Theoretical Studies of S100 Proteins using the Accelerated Molecular Dynamics and Nudged Elastic Band Methods

By

Austin Gates

July 2024

Director of Thesis: Dr. Yumin Li

Major Department: Chemistry

ABSTRACT

The use of calcium as a secondary messenger is an important aspect of many biological pathways. Through binding calcium ions, some proteins can undergo conformation changes required for biological activity. Understanding the mechanisms by which calcium binding and conformation changes occur, we can better treat pathologies associated with these proteins. The S100 protein family is highly conserved and is implicated in a number of biological systems due to their interactions with effector proteins. These interactions are dependent on both calcium and target protein concentrations, inducing the formation of a hydrophobic patch region.

In this work, accelerated molecular dynamics and nudged elastic band methods were applied to investigate the calcium induced conformation pathways of the S100A6 protein. Accelerated molecular dynamics was used in the isolation of structures suitable to act as starting and end points in nudged elastic band pathways and in mutation studies. Multiple nudged elastic band simulations were completed to obtain adequate sampling of the S100A6 conformation

pathway. Pairwise distance calculations revealed distinct calcium binding site formation, occurring in a stepwise fashion. The calculation of distance and angle of the hydrophobic patch constituents revealed the presence of a long-lived intermediate conformation. MMGBSA calculations allowed for the identification of K31, D50, and E67 as important residue contributors to patch formation. Residues K31 and D50 were identified as key mechanisms responsible for secondary structure rearrangement and stabilization. Residue E67 was identified as a potential communication mechanism between calcium binding loops, potentially assisting in cooperative binding. Mutation to these residues strongly supported the identified roles of these residues. The application of SASA calculations in these mutated systems revealed significant changes due to the absence of these residues in both apo and holo conformations.

Further, accelerated molecular dynamics and nudged elastic band methods were also applied to the S100B system. Structural start and end points used in nudged elastic band simulations were identified through accelerated molecular dynamics simulations. The goal of this study was to identify mechanisms comparable to those discovered in S100A6. Pairwise distance calculations revealed similar yet unique pathways of calcium binding site formation due to structural differences. The presence of a long-lived intermediate conformation was also identified in S100B through distance and angle calculations of secondary structure elements. MMGBSA calculations revealed mechanisms similar to those in S100A6 that include comparable residues. Residue K30 in S100B was identified to have remarkably similar behavior to K31 in S100A6, having key roles in both sensing initial calcium binding and formation of the hydrophobic patch region. Residue E52 and D55 in S100B were identified to have similar function to D50 in S100A6, functioning to stabilize the hydrophobic patch in conjunction with K30. Initial review of other S100 protein systems identified residues that appear comparable to

both K31 and K30 in S100A6 and S100B, respectively. Further study of the S100 protein family is vital to understanding the role of these proteins in various pathologies and we believe many of the mechanisms identified in this work are present in other S100 proteins.

Theoretical Studies of S100 Proteins using the Accelerated Molecular Dynamics and Nudged Elastic Band Methods

A Master's Thesis

Presented to the Faculty in the Department of Chemistry

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master's of Science Degree in Chemistry

By

Austin Nathaniel Gates

July 2024

Director of Thesis: Yumin Li, PhD

Thesis Committee Members:

Anne Spuches, PhD

Adam Offenbacher, PhD

Xiaoping Pan, PhD

Acknowledgements

In my time at ECU, I have both grown more than I thought possible and met those who are now friends and family to me. First and foremost, I would like to thank my thesis advisor, Dr. Yumin Li. I have known Dr. Li for many years, and when recruited to her lab in my senior year of undergraduate study, I knew it was an amazing opportunity that I could not miss. Over the two years I have been in the lab, she has been a mentor for me and helped me not only hone my academic skills but also those pertaining to life. I would not be the person I am today without her help, and for that, I am extremely grateful.

I would also like to thank my committee, Dr. Adam Offenbacher, Dr. Anne Spuches, and Dr. Xiaoping Pan, for their support. They have been wonderful role models and helped me better understand my own work. I would like to thank the entirety of the ECU chemistry faculty. My time in the department has been extremely beneficial and I have learned much from the work ethic of everyone I have met and worked with. Dr. Andrew Sargent and Dr. Thomas Shoopman have been great influences and genuinely cared about my education and what I had to say. My time in the program has been incredibly humbling and I am proud of the time I have spent here.

Joining Dr. Li's lab introduced me to Sam, Wesley, and Andrew who have grown to be close friends of mine. I will never forget the help and guidance that Sam has provided me. Without him, my work would not be the same. Being able to both joke with and be a nerd around Sam made my time in the lab much more enjoyable. Never did I expect to meet a character like Wesley, but I am grateful that I did. Immediately upon joining the lab, Wesley made me feel welcome and his dedication to learning helped shape my own. Wesley's witty remarks kept me sane on some days and his presence made the lab much livelier. While Andrew only joined the

lab more recently, he has been a wonderful lab-mate the last few months. Andrew is on a similar path to my own, and having someone to talk to about it has done much for me. I am extremely appreciative of each of them, and I am proud to have watched them all grow and succeed.

I am also appreciative of my friends outside of the lab, Brandon, Collin, Jack, and Sid, the Napast family, and everyone I have met through the robotics team. I could not ask for better mentors than Mr. Bill and Mrs. Ann McClung and I cannot express how thankful I am for everything they have done for me. Last but definitely not least, I would like to thank my family. My parents and grandmother have been so supportive of me through my entire life, and I cannot thank them enough. My brother and sister-in-law have been outstanding role models, and I am proud to call them family. They all have helped me in ways they do not even know. Thank you to everyone for all of the love and support that you have given me.

Table of Contents

Table of Figures

Table of Equations

Table of Tables

Chapter 1: Introduction

1.1 Calcium Binding Proteins

Calcium binding is an important, integral aspect of many proteins within the body. The binding of calcium is used in a variety of roles including homeostasis and protein utility.[1] Calcium binding proteins (CaBPs) can serve to store calcium in addition to utilizing calcium as secondary messengers.[2] Calcium levels within cells are highly regulated and a common effect of signal transduction is the release of calcium from cellular stores, such as bone and sarcoplasmic reticula.[1] This release of calcium can drastically alter both intracellular and extracellular levels and promotes calcium binding by CaBPs for various use. Calcium binding in CaBPs typically effect conformation change associated with the activation or deactivation of CaBP function and is usually reversible.[1] The conformation changes associated with CaBPs allow for the proteins to form various interactions with cellular components including other proteins, membranes, and ligands.[1,3]

CaBPs commonly interact with calcium via motifs, or highly conserved structural elements.[4] These structures utilize loops highly composed of negatively charged residues that allow for interaction with the positively charged calcium ions.[5] The highly conserved nature of motifs assists in ensuring proper binding of calcium or other metals. Damaged or mutated loops are commonly associated with decreased binding capacity and disrupted CaBP activity.[3] A common motif observed in a wide array of proteins is the EF-hand motif which is composed of a helix-loop-helix structure, as shown in **Figure 1.1**.[5–8] The loop within the EF-hand motif is primarily responsible for calcium binding and undergoes a conformation change as binding occurs, displacing the associated helices.[6] This conformation change has sequential effects within the protein as various structures are then also displaced, allowing for changes in protein activity.[8]

The CaBP family is extremely large, including the proteins of the annexin family, musculoskeletal proteins such as the troponin-tropomyosin complex, and the S100 family.[5–7,9] In this work, we studied two CaBPs of the S100 family. These two proteins share both similar structure and utility with both requiring the calcium binding associated conformation changes for full functionalization.[10,11]

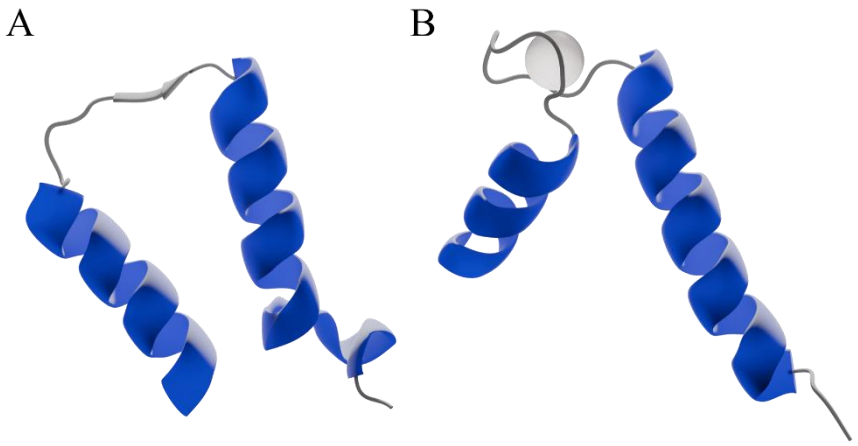

Figure 1.1: Cartoon representation and diagram of EF-hand motif in the calcium absent (A) and present (B) conformations

1.2 S100 Protein Family

S100 proteins were first discovered by B.W. Moore in 1965 and were given S100 designations as the protein constituents were soluble in 100% saturated ammonium sulfate in neutral pH.[10,12] Many members of the S100 protein family had been discovered prior to the identification of the family itself and also belong to the CaBP super family.[11,13] Due to this, many

individual S100 proteins have been known under various names including S100B as NEF, S100A8 as calgranulin-A, and S100A6 as calcyclin.[11] As the proteins were reclassified into the S100 family, they were reassigned with S100- designations. Since the original identification of the S100 protein, over 20 members within a wide range of tissues and sharing similar structural patterns amino acid sequence have been identified within the human body.[11,14] Debate exists surrounding the classification of some S100 proteins and whether they belong to the family due to significant structural differences despite containing elements characteristic of S100 proteins.[7,11,12] S100 proteins have only been observed within vertebrates but share similar roles within non-human species.[11]

Protein activation in most S100 proteins requires the binding of metal ions, typically calcium. The roles of S100 proteins share some similarity between members and most often function in cell growth, proliferation, differentiation, cytoskeletal dynamics and are often associated with cell stress response.[7,11,12,14,15] For example, S100B is found most highly in neuronal cells and is known for its role in stimulating cell proliferation while preventing differentiation and apoptosis.[16] Both S100A4 and A6 are involved in cytoskeletal dynamics.[17,18] However, these proteins also differ in some roles, with S100A4 being involved in macrophage activity and transcriptional regulation while S100A6 is involved in protein ubiquitination and heat shock response.[11] S100A8 functions in stimulating differentiation in both myeloid and keratinocyte cells.[19,20] S100A10 is commonly associated with membrane activity, forming complexes with various proteins, including annexin a2, to function in membrane repair and trafficking of proteins.[11,21] The S100 proteins share realms of functionality but are primarily

tissue and cell specific where they fulfill numerous roles.[10–12] A detailed diagram of S100 protein function is shown in **Figure 1.2**.[22]

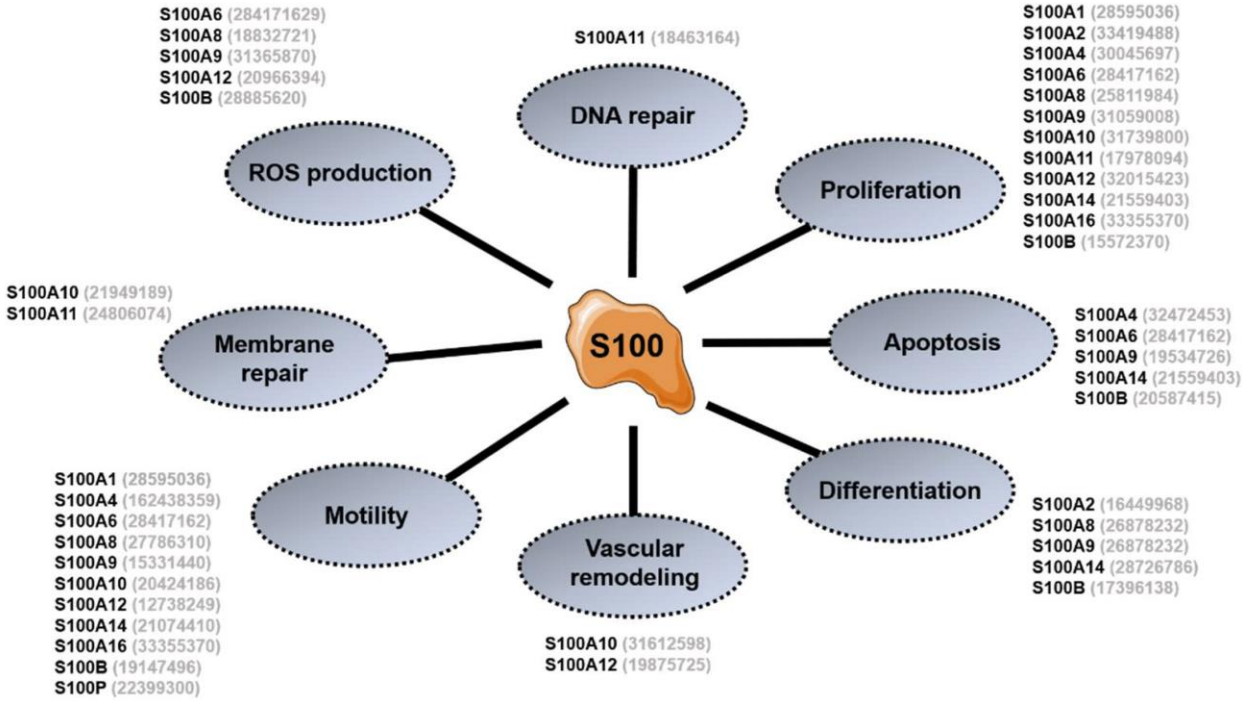

Figure 1.2: Outline of individual S100 protein involvement in cellular pathways. Reprinted from "S100 Proteins in Fatty Liver Disease and Hepatocellular Carcinoma" by Delangre, E. *et al. Int J Mol Sci* **23**, 11030 (2022).

Extracellularly, S100 proteins are observed in various bodily fluids and secretions where they can be used as signaling molecules or biomarkers for disease.[11,15] Functioning within a variety of pathways, especially S100B, these proteins have regulatory effects on various tissues.[11,16,23] S100 proteins are believed to have a non-classical secretion mechanism and follow cell stress events.[11] Many of the extracellular roles of S100 proteins are unproven and pose avenues for future study.

1.2.1 *Structure*

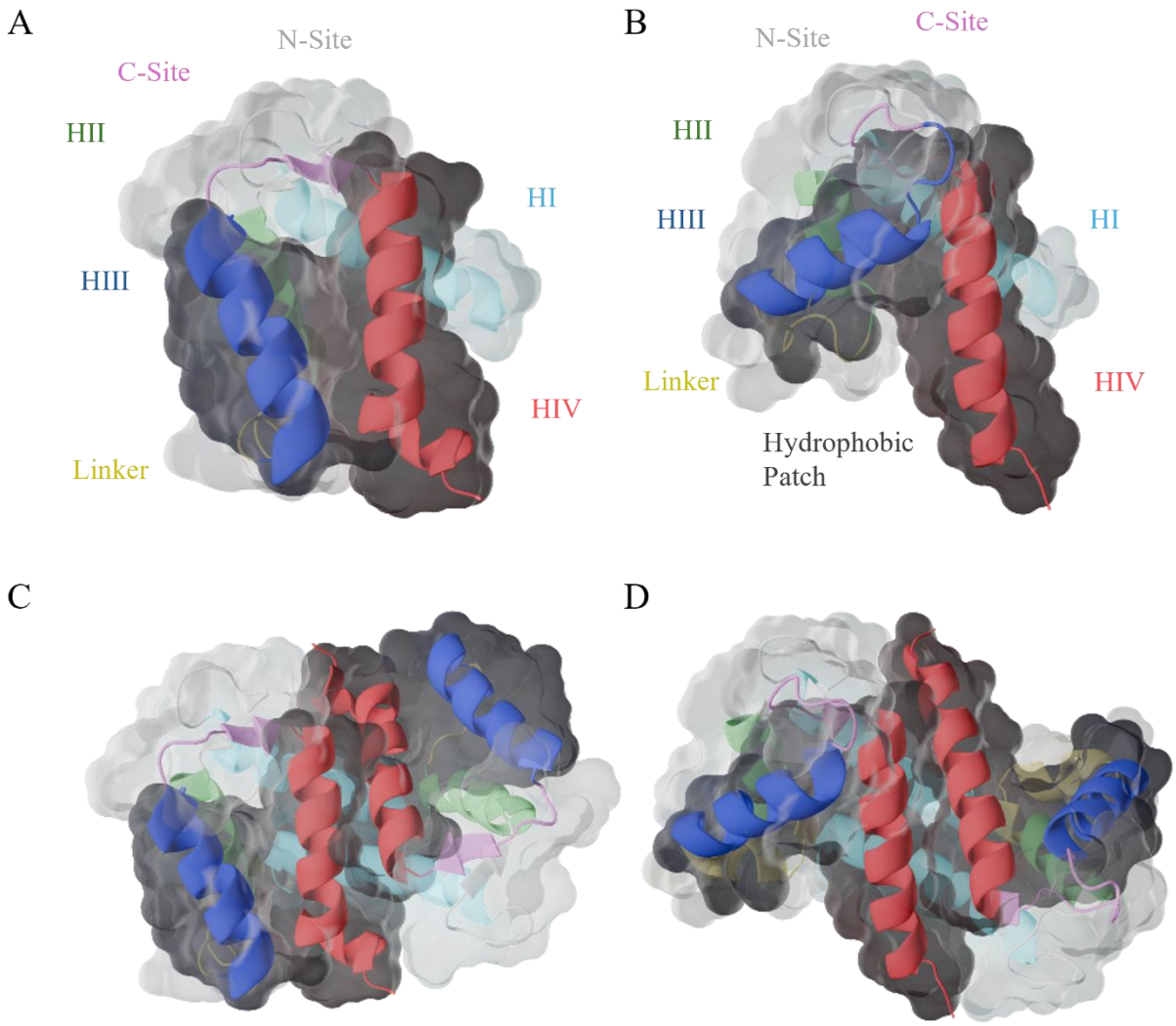

Figure 1.3: (A) Cartoon and surface representations of an S100A6 monomer in the apo conformation. (B) Cartoon and surface representations of an S100A6 monomer in the holo conformation. Labels have been added colored with respect to the individual helices. (C & D) Dimeric cartoon and surface representations of S100A6 in the apo and holo conformations, respectively.

S100 proteins are small (9-12 kDa) proteins containing two EF-hand motifs.[11] These motifs compose an EF-lobe which is responsible for binding two individual calcium ions.[15] The two EF-hand motifs effectively compose almost the entirety of an S100 protein. These proteins

are composed of four helices, with the first two (HI and HII) being directly connected to the N-terminal binding site on either site, with the remaining helices (HIII and HIV) connected to the C-terminal binding sites in a similar manner.[15] See **Figures 1.3A and 1.3B** for visual representation of the generalized structural ordering of S100 proteins. Connecting the two helices is a linker, which varies in composition between individual S100 proteins. While EF-hand motifs are commonly antiparallelly paired, the EF-hand lobe in S100 proteins is unique. A canonical EF-hand motif contains 12 residues associated with calcium interactions.[11] While one such motif exists in S100 proteins, a unique species of EF-hand motif is also present, the S100-hand motif. This motif differs from the canonical motif in that it contains 14 residues and primarily interacts with calcium via carbonyl oxygen atoms.[10,11] The two binding sites are reported to have varying calcium binding affinities.[7,15,24] The canonical motif (cEF) is believed to have a higher affinity for metal ions compared to the S100-hand motif (pEF).[7,11] It has been suggested by Otterbein et al. that cooperative binding may be utilized by members of the S100 protein family.[24] This was, in part, due to these proteins sharing similar structural elements with proteins known to undergo cooperative binding.[25,26] However, it was stated that while there may be some evidence for cooperativity, most models best support a two-step binding mechanism.[24]

While most S100 proteins are CaBPs, many do not have high binding affinities for calcium in the absence of a binding target, a common mechanism among EF-hand proteins.[7,11,12] This mechanism prevents the unnecessary capture of calcium which would decrease calcium levels and potentially lead to accidental activation of S100 and disruption of other CaBP activity. This indicates that most S100 proteins require sufficient concentrations of calcium and target binding proteins. However, S100A10 acts as an exception to this mechanism. S100A10 cannot

interact with calcium ions, making the protein activity calcium independent and solely reliant upon its target interaction partner. [11]

1.2.2 *Calcium Induced Conformation Change*

Upon binding calcium, the primary conformation change is associated with a turn of HIII and the formation of a hydrophobic patch. This patch typically consists of residues in the regions of the linker, HIII, and HIV. Within the apo conformation, HIII rests roughly perpendicular to HIV and sharing hydrophobic interactions, as shown in **Figures 1.3A and 1.3C**. While the composition of the linker region does vary between proteins, the linker region itself is typically proximal to the point where HIII and HIV interact.[7,11,24] These sets of interactions represent the closed state of the hydrophobic patch. As the conformation change occurs, HIII undergoes a large turn, typically turning around 90°, while HIV remains primarily fixed in space, as shown in **Figures 1.3B and 1.3D**.[7,11,15,27] The linker region of the proteins also adjusts, shifting away from HIV. These changes allow for interactions between the patch and various partner proteins. The composition of the patch residue, including the linker, is key to determining what partners the protein can bind with.[7]

Unlike with the cEF, the pEF does not undergo such significant movement as it typically rests in a fixed, open conformation. Instead, the effect of calcium binding here is less clear, but it is believed to play a role in anchoring various protein structures throughout the conformation.[5,28] Within both apo and holo conformations, the pEF and cEF binding loops interact. However, this interaction differs between the apo and holo conformations, with a decrease in interactions commonly being observed as the conformation change occurs.[24] Upon turning, HIII is positioned

alongside HII, which likely functions to assist in stabilizing the position of both HIII and the linker. Due to the lack of clarity, the function of the pEF remains an avenue for research.

1.2.3 *Dimerization*

S100 proteins are most commonly found in a dimerized state, and typically in a homodimeric state although some S100 proteins are known to be capable of forming heterodimers with other proteins.[29] S100G is an exception to this, as it has only been observed in the monomeric state.[11] S100 homodimers are typically connected via non-covalent bonds between HI-HI and HIV-HIV of either monomer, M1 and M2. In these conformations, calcium binding occurs on either end of the dimer and results in the formation of two, identical hydrophobic patches. However, the formation of one hydrophobic patch is not reliant upon the formation of both, meaning that patch formation can be independent of the opposing monomer.

