SHORT PEPTIDE BINDING IN S100A5 AND A6: A PARADIGM FOR THE BIOMOLECULAR EVOLUTION OF SPECIFICITY

by

Caitlyn Wong

A THESIS

Presented to the Department of Chemistry and Biochemistry
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

December 2017

An Abstract of the Thesis of

Caitlyn Wong for the degree of Bachelor of Science
in the Department of Biochemistry to be taken December 2017

Title: Short Peptide Binding in S100A5 and A6: A Paradigm for the Biomolecular Evolution of Specificity

Approved: _____

Michael J. Harms, Ph.D.

S100A5 and S100A6 provide an excellent model for study of the evolution of binding specificity in proteins because they are structurally similar yet functionally diverse. We set out to determine the biochemical attributes and biological constraints responsible for the divergence of specificity between these paralogs that occurred post gene duplication of the S100A5/A6 ancestor. We first characterized the evolution of protein structure in this subfamily by sampling extant amniote orthologs, S100A5 and A6 ancestors, and the shared S100A5/A6 ancestor. We used circular dichroism (CD) spectroscopy to show a phylogenetic pattern of secondary structure and conformation change upon Ca^{2+} binding. The increase in response to Ca^{2+} binding in S100A5 orthologs and their shared ancestor is a derived trait relative to S100A6 and ancestral S100A5/A6. We looked to crystallography to identify the molecular mechanism for peptide binding in S100A5 and reveal the chemical basis for this observed divergence. We report crystallization conditions for human S100A5 in both the Ca^{2+}-bound and Ca^{2+} and peptide-bound states that require further optimization before submission for diffraction. We then developed a co-immunoprecipitation protocol to isolate and identify targets of mouse S100A5 from olfactory bulb tissue. We report a maximal elution fraction of 0.54 for mouse S100A5 using our primary antibody and, despite cross-reactivity with mouse S100A6, an expectation of 60% of binding partners isolated to be of mouse S100A5. Our protocol provides a starting point for identifying targets bound by S100A5 in olfactory tissue.

Acknowledgements

I would like to thank Professor Mike Harms and Ph.D. student Luke Wheeler for their guidance throughout my time in the Harms Lab. I have grown as a student and a scientist, and am grateful for Luke's mentorship. Gratitude is also extended to all members of the Harms Lab for being so fun and welcoming. I feel very lucky to have learned from a group of talented and well-rounded people.

Outside of the Harms Lab, there are many people who have contributed to my journey as an undergraduate. I am eternally grateful for the compassion and support of Professor Louise Bishop since I arrived as a wide-eyed freshman. She served as my Honors College Representative on my Thesis Committee and has eased the pressures and stress of being a student-athlete in the Clark Honors College.

Jennifer Jackson, Katie Harbert, and Resa Lovelace are key members of the academic support staff that have helped me to succeed off the field. I would like to thank them in addition to Lisa Peterson for supporting my endeavors on and off the field. I am also eternally grateful for Heather Halseth, whose insight and care has kept me healthy and allowed me to compete in the sport I love. Finally, the Oregon Soccer coaching staff, including Kat Mertz, Manny Martins, Katie Hultin, and Tom Serratore, has enriched my journey as a student-athlete. Thank you to them as well as my past coaches for helping me to develop as a soccer player, teammate, and person.

Above all, I owe my success, drive, and happiness to my family. Words cannot describe the depth of my gratitude for their love and support. Everything I do is for them.

Table of Contents

List of Figures

List of Tables

Introduction

History facilitates understanding, development, and prediction. For example, it would be frightening to see people entranced by the black screen of a small rectangular device if we did not know of the path that cellular technology has taken since Alexander Graham Bell's invention of the telephone. Since then, innovators have repeatedly adjusted and improved the technology to meet the demands of its time and users. The telephone has evolved over the last century so that each generation is more portable and practical than the one before, and we can extract themes from this history to predict the characteristics of future devices. Overall, we cannot understand the modern telephone without considering how it has changed over time. Likewise, evolutionary thinking adds historical context to any scientific research problem. We hold a narrow view of the world and its complexities without consideration of the how and the why. To gain more thorough insight regarding the questions we have, history must be applied to science.

Evolutionary biochemists study the "stories" of proteins[1] and other biomolecules in hope of understanding why they have the properties that they do. Proteins play irreplaceable roles in the cell and body and are essential to biological life in general. They are comprised of a sequence of amino acids[2], the unique molecular building blocks encoded by a protein's genetic DNA[3]. In the cell, DNA is transcribed into RNA[4], which is then translated into protein. The properties of the amino acids in a

---

[1] Molecules made of amino acids in a linear chain that fold into a structure to perform many biological functions.
[2] An organic compound that contains a carboxyl and an amino group.
[3] Deoxyribonucleic acid, the self-replicating material that acts as the carrier of genetic information in the cell.
[4] Ribonucleic acid, the result of DNA transcription, encodes protein sequence.

protein determine its structure and function. While the complexity of this process leaves many opportunities for mistakes, it has contributed to the biological diversity we see macroscopically as well as on the molecular level.

Genetic mutation drives protein evolution by causing changes in sequence and function [1]. One downstream consequence is binding specificity, which is essential to the efficiency and accuracy of protein function. Binding occurs through favorable chemical interactions between the protein and target, and specificity is determined by the participating amino acids and other biomolecules[5]. Both sequence and structure contribute to a protein's ability to select its target from a wide swath of possible interaction partners [2]. Since binding requires physical interaction, the characteristics of a protein's biological environment can attenuate its specificity and function. One of the most widely recognized theories of protein evolution is the concept of promiscuity, which evolutionary biochemists define as the ability of a protein to bind many interaction partners that are not biologically relevant [2]. These latent interactions can act as starting points for future evolution. While a protein may be capable of binding to many partners, biological constraints such as co-localization and sterics reduce this massive collection of biochemically possible interaction partners to a much smaller set of biologically realized targets. Mutation drives shifts in binding specificity by exploiting these latent interactions [2,3].

Ultimately, evolution has produced a vast repertoire of proteins diverse in structure, environment, and function. Alone, biochemistry can tell us what a protein does and how it does it. However, an evolutionary approach also tells us why.

[5] Any molecule involved in biology or in a living organism. Can be a protein, lipid, or vitamin, etc.

Evolutionary biochemistry and its recent resurfacing in scientific discourse has provided a more complete understanding of the influences that mutations pose on central protein processes such as folding[6], binding, and stability [3]. Amino acid substitutions can alter the mode of binding in a protein, contributing to the divergence of paralogs[7] and the evolutionary path of a protein family [1]. The most ubiquitous result is sub-functionalization[8], or the divergence of function between paralogs following gene duplication[9] [4]. Many biochemists have aimed to dissect the mechanism behind specificity for drug engineering purposes, and have only recently included an evolutionary perspective. By characterizing the evolutionary history of a protein, one can understand the molecular basis for its present function and the engineering principles for manipulating it.

In the Harms Lab we use a combination of biophysics and phylogenetics[10] to study the evolutionary trajectories of the S100s, a protein family implicated in many biological functions and diseases [5,6]. Figure 1 outlines the evolution of this family in a phylogenetic tree.

---

[6] The process of forming a 3-dimensional structure that a protein undergoes after translation.
[7] Gene "relatives" that evolved through gene duplication.
[8] Degenerate mutations that result in a gene and its duplicated copy sharing the burden of one function [4].
[9] Duplication of a gene in DNA so that there are two copies of the same gene. One copy can acquire a mutation thus leading to a different protein.
[10] The study of biological evolution. In evolutionary biochemistry, phylogenetics involves the study of evolution at the molecular level.

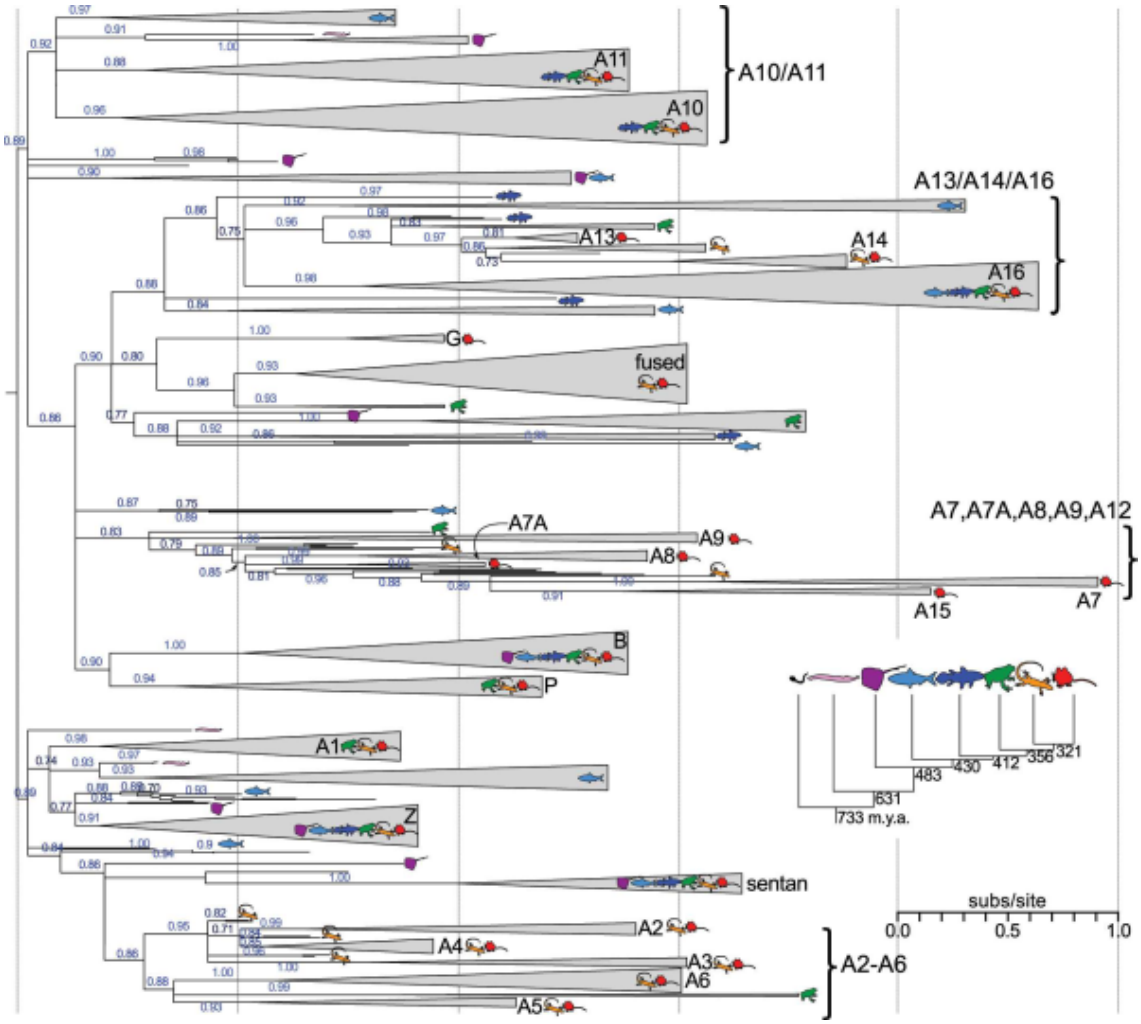

Figure 1. The S100 protein family exhibits great sequence diversity.

