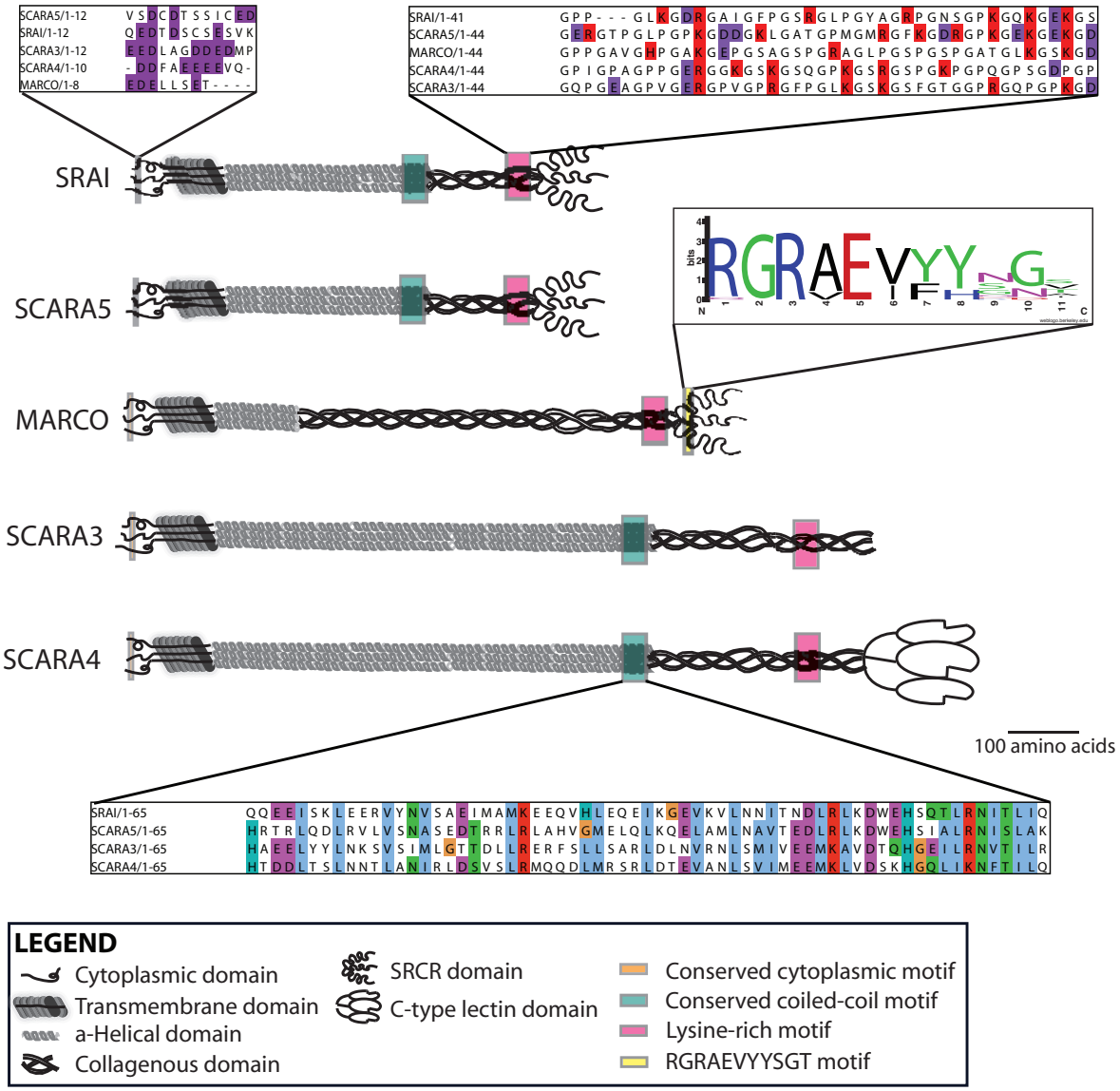
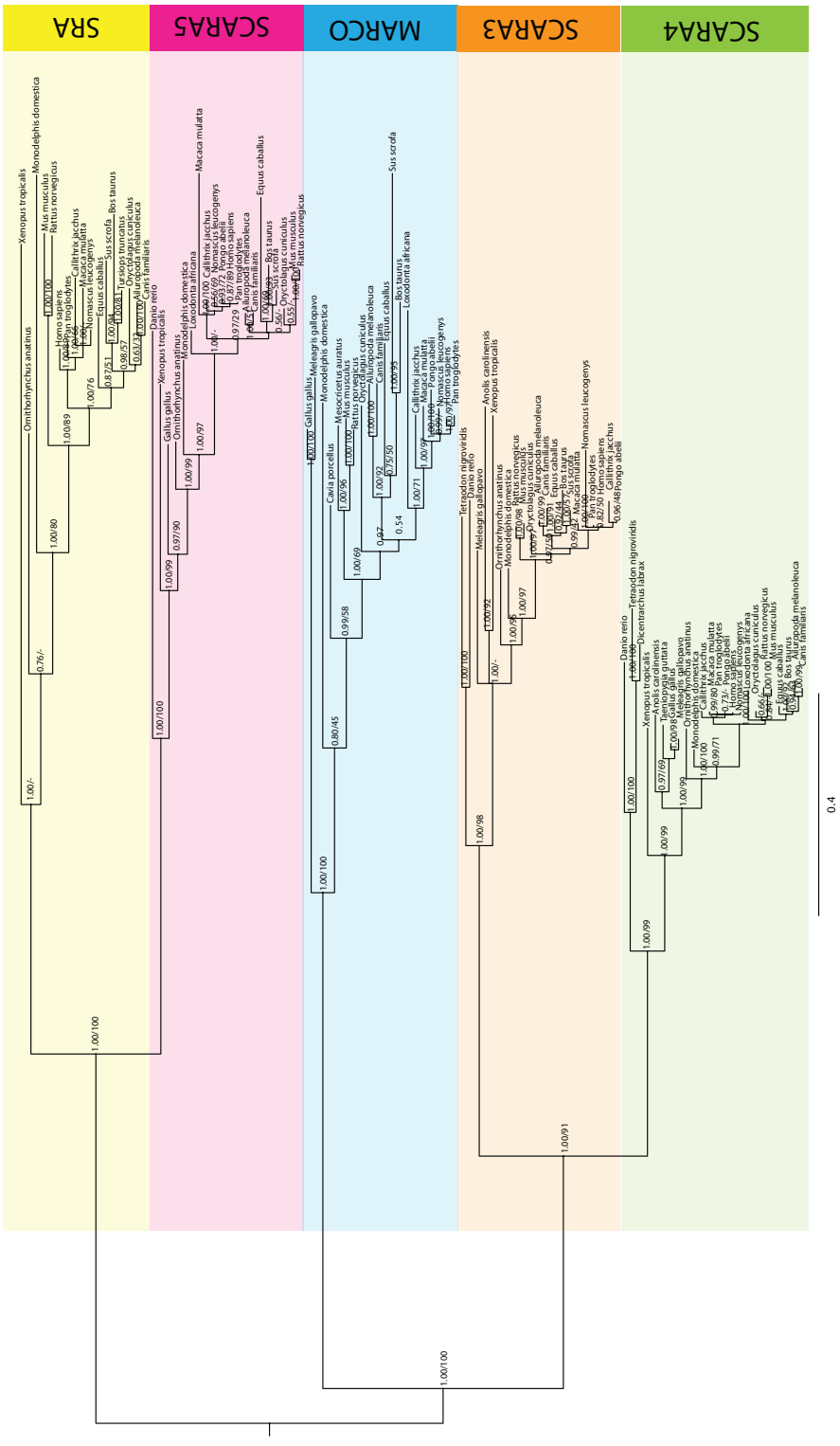


Figure 3.6: **A summary of the common motifs in the class A Scavenger receptors.** Conserved motifs present in these receptors are indicated with coloured boxes at their approximate position within the protein with shout-out boxes used to show the level of conservation across the aligned *Homo sapiens* sequences.

Figure 3.7 *(following page)*: **Phylogeny of all the common domains shared by the class A Scavenger receptors.** Bayesian and maximum likelihood phylogenetic analyses of SRAI (yellow), SCARA5 (red), MARCO (blue), SCARA3 (orange), and SCARA4 (green) show a possible evolutionary history of this protein family. Node labels indicate posterior probabilities generated from the Bayesian analysis and bootstrap values of the maximum likelihood tree [BY/ML]. Phylogeny is midpoint rooted; scale bar indicates the number of substitutions per site.


SRA
SCARA5
MARCO
SCARA3
SCARA4
0.4

3.2.7 The evolutionary history of the cA-SRs

In order to specify the exact relationships amongst the members of the cA-SR gene family, a phylogeny was established using the 4 domains shared across these receptors (*Fig. 3.7*). This phylogeny suggested a strong relationship amongst SCARA3 and SCARA4 in addition to between SRAI and SCARA5, and that MARCO amino acid sequences cluster between the non-SRCR containing receptors and SRAI and SCARA5. Additionally, pairwise identity scores were calculated between each full-length *Homo sapiens* cA-SR protein sequences (*Table 3.1*) which identify a higher level of similarity between MARCO and the other SRCR-containing receptors when compared to between MARCO and the non-SRCR-containing proteins.

Table 3.1: **Percent identity and permutation test scores between the full-length *Homo sapiens* cA-SR amino acid sequences.** The percent identity (dark gray) between each cA-SR was calculated using EBI's EMBOSS Needle global alignment algorithm to quantify the amount of sequence similarity shared amongst these receptors. Additionally, permutation tests were measured using the PRSS algorithm (part of the FASTA package). The probabilities displayed (light gray) are the probabilities that these receptors share sequence alignment similarity scores with each other outside of the distribution when compared to sequence alignments of randomly permutated sequences with the same amino acid contribution.

	SRAI	MARCO	SCARA3	SCARA4	SCARA5
SRAI		25.0%	13.5%	14.7%	32.3%
MARCO	5.37e-18		12.9%	21.8%	25.0%
SCARA3	4.26e-13	1.04e-12		26.6%	18.5%
SCARA4	3.66e-17	1.34e-16	7.23e-54		15.7%
SCARA5	4.88e-52	1.06e-26	8.70e-15	5.34e-25	

3.2.8 Investigation of the cytoplasmic region of MARCO proteins indicate conserved motifs

Specific analysis of the intracellular region of MARCO proteins across mammalian and avian species identified further areas of interest. In addition to the cluster of negatively charged amino acids conserved in all cA-SRs (residues 8-11, *Fig. 3.8*), a motif of positively charged residues is also conserved proximal to the transmembrane domain (*Fig. 3.8*). Additionally, there is a conserved serine residue at position 12-14 in all but 1 sequence that may be an important site of phosphorylation. This

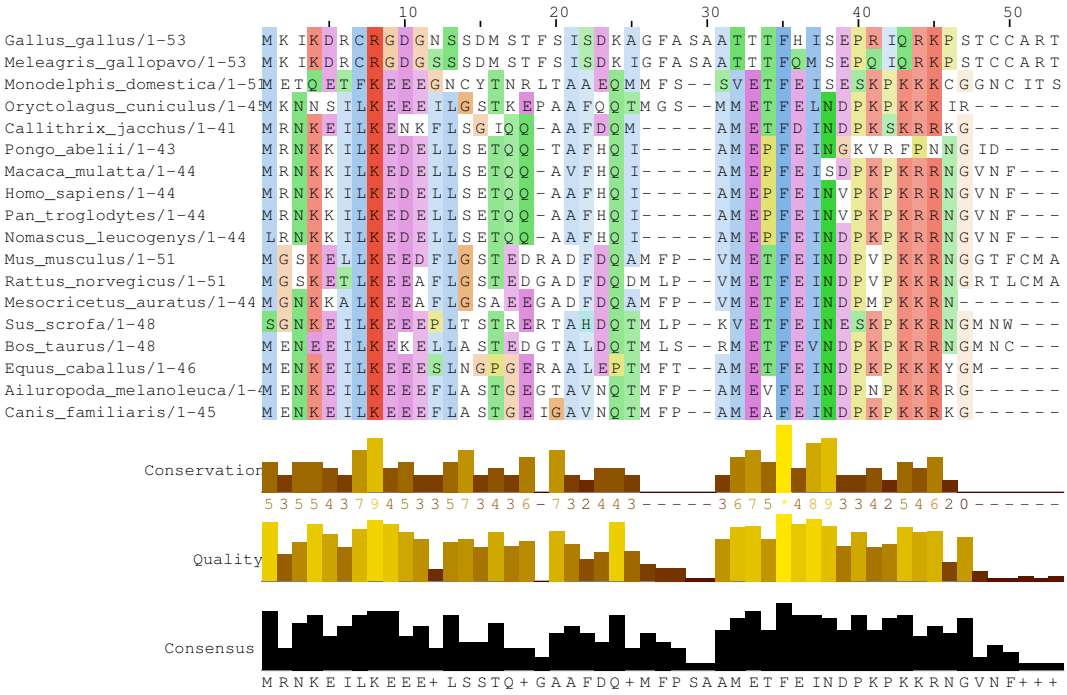


Figure 3.8: **Multiple sequence alignment of the cytoplasmic region of all identified MARCO protein sequences.** Analysis of this domain across species identified areas of conservation that may be imperative to the functionality of MARCO *in vivo*.

prediction was tested in *Chapter 5*. There is additionally a conserved lysine residue at position 8 that may be important in other post-translational modifications such

as ubiquinitation and sumoylation (*Fig. 3.8*) and conserved motifs in the region of residues 35-45.

3.3 Conclusions

I have established that all 5 members of the cA-SRs are evolutionarily related to each other. Specifically, SCARA3 and SCARA4 are closely related which is evident from examination of their protein domain architecture. Similarly, SCARA5 and SRAI share a high level of similarity, notably in their SRCR domains. Furthermore, all 5 members of the cA-SRs share specific motifs in many of their domains, which suggests that all of these proteins share a common ancestry.

Strikingly, a cluster of negatively charged amino acids were identified in the cytoplasmic region of these proteins that was conserved across all members of the cA-SRs. Specific analysis of MARCO proteins revealed notable areas of conservation that may be important as sites of post-translational modifications or docking motifs for interacting proteins.

3.4 Tables

Table 3.2: **Class A Scavenger receptor mRNA and protein sequence information.** Novel sequences are indicated with bold font. Those sequences marked as only predicted in GenBank and Ensembl databases are labelled with an asterisk. Proteins for which only partial sequence information is available are indicated in italics.

mRNA Identifier	**Protein Identifier**	**Organism**	**cA-SA**
NM_138715.2	NP_619729	*Homo sapiens*	SRAI
XM_001140701.2	XP_001140701	*Pan troglodytes*	SRAI*
XM_001097884.2	XP_001097884	*Macaca mulatta*	SRAI*
XM_002756906.1	XP_002756952.1	*Callithrix jacchus*	SRAI*
NM_001113326.1	NP_001106797.1	*Mus musculus*	SRAI
NM_001191939.1	NP_001178868	*Rattus norvegicus*	SRAI
NM_001082248.1	NP_001075717	*Oryctolagus cuniculus*	SRAI
ENSTTRT00000000290	ENSTTRT00000000290	*Tursiops truncatus*	SRAI
NM_001113240.1	NP_001106711	*Bos taurus*	SRAI
XM_003134186.2	XP_003134234	*Sus scrofa*	SRAI
XM_001488563.2	XP_001488613	*Equus caballus*	SRAI*
XM_002928609.1	XP_002928655	*Ailuropoda melanoleuca*	SRAI*
XM_843168.1	XP_848261	*Canis familiaris*	SRAI*
XM_001374058.2	XP_001374095	*Monodelphis domestica*	SRAI*
XM_001512826.1	XP_001512876	*Ornithorhynchus anatinus*	SRAI*
ENSXETG00000017345	ENSXETTT00000037776	*Xenopus tropicalis*	SRAI
XM_003256699.1	XP_003256747	*Nomascus leucogenys*	SRAI*
NM_006770.3	NP_006761	*Homo sapiens*	MARCO
XM_515756.3	XP_515756	*Pan troglodytes*	MARCO*
XM_001083118.2	XP_001083118	*Macaca mulatta*	MARCO*
XM_002749526.1	XP_002749572	*Callithrix jacchus*	MARCO*
NM_010766.2	NP_034896	*Mus musculus*	MARCO
NM_001109011.1	NP_001102481	*Rattus norvegicus*	MARCO
XM_003478607	XP_003478655.1	*Cavia porcellus*	MARCO
XM_002712403.1	XP_002712449	*Oryctolagus cuniculus*	MARCO*
XM_864315.4	XP_869408	*Bos taurus*	MARCO*
ENSSSCG00000015716	ENSSCT00000017116	*Sus scrofa*	MARCO
ENSECAT00000012884	ENSECAT00000012884	*Equus caballus*	MARCO