The formation of heterodimers is common in S100 proteins, with S100A8/A9 commonly being observed in addition to S100A6/B also being observed.[30] Unlike the formation of homodimers, which typically allow for the canonical function of each S100 protein, the formation of heterodimers is accompanied by altered functionality. For example, S100A8 and S100A9 have antagonistic roles regarding cell differentiation in their homodimeric forms but

function in fatty acid transportation, oxidative and phagocytic activity, and cytoskeleton dynamics when combined.[11,20,31,32]

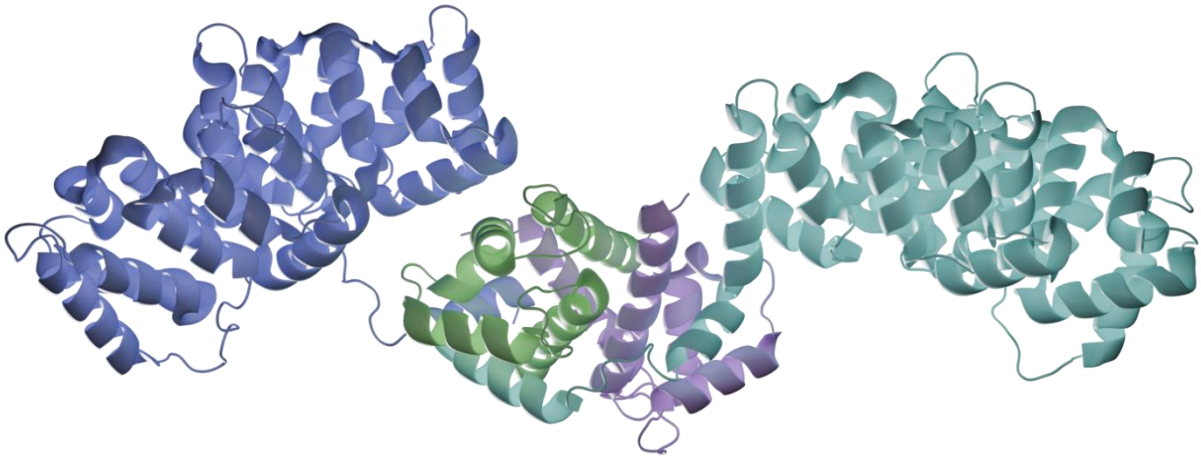

Figure 1.4: Cartoon representation of an S100A4/Annexin A2 heterotetramer. Annexin monomers are colored blue and cyan. S100 monomers are colored green and purple.

An important aspect of S100 proteins is the formation of heterocomplexes, like in the case of S100A10 and annexin A2. This complex consists of homodimeric S100A10 with each monomer being bound to an individual annexin A2 monomer. This results in the formation of a heterotetramer which, as previously described, functions in membrane dynamics.[33,34] As observed in **Figure 1.4**, S100A4 is capable of similar heterotetramer formation.[35] This characteristic is common among many S100 proteins and allows for the proper functionality.

1.2.4 *S100 Proteins in Disease*

S100 proteins are implicated in a wide range of diseases.[7,11,36,37] Given their wide ranges of functions regarding cell proliferation, migration, and stress response, alteration to protein function can have detrimental health consequences. Some S100 proteins have been identified as

having non-redundant functions, namely S100A8 in which gene deletion was observed to result in embryonic death.[38] Most other proteins are heavily implicated in a wide array of cancers and other diseases. While almost all S100 proteins are involved in one disease or another, the type of disease typically differs relating back to tissue specificity of the proteins themselves. For example, deficiencies of S100A1, which is most commonly associated with heart tissues, are related to poor cardiovascular health and heart failure.[39]

S100 proteins associated with cell migration and cytoskeletal dynamics are also commonly upregulated within many cancers, such as S100A4 and S100A6.[40–43] Several S100 proteins are implicated as having roles in breast cancers. S100A7 is often upregulated in some breast cancers while also reducing cell proliferation, while both S100A8 and S100A9 have been associated with poor prognoses.[44–46] S100 proteins also commonly interact with or regulate p53, often altering the effectiveness of its tumor suppressing functionality.[7,11,47–50]

Given the presence of S100 proteins in many diseases, they are often used clinically as a biomarker for their associated diseases.[11,14] The most well-known example of an S100 protein being used as a biomarker is with S100B and a number of neurological disorders.[23,36] S100B concentrations are often used in determining prognoses, such as with elevated levels being associated with a poor prognosis after a subarachnoid hemorrhage.[16,51] S100A8, S100A9, and S100A12 are other commonly used biomarkers and can distinguish between forms of arthritis, with the S100A8/A9 heterodimer itself being used in identifying erosive disease and determining prognoses.[11,52–54] Additionally, S100A12 can be used in identifying several lung pathologies and can be used to indicate prognoses.[55,56] S100 proteins as biomarkers are powerful tools in clinical settings and may prove to be key in the treatment of many pathologies.

1.2.5 *Protein Interactions*

As previously discussed, the formation of interactions between S100 proteins and various proteins are an important aspect of S100 function. Given that shape determines function within protein systems, varied composition within the hydrophobic patch, and thus varied shape, is an important regulator for determining what a S100 protein can interact with. This variety is most commonly found within the linker which is typically observed to have the highest levels of composition differences between S100 proteins.[57] Many of these proteins bind via burying a small peptide sequence within the hydrophobic patch. However, this is not the case for all S100 interacting proteins. One such protein is the receptor for advanced glycation end products complex (RAGE). [58–60]

RAGE proteins are important interaction partners for many S100 proteins, such as S100A6, S100B, and S100P.[11] S100A6 and S100B have been observed to interact with RAGE proteins via the C-terminal binding loop and C-terminal residues of HIV rather than the hydrophobic patch alone.[58,60] However, some variation has been observed between the regions with which either of these S100 protein interacts. Conversely, S100P has been observed to interact with RAGE via a more canonical mechanism in which interactions form between RAGE and the linker region of S100P.[61] While these interactions may vary between S100 proteins, this is also due to the conformation of the complex containing the S100 protein. For example, the regions available for interaction will differ in the case of an S100 protein existing as a homodimer or as a tetrameric complex. These complexes significantly affect the affinity of target proteins as in the case of S100B, which has a very strong affinity for RAGE when found extracellularly in a tetrameric complex.[62] S100 proteins are diverse in both proteins they interact with and how these interactions occur.

1.2.6 *S100A6*

Unlike many other S100 proteins, S100A6 has been observed within most tissue types but primarily resides in fibroblasts and epithelial tissues.[18,27] S100A6 expression has been observed in neural and muscular tissues. Interactions between effector proteins and S100A6 are key to the biological roles of S100A6 and link S100A6 to specific roles. Intracellularly, S100A6 primarily exists within the cytoplasm where it interacts with a wide array of proteins including tropomyosin, cofilin-1, and a number of proteins relating to cytoskeletal dynamics.[27]

Reports indicate that S100A6 is also involved in nuclear activity.[27,37] Interactions between S100A6 and Lamin A/C, importin α, and ribosomal S6 kinase 1 are believed to regulate chromatin organization, nuclear transport, and cell proliferation via cell cycle regulation, respectively.[27,37,43,63] Like other S100 proteins, S100A6 is known to interact with several annexin proteins, namely annexin A1, A2, A6, and A11, where they are involved in membrane associated roles including secretion and endocytosis.[11] S100A6 has also been reported to function in a protein chaperone capacity. The proteins to which this function has been linked to are primarily associated with cellular stress response and include many of the proteins that interact with Hsp70 and Hsp90.[64,65] Although there is less documentation surrounding the extracellular functions of S100A6, it is known that these proteins are used for a number of signaling processes resulting in various secretory events. These include the release of insulin and histamine from pancreatic and mast cells, respectively.[18]

Structurally, S100A6 proteins are small (10-12 kDa) and similar to most other S100 proteins and are typically found homodimerized. S100A6 contains two EF-hand motifs, one canonical motif and one s100-hand motif.[24] In humans S100A6 contains 90 residues, with slight variations in length observed in non-human vertebrates. The pEF is comprised of HI and HII

which connect to the N-terminal binding site. The cEF, responsible for the most significant conformation change, contains HIII and HIV which surround the C-terminal binding site. These motifs are connected via a linker region containing 8-9 residues. Upon calcium binding, HIII undergoes a turn of roughly 86°.[24] This turn also results in the linker region swinging away from the core of the protein and reveals a large hydrophobic patch with which other proteins can interact. S100A6 is known to interact with other metal ions, commonly zinc, but the resulting changes are believed to differ from calcium binding and are not as well understood.

S100A6 is heavily involved in several cancers, including pancreatic and gastric cancers, carcinomas, and melanomas.[37,66–68] In many of these cancers, the expression of S100A6 correlates directly with the properties of the cancer itself, including tumor size and metastasis. The functions of S100A6 associated with cell proliferation, growth, and motility are directly attributed to its role within these cancers in which it is typically upregulated. However, it is unknown whether this overexpression is a result of the immune response, or if it is by cancer cells to assist in proliferation. Reportedly, knockout of S100A6 was observed to result in the relief of symptoms associated with the cancers. However, S100A6 was recently linked to the regulation of hematopoietic stem cells (HSC), and it is believed that S100A6 regulation of Hsp90 activity has indirect, downstream effects relating to the p-Akt pathway.[69] In the event of S100A6 loss, it was reported that one of many consequences involved a decrease in HSC viability alongside various other effects associated with increased apoptotic activity, including decreased mitochondrial metabolism. These results and the detection of S100A6 in other various cancer stem cells do suggest that S100A6 is directly involved in and contributing to many cancers.

Non-cancerous pathologies have also been identified to result in elevated levels of S100A6. Many of these diseases are related to fibroblasts, in which S100A6 proteins are known

to localize in.[37] However, S100A6 is also implicated in several neurodegenerative diseases as a result of its link to astrogliosis. In many of the described pathological related cases, S100A6 is often used as a biomarker for both detection and prognosis determination. Clinically, the detection of increased levels of S100A6 often correlates to the progression of the pathology.

1.2.7 *S100B*

S100B is one of the most well studied S100 proteins. This protein localizes primarily in cells of the nervous system where it functions in enhancing cell proliferation and migration in addition to inhibiting differentiation and apoptosis.[11] S100B proteins act both intracellularly and extracellularly. Intracellularly, S100B is known to interact with several binding partners to effect function. Some of these partners include annexins A2 and A6, p53, and a number of cytoskeletal elements including caldesmon and type III intermediate filaments.[11,16] Many of these interactions serve roles associated with cytoskeletal dynamics and protein phosphorylation. S100B is also believed to be a key regulator of calcium homeostasis in many nervous system cells.[36,70]

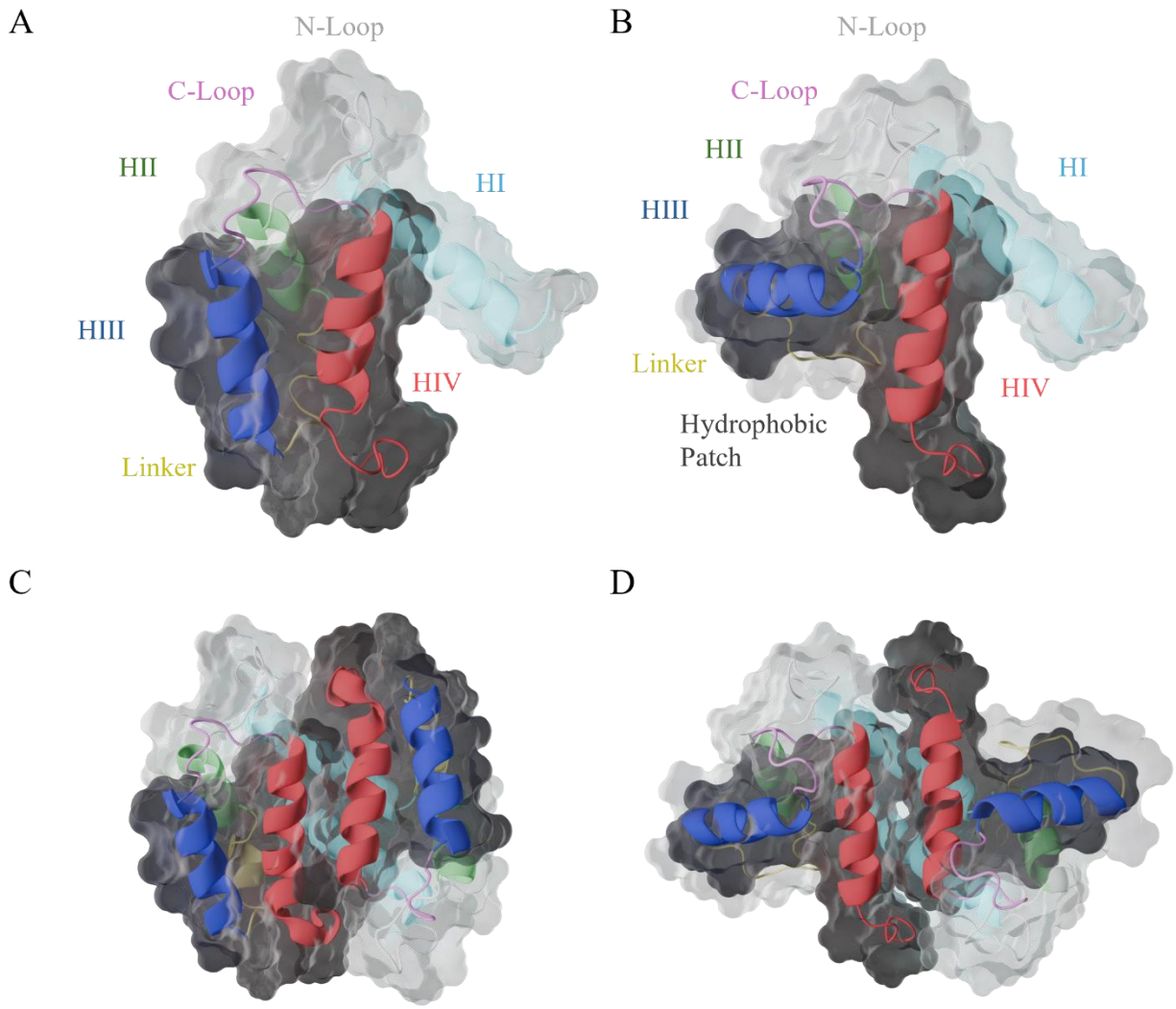

Figure 1.5: (A) Cartoon and surface representations of an S100B monomer in the apo conformation. (B) Cartoon and surface representations of an S100B monomer in the holo conformation. Labels have been added and colored with respect to the individual helices. (C & D) Dimeric cartoon and surface representations of S100B in the apo and holo conformations, respectively.

Extracellularly, S100B has been linked to activity associated with astrocyte proliferation and apoptosis, neuronal apoptosis, neurite extension, and peripheral nerve regeneration.[11,71] While the mechanisms by which the described activity occurs are not known in all cases, S100B proteins are understood to be instrumental in maintaining proper neural function. Reportedly, S100B proteins are linked to cognitive behavior.[16] However, the exact mechanism by which

S100B affects behavior is unknown but faces continued research. Like S100A6, S100B is known to interact with RAGE proteins and a significant portion of S100B extracellular activity is attributed to RAGE-dependent pathways.[62]

S100B contains similar structural characteristics to most other S100 proteins, containing pEF and cEF motifs connected via a linker region.[72] The pEF itself is composed of two helices, HI and HII, surrounding the N-terminal binding site. The cEF is composed of the remaining helices, HIII and HIV, surrounding the C-terminal binding site. The primary conformation change associated with calcium binding in S100B is the turn of HIII, as with most other S100 proteins despite somewhat varied composition. This turn, between 80-90°, allows for the interaction with many proteins. [16]. The linker region is similar in both location and function while being composed of a unique residue composition, as is characteristic of S100 proteins. However, this linker region has been observed to form a short helix in the absence of calcium.[72] A detailed overview of S100B is shown in **Figure 1.5**.

S100B is widely reported as being involved in many pathologies, including cancers and neurodegenerative disorders in addition to traumatic brain injuries.[23,51,73] Given the high localization of S100B in nervous tissue, it is only logical that many of these pathologies pertain to the nervous system as well. These include Parkinson's disease, Alzheimer's disease, schizophrenia, and multiple sclerosis. However, S100B is implicated in some non-neurological pathologies such as diabetes or cardiomyopathy. S100B is often used in clinical evaluations for patients exhibiting symptoms of these pathologies. However, the role S100B proteins play in these cases is not completely understood and it can be difficult to ascertain whether S100B level fluctuation due to release from damaged cells or as a response by the body. While not involved in

as many cancers as some S100 proteins, S100B has been implicated in both cutaneous melanomas and breast cancers.[74,75]

Chapter 2: Computational Methodology

2.1 Molecular Dynamics

In the study of proteins, it is well known that the function of a protein is reliant upon the structure of the protein itself. However, direct visualization and characterization of proteins can be difficult and prohibitively expensive. Due to this, computational methods have become a powerful instrument in providing insight in understanding protein function. Conventional molecular dynamics (cMD) has become of the more frequently used methods, utilizing known structures and classical mechanics to approximate conformation changes.[76,77] Conventional molecular dynamics treats atoms, or groups of atoms, as beads connected via springs. Motion in these systems is derived from Newton's second law of motion as described in **Equation 2.1**

$$F = ma \qquad \text{Equation 2.1}$$

where force (F) is equal to mass (m) times its acceleration (a). In molecular dynamics simulations, forces are derived as gradients of the potential energy using force fields.

$$U = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{Vn}{2}[1 + \cos(n\psi - \varphi)] + \sum_{electrostatic} \frac{q_i q_j}{\varepsilon_d r_{ij}} + \sum_{vdW} \left\{ \varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] \right\} \qquad \text{Equation 2.2}$$

This equation describes a generalized potential energy function to characterize proteins and other large biomolecular systems. Potential energy itself is determined by terms representing bonded and non-bonded interactions. The bonded terms typically include bond length, angle, and dihedral angles terms. The bonds term is responsible for energy associated with changes in bond lengths, treating K_b as the force constant and $b\text{-}b_0$ as the changes in length where b_0 acts as reference. The angle term calculates changes in bond angle in reference to θ_0, treating K_θ as the angular force constant. Potential energy associated with dihedral angles are described as periodic in which V_n and n represent barrier height and periodicity, respectively. The angle between planes is represented by ψ, with ϕ representing the phase shift angle. The non-bonded terms typically describe electrostatic, van der Waals, and repulsive terms, with the latter two typically being combined. Electrostatic interactions are typically calculated according to Coulomb's law which summates interactions between charges q_i and q_j within r_{ij} distance. The dielectric constant is also included in this term as ε_{ij}. The Lennard-Jones potential is used as a combination of van der Waals and repulsive interactions and calculation utilizes the depth of the energy well, ε_{ij}, distance between atoms, r_{ij}, and the distance where the potential energy is zero, $R_{min,ij}$.

The goal of cMD simulations is to watch biomolecular systems evolve over time. Thus, the system must constantly calculate these forces as the system changes. These simulations usually follow a similar sequence of calculation. Initially, systems must be defined according to atomic positions and initial velocities. The forces on each atom are then calculated according to their interactions with neighboring atoms. The calculated acceleration is applied to the atoms velocity, resulting in a new position and velocity at t + dt, with dt being the fixed time-step.

These timesteps are typically very short, with AMBER using a 1 femtosecond (fs) timestep, to avoid issues relating to instability of light atoms, such as hydrogen.[78] Short timesteps

do often limit the ability of a system to undergo large conformation changes as access to compute time is often limited. However, many of these programs incorporate various methods that allow for increased timesteps. For example, simulations using AMBER can utilize the SHAKE algorithm and hydrogen mass repartitioning (HMR) to expedite simulation speed and allow for larger changes within the system to be observed.[78–80] The SHAKE algorithm functions by applying constraints on light atoms that prevent them from moving too fast and becoming unstable, permitting a timestep of 2 fs. HMR functionality is such that mass from larger atoms, such as carbon, are taken and given to hydrogen atoms.[79] The increased mass assists in preventing hydrogen instability. HMR has been shown to reproduce conformation changes within many biomolecular systems while also allowing for a 4 fs timestep when used in conjunction with the SHAKE algorithm.[79,80] There are many force fields used in cMD simulations but some of the more commonly used force fields are those of AMBER, CHARMM, and GROMOS.[78,81,82] The AMBER suite of programs and associated all-atomic force field, ff19sb, were used.[78,83]

2.2 *Accelerated Molecular Dynamics*

Given the difficulties in simulating large conformation changes using cMD, various methods have been designed to assist in this process. The previously described SHAKE algorithm and HMR are effective, but still exhibit some difficulty in characterizing conformation changes requiring simulations requiring upwards of a microsecond.[78–80] Accelerated molecular dynamics (aMD) is a branch of cMD developed to assist in simulating system that require significant compute time.[84] In aMD, a boosted potential is applied to systems of interest, in which the boost can be applied to the total and dihedral potential.[84] This is accomplished using

boost parameters partially derived from cMD simulations. The boosted potential is calculated in accordance with the following equation.

$$\Delta U(r) = \frac{(E - U(r))^2}{\alpha + (E - U(r))} \quad \text{Equation 2.3}$$

Included in these calculations are the total biased potential, ΔU(r), boost threshold, E, unbiased potential U(r), and the boost factor, α. Calculating the boost for dihedral populations utilizes a very similar equation with the only differences being that unbiased dihedral energy is used, instead of total potential energy, in addition to a dihedral specific boost factor. Guidelines have been established to assist in reasonable calculation of the energy thresholds and boost factors. The calculations for these are described in the following equations.

$$E_{tot} = U(r) + (0.16\ kcal\ mol^{-1}\ atom^{-1} * n_{atoms}) \quad \text{Equation 2.4}$$

$$\alpha = 0.16\ kcal\ mol^{-1}\ atom^{-1} * n_{atoms} \quad \text{Equation 2.5}$$

$$E_{dih} = U(r)_{dih} + (4\ kcal\ mol^{-1}\ residue^{-1} * n_{sol}) \quad \text{Equation 2.6}$$

$$\alpha_{dih} = (1/5) * (4\ kcal\ mol^{-1}\ residue^{-1} * n_{sol}) \quad \text{Equation 2.7}$$

These calculations include U(r), unbiased total potential energy, n_{atoms}, the number of atoms present in the system, n_{sol}, number of solute residues. The first two of these equations are used in establishing boost parameters for the total potential energy. Calculations for the dihedral energy boost are described by the two remaining equations. All aMD simulations completed in this work followed these guidelines in the calculation of boost parameters.