Wheeler et al used model-based phylogenetics to produce a maximum likelihood phylogeny[11] for the S100 protein family. 564 S100 proteins were drawn from 52 *Olfactores* species. The tree is rooted arbitrarily and the support values are SH supports derived from an approximate likelihood ratio test. Taxonomic classes are represented by the colored icons, and the timing of speciation is outlined in the small phylogeny at the bottom right. The scale for branch length (subs/site) represents the probably that the amino acid at each site in the sequence will undergo a substitution. Wedge height denotes the number of taxa in the respective clade and wedge length denotes the longest branch length with the clade [5].

[11] The evolutionary path of the ancestral and extant members of a protein family.

The S100 protein family serves as an excellent model for studying protein evolution at the molecular level because its members are structurally similar yet functionally diverse [6]. The S100s have evolved by both gene duplication and speciation[12] from a single initial ancestor into a family of widespread sequence diversity, as shown in Figure 1. This has allowed the S100s to fulfill various biological roles and operate in various biological environments. Through binding to their targets, the extant[13] proteins in the S100 family contribute to the regulation of biological processes such as cell proliferation[14], apoptosis[15], metabolism[16], inflammation[17], and immunity[18] [6].

My research focuses on the problem of protein specificity and uses the S100A5/A6 subfamily as a model system. S100A5 (A5) and S100A6 (A6) are "sister proteins," which means that they share a common ancestor and are therefore more closely related to each other than to other S100s. Figure 2 shows the divergence of A5 and A6 that occurred after a gene duplication event roughly three hundred million years ago [7]. Amniote[19] speciation that occurred post gene duplication of S100A5/A6 has also contributed to the diversification of this clade[20].

---

[12] The evolution of a new species.
[13] The members of a protein family that are still in existence.
[14] Cell replication and division.
[15] Programmed cell death.
[16] The cellular process of breaking down resources to use for energy or in the building of other molecules.
[17] A biological response that causes localization of redness and swelling due to injury or infection.
[18] An organism's ability to fight infection and disease. There is both innate and adaptive immunity; both contribute to the health and longevity of the organism.
[19] A mammal, bird, or reptile. Amniotes are vertebrate animals that have an amnion (membrane enclosing the embryo) during the embryonic stage.
[20] A group of organisms related by a common ancestor. All the species or homologs that have evolved after a "node" in a phylogenetic diagram.

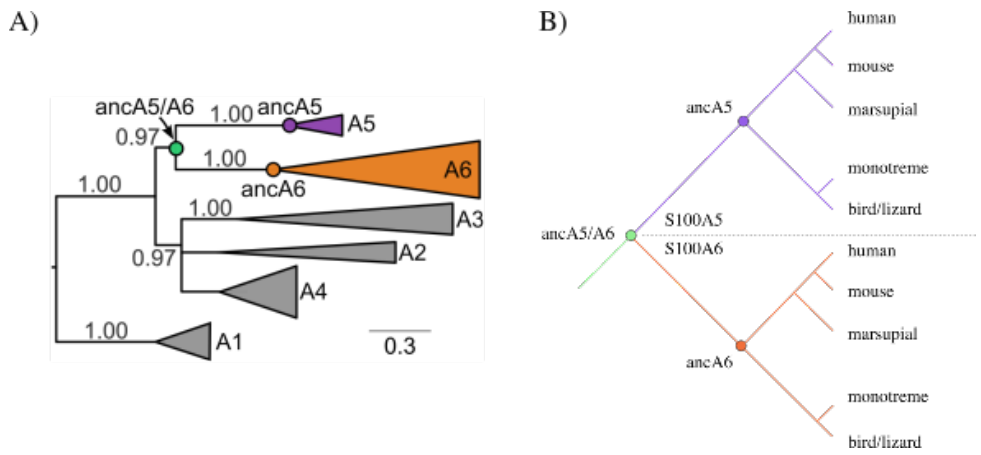

Figure 2. Schematic representation of the S100A5 and S100A6 clade.

A) Maximum likelihood phylogeny for S100A5, A6, and their close homologs (S100A1, S100A2, S100A3, S100A4). Wedge height denotes the number of sequences in the clade and wedge length denotes the maximum branch length for the clade. The SH supports, listed above each branch, were estimated using an approximate likelihood ratio test. The circles represent reconstructed ancestors (ancestral S100A5/A6, ancestral S100A5, and ancestral S100A6). The scale represents the probability that each site in the protein sequence will undergo a substitution [7]. B) Divergence of S100A5 (purple) and S100A6 (orange) occurred after a gene duplication of the ancestral protein (green) about 300 million years ago. The species listed on the tips of the tree are members of the amniote family and represent extant species that contain copies of each protein.

We selected representative species for each extant protein in order to conduct a complete survey of peptide binding along the A5/A6 lineage (Figure 3). Additionally, Luke Wheeler, a Ph.D. candidate in the Harms Lab and my graduate student mentor, cloned the ancestral S100A5, S100A6, and S100A5/A6 proteins using a process called Ancestral Sequence Reconstruction[21]. Expression and purification of these constructs allowed us to perform biophysical characterization experiments on representatives from the S100A5/A6 clade shown above. By focusing on the S100A5/A6 subfamily, we aimed to extrapolate general principles of binding specificity and its evolution.

[21] The process by which the sequence of an ancestral protein is inferred from present data.

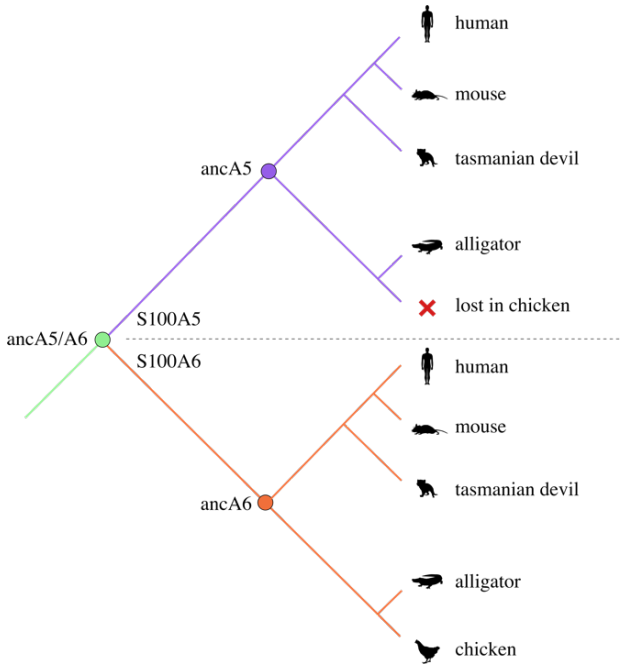

Figure 3. A wide range of amniote species in addition to ancestral proteins allows for comprehensive evolutionary study.

Members chosen for study represent every extant sequence for S100A5 and S100A6 in addition to ancestral nodes in order to determine the points of differentiation along this lineage. Reconstructed ancestors allow for the determination of ancestral and derived characteristics.

S100A5 and A6 function as homodimers[22], meaning that they are comprised of two identical subunits. Both proteins are allosterically regulated[23] by the binding of calcium, and the resulting conformation change exposes a hydrophobic[24] cleft in each

[22] Proteins that are made of two identical subunits. The subunits are separate sequences of the same amino acids but are held together through molecular attractions.
[23] When binding at one site of a protein has an effect on binding at another site, often contributing to regulation of function.
[24] The tendency of a nonpolar protein, amino acid, or biomolecule to fail to mix with water.

monomer[25] [7-10]. An example is shown in human S100A5 (hS100A5) alongside the structure of S100A6 bound to calcium and two peptides[26] (Figure 4).

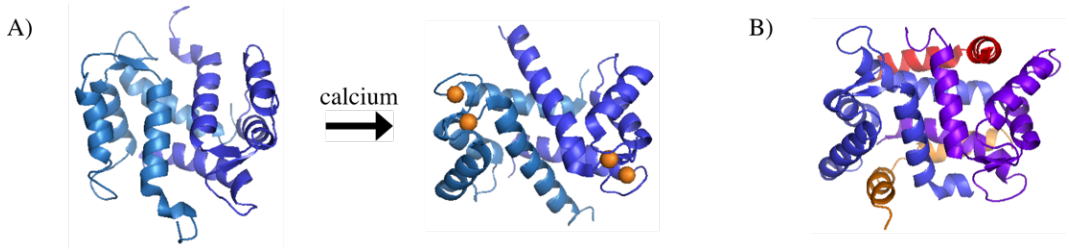

Figure 4. Calcium binding induces a conformational change in S100A5 and S100A6.