Continued on next page

Table 3.2 – *Continued from previous page*

mRNA Identifier	**Protein Identifier**	**Organism**	**cA-SR**
ENSAMET00000009311	ENSAMET00000009311	*Ailuropoda melanoleuca*	MARCO
XM_533324.2	XP_533324	*Canis familiaris*	MARCO*
XM_003416671	XP_003416719	*Loxodonta africana*	MARCO
XM_001368441.2	XP_001368478	*Monodelphis domestica*	MARCO*
NM_204736.1	NP_990067	*Gallus gallus*	MARCO
XM_003207707.1	XP_003207755	*Meleagris gallopavo*	MARCO*
XM_002944561.1	*XP_002944607*	*Xenopus tropicalis*	*MARCO**
ENSPPYT00000014855	ENSPPYT00000014855	*Pongo abelii*	MARCO
ENSNLET00000015139	ENSNLET00000015139	*Nomascus leucogenys*	MARCO
AF125191.1	Q9WUB9	*Mesocricetus auratus*	MARCO
NM_016240.2	NP_057324	*Homo sapiens*	SCARA3
XM_519678.3	XP_519678	*Pan troglodytes*	SCARA3*
ENSMMUT00000030287	**ENSMMUT00000030287**	***Macaca mulatta***	**SCARA3**
ENSCJAT00000038371	**ENSCJAT00000038371**	***Callithrix jacchus***	**SCARA3**
NM_172604.3	NP_766192	*Mus musculus*	SCARA3
NM_001108870.1	NP_001102340	*Rattus norvegicus*	SCARA3
ENSOCUT00000009916	ENSOCUT00000009916	*Oryctolagus cuniculus*	SCARA3
ENSBTAT00000010075	ENSBTAT00000010075	*Bos taurus*	SCARA3
XM_003359041.1	XP_003359089	*Sus scrofa*	SCARA3*
ENSECAT00000011732	ENSECAT00000011732	*Equus caballus*	SCARA3
-	**EFB22765**	***Ailuropoda melanoleuca***	**SCARA3**
XP_543225.2	XP_543225	*Canis familiaris*	SCARA3*
ENSMODT00000020278	ENSMODT00000020278	*Monodelphis domestica*	SCARA3
ENSOANT00000002002	**ENSOANT00000002002**	***Ornithorhynchus anatinus***	**SCARA3**
XM_003204578.1	XP_003204626	*Meleagris gallopavo*	SCARA3*
XM_003229633.1	XP_003229681	*Anolis carolinensis*	SCARA3*
XM_002938177.1	XP_002938223	*Xenopus tropicalis*	SCARA3*
ENSTNIT00000015378	ENSTNIT00000015378	*Tetraodon nigroviridis*	SCARA3
XM_687918.4	XP_693010	*Danio rerio*	SCARA3*
XM_002818936.1	XP_002818982	*Pongo abelii*	SCARA3*
XM_003272942.1	XP_003272990	*Nomascus leucogenys*	SCARA3*
NM_130386.2	NP_569057	*Homo sapiens*	SCARA4
ENSPTRT00000018035	ENSPTRT00000018035	*Pan troglodytes*	SCARA4

Continued on next page

Table 3.2 – *Continued from previous page*

mRNA Identifier	**Protein Identifier**	**Organism**	**cA-SR**
ENSMMUT00000006592	**ENSMMUT00000006592**	***Macaca mulatta***	**SCARA4**
ENSCJAT00000000261	ENSCJAT00000000261	*Callithrix jacchus*	SCARA4
NM_130449.2	NP_569716	*Mus musculus*	SCARA4
NM_001025721.1	NP_001020892	*Rattus norvegicus*	SCARA4
ENSOCUT00000029756	ENSOCUT00000029756	*Oryctolagus cuniculus*	SCARA4
NM_001101843.1	NP_001095313	*Bos taurus*	SCARA4
ENSECAT00000020273	**ENSECAT00000020273**	***Equus caballus***	**SCARA4**
XM_002922625.1	XP_002922671	*Ailuropoda melanoleuca*	SCARA4*
XM_843964.1	XP_849057	*Canis familiaris*	SCARA4*
XM_003406773	XP_003406821	*Loxodonta africana*	SCARA4
ENSMODT00000027319	**ENSMODT00000027319**	***Monodelphis domestica***	**SCARA4**
ENSOANT00000018414	**ENSOANT00000018414**	***Ornithorhynchus anatinus***	**SCARA4**
NM_001039599.1	NP_001034688	*Gallus gallus*	SCARA4
XM_003204989.1	XP_003205037	*Meleagris gallopavo*	SCARA4*
ENSTGUT00000010631	ENSTGUT00000010631	*Taeniopygia guttata*	SCARA4
XM_003219678.1	XP_003219726	*Anolis carolinensis*	SCARA4*
XM_002934123.1	XP_002934169	*Xenopus tropicalis*	SCARA4*
GSTENT10018049001	GSTENT10018049001	*Tetraodon nigroviridis*	SCARA4
NM_001122840.1	NP_001116312	*Danio rerio*	SCARA4
ENSPPYT00000010562	ENSPPYT00000010562	*Pongo abelii*	SCARA4
XM_003262015.1	XP_003262063	*Nomascus leucogenys*	SCARA4*
FQ310508.1	CBN82070	*Dicentrarchus labrax*	SCARA4
EU008733.1	*ABV44703*	*Oncorhynchus mykiss*	*SCARA4*
NM_173833.5	NP_776194	*Homo sapiens*	SCARA5
XM_519680.3	XP_519680	*Pan troglodytes*	SCARA5*
XM_002805297.1	XP_002805343	*Macaca mulatta*	SCARA5*
XM_002756806.1	XP_002756852	*Callithrix jacchus*	SCARA5*
NM_028903.2	NP_083179	*Mus musculus*	SCARA5
NM_001135855.1	NP_001129327	*Rattus norvegicus*	SCARA5
XM_002709450.1	XP_002709496	*Oryctolagus cuniculus*	SCARA5*
NM_001102499.1	NP_001095969	*Bos taurus*	SCARA5
XM_003132819.2	XP_003132867	*Sus scrofa*	SCARA5*
ENSECAT00000014531	ENSECAT00000014531	*Equus caballus*	SCARA5*

Continued on next page

Table 3.2 – *Continued from previous page*

mRNA Identifier	**Protein Identifier**	**Organism**	**cA-SR**
XM_002914411.1	XP_002914457	*Ailuropoda melanoleuca*	SCARA5*
XM_543223.2	XP_543223	*Canis familiaris*	SCARA5*
XM_003412459	XP_003412507	*Loxodonta africana*	SCARA5
XM_001370497.1	XP_001370534	*Monodelphis domestica*	SCARA5*
XM_001507041.1	XP_001507091	*Ornithorhynchus anatinus*	SCARA5*
XM_001234365.1	XP_001234366	*Gallus gallus*	SCARA5*
ENSACAT00000014115	*ENSACAT00000014115*	*Anolis carolinensis*	*SCARA5**
XM_002941449.1	XP_002941495	*Xenopus tropicalis*	SCARA5*
CAAE01015120.1	***CAG12980***	***Tetraodon nigroviridis***	***SCARA5***
NM_001030190.1	NP_001025361	*Danio rerio*	SCARA5
XM_002818942.1	XP_002818988	*Pongo abelii*	SCARA5*
XM_003272915.1	XP_003272963	*Nomascus leucogenys*	SCARA5*

Table 3.3 *(following page)*: **Domain boundaries of the representative class A Scavenger receptors in *Homo sapiens*.** Probabilities for the cytoplasmic and transmembrane domains were determined using the TMHMM software tool. The α-helical domain with coiled-coil motifs was determined using the JUFO Server and PSIPred [JUFO;PSIPRED]. The collagenous, SRCR, and C-type lectin domains were determined using NCBI's CDD. "$\wedge$" indicates that probabilities were measured by the corresponding software for each amino acid in the domain and a range is given. P(H) and P(C) represent the probability of a helix or coil at each amino acid (aa) in the domain. "$*$" indicates that there were multiple hits in NCBI's CDD, for which a range of E-values is presented.

		Cytoplasmic			Transmembrane			a-Helical with coiled-coil motifs			Collagenous			SRCR			C-Type Lectin		
cA-SR	NCBI accession	Start	End	Probability	Start	End	Probability	Start	End	Probability	Start	End	E-Value	Start	End	E-Value	Start	End	E-Value
MARCO	NP_006761	0	44	0.67475<P <0.99966^	45	67	0.5591<P <0.99999^	68	147	P(H)>0.333 at 0% of aa; P(H) at >1% of aa	148	420	2.08E-07<E <5.16E-03*	424	520	1.73E-40	-	-	-
										P(C)>0.333 at >95% of aa; P(C) at >95% of aa									
SRA	NP_619729	0	52	0.68085<P <0.94940^	53	75	0.42892<P <0.99990^	76	272	P(H)>0.333 at >75% of aa; P(H) at >60% of aa	273	340	5.68E-22<E <1.41E-08*	350	451	7.38E-48	-	-	-
										P(C)>0.333 at >15% of aa; P(C) at >35% of aa									
SCARA3	NP_057324	0	54	0.77223<P <0.99666^	55	77	0.59362<P <0.99799^	78	457	P(H)>0.333 at >25% of aa; P(H) at >15% of aa	458	601	6.09E-17<E <1.17E-08*	-	-	-	-	-	-
										P(C)>0.333 at >45% of aa; P(C) at >60% of aa									
SCARA4	NP_569057	0	34	0.76775<P <0.99826^	35	56	0.41555<P <0.99977^	57	442	P(H)>0.333 at >25% of aa; P(H) at >25% of aa	441	589	1.26E-21<E <2.84E-17*	-	-	-	606	732	1.30E-61
										P(C)>0.333 at >45% of aa; P(C) at >50% of aa									
SCARA5	NP_776194	0	60	0.63730<P <0.97331^	61	82	0.53247<P <0.99979^	83	305	P(H)>0.333 at >75% of aa; P(H) at >60% of aa	306	380	2.41E-14<E <2.88E-08*	393	495	2.32E-47	-	-	-
										P(C)>0.333 at >20% of aa; P(C) at >30% of aa									

Chapter 4

In silico analysis and prediction of post-translational modifications and putative adaptor binding proteins associated with MARCO's cytoplasmic domain

4.1 Introduction

4.1.1 Rationale

Research previously conducted in Dr. Bowdish's laboratory provided the first data indicating the significance of MARCO's intracellular region. Specifically, it was shown that removing the intracellular domain by mutation impaired the receptor's

ability to internalize bound ligands and adhere to surfaces as compared to the wild-type molecule (Kroos, 2008). Since, the cytoplasmic region of SRAI was shown to be important for endocytosis (Fong, 1996; Yu *et al.*, 2011) and MARCO and SRAI to be closely related at both the sequence and functional levels, I hypothesize that the cytoplasmic domain of MARCO may be essential to the endocytic function of this receptor.

In silico methods have many advantages when they are used to guide *in vitro* experimental designs. By first using easily accessible bioinformatic tools to predict sites of post-translational modifications (PTMs) and protein interactions, I can direct my laboratory investigations in a fashion that results in a more efficient use of resources and time.

4.2 Predictions of post-translational modifications

Of the *in silico* prediction tools used to identify putative PTMs in the cytoplasmic domain of *Homo sapiens* MARCO (hMARCO), predictions were made for ubiquitination, sumoylation, and phosphorylation.

4.2.1 A possible site of ubiquitination at residue 8 in the intracellular region of hMARCO

Ubiquitin, named for its ubiquitous expression in eukaryotic cells (Welchman *et al.*, 2005), is a small, 76 amino acid long protein capable of covalently binding to other proteins in an event termed ubiquitination (Weissman, 2001). Traditionally, ubiquitination has been considered to be a flag for destruction of a protein via the

proteasome (Hicke, 2001), but more recently the PTM has been equated to the multi-functionality of phosphorylation (Weissman, 2001), having confirmed roles in histone regulation (Hicke, 2001), DNA repair, transcription, and signal transduction (Welchman *et al.*, 2005). Importantly for my research, ubiquitination has also been shown to be necessary in regulating endocytosis of certain receptors (e.g. growth hormone receptor (Strous *et al.*, 1996)) and turn over of cell-surface receptors; the binding of ubiquitin can cause internalization of a receptor into an endocytic vesicle directly (e.g. the epithelial Na+ channel) (Hicke, 2001), or can encourage the binding of various ubiquitin-binding proteins (e.g. epidermal growth factor receptor) (Welchman *et al.*, 2005).

There are a limited number of bioinformatic tools available to predict ubiquitination. It is known that ubiquitin forms covalent bonds with lysine residues on target proteins (Hicke, 2001); however, not all lysines are ubiquitinated. In an effort to predict potential ubiquitination sites, Radivojac *et al.* undertook a protein-wide analysis of ubiquitinated proteins from *Saccharomyces cerevisiae* via liquid chromatography and mass spectrometry (Radivojac *et al.*, 2010). Radivojac and colleagues analysed the sequence and structural preferences of ubiquitinated regions and discovered that, in general, areas of disorder and flexibility were more likely to be ubiquitinated (Radivojac *et al.*, 2010). The findings from this study were used to create UbPred, an online bioinformatic tool to predict ubiquitination (Radivojac *et al.*, 2010).

Upon applying the amino acid sequence of hMARCO to this tool, the lysine at residue 8 was predicted to be ubiquitinated with medium confidence (a score of 0.72 out of a possible 1) (*Fig. 4.1*). No other intracellular residues were predicted to have this particular PTM.

One caveat of this prediction tool is that patterns in the sequence surrounding a potentially ubiquitinated lysine residue are species-specific, indicating that the research completed with *S. cerevisiae* may not translate to *Homo sapiens* proteins.

>hMARCO_tail
MRNKKILKEDELLSETQQAAFHQIAMEPFEINVPKPKRRNGVNFSLA

Output:

Residue	Score	Ubiquitinated
4	0.37	No
5	0.39	No
8	0.72	Yes Medium confidence
35	0.33	No
37	0.28	No

Legend:

Label	Score range	Sensitivity	Specificity
Low confidence	0.62 ≤ s ≤ 0.69	0.464	0.903
Medium confidence	0.69 ≤ s ≤ 0.84	0.346	0.950
High confidence	0.84 ≤ s ≤ 1.00	0.197	0.989

Figure 4.1: **UbPred predicts a ubiquitination site at the lysine present at position 8 of hMARCO.** This prediction was made with medium confidence, having a score of 0.72 of a possible 1.0.

4.2.2 Predictions of sumoylation in the intracellular region of hMARCO

In addition to ubiquitin, other ubiquitin-like proteins have been identified that share some level of protein sequence similarity (Welchman *et al.*, 2005). One such set of proteins are the small ubiquitin-related modifiers (SUMOs) which are responsible for the sumoylation of proteins. SUMO-1 through -3 are expressed in a variety of cell types; however, SUMO-4 expression is limited to the kidney, lymph-node, and spleen (Geiss-Friedlander and Melchior, 2007). Similar to ubiquitination, sumoylation is a reversible covalent attachment of a SUMO protein to a specific lysine residue (Geiss-Friedlander and Melchior, 2007). However, unlike ubiquitination, the specific residues surrounding a lysine are reliable predictors of whether it will be sumoylated. A 4 residue SUMO-binding motif has been identified and can be expressed as ψKXE,

where ψ represents a hydrophobic residue, X any amino acid, and K and E a lysine and glutamic acid, respectively (Anckar and Sistonen, 2007). The effect of sumolyation on a protein can be one of many: sumoylation has been shown to encourage protein stability (in contrast to ubiquitin's role in degrading proteins), the localization of a protein to the cytosol, nucleus, or plasma membrane, and the encouragement or inhibition of protein-protein interactions (Geiss-Friedlander and Melchior, 2007). Because of the various contradictions as to the effect of sumoylation on the function of a protein, it is impossible to predict how this modification might affect MARCO.

SUMOplot, an online tool available from Abgent, a biotechnology company specializing in PTMs, uses the known SUMO-binding motif to predict sites of sumoylation in sequences (http://www.abgent.com/tools/). This tool predicted a SUMO-binding motif with high probability at the 8th residue lysine within the cytoplasmic domain of hMARCO, with a score of 0.91 out of a possible 1 (*Fig. 4.2*). No other intracellular residues were predicted to have this particular PTM.

No.	Pos.	Group	Score
1	K18	RNKKI LKED ELLSE	0.91

Figure 4.2: **SUMOplot predicts a single sumoylation site at the lysine at residue 8 in the intracellular region of hMARCO.** This prediction was made with high confidence, having a score of 0.91 out of a possible 1.0.

4.2.3 Serine-phosphorylation predictions in the cytoplasmic domain of hMARCO

Arguably, the most well-studied PTM, protein phosphorylation is the transfer of a phosphate group usually from adenosine triphosphate (ATP) onto a serine, threonine, tyrosine, or, less commonly, other residues (Marks, 2008; Rubin and Rosen, 1975). The addition of this phosphate is mediated by protein kinases, which are a family of proteins with almost 600 identified members in the human genome (Manning *et al.*, 2002). Traditionally, phosphorylation is seen as a means of communication. The addition of the phosphate group can cause significant structural changes to the protein, allowing it to be turned 'on' or 'off' (Marks, 2008). Furthermore, this modification can be reversed by protein phosphatases, resulting in protein phosphorylation being an excellent modulator of protein function (Marks, 2008).

Although there is much diversity in the sequence and structure of protein kinases, there is some level of conservation across their catalytic domains (Blom *et al.*, 1999). This conservation corresponds with a preference to phosphorylate protein sequences with acidic, basic, or hydrophobic residues surrounding a targeted serine, threonine, or tyrosine residue (Blom *et al.*, 1999). Using this knowledge, Blom *et al.* developed NetPhos, an online prediction server of protein phosphorylation. This tool predicts a single phosphorylation site in the cytoplasmic domain of hMARCO at the 14th serine residue with a probability above the threshold defined by the tool (*Fig. 4.3*). Two other residues at positions 16 and 45 were also highlighted by the tool but both had very low predictive scores.