The boosted potential is not applied under all conditions, but rather, it serves to prevent the system from falling into energy wells. This permits the system to more easily traverse the energy landscape. Due to this, conformation changes must be viewed as independent of time as systems may not follow established timeframes. The protocol by which aMD is applied is described by the following function.

$$U^*(r) = \begin{cases} U(r), & U(r) \geq E \\ U(r) + \Delta U(r), & U(r) < E \end{cases} \quad \text{Equation 2.8}$$

The modified potential is represented by $U^*(r)$, with U(r), the unbiased potential, $\Delta U(r)$, the boost to potential, and E, the energy threshold, also being represented here. When the system falls below the calculated threshold, the system applies the boost to the potential. The primary use of aMD is to survey the possible conformations of biomolecular systems. The ease with which these systems can progress through conformations allows for the study of conformation changes previously inaccessible to us. aMD can also be applied to smaller systems. For the structural characterization of protein structures, they are typically crystallized, resulting in tight packing useful for imaging. However, this packing can deform proteins from their physiological shape. aMD provides a tool to return protein structure to more physiologically relevant conditions. Measuring the tendency of the protein to rest within specific states can indicate increased stability.

2.3 Molecular Dynamics Simulation Details

2.3.1 *System preparation*

Simulations utilizing cMD require preparation prior to completing final simulations. Initially individual structures must be identified for each system of interest. While some

structures contain missing regions that require additional preparation, this was not needed in this work. However, in some systems, PyMol was used in the application of point mutations.[85] Hydrogen atoms are then removed from the system to prevent issues regarding the AMBER naming.[78] TLEaP, native to AMBER, was used in the preparation of topology and coordinate files and was also responsible for adding missing hydrogens according to the residue database.[78] These files contain the raw coordinates of atoms and their associations required for simulation. The systems were then solvated with water alongside sodium and chlorine ions to establish physiological conditions.

Due to these conditions in which systems are simulated in, it would typically be inevitable that water or any of the various ions would escape physiologically relevant bounds. Increasing the system size would also increase the compute requirements for the system, reducing the speed at which a simulation progresses. The introduction of a strict box would result in issues regarding temperature and pressure regulation. In these simulations, periodic bounds are used to solve this issue. These bounds work by introducing a defined boundary around the system, with a 10 Å non-bonded cutoff being used here. In the case of a water molecule crossing the bounds, it is then wrapped around to the opposite side of the box. This allows for the system to act as if it has infinite bounds and prevents solvent escape. While the application of periodic bound conditions is useful, it introduces challenges associated with handling long-range electrostatic interactions. However, the use of the Particle mesh Ewald method effectively manages these challenges.

2.3.2 *Minimization*

Prepared systems underwent energy minimization, reducing strain and removing problematic contacts present in the initial structures. In all simulations, minimizations were completed over two stages. In the first stage, the protein itself was kept constrained for 10,000 steps of minimization. The steepest descent method was used for the first 5,000 steps, with the remaining steps utilizing the conjugate gradient method. In the second stage, the entire system was allowed to minimize over 1000 steps. The first 500 steps utilized the steepest descent method, with the remaining steps using the conjugate gradient method. From these minimizations, we obtained restart and trajectory files necessary for further simulation.

2.3.3 *System Warming*

The process of system warming allows for the system to reach physiological temperatures and assists in equilibration. This was completed over 25,000 steps, in which the system is raised from 0 K to 300 K. Heating occurred gradually, to prevent the formation of pockets within the solvent and maintain an even transition to physiological temperatures. The protein structure was held fixed in this process, leaving solvent molecules unrestrained. This process utilized a 1 fs timestep to reduce any instability resulting from the heating protocol. The final coordinates and restart file were used in the next step.

2.3.4 *Constant Pressure Dynamics (NPT)*

NPT steps were conducted to allow for further equilibration. NPT simulations utilize fixed moles (N), pressure (P), and temperature (T) and allow for the density equilibration of the system. When solvent molecules are added, this is done randomly within the defined boundaries. This can result in uneven distribution of molecules. In this protocol, the volume of the system can fluctuate to facilitate a more even distribution. Two steps of NPT simulations were completed, totaling 200,000 steps at a 1 fs timestep. The coordinate and restart files created in these steps were used in running production simulations of the system.

2.3.5 *Constant Volume Dynamics (NVT)*

NVT simulations were completed once density equilibration was complete. These simulations are similar to that of NPT simulations, except moles (N), volume (V), and temperature (T) are held fixed, allowing for pressure variation. These simulations consisted of 10,000,000 steps at a 2fs timestep, or 20 nanoseconds (ns), of simulation time in each stage. The number of stages can be increased or decreased depending on the desired simulation length. The SHAKE algorithm was used in these simulations to allow for the aforementioned 2 fs timestep.[80] Further, it is within this step that aMD and cMD simulations diverge. Given that aMD requires parameters obtained from cMD simulations, 50 ns cMD simulations were completed. Once the parameters were obtained and boost parameters were calculated, the aMD simulations could start over in this step. These simulations were used to obtain trajectories utilized in data analysis.

2.4 Nudged Elastic Band

Conventional and accelerated molecular dynamics are both advantageous in observing protein dynamics due to the lack of a defined end-state, allowing for the system to evolve in an unbiased manner. However, they can prove difficult when one wants to study a specific conformation change. The nudged elastic band method (NEB) provides a strong tool to directly study the transition of a structure between two states and is derived from the plain elastic band method (PEB).[86,87] PEB simulations functioned as an attempt to identify the lowest energy paths between structures. However, this method was inclined to produce pathways that cut corners in the energy landscape.[86] The NEB method was developed as a fix to this issue and avoids it by applying spring forces perpendicular to the tangent of the path.

NEB simulations are most highly utilized in the observation of small molecules where it is possible to use more computationally demanding but more accurate methodologies. However, it is possible to complete NEB simulations using programs native to AMBER.[78] While these NEB simulations are typically more restricted to small molecule systems, it does provide an avenue to assist in characterizing larger systems, such as proteins.

Typically, the NEB method requires both starting and endpoints in addition to hypothesized structures along the pathway. This method treats these structures as images and distributes them along a path between two points. Within AMBER, the implementation of simulated annealing permits the use of starting and endpoints only.[78] In this method, the first half of the images are all identical to the starting structure with the second half being identical to the endpoint. By connecting the images with springs, simulated annealing can permit the images to explore conformational space without images evolving independently due to being tethered by

the starting and end points. The springs prevent the images from falling back down into energy minima and maintain even distribution.

The connection of images by springs is a critical aspect of the NEB method and allows for the path to be accurately approximated. The spring forces exerted upon the images have parallel and perpendicular components.

$$F^{NEB} = F^{\parallel} + F^{\perp} \quad \text{Equation 2.9}$$

$$F^{\parallel} = [(k_{i+1}|P_{i+1} - P_i| - k_i|P_i - P_{i-1}) * \tau]\tau \quad \text{Equation 2.10}$$

$$F^{\perp} = -\nabla U(P) + (\nabla U(P) * \tau)\tau \quad \text{Equation 2.11}$$

The equations above calculate the forces, F, acting on each image, i, in both parallel, $F^{\parallel}$, and perpendicular, $F^{\perp}$, directions. Included in these calculations are the 3N-dimensional vectors of the images, the spring constant, k_i, between each image, i-1, and image i, with τ describing the 3N-dimensional tangent unit vector of the path. The potential energy of the system, as described by the force field, contributes to these calculations as U.

While spring constants are often fixed, they can be scaled to allow for higher resolution transition states. In this work, all NEB simulations utilized scaled spring constants. This method utilizes a climbing image technique in which all springs scale in reference to the highest energy image. This is described by the following equation.

$$k_i = \begin{cases} k_{max} - \Delta k(E_{max} - E_i)/(E_{max} - E_{ref}, & E_i > E_{ref} \\ k_{max} - \Delta k, & E_i \leq E_{ref} \end{cases} \quad \text{Equation 2.12}$$

The application of scaled spring constant is relatively insensitive to the magnitude of the spring constant, ensuring proper application of the scaled spring force as needed.[78] The generalized Born implicit solvation model is reported to provide a good approximation to explicit solvent calculations and was used in all NEB simulation included in this work. This allowed for reduced computation time and complexity regarding the conformation path.

2.5 NEB Simulation Details

2.5.1 *Minimization*

Minimizations were completed on both starting and ending structures used in the NEB simulations. The minimization protocol in this work is derived from previous systems studied using the NEB pathway.[88] Two stages of minimization were completed for each structure, the first of which held all atoms restrained except hydrogen. This was completed over 10,000 steps, split evenly between steepest descent, then conjugate gradient methods. The second stage minimization held no atoms restrained, allowing the problematic contacts and strain to be fully relieved. The structures obtained from these minimizations were then used in further equilibration steps.

2.5.2 *Heating & equilibration*

Once minimizations were completed, the images were distributed between the starting and ending images of the path. There were 30 images placed between the points at either end, with a total of 32 images present. The protocol for simulated annealing dictates that, in total, 16 images of both the starting and ending structure are placed. The first 16 are composed of the starting structure while the remaining are composed of the ending structure. These structures were heating from 0 K to 300 K over 40,000 steps at a 0.5 fs timestep. The systems then undergo

further equilibration to ensure that the remaining structures have no unresolved issues. This allows for a smooth transition into simulated annealing.

2.5.3 *Simulated Annealing*

The process of simulated annealing increases the temperature of the systems much higher than observed under physiological conditions. This step allows for the structures to freely conform and sample the conformation space. However, the attachment of these structures to the endpoints acts as a tether, ensuring the structure remains on the path between starting and ending conformations. This step requires the most computational time to ensure that the structures have ample time to explore the landscape. The temperature was raised in increments of 100 K, starting at 300 K, in steps of 100,000 until the system reached 500 K. The system was held at 500 K for 100,000 steps before returning to 300 K at which it was held for another 100,000 steps.

2.5.4 *Cooling*

Cooling the system allows for the isolation of structures along the path as they are returned to local minima. Cooling occurred over two stages. The first stage of cooling reduced the temperature of each structure by increments of 50 K before being allowed to equilibrate for a short period. This process was repeated until the temperature of the systems reached 0 K. The second stage of cooling utilized the quenched velocity method, ensuring that structure energies were properly minimized. The structures identified here were used in further data analysis.

2.6 Data Analysis

All structural visualization and preparation were completed using PyMol and TLEaP.[78,85] Trajectory analysis was completed using CPPTRAJ, MDAnalysis, and NumPy modules.[89–92] The matplotlib module was used for plotting and creating figures.[93] The Molecular Nodes plugin for Blender was used to create structural figures of each system.[94]

Chapter 3: Characterization of Calcium Induced Conformation Changes and Hydrophobic Patch Formation in S100A6

3.1 Background

The major protein conformation associated with S100 proteins is the turn of HIII upon calcium binding. However, the many mechanisms by which this process occurs have not been completely characterized. A previous study by Okazaki et al. included an investigation into the S100A6 conformation change.[95] In that work, a coarse-grained simulation was completed that allowed for the description of the pathway between apo and holo states. It was reported that, as the system progressed, native contacts involving HIII within the apo state were broken before contacts native to the holo state formed, occurring sequentially. The general motion of HIII was also described as a shearing motion. While this work contributed significantly to the understanding of S100 protein conformation change, it did not completely describe mechanisms responsible for the change itself. The use of coarse-grained models are a powerful method to simulate large conformations quickly by sacrificing system resolution for simulation speed. While this method is known to produce accurate results regarding large-scale conformation changes, interactions between residues are effectively indescribable.

Given that EF-hand motifs have been reported as potentially using cooperative binding, Otterbein et al. has suggested E67 as a potential mechanism for this in S100A6.[24] It was reported that E67 forms interactions between both solvent water molecules and residues of the C-terminal binding site, assisting in interactions between water and Ca^{2+} ions. While these interactions were only suggested to play a role in cooperative binding, further study may allow for better elucidation of the mechanism by which S100A6 conforms upon calcium binding. In this work,

we present the results of a work using the NEB and aMD methods to study the Ca^{2+}-induced conformation change of S100A6. To the best of our knowledge, neither NEB nor aMD methods have been applied to this system. Specifically, we sought to understand the conformation change of HIII and how various residues, including E67, contribute to the pathway.

3.2 Computational Methods

3.2.1 *System Details*

Accelerated molecular dynamics and NEB simulations were used in conjunction to study the S100A6 system. Previously completed work in our lab uncovered difficulties associated with simulating S100A6 using only cMD. In that work, cMD proved unreliable to simulate the conformation change itself or to maintain known conformations. Thus, aMD was chosen as it proved more reliable in maintaining patch formation. To identify conformations suitable to act as the starting and ending structures of the pathway, aMD simulations were completed on S100A6 PDB crystal structures in both apo (1K9P) and holo (1K9K) conformations.[24] Principal component analysis coupled with a kernel density estimate were used to identify the most representative structures in a reduced space, using the first two principal components. The regions with the highest sampling of similar structures are believed to be the most stable and therefore more representative of the physiological forms of S100A6. From this, we identified structures most representative of the apo and holo conformations.

3.2.2 *Simulation Details*

The aMD simulations were conducted utilizing the previously described protocol and force field. Unlike with the NEB simulations, an explicit solvent was used here. The OPC water box and 12-6 OPC ion parameters were utilized in this solvent. 200 ns of production simulation were completed for each aMD simulation, allowing ample time for the system to sample conformation space. Structures isolated from these simulations were then minimized and passed along to the NEB simulation protocol. The starting and ending structures are shown in **Figure 3.1**, represented by blue and orange points, respectively.

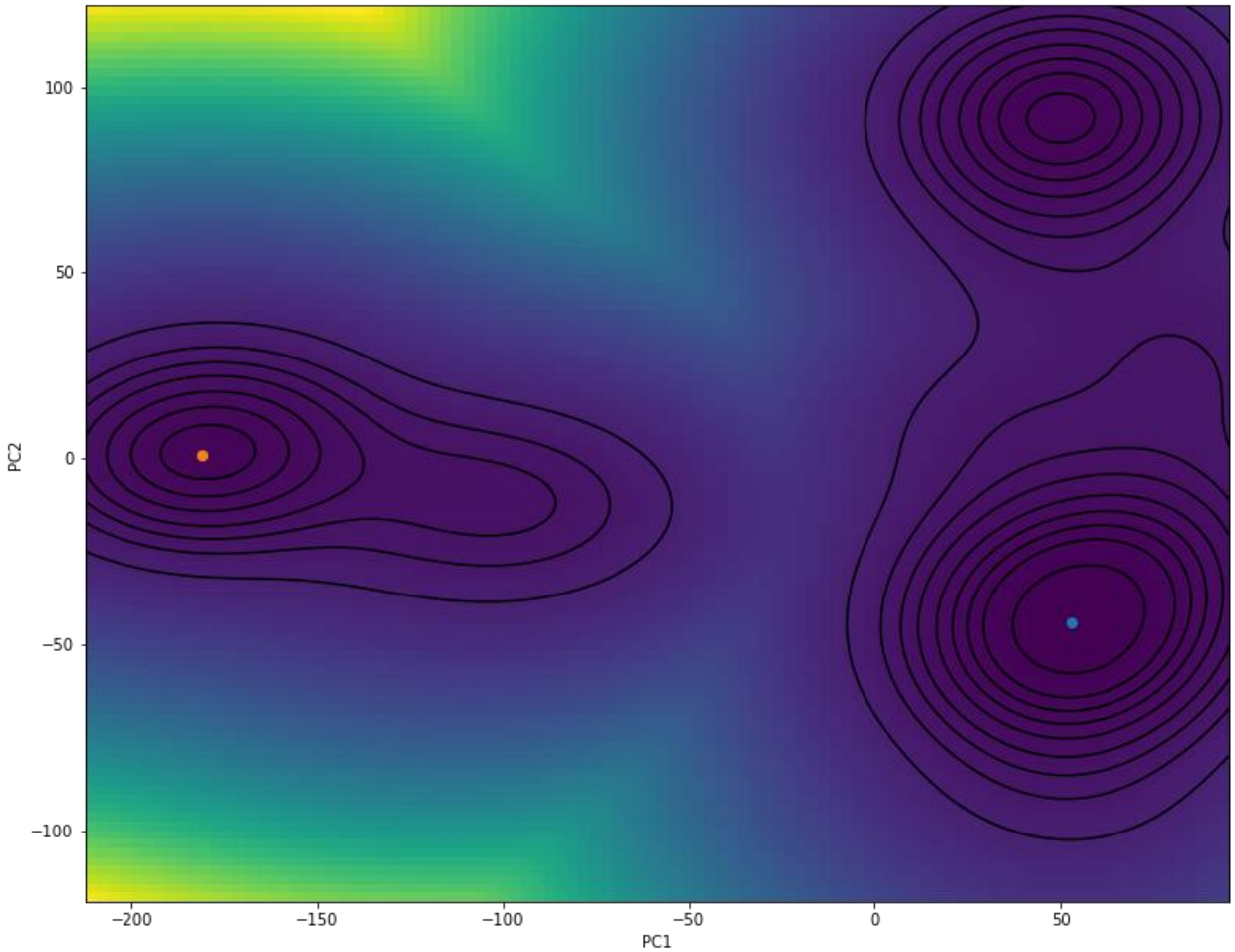

Figure 3.1: Diagram of the first two principal components combined with a KDE. The chosen apo and holo conformations are displayed as blue and orange dots, respectively.

NEB simulations were completed to characterize the pathway between apo and holo S100A6 conformations. Differences between initial velocities of atoms are capable of biasing results of MD simulations. To avoid this, these NEB simulations were completed in pentaplicate to ensure the validity of simulation results. As ion behavior often proves unreliable in implicit solvent models, NEB simulations completed in this work were done so in the absence of Ca^{2+} ions.[96]

3.2.3 *Mutated Simulation Details*

To better understand the role of specific residues, point mutations were utilized to characterize the effect of their absence. These residues are K31, E67, and D50, which were all mutated to alanine. Simulations were conducted for each of these systems in which both apo and holo conformations were simulated independently. This allowed us to characterize the effects of the residue at either end of the pathway. Additional simulations were completed using non-mutated structures to provide a reasonable comparison to determine any changes associated with the mutations.

3.3 Results

3.3.1 *Ca^{2+} Binding Site Formation*

The primary conformation changes of interest, in S100A6, results from the binding of Ca^{2+} to binding sites. Given that our simulations were completed in the absence of Ca^{2+}, we calculated the pairwise distance between known Ca^{2+} binding atoms as a proxy for binding site formation. To identify these atoms, the holo-PDB structure (1K9K) was used. The atoms used in the calculation for the pEF were S20@O, G24@O, D25@O, T28@O, E33@OE1, E33@OE2.

The atoms used in the calculation for the cEF were D61@OD1, D61@OD2, N63@OD1, D65@OD2, E67@O, E72@OE1, E72@OE2.

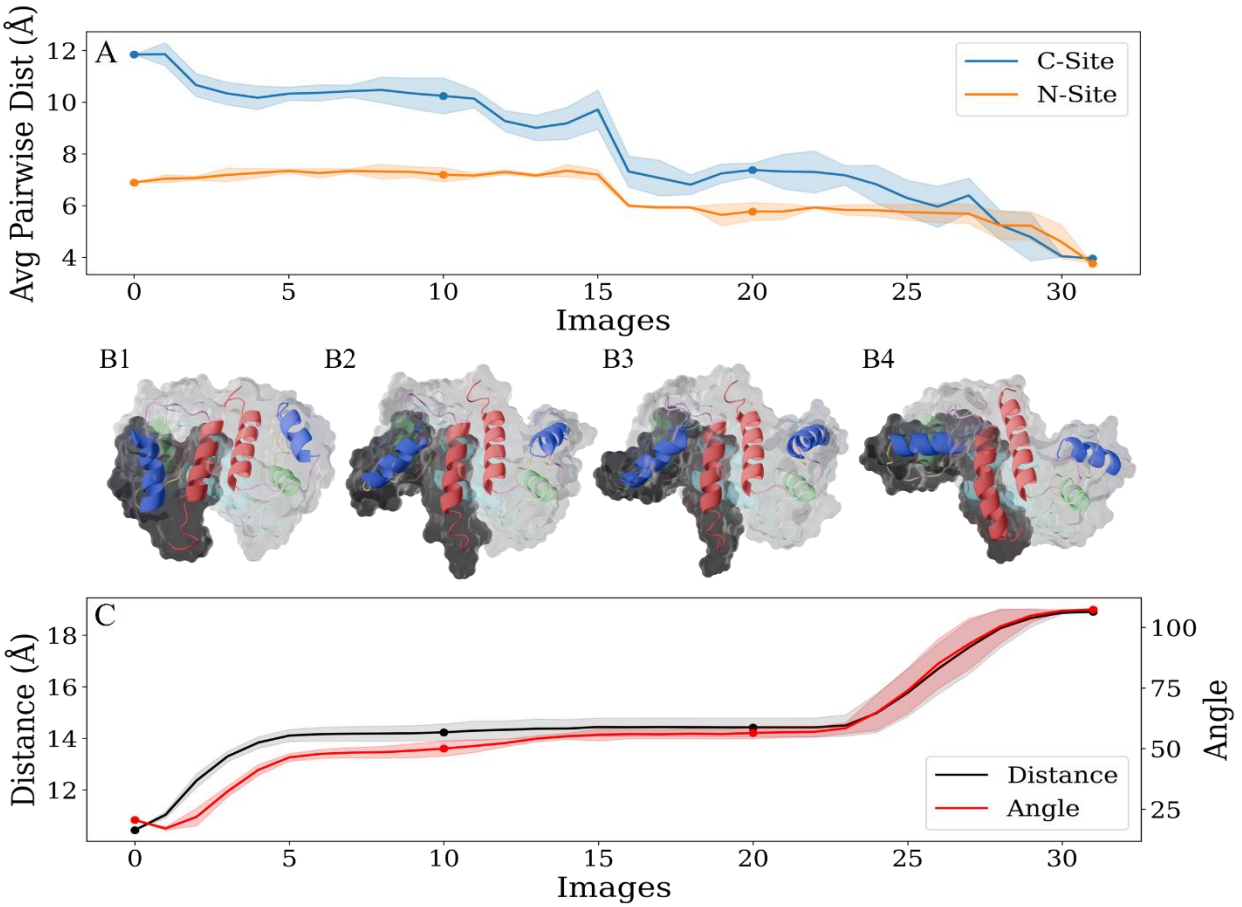

Figure 3.2: (A) Average pairwise distance between Ca^{2+} binding atoms in both pEF (orange) and cEF (blue) binding loops in S100A6. (B1-4) Cartoon and surface representations of the structures throughout the path. (C) Distances (black) and angles (red) between HIII and HIV throughout the pathway in S100A6.