A) Helices in both subunits of hS100A5 (blue) shift upon binding four calcium ions (orange spheres) [9]. B) S100A6 (blue/purple) binds two fragments of the Siah-1 interacting protein (SIP) peptide in the hydrophobic pockets exposed upon calcium binding [10].

The "binding pockets" are the clefts exposed upon calcium binding and are where the protein makes contact with short target peptide fragments [7,9,10]. S100A5 and A6 bind peptides in these pockets to regulate calcium-dependent signaling. However, despite their structural homology[27], these proteins fulfill very different biological roles.

S100A6 is well characterized and has been shown to interact with a variety of cytosolic[28] proteins [10]. It promotes cell proliferation through binding with p53[29], and consequentially serves as a cancer precursor if overexpressed [6,10]. Crystal structures have been published for S100A6 in the apo[30], calcium-bound, and calcium and peptide-bound states, revealing the mechanism for target binding in this protein [10]. The amino

[25] A single unit of a protein.
[26] Short chains of two or more amino acids. Can have arbitrary length and does not normally possess enzymatic function.
[27] Similarity due to shared ancestry.
[28] Existing in the aqueous component within the cell
[29] An infamous gene linked to many different types of cancers.
[30] A state in which the proteins under study are completely neglect of calcium.

acids exposed at the binding interface exploit hydrophobic and electrostatic[31] interactions to drive binding in S100A6 [10].

In contrast, less is known about S100A5. It was recently shown to interact with the receptor for advanced glycation end product (RAGE), a cell surface receptor implicated in signaling pathways for many diseases [11], and a fragment of NCX1, a sodium-calcium exchanger [12]. Like many proteins, misregulation of S100A5 is linked to oncogenesis [9]. However, there is currently no crystal structure for the S100A5 protein in a calcium and peptide-bound state. For this reason, the structural mechanism for binding in S100A5 is unknown.

S100A5 has been implicated in olfactory signaling, the process by which organisms react and adapt to the odorous stimuli from the environment [13-15]. Previous work has shown that it localizes in the cilia[32] of olfactory sensory neurons[33] (OSNs) and is therefore hypothesized to be involved in transmission of external information to the olfactory bulb[34] for downstream processing [13]. Additionally, its expression is tightly controlled and activity-dependent, suggesting that this protein is a key player in this process [13]. Many hypothesize that S100A5 contributes to olfactory signaling through calcium regulation, since homeostasis[35] promotes the survival and maintenance of OSNs [14,16]. However, the targets of this protein are poorly characterized, limiting our understanding of its role in the signaling process. Since a

---

[31] Interactions involving electric charges on amino acids or other biomolecules.
[32] A hairlike structure that occurs on the surfaces of receptor cells.
[33] Cells that transduce signals from the external environment to the rest of the olfactory system.
[34] Neural structure involved in olfaction (smell).
[35] The relatively stable equilibrium of elements, essential to maintenance of biological processes.

properly functioning olfactory system is essential to health, it is important to understand the mechanism of S100A5's contribution.

Luke Wheeler has worked to characterize the evolution of binding specificity in the S100A5/A6 subfamily. His experiments have shown that both S100A5 and S100A6 are biochemically promiscuous, meaning that they can bind a wide range of short peptides *in vitro*[36] [7,18]. Despite this general motif, these paralogs show distinct binding specificities. Wheeler's phage display[37] experiments demonstrated that S100A5 has a smaller and more divergent interaction set than S100A6 and the reconstructed[38] A5/A6 ancestor [18]. His results suggest that mutation in the A5/A6 ancestral gene post duplication has altered the selectivity of the S100A5 protein [7,18]. However, genetic mutations and the resulting amino acid substitutions can do so through many avenues. The differences between the S100A5 and A6 sequences may contribute to different responses to calcium, differences in binding interface, or differences in overall structure. These discrepancies are further refined by each protein's biological environment. Ultimately, there are numerous layers to the evolutionary story of target peptide binding in this clade. My research aims to elucidate the biochemical attributes and biological constraints that explain the divergence in peptide specificity between S100A5 and S100A6.

---

[36] Refers to experiments performed in a test tube or culture dish, etc.
[37] Use of the random peptide sequences displayed by bacteriophages to survey peptide binding in proteins. The phages that contain bound peptide sequences can be isolated and replicated. Ultimately, the sequence of the peptide can be found through a process called high-throughput sequencing. This method is an easy way to survey all possible binding partners *in vitro*.
[38] A sequence inferred using Ancestral Sequence Reconstruction.

Structural knowledge is key to understanding protein evolution at the molecular level because it reveals the chemical determinants of binding specificity. The results of Wheeler's work shows that, despite low overall specificities, the S100A5 protein interacts with a set of peptides that is more divergent from the ancestral A5/A6 protein than S100A6 [7]. To gain structural insight into this evolution, we first used far-UV circular dichroism (CD) spectroscopy[39] to determine the pattern of structural change in response to calcium binding across the S100A5/A6 clade. Since these proteins undergo a conformational change upon calcium binding, differences in their secondary structure[40] would likely contribute to differences in their binding specificities. In fact, our results recapitulate the phylogenetic pattern of specificity determined by Wheeler's work. We found that the S100A5 and S100A6 amniote orthologs, in addition to their respective ancestors, exhibit a conserved pattern of secondary structure in both apo and calcium-bound states. However, since CD spectroscopy provides a relatively low resolution measurement of protein structure, we set out to solve the 3-dimensional structure of S100A5 when bound to a target peptide. This would identify the direct molecular determinants of specificity in this protein.

Crystallography and the subsequent back-calculation of protein structure through X-ray diffraction pattern analysis provide information such as binding sites and the amino acids that characterize them. There is currently no crystal structure of S100A5 bound to a target published, which limits our understanding of binding in this

[39] A method that provides spectroscopic data regarding the structure of optically active biomolecules.
[40] The local 3-dimensional structure of beta sheets, alpha helices or other forms adopted by a sequence of amino acids after translation. Such contributes to its tertiary structure, the formation of a globular unit.

protein. While previous experiments have shown that peptides bind in the hydrophobic clefts as expected [7], a three-dimensional model would prove so definitively. More importantly, it would show the specific contacts made between the protein and peptide, and allow for comparison to those in S100A6. Human S100A5 crystals were formed in the calcium-bound state and in the calcium and peptide-bound state during initial trials. However, further optimization is required in order to produce crystals large enough for x-ray diffraction. Such work is in progress and will reveal the amino acids directly involved in peptide binding and peptide binding specificity.

The second goal of this project was to characterize the influence that a biological environment has on a protein's binding specificity. Wheeler's experiments showed that S100A5 has the biochemical ability to interact with many partners [7]. We therefore set out to determine how this *in vitro* promiscuity is realized biologically. To date, two biological targets have been identified, allowing for incomplete comparison [7,11,12]. Our goal was to develop a relatively high throughput assay in order to add to this list. We developed the first co-immunoprecipitation[41] protocol for mouse S100A5 to be performed on the brains of sacrificed mice. Application of this protocol will identify biological targets of this protein *in vivo*[42] by physically pulling them from their biological environment, the olfactory bulb. The identification of new targets will allow for further characterization of peptide binding in S100A5 and facilitate a more complete understanding of how specificity has evolved along this lineage.

[41] The method of pulling a protein under study from its biological environment and analyzing the binding targets that come with it.
[42] Refers to experiments performed on a living organism or its tissue.

Results

S100A5 and A6 exhibit an evolutionary pattern of secondary structure

Far-UV Circular Dichroism (CD) Spectroscopy provides a signal specific to the secondary structure of a protein by measuring the difference between its absorption of left-handed circularly polarized light and right-handed circularly polarized light [17]. The secondary structures of S100 proteins are rich in alpha helices that shift upon calcium binding [6], allowing for a characterizing signal at this resolution. We use CD spectroscopy to measure the Ca^{2+}-induced conformation change in the ancestral S100A5/A6, S100A5, and S100A6 proteins as well as in their extant amniote orthologs. Spectra were recorded from 200 to 250 nanometers (nm) for each protein in both the apo and Ca^{2+}-bound states. Figure 5 compares the spectra of the human S100A5 (hS100A5) and human S100A6 (hS100A6) proteins.

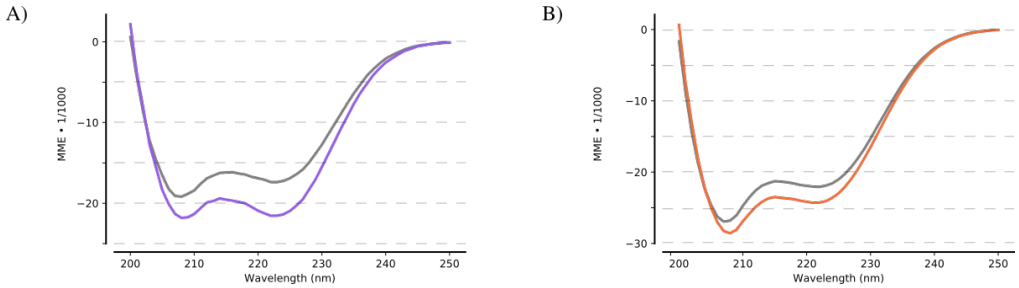

Figure 5. CD spectroscopy reveals differences in structural response to calcium in hS100A5 and hS100A6.

CD signals were recorded from 200 to 250 nm and are reported in MME (° cm $dmol^{-1}$ $resid^{-1}$) • 1/1000. A) Far-UV spectra collected for hS100A5 in the apo conformation (grey) and Ca^{2+}-bound conformation (purple). B) Spectra for apo hS100A6 (grey) and Ca^{2+}-bound hS100A6 (orange).

Even though both proteins show the general signature of the alpha helix spectrum, with minima at 208 and 222 nm, there are sharp differences in the response to Ca^{2+} binding.