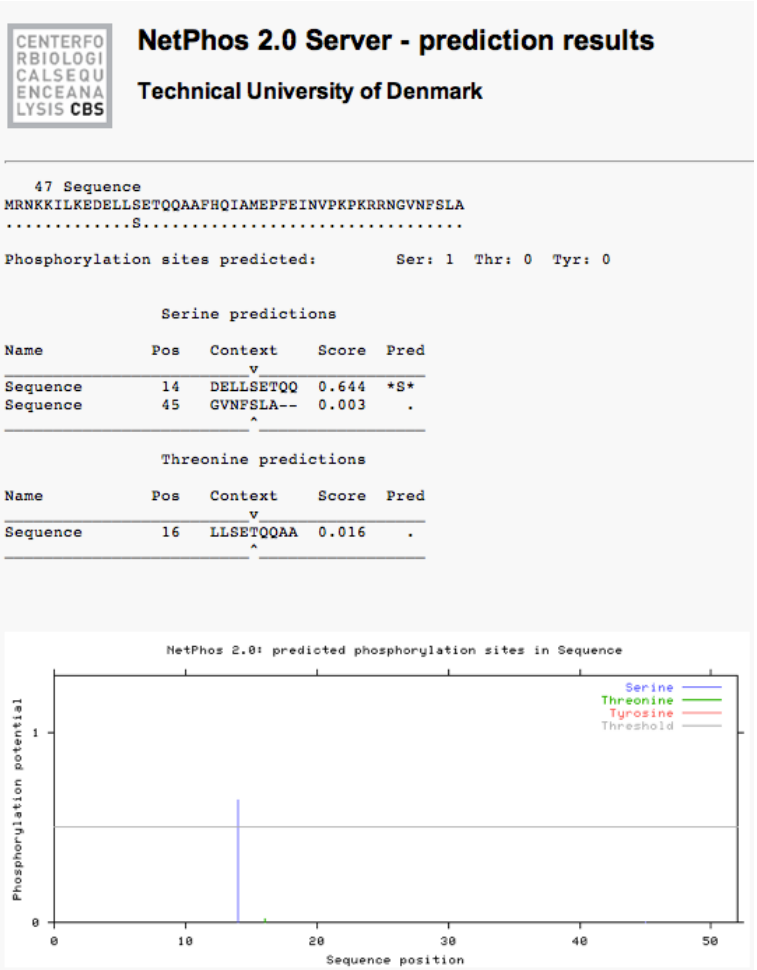


Figure 4.3: **NetPhos predicts a phosphorylation site at the serine at residue 14 in the cytoplasmic domain of hMARCO.** Additionally, 2 residues at position 16 and 45 were predicted to have a low probability of phosphorylation.

4.3 Predictions of adaptor binding proteins

During signal transduction, adaptor proteins bind to form protein-protein interactions which induce down-stream function (Murphy *et al.*, 2011). Protein complexes have traditionally been understood as the interaction of 2 large, globular proteins; however, many recently studied interactions show that these complexes are often

formed between the interaction of a singular domain on one protein with a small (3-8 amino acids) peptide sequence on another (Neduva and Russell, 2006). These sequences, called short linear motifs, are often found in unstructured areas of a protein and bind with lower affinity when compared to typical domain-domain interactions, allowing for their reversible phenotype and quick responses to changing environments (Diella *et al.*, 2008). These motifs are typically loosely conserved across a variety of proteins that bind the same adaptor protein. Alignments of a short linear motif across such proteins will indicate that some positions are highly conserved, while others need only have specific physiochemical properties (Diella *et al.*, 2008). Further, in these motifs some positions serve only as spacers between a set of other interacting amino acids (Diella *et al.*, 2008). This, along with the fact that the short length of these sequences makes it extremely difficult to use statistics effectively, has made the production of bioinformatic prediction tools difficult (Diella *et al.*, 2008). Despite this, a few have been developed.

Of the tools used, only one found evidence of short linear motifs within the cytoplasmic domain of hMARCO despite its lack of structured regions. However, this motif was predicted with a relatively low probability and was only distinguished by 1 of the 4 tools used. The lack of highly probable predictions could be due to the challenges these tools face in making these types of predictions or simply by the fact that there are no short linear motifs in hMARCO's cytoplasmic domain.

4.3.1 Mitogen-activated protein kinase docking motif

The lone prediction made was of a Mitogen-activating protein (MAP) kinase docking sequence by the Eukaryotic Linear Motif (ELM) tool with a very low probability of occurrence (*Fig. 4.4*). MAP kinases are signal transducing molecules which are responsible for the phosphorylation of serine/threonine residues present on a wide range of substrates (Coulombe and Meloche, 2007),

Elm Name	Instances (Matched Sequence)	Positions	View in Jmol	Elm Description	Cell Compartment	Pattern	PHI-Blast Instance Mapping	Structural Filter Info	
LIG_MAPK_1	KPKRRNGVNF KRRNGVNF RRNGVNF	35-44 [A] 37-44 [A] 38-44 [A]	- - -	MAPK interacting molecules (e.g. MAPKKs, substrates, phosphatases) carry docking motif that help to regulate specific interaction in the MAPK cascade. The classic motif approximates (R/K)xxxx#x# where # is a hydrophobic residue.	nucleus, cytosol	[KR]{0,2}[KR].{0,2}[KR].{2,4}[ILVM].[ILVF]	-	-	0.0043

Figure 4.4: **The database of Eukaryotic Linear Motifs (ELM) predicts a MAP kinase docking site in the cytoplasmic domain of hMARCO.** This short linear motif was scored with a very low probability of occurrence of 0.0043 of a possible 1.0.

(Pearson *et al.*, 2001). Typically, these signalling cascades consist of a MAP kinase kinase kinase which acts on a MAP kinase kinase effecting the MAP kinase itself which then acts on a substrate (Coulombe and Meloche, 2007). It is possible in this case that hMARCO is a substrate for a MAP kinase, acting to phosphorylate or dephosphorylate a serine or threonine residue in its cytoplasmic domain, which corresponds well with the predictions of serine-phosphorylation at residue 14. MAP kinase substrates tend to be transcription factors although, as the field develops, other substrates have been discovered (e.g. a protein involved in spindle stability (Lefebvre, 2002), and another which aids in the proliferation of osteoblasts and T cells (Maekawa, 2002)), making it unlikely, though possible, that there may be a MAP kinase docking site in the intracellular region of hMARCO.

As indicated by ELM, the MAP kinase docking motif consists of a short regular expression of the form [KR]{0,2}[KR].{0,2}[KR].{2,4}[ILVM].[ILVF] (*Fig. 4.4*). Upon examination of the cytoplasmic domain of MARCO across all species, this motif was found in only 35% of sequences (*Table 4.1*), suggesting that it may be

Table 4.1: **Identification of the presence of a MAP kinase docking motif in the cytoplasmic domain of MARCO across species.** The short linear motif of the MAP kinase docking sequence is a regular expression of the form [KR]{0,2}[KR].{0,2}[KR].{2,4}[ILVM].[ILVF]. This motif was discovered in hMARCO along with select other organisms, indicated here by "yes" or "no". Matching amino acids are presented in uppercase, whereas misses are indicated in lowercase.

	[KR]{0,2}	[KR]	.{0,2}	[KR]	.{2,4}	[ILVM]	.	[ILVF]	
Homo sapiens	-	K	P	K	RRNG	V	N	F	yes
Pan troglodytes	-	K	P	K	RRNG	V	N	F	yes
Nomascus leucogenys	-	K	P	K	RRNG	V	N	F	yes
Macaca mulatta	-	K	P	K	RRNG	V	N	F	yes
Callithrix jacchus	-	K	S	K	RRKG	V	N	F	yes
Pongo abelli	-	R	FP	n	G	I	D	q	no
Equus caballus	-	K	P	K	KKYG	M	S	c	no
Ailuropoda melanoleuca	-	n	P	K	KRRG	V	N	c	no
Canis familiaris	-	K	P	K	KRKG	V	N	c	no
Sus scrofa	-	K	P	K	KRNG	M	N	w	no
Bos taurus	-	K	P	K	RRNG	M	N	c	no
Oryctolagus cuniculus	-	K	P	K	KKIR	V	N	F	yes
Rattus norvegicus	-	v	P	K	KRNG	r	T	L	no
Mus musculus	-	v	P	K	KRNG	g	T	F	no
Mesocricetus auratus	-	m	P	K	KRNW	g	S	F	no
Loxodonta africana	-	K	P	K	KRNW	V	T	c	no
Cavia porcellus	-	K	P	R	KKNG	M	N	V	yes
Monodelphis domestica	-	K	P	K	KKCG	g	N	c	no
Meleagris gallopavo	-	q	IQ	R	KPS	t	C	c	no
Gallus gallus	-	R	IQ	R	KPS	t	C	c	no

present by chance as opposed to being a functional MAP kinase docking sequence.

4.4 Conclusions

Bioinformatic methods for predicting PTMs are fairly well established; however predicting short linear motifs of docking sequences for protein-protein interactions is a much harder task due to the short length of these motifs and their variable composition. Three possible sites of PTMs were identified in the intracellular region of hMARCO including a target residue for ubiquitination, sumoylation, and phosphorylation. Predictions of adaptor binding proteins were less fruitful, though there is reason to believe that there may still be proteins interacting with this domain because of the difficulties with these types of *in silico* predictions.

Chapter 5

In vitro analysis of predicted post-translational modifications and interacting proteins in the cytoplasmic domain of MARCO

5.1 Introduction

Based on the predictions made in *Chapter 4*, I examined 3 post-translational modifications (PTMs): ubquitination, sumoylation, and serine-mediated phosphorylation. Additionally, I assessed the validity of the *in silico* predictions of potential adaptor proteins binding to the intracellular region of hMARCO. Since the bioinformatic methods used to predict the small linear motifs that adaptor proteins recognize are under-developed, a broad approach was taken to identify adaptor proteins by using column chromatography in conjunction with mass spectrometry to examine a

larger spectrum of potential interacting partners.

5.1.1 Rationale

Previous research suggests that the cytoplasmic domain of MARCO is essential for function, including the uptake of ligands (Kroos, 2008). I predicted that functional modifications, such as PTMs or the binding of adaptor proteins, may occur in this domain and could contribute to the signalling required to internalize bound ligands.

5.2 The hMARCO-T4-construct system: production & testing

5.2.1 The hMARCO-T4-construct system

Before being able to test *in silico* predictions *in vitro*, I developed a system that would allow me to isolate my studies to the cytoplasmic region of hMARCO. To do this, the hMARCO-T4-construct was created as described below, which consists of only these domains of the hMARCO protein.

The creation of the hMARCO-T4-construct provided an opportunity to overcome limitations of using the full-length protein. First, I wanted to limit all identified modifications and interacting partners to the N-terminal of hMARCO. For example, there may have been background binding of extraneous proteins to the collagenous and SRCR domains if the full-length protein had been used. Additionally, at the time of production, there was not a suitable antibody specific for hMARCO which would allow for its identification via Western blotting, immunoprecipitation (IP),

or column chromatography. Cloning this synthetic construct afforded the opportunity to include epitopes such as the Myc-tag for which highly specific antibodies exist.

5.2.2 Cloning of the hMARCO-T4-construct

I produced a synthetic construct containing the cytoplasmic and transmembrane domains of hMARCO connected to the T4-Foldon oligonucleotide based on that described in (Kleber *et al.*, 2008). The T4-Foldon domain was chosen because of its characteristic trimeric structure which I predicted would allow the hMARCO-T4-construct itself to form a trimer, ensuring that the N-terminal domains of hMARCO be representative of physiological conditions.

The sequence of the T4-Foldon oligonucleotide was based on that used by Kleber *et al.* (2008) with additional polyhistidine-tag, streptavidin, and Myc-tags added to the C-terminal end. The oligonucleotide was ordered from Integrated DNA Technologies and subsequently resuspended in TE buffer (10mM Tris (pH 8), 1mM ethylenediaminetracetic acid (EDTA)). The plasmid was transformed into competent DH5α *Escherichia coli* cells and cut out of the pIDTSMART plasmid via restriction digest (*Fig. 5.1*). The pieces were subsequently purified and the T4-Foldon ligated into pcDNA3.1 and verified by restriction digest (*Fig. 5.2*). Concurrently, the cytoplasmic and transmembrane domains of hMARCO were amplified via polymerase chain reaction (PCR) with specific primers. This PCR product was cut with restriction enzymes, purified, and ligated into cut and purified T4-Foldon-pcDNA3.1, and further verified by restriction digest (*Fig. 5.3*). From this point forward the resulting product is referred to as the hMARCO-T4-construct.

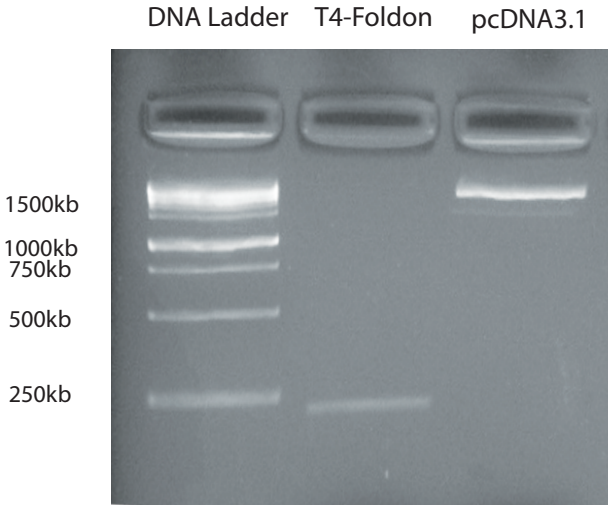


Figure 5.1: **Restriction digest to obtain the T4-Foldon and cut pcDNA3.1.** DNA isolated from transformed *Escherichia coli* cells were cut with appropriate restriction enzymes to obtain the T4-Foldon oligonucleotide from the pIDTSMART vector and, secondly, a linearized pcDNA3.1 such that the T4-Foldon could later be inserted.

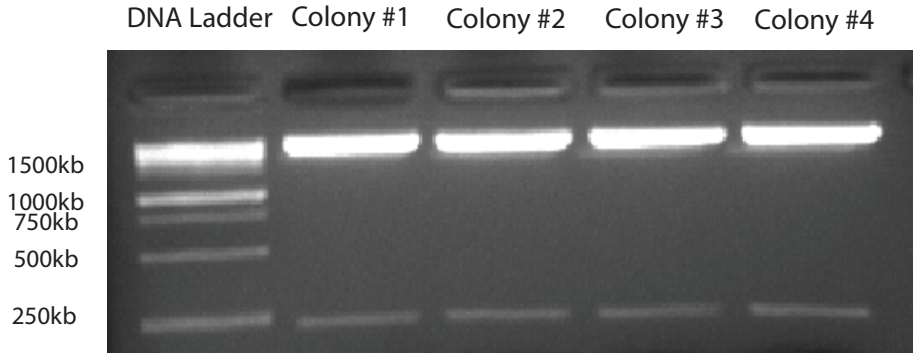


Figure 5.2: **Restriction digest to verify the successful ligation of the T4-Foldon oligonucleotide into pcDNA3.1.** Four colonies were isolated from agar plates streaked with varying concentrations of the T4-Foldon insert, pcDNA3.1 vector along with ligation buffer and T4 ligase. All 4 colonies chosen present successful ligation reactions.

5.2.3 The hMARCO-T4-construct is expressed in transiently transfected HEK293T cells

The hMARCO-T4-construct was transiently transfected into HEK293T cells and expression was examined via Western blot by probing with the 9E10 mouse anti-human Myc antibody. Expression relative to 2 known Myc-tagged proteins is displayed in *Fig. 5.4*.

5.2.4 Full-length hMARCO and the hMARCO-T4-construct are expressed in transiently transfected HEK293T cells as a trimeric protein

It is important that the hMARCO-T4-construct express the N-terminal domains of hMARCO as close to physiological conditions as possible in order to better simulate the *in vivo* function of the protein. The protocol for SDS-PAGE usually includes the addition of an aqueous reducing agent, dithiothreitol (DTT), accompanied with bringing the proteins to a near boil. DTT, along with SDS, denature proteins by breaking disulfide bonds and binding with the heated and denatured proteins, preventing the reformation of tertiary structures (Shapiro *et al.*, 1966). By omitting DTT, I hypothesized that a proportion of the protein would keep its tertiary structure even in the presence of a small amount of SDS in the running gel.

To test this, I transiently transfected hMARCO and the hMARCO-T4-construct into a panel of cell lines under a variety of conditions, including different lysis procedures, to determine whether any allowed for the maintenance of a trimer (*Fig. 5.5*). In samples that were not supplemented with DTT, distinct bands were visible at 150kD, the approximate size of the trimeric hMARCO protein. When similar conditions were applied to cell lines transiently transfected with the hMARCO-T4-construct, a band approximately triple the size of the construct monomer was observed when DTT was not included (*Fig. 5.5*).

Based on these results, samples were not reduced as part of further experimental protocols.

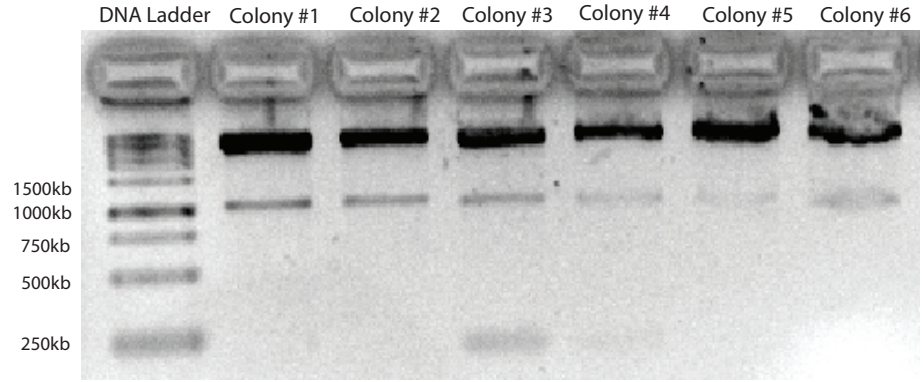


Figure 5.3: **Restriction digest to verify the successful ligation of the N-terminal domains of hMARCO into T4-Foldon-pcDNA3.1.** The cytoplasmic and transmembrane domains of hMARCO were isolated via PCR, purified, and ligated into the T4-Foldon-pcDNA3.1 vector to insert the hMARCO-T4-construct in the pcDNA3.1 plasmid. Six colonies were examined with two, Colony #3 and Colony #4, having successfully integrated the hMARCO fragment.