Variability in binding site formation between the pEF and cEF was observed. While both sites were observed to form in a stepwise fashion, the cEF exhibited more steps than with the pEF, as shown in **Figure 3.2A**. The cEF site appeared to contain three distinct periods in which the system rapidly progressed towards formation. These were located near the beginning, middle,

and end of the pathways, with two relatively stable intermediate stages being observed. Conversely, the pEF was observed to have only two periods of rapid formation progress which occurred in the middle and end of the path. While it is unclear whether cooperative binding is present from this evidence alone, the presence of concerted motion at the middle and end of the formation pathways does lend itself to supporting this theory. However, further evidence would be required for it to be conclusive and the lack of concerted motion at the beginning supports a two-step mechanism.

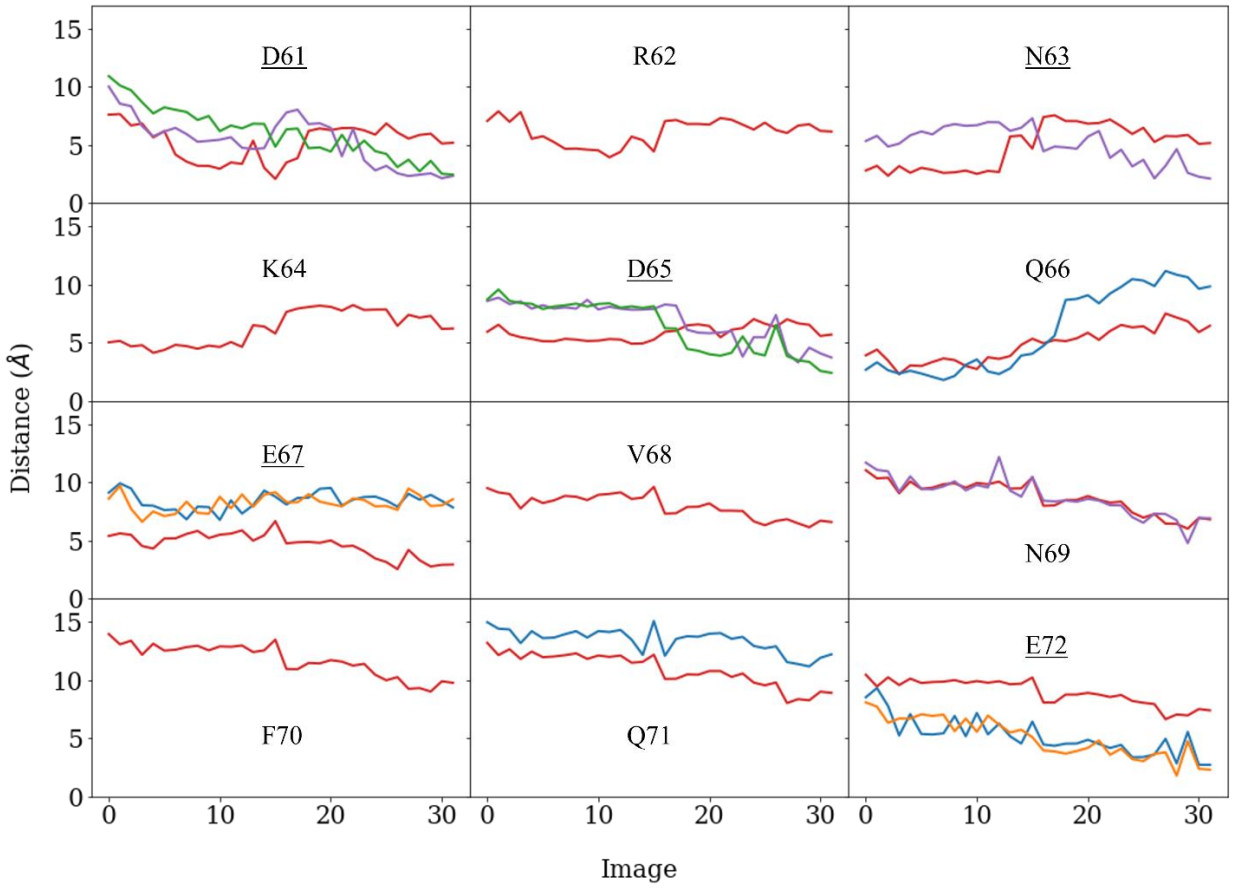

Figure 3.3: Plot of the distance between each oxygen atom and the average pairwise distance between Ca2+ coordinating atoms within the cEF binding site. The oxygens are colored according to each residue: Backbone (red), OE1 (Blue), OE2 (orange), OD1 (purple), OD2 (green), OH (yellow), and OG (cyan). Known Ca2+ interacting residues are underlined.

The oxygen atoms of both the pEF and cEF were tracked throughout the pathway in their relation to the average pairwise distance to characterize their potential contribution to binding site formation. These results of the tracked oxygen atoms within the cEF are displayed in **Figure 3.3**. We see that, within the cEF, a number of residues gradually progress towards site formation at different stages which are representative of the three stages observed in **Figures 3.2A**. Namely, D61 and E72 undergo rearrangements early in the path, resulting in the average pairwise distance decreasing by 2 Å. While not included in the average pairwise distance calculation, residue R62 was observed to shift towards the binding site. R62 is positioned on the distal side of HIII, away from the binding site, but contributes to a series of residue interactions. It may also partially contribute to the formation of the hydrophobic patch, but the exact role of R62 is unclear. A brief period of relative stability followed this initial period of progress. In this stable period, less movement of the residues was observed.

The second period of movement followed the stabilized period and initially consisted of slow shifting by multiple residues. A larger set of shifts also occurred within this period of movement due to a rotation within the binding loop. Overall, a decrease of 3.5 Å in average pairwise distance was observed. The residues of D61-Q66 all exhibited some capacity to rotate away from the binding site while residues E67-Q71 were all observed to rotate towards the binding site. The exception to this was residue E72, which remained primarily stationary throughout the pathway and instead of moving in relation to the binding site, the binding site moved in relation to this residue. One of the most significant shifts within the cEF belonged to Q66. It was observed that Q66 rotated through the cEF binding site to reach its final position.

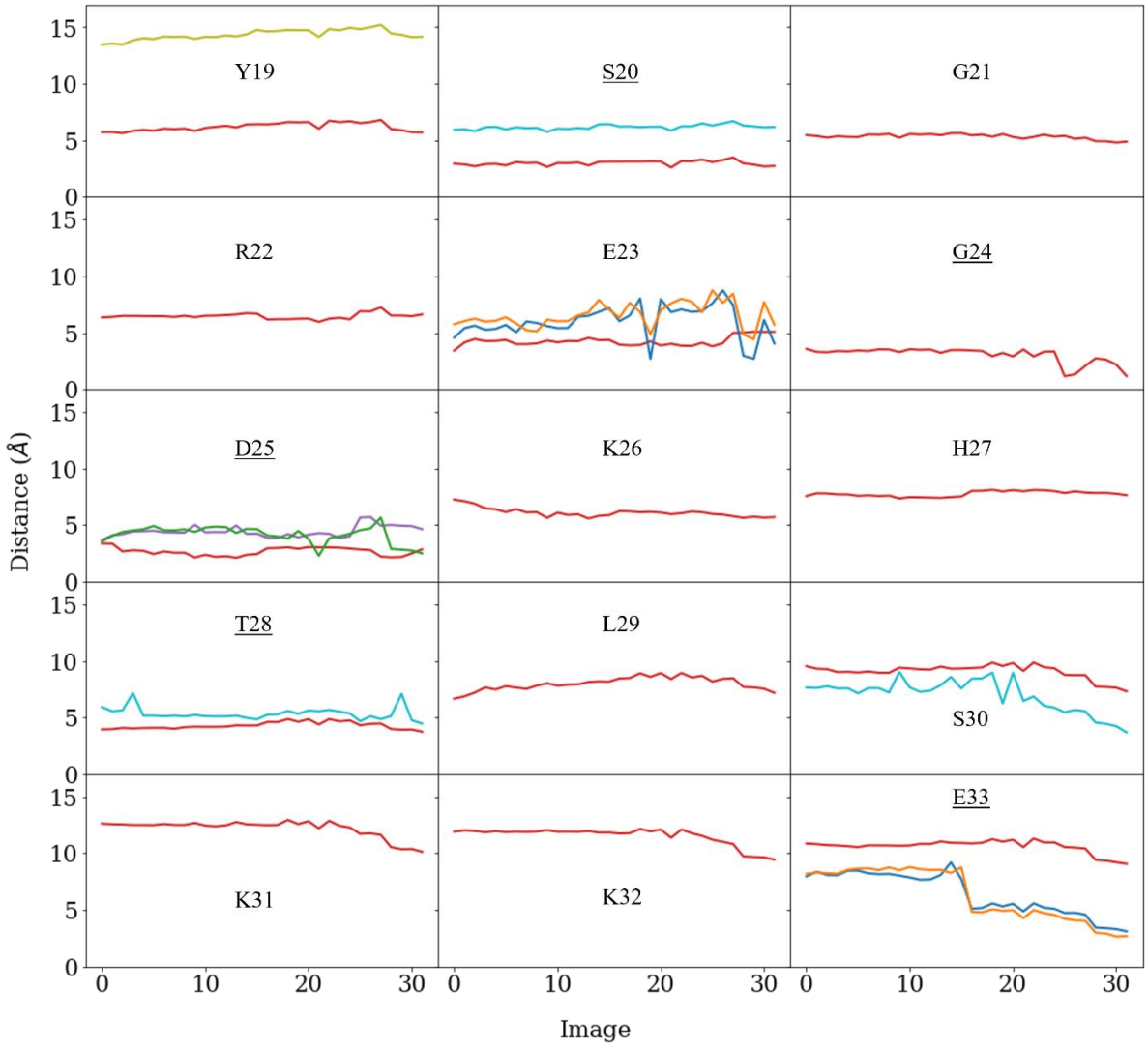

Figure 3.4: Plot of the distance between each oxygen atom and the average pairwise distance between Ca2+ coordinating atoms within the pEF binding site. The oxygens are colored according to each residue: Backbone (red), OE1 (Blue), OE2 (orange), OD1 (purple), OD2 (green), OH (yellow), and OG (cyan). Known Ca^{2+} interacting residues are underlined.

Interestingly, this rotation itself and second period of stability appeared to assist in positioning the binding loop to prepare it for the final period of movement. Specifically, the binding loop was observed to roll over and form interactions with HIII during this period. The final period of movement consisted of condensing movement of the entire loop itself. Many of the residues contributing to this movement were already shifting but began to move more in

concert with surrounding residues, leading to a decrease of another 3.5 Å in average pairwise distance.

Conversely, the pEF was observed to be much more positionally stable, as shown in **Figure 3.4**. Unlike with the cEF, the pEF exhibited initial stability. During this period, very little movement was observed in the majority of the residues. However, the residues of T28 and E23 were observed to undergo slight readjustments. E23, specifically, appeared relatively unstable as it interacted with a residue complex including both R22 and E33. Residue R22 has been observed to contribute to some RAGE interactions, and it is possible that this complex of residues is also involved but the extent is unknown.[59] However, in S100A6, this residue complex appeared to directly contribute to the stability of the pEF binding loop. The first major change within the pEF consisted of E33 shifting away from this complex. Once this shift occurred, the binding loop was observed to begin condensing around E33. This action was contributed to by residues E23-D25 and L29-K32. This final motion corresponded with the second major period of motion within the pEF and acted in concert with the final period of motion within the cEF.

3.3.2 *Hydrophobic Patch Formation*

In each of the completed NEB simulations, S100A6 was observed to undergo similar pathways between apo and holo conformations. HIII was, in fact, observed to undergo a shearing motion that revealed the hydrophobic patch, as described by Okazaki et al. To characterize the turn of HIII, the distance between residues located on the N-terminal end of HIII and C-terminal end of HIV were calculated for each image. A similar calculation was performed to determine the change in angle between the same helices. The results of these analyses are displayed in **Figure**

3.2C. Snapshots of the system at images 0, 11, 20, 32 and are displayed in **Figures 3.2B1-4 and 3.5-3.8**. These snapshots are representative of four distinct regions of the pathways. HIII was observed to undergo two major shifts in both angle and distance relative to HIV. Along the pathway and between these two shifts is an extended period of HIII stabilization. However, this period is distinctly different between its initiation and termination.

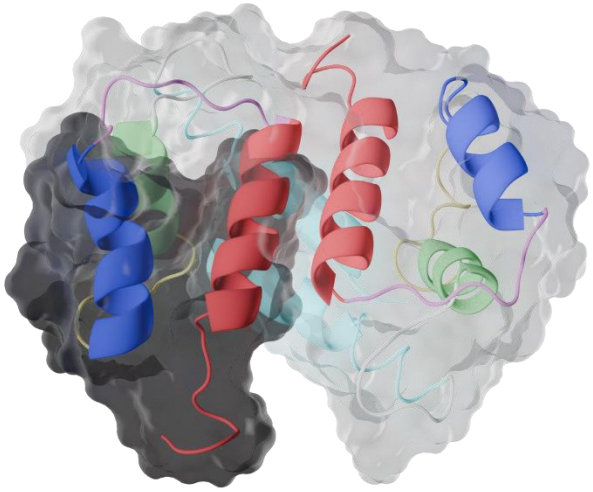

Figure 3.5: Cartoon and surface representations of S100A6 along the path at Image 0.

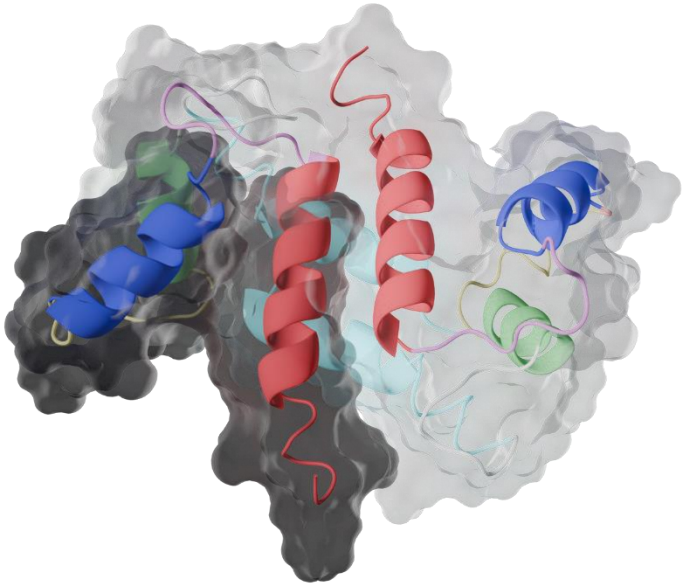

Figure 3.6: Cartoon and surface representations of S100A6 along the path at Image 10.

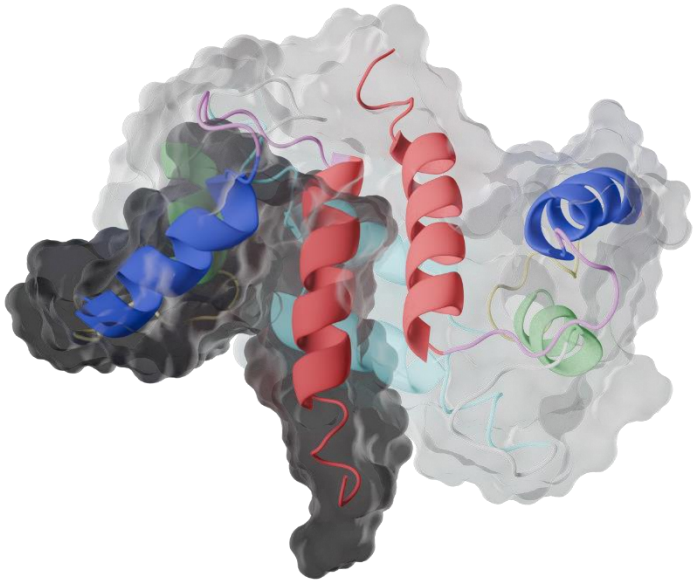

Figure 3.8: Cartoon and surface representations of S100A6 along the path at Image 20.

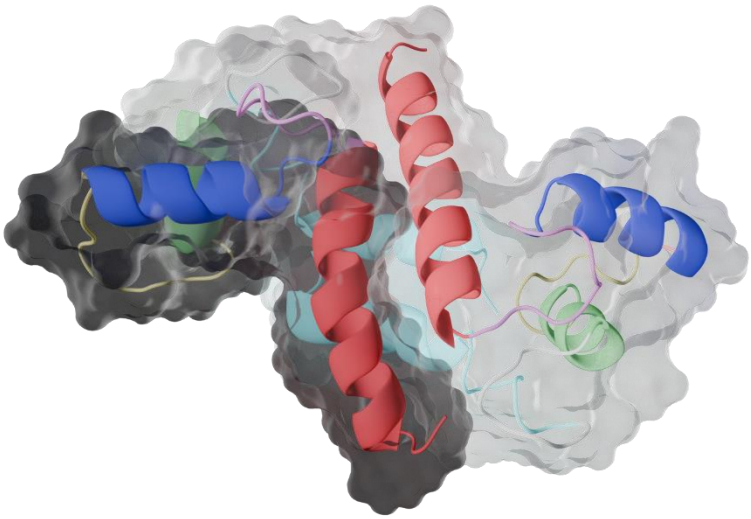

Figure 3.7: Cartoon and surface representations of S100A6 along the path at Image 31.

We see that within the apo conformation, HIII rests roughly perpendicular to HIV, as shown in **Figure 3.5**. As the system progresses, we see a rapid initial shift of HIII. This shift

consisted of HIII moving roughly 4 Å away from HIV and turning almost 35° until the system had entered the described period of stability. The turn itself was anchored near the cEF binding site, with the N-terminal end of HIII experiencing the greatest degree of movement. While very little change in HIII was observed during this time, it was observed to slowly reorient by almost 10°. The intermediate conformation can be observed in both **Figure 3.6 and 3.7**. Following this intermediate conformation, HIII was observed to undergo a second period of rapid conformation change as the system reached the holo conformation, as shown in **Figure 3.8**. However, these analyses cannot completely characterize the formation of the hydrophobic patch.

Previous work completed by Lee et al. utilized Siah-1 Interacting Protein (SIP) to develop a novel mode for S100 protein recognition.[97] Specifically, a peptide sequence of SIP was identified to interact with the hydrophobic patch of S100A6 using residues L42, I44, E52, I53, L56, F76, L80, and I83. Using these identified residues, we calculated the solvent accessible surface area (SASA) to understand the possibility of interactions between S100A6 and partner proteins throughout the path. We see that the initial conformation change of HIII corresponds to a rapid increase in surface accessibility and partial formation of the hydrophobic patch, as shown in **Figures 3.6 and 3.9**. The accessible surface area was observed to fluctuate around 700 $Å^2$ during the intermediate conformation. Immediately prior to and during the second major conformation change of HIII, the SASA was observed to more steadily decrease until it reached 590 $Å^2$.

Throughout the course of the intermediate conformation, the described fluctuation was observed to trend towards greater accessibility despite a single period of rapid decrease. We attribute this to the behavior of the linker region connecting the pEF and cEF, in which it slowly

peels away from the hydrophobic patch region. This introduced the question of what is the physiological utility of the extended intermediate conformation? A potential explanation for this lies in the fact that S100 proteins most easily undergo the major hydrophobic patch formation in the presence of partner proteins.[7,11,12]

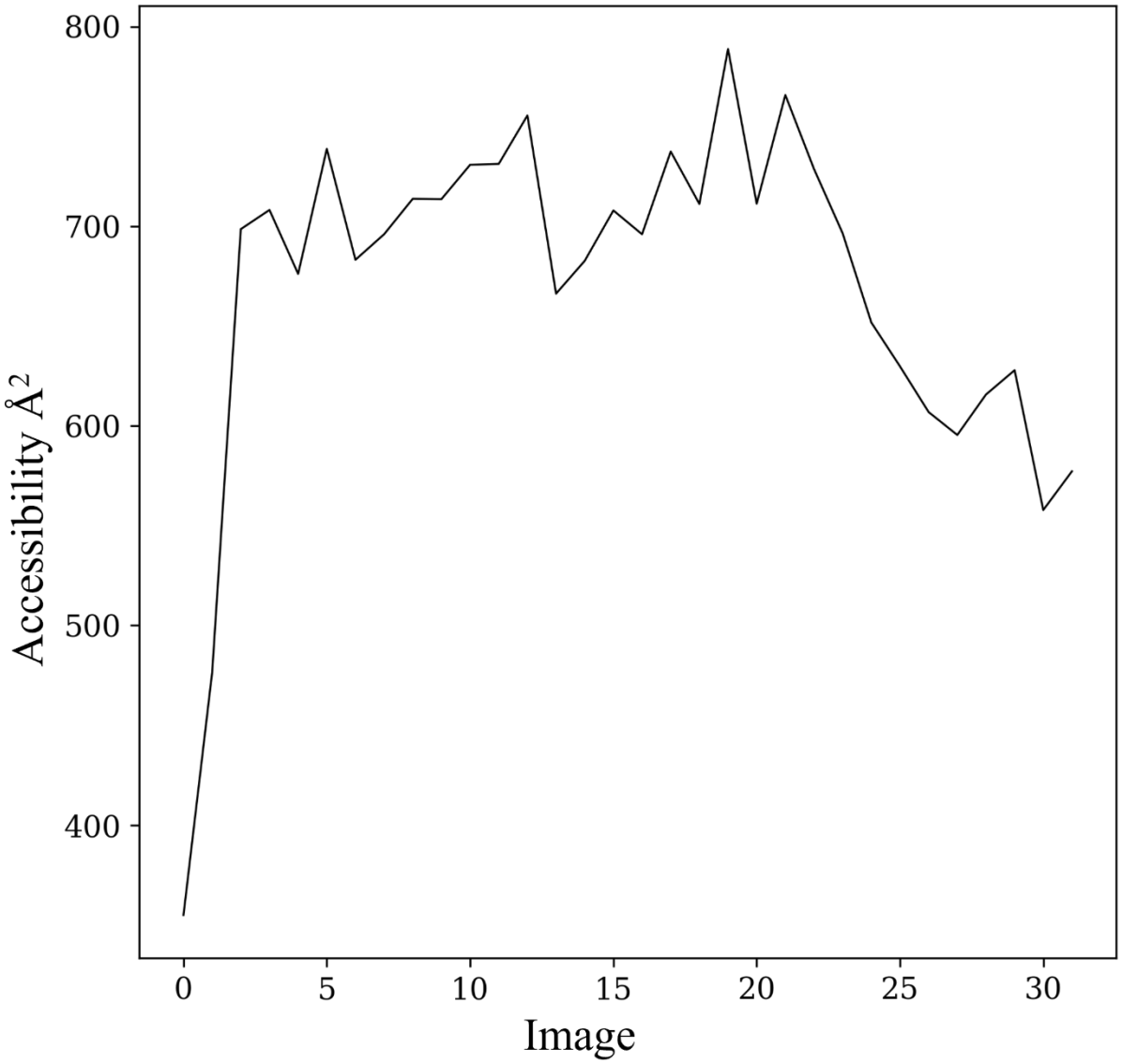

Figure 3.9: Plot of the solvent accessible surface area throughout the S100A6 pathway.