Ca^{2+} saturation causes a 13.6% increase in helical content hS100A5 at 208 nm and a 24.0% increase at 222 nm. In contrast, hS100A6 only exhibits 6.6% and 10.0% changes at 208 and 222 nm, respectively. This pattern is conserved in each pair of S100A5 and S100A6 paralogs. Every S100A5 protein studied had a stronger response to Ca^{2+} than that of its A6 paralog (Table I).

	Percent Change at 222nm
Anc S100A5/A6	7.5
AltAll Anc S100A5/A6	10.1
Anc S100A5	12.7
human S100A5	24.0
mouse S100A5	14.9
t. devil S100A5	16.4
alligator S100A5	20.0
--S100A5 lost in chicken--	--
Anc S100A6	5.0
human S100A6	10.0
mouse S100A6	6.1
t. devil S100A6	9.0
alligator S100A6	11.7
chicken S100A6	6.0

Table I. S100A5 proteins exhibit a stronger response to calcium binding than both their S100A6 paralogs and the shared S100A5/A6 ancestor.

Percent changes in secondary structure between the apo conformation and Ca^{2+}-bound conformation were calculated from the CD signal at 222 nm. AltAll represents an alternative reconstruction of ancestral S100A5/A6.

We observed a general pattern for the curvature of the spectra of this clade that also resembled its phylogeny. Figure 6 overlays the data collected on the phylogenetic tree for S100A5 and S100A6 amniotes. Both the S100A5 and S100A6 extant orthologs show signals comparable to those of their respective ancestors. Additionally, the structural response to calcium in the ancestral A5/A6 protein is more similar to that of S100A6 than S100A5.

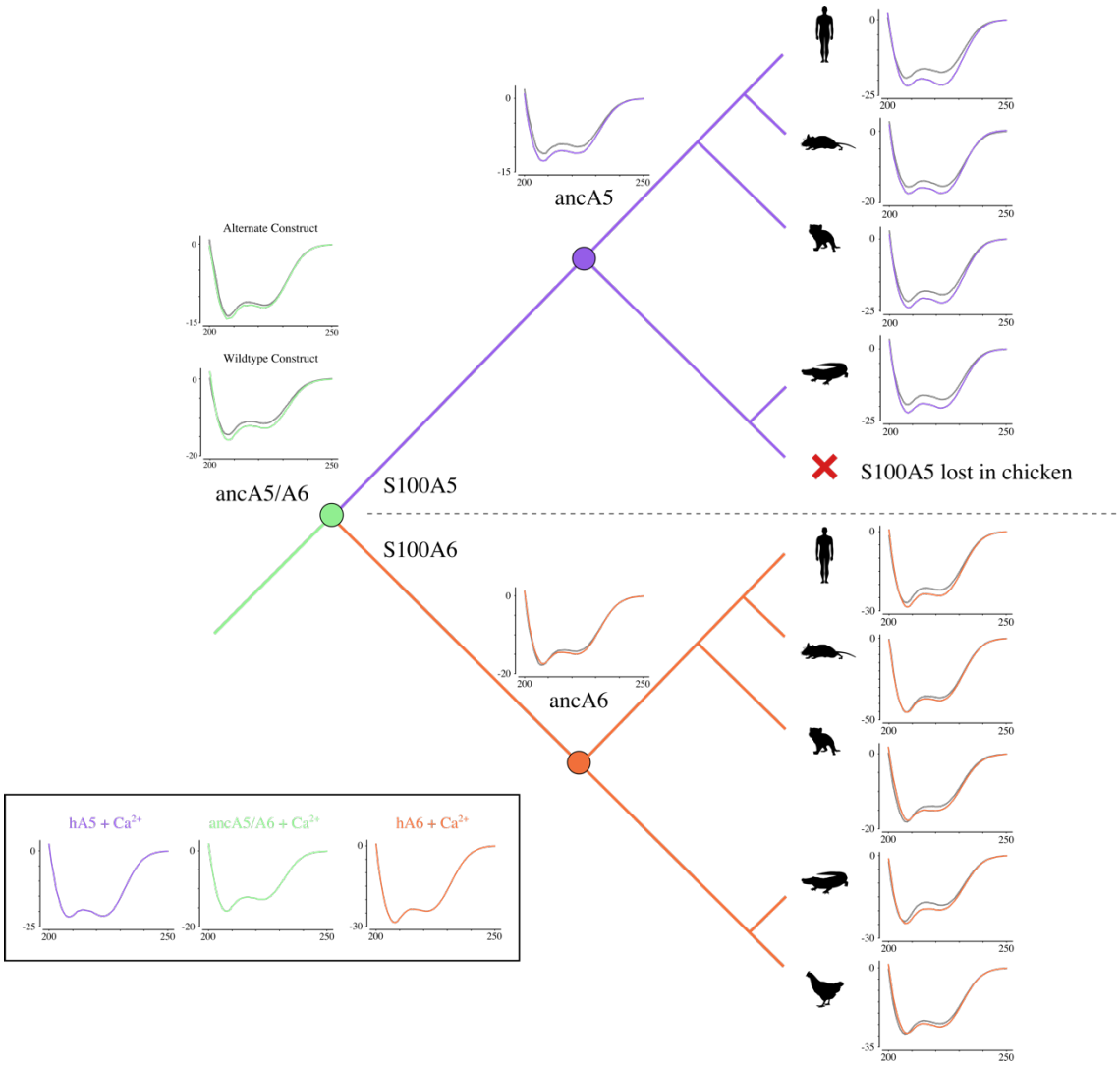

Figure 6. Far-UV CD Spectra recapitulate the phylogeny of the S100A5/A6 clade.

The far-UV CD spectra, reported in MME (° cm dmol^{-1} resid^{-1} • 1/1000), were recorded from 200 to 250 nm. Sequences used were extant S100A5 and A6 amniote orthologs in addition to reconstructed S100A5, S100A6, and S100A5/A6 ancestral sequences. The curves for the apo state of each protein are shown in gray, and the colors represent the Ca^{2+}-bound spectra for the S100A5 (purple), S100A6 (orange), and ancestral A5/A6 (green) proteins. The boxed inset compares spectra for calcium-bound human S100A5 and A6 extant paralogs (purple and orange, respectively) and the ancestral A5/A6 construct (green).

Our results indicate that the increase in response seen in the S100A5 ancestor and orthologs is a derived feature in this clade. The pattern of secondary structure is attributed to the mutations that occurred after the duplication of the ancestral S100A5/A6 gene. Notably, the pattern of binding specificity measured in Wheeler's isothermal titration calorimetry[43] (ITC) experiments also highlights the duplication event as the point of differentiation. Figure 7 summarizes Wheeler's ITC data alongside the results from our CD experiments.

[43] A method that measures heats of binding by comparing the heat released in the protein-containing sample cell during titration of ligand against an idle reference cell.

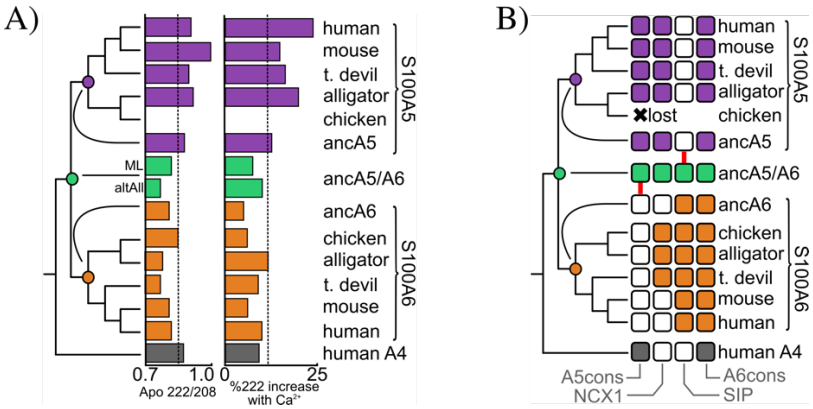

Figure 7. Secondary structure and binding specificity diverge after gene duplication of S100A5/A6 ancestor.

A) CD Spectroscopy data mapped onto the phylogeny of the S100A5/A6 clade. The left column shows the ratio of signal at 222 nm to signal at 208 nm in the apo conformation of S100A5 (purple), S100A6 (orange), ancestral S100A5/A6 (green), and human S100 A4 (grey). The right column shows the percent increase in signal upon calcium binding for S100A5 (purple), S100A6 (orange), ancestral S100A5/A6 (green), and human S100 A4 (grey). The dashed lines show mean values for each measurement; Results show a mean ratio of apo 222/208 (0.85) and mean increase at 222 nm upon calcium binding (11.6%) [7]. B) Luke Wheeler's ITC data mapped onto the phylogeny of the S100A5/A6 clade. Each square denotes binding of the peptide listed at the bottom and the ortholog listed to the right. The square is colored in if binding was observed and the measured affinity was greater than or equal to 100μM. The data for the ancestral protein is shown in green, with red line indicating changes in binding specificity post duplication. The data for S100A5 is shown in purple and the data for S100A6 is shown in orange. The outgroup, human S100A4, is again shown in grey [7].

The observed increase in response to Ca^{2+} binding in the S100A5 family correlates with its change in binding specificity. Such may also contribute to the sub-functionalization of S100A5 observed in Wheeler's phage display experiments. Ultimately, our results suggest that structure contributes to specificity for our model system, as is expected. However, they do not implicate secondary structure as the definitive cause.

X-ray crystallography will reveal the chemical determinants of specificity in human S100A5

The amino acids that populate a binding interface determine a protein's relative affinities for different interaction partners. For this reason, it is likely that the differences in S100A5 and A6 specificities are at least in part due to differences in the biochemistry of their binding sites. X-ray crystallography has been used to produce 3-dimensional pictures of biological macromolecules at a resolution high enough to reveal their molecular compositions and overall structures. We set out to crystallize hS100A5 bound to a target peptide isolated from Wheeler's phage display experiments to determine the amino acids involved.