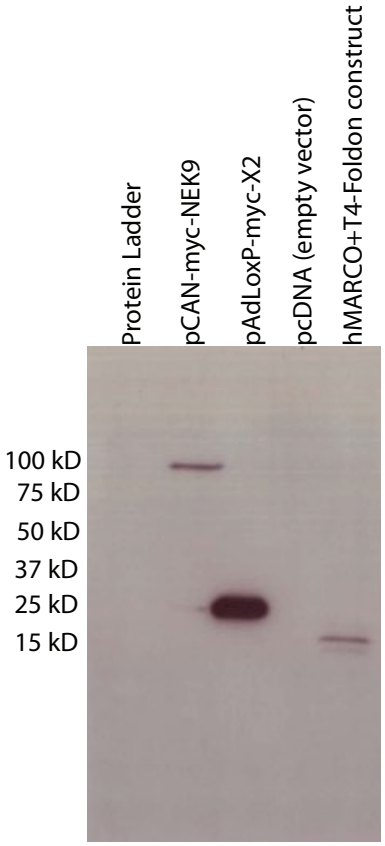


Figure 5.4: **Western blot showing the expression of hMARCO-T4-construct in HEK293T cells upon transient transfection.** The hMARCO-T4-construct (15 kDa) was probed with the 9E10 antibody in transiently transfected HEK293T cells. As positive controls for the affinity of 9E10, 2 proteins known to contain the Myc epitope were also transfected into HEK293T cells and their expression was simultaneously verified by Western blot.

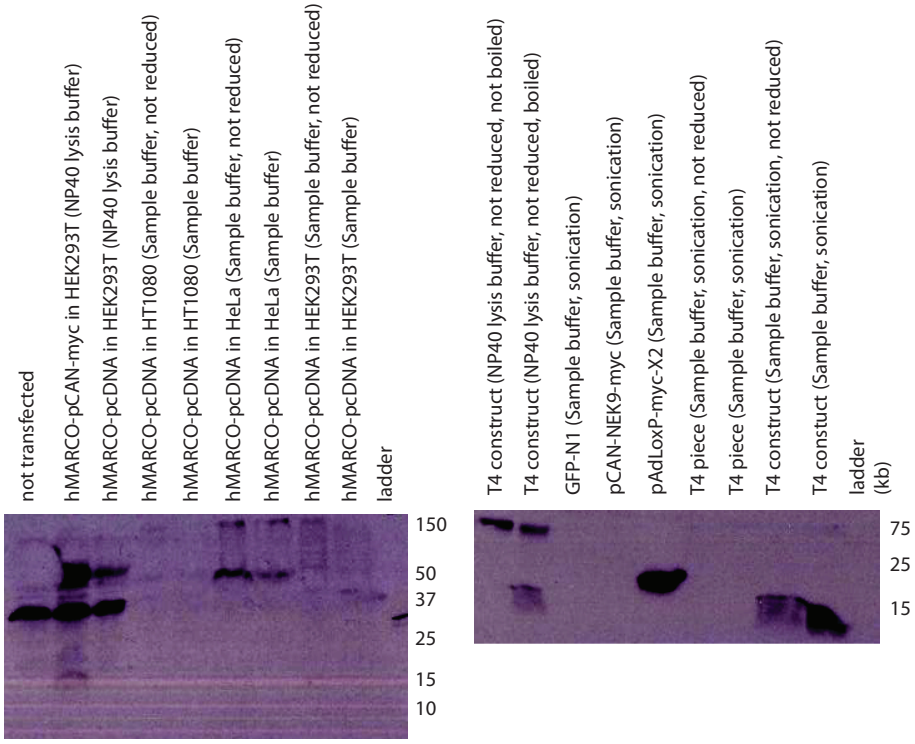


Figure 5.5: **hMARCO and the hMARCO-T4-construct are expressed as trimeric proteins.** A panel of cell lines and lysis protocols were tested to determine whether certain conditions allowed for the expression of some proportion of the protein as a trimer. Results indicate that, in conjunction with specific lysis methods, these proteins can remain trimeric in the absence of DTT.

5.2.5 The hMARCO-T4-construct is expressed on the surface of transiently transfected HEK293T cells

In order to establish whether the hMARCO-T4-construct is expressed on the cell surface, I performed immunofluorescene in conjunction with microscopy to visualize the cellular location of the construct. The geography of the protein is important when we consider the study of PTMs and potential adaptor binding proteins as I would like the hMARCO-T4-construct to be located as similarly to the wildtype hMARCO

protein as possible. HEK293T cells transiently transfected with the hMARCO-T4-construct were stained with phalloidin and a fluorescent antibody specific to its Myc epitope. Images taken with a confocal microscope indicate that a large proportion of the construct is localized to the cellular membrane (*Fig. 5.6*).

5.3 Assessing *in silico* predictions using protein immunoprecipitation

5.3.1 Rationale

Immunoprecipitation (IP) is a method of protein isolation via conjugation of an antibody specific to a protein of interest to protein A or G sepharose beads. In this work, I used the 9E10 mouse anti-human Myc antibody to immunoprecipitate the hMARCO-T4-construct and an artificially Myc-tagged hMARCO protein. Once isolated, I probed the immunoprecipitated material with antibodies specific to the PTMs of interest via Western blotting.

Additionally, surplus immunoprecipitated material was run on a gradient gel, isolated, and analysed via mass spectrometry in order to further identify PTMs and adaptor binding proteins.

5.3.2 Myc-hMARCO can be successfully immunoprecipitated

Both the hMARCO-T4-construct and the Myc-hMARCO protein were independently immunoprecipitated using the protocol described in *2.2.9*. The immunoprecipitated material was analysed via Western blotting using the rabbit anti-human MARCO and rabbit anti-human His antibodies to detect Myc-hMARCO and the

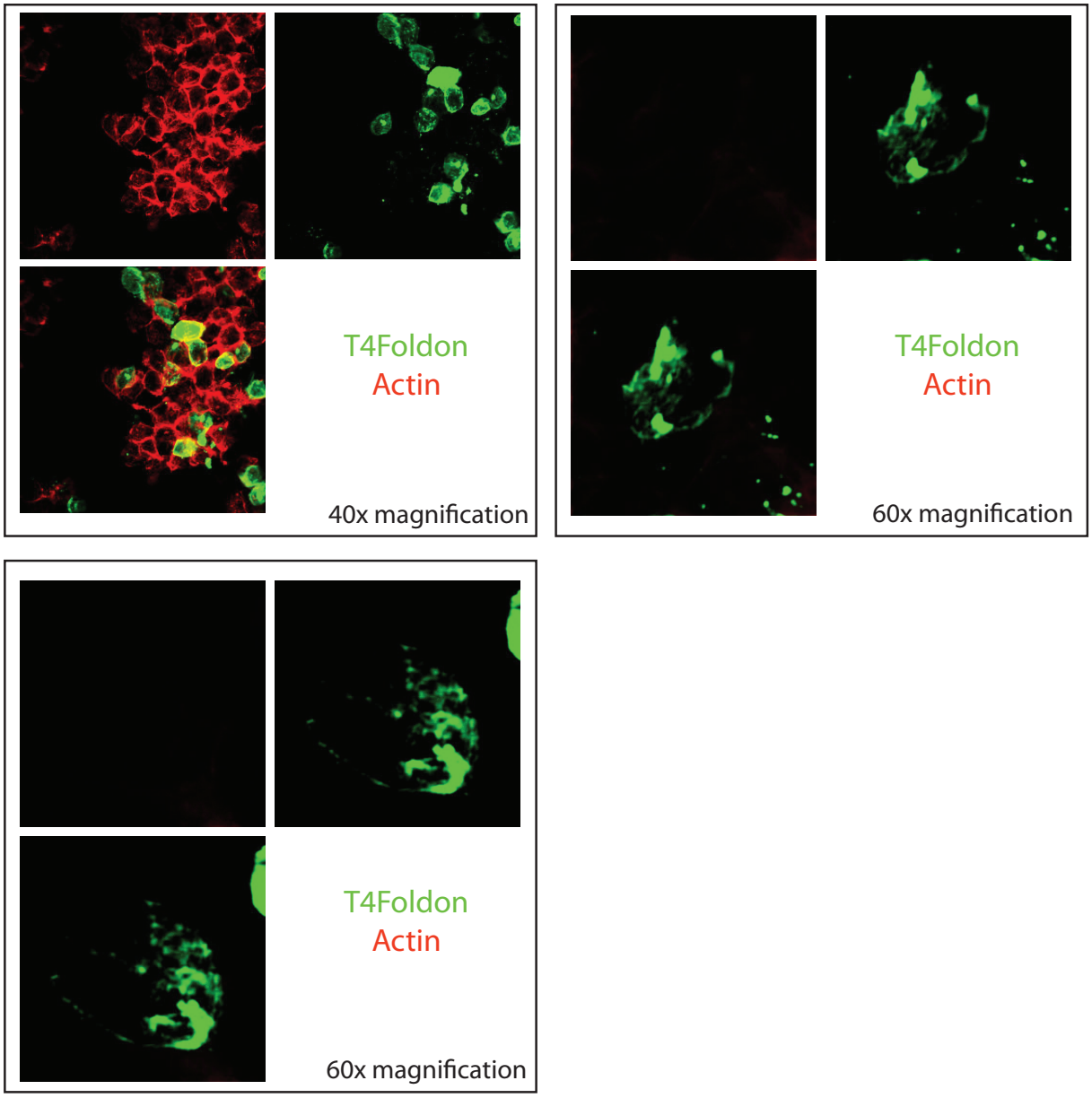


Figure 5.6: **Immunofluorescence microscopy of HEK293T cells transiently transfected with hMARCO-T4-construct demonstrate that the majority of the protein localizes to the cellular membrane.** HEK293T cells were stained in suspension with antibody specific for the Myc epitope of the hMARCO-T4-construct using an Alexa 488 (Fluorescein isothiocyanate, green) secondary antibody in addition to phalloidin (red) to determine its cellular geography. Staining indicates that most of the protein was localized to the cellular membrane.

hMARCO-T4-construct, respectively. Three samples were run per gel; the first, cell lysate containing the protein transiently transfected into the cells without further isolation. The second and third lanes represent samples from the immunoprecipitation: the second lane consists of supernatant from the IP procedure and the third lane the resultant isolated material. The non-immunoprecipitated samples acted as a positive control for the blotting and antibody detection procedures and the supernatant acts as a control to ensure that the IP procedure was successful. Immunoprecipitations were performed over a panel of 3 lysis buffers in order to increase the probability of success.

Results show that Myc-hMARCO can be successfully immunoprecipitated (*Fig. 5.7*) as indicated by its presence in the transiently transfected and immunoprecipitated samples and lower abundance in the immunoprecipitated supernatant. Additionally, it appears that both RIPA and Triton X-100 lysis buffers are better able to isolate hMARCO than NP40.

Unfortunately, I was unable to immunoprecipitate the hMARCO-T4-construct. There was no detection of the construct with the rabbit anti-human His antibody in either the transiently-transfected or immunoprecipitated samples; however, protein with a known His-tag epitope was successfully detected, indicating that the antibody was functional (data not shown). It is possible that hMARCO-T4-construct's His epitope is not exposed on the surface of the protein or that it is hidden when the protein is trimerized and, thus, cannot be detected by the antibody. Additionally, it is possible that when the Myc epitope of the construct is only visible when the protein is denatured.

As previously described, the hMARCO-T4-construct was originally developed

for 2 reasons: to isolate investigations to hMARCO's cytoplasmic domain, and

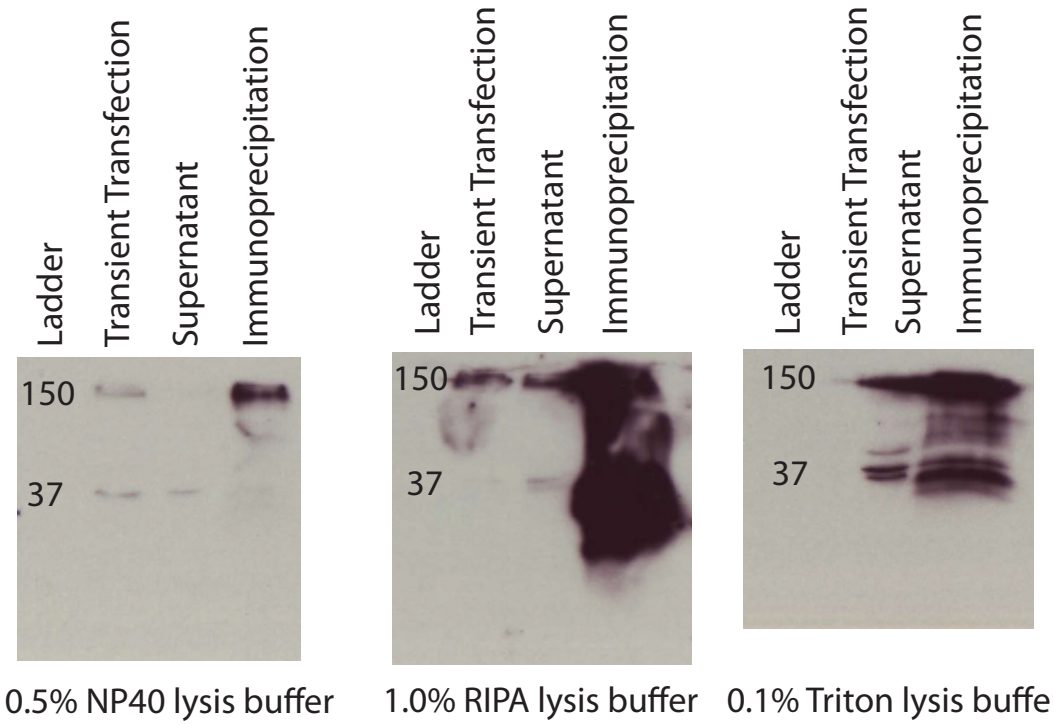


Figure 5.7: **hMARCO can be successfully immunoprecipitated using the 9E10 mouse anti-human Myc antibody.** PVDF membranes were probed with rabbit anti-human MARCO antibodies. A panel of 3 lysis buffers were used in the IP procedure. In addition to the immunoprecipitated samples, transient transfected samples were included to control for the blotting and probing procedures. Further, the supernatant from the IP was probed as a negative control to ensure the stringency of the protocol. Photographic films were developed after a 2 minute exposure to ECL-activated membranes.

to create a construct that could be detected with commercially verified antibodies because at the time there was not an antibody specific for hMARCO. Since the onset of this project, colleagues in the Bowdish laboratory have developed a rabbit anti-human MARCO antibody specific to hMARCO. Because the cytoplasmic domain of MARCO is its only intracellular domain, instead of optimizing the IP protocol or altering the sequence of the hMARCO-T4-construct to include additional epitopes, I chose to assess the bioinformatic predictions of PTMs and adaptor binding proteins

using the full-length hMARCO protein with the understanding that this may result in higher background binding in domains other than the cytoplasmic.

5.3.3 Analysis of post-translational modifications

In order to test predictions of ubiquitination, sumoylation, and phosphorylation, gels were conducted and probed with antibodies specific for these modifications. As with those used in *5.3.2*, gels were run with transiently transfected controls expressing hMARCO, immunoprecipitated samples specific for hMARCO, and a sample of the immunoprecipitated supernatant. Additionally, 3 lysis buffers - 0.5% NP40, 1% RIPA, and 0.1% Triton X-100 - were used as part of independent IPs.

Ubiquitination

To detect ubiquitination, a rabbit anti-human ubiquitin antibody was used in conjunction with N-ethylmaleimide, an agent that blocks deubiquitinating enzymes (Alwan and van Leeuwen, 2006). *Figure 5.8a* shows that while ubiquitinated proteins were detected in the whole cell lysates, no ubiquitination was found in the samples specific for hMARCO protein. This indicated that, contrary to *in silico* predictions, hMARCO is not ubiquitinated under the conditions tested.

Sumoylation

A rabbit anti-human sumoylation antibody specific to SUMO-1 was used to test for the predicted site of sumoylation in the cytoplasmic tail of hMARCO. Similar to ubiquitination, sumoylated proteins were detected in the transiently transfected

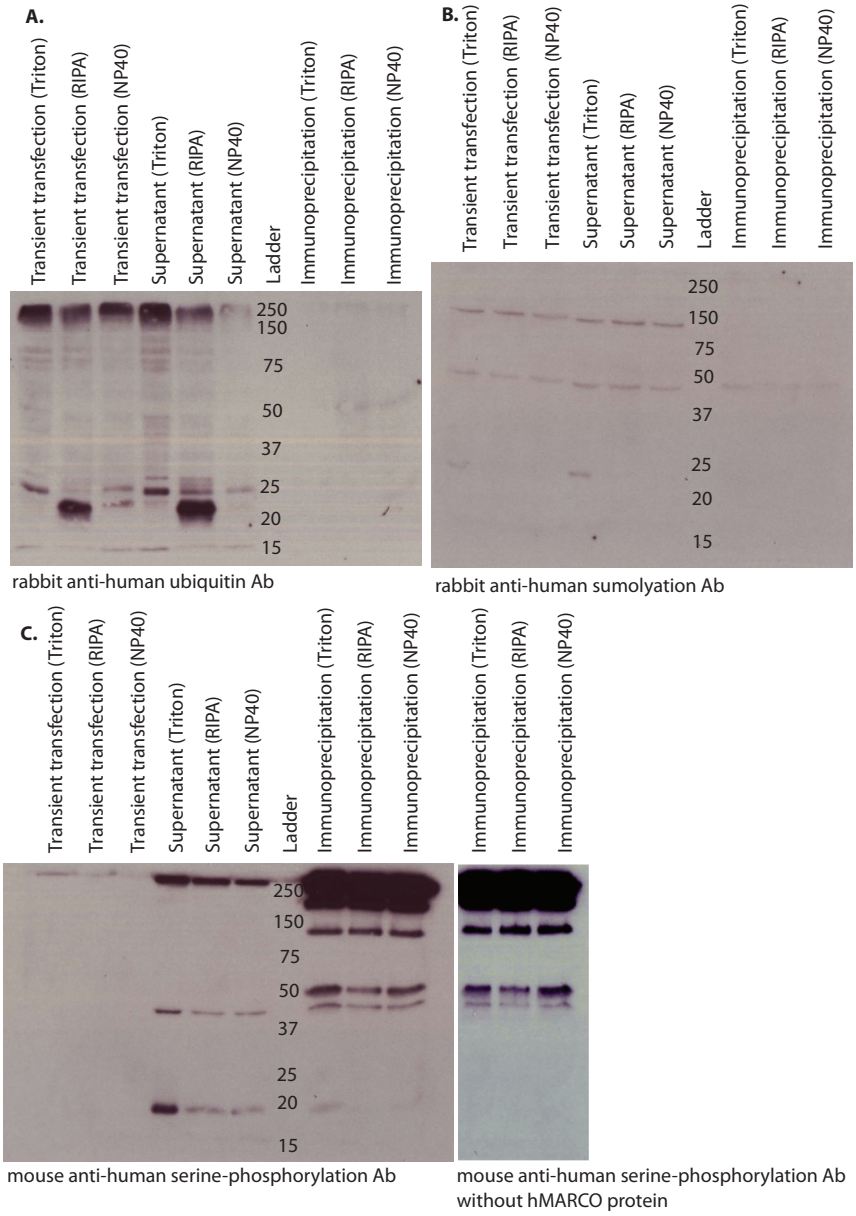


Figure 5.8: **Immunoprecipitations to test the predictions of post-translational modifications to hMARCO.** PVDF membranes were probed with antibodies designed to recognize ubiquitination, sumoylation, and serine-phosphorylation. A panel of 3 lysis buffers were used in the IP procedure. Photographic films were developed after a 30 minute exposure to ECL-activated membranes.

and IP supernatant samples, but not in the immunoprecipitated material (*Fig. 5.8b*) save for a faint band at 50kDa. This band may indeed be hMARCO but because similar bands were detected in other samples, further analyses must be accomplished. Additionally, it is possible, that hMARCO is sumoylated by SUMO-2 or -3, or in conditions that were not examined, such as in the presence of a co-receptor or bound ligand.