While it is known that sufficient concentrations of Ca^{2+} ions are capable of causing patch formation without partner protein assistance, it is also known that the presence of partner proteins greatly increases the affinity with which S100 proteins bind Ca^{2+}. It is possible that patch residues, in this highly accessible state, can form interactions with target proteins. The formation of these interactions can assist in positioning HIII such that the interactions between the cEF binding site and HIII occur more rapidly, thus decreasing the possibility of Ca^{2+} ions escaping. It is observed that within this period, binding site formation has only partially progressed and has not met full formation. In the absence of a fully formed binding site, Ca^{2+} ions are less restricted and more easily able to escape the binding site. By more easily forming, the cEF binding site would more often retain ions, increasing the binding affinity of Ca^{2+}. Given the slow rate at which the intermediate conformation progresses, internal mechanisms are likely slow to meet conditions required for the final turn of HIII. In the presence of a partner, many of these mechanisms can be bypassed while also increasing binding affinity for Ca^{2+}.

3.3.3 *Contact Analysis*

The work from Okazaki et al. has, thus far, been supported by several completed analyses in this study.[95] Specifically, HIII appears to transition to between apo and holo conformations in a shearing motion although we observed this to occur in steps. This has been confirmed both visually and through the characterization of HIII motion, as shown in **Figure 3.2C**. Additionally, the system appears to break initial contacts prior to forming contacts native to the holo conformation. To investigate the sequence in which the conformation occurs, the

pairwise distance between all residues was calculated for the beginning, middle, and end of the pathway, as shown in **Figure 3.10**

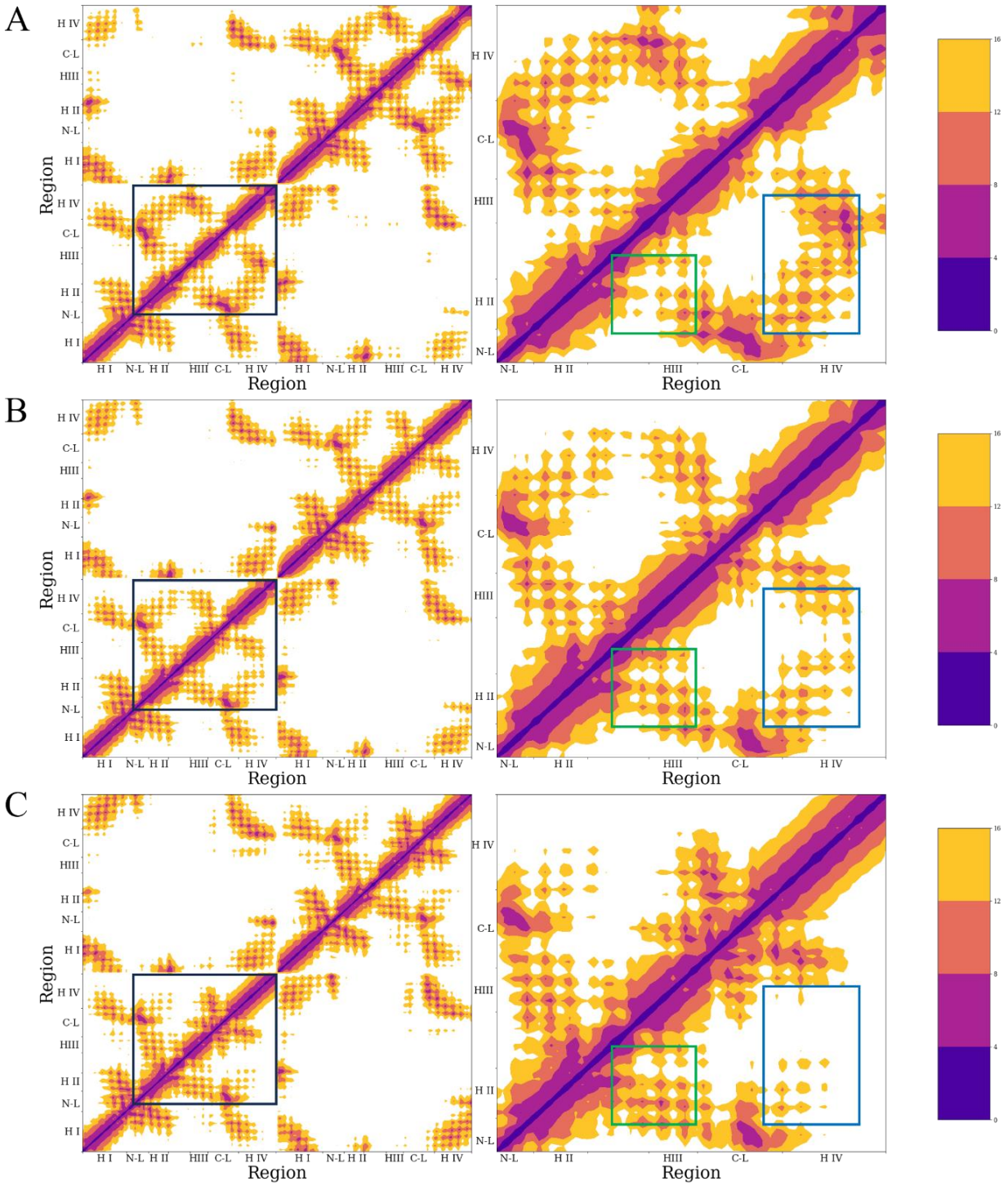

Figure 3.10: Pairwise distance between residues at Images 0 (A), 15 (B), and 31 (C). The region within the black square has been enhanced for greater detail. The regions within the blue and green squares indicate regions of contact loss and formation, respectively.

We see in **Figure 3.10A**, the linker-HIII region exists primarily in contact with HIV while maintaining loose association with HII. Although close contact is not observed, it can also be observed that HII is in relative proximity to HIV. These interactions contributed to a condensed shape in which the protein's surface appeared rounded. In this shape, the hydrophobic regions of the protein were well protected from solvent interactions. For reference to the shape of the protein surface throughout the conformation pathways, see **Figures 3.5-3.8**. Due to this, the conformation of the protein was not conducive to the formation of partner protein interactions.

We see that as the system progresses, there is a significant decrease in the contacts between HIII and HIV. Conversely, we also see a gradual strengthening of contacts between the linker-HIII region and HII. As these changes occured, HII shifted away from HIV, elongating the protein structure. Given that this calculation occurred within the intermediate conformation, partial hydrophobic patch formation had occurred, and the patch was, to some extent, revealed. The small reorientation of HIII had also been completed and the surface of the protein was similar to what is shown in **Figure 3.7**. It was at this point that we observed the largest degree of patch residue availability, as observed in **Figure 3.9**. Having not yet fully broken contacts with HIV or fully forming contacts with HII, HIII was positioned in a more exposed state, resulting in the residues of the patch becoming more accessible.

This trend of decreasing contact with HIV and increasing contact with HII continues as the system progresses to the fully formed holo conformation. Here, it can be observed that almost all associations between the linker-HIII region and HIV have been broken, with strong contacts having formed between the linker-HIII region and HII. Similarly, the largest distance between HII and HIV was observed here. This provided the protein with an elongated shape, differing from the rounded shape observed in the first calculation. Although this calculation

includes the fully formed hydrophobic patch, a decrease in availability is observed in this conformation. However, it must be noted that the patch residue accessibility is significantly higher here than in the apo conformation, as shown in **Figure 3.9**. Given the correlation between contact formation and decrease in accessibility, it is likely that linker-HIII and HII interactions are directly responsible. Specifically, HIII appears to exist in a more stable position given its interactions with HII that likely serve to prevent hydrophobic patch closure. This conformation also corresponds to a fully formed binding site, which as previously mentioned, likely assists in retaining Ca^{2+} ions.

3.3.4 *Residue Contributions to Hydrophobic Patch Formation*

Large scale protein conformations are key to the overall function of the protein itself. By understanding what conformation changes occur, we can better grasp how to treat various pathologies. However, understanding what conformation changes occur does not complete the entire story. To do this, we must also identify the underlying mechanisms in place that allow for the conformations to occur. We completed MMGBSA calculations to assist in better understanding the residue contributions to hydrophobic patch formation in our system. Combined with visual inspection, we identified several residues that we believe are essential in most S100 proteins, not just S100A6. These MMGBSA calculations were used to determine the differences in pairwise residue interaction energies between the apo and holo conformations. These results, along with a table of all interactions with an absolute energy value greater than 5 kcal/mol are reported in **Figure 3.11 and Table 3.1**. While this calculation does not fully describe interactions along the pathway, it can be used to identify some of the most important residues involved in the formation of the hydrophobic patch. From these calculations, we

identified two residues that are significant throughout the path. Namely, these residues are K31 and D50. However, there are a number of other residues that contribute to strong interactions.

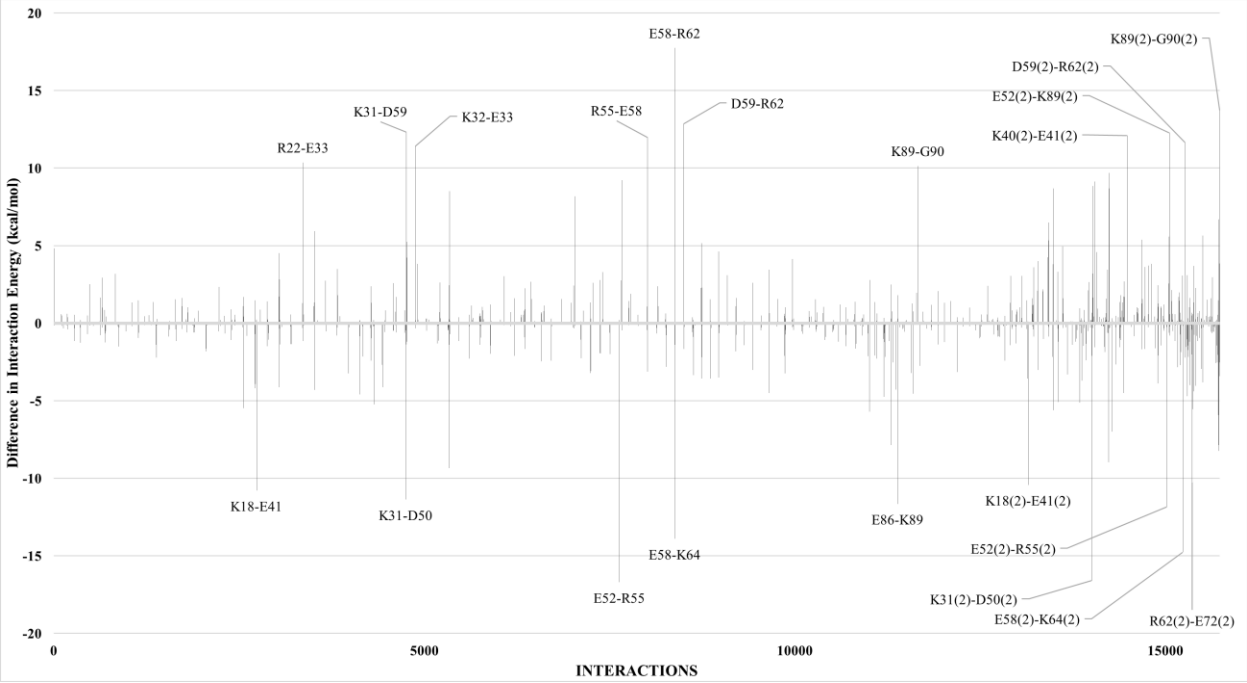

Figure 3.11: Diagram of the differences in pairwise interaction energies in S100A6. Interactions with absolute energy values greater than 10 kcal/mol have been labeled.

Table 3.1: Pairwise interactions in S100A6 with absolute energy value greater than 5 kcal/mol. Bolded residues indicate discussed interactions.

Residue 1	Residue 2	Difference (kcal/mol)
GLU 58	**ARG 62**	**17.73854022**
GLU 52	ARG 55	**-16.70533702**
LYS 31(2)	**ASP 50(2)**	**-16.60712841**
GLU 58(2)	LYS 64(2)	**-14.76811368**
GLU 58	LYS 64	**-13.9044986**
LYS 89(2)	GLY 90(2)	**13.713246**
ASP 59	ARG 62	**12.84153648**
LYS 31	**ASP 59**	**12.30901863**
GLU 52(2)	LYS 89(2)	**12.28442344**
LYS 40(2)	GLU 41(2)	**12.09753023**

ARG 55	GLU 58	**11.99537689**
GLU 52(2)	ARG 55(2)	**-11.8370138**
GLU 86	LYS 89	**-11.65111798**
ASP 59(2)	ARG 62(2)	**11.64847626**
LYS 32	GLU 33	**11.40481279**
LYS 31	**ASP 50**	**-11.37472859**
LYS 18	GLU 41	**-10.79075433**
LYS 18(2)	GLU 41(2)	**-10.42526588**
ARG 22	**GLU 33**	**10.35993303**
ARG 62(2)	GLU 72(2)	**-10.25760889**
LYS 89	GLY 90	**10.13397593**
LYS 35(2)	GLU 58(2)	**9.678010942**
LYS 35	**ASP 50**	**-9.336346639**
GLU 52	LYS 89	**9.211958879**
LYS 32(2)	GLU 33(2)	**9.129797832**
LYS 35(2)	**ASP 50(2)**	**-8.959258382**
LYS 31(2)	**GLN 66(2)**	**8.837407026**
LYS 35(2)	ASP 61(2)	**8.69157647**
GLU 23(2)	LYS 32(2)	**8.687060795**
LYS 35	ASP 61	**8.502230489**
ASN 85(2)	GLU 86(2)	**-8.233643394**
LYS 46	ASP 6(2)	**8.184248358**
ASN 85	GLU 86	**-7.862288661**
ASN 85(2)	GLY 90(2)	**-7.856896876**
ASP 59(2)	ASN 63(2)	**7.187606411**
GLU 36(2)	LYS 40(2)	**-6.9867035**
ASN 85(2)	LYS 89(2)	**6.692690736**
ARG 22(2)	**GLU 33(2)**	**6.479774738**
GLU 23	**LYS 32**	**5.941028376**
TYR 84(2)	ASN 85(2)	**-5.924957246**
LEU 82	ASN 85	**-5.709020514**

ASN 69(2)	GLU 72(2)	**5.650732778**
GLU 23(2)	GLU 33(2)	**-5.60485398**
GLU 52(2)	ASN 85(2)	**5.588972414**
ARG 62(2)	THR 75(2)	**-5.540135666**
HIE 17	LYS 26	**-5.479461078**
ILE 44(2)	GLN 49(2)	**5.371881975**
ARG 22(2)	GLU 36(2)	**5.342271378**
LYS 31	**GLN 66**	**5.251493721**
THR 28	GLU 67	**-5.238533679**
ASP 61	LYS 64	**5.150710314**
THR 28(2)	ASN 69(2)	**-5.120776602**
GLY 24(2)	GLU 33(2)	**-5.079015428**

While previously discussed, both R22 and R62 were implicated in strong interactions due to their association with residue complexes. Further inspection of R22 revealed that it remains positionally stable throughout most of the path, sharing strong interaction with E33 in both monomers (-10.4 & -9.1 kcal/mol). However, a shift of residue E33 from this complex to the pEF binding site corresponded to movement of R22. Specifically, R22 more strongly interacted with E35 in the absence of E33. While the exact effect of this change is hard to characterize, we believe that R22 assists in stabilizing the nearby pEF binding loop and that this residue complex, as a whole, is involved.

Further review of R62 confirmed its initial association with a residue complex located on HIII. While we initially thought that this residue may assist in stabilizing HIII within the apo conformation, very little correlation was found between R62 contact loss and movement of HIII. These results indicate that this residue may assist in the positioning of the cEF binding loop in

the absence of Ca^{2+} alongside its role in membrane interactions, as previously discussed. In the absence of Ca^{2+}, R62 strongly interacts with D59 in both monomers (-12.8 & -11.6 kcal/mol). In this position, R62 is located at the junction between HIII and the cEF binding loop itself such that the binding loop is held open. As S100A6 transitioned into the holo conformation, some secondary structure was lost as the loop containing R62 began to rearrange. Once this interaction was broken, the loop was more easily able to roll over and condense around the binding site.

Visual inspection revealed that residue D50 is located at the point which the linker and HIII regions meet. This point is at the opposing end of HIII from where the cEF binding site is positioned, meaning that it experiences strong movement as HIII turns. Prior to the turn, D50 was observed to interact with very few residues in a stable manner, indicating a very minor role outside of the intermediate and holo conformations. Furthermore, D50 was identified to interact strongly with K31 once the turn of HIII was complete (-16.6 & -11.3 kcal/mol). Once this interaction formed, both ends of HIII were anchored in place, fully forming the hydrophobic patch. Similarly, D50 was observed to interact strongly with K35, but not as strongly as with K31, within the holo conformations (-9.3 & -9.0 kcal/mol). While both K31 and K35 are involved in stabilizing HIII through interactions with D50, our results indicate K31 is the primary interaction partner.

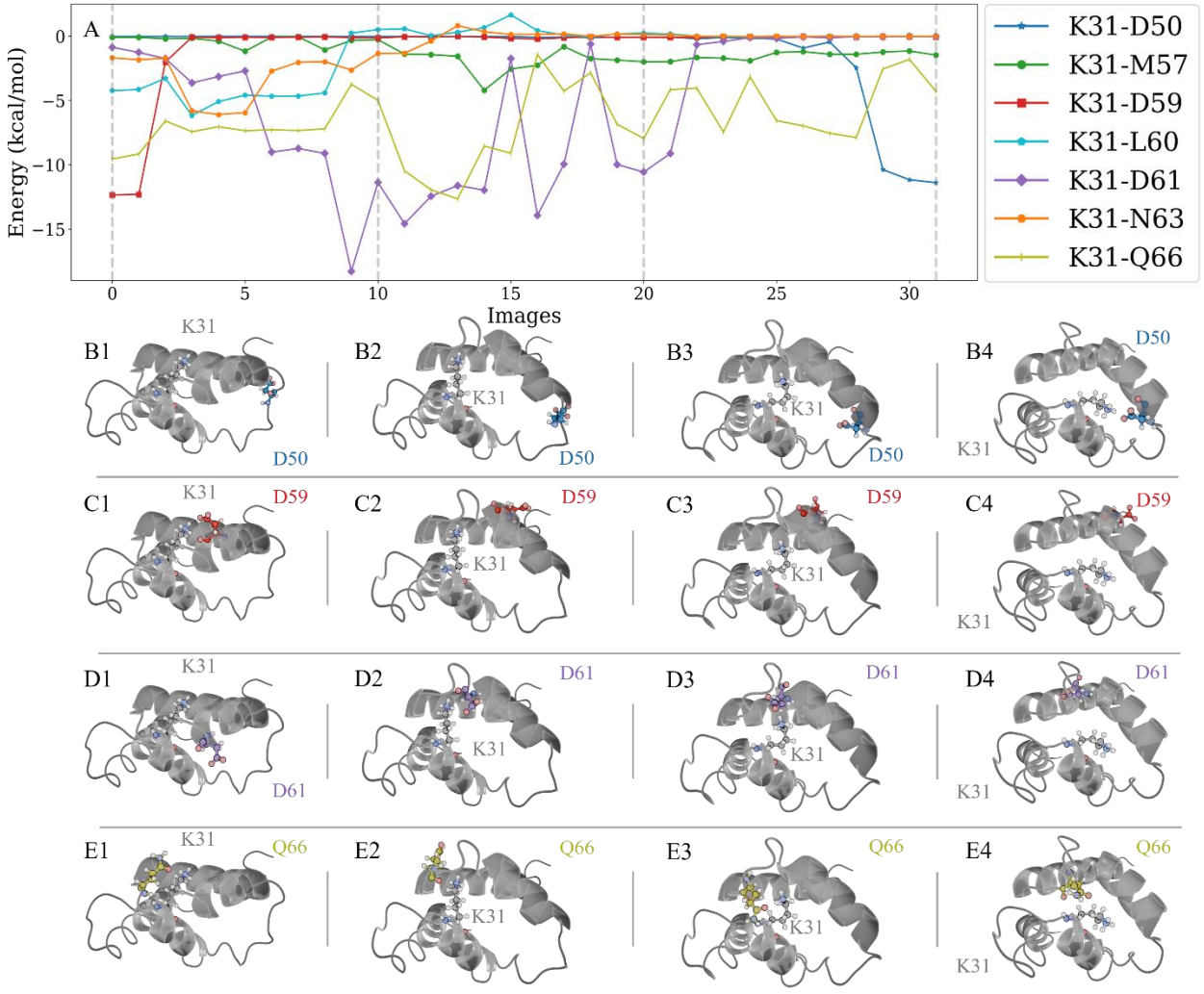

Figure 3.12: (A) Plot of the 7 strongest pairwise interaction energies between K31 and residues along its conformation path in S100A6. (B-E) Diagrams of the residue interactions of K31 throughout Images 0, 10, 20, and 32. These residues include D50 (B), D59 (C), D61 (D), and Q66 (E). Only a single monomer was included in the diagrams to ensure clarity.