We first screened ninety-six initial conditions for samples of hS100A5 containing both Ca^{2+} and a target peptide. Planar crystals were formed in a 100 nanoliter (nL) droplet above 100 microliters (μL) of 2.0 M[44] $(NH_4)_2SO_4$ and 0.1 M Tris-HCl, pH 8.5 (Figure 8A, left). Optimization around $(NH_4)_2SO_4$ concentration and pH in 10 μL droplets hanging above a 1-mililiter (mL) reservoir resulted in the final crystal shown in Figure 8A, right. This crystal was harvested and submitted for x-ray diffraction at the Advanced Light Source in Berkeley, California, but the results were inconclusive. The diffraction pattern did not indicate a protein crystal nor did it identify a salt crystal. Possible explanations might be that the crystal contained a protein aggregate, or that the quality of the crystal was too poor to produce a diffraction pattern indicative of a protein complex.

[44] Molarity; A measurement of concentration described by the number of moles of solute divided by the volume of the solution in liters.

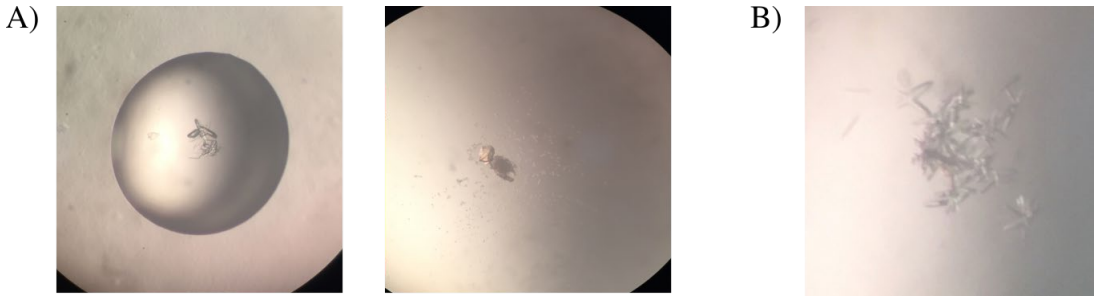

Figure 8. Initial screens reveal crystallography conditions for optimization.

All droplets were visualized using a Nikon microscope at 4 °C. A) A solution containing 1.4 milimolar (mM) hS100A5, 7.5 mM Ca^{2+}, and mM NewA5con peptide was diluted 1:1 with the reservoir solution (2.0 M $(NH_4)_2SO_4$ and 0.1 M Tris-HCl, pH 8.5) (left). Optimization around [$(NH_4)_2SO_4$] and pH resulted in the final crystal (right) formed from a solution of 1.5 mM hS100A5, 7.5 mM Ca^{2+}, mM peptide above 2.8 M $(NH_4)_2SO_4$ 0.1 M Tris-HCl, pH 8.5. B) Microcrystals formed in a 100 nL droplet of 3.5 mM hS100A5, 7.5 mM Ca^{2+} diluted 1:1 with 0.05 M CsCl, 30% (v/v) Jeffamine M-600, 0.1 M MES monohydrate, pH 6.5.

We turned to a peptide-soaking method in hope of improving our chances of crystallization. Microcrystals of Ca^{2+}-bound hS100A5 were formed in our initial screen in a 100-nL sized droplet hanging above 0.05 M CsCl, 30% (v/v) Jeffamine M-600, 0.1 M MES monohydrate, pH 6.5 (Figure 8B). Since microcrystals suggest that the nucleation occurred too rapidly, often due to high protein concentration [xx], we optimized around concentrations of protein and precipitating agent, Jeffamine M-600. This process is ongoing and has not yet resulted in crystals large enough to soak with the target peptide.

Mouse S100A5 is precipitated in calcium-saturating conditions

Luke Wheeler's high-throughput phage display experiments generated sets of biochemically possible interaction partners for A5, A6, and ancestral A5/A6 by surveying their affinities for synthetic peptides of various sequences [18]. Random

sampling from a library of all of the potential sequences for a twelve amino acid long peptide generated interaction "fingerprints" for each protein [18]. We aimed to elucidate the relationship between this "biochemically possible" set and the "biologically realized" set by identifying targets of S100A5.

Co-immunoprecipitation involves the physical pulling down of a protein from a solution through interaction with microscopic magnetic beads. When performed on a cell culture, one can simultaneously isolate the binding partners of the protein under study. We developed the first known co-immunoprecipitation protocol of mouse S100A5 (mS100A5) with the primary antibody, polyclonal rabbit anti-S100A5. This antibody was used in a previous immunohistochemical analysis of olfactory sensory cilia [13]. Antibody concentrations and blocking conditions for western blot[45] analysis were optimized using a purified sample of mS100A5 (Figure 9A-C). Mouse S100A5 was successfully precipitated even in calcium-saturating conditions. This is important for the identification of targets of this protein, since they bind in a calcium dependent manner.

[45] A method for selective detection of the location of a protein after transfer from an SDS-PAGE gel. Allows for the identification and quantification of samples that contain the desired protein.

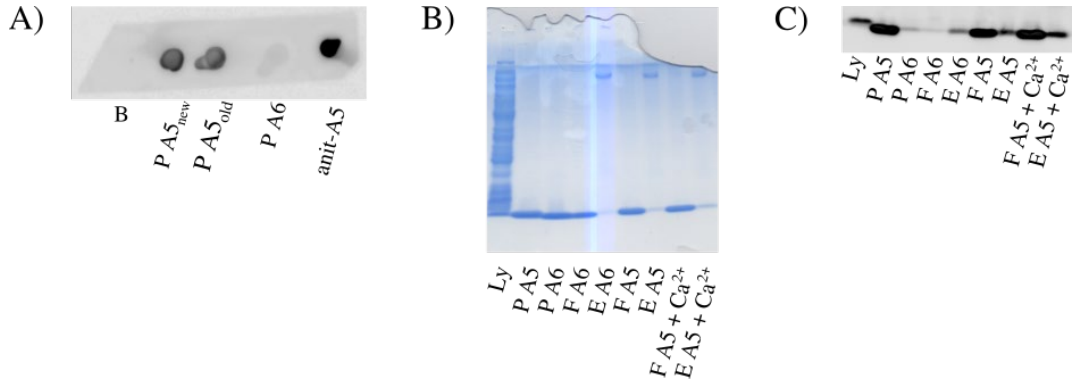

Figure 9. Calcium saturation allows for a greater yield of mouse S100A5 from a purified sample.

A) A western blot of the following: B- buffer; $PA5_{new}$- recently purified mS100A5; $PA5_{old}$- old prep of mS100A5 used in prior optimizations; PA6- purified mS100A6; anti-A5- primary antibody; polyclonal rabbit anti-S100A5. B) An SDS-PAGE containing: Ly- lysate; PA5- pure mS100A5; PA6- pure mS100A6; FA6- mS100A6 flow through; EA6- mS100A6 elution; FA5- mS100A5 flow through; EA5- mS100A5 elution; FA5 + Ca^{2+}- flow through for mS100A5 saturated with calcium; EA5 + Ca^{2+}- elution for mS100A5 saturated with calcium. C) A western blot of the same samples shown in 9B, labeled with the same abbreviations.

Since the ultimate goal was to perform the co-immunoprecipitation of mS100A5 using tissue from the mouse olfactory bulb, it was important to test the validity of the protocol in more complex conditions. We used a bacterial lysate containing overexpressed mS100A5 to confirm that our chosen antibody and its concentration allow for selective immunoprecipitation of this protein (Figure 10).

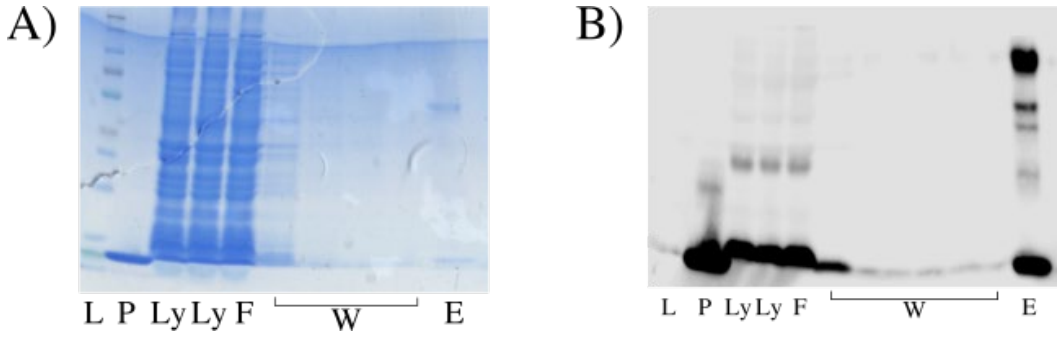

Figure 10. Mouse S100A5 is selectively immunoprecipitated from a bacterial lysate.

A) An SDS-PAGE gel containing the following samples: L- protein ladder; P- purified mS100A5; Ly- input bacterial lysate; F- immunoprecipitation flow through; W- subsequent washes; and E- elution. B) A western blot of the same samples shown on the SDS-PAGE gel, labeled with the same abbreviations.

Polyclonal rabbit anti-S100A5 shows cross-reactivity with mouse S100A6

As shown in Figure 9, our initial results indicated that the primary antibody exhibits cross-reactivity with mouse S100A6 (mS100A6). To determine the selectivity of our protocol, we performed a titration[46] analysis for both paralogs at varying input concentrations. Our antibody exhibits a greater affinity for mS100A5 (Figure 11). All trials were performed with the same antibody concentration. As shown above, mS100A5 precipitates at concentrations on the 0.1mM scale (Figure 11A). In contrast, it appears that a noticeable yield of mS100A6 requires a minimal concentration of 3mM (Figure 11B).

[46] Addition of a solute into a solution in a slow and controlled manner. Often useful for measurements characterized by the concentration of solute.