Serine Phosphorylation

The possibility of serine phosphorylation was examined by use of a mouse anti-human serine phosphorylation antibody. This reagent was not ideal as it was produced in the same species as the 9E10 Myc antibody that was used to IP Myc-hMARCO. As a result, the anti-mouse secondary antibody used to probe the PVDF membrane cross-reacted with the murine antibody present in the immunoprecipitated material (*Fig. 5.8c*). Regardless, the antibody was able to detect phosphorylated protein in the control cell lysates. Additionally, the band patterns of the hMARCO immunoprecipitated samples match identically to additional immunoprecipitated samples prepared in the absence of hMARCO, indicating that these bands are not specific to hMARCO and, instead, are likely fragments of the mouse anti-human Myc antibody. These results, however, are not conclusive; because hMARCO is approximately 50kDa as a monomer and 150kDa as a trimer, it is possible that phosphorylated protein is present on this gel and masked by other bands.

By use of Mass spectrometry

Surplus immunoprecipitated material, along with the necessary controls, were run on a pre-cast gradient gel and stained with Coommassie Blue Silver (*Fig. 5.9*). Of the bands present in the immunoprecipitated material, some were hypothesized to be hMARCO based on size (*Fig. 5.9, yellow boxes*). Of the 3 candidate bands, 2 were identified as hMARCO through mass spectrometry (*Table 5.1*); however, neither of these samples contained peptides from the intracellular region which precluded analysis of PTMs.

5.3.4 Analysis of adaptor binding partner predictions

In addition to using immunoprecipitated material to test for PTMs, it was also used to search for potential binding proteins. As previously described, immunoprecipitated material specific to hMARCO was stained with Coommassie Blue Silver and analysed via a pre-cast SDS-PAGE gel. Protein bands present in the hMARCO immunoprecipitated material at the approximate size of hMARCO were picked (*Fig. 5.9, yellow boxes*) in addition to those which were unique to the hMARCO samples when compared to the non-transfected controls hypothesized to be adaptor binding proteins (*Fig. 5.9, red boxes*). Results are summarized in *Table 5.1*.

As discussed in section *5.3.3*, 3 bands were chosen as potential representatives of hMARCO. Band 1 was uniquely identified as hMARCO with high probability. Band 2 was classified, also with high probability, as 3 proteins: hMARCO, actin gamma 1 (ACTG1), and RNA binding motif protein, X-linked (RMBX). ACTG1 is a cytoplasmic actin, of which specific isoforms are associated with autosomal dominant hearing loss (Van Wijk *et al.*, 2003). RMBX is a pre-mRNA splicing protein associated with

proper germ cell development (Elliott, 2004).

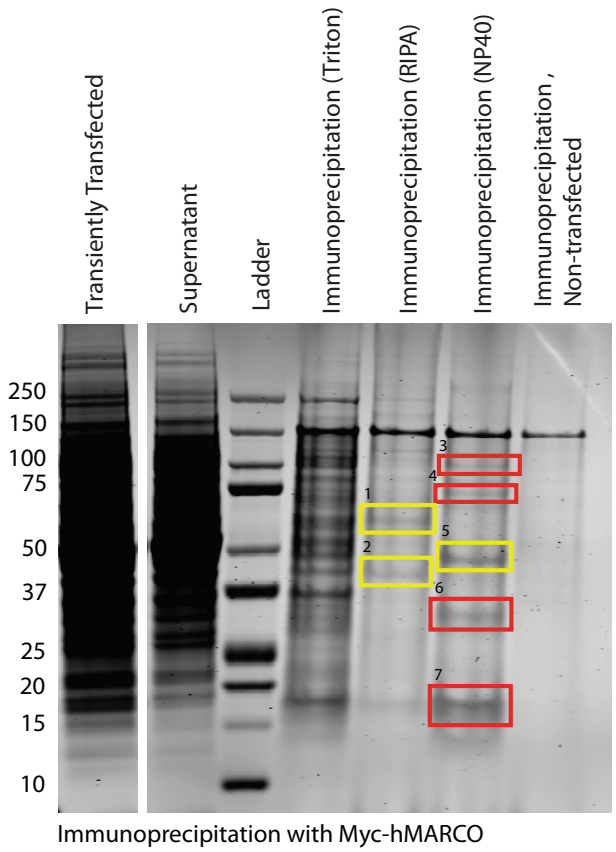


Figure 5.9: **An SDS-PAGE gradient gel to determine post-translational modifications of hMARCO in addition to potential adaptor binding proteins.** The gel was stained with Coommassie Blue Silver and bands distinct from the non-transfected control (boxed) were cut and analysed by mass spectrometry. Some bands chosen were hypothesized as hMARCO based on size (*yellow boxes*); others were analysed as potential adaptor binding proteins (*red boxes*). Band identification numbers correspond to those used in *Table 5.1*.

Of the other bands analysed, only 1 other returned significant results. Band 6 was identified as 3 proteins. Soluble carrier family 25, member 5 (SLC25A5) is an ADP transporter responsible for translocating the molecule from the mitochondria into the cytosolic space (Yang *et al.*, 2009). PHB2 is a protein associated with mitochondrial metabolism (Coates, 2001). Lastly, RPL7A encodes a gene that is part of

the 60S RNA subunit (Fried and Colombo, 1992).

These results are likely not true interacting proteins but rather false-positives associated with using the full-length hMARCO protein at artificially high levels. The coiled-coil and collagenous domains are common elements of other proteins and may result in interactions *in vitro* that are not seen when hMARCO is expressed at normal *in vivo* levels. RMBX, for example, normally binds another coiled-coil containing protein (Rual *et al.*, 2005). Additionally, since the collagenous and SRCR domains are defined by their ability to bind numerous polyanions, it was expected that non-specific interactions would occur in this region. ACTG1 contains a small cluster of negative amino acids at its N-terminus, raising the possibility that its appearance in the immunoprecipitated material is due to a charge-charge interaction between this region and the ligand binding domains of hMARCO. Therefore, mass spectrometry analyses of immunoprecipitated hMARCO material was unable to successfully predict adaptor binding partners of hMARCO.

5.4 Assessing *in silico* predictions using column chromatography

5.4.1 Rationale

By utilizing the polyhistidine epitope present as part of the hMARCO-T4-construct, a nickel column was used to isolate the construct as well as any potential adaptor binding partners from a collection of whole cell lysate. Column chromatography was chosen due to its ability to purify a protein of interest from a heterogenous solution.

Table 5.1: **Summary of mass spectrometry results from the immunoprecipitated hMARCO samples.** Band identifiers correspond with those in *Figure 5.9.* Significant results are summarized below. Results were considered significant if the log expectation was greater than -10, percent coverage greater than 5%, and the molecular size of the protein matched that of the band sampled. Results of keratin or histone proteins, which are common false positives in mass spectrometry analyses, are omitted. **log(e)** represents the log expectation that the result be by chance. **%** is the percentage of the protein that is covered by peptides recovered from the mass spectrometry experiment. Full results can be seen in *Appendix B, Table 1.*

Band Identifier	**Protein name and description**	**log(e)**	**%**
1	hMARCO - cA-SR	-29.6	12.0
2	ACTG1 - actin gamma 1	-74.0	34.0
	RBMX - RNA binding motif protein, X-linked	-13.5	8.2
	hMARCO - cA-SR	-12.3	6.7
3	*no significant results*		
4	*no significant results*		
5	*no significant results*		
6	SLC25A5 - soluble carrier family 25, member 5	-32.4	15.0
	PHB2 - prohibitin 2	-23.6	12.0
	RPL7A - ribosomal protein L7a	-11.2	8.6
7	*no significant results*		

5.4.2 Column chromatography

Fast protein liquid chromatography (FPLC) was used in conjunction with a nickel column to separate the hMARCO-T4-construct, along with its interacting partners, from other material in the cell lysate from which they were amplified. A second separation with cell lysate from cells not transiently transfected with the construct

was also conducted to be used as a control. An ELISA was used to determine in which of the 200 resultant fractions the hMARCO-T4-construct was eluted from the column. It was determined that the construct eluted over approximately 15 fractions which were subsequently pooled and concentrated to a volume of 40μl.

SDS-PAGE pre-cast gradient gels were used to compare samples from the pooled fractions positive for the hMARCO-T4-construct to their corresponding fractions in the control run, which controlled for the presence of other proteins that may have eluted under the same conditions as the construct (*Fig. 5.10*).

5.4.3 Analysis of post-translational modifications

On the resultant gel, there was a strong band present at approximately 45kDa that was hypothesized to be the trimeric form of the hMARCO-T4-construct, and another at approximately 15kDa hypothesized to be the monomeric form (*Fig. 5.10, yellow boxes*). These bands were excised and analysed via mass spectrometry. Unfortunately, the hMARCO-T4-construct was not identified in these samples and, as such, no analysis of PTMs could be conducted. It is surprising that the construct could not be determined since a previous ELISA specific for the protein concluded its presence in these samples. The construct may have been hard to determine, however, since the cytoplasmic region of hMARCO was described as unfavorable for standard peptide degradation techniques.

5.4.4 Analysis of adaptor binding partners

In addition to these bands hypothesized to be the hMARCO-T4-construct itself, another band which differed from the non-transfected control was also analysed by

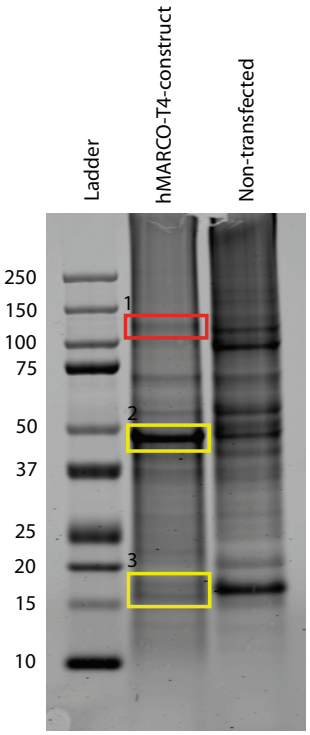


Figure 5.10: **An SDS-PAGE gradient gel to determine post-translational modifications of the hMARCO-T4-construct and potential interacting partners.** A gel was used to compare the pooled resultant fractions from the column chromatography that were positive for the hMARCO-T4-construct. The gel was stained with Coommassie Blue Silver and bands estimated to be the construct (*yellow boxes*) or potential binding partners (*red boxes*) were further analysed via mass spectrometry.

mass spectrometry as a potential adaptor binding protein (*Fig. 5.10, red box*). Of the 3 bands analysed, several proteins were discovered. SF3B3 is a subunit of splicing factor 3b which is involved in prespliceosome formation as part of the removal of introns during pre-mRNA processing (Igel *et al.*, 1998). TXNDC5 is a protein which is upregulated in hypoxic conditions and enduces apotosis (Chang *et al.*, 2011). NME1 encodes a nucleoside diphosphate kinase which catalyzes the exchange of phosphate groups of nucleoside diphosphates (Boissan and Lacombe, 2011). EIF5A, unique for its inclusion of a hypusine amino acid in its sequence, is responsible for the elongation of mRNA during translation (Saini *et al.*, 2009). C20orf27 is a protein of

Table 5.2: **Summary of mass spectrometry results from the hMARCO-T4-construct column chromatography.** Band identifiers correspond with those in *Figure 5.10*. Significant results are summarized below. Results were considered significant if the log expectation was greater than -10, percent coverage greater than 5%, and the molecular size of the protein matched that of the band sampled. Results of keratin or histone proteins, which are common false positives in mass spectrometry analyses, are omitted. **log(e)** represents the log expectation that the result be by chance. **%** is the percentage of the protein that is covered by peptides recovered from the mass spectrometry experiment. Full results can be seen in *Appendix B, Table 2*.

Band Identifier	**Protein name and description**	**log(e)**	**%**
1	SF3B3 - splicing factor 3b, subunit 3	-102.3	14
2	TXNDC5 - thioredoxin domain containing 5	-20.9	9.3
3	NME1 - NME/NM23 nucleoside diphosphate kinase 1	-75.9	8.6
	EIF5A - eukaryotic translation initiation factor 5A	-40.0	33.0
	C20orf27 - Uncharacterized protein C20orf27	-22.3	29.0
	RPS18 - ribosomal protein S18	-12.1	16.0

unknown function located on chromosome 20. Lastly, RPS18 encodes a gene that is part of the 40S RNA subunit (Yu *et al.*, 2009).

These results are likely not true interacting proteins but rather false-positives associated with the use of the nickel column. The non-transfected control lane (*Fig. 5.10*) had an abundance of protein material which made it difficult to choose proteins unique to the hMARCO-T4-construct for this analysis. However, the prediciton of C20orf27, an unidentified protein, may be a line of further study. Even though a similar band was not identified in the hMARCO immunoprecipitated material, it is interesting that a completely novel protein be identified in this fashion.

5.5 Conclusions

Two independent methods were employed in order to determine whether the predictions of PTMs in the intracellular region of hMARCO made in *Chapter 4* could be illustrated *in vitro*. Using immunoprecipitated material, predictions of ubiquitination, sumoylation, and serine phosphorylation could not be resolved via Western blotting or mass spectrometry analyses.

Additionally, *in silico* predictions were made concerning possible protein-protein interactions in the cytoplasmic domain of hMARCO. Again, 2 independent methods were used to study these predictions *in vitro*. First, immunoprecipitating the full-length protein and analysing proteins of interest in the resulting solution did not discover any potential adaptor binding proteins of interest. Second, column chromatography with the hMARCO-T4-construct did not result in any proteins that are likely to interact with hMARCO *in vivo*.

Chapter 6

Discussion

Since their discovery in 1979, Scavenger receptors (SRs) have been defined by their ability to 'scavenge' modified LDL from their environment for internalization and subsequent degradation (Goldstein *et al.*, 1979). As more proteins sharing this description were discovered, the SRs came to represent a polyphyletic group of receptors with varying domain architectures and protein structures that appear to have independently evolved similar functions (for example, although CD36, a class B SR, also binds modified lipids, permutation tests show that it is unrelated to SRAI (*Appendix A, Table 1*). This prompted the introduction of subclasses to group structurally similar proteins (Krieger, 1997). However, even within the class A subclass, there is considerable variability. MARCO, for example, can bind oxLDL (Plüddemann *et al.*, 2009), whereas SRAI can bind both oxLDL and acLDL (Suzuki *et al.*, 1997) and SCARA5 can bind neither (Jiang *et al.*, 2006). Structurally, the cA-SRs differ at their C-terminal region and in the lengths of their other domains (*Fig. 3.1*). There is little justification for grouping the cA-SRs together based on the original definition of ligand binding unless there is an evolutionary relationship amongst the members.

To uncover an evolutionary link between the cA-SRs, I first needed to definitively characterize the domain architecture of these proteins. Domain boundaries had been previously defined for the individual members of the cA-SRs; however, this was often done in comparison to SRAI and were not based on current tools. My findings (*Fig. 3.1*, *Table 3.3*) suggest that there are 4 domains - cytoplasmic, transmembrane, α-helical, and collagenous - shared by all members of the cA-SRs. While the lengths and consistency of the cytoplasmic and transmembrane domains remain mostly fixed, the α-helical and collagenous domains vary in length across the receptors in a manner consistent with the possibility of repeats brought about by recombination or duplication events (Andrade *et al.*, 2001). In contrast, the fifth terminal domain differs or is absent in the cA-SRs. While SRAI, MARCO, and SCARA5 share a SRCR domain at their terminus, SCARA4 possesses a C-type lectin domain and SCARA3 terminates at its collagenous region. The SRCR and C-type lectin domains are both able to recognize pathogens (Drickamer, 1999; Martínez *et al.*, 2011), suggesting that the radiation in this region may be due to a domain swapping event that may have allowed for the diversification of host pathogen recognition.

Data mining was used to identify known and novel cA-SRs in publicly available databases. Conservation of these proteins across vertebrate species was examined via phylogenetics. No cA-SRs were identified in available invertebrate genomes, implying that although the individual domains that make up these receptors - specifically the SRCR and C-type lectin domains - are ancient, the modern cA-SR domain architecture likely arose after the divergence of vertebrates from invertebrate species. Using these sequences, the relationships between the 5 members of the cA-SRs were analysed.