While the aforementioned residues are likely involved in a number of functions, visual review of K31 revealed that it is likely key to the mechanism by which HIII turns. This residue is initially located within the binding region of the pEF. However, the sidechain of K31 is positioned within the cEF binding site. Throughout the pathway, it was observed that K31

undergoes a long interaction path where it forms the described interaction with D50. In the path itself, K31 appeared to interact with a number of residues along the side of HIII. These observations warranted further study to understand the path of K31 and identify the role of the residue itself. To accomplish this, MMGBSA calculations were performed for the total pathway. In this K31 interacted strongly with several residues. The strongest of these have been plotted against each other to establish a timeline of K31 interactions, as shown in **Figure 3.12**. The difference in pairwise residue interaction energy between apo and holo conformations was useful in identifying both D50 and D59 as key stabilizing interactions at the start and end of the pathway, respectively. Further analysis confirms these residues as being the primary interacting agent in their respective conformations.

While D59 was the predominant interaction partner of K31 in the apo conformation, several other residues were involved in a complex. These residues include L60, D61, N63, and Q66. We observed that this complex acts as a mechanism by which the structure of the cEF loop is maintained in the absence of Ca^{2+} in addition to being the initial Ca^{2+} sensor. As binding site formation occurs within the cEF, an ejection of K31 from the site occurred (see **Figures 3.2A, 3.12A, and 3.12B1-2**). During this ejection, the first turn of HIII was also observed in addition to a shift in position of HII.

The position of K31 following the ejection was such that it could form a stronger interaction with D61 once the ejection occurred, as shown in **Figures 3.12A and 3.12D1-3**. While D61 is local to the cEF binding loop, this interaction corresponded to the small reorientation of HIII that occurred during the intermediate conformation as the K31-D61 interaction reached its most stable point (<-15 kcal/mol). This correspondence indicates that the K31-D61 interaction may assist in positioning HIII in the exposed state. Given the possibility

that the exposure of HIII may increase the efficiency of S100A6 by allowing protein interactions, it is also possible that this interaction may act as a checkpoint to prevent HIII from falling back to the apo conformation.

While the interaction between K31 and D61 exists during most of the intermediate conformation, a rapid decrease in interaction energy was observed at image 22. Structurally, this did not appear to coincide with any major events within the pathway. However, this change occurred almost simultaneously to the start of a steady decline of accessibility observed in the pathway's later stages (see **Figures 3.9 and 3.12A**). As the K31-D61 interaction diminished, Q66 became the primary interaction partner of K31 (see **Figures 3.12A, 3.12D4, and 3.12E1-4**). While this interaction was not as strong as many of the other K31 interactions, it was by far the strongest occurring at that point in the pathway. As previously mentioned, Q66 was observed to rotate through the cEF binding site. We observed the rotation to allow for interactions between Q66 and K31, where it appeared to act as a guiding mechanism. Given that K31 is known to have a long interaction path, it would be reasonable to expect some mechanism to assist in this process. We observe Q66 to fulfill this role in S100A6, allowing for K31 to regulate the position of HIII throughout the entire pathway. In the absence of Q66, the only remaining interaction partners for K31 during this period would have much smaller interaction energies and a decreased ability to maintain K31 position and orientation.

The conclusion of the K31 pathway is composed of the interaction formed with D50, anchoring the position of HIII, and maintaining the full hydrophobic patch. From apo to holo conformations, K31 is directly involved through all stages of the conformation. While contributions from a number of residues assist in hydrophobic patch formation, the four major contributors are those discussed here: K31, D50, D59, and D61, as shown in **Figure 3.12A**.

Interactions between these residues and K31 correspond to four major stages throughout the pathway and are integral to the process as a whole. Specifically, D50 and D59 define the apo and holo conformations, respectively, while D61 and Q66 define two stages associated with the intermediate conformation. While we have proposed one answer as to why the intermediate conformation exists, results indicate that this period also allows for K31 to traverse this path

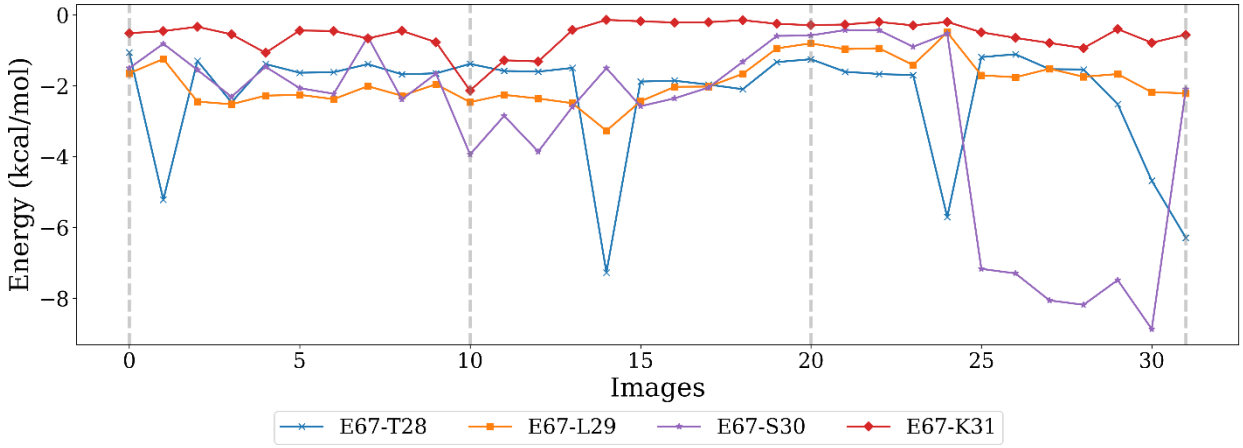

Figure 3.13: Pairwise interaction energy of E67 with the four strongest interaction partners in S100A6.

without external assistance.

While residue E67 was not identified as having any strong differences in interaction energy between apo and holo conformations, it is believed to have a role in communicating Ca^{2+} binding between the binding loops. To further study this, MMGBSA calculations were extended to include this residue. The results from these calculations are shown in **Figure 3.13**. Positionally, E67 is located within the cEF binding loop and resides near binding loop residues of the pEF. Throughout the path, E67 held a relatively fixed orientation. Instead of directly interacting with the binding site, the side chain of E67 was observed to interact strongly with residues T28-K31. E67 interactions with L29 and K31 were observed to fluctuate throughout the

path but did not exhibit any significant changes. However, both T28 and S30 were observed to peak at various stages of the conformation change.

The interactions between E67 and T28 appeared somewhat random, although they corresponded to the major periods of cEF binding site formation. Namely, this interaction energy was observed to peak in the range of -4 to -7 kcal/mol during or immediately prior to the major periods of formation (see **Figure 3.2A and Table 3.1**). It is possible that these interactions serve as signaling events. However, the exact cause of the signal initiation is unclear. An explanation for this may lie, once again, with the positioning of K31. Given the relative proximity of T28 to K31, some indirect influence may be present. At each of the points of T28 interaction energy peaking, we also observe some change in position of K31. Specifically, the first peak corresponded to the ejection of K31 while peak two corresponded with the small reorientation of HIII caused by the K31-D61 interaction. The third peak did not occur in direct conjunction with any spike of K31 interaction energy but does occur as the cEF binding residues undergo a reorientation. An effect of this rotation was that of Q66 being brought closer to K31 which, as previously described, has been indicated to guide K31 to D50.

We see that as the second turn of HIII is in progress, interactions between E67 and S30 strongly and stably interact throughout its duration. While these results may further indicate that E67 may be involved in signaling Ca^{2+} binding, they are also not completely conclusive. Instead, we believe that K31 is more directly involved in signaling cEF binding to the pEF binding loop, occurring via the initial binding of Ca^{2+} to the cEF. As previously mentioned, the ejection of K31 was observed to correlate with a positional shift of HII. This is part of a rearrangement that is likely involved in Ca^{2+} binding. Once K31-D61 interactions had reoriented HIII, we also

observed the shift of E33 towards the cEF binding site. Only then did the pEF show signs of binding site formation, however gradual.

3.3.5 *Mutation Studies*

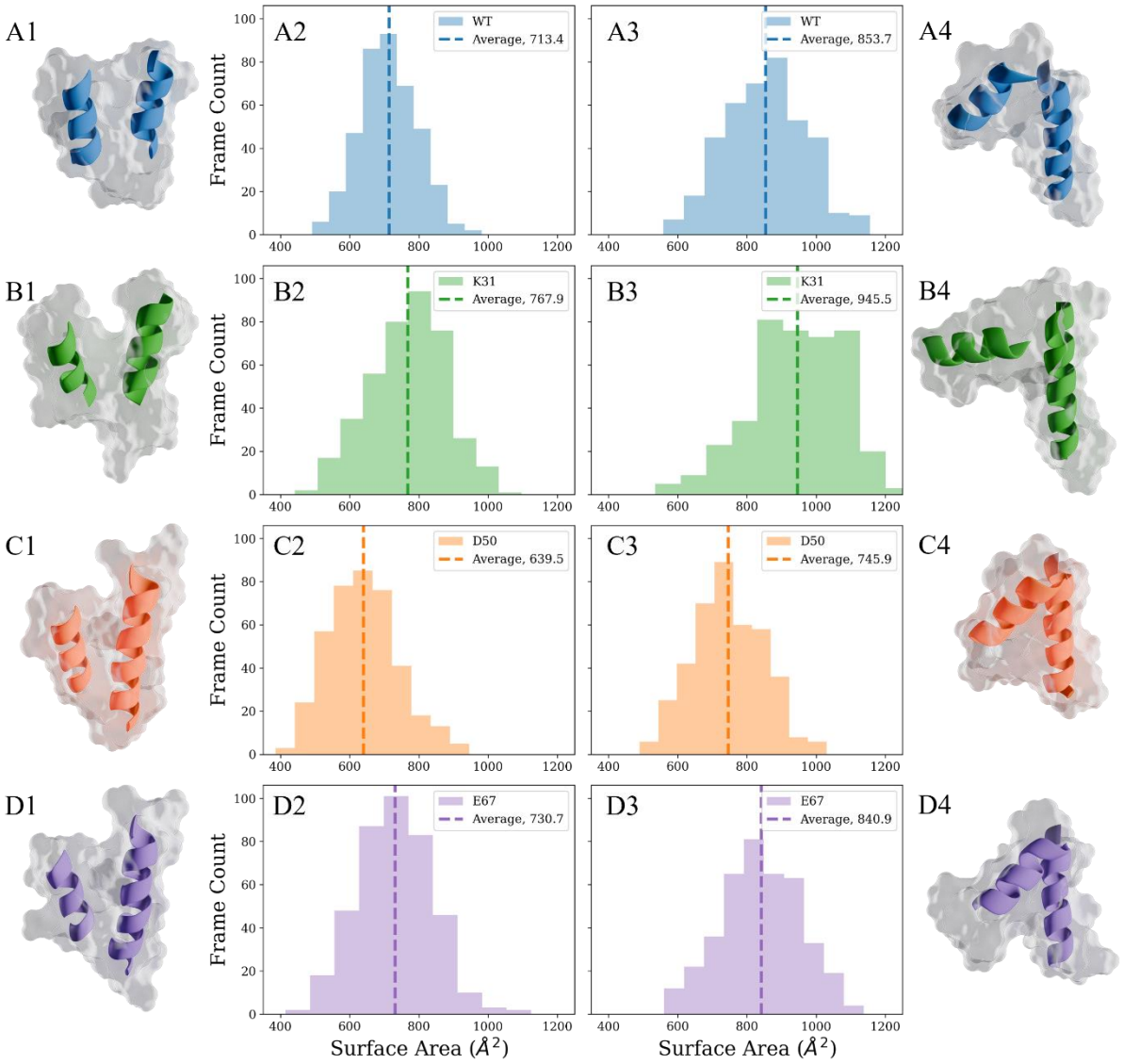

Figure 3.14: Diagrams and SASA calculations of the WT (A), K31A (B), D50A (C) and E67A (D) systems. These include the apo (1-2) and holo (3-4) conformations. Only HIII and HIV have been shown in the cartoon representation to ensure clarity.

Having identified several key residues and their potential roles, further study of these residues was warranted. To accomplish this, mutations to alanine were completed for the residues of K31 (K31A), D50 (D50A), and E67 (E67A). Our goal with these mutations was to observe resulting changes in residue availability. The results from these calculations are shown in **Figure 3.14**. Similarly, we calculated the average angle between HIII and HIV in all systems to determine how each mutation may have affected the position of HIII. These results are shown in **Table 3.2.** Accelerated molecular dynamics simulations were completed for both apo and holo starting conformations for all systems. Through these simulations, we were able to compare differences with wild type systems (WT).

Table 3.2: Average angle between HIII and HIV in apo (A) and holo (H) systems.

SYSTEM	AVERAGE ANGLE (°)
WT-H	95.4557620784005
WT-A	26.179378862643
K31A-H	104.094922259832
K31A-A	36.9889839489173
D50A-H	41.8179449130769
D50A-A	31.665811305381
E67A-H	60.2631253636774
E67A-A	19.4227677027816

The WT apo starting conformation system (WT-A) was not observed to exhibit any major conformation changes. However, both K31A and D50A apo starting conformations systems (K31A-A and D50A-A, respectively) displayed significant increases and decreases of their patch residue availability, respectively (see **Figures 3.14A1-2, 3.14B1-2, 3.14C1-2**). The K31A-A system was observed to have decreased stability in the region surrounding the cEF binding loop. Normally, K31 would rest within this binding site to stabilize this region with contributions from R62, as indicated previously in **Figure 3.12A**. These results indicate that either R62 does not

assist in stabilizing or is not capable of stabilizing the loop, alone, in the absence of K31. This destabilization resulted in an increase in surface area within this region as the C-terminal end of HIII was observed to shift further away from HIV, resulting in an increased average angle between the helices compared to the WT-A system (26.2° and 37.0°, respectively). However, the patch was still observed to be in a closed state, with the rest of HIII adopting a position typical to the apo conformation. It is possible that the lack of K31 affects the binding of Ca^{2+} in some way, given that the loop can more easily conform. It is also possible that this may have some detrimental effect if the loop tends toward some conformation that is suboptimal for Ca^{2+} binding. Most importantly, this mutation may disrupt or alter progress along the path between apo and holo conformations. While we are unable to fully prove this given the difficulties in simulating the turn of HIII, the prior implication of K31 throughout the path indicates that it may be required or at least beneficial.

The D50A-A system was observed to significantly decrease in patch residue accessibility. While the causes of this change were more difficult to identify than in the K31A-A system, it appeared to be due to a rearrangement associated with residues in the linker region. Within the NEB pathways, we observed a number of hydrophobic interactions between the linker-HIII region and HIV, as indicated by the blue box in **Figure 3.10A**. In the absence of a functioning D50 residue, this region appeared to interact more tightly and caused a decrease in residue availability as the linker-HIII junction interacted more strongly than in the WT system. Interestingly, a slight increase in angle was observed here compared to the WT-A system (26.2° and 31.7°, respectively) despite the decrease in SASA.

The E67A apo starting conformation system (E67A-A) was not observed to correlate with any significant changes in residue accessibility (see **Figure 3.14D1-2**). In this system, HIII was

observed to position itself in a manner somewhat different from the WT-A system, resulting in a decreased angle in relation to HIV compared to that of the WT-A system (26.2° and 19.4°, respectively). Given the lack of residue accessibility change, it is possible that this conformation change is within physiological bounds as HIII itself was observed to simply fluctuate in position. This fluctuation can also be observed in **Figure 3.14E2** in which residue accessibility varied more here than compared to the WT-A system.

In the WT holo starting conformation (WT-H) we observed the maintenance of the hydrophobic patch throughout the entire simulation (see **Figure 3.14A3-4**). While some fluctuation was observed, the patch remained open for the entire of the trajectory. However, in both K31A and D50A holo starting conformation systems (K31A-H and D50A-H, respectively), significant changes were observed. In the K31A-H system, we observed that the hydrophobic patch actually increased in availability, as shown in **Figures 3.14A3-4 and 3.14B3-4**. Throughout the NEB pathways, K31 was observed to interact significantly with HIII. Previous results indicate this relationship serves to assist in regulating the position of the HIII and to assist in the turn itself and prevent it from falling back into the closed conformation. Here, in the absence of K31, we observed no such interactions and HIII was allowed to turn further here than in the WT-H system (104.1° and 95.5°, respectively). We observed D50, the typical holo conformation interaction partner of K31, and it appeared to search for a new interaction partner that would allow for the stabilization of HIII. These results suggest that K31 is involved in both assisting in turning HIII in addition to preventing an overturn. Physiologically, there is reason to believe that a stop mechanism would be utilized. Given that protein shape determines function, a structure undergoing too large of a turn may disrupt the ability of a partner protein or molecule to

interact properly. Thus, our results indicate that K31 acts as a stop mechanism to prevent this problem from arising.

The D50A-H system displayed a significant decrease in patch residue availability, as shown in **Figures 3.14A3-4 and 3.14C3-4**. Upon viewing, it was noted that K31 formed interactions with residues along HIII in the absence of D50. These newly interacting residues were located in closer proximity to both K31 and the cEF binding loop itself and resulted in a significantly decreased angle compared to that of the WT-H system (41.8° and 95.5°, respectively). Given that K31 may prevent turning just as much as assisting in it, it is possible that these interactions result in premature termination of the turn. This change in conformation occurred rapidly within the simulation and was stable throughout it. These results confirm our belief that D50 is significant and is required for proper hydrophobic patch stabilization.

Similar to the E67A-A system, the E67A holo starting conformation system (E67A-H) did not display significant changes in patch residue accessibility. However, the patch did appear to close somewhat regarding the angle of HIII in which a 60.3° turn was observed, rather than the 95.5° of the WT-H system. This indicates that E67 may indeed have some role in stabilizing the patch. However, it is possible that this is limited to its influence upon the binding loop, where any change may disrupt the ability of the loop to effect proper conformation change. While HII did appear to adopt a conformation like that of the D50A-H system, the protein's surface was not observed to close in the patch region. The lack of change in patch residue accessibility here combined with the lack of changes observed in the E67A-A system indicate that while E67 may be involved, no conclusion can yet be derived as to E67 roles. It is possible, however, that E67 would have an increased role within the system's path and that our simulations were unable to capture the full extent of the mutations effect.

3.4 Conclusions

In this work, aMD and NEB methods were applied to S100A6. Given the highly conserved nature of S100 proteins and strong similarity within the conformation changes in these proteins, we believe that S100A6 provided a suitable model protein. In these simulations, hydrophobic patch formation was observed to occur along a similar pathway. The results of this study agree with the literature as HIII was observed to undergo a shearing motion throughout the path. However, we observed the presence of an intermediate conformation. The described shearing motion led both into and out of the intermediate conformation. It was also observed that contact breakage and formation occurred similar to what was described in the literature. Native apo contacts involving HIII were observed to break prior to the formation of native holo contacts between HII and HII. However, there is a degree of overlap that occurs while the system is within the intermediate conformation in which contacts are breaking and forming in concert.

Residue contributions to hydrophobic patch formation were also studied here. We identified the residues of K31 and D50 as having strong influence upon patch formation. While we believe that D50 is primarily involved in stabilizing the holo conformation, K31 is strongly implicated as being involved throughout the entire pathway. It is likely that K31 assists in the stabilization of the cEF binding loop and has a role in Ca^{2+} binding communication, positional regulation of HIII, and stabilization. E68, having been identified through the literature, was also studied here. We believe that there is some evidence to support that E68 is involved in Ca^{2+} binding communication between loops. However, this evidence did not conclusively prove specific function. Mutations to these residues further support the described roles of these residues.

Chapter 4: Characterization of Ca^{2+} Induced Conformation Changes in S100B and the Comparison of Internal Mechanisms Common to S100 Proteins

4.1 Purpose

The characterization of calcium-induced conformation changes in S100A6 revealed several mechanisms involved in hydrophobic patch formation. The purpose of this study was to further investigate similar mechanisms in other model S100 systems. In this work, we applied similar methodologies to find evidence of mechanisms involving residues here in S100B that are comparable to those of K31, D50, and E67 in S100A6. Similarly, this study also had the goal of characterizing the turn of HIII in S100B.

4.2 Computational Methods

4.2.1 *System Details*

In this work, aMD and NEB methods were used to study the path between apo and holo conformations in S100B. To identify relevant starting and ending conformations, a similar protocol was performed here as with S100A6. The use of aMD simulations paired with PCA and KDE allowed for the identification of these structures. In the aMD simulations, the crystal structures for S100B in the apo (1B4C) and holo (1QLK) were used.[72,98] PCA and KDE were used to identify the most highly sampled structures in the aMD simulations.

4.2.2 *Simulation Details*

Both aMD and NEB simulations utilized the protocols established in the study of S100A6. Similarly, the NEB simulations were completed in pentaplicate, allowing for proper sampling of the conformational pathway. The aMD simulations utilized an explicit solvent, using OPC water models and 12-6 OPC ion parameters.[99,100] The boost factors were calculated using

the same parameters as in S100A6. However, the energy values used in these calculations differed, given that they were calculated from S100B specific cMD simulations. Simulations using the NEB method utilized an implicit solvent with no Ca^{2+} ions being present. No additional simulations were completed in which various residues had been mutated.

Unlike with S100A6, the structure chosen for the endpoint exhibited differences between either monomer. Specifically, the conformation of M2 was similar to that observed in S100A6 with HIII undergoing a full turn and the linker region peeling away from the hydrophobic patch. This allowed us to compare internal mechanisms between M2 and those of S100A6. M1 exhibited similar behavior but contained slight variations, primarily regarding the linker region. Here, we observed the linker region in M1 to peel away from the patch area in a more limited fashion. It is possible that the different conformation in M1 allows for some physiological function, warranting investigation. The differences between these monomers allowed for both comparisons to suspected mechanisms in S100A6, and between fully and partially formed hydrophobic patches.