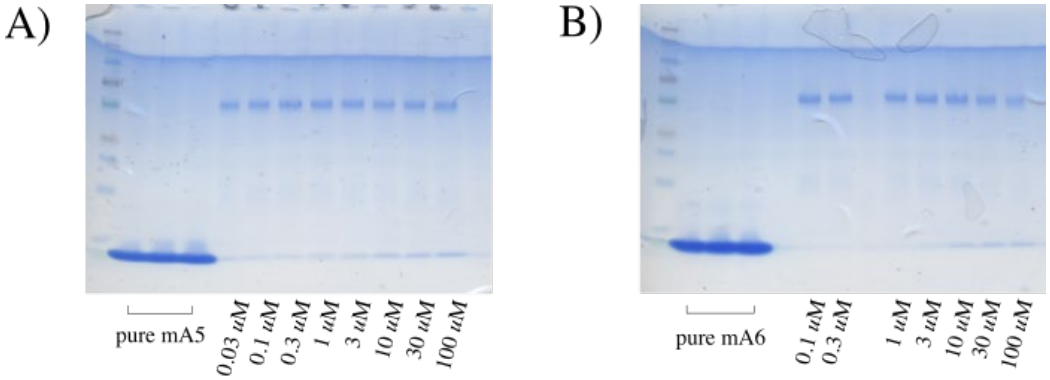

Figure 11. Despite cross-reactivity, polyclonal rabbit anti-S100A5 exhibits selectivity for mouse S100A5.

Eluted samples from immunoprecipitation experiments for different protein concentrations are shown for mouse S100A5 (A) and mouse S100A6 (B). Concentrations increased from 0.03μM to 100μM for mS100A5 and 0.1μM to 100μM for mS100A6. Three pure samples of known protein concentration were analyzed in both experiments to control for unequal intensity of staining. For the subsequent lanes, the bands at the top of the gel represent the antibody, and the bands at the bottom of the gel represent the protein eluted.

We used the ImageJ processing software to integrate the intensity of the SDS-PAGE gel bands. The resulting value would serve as a surrogate for protein concentration. A plot of fraction of protein eluted versus input concentration is shown for both paralogs in Figure 12. Our results show a progression toward saturation of the antibodies with increasing protein concentration as is expected.

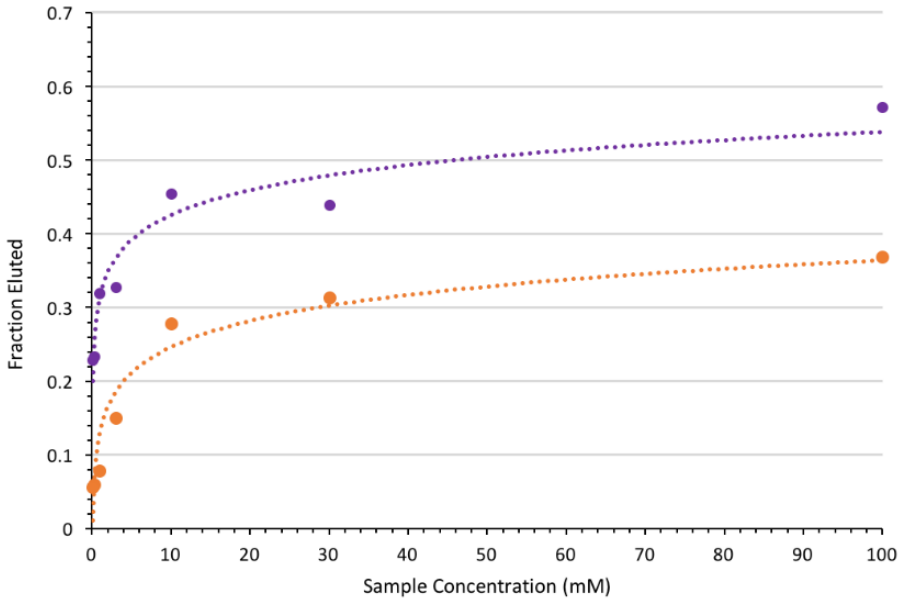

Figure 12. Polyclonal rabbit anti-S100A5 exhibits a 6.9-fold selectivity for mouse S100A5 over mouse S100A6.

Fraction eluted was calculated by dividing the intensity of protein eluted over the intensity of antibody eluted. Intensities were calculated using the ImageJ processing software. Data points and the logarithmic fits are shown for mouse S100A5 (purple) and mouse S100A6 (orange), with the following equations: $y = 0.049\ln(x) + 0.3124$ ($R^2 = 0.932$) for mouse S100A5 and $y = 0.051\ln(x) + 0.1287$ ($R^2 = 0.933$) for mouse S100A6.

It is important to note that the maximal fraction eluted for both mS100A5 and mS100A6 is below saturation. We back-calculated maximal fractions eluted to be 0.54 and 0.36 for mS100A5 and A6, respectively. Therefore the antibody shows some selectivity for mS100A5 over mS100A6, though not much.

Discussion

Secondary structure and S100 binding specificity

Both S100A5 and S100A6 have been previously shown to reorient helices in their structures and expose a hydrophobic interface upon calcium binding [6-10]. Our data affirm this mechanism and shed light on the different responses in these paralogs. The resulting exposure of these binding pockets increases the alpha helical signal in the Ca^{2+}–bound states of both proteins. This process is characteristic of every ortholog studied in this experiment. However, S100A5 proteins exhibit a stronger response to Ca^{2+} binding than their A6 paralogs. This is likely due to the induced elongation of the C-terminal helix of the S100A5 protein [11]. More importantly, this pattern is conserved along the A5 lineage, suggesting that the structural response to calcium binding is an evolutionary feature of our model system.

Ultimately, the pattern of structural response to calcium binding in extant S100A5 and A6 amniotes recapitulates both Wheeler's binding specificity data and the phylogenetic relationship of this clade. Additionally, the S100A5/A6 ancestor shows a structural response more similar to the S100A6 proteins than to S100A5, indicating that mutation post gene-duplication of the S100A5/A6 common ancestor likely contributed to the development of different secondary structure in the S100A5 ancestor and extant proteins. This agrees with Wheeler's proposed theory of sub-functionalization in S100A5. Our results suggest that secondary structure is important to the pattern of specificity we see in the S100A5/A6 clade. However, they do not identify the mechanism behind its contribution.

Chemistry at the binding interface

Crystallization of biological macromolecules is a stochastic process. So far, optimization of initial screens has failed to produce crystals of sufficient quality for downstream analysis and structure determination. However, we have identified promising conditions in our initial screens. Moving forward, we expect that the peptide-soak method will be more successful. The crystallization of S100A5 is much simpler without the addition of a binding target. If we can crystallize the S100A5 protein in the Ca^{2+}-bound conformation, we hypothesize that our target peptide will diffuse into the protein's binding pockets upon careful addition to the crystal in solution.

The amino acids involved in binding provide the chemical basis for specificity through structure, charge, and hydrophobicity. Wheeler has shown that binding in S100A5 is driven primarily by the hydrophobic effect [7]. This means that packing is more important to binding in this protein than electrostatics [7]. The consequence is that mutations that alter the structure of the binding interface could easily yield changes in binding specificity. In fact, previous experiments have shown that a single amino acid reversion to the ancestral state within the binding interface is sufficient to alter the binding specificity in S100A5 [7]. In contrast, the crystal structure of S100A6 bound to the SIP peptide reveals that S100A6 employs both hydrophobic and electrostatic interactions to bind its targets. The differences between S100A5 and A6 may be responsible for the differences in binding specificity that we have observed.

Identification of biological targets will help to characterize S100A5

We developed a co-immunoprecipitation protocol for mS100A5 to compare the range of its biological targets with the set of biochemically feasible interaction partners.

It is useful that our protocol is successful in calcium-saturating conditions, since peptide binding in S100A5 requires allosteric activation by Ca^{2+}. Furthermore, the additional bands in the elution sample for co-immunoprecipitation out of bacterial lysate (Figure 10) suggest that the antibody does not inhibit binding of mS100A5 by competing for the active site or inducing other steric interactions. However, of those bands, we do not know which targets were precipitated through binding with mS100A5 as desired versus which may have interacted directly with the antibody. Additional controls and downstream biochemical characterization experiments will be needed to confirm that the species isolated are targets of S100A5. A pulldown out of bacterial lysate in which expression of mS100A5 is not induced will reveal the targets precipitated in the absence of the protein under study. Subtraction of this control will be important when performing the co-immunoprecipitation protocol on mouse olfactory tissue.

It is also important to consider the antibody's cross-reactivity with mS100A6. Ultimately, our results suggest that 60% of the species precipitated with this antibody are due to interaction with mS100A5. However, this confidence threshold assumes that expression levels of mS100A5 and A6 in the olfactory bulb are comparable, when in fact expression of mS100A5 in this tissue is substantially greater than that of its paralog [13]. Our confidence that the species precipitated are targets of this protein is therefore increased by a factor proportional to the expression upregulation.

With appropriate controls, our protocol will isolate biological targets of mS100A5. Eluted samples will be submitted for analysis by proteomics[47]. The ultimate

[47] A high throughput method for separating and identifying proteins from a populous sample.

goal is to measure the effects that activation of the olfactory signaling cascade has on the library of S100A5 interaction partners. The identification and characterization of these targets will improve our understanding of the biology of S100A5 as well as of the olfactory system.

Environmental factors may produce biological specificity from biochemical promiscuity

Our work demonstrates the powerful influence that a biological environment has on proteins. S100A5 functions in the cilia of olfactory sensory neurons and is associated with the cell membrane [13]. Though Wheeler's phage display experiments show relatively low specificity in S100A5, its gene is significantly upregulated upon olfactory activation [7,13]. This suggests that specific function for this protein is important despite its low overall specificity. It is likely that expression regulation and protein localization transform S100A5 from a biochemically sloppy protein into an essential component of olfactory signaling. Our co-immunoprecipitation protocol coupled with downstream proteomics will reveal the biological targets for S100A5 and provide insight into how this set arises from and compares to the biochemical set provided by phage display. Additionally, the results of such experiment will reveal the mechanism by which the latent binding partners of S100A5 and A6 can shape the evolution of these proteins. Gene mutations in DNA can lead to sequence changes in the respective protein. Changes in sequence can yield changes in function, and thereby open or close evolutionary paths.