To determine a shared evolutionary ancestry amongst all 5 members of the cA-SRs, permutation tests were performed using the representative *Homo sapiens* protein sequences, which revealed significant sequence similarity between all proteins (*Table 3.1*). Additionally, notable motifs shared amongst all or most receptors were identified (*Fig. 3.6*), giving definitive reason for these proteins to be classified as a protein family.

Phylogenetic analyses allowed me to hypothesize regarding the evolutionary history of this protein family. First, analyses presented in *Figs. 3.3, 3.5, 3.6, & 3.7* indicate that SRAI and SCARA5 are most closely related to each other than to the other cA-SRs. This finding is further supported in the fact that the highest amount of sequence similarity is shared between SRAI and SCARA5 (*Table 3.1*). This is not surprising given what is known biologically about these 2 proteins. Although little research has been completed on SCARA5, it is known that both it and SRAI bind Gram-positive and -negative bacteria (Jiang *et al.*, 2006; Suzuki *et al.*, 1997; Peiser *et al.*, 2000) and are both hypothesized to be involved in host defence (Jiang *et al.*, 2006; Tomokiyo *et al.*, 2002). Second, SCARA3 and SCARA4 were also identified as closely related proteins. Not only are their domain lengths similar (*Fig. 3.1*), but these proteins are also presented as an independent cluster in the phylogenetic analysis of all cA-SRs (*Fig. 3.7*). Although they are not well studied, from what we know these 2 proteins do not have common functions. From what little is known regarding SCARA4, this receptor appears to function in a similar fashion to the SRCR-containing cA-SRs by binding Gram-positive and -negative bacteria and being expressed on cells involved in host defence (Nakamura *et al.*, 2001; Selman *et al.*, 2008). In contrast, SCARA3 is expressed on fibroblasts and has been

proposed to protect against reactive oxygen species by binding and internalizing oxidative molecules (Han *et al.*, 1998; Martínez *et al.*, 2011). However, the lengths and general composition of SCARA3 and SCARA4 proteins are very similar as indicated by a shared percent identity of 26.6% across the full-lengths of their proteins (*Table 3.3*). Perhaps the differences in their biological functions are restricted to the presence of a C-terminal C-type lectin domain in SCARA4 and the potentially lost terminal domain in SCARA3.

Lastly, the phylogenetic position of MARCO is intermediate between the SRAI/SCARA5 and SCARA3/SCARA4 clusters. The phylogenetic evidence presented in *Figure 5* suggests that this protein sequence is most similar to SCARA3/SCARA4 with high posterior probabilities and bootstrap support. However, measurements of percent identity (*Table 3.3*) as well as functional evidence suggests that it is most similar to the other SRCR-containing receptors. For example, research conducted by Arredouani *et al.* demonstrates that both SRAI and MARCO are essential for clearance of bacteria and inert particles from the lungs (Arredouani *et al.*, 2004, 2006), indicating that, even though MARCO is more evolutionarily related to SCARA3 and SCARA4, MARCO is more functionally related to the SRCR-containing receptors. Further analysis of the exon gene structures of the cA-SRs or further functional analyses of all 5 members may help resolve this inconsistency.

These data support the hypothesis of a single ancestral cA-SR from which duplication events occurred allowing for the diversification of this group. I propose that 4 independent gene duplication events occurred allowing for the presence of 5 cA-SRs in vertebrate species. This common ancestor likely included most of the common

features of the cA-SRs including the transmembrane, α-helical, and collagenous domains, and may have also contained the SRCR domain shared by 3 of the 5 class A scavenger receptors. The most recent duplications are apparent in *Figs. 3.6 & 3.7*, in which a duplication event produced SRAI and SCARA5, and a separate duplication produced SCARA3 and SCARA4 followed by gain or loss of the C-type lectin domain. The placement of MARCO in the tree in *Fig. 3.4* supports MARCO diverging from the common ancestor prior to both of the above duplications, with support for a closer relationship with SCARA3 and SCARA4. Given this, it is likely that the ancestral protein for this family contained a SRCR domain, and the C-type lectin domain was the result of either a domain swap or domain fusion. It is not clear whether the SRCR domain was first replaced by a C-type lectin domain that was later lost by SCARA3, or whether the SRCR domain was lost first, followed by a domain fusion resulting in the C-type lectin of SCARA4. This change to the cA-SR sequence may account for the stronger functional relationship between the SRCR-containing receptors, even though MARCO is more evolutionarily related to SCARA3 and SCARA4.

Despite the broad, general definition that brought the 5 members of the cA-SRs into the same subclassification of proteins capable of recognizing modified lipoproteins, I have shown significant evidence that these 5 proteins are, indeed, a protein family. There is considerable evidence of a common origin for these proteins, which may in turn provide insight when performing functional studies on members of this family.

One member to which the results of these evolutionary analyses can be applied is MARCO. Using this work, I was able to conclude that there are specific conserved motifs in the MARCO proteins. By examining areas of conservation in this protein

across a panel of organisms, in addition to across the cA-SR family as a whole, I was able to identify motifs that are likely essential for function. For example, research by Brannstrom *et al.* demonstrated a region in MARCO necessary for ligand binding (Brännström *et al.*, 2002). My analyses found that this motif (RGRAEVYYSGT) (*Fig. 3.6, yellow box*) is highly conserved in MARCO SRCR domains across all species but, however, is not present in the SRCR domains of SRAI or SCARA5. This is consistent with the different ligand binding abilities of MARCO compared to SRAI and SCARA5.

Given my research question, however, the conserved motifs in the cytoplasmic domain of hMARCO were of more interest. A conserved lysine at position 8 was identified as a potential site of ubiquitination and sumoylation in the *Homo sapiens* sequence; a conserved serine at residue 14 was predicted to be a site of serine phosphorylation. There are additional areas of conservation in the cytoplasmic domain, such as the cluster of positive amino acids proximal to the transmembrane domain and negatively charged amino acids clustered at the N-terminus, that I think may be essential for the function of hMARCO due to their high level of conservation. Despite this conservation, however, there was no similarity to any predicted functional motifs when analysed via bioinformatic predictive softwares.

Of the *in silico* predictions of post-translational modifications in the intracellular region, none were verified with *in vitro* inquires. Immunoprecipitations of an artifically Myc-tagged hMARCO protein were completed and analysed by Western blotting with antibodies specific for these modifications. While no PTMs were observed, I think that a lack of PTMs in the intracellular region of hMARCO cannot be conclusively determined from this work alone. Because PTMs are transient and

often act as 'toggle-switches' for cellular activity, it is possible that the PTMs hypothesized to act on hMARCO only take affect when this protein is exposed to a bound ligand. The experiments completed as part of this research would not be effective in identifying these types of modifications. Second, recent research from the Bowdish laboratory has shown that hMARCO acts in conjunction with TLR2 and CD14 in its identification of pathogenic substances (Dorrington *et al., In review*). As such, PTMs of hMARCO may rely on the interactions with these co-receptors at the plasma membrane.

To further this research, I wish to outline 3 sets of experiments which I believe would provide additional information as to how conserved motifs in the cytoplasmic domain of hMARCO affect the function of the protein. First, an immunoprecipitation of cells expressing hMARCO, TLR2, and CD14 which have been subsequently exposed to a ligand specific for hMARCO would further the investigations of PTMs in the cytoplasmic domain of hMARCO. Not only would the presence of TLR2 and CD14 produce conditions closer to *in vivo*, but also the addition of a ligand would begin the signalling cascade that these PTMs are hypothsized to be associated with. Second, these immunoprecipitations would be interesting to study via a mass spectrometry approach. Similar to that stated above, the induction of a signalling cascade that would begin when hMARCO binds a ligand may induce protein-protein interactions that could have been missed with the mass spectrometry analysis I conducted as part of this work. The SDS-PAGE analyses presented could be compared to those completed in the presence of a pathogenic ligand to identify bands that differ between these 2 states, and, therefore, bands of interest. Lastly, I think mutagenesis experiments could provide instructive results. A panel of mutants could be developed, each

with independently mutated conserved motifs, including those predicted to be sites of PTMs, in the cytoplasmic domain. With these mutants, various functional assays could be conducted and the results compared in order to elucidate the functions of these motifs. Colleagues in the Bowdish laboratory have developed assays to study the ligand binding, ligand internalization, and killing ability of cells expressing wild-type hMARCO which could be applied to these mutants.

Although the identification of PTMs and adaptor binding proteins resulted in a chapter of negative results, this research is still very important as a base line for future experiments. I have shown that PTMs do not occur on hMARCO in the absence of a ligand or in the absence of co-receptors such as TLR2 and CD14. The outlined subsequent experiments, may show that similar results are found in these conditions, or instead that the cytoplasmic domain confers a different role when its environment is altered.

Chapter 7

Conclusions

Through this work, I have been able to postulate the shared evolutionary history of all 5 members of the cA-SRs in addition to identifying which cA-SRs are present in which organisms. Sequence analyses of these receptors identified several conserved motifs amongst the cA-SRs, specifically a negative cluster of amino acids within the cytoplasmic domains of all members of this family. Additional analysis of the cytoplasmic region of MARCO has identified 3 sites of potential post-translational modifications, but was unsuccessful in predicting the presence of any protein-protein interactions. Further *in vitro* analyses were unable to identify these predicted post-translational modifications in the intracellular region when examined both by Western blotting and mass spectrometry techniques. Similarly, the existence of adaptor binding proteins in the absence of ligand was inconclusive; however, further studies investigating adaptor protein binding in the context of ligand binding or co-receptors are warranted.

Appendix A

Table 1: **Permutation test scores between the class A Scavenger receptors and a representative class B Scavenger receptor, CD36.** Permutation tests were measured using the PRSS algorithm (part of the FASTA package). The probabilities displayed are the probabilities that these receptors share sequence similarity with each other by chance.

cA-SR	Permutation test score against CD36
SRAI	0.81702
MARCO	0.30675
SCARA3	0.92749
SCARA4	0.052749
SCARA5	0.35774

Appendix B

Table 1: **Raw mass spectrometry results from hMARCO immunoprecipitations. Rank** represents the relative position of the result when ranked by log(e). **log(e)** is the base -10 log of probability of seeing the protein in the sample. **log(I)** is the base -10 log of the sum of ion intensities of the samples. %/% is the amino acid coverage of the protein and of that predicted to be able to be identified with the digest used. # is the number of peptide sequences identified. **total** is the number of mass spectra identified to this protein. **Mr** is the molecular mass of the protein in kDa.

Band	rank	log(e)	log(I)	%/%	#	total	Mr	Accession
1	1	-29.6	3.34	12/16	5.5	6.5	52.6	ENSP00000318916 MARCO, macrophage receptor with collagenous structure (Source: HGNC 6895)
2	1	-74	4.09	34/38	9	17	41.8	ENSP00000331514 ACTG1, actin, gamma 1 [Source: HGNC 144]
2	2	-66.6	4.08	4.3/5	1	2	41.7	ENSP00000349960 ACTB, actin, beta [Source: HGNC 132]

Continued on next page

Table 1 – *Continued from previous page*

Band	rank	log(e)	log(I)	%/%	#	total	Mr	Accession
2	3	-26.1	3.49	4.0/5	1	1	42	ENSP00000416706 ACTBL2, actin, beta-like 2 [Source: HGNC 17780]
2	4	-15.7	3.43	5.1/9	3	5.5	66	ENSP00000252244 KRT1, keratin 1 [Source: HGNC 6412]
2	5	-13.5	3.28	8.2/12	3.5	3	42.3	ENSP00000359645 RBMX, RNA binding motif protein, X-linked [Source: HGNC 9910]
2	6	-13.4	3.53	15/19	3	6	24.4	TRYP PIG Trypsin; EC 3.4.21.4
2	7	-12.3	3.21	6.7/9	3	3	52.6	ENSP00000318916 Macrophage receptor MARCO (Macrophage receptor with collagenous structure). [Source: SWISSPROT (Q9UEW3)]
2	8	-9.2	2.83	9.0/13	2	2	32	ENSP00000338095 HNRNPC, heterogeneous nuclear ribonucleoprotein C (C1/C2) [Source: HGNC 5035]
2	9	-6.7	2.9	1.7/2	1	1	65.4	ENSP00000310861 KRT2, keratin 2 [Source: HGNC 6439]
2	10	-5.5	2.62	2.3/4	1	1	47.7	ENSP00000311430 RPL4, ribosomal protein L4 [Source: HGNC 10353]
2	11	-1.7	2.71	2.5/4	1	1	37.4	ENSP00000346694 HNRNPA2B1, heterogeneous nuclear ribonucleoprotein A2/B1 [Source: HGNC 5033]
2	12	-1.5	2.7	1.4/2	1	1	47.2	ENSP00000286096 JMJD5, No description
2	13	-1.4	2.7	2.0/3	1	1	46.1	ENSP00000346001 RPL3, ribosomal protein L3 [Source: HGNC 10332]
2	14	-1.4	2.42	1.9/3	1	1	58.8	ENSP00000269576 KRT10, keratin 10 [Source: HGNC 6413]
2	15	-1.3	2.69	1.2/2	1	1	61.4	ENSP00000386229 RGS14, regulator of G-protein signaling 14 [Source: HGNC 9996]
2	16	-1.1	2.43	1.7/2	1	1	64	ENSP00000350078 ARID5A, AT rich interactive domain 5A (MRF1-like) [Source: HGNC 17361]
3	1	-19	3.22	4.8/8	3	3	66	ENSP00000252244 KRT1, keratin 1 [Source: HGNC 6412]

Continued on next page

Table 1 – *Continued from previous page*

Band	rank	log(e)	log(I)	%/%	#	total	Mr	Accession
3	2	-16.9	2.85	3.8/7	3	3	112.8	ENSP00000295598 ATP1A1, ATPase, Na+/K+ transporting, alpha 1 polypeptide [Source: HGNC 799]
3	3	-13.6	3.07	3.4/4	2.5	2	65.4	ENSP00000310861 KRT2, keratin 2 [Source: HGNC 6439]
3	4	-8.4	3.14	3.9/5	2	3.5	55.9	ENSP00000407516 SLC3A2, solute carrier family 3 (activators of dibasic and neutral amino acid transport), member 2 [Source: HGNC 11026]
3	5	-8.2	3.5	11/14	2	5	24.4	TRYP PIG Trypsin; EC 3.4.21.4
3	6	-5.5	2.57	5.0/9	2	2	58.8	ENSP00000269576 KRT10, keratin 10 [Source: HGNC 6413]
3	7	-2.6	2.34	2.4/4	1	1	62	ENSP00000246662 KRT9, keratin 9 [Source: HGNC 6447]
3	8	-2.6	2.49	1.3/2	1	1	87.3	ENSP00000346120 DDX21, DEAD (Asp-Glu-Ala-Asp) box helicase 21 [Source: HGNC 2744]
4	1	-46.3	3.61	12/21	7	9	66	ENSP00000252244 KRT1, keratin 1 [Source: HGNC 6412]
4	2	-27	3.1	8.2/15	4.5	4.5	58.8	ENSP00000269576 KRT10, keratin 10 [Source: HGNC 6413]
4	3	-21.7	3.18	4.7/6	3	4	65.4	ENSP00000310861 KRT2, keratin 2 [Source: HGNC 6439]
4	4	-12.6	3.11	12/15	2	2	24.4	TRYP PIG Trypsin; EC 3.4.21.4
4	5	-5.5	3.1	2.7/4	2	2	62	ENSP00000246662 KRT9, keratin 9 [Source: HGNC 6447]
4	6	-1.9	2.37	1.7/2	1	1	73.2	ENSP00000382840 DDX3X, DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked [Source: HGNC 2745]
5	1	-35.5	3.3	8.5/14	5	6	66	ENSP00000252244 KRT1, keratin 1 [Source: HGNC 6412]
5	2	-20.6	2.84	6.8/12	3	3	58.8	ENSP00000269576 KRT10, keratin 10 [Source: HGNC 6413]

Continued on next page

Table 1 – *Continued from previous page*

Band	**rank**	**log(e)**	**log(I)**	**%/%**	**#**	**total**	**Mr**	**Accession**
5	3	-9.9	2.95	4.0/6	2.5	2.5	62	ENSP00000246662 KRT9, keratin 9 [Source: HGNC 6447]
5	4	-9.7	2.91	4.2/7	2	2	47.7	ENSP00000311430 RPL4, ribosomal protein L4 [Source: HGNC 10353]
5	5	-4.9	2.48	0.6/2	1	1	116.7	ENSP00000299366 CDH23
5	6	-2.5	2.63	2.4/3	1	1	42	ENSP00000355645 ACTA1, actin, alpha 1, skeletal muscle [Source: HGNC 129]
5	7	-2.3	2.48	8.2/10	1	1	24.4	TRYP PIG Trypsin; EC 3.4.21.4
hline 5	8	-2.3	2.56	0.4/1	1	1	216.7	ENSP00000296043 SHROOM3, shroom family member 3 [Source: HGNC 30422]
5	9	-2.1	2.65	0.7/1	1	1	97	ENSP00000362306 EIF2C4, eukaryotic translation initiation factor 2C, 4 [Source: HGNC 18424]
5	10	-1.5	2.56	9.9/20	1	1	11.8	ENSP00000427203 SLC1A3
5	11	-1.3	2.61	0.4/1	1	1	213.9	ENSP00000365427 CACNA1F, calcium channel, voltage-dependent, L type, alpha 1F subunit [Source: HGNC 1393]
6	1	-40.3	3.79	14/26	6	10	22.3	ENSP00000244534 HIST1H1D, histone cluster 1, H1d [Source: HGNC 4717]
6	2	-35.3	3.43	9.5/16	6	7	66	ENSP00000252244 KRT1, keratin 1 [Source: HGNC 6412]
6	3	-32.4	3.91	15/21	5	7	32.8	ENSP00000360671 SLC25A5, solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 5 [Source: HGNC 10991]
6	4	-23.6	3.43	12/15	4	5	33.3	ENSP00000441875 prohibitin 2 [Source: HGNC 30306]
6	5	-17	3.12	5.7/10	4	4.5	58.8	ENSP00000269576 KRT10, keratin 10 [Source: HGNC 6413]
6	6	-15.6	3.08	9.2/12	3.5	3	29.8	ENSP00000300408 PHB, prohibitin [Source: HGNC 8912]
6	7	-12.1	2.98	3.0/4	2	2	65.4	ENSP00000310861 KRT2, keratin 2 [Source: HGNC 6439]