4.3 Results

4.3.1 *Ca^{2+} Binding Site Formation*

To understand the direct cause of the conformation change in S100B, we calculated the average pairwise distance between known Ca^{2+} binding residues. These results are shown here in **Figure 4.1A**. The atoms used in the calculation for the pEF were S19@O, E22@OE1, E22@OE2, D24@O, K27@O, E32@OE1, and E32@OE2. The atoms used in the calculation for the cEF were D62@OD1, D64@OD2, D66@OD1, E68@O, D70@OD1, D70@OD1, E73@OE1, and E73@OE2.

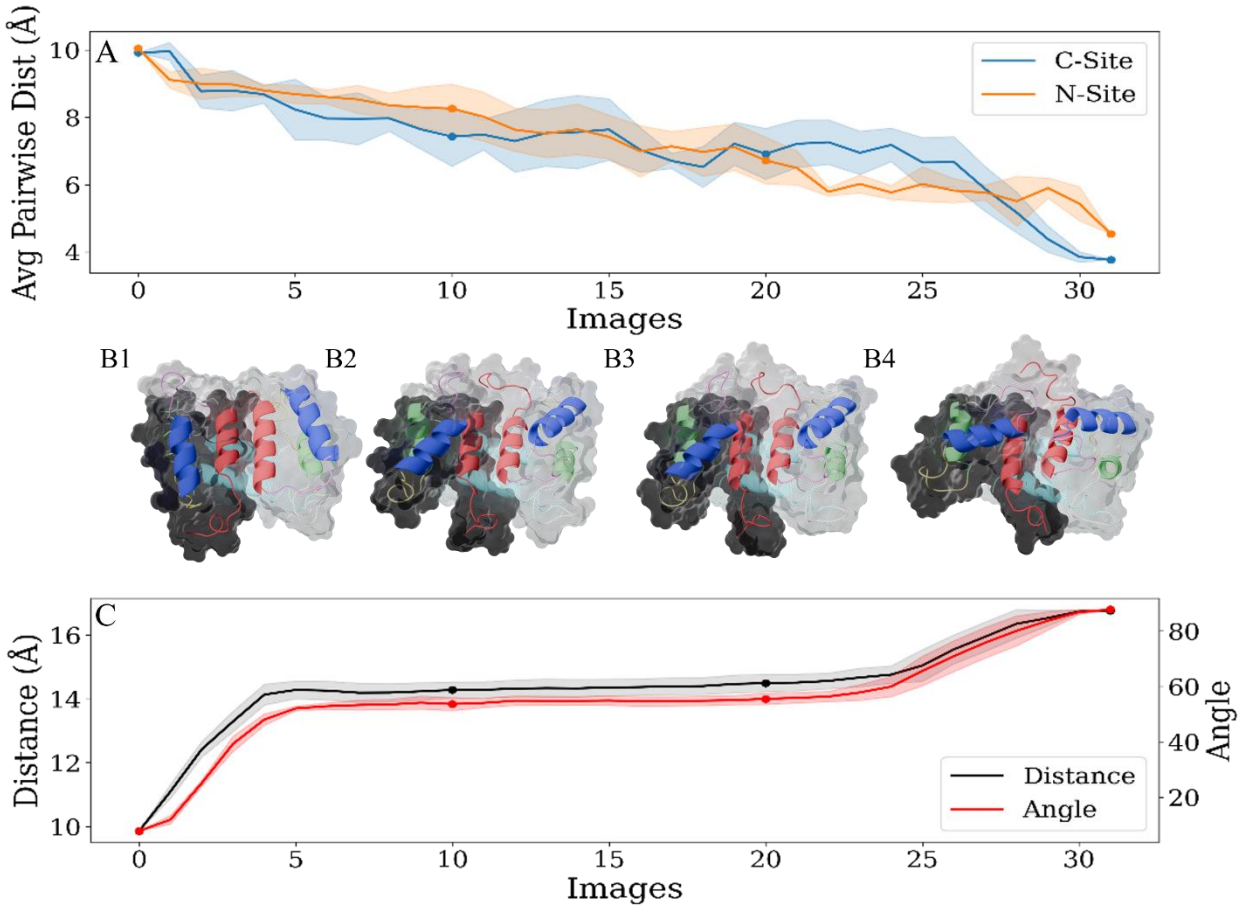

Figure 4.1: (A) Average pairwise distance between Ca^{2+} binding atoms in both pEF (orange) and cEF (blue) binding loops in S100B. (B1-4) Cartoon and surface representations of the structures throughout the path. (C) Distances (black) and angles (red) between HIII and HIV in S100B throughout the pathway.

We see that, unlike in S100A6, a more gradual formation throughout the entirety of the path, rather than periods of rapid formation progress. The cEF binding loop, while more gradual, remained similar in shape to that of the S100A6 cEF binding loop. The periods with the highest rate of formation correspond to those in S100A6 but the rate itself is decreased here. The most significant difference lies in the formation of the pEF binding site. The differences between the distance of each oxygen atom in the cEF binding site and the average pairwise distance was calculated for each image and are shown in **Figure 4.2**. These results allowed us to characterize binding site formation on the per-residue level.

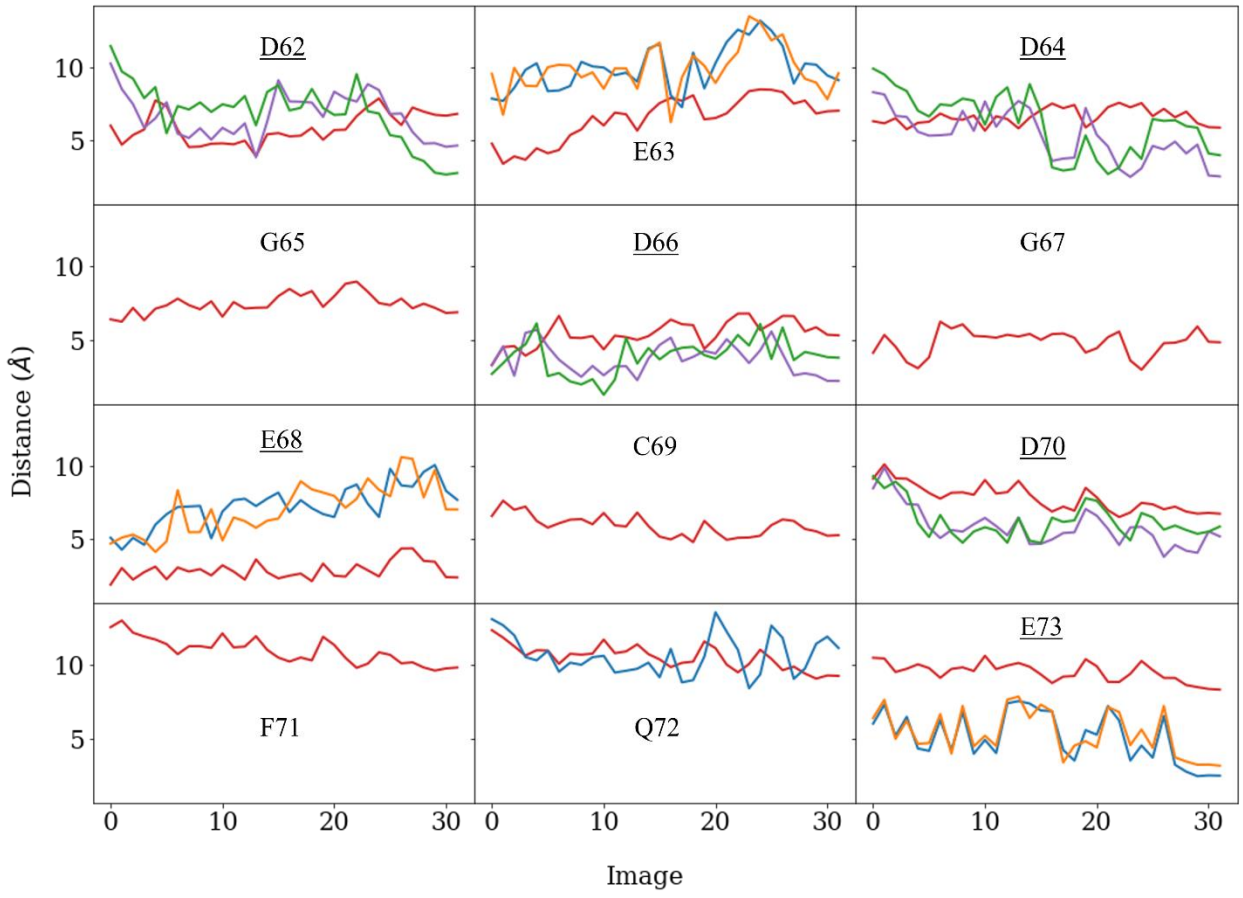

Figure 4.2: Plot of the distance between each oxygen atom and the average pairwise distance between Ca2+ coordinating atoms within the cEF binding site. The oxygens are colored according to each residue: Backbone (red), OE1 (Blue), OE2 (orange), OD1 (purple), OD2 (green), OH (yellow), and OG (cyan). Known Ca2+ interacting residues are underlined.

The cEF binding loop contains a high concentration of aspartic acid residues and they contribute significantly to formation via side chain interactions. Similar to in S100A6, this binding loop appears to have three distinct periods which are located at the beginning, middle, and end. However, most of these residues contribute primarily at the beginning and end of the path, with only D64 and E73 exhibiting progress in the middle of the path. Interestingly, the movement of D64 corresponds to the surrounding region rolling over the rest of the binding loop, similar to what was observed in S100A6. Similarly, we also saw that the binding loop condensed around E73 rather than E73 conforming to the binding loop. This is due to its positioning on HIV which is positionally restrained.

Within the pEF, we observed a greater degree of residue rearrangement than in S100A6. While some of these residues do not exhibit significant rearrangement, we saw a more gradual movement in the residues that did. Namely, the residues exhibiting the highest contribution to binding site formation were E22, D24, K25, and K31-E32, as shown in **Figure 4.3**. The lack of rapid formation progress differs significantly from what is observed in S100A6, in which E33 was observed to rapidly shift from its apo to holo conformations. We also saw differences at the junction between HI and the pEF binding loop. Here, those junction residues exhibited a behavior to gradually shift away from the center of the binding site. While S19 has been identified as residue capable of interacting with the loop, it was not observed to remain in proximity here. Rather, it was observed to interact with F15. This was due to the position of the pEF binding site, which repositioned farther from S19 to more tightly interact with the cEF

binding loop. The period in which the repositioning occurs corresponded to a gradual condensing of the pEF binding loop combined with the system undergoing a final conformation change.

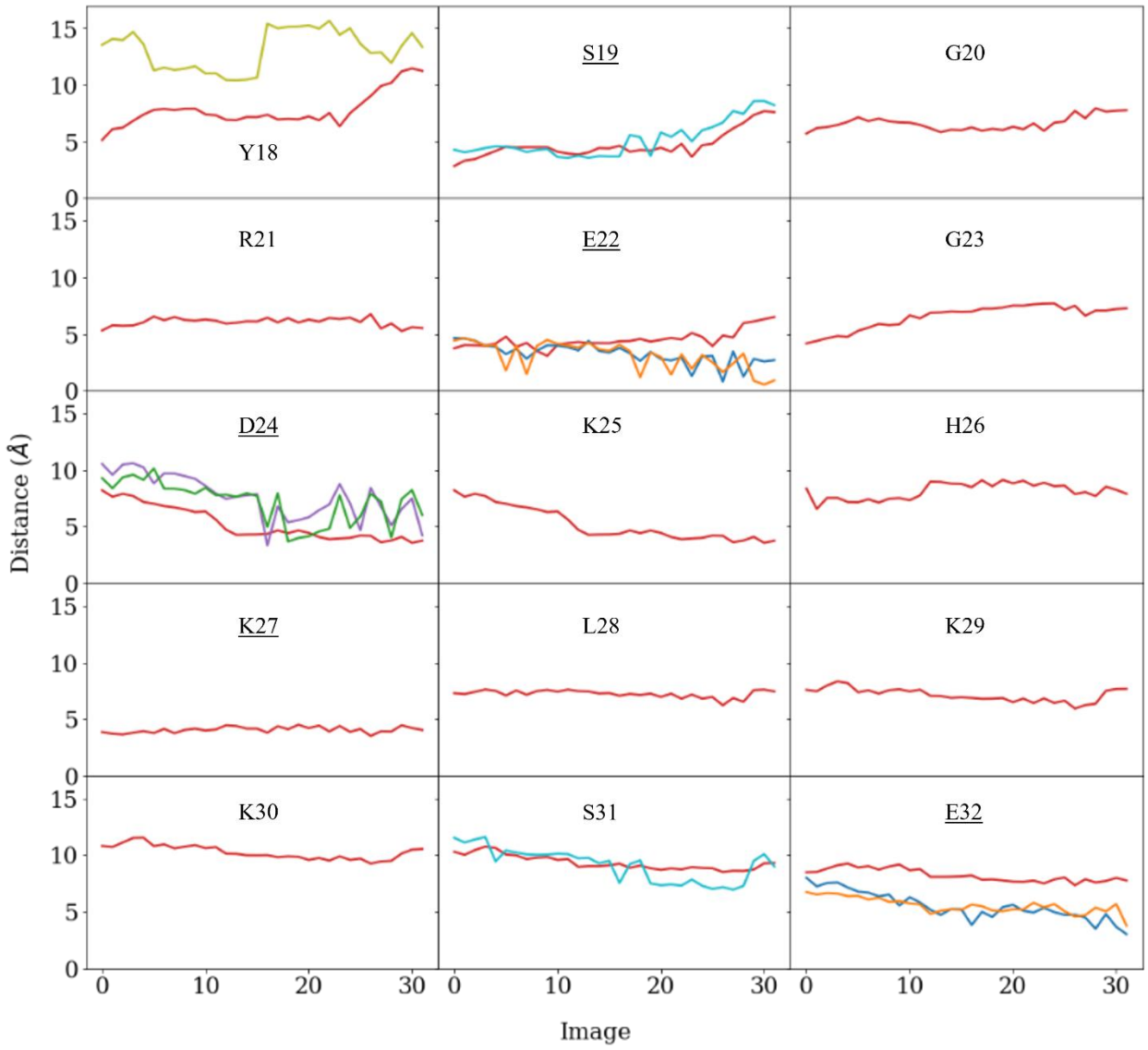

Figure 4.3: Plot of the distance between each oxygen atom and the average pairwise distance between Ca2+ coordinating atoms within the pEF binding site. The oxygens are colored according to each residue: Backbone (red), OE1 (Blue), OE2 (orange), OD1 (purple), OD2 (green), OH (yellow), and OG (cyan). Known Ca^{2+} interacting residues are underlined.

There are several potential reasons for this difference in binding site formation. Typically, this could be attributed to variation in residue composition between these two proteins. However, the residues of these binding loops share a high degree of similarity, as in most S100 proteins. While the composition of the binding loops remained similar, some difference was observed in both HI and HIV. Interactions between the pEF binding loop and the C-terminal residues of the opposing monomer were present in S100A6. However, this appeared to be involved in homodimerization where it would play a role in stabilizing the association between the two monomers.

Here, a difference in this interaction strongly influences the structure of the pEF binding loop. In S100A6, residue H28 of each monomer interacted with G90 of the opposing monomer. This appeared to stabilize the position of G90, which is in a region of high flexibility and conforms to the position of H28 and subsequently the pEF binding loop. In S100B, HI and HIV appeared to more tightly interact due to the composition differences in these helices with the presence of both F87 and F88, in S100B, inducing in this tight interaction. The result of this is that the C-terminal residues of the opposing monomer are more positionally restrained. H25 in S100B, which is comparable to H28 in S100A6, must conform to the positionally restrained C-terminal residues. Subsequently, the pEF binding loop is in a more open conformation here than in S100A6 and must undergo a longer path for binding site formation.

As previously discussed, a complex of residue interactions was located at either end of the pEF binding site. These residues, including R22 and E33, appeared to serve a role in stabilizing the binding loop in a more restricted position. The period of greatest formation corresponded to a change in this complex, in which E33 rearranged to face the binding site. However, this complex was observed to differ in S100B, with R21, the most comparable residue

to R22 in S100A6, instead interacting with E40 in one monomer and E32 in the other. These interactions seemed more direct, lacking the full residue complex observed in S100A6.

4.3.2 *Hydrophobic Patch Formation*

S100B was observed to undergo a shift of HIII and hydrophobic patch formation along a similar pathway in all NEB simulations, as shown in **Figure 4.1B1-4 and 4.1C**. The shift of HIII consisted of a turn of 80°, with the patch itself opening by 7 Å. These results are similar to that of S100A6, in which the turn was 86° and the patch opened by 8 Å. Whilst the linker of M1 was not observed to fully peel away from the patch region, we still observed evidence of hydrophobic patch formation.

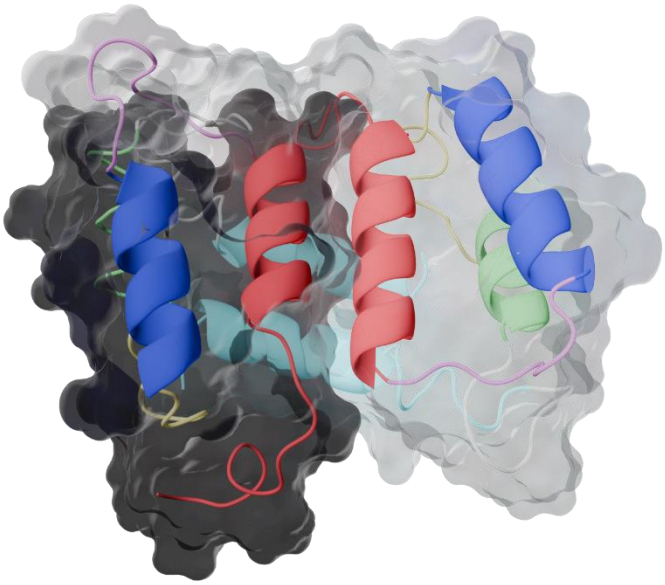

Figure 4.4: Cartoon and surface representations of S100B along the path at Image 0.

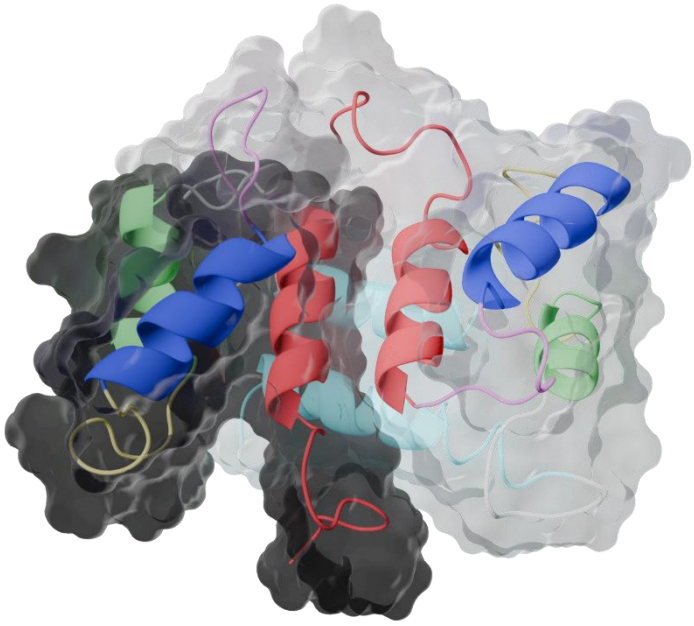

Figure 4.6: Cartoon and surface representations of S100B along the path at Image 20.

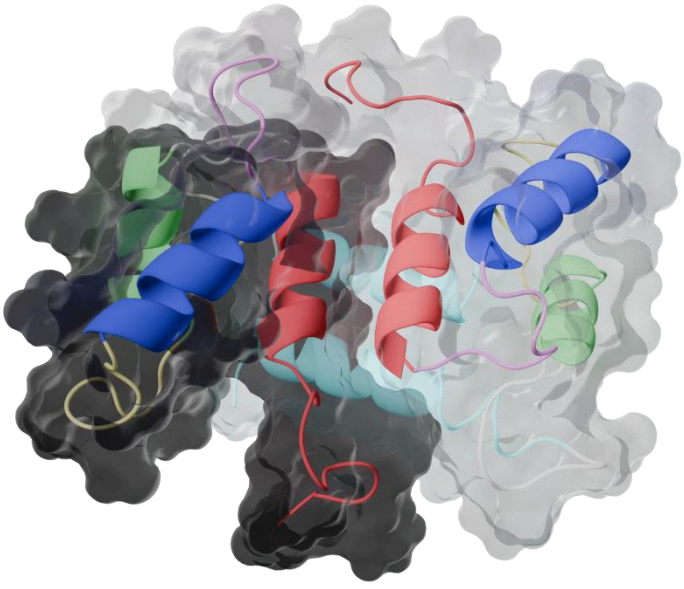

Figure 4.5: Cartoon and surface representations of S100B along the path at Image 10.

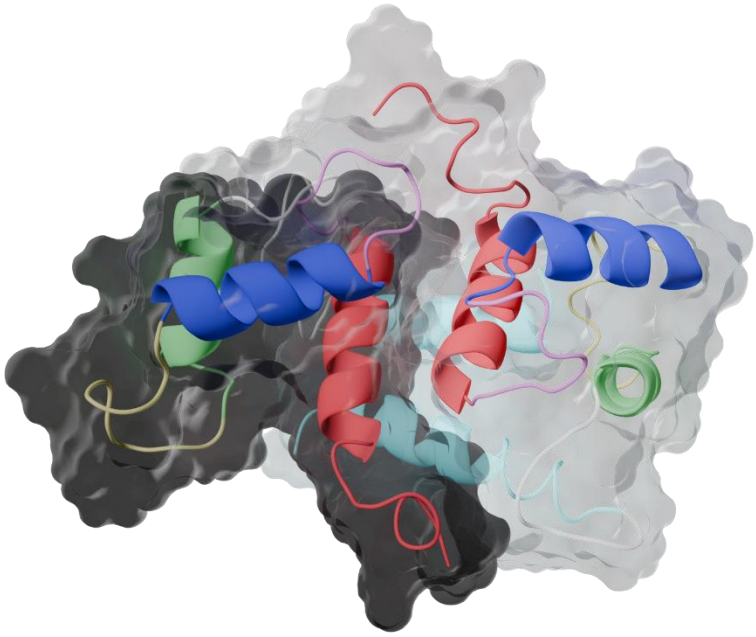

Figure 4.7: Cartoon and surface representations of S100B along the path at Image 31.