Binding specificity, evolutionary biochemistry, and bioengineering

Both S100A5 and A6 are implicated in multiple necessary biological functions, thus insight into their evolution is useful for understanding their present contributions to biology and disease. More broadly, however, evolutionary study of S100A5 and A6 serves as an excellent model for the evolution of binding specificity in general. Our evolutionary study will reveal general engineering principles for specificity, which is essential to the utility of drug therapeutics. Such will help to produce safer and more effective drug therapies for a wide range of diseases. Understanding of the mechanism for S100A5 and A6 evolution helps us to explain why these proteins have the functions that they do, but the themes identified will prove useful beyond the scope of this subfamily.

Methods

Gene Cloning and Protein Expression/Purification

S100 protein expression was induced[48] in a culture of bacteria. The desired genes were cloned into a pET28/30 vector and then absorbed by Rosetta (DE3) pLysS E. coli cells. 1.5 liter cultures were inoculated with 20 μL of saturated overnight culture and grown to high log-phase (OD_{600} ~ 0.6-1.0) by shaking at 250rpm at 37 °C. After induction with 1mM IPTG and addition of 0.2% glucose, cultures were stabilized overnight at 16 °C with shaking at 250rpm. Cells were harvested by centrifugation and frozen at -20 °C before lysis and purification. All genes were codon-optimized for expression in E. coli and contained an N-terminal His tag with a TEV protease cleavage sit (Millipore). This cleavage site was used to purify the desired protein from the mixture of unwanted protein, DNA, and other macromolecules in the bacterial lysate. Proteins can be purified through manipulation and exploitation of their characteristics.

Cells were lysed via sonication in 25mM Tris (Fisher Chemical), 100mM NaCl (Fisher Chemical), 25mM imidazole (Fisher Chemical), pH 7.4 with addition of 2 μL of Dnase (Thermo Scientific) and lysozyme (Thermo Scientific) per μg cells. We then employed a three-step purification process using an Äkta PrimePlus FPLC (GE Health Science) to fraction and isolate the S100s. First, we eluted the proteins using a 25-mL gradient from 25 to 500mM imidazole in 25mM Tris, 100mM NaCl, pH 7.4 on a 5-mL HiTrap Nickel-affinity column (GE Health Science). A chromatogram revealed the

[48] A term referring to the process by which cells are triggered to express certain genes in the presence of a signal molecule.

peak fractions which were pooled and incubated overnight at 4 °C with about 1:5 TEV protease (produced in lab). For the second step, hydrophobic interaction chromatography (HIC) was used to further purify the S100s. After saturation with 2mM Ca^{2+}, the solutions were eluted over a 5-mL HiTrap phenyl-sepharose column (GE Health Science) in a 25-mL gradient from 0mM to 5mM EDTA in 25mM Tris, 100mM NaCl, pH 7.4. Peak fractions were pooled and dialyzed against 4 L of 25mM Tris, 100mM NaCl, pH 7.4 overnight at 4 °C. The final step involves another passage over the 5-mL HiTrap Nickel-affinity column (GE Health Science) to remove His-tagged protein that was not cleaved during the incubation with TEV protease. Proteins were collected in the flow through, and the remaining contaminants were eluted from the column using a step to the 500mM imidazole buffer.

Purity was verified using a method called gel electrophoresis. In this process, the proteins are loaded at the top of an SDS-PAGE gel. The SDS is a detergent that denatures or unfolds the proteins and coats them with a negative charge. Because of this charge, the proteins in solution will move across the gel during the application of an electrical gradient. The method separates the proteins based on size because the bigger proteins move slower. The gels are then stained with Bio-Safe Coomassie G-250 Stain (Bio-Rad) for analysis, resulting in blue bands identifying the location of each protein. The amount and purity of the desired proteins were visualized by comparison with a broad range protein ladder (Thermo Scientific).

Far-UV CD Spectroscopy

Protein samples were exchanged into a chelex (Bio-Rad) treated, 25mM TES (Sigma-Aldrich), 100mM NaC (Fisher Chemical), pH 7.4 buffer and concentrated by centrifugation at 4500 x g at 16 °C in a temperature-controlled centrifuge (Eppendorf) using 15-mL 3K Microsep Filter Centrifuge Conicals (Pall). Far-UV CD spectra were collected from 200 to 250nm on a J-815 CD spectrometer (Jasco) with a 1-mm quartz cell (Starna Cells, Inc.). Spectra were recorded for samples in the absence and presence of 1mM Ca^{2+}. Saturation with 2mM verified the reversibility of the Ca^{2+}-induced conformation change. A buffer blank spectrum was subtracted with the built-in subtraction feature of the Jasco spectra analysis software. Raw ellipticity data were converted to mean molar ellipticity after this subtraction.

Hanging Drop Crystallization

To crystallize a protein out of solution, one manipulates protein/solvent system parameters to slowly reduce the solubility of the protein, often by increasing the concentration of the solutes[49]. The hanging drop method involves a droplet that hangs above its reservoir solution in a closed system called a "well." The drop, held on a glass coverslip or other surface, contains the protein-peptide solution and is diluted one-to-one with the reservoir solution in the well. The droplet contains the desired protein (with its binding partner) and the same solutes as in the reservoir, just in lower concentration. All biological systems prefer to be at equilibrium[50], or in balance. Thus

[49] The compound or molecule that is dissolved in a solution.
[50] When all conditions or opposing forces are equal. Biological systems prefer this state because it is low in energy. Just as ice melts in water to form a solution of

over time, water diffuses from the droplet to the reservoir in order to equalize the solute concentrations and move the system toward equilibrium. As water leaves the drop, the protein becomes more concentrated and will eventually crash out of solution. If the conditions are correct, this will happen through the aggregation and ordering of protein molecules into a crystalline solid.

We used the hS100A5 C43SC79S double mutant for this process because it removed the potential for disulfide bonds[51]. These mutants do not affect the function of the protein, but simply make it easier to work with in the lab. Purified samples of hA5 C43SC79S were dialyzed for several days into 4 L of 1mM Hepes (Sigma-Aldrich) pH 7.5 at 4 °C and then concentrated to ~ 6mM by centrifugation in a temperature-controlled centrifuge (Eppendorf) at 4500 x g in 15-mL 3K Microsep Filter Centrifuge Conicals (Pall). The peptide used was the A5consensus peptide identified in Wheeler's phage display (mw = 2374.5 g/mol) [7]. Initial screens were done using a 96-well plate (Thermo Scientific) and drops were laid using the Mosquito Crystal Robot (Ttplabtech). Optimization trials were performed using 6 x 4 well plates (Hampton Research) and 22 mm siliconized glass cover slides (Hampton Research). Reagents for optimization included $(NH_4)_2SO_4$ (Fisher Chemical), Tris-HCl (Fisher Chemical), CsCl (Fisher Biotech), and Jeffamine M-600 (Hampton Research). Trays were sealed and stored at 4 °C for at least 5 days before examination using a Nikon Microscope. Crystals harvested

---

even temperature, the system in hanging drop crystallization wants to have solutions at uniform concentration.

[51] A strong bond formed between two Sulfide atoms. This is common in proteins that have the amino acid cysteine, because it contains a Sulfide in its side chain.

were flash frozen in liquid nitrogen and submitted for diffraction pattern collection at the Advanced Light Source in Berkeley, California.

Co-immunoprecipitation

For experiments with pure protein, samples were exchanged into 25mM TES (Sigma-Aldrich), 100mM NaCl (Fisher Chemical). For experiments with mS100A5 bacterial lysates, pellets were lysed in 5 mL B-PER (Thermo Scientific) and 10 μL Dnase (Thermo Scientific) and lysozyme (Thermo Scientific) per 1 g cells. 25 μL of Protein G magnetic beads (Thermo Scientific) were washed with 175 μL of TBS-T (20mM Tris-HCl, 500mM NaCl, pH 7.5, containing 0.05% Tween), separated in a magnetic rack, then washed in 1 mL of TBS-T. Beads were primed with 5-10 μg of the primary antibody (polyclonal rabbit anti-S100A5; Proteintech 17924-1-AP) in 500 μL of "lysis buffer" (TES buffer of B-PER) by a 1-hour incubation at room temperature with slow stirring. After isolation of the beads by magnet and removal of the supernatant, 500 μL of each sample was incubated with the antibody-bound beads overnight at 4 °C with slow stirring. Beads were then isolated by magnet and the flow through was removed and saved for analysis. Beads were washed 3 times with 1 mL TBS-T and 1 time with 1 mL ddH_2O. Samples were eluted with 1x Laemmli (Bio-Rad) and 2.5% BME (Sigma-Aldrich) in the lysis buffer for 2 hours at room temperature with slow stirring. Results were visualized by SDS-PAGE and western blot.

Western Blot Analysis

After electrophoresis, SDS-PAGE gels were transferred to a nitrocellulose membrane for 1 hour at a constant voltage of 110 V in 4 °C. The gel, membrane, and

transfer materials, were soaked in transfer buffer (25mM Tris-HCl, 192mM glycine, 20% MeOH) prior to assembly of the apparatus. The membrane was washed in TBS-T with shaking for 5 minutes and then blocked in 20% Seablock (Thermo Scientific)/TBST-T for 1 hour at room temperature with shaking to mask non-specific interactions. After washing 3 times with TBST-T, the membrane was incubated with the primary antibody (polyclonal rabbit anti-S100A5; Proteintech 17924-1-AP) at a 1:2000 dilution in blocking buffer overnight at 4 °C with shaking. The membrane was washed 3 times in TBS-T then incubated with the secondary antibody (goat anti-rabbit IgG H & L HRP; abcam ab205718) at a 1:25000 dilution in blocking buffer for 1 hour at room temperature with shaking. The membrane was washed 3 times in TBS-T then 2 times in TBS to remove excess tween. Luminol Reagent Solutions A and B (Santa Cruz Biotechnology) were added in a 1:1 amount to the membrane and incubated with shaking for 2 minutes. The membrane was imaged using a LI-COR system with 2-minute exposures at 600, 700, and 800nm and a 10-minute chemiexposure.