Continued on next page

Table 1 – *Continued from previous page*

Band	**rank**	**log(e)**	**log(I)**	**%/%**	**#**	**total**	**Mr**	**Accession**
6	8	-11.2	3.09	8.6/15	3	3	30	ENSP00000361076 RPL7A, ribosomal protein L7a [Source: HGNC 10364]
6	9	-7.1	2.72	4.8/6	1	1	34	ENSP00000225665 SLC25A11, solute carrier family 25 (mitochondrial carrier; oxoglutarate carrier), member 11 [Source: HGNC 10981]
6	10	-6.9	3.77	2.7/3	1	1	32.8	ENSP00000370808 SLC25A6, solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6 [Source: HGNC 10992]
6	11	-6.9	2.88	4.0/6	2	2	62	ENSP00000246662 KRT9, keratin 9 [Source: HGNC 6447]
6	12	-5.5	3.09	12/15	2	2	24.4	TRYP PIG Trypsin; EC 3.4.21.4;
6	13	-4.9	2.86	3.0/4	2	2	39.9	ENSP00000383898 SLC25A3, solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 3 [Source: HGNC 10989]
6	14	-4.8	3	6.7/12	2	2	30.3	ENSP00000298468 VDAC2, voltage-dependent anion channel 2 [Source: HGNC 12672]
6	15	-4.3	2.48	5.6/15	1	1	28.7	ENSP00000369757 RPS6, ribosomal protein S6 [Source: HGNC 10429]
6	16	-3.7	2.69	3.9/6	1	1	28	ENSP00000378378 RPL8, ribosomal protein L8 [Source: HGNC 10368]
6	17	-3.4	2.43	8.3/13	1	1	16.5	ENSP00000420961 SFXN1
6	18	-3.2	2.56	4.4/7	1	1	34.3	ENSP00000359345 RPL5, ribosomal protein L5 [Source: HGNC 10360]
6	19	-2.1	2.84	2.7/4	1	1	29.9	ENSP00000346050 RPS3A, ribosomal protein S3A [Source: HGNC 10421]
6	20	-2	2.7	3.5/4	1	1	30.8	ENSP00000265333 VDAC1, voltage-dependent anion channel 1 [Source: HGNC 12669]
7	1	-26.6	3.3	6.4/11	4	4	66	ENSP00000252244 KRT1, keratin 1 [Source: HGNC 6412]

Continued on next page

Table 1 – *Continued from previous page*

Band	rank	log(e)	log(I)	%/%	#	total	Mr	Accession
7	2	-22.7	3.81	25/35	3	6	13.9	ENSP00000375736 HIST3H2BB, histone cluster 3, H2bb [Source: HGNC 20514]
7	3	-22.6	3.48	39/67	4.5	5.5	11.4	ENSP00000376669 HIST2H4A, histone cluster 2, H4a [Source: HGNC 4794]
7	4	-14.3	3.02	3.8/5	2	2	65.4	ENSP00000310861 KRT2, keratin 2 [Source: HGNC 6439]
7	5	-14.1	3.26	25/65	3	4	14.1	ENSP00000355656 HIST3H2A, histone cluster 3, H2a [Source: HGNC 20507]
7	6	-6.4	3.22	11/14	2	3	24.4	TRYP PIG Trypsin; EC 3.4.21.4
7	7	-3.6	2.39	3.1/5	1	1	58.8	ENSP00000269576 KRT10, keratin 10 [Source: HGNC 6413]
7	8	-3.1	3.05	0.3/0	1	1	301.2	ENSP00000343741 ATR, ataxia telangiectasia and Rad3 related [Source: HGNC 882]

Table 2: **Raw mass spectrometry results from T4-construct column chromatography. Rank** represents the relative position of the result when ranked by log(e). **log(e)** is the base -10 log of probability of seeing the protein in the sample. **log(I)** is the base -10 log of the sum of ion intensities of the samples. **%/%** is the amino acid coverage of the protein and of that predicted to be able to be identified with the digest used. **#** is the number of peptide sequences identified. **total** is the number of mass spectra identified to this protein. **Mr** is the molecular mass of the protein in kDa.

Band	rank	log(e)	log(I)	%/%	#	total	Mr	Accession
1	1	-102.3	4.03	14/22	14	25	135.5	ENSP00000305790 SF3B3, splicing factor 3b, subunit 3, 130kDa [Source: HGNC 10770]
1	2	-24.2	3.46	5.2/8	4	5.5	83.2	ENSP00000360609 HSP90AB1, heat shock protein 90kDa alpha (cytosolic), class B member 1 [Source: HGNC 5258]
1	3	-18.5	3.16	16/19	3.5	4	41.7	ENSP00000349960 ACTB, actin, beta [Source: HGNC 132]

Continued on next page

Table 2 – *Continued from previous page*

Band	rank	log(e)	log(I)	%/%	#	total	Mr	Accession
1	4	-15.2	3.35	5.4/9	3	5	66	ENSP00000252244 KRT1, keratin 1 [Source: HGNC 6412]
1	5	-11.2	3.08	0.8/1	1	1	84.6	ENSP00000216281 HSP90AA1, heat shock protein 90kDa alpha (cytosolic), class A member 1 [Source: HGNC 5253]
1	6	-11.2	3.02	23/29	2	2	12.8	ENSP00000428085 RPL30, ribosomal protein L30 [Source: HGNC 10333]
1	7	-6.6	2.69	4.9/7	2	2	49.9	ENSP00000369703 TUBB2A, tubulin, beta 2A class IIa [Source: HGNC 12412]
1	8	-4.5	3.11	8.2/10	1	2	24.4	TRYP PIG Trypsin; EC 3.4.21.4;
1	9	-2.5	2.34	2.3/4	1	1	54.2	ENSP00000276079 NONO, non-POU domain containing, octamer-binding [Source: HGNC 7871]
1	10	-2.1	2.69	1.7/2	1	1	66	ENSP00000367265 CKAP4, cytoskeleton-associated protein 4 [Source: HGNC 16991]
1	11	-1.9	2.75	0.7/1	1	1	86.3	ENSP00000356559 KIAA1614
1	12	-1.9	2.36	3.9/6	1	1	36	ENSP00000229239 GAPDH, glyceraldehyde-3-phosphate dehydrogenase [Source: HGNC 4141]
1	13	-1.6	2.64	1.4/2	1	1	79.4	ENSP00000448182 TBC1 domain family, member 15 [Source: HGNC 25694]
1	14	-1.5	2.49	3.3/4	1	1	36.7	ENSP00000395337 LDHA, lactate dehydrogenase A [Source: HGNC 6535]
1	15	-1.5	2.55	1.1/2	1	1	92.4	ENSP00000354125 eukaryotic translation initiation factor 3, subunit B [Source: HGNC 3280]
1	16	-1.4	2.69	1.7/2	1	1	64.6	ENSP00000445379 eukaryotic translation elongation factor 2 [Source: HGNC 3214]
1	17	-1.1	3.01	1.3/2	1	1	71.5	ENSP00000325448 KARS, lysyl-tRNA synthetase [Source: HGNC 6215]
2	1	-23.6	3.5	6.2/10	4	5	80.4	ENSP00000380033 DDX17, DEAD (Asp-Glu-Ala-Asp) box helicase 17 [Source: HGNC 2740]

Continued on next page

Table 2 – *Continued from previous page*

Band	**rank**	**log(e)**	**log(I)**	**%/%**	**#**	**total**	**Mr**	**Accession**
2	2	-20.9	3.23	9.3/11	4	4	47.6	ENSP00000369081 TXNDC5, thioredoxin domain containing 5 (endoplasmic reticulum) [Source: HGNC 21073]
2	3	-14.6	3.37	4.5/8	3	4	76.1	ENSP00000349748 SFPQ, splicing factor proline/glutamine-rich [Source: HGNC 10774]
2	4	-10.9	3.52	11/14	2	3	24.4	TRYP PIG Trypsin; EC 3.4.21.4;
2	5	-9.2	3.11	23/29	2	3	12.8	ENSP00000428085 RPL30, ribosomal protein L30 [Source: HGNC 10333]
2	6	-8.5	3.1	3.1/5	2	2	76.6	ENSP00000318195 NCL, nucleolin [Source: HGNC 7667]
2	7	-7.7	3.13	1.9/3	2.5	2.5	119.8	ENSP00000216297 SUPT16H, suppressor of Ty 16 homolog (S. cerevisiae) [Source: HGNC 11465]
2	8	-6.5	3.14	2.3/4	1	1	54.2	ENSP00000276079 NONO, non-POU domain containing, octamer-binding [Source: HGNC 7871]
2	9	-6.4	3.32	1.7/3	1	3	66	ENSP00000252244 KRT1, keratin 1 [Source: HGNC 6412]
2	10	-3.3	3.59	1.9/2	1	2	70.5	ENSP00000417719 ALAS1, aminolevulinate, delta-, synthase 1 [Source: HGNC 396]
2	11	-2.8	3.03	2.8/5	1	1	45.5	ENSP00000436038 PSPC1, paraspeckle component 1 [Source: HGNC 20320]
2	12	-2.6	2.66	3.0/4	1	1	44.8	ENSP00000369127 DNAJA1, DnaJ (Hsp40) homolog, subfamily A, member 1 [Source: HGNC 5229]
2	13	-2.1	2.76	2.0/4	1	1	48.2	ENSP00000313829 KHDRBS1, KH domain containing, RNA binding, signal transduction associated 1 [Source: HGNC 18116]

Continued on next page

Table 2 – *Continued from previous page*

Band	rank	log(e)	log(I)	%/%	#	total	Mr	Accession
2	14	-2.1	2.57	1.9/2	1	1	56.5	ENSP00000262030 ATP5B, ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide [Source: HGNC 830]
2	15	-1.8	2.64	0.6/1	1	1	273.5	ENSP00000384018 NCOR2, nuclear receptor corepressor 2 [Source: HGNC 7673]
2	16	-1.6	2.83	7.2/8	1	1	17.1	ENSP00000376892 NME1, NME/NM23 nucleoside diphosphate kinase 1 [Source: HGNC 7849]
2	17	-1.5	3.32	0.5/1	1	2	101.4	ENSP00000388910 NFATC4
2	18	-1.4	2.67	3.1/5	1	1	58.8	ENSP00000269576 KRT10, keratin 10 [Source: HGNC 6413]
2	19	-1.4	2.49	1.9/3	1	1	54.8	ENSP00000372804 RNMT, RNA (guanine-7-) methyltransferase [Source: HGNC 10075]
2	20	-1.4	2.84	0.6/1	1	1	147.4	ENSP00000311502 HEG1
3	1	-101.7	4.71	47/54	14	28	32.6	ENSP00000451932 NME/NM23 nucleoside diphosphate kinase 2 [Source: HGNC 7850]
3	2	-75.9	4.62	8.6/10	1	2	17.1	ENSP00000376892 NME1, NME/NM23 nucleoside diphosphate kinase 1 [Source: HGNC 7849]
3	3	-40	4.16	33/40	6	11	16.8	ENSP00000336776 EIF5A, eukaryotic translation initiation factor 5A [Source: HGNC 3300]
3	4	-40	3.86	8.5/14	6	9	66	ENSP00000252244 KRT1, keratin 1 [Source: HGNC 6412]
3	5	-34.7	3.6	8.5/10	5	6	65.4	ENSP00000310861 KRT2, keratin 2 [Source: HGNC 6439]
3	6	-29.3	3.76	37/48	4	9	12.8	ENSP00000428085 RPL30, ribosomal protein L30 [Source: HGNC 10333]
3	7	-27.7	3.54	6.9/12	5	5.5	76.1	ENSP00000349748 SFPQ, splicing factor proline/glutamine-rich [Source: HGNC 10774]

Continued on next page

Table 2 – *Continued from previous page*

Band	**rank**	**log(e)**	**log(I)**	**%/%**	**#**	**total**	**Mr**	**Accession**
3	8	-26.2	3.63	9.1/12	5	6	61.4	ENSP00000277865 GLUD1, glutamate dehydrogenase 1 [Source: HGNC 4335]
3	9	-22.3	3.87	29/36	4.5	8	19.3	ENSP00000369097 C20orf27, Uncharacterized protein C20orf27.Source: Uniprot/SWISSPROT Q9GZN8
3	10	-21.4	3.42	21/23	4	4	25.4	ENSP00000395605 NME4
3	11	-19.1	3.63	6.3/10	4	4	62	ENSP00000246662 KRT9, keratin 9 [Source: HGNC 6447]
3	12	-14.7	3.57	15/19	3	5	24.4	TRYP PIG Trypsin; EC 3.4.21.4;
3	13	-13.9	2.96	4.5/6	2	2	65.1	ENSP00000357284 LMNA, lamin A/C [Source: HGNC 6636]
3	14	-12.7	3.39	4.9/15	3	4	53.4	ENSP00000254108 FUS, fused in sarcoma [Source: HGNC 4010]
3	15	-12.1	3.25	16/24	3	3	17.7	ENSP00000393241 RPS18, ribosomal protein S18 [Source: HGNC 10401]
3	16	-11.5	3	15/18	2	2	19.2	ENSP00000434190 MRPL49, mitochondrial ribosomal protein L49 [Source: HGNC 1176]
3	17	-8.4	3.06	2.8/4	1	2	54.2	ENSP00000276079 NONO, non-POU domain containing, octamer-binding [Source: HGNC 7871]
3	18	-8	2.65	1.4/2	1	1	113	ENSP00000355759 PARP1, poly (ADP-ribose) polymerase 1 [Source: HGNC 270]
3	19	-7.7	3.02	11/15	2	2	16.1	ENSP00000221975 RPS19, ribosomal protein S19 [Source: HGNC 10402]
3	20	-7	3.39	2.6/5	2	2	58.8	ENSP00000269576 KRT10, keratin 10 [Source: HGNC 6413]

Bibliography

Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics (Oxford, England)*, **21**(9), 2104.

Aderem, A. and Ulevitch, R. J. (2000). Toll-like receptors in the induction of the innate immune response. *Nature*, **406**(6797), 782–787.

Akira, S., Takeda, K., and Kaisho, T. (2001). Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat Immunol*, **2**(8), 675–680.

Alberts, B. (2008). *Molecular Biology of the Cell*. Garland Science.

Allen, L. A. H. and Aderem, A. (1996). Mechanisms of phagocytosis. *Current opinion in immunology*, **8**(1), 36–40.

Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool. *Journal of molecular biology*, **215**(3), 403–410.

Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**(17), 3389.

Alwan, H. A. J. and van Leeuwen, J. E. M. (2006). UBPY-mediated Epidermal Growth Factor Receptor (EGFR) De-ubiquitination Promotes EGFR Degradation. *The Journal of biological chemistry*, **282**(3), 1658–1669.

Anckar, J. and Sistonen, L. (2007). SUMO: getting it on. *Biochemical Society transactions*, **35**(Pt 6), 1409–1413.

Anderson, K. V., Jürgens, G., and Nüsslein-Volhard, C. (1985). Establishment of dorsal-ventral polarity in the Drosophila embryo: genetic studies on the role of the Toll gene product. *Cell*, **42**(3), 779–789.

Andrade, M. A., Perez-Iratxeta, C., and Ponting, C. P. (2001). Protein Repeats: Structures, Functions, and Evolution. *Journal of Structural Biology*, **134**(2-3), 117–131.

Arredouani, M., Yang, Z., Ning, Y., Qin, G., Soininen, R., Tryggvason, K., and Kobzik, L. (2004). The scavenger receptor MARCO is required for lung defense against pneumococcal pneumonia and inhaled particles. *The Journal of experimental medicine*, **200**(2), 267–272.

Arredouani, M., Palecanda, A., Koziel, H., Huang, Y., Imrich, A., Sulahian, T., Ning, Y., Yang, Z., Pikkarainen, T., Sankala, M., Vargas, S., Takeya, M., Tryggvason, K., and Kobzik, L. (2005). MARCO is the major binding receptor for unopsonized particles and bacteria on human alveolar macrophages. *Journal of immunology (Baltimore, Md : 1950)*, **175**(9), 6058–6064.

Arredouani, M., Yang, Z., Imrich, A., Ning, Y., Qin, G., and Kobzik, L. (2006). The

macrophage scavenger receptor SR-AI/II and lung defense against pneumococci and particles. *Am J Respir Cell Mol Biol*, **35**(4), 474–478.

Bairoch, A. (1991). PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Research*, **19**(Suppl), 2241.

Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites1. *Journal of molecular biology*, **294**(5), 1351–1362.

Boissan, M. and Lacombe, M.-L. (2011). Learning about the functions of NME/NM23: lessons from knockout mice to silencing strategies. *Naunyn-Schmiedeberg's Archives of Pharmacology*, **384**(4-5), 421–431.

Bowdish, D. and Gordon, S. (2009). Conserved domains of the class A scavenger receptors: evolution and function. *Immunol Rev*, **227**(1), 19–31.

Bowdish, D., Sakamoto, K., Kim, M., Kroos, M., Mukhopadhyay, S., Leifer, C., Tryggvason, K., Gordon, S., and Russell, D. (2009). MARCO, TLR2, and CD14 are required for macrophage cytokine responses to mycobacterial trehalose dimycolate and Mycobacterium tuberculosis. *PLoS Pathog*, **5**(6), e1000474.

Brännström, A., Sankala, M., Tryggvason, K., and Pikkarainen, T. (2002). Arginine Residues in Domain V Have a Central Role for Bacteria-Binding Activity of Macrophage Scavenger Receptor MARCO* 1. *Biochemical and Biophysical Research Communications*, **290**(5), 1462–1469.

Brown, M. S., Basu, S. K., Falck, J., Ho, Y., and Goldstein, J. L. (1980). The scavenger cell pathway for lipoprotein degradation: Specificity of the binding site

that mediates the uptake of negatively-charged LDL by macrophages. *Journal of Supramolecular Structure*, **13**(13), 67–81.