Data Analysis and Intensity Integration

Scanned images of SDS-PAGE gels were processed using the ImageJ software. Vertical boxed sections of each well were integrated to provide the color intensity. Areas under peaks, which represented bands on the gel, were calculated using the software after subtraction of noise after linear fit. Fraction eluted was calculated by dividing the area representative of the mS100A5 elution band from the area representative of the antibody elution band. Logarithmic fits for the data in Figure 12 were calculated using Microsoft Excel.

Glossary

Allosterically Regulated: When binding at one site of a protein has an effect on binding at another site, often contributing to regulation of function.

Amino Acids: An organic compound that contains a carboxyl and an amino group. Generally, there are twenty amino acids that make up biological proteins. The linear arrangement of these amino acids is called a sequence.

Amniotes: A mammal, bird, or reptile. Amniotes are vertebrate animals that have an amnion (membrane enclosing the embryo) during the embryonic stage.

Ancestral Sequence Reconstruction: The process by which the sequence of an ancestral protein is inferred from present data.

Apo: A state in which the proteins under study are completely neglect of calcium.

Apoptosis: Cell death. This is a regulated process and normal to the physiological function and life cycle of a cell.

Bacterial Lysate: The product of the physical destruction of bacterial cells. A cell is lysed so that its walls are broken down and the proteins and molecules inside are accessible.

Binding Affinity: The strength of the interaction between a protein and its binding partner. This is determined by the elements involved and their chemical properties. The symbol for binding affinity is K_D and the units are M^{-1}, where M is molarity. (See Molarity).

Biomolecules: Any molecule involved in biology or in a living organism. Can be a protein, lipid, or vitamin, etc.

Cilia: A hairlike structure that occurs on the surfaces of receptor cells.

Clade: A group of organisms related by a common ancestor. All the species or homologs that have evolved after a "node" in a phylogenetic diagram.

Co-immunoprecipitation: The method of pulling a protein under study from its biological environment and analyzing the binding targets that come with it.

Disulfide Bond: A strong bond formed between two Sulfide atoms. This is common in proteins that have the amino acid cysteine, because it contains a Sulfide in its side chain.

DNA: Deoxyrybonucleic acid, the self-replicating material that acts as the carrier of genetic information in the cell.

Electrostatic: Interactions involving electric charges on amino acids or other biomolecules.

Equilibrium: When all conditions or opposing forces are equal. Biological systems prefer this state because it is low in energy. Just as ice melts in water to form a solution of even temperature, the system in hanging drop crystallization wants to have solutions at uniform concentration.

Extant: The members of a protein family that are still in existence.

Far-UV Circular Dichroism Spectroscopy: A method that provides spectroscopic data regarding the structure of optically active biomolecules.

Folding: The process of forming a 3-dimensional structure that a protein undergoes after translation.

Gene Duplication: Duplication of a gene in DNA so that there are two copies of the same gene. One copy can acquire mutations thus leading to a different protein.

Genetic Mutation: An alteration in the DNA sequence of a gene.

Homeostasis: The relatively stable equilibrium of elements, essential to maintenance of biological processes.

Homodimers: Proteins that are made of two identical subunits. The subunits are separate sequences of the same amino acids but are held together through molecular attractions.

Homology: Similarity due to shared ancestry.

Hydrophobic: The tendency of a nonpolar protein, amino acid, or biomolecule to fail to mix with water.

Immunity: An organism's ability to fight infection and disease. There is both innate and adaptive immunity; both contribute to the health and longevity of the organism.

In vitro: Refers to experiments performed in a test tube or culture dish, etc.

In vivo: Refers to experiments performed on a living organism or its tissue.

Induced: A term referring to the process by which cells are triggered to express certain genes in the presence of a signal molecule.

Inflammation: A biological response that causes localization of redness and swelling due to injury or infection.

Isothermal Titration Calorimetry: A method that measures heats of binding by comparing the heat released in the protein-containing sample cell during titration of ligand against an idle reference cell.

Ligand: A molecule that binds to a protein or other molecule.

Metabolism: The cellular process of breaking down resources to use for energy or in the building of other molecules. Any chemical process involving energetic absorption or exertion and the maintenance of life.

Molarity: A measurement of concentration described by the number of moles of solute divided by the volume of the solution in liters.

Monomer: A single unit of a protein.

Olfactory Bulb: Neural structure involved in olfaction (smell).

Olfactory Receptor Neurons: Cells that transduce signals from the external environment to the rest of the olfactory system.

p53: An infamous gene linked to many different types of cancers.

Paralogs: Gene "relatives" that evolved through gene duplication.

Peptide: A chain of two or more amino acids. Can have arbitrary length but does not normally possess enzymatic function.

Phage Display: Use of the random peptide sequences displayed by bacteriophages to survey peptide binding in proteins. The phages that contain bound peptide sequences can be isolated and replicated. Ultimately, the sequence of the peptide can be found through a process called high-throughput sequencing. This method surveys all possible binding partners *in vitro*.

Phylogenetic Tree: A diagram that uses lines and branches to describe the relation of a group of genes, species, or both. The tree can often have a temporal organization, showing the present genes and their relationships to each other and their ancestors.

Phylogenetics: The study of biological evolution. In evolutionary biochemistry, phylogenetics involves the study of evolution at the molecular level.

Phylogeny: The evolutionary path of the ancestral and extant members of a protein family.

Proliferation: Cell replication and division. This is a normal part of the life cycle of most cells, and is regulated by numerous factors. Uncontrollable proliferation is a characteristic of tumor cells.

Protein Sequence: The linear organization of amino acids that is unique to the protein. This determines the structure and function/ability of the protein. (See Amino Acids).
Proteins: Molecules made of amino acids in a linear chain that fold into a structure to perform many biological functions.

Proteomics: A high throughput method for separating and identifying proteins from a populous sample.

RNA: Ribonucleic acid, the result of DNA transcription, encodes protein sequence.

Reconstructed: A sequence inferred using Ancestral Sequence Reconstruction.

SDS-PAGE Gel: The polyacrylamide gel used in SDS-PAGE analysis. The acronym SDS stands for sodium dodecyl sulfate, and it is a negatively charged detergent. The acronym PAGE stands for polyacrylamide gel electrophoresis.

Secondary Structure: The local 3-dimensional structure of beta sheets, alpha helices or other forms adopted by a sequence of amino acids after translation. Such contributes to its tertiary structure, the formation of a globular unit.

Solutes: The compound or molecule that is dissolved in a solution.

Speciation: The evolution of a new species.

Sub-functionalization: Degenerate mutations that result in a gene and its duplicated copy sharing the burden of one function [4].

Titration: Addition of a solute into a solution in a slow and controlled manner. Often useful for measurements characterized by the concentration of solute.

Orthologs: Gene “relatives” that evolved through speciation.

Westerm Blot: A method for selective detection of the location of a protein after transfer from an SDS-PAGE gel. Allows for the identification and quantification of samples that contain the desired protein.

Bibliography

[1] Sikosek, T. & Chan, H. S. Biophysics of protein evolution and evolutionary protein biophysics. J. R. Soc. Interface 11, 20140419–20140419 (2014).

[2] Copley, S. D. An evolutionary biochemist's perspective on promiscuity. *Trends in Biochemical Sciences* **40,** 72–78 (2015).

[3] Harms, M. J. & Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* **14,** 559–571 (2013).

[4] Soskine, M. & Tawfik, D. S. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* **11,** 572–582 (2010).

[5] Wheeler, L. C., Donor, M. T., Prell, J. S. & Harms, M. J. Multiple evolutionary origins of ubiquitous Cu2+ and Zn2+ binding in the s100 protein family. *PLoS One* **11,** (2016).

[6] Donato, R. *et al.* Functions of S100 proteins. *Curr. Mol. Med.* **13,** 24–57 (2013).

[7] Wheeler, L.C., *et al*. Conservation of specificity in two low-specificity proteins. bioRxiv. https://doi.org/10.1101/207324.

[8] Zimmer, D. B., Eubanks, J. O., Ramakrishnan, D. & Criscitiello, M. F. Evolution of the S100 family of calcium sensor proteins. *Cell Calcium* **53,** 170–179 (2013).

[9] Bertini, I. *et al.* Solution structure and dynamics of S100A5 in the apo and Ca2+-bound states. *J. Biol. Inorg. Chem.* **14,** 1097–107 (2009).

[10] Lee, Y. T. *et al.* Structure of the S100A6 complex with a fragment from the C-terminal domain of Siah-1 interacting protein: A novel mode for S100 protein target recognition. *Biochemistry* **47,** 10921–10932 (2008).

[11] Kim, I. *et al.* Biophysical characterization of Ca2+-binding of S100A5 and Ca2+-induced interaction with RAGE. *Biochem. Biophys. Res. Commun.* **483,** 332–338 (2017).

[12] Liriano MA. Structure, Dynamics and Function of S100B and S100A5 Complexes [Ph.D.]. University of Maryland, Baltimore. United States – Maryland; 2012.

[13] Kuhlmann, K. *et al.* The Membrane Proteome of Sensory Cilia to the Depth of Olfactory Receptors. *Mol. Cell. Proteomics* **13,** 1828–1843 (2014).

[14] Schäfer, B. W. *et al.* Brain S100A5 is a novel calcium-, zinc-, and copper ion-binding protein of the EF-hand superfamily. *J. Biol. Chem.* **275,** 30623–30630 (2000).

[15] Serizawa, S. *et al.* A Neuronal Identity Code for the Odorant Receptor-Specific and Activity-Dependent Axon Sorting. *Cell* **127,** 1057–1069 (2006).

[16] Gomez, G., Rawson, N. E., Hahn, C. G., Michaels, R. & Restrepo, D. Characteristics of odorant elicited calcium changes in cultured human olfactory neurons. *J. Neurosci. Res.* **62,** 737–749 (2000).

[17] Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1,** 2876–90 (2006).

[18] Wheeler, L.C., Harms, M.J. (in prep).