Chang, X., Zhao, Y., Yan, X., Pan, J., Fang, K., and Wang, L. (2011). Investigating a pathogenic role for TXNDC5 in rheumatoid arthritis. *Arthritis Research & Therapy*, **13**(4), R124.

Coates, P. (2001). Mammalian Prohibitin Proteins Respond to Mitochondrial Stress and Decrease during Cellular Senescence. *Experimental cell research*, **265**(2), 262–273.

Coulombe, P. and Meloche, S. (2007). Atypical mitogen-activated protein kinases: Structure, regulation and functions. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, **1773**(8), 1376–1387.

DeWitte-Orr, S. J., Collins, S. E., Bauer, C. M. T., Bowdish, D. M., and Mossman, K. L. (2010). An Accessory to the 'Trinity': SR-As Are Essential Pathogen Sensors of Extracellular dsRNA, Mediating Entry and Leading to Subsequent Type I IFN Responses. *PLoS Pathog*, **6**(3), e1000829.

Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N. P., Trave, G., and Gibson, T. J. (2008). Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Frontiers in bioscience : a journal and virtual library*, **13**, 6580–6603.

Doi, T., Higashino, K., Kurihara, Y., Wada, Y., Miyazaki, T., Nakamura, H., Uesugi,

S., Imanishi, T., Kawabe, Y., and Itakura, H. (1993). Charged collagen structure mediates the recognition of negatively charged macromolecules by macrophage scavenger receptors. *The Journal of biological chemistry*, **268**(3), 2126–2133.

Drickamer, K. (1999). C-type lectin-like domains. *Current opinion in structural biology*, **9**(5), 585–590.

Duckert, P., Brunak, S., and Blom, N. (2004). Prediction of proprotein convertase cleavage sites. *Protein Engineering Design and Selection*, **17**(1), 107–112.

Eddy, S. (1998). Profile hidden Markov models. *Bioinformatics (Oxford, England)*, **14**(9), 755.

Edgar, R. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**(5), 1792.

Elliott, D. J. (2004). The role of potential splicing factors including RBMY, RBMX, hnRNPGT and STAR proteins in spermatogenesis*. *International journal of andrology*, **27**(6), 328–334.

Elomaa, O., Kangas, M., Sahlberg, C., Tuukkanen, J., Sormunen, R., Liakka, A., Thesleff, I., Kraal, G., and Tryggvason, K. (1995). Cloning of a novel bacteria-binding receptor structurally related to scavenger receptors and expressed in a subset of macrophages. *Cell*, **80**(4), 603–609.

Elomaa, O., Sankala, M., Pikkarainen, T., Bergmann, U., Tuuttila, A., Raatikainen-Ahokas, A., Sariola, H., and Tryggvason, K. (1998). Structure of the human macrophage MARCO receptor and characterization of its bacteria-binding region. *The Journal of biological chemistry*, **273**(8), 4530–4538.

Fong, L. G. (1996). Modulation of macrophage scavenger receptor transport by protein phosphorylation. *J Lipid Res*, **37**(3), 574–587.

Fong, L. G. and Le, D. (1999). The processing of ligands by the class A scavenger receptor is dependent on signal information located in the cytoplasmic domain. *The Journal of biological chemistry*, **274**(51), 36808–36816.

Fried, M. and Colombo, P. (1992). Functional elements of the ribosomal protein L7a (rpL7a) gene promoter region and their conservation between mammals and birds. *Nucleic Acids Research*, **20**(13), 3367–3373.

Fujita, T. (2002). Evolution of the lectin–complement pathway and its role in innate immunity. *Nat Rev Immunol*, **2**(5), 346–353.

Geiss-Friedlander, R. and Melchior, F. (2007). Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol*, **8**(12), 947–956.

Goh, J. W. K., Tan, Y. S., Dodds, A. W., Reid, K. B. M., and Lu, J. (2010). The class A macrophage scavenger receptor type I (SR-AI) recognizes complement iC3b and mediates NF-B activation. *Protein & cell*, **1**(2), 174–187.

Goldstein, J., Ho, Y., Basu, S., and Brown, M. (1979). Binding site on macrophages that mediates uptake and degradation of acetylated low density lipoprotein, producing massive cholesterol deposition. *Proceedings of the National Academy of Sciences of the United States of America*, **76**(1), 333.

Gordon, S. (2002). Pattern recognition receptors: doubling up for the innate immune response. *Cell*, **111**(7), 927–930.

Han, H. J., Tokino, T., and Nakamura, Y. (1998). CSR, a scavenger receptor-like protein with a protective role against cellular damage causedby UV irradiation and oxidative stress. *Human molecular genetics*, **7**(6), 1039–1046.

Heider, H. and Wintergerst, E. (2001). Mimicking phosphorylation at Ser-48 strongly reduces surface expression of human macrophage scavenger receptor class A: implications on cell motility. *FEBS Lett*, **505**(1), 185–190.

Hicke, L. (2001). Protein regulation by monoubiquitin. *Nature Reviews Molecular Cell Biology*, **2**(3), 195–201.

Holmskov, U., Malhotra, R., Sim, R. B., and Jensenius, J. C. (1994). Collectins: collagenous C-type lectins of the innate immune defense system. *Immunology today*, **15**(2), 67–74.

Huelsenbeck, J. P. and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)*, **17**(8), 754–755.

Igel, H., Wells, S., Perriman, R., and Ares Jr, M. (1998). Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits. *Rna*, **4**(1), 1–10.

Jiang, Y., Oliver, P., Davies, K., and Platt, N. (2006). Identification and characterization of murine SCARA5, a novel class A scavenger receptor that is expressed by populations of epithelial cells. *Journal of Biological Chemistry*, **281**(17), 11834.

Jones, D. (1999). Protein secondary structure prediction based on position-specific scoring matrices1. *Journal of molecular biology*, **292**(2), 195–202.

Julenius, K. (2007). NetCGlyc 1.0: prediction of mammalian C-mannosylation sites. *Glycobiology*, **17**(8), 868–876.

Kent, W. (2002). BLAT—the BLAST-like alignment tool. *Genome research*, **12**(4), 656–664.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome research*, **12**(6), 996–1006.

Kleber, S., Sancho-Martinez, I., Wiestler, B., Beisel, A., Gieffers, C., Hill, O., Thiemann, M., Mueller, W., Sykora, J., Kuhn, A., Schreglmann, N., Letellier, E., Zuliani, C., Klussmann, S., Teodorczyk, M., Gröne, H.-J., Ganten, T. M., Sültmann, H., Tüttenberg, J., von Deimling, A., Regnier-Vigouroux, A., Herold-Mende, C., and Martin-Villalba, A. (2008). Yes and PI3K bind CD95 to signal invasion of glioblastoma. *Cancer cell*, **13**(3), 235–248.

Kodama, T., Freeman, M., Rohrer, L., Zabrecky, J., Matsudaira, P., and Krieger, M. (1990). Type I macrophage scavenger receptor contains alpha-helical and collagen-like coiled coils. *Nature*, **343**(6258), 531–535.

Kraal, G., van der Laan, L., Elomaa, O., and Tryggvason, K. (2000). The macrophage receptor MARCO. *Microbes Infect*, **2**(3), 313–316.

Krieger, M. (1992). Molecular flypaper and atherosclerosis: structure of the macrophage scavenger receptor. *Trends in biochemical sciences*, **17**(4), 141–146.

Krieger, M. (1997). The other side of scavenger receptors: pattern recognition for host defense. *Current opinion in lipidology*, **8**(5), 275.

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes1. *Journal of molecular biology*, **305**(3), 567–580.

Kroos, M. (2008). *The role of the cytoplasmic domain of MARCO in adhesion and uptake.* Ph.D. thesis, University of Oxford, Sir William Dunn School of Pathology and Lincoln College.

Lefebvre, C. (2002). Meiotic spindle stability depends on MAPK-interacting and spindle-stabilizing protein (MISS), a new MAPK substrate. *The Journal of Cell Biology*, **157**(4), 603–613.

Maekawa, M. (2002). Identification of the Anti-proliferative Protein Tob as a MAPK Substrate. *Journal of Biological Chemistry*, **277**(40), 37783–37787.

Manning, G., Whyte, D., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science's STKE*, **298**(5600), 1912.

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchler, G. H., Mullokandov, M., Omelchenko, M. V., Robertson, C. L., Song, J. S., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N., Zheng, C., and Bryant, S. H. (2010). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, **39**(Database), D225–D229.

Marks, F. (2008). *Protein Phosphorylation.* John Wiley & Sons.

Martínez, V. G., Moestrup, S. K., Holmskov, U., Mollenhauer, J., and Lozano, F. (2011). The conserved scavenger receptor cysteine-rich superfamily in therapy and diagnosis. *Pharmacological reviews*, **63**(4), 967–1000.

Matsushita, M., Matsushita, A., Endo, Y., Nakata, M., Kojima, N., Mizuochi, T., and Fujita, T. (2004). Origin of the classical complement pathway: Lamprey orthologue of mammalian C1q acts as a lectin. *Proc Natl Acad Sci U S A*, **101**(27), 10127–10131.

Maurer-Stroh, S. and Eisenhaber, F. (2005). Refinement and prediction of protein prenylation motifs. *Genome Biology*, **6**(6), R55.

McAlinden, A. (2003). -Helical Coiled-coil Oligomerization Domains Are Almost Ubiquitous in the Collagen Superfamily. *The Journal of biological chemistry*, **278**(43), 42200–42207.

Monigatti, F., Gasteiger, E., Bairoch, A., and Jung, E. (2002). The Sulfinator: predicting tyrosine sulfation sites in protein sequences. *Bioinformatics (Oxford, England)*, **18**(5), 769–770.

Murphy, K., Travers, P., and Walport, M. (2011). *Janeway's Immunobiology*. Seventh Edition. Garland Science, 7 edition.

Nakamura, K., Funakoshi, H., Miyamoto, K., Tokunaga, F., and Nakamura, T. (2001). Molecular Cloning and Functional Characterization of a Human Scavenger Receptor with C-Type Lectin (SRCL), a Novel Member of a Scavenger Receptor Family* 1,* 2. *Biochemical and Biophysical Research Communications*, **280**(4), 1028–1035.

Neduva, V. and Russell, R. B. (2006). Peptides mediating interaction networks: new leads at last. *Current Opinion in Biotechnology*, **17**(5), 465–471.

Nylander, J., Wilgenbusch, J., and Warren, D. (2008). AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. ....

Obenauer, J. C. (2003). Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Research*, **31**(13), 3635–3641.

Parks, D. and Beiko, R. (2010). Identifying biologically relevant differences between metagenomic communities. *Bioinformatics (Oxford, England)*, **26**(6), 715.

Parry, D. A. D., Fraser, R. D. B., and Squire, J. M. (2008). Fifty years of coiled-coils and -helical bundles: A close relationship between sequence and structure. *Journal of Structural Biology*, **163**(3), 258–269.

Pearson, A. M. (1996). Scavenger receptors in innate immunity. *Current opinion in immunology*, **8**(1), 20–28.

Pearson, G., Robinson, F., Beers Gibson, T., Xu, B. E., Karandikar, M., Berman, K., and Cobb, M. H. (2001). Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. *Endocrine reviews*, **22**(2), 153–183.

Pearson, W. R. and Lipman, D. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A*, **85**(8), 2444.

Peiser, L., Gough, P. J., Kodama, T., and Gordon, S. (2000). Macrophage Class A

Scavenger Receptor-Mediated Phagocytosis of Escherichia coli: Role of Cell Heterogeneity, Microbial Strain, and Culture Conditions In Vitro. *Infection and Immunity*, **68**(4), 1953–1963.

Plewczynski, D., Tkacz, A., Wyrwicz, L. S., and Rychlewski, L. (2005). AutoMotif server: prediction of single residue post-translational modifications in proteins. *Bioinformatics (Oxford, England)*, **21**(10), 2525–2527.

Plüddemann, A., Mukhopadhyay, S., and Gordon, S. (2006). The interaction of macrophage receptors with bacterial ligands. *Expert reviews in molecular medicine*, **8**(28), 1–25.

Plüddemann, A., Neyen, C., and Gordon, S. (2007). Macrophage scavenger receptors and host-derived ligands. *Methods*, **43**(3), 207–217.

Plüddemann, A., Mukhopadhyay, S., Sankala, M., Savino, S., Pizza, M., Rappuoli, R., Tryggvason, K., and Gordon, S. (2009). SR-A, MARCO and TLRs Differentially Recognise Selected Surface Proteins from Neisseria meningitidis: an Example of Fine Specificity in Microbial Ligand Recognition by Innate Immune Receptors. *Journal of Innate Immunity*, **1**(2), 153–163.

Posada, D. (2008). jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*, **25**(7), 1253–1256.

Puntervoll, P. (2003). ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Research*, **31**(13), 3625–3630.

Radivojac, P., Vacic, V., Haynes, C., Cocklin, R. R., Mohan, A., Heyen, J. W., Goebl, M. G., and Iakoucheva, L. M. (2010). Identification, analysis, and prediction of

protein ubiquitination sites. *Proteins: Structure, Function, and Bioinformatics*, **78**(2), 365–380.

Rambaut, A. and Drummond, A. (2009a). FigTree v1.3.1: Tree figure drawing tool. Available: http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut, A. and Drummond, A. (2009b). Tracer v1.5, Available from http://beast.bio.ed.ac.uk/software/Tracer.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends in Genetics*, **16**(6), 276–277.

Roach, J., Glusman, G., Rowen, L., Kaur, A., Purcell, M., Smith, K., Hood, L., and Aderem, A. (2005). The evolution of vertebrate Toll-like receptors. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(27), 9577.

Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)*, **19**(12), 1572–1574.

Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P., and Vidal, M. (2005). Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, **437**(7062), 1173–1178.

Rubin, C. S. and Rosen, O. M. (1975). Protein phosphorylation. *Annual review of biochemistry*, **44**, 831–887.

Saini, P., Eyler, D. E., Green, R., and Dever, T. E. (2009). Hypusine-containing protein eIF5A promotes translation elongation. *Nature*, **459**(7243), 118–121.

Schulenburg, H., Kurz, C. L., and Ewbank, J. J. (2004). Evolution of the innate immune system: the worm perspective. *Immunol Rev*, **198**, 36–58.

Selman, L., Skjødt, K., Nielsen, O., Floridon, C., Holmskov, U., and Hansen, S. (2008). Expression and tissue localization of collectin placenta 1 (CL-P1, SRCL) in human tissues. *Molecular immunology*, **45**(11), 3278–3288.

Shapiro, A. L., Scharff, M. D., Maizel, J. V., and Uhr, J. W. (1966). Polyribosomal synthesis and assembly of the H and L chains of gamma globulin. *Proc Natl Acad Sci U S A*, **56**(1), 216–221.

Smith, T. F., Waterman, M. S., and Fitch, W. M. (1981). Comparative biosequence metrics. *Journal of molecular evolution*, **18**(1), 38–46.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)*, **22**(21), 2688–2690.

Strous, G. J., van Kerkhof, P., Govers, R., Ciechanover, A., and Schwartz, A. L. (1996). The ubiquitin conjugation system is required for ligand-induced endocytosis and degradation of the growth hormone receptor. *The EMBO journal*, **15**(15), 3806–3812.

Suzuki, H., Kurihara, Y., Takeya, M., Kamada, N., Kataoka, M., Jishage, K., Ueda, O., Sakaguchi, H., Higashi, T., Suzuki, T., Takashima, Y., Kawabe, Y., Cynshi, O., Wada, Y., Honda, M., Kurihara, H., Aburatani, H., Doi, T., Matsumoto, A., Azuma, S., Noda, T., Toyoda, Y., Itakura, H., Yazaki, Y., and Kodama, T. (1997). A role for macrophage scavenger receptors in atherosclerosis and susceptibility to infection. *Nature*, **386**(6622), 292–296.

Tomokiyo, R.-i., Jinnouchi, K., Honda, M., Wada, Y., Hanada, N., Hiraoka, T., Suzuki, H., Kodama, T., Takahashi, K., and Takeya, M. (2002). Production, characterization, and interspecies reactivities of monoclonal antibodies against human class A macrophage scavenger receptors. *Atherosclerosis*, **161**(1), 123–132.

Van Wijk, E., Krieger, E., Kemperman, M. H., De Leenheer, E., Huygen, P., Cremers, C., Cremers, F., and Kremer, H. (2003). A mutation in the gamma actin 1 (ACTG1) gene causes autosomal dominant hearing loss (DFNA20/26). *Journal of medical genetics*, **40**(12), 879–884.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*, **25**(9), 1189–1191.

Weissman, A. M. (2001). Themes and variations on ubiquitylation. *Nature Reviews Molecular Cell Biology*, **2**(3), 169–178.

Welchman, R. L., Gordon, C., and Mayer, R. J. (2005). Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat Rev Mol Cell Biol*, **6**(8), 599–609.

Yang, L., He, Y., Kong, Q., Zhang, W., Xi, D., Mao, H., and Deng, W. (2009).

Isolation, nucleotide identification and tissue expression of three novel ovine genes—SLC25A4, SLC25A5 and SLC25A6. *Molecular Biology Reports*, **37**(6), 2743–2748.

Yu, X., Yi, H., Guo, C., Zuo, D., Wang, Y., Kim, H. L., Subjeck, J. R., and Wang, X. Y. (2011). Pattern Recognition Scavenger Receptor CD204 Attenuates Toll-like Receptor 4-induced NF- B Activation by Directly Inhibiting Ubiquitination of Tumor Necrosis Factor (TNF) Receptor-associated Factor 6. *The Journal of biological chemistry*, **286**(21), 18795–18806.

Yu, Y., Ji, H., Doudna, J. A., and Leary, J. A. (2009). Mass spectrometric analysis of the human 40S ribosomal subunit: Native and HCV IRES-bound complexes. *Protein Science*, **14**(6), 1438–1446.