The difference in hydrophobic patch formation between each monomer can be most easily observed in **Figures 4.6 and 4.7**. While the patch in M1 was smaller, we believe that it would still be possible for a partner protein to partially interact and possibly induce full formation.

Like S100A6, we saw a long intermediate conformation in S100B that appeared to fill a similar role in both proteins. This conformation likely allows for internal mechanisms to slowly transition the protein between apo and holo conformations, while also allowing for partner proteins to assist in this process by interaction prior to full patch formation. It appeared as if M1 was restricted into a conformation containing elements characteristic of the intermediate

conformation. Specifically, this refers to the lack of the linker having transitioned to an open conformation.

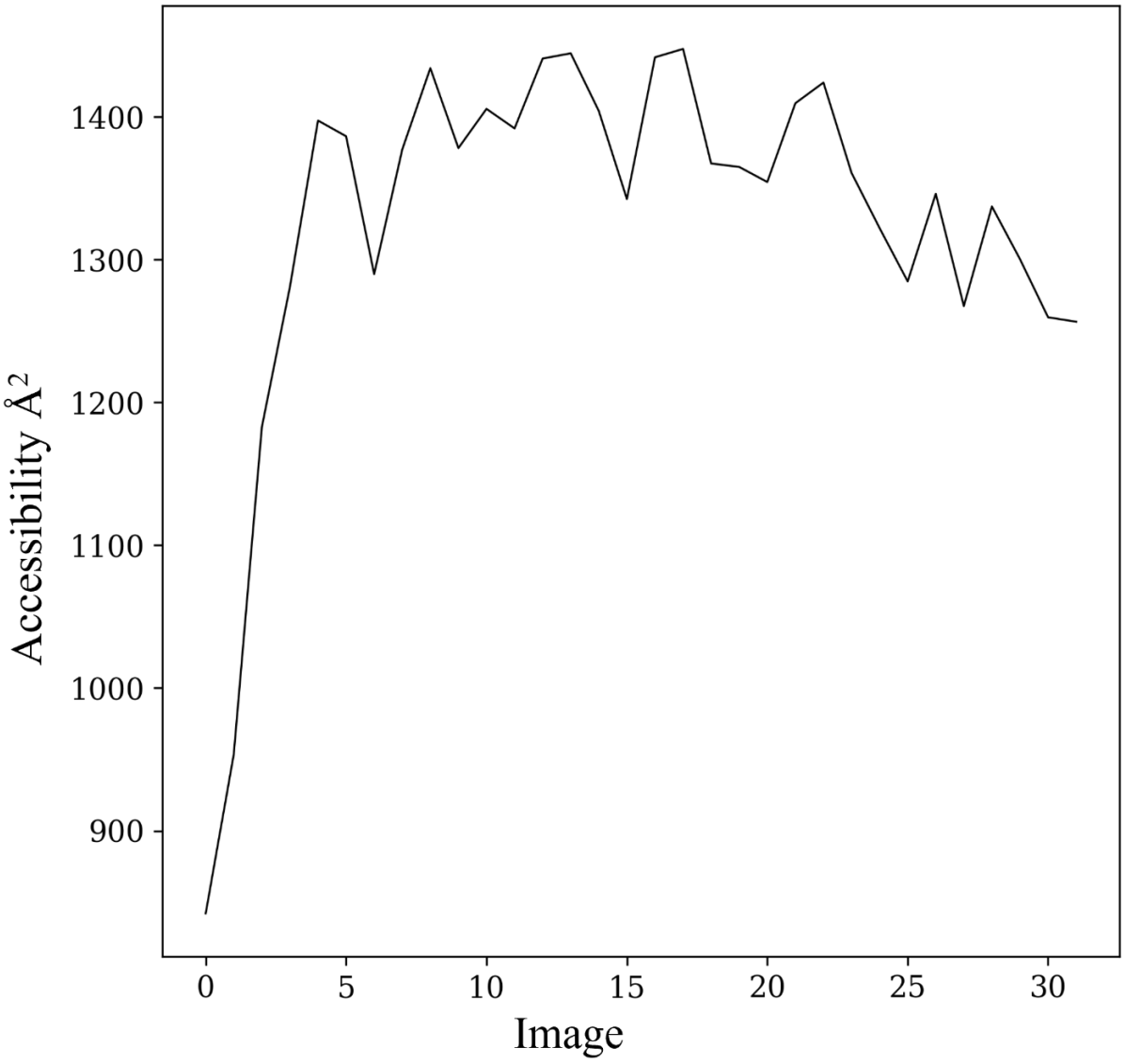

Figure 4.8: Plot of the solvent accessible surface area throughout the S100B pathway.

To characterize the availability of the hydrophobic patch, SASA calculations were completed in this work. For this, we identified residues known to interact with S100B in the hydrophobic patch region, including the linker region. These were identified by McClintock et

al., in which TRTK-12 was utilized in target recognition for S100B.[101] The residues included in this calculation are I37, L41, H43, L45, E46, I46, V51, K56, V57, T60, M80, V81, A83, C84, and F89. With these calculations, we observed the patch residues to rapidly become more accessible, prior to fluctuating around 1350 $Å^2$. We do see a gradual decrease in accessibility coinciding with the final turn of HIII, although the residues are still significantly more accessible in the holo conformation than in the apo conformation. However, this decrease does not initially appear as large as seen in S100A6 (see **Figure 4.8**). This is likely due to the number of residues participating in the calculation itself, given that many more residues were included here than in the S100A6 calculation. The apparent accessibility of the patch residues, combined with the presence of the intermediate conformation, indicates that S100B may bypass many of the internal mechanisms via partner interactions, as occurs in S100A6.

4.3.3 *Residue Contributions to Hydrophobic Patch Formation*

Here, we completed MMGBSA calculations to identify the differences in interaction energies between apo and holo conformations with the goal of examining similarities between S100A6 and S100B. The results from these calculations are reported in **Figure 4.9 and Table 4.1**.

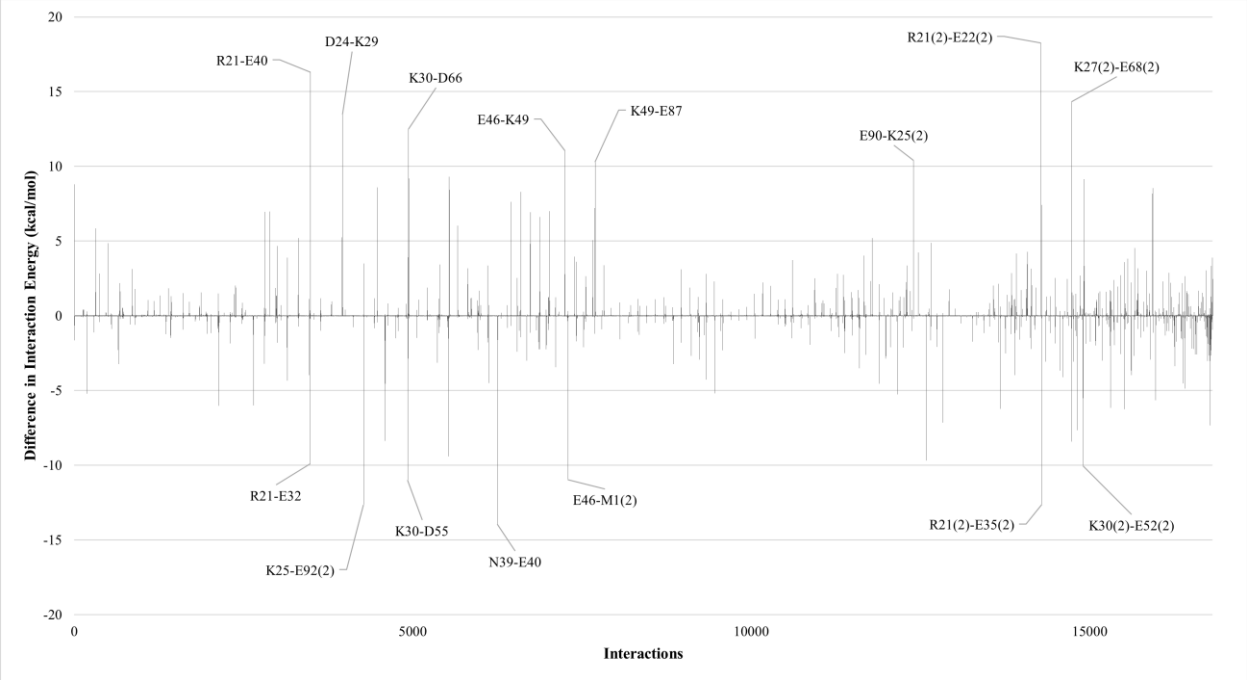

Figure 4.9: Diagram of the differences in pairwise interaction energies in S100B. Interactions with absolute energy values greater than 10 kcal/mol have been labeled.

Table 4.1: Pairwise interactions in S100B with absolute energy value greater than 6 kcal/mol. Bolded residues indicate discussed interactions.

Residue 1	**Residue 2**	**Difference (kcal/mol)**
ARG 21(2)	GLU 22(2)	**18.25796**
ARG 21	**GLU 40**	**16.30307**
LYS 27(2)	GLU 68(2)	**14.31636**
ASN 39	GLU 40	**-13.9408**
ASP 24	LYS 29	**13.465**
LYS 25	GLU 92(2)	**-12.6398**
ARG 21(2)	**GLU 35(2)**	**-12.6109**
LYS 30	ASP 66	**12.50727**
GLU 46	LYS 49	**11.07667**
LYS 30	**ASP 55**	**-10.9911**
GLU 46	MET 1(2)	**-10.9672**
GLU 90	LYS 25(2)	**10.41281**

LYS 49	GLU 87	**10.29832**
LYS 30(2)	**GLU 52(2)**	**-10.0299**
ARG 21	**GLU 35**	**-9.9036**
GLU 92	LYS 27(2)	**-9.67795**
LYS 34	GLU 52	**-9.41232**
LYS 34	GLU 59	**9.298822**
LYS 30	ASP 64	**9.248868**
LYS 30	GLU 73	**9.186758**
LYS 30(2)	**GLU 63(2)**	**9.157172**
LYS 34	ASP 55	**-8.80603**
MET 1	GLU 5	**8.789495**
LYS 27	GLU 68	**8.592283**
LYS 49(2)	GLU 92(2)	**8.547875**
LYS 34	ASP 62	**8.437663**
LYS 27(2)	ASP 66(2)	**-8.40355**
LYS 27	GLU 87(2)	**-8.37712**
LEU 41	GLU 5(2)	**8.290999**
LYS 27(2)	ASP 70(2)	**8.175378**
LYS 49(2)	GLU 87(2)	**8.172947**
LYS 39(2)	ASP 55(2)	**-8.04145**
LYS 29(2)	GLU 32(2)	**-7.6536**
GLU 40	SER 2(2)	**7.633263**
ARG 21(2)	**GLU 32(2)**	**7.418415**
ALA 84(2)	CYS 85(2)	**-7.32456**
LYS 49	HIE 86	**7.218516**
GLU 3(2)	LYS 6(2)	**-7.14307**
LYS 30(2)	**ASP 62(2)**	**7.132523**
ASP 24	LYS 27	**7.038209**
PHE 44	GLU 5(2)	**6.994476**
GLN 17	GLU 3(2)	**6.987113**
GLN 17	ARG 21	**6.967849**

SER 42	GLU 5(2)	**6.937212**
LYS 30(2)	**ASP 66(2)**	**6.779074**
ARG 21	GLU 32	**-6.76546**
HIE 43	GLU 5(2)	**6.62189**
LYS 25	HIE 92(2)	**-6.33565**
LEU 42(2)	LYS 49(2)	**-6.24218**
ASP 13(2)	GLN 17(2)	**-6.23473**
ILE 37(2)	LYS 49(2)	**-6.1458**
GLU 35	ASN 39	**6.041705**
ASP 13	GLN 17	**-6.02885**

Here, we see a number of interactions involving similar residues as to those observed in S100A6 (see **Figures 3.11 and 4.9, Tables 3.1 and 4.1**). These residues include R21, K30, E32, and residues located along HIII itself. Although previously discussed, both R21 and E32 were once again identified due to their contribution to residue interactions near the pEF. We observed significant differences between R21 in either monomer, with stronger interaction occurring between R21 and E40 in M1 than R21 and E32 in M2 (-16.3 and -7.1 kcal/mol, respectively). Additionally, we observed that R21-E32 interactions occurred within the holo conformation of M1 (-6.7 kcal/mol) whereas E32 shifted towards the pEF binding site in M2. These results, combined with the M1 exhibiting other characteristics of the intermediate conformation, as shown in **Figures 4.4-4.7**, indicate that M1 is simply further behind on the conformation pathway. This further indicates, along with results from S100A6, that these residues are involved in formation of the pEF binding site. This is to stabilize the loop itself for Ca^{2+} binding while also being involved in the smaller pEF conformation changes.

Given the indicated significance of K31 in S100A6, we believe that the comparable residue, in S100B, may have similar utility. To study this further, MMGBSA calculations were extended to include the entirety of the pathway regarding these residues. The results of these calculations are displayed in **Figures 4.10**. Additionally, given the lack of conformation change in M1, these calculations were performed using M2.

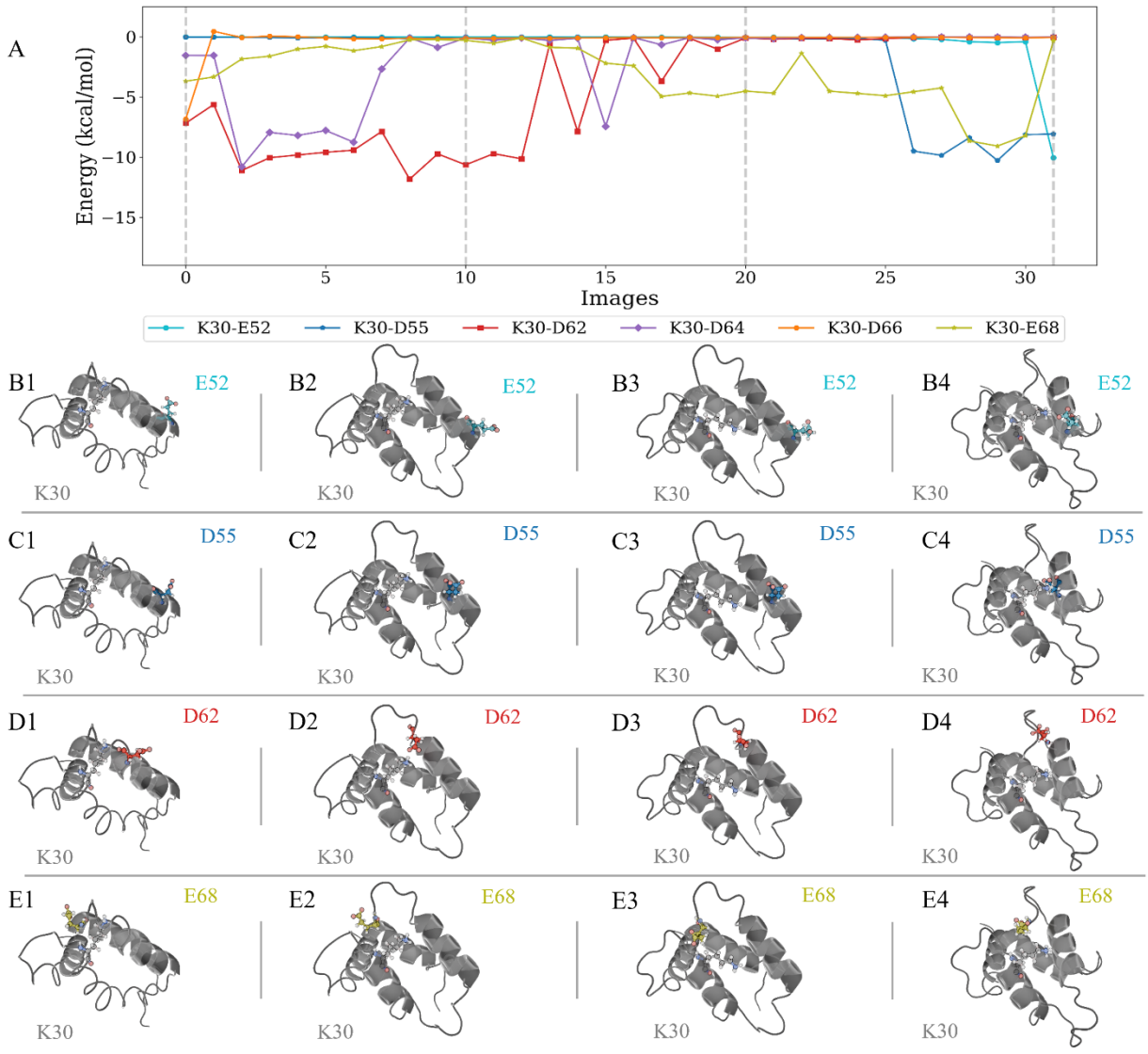

Figure 4.10: (A) Plot of the 6 strongest pairwise interaction energies between K31 and residues along its conformation path in S100B. (B-E) Diagrams of the residue interactions of K30 throughout Images 0, 10, 20, and 31. These residues include E52 (B), D55 (C), D62 (D), and E68 (E). Only a single monomer was included in the diagrams to ensure clarity.

In S100B, K30 was observed to exhibit behavior similar to that of K31 in S100B. We see that the side chain of K30 is positioned within the cEF binding site while in the apo conformation where it interacts with several binding site residues. As binding site formation occurs, we observed an ejection of this side chain. In Figure 4.10A, we see K30 interact with several residues along the pathway. The most notable of these are E52, D55, D62, and E68. Several other residues were observed to interact but were generally observed to interact less strongly than the aforementioned residues. These include both D64 and D66. Within the apo conformation, a complex of interactions served to form around K30, suggesting its role in stabilizing the loop. As the system progressed, interactions with D62 were observed to become stronger while many of the other interactions weakened. This K30-D62 had a long lifespan and was only observed to break apart once the system had progressed well into the intermediate conformation, as shown in **Figures 4.10A and 4.10D1-4**.

In M2, which exhibited the largest degree of hydrophobic patch formation, a K30-E68 interaction took place in a manner similar to the K31-Q66 interaction in S100A6 (see **Figures 3.12E1-4 and 4.10E1-4**). Specifically, E68 appeared to rotate outside of the binding loop where it was observed to interact with the backbone residues of K30. This interaction appeared to push the side chain of K30 towards HIII. The resulting position of K30 allowed for interactions with both E52 and D55, occurring in the latter stage of the pathway and existing in the holo conformation (see **Figures 4.10B1-4 and 4.10C1-4**). This interaction served to stabilize the position of HIII itself, given the positions of these residues toward the N-terminal end of HIII and close to the linker region.

Given that hydrophobic patch formation differed between monomers, initial review of K30 and E68 behavior in M1 presented additional differences. K30 did not appear to fully

interact with E52 nor D55. Rather, K30 appeared fixated on D62. Additionally, unlike in M2, we observed no guiding mechanism, assisting in the reorientation of K30. While E68 completed this in M2, we observed E68 remain fixed in an upright position in M1 with no evidence of a rotation occurring within the cEF. Similarly, we observed K30 of M1 to never reach and form interactions with E52. Rather, the strongest holo conformation interaction of K30 in M1 was with D55 (-11.0 kcal/mol), further suggesting that M1 was still within the latter stages of hydrophobic patch formation and the pathway as a whole.

4.4 Conclusions

In this work, aMD and NEB methods were applied to S100B systems with the goal of identifying mechanisms of similar function to those previously identified in S100A6. We observed the formation of the hydrophobic patch region along similar pathways in all NEB simulations. An intermediate conformation of S100B was identified along the path. This conformation had an extended lifespan, accounting for most of the path. Differences between monomers were observed to be present in the holo conformation, with M1 exhibiting different linker region behavior. However, the full hydrophobic patch formation in M2 allowed us to draw comparison between mechanisms in S100A6 and S100B. Similarly, we were able to compare mechanical differences between the functional and potentially nonfunctional monomers.

Here, we observed a number of comparable residues in S100A6 and S100B that we believe are functionally similar. The most important of these residues is K30/K31. While the behavior of these residues slightly differed between proteins, we observed a remarkable similarity in function and pathway between these residues. Similarly, we believe that residues comparable to D50 or E52 are vital to proper hydrophobic patch stabilization. The interaction of residues located proximal to the pEF binding loop and those at the junction between HIII and the

linker are likely vital in the turn of HIII. Residues displaying residues similar to those described in this study have been identified in other S100 systems. These include but are not limited to S100A2, S100A4, S100A5, S100A10, S100P, and S100Z. [102–107]

References

1. Yáñez, M., Gil-Longo, J. & Campos-Toimil, M. Calcium Binding Proteins. in 461–482 (2012). doi:10.1007/978-94-007-2888-2_19.

2. Katz, A. K., Glusker, J. P., Beebe, S. A. & Bock, C. W. Calcium Ion Coordination: A Comparison with That of Beryllium, Magnesium, and Zinc. *J Am Chem Soc* **118**, 5752–5763 (1996).

3. Maurer, P., Hohenester, E. & Engel, J. Extracellular calcium-binding proteins. *Curr Opin Cell Biol* **8**, 609–617 (1996).

4. Heizmann, C. W. & Hunzlker, W. Intracellular calcium-binding proteins: more sites than insights. *Trends Biochem Sci* **16**, 98–103 (1991).

5. Grabarek, Z. Structural Basis for Diversity of the EF-hand Calcium-binding Proteins. *J Mol Biol* **359**, 509–525 (2006).

6. Lewit-Bentley, A. & Réty, S. EF-hand calcium-binding proteins. *Curr Opin Struct Biol* **10**, 637–643 (2000).

7. Schäfer, B. W. & Heizmann, C. W. The S100 family of EF-hand calcium-binding proteins: functions and pathology. *Trends Biochem Sci* **21**, 134–140 (1996).

8. Kawasaki, H. & Kretsinger, R. H. Structural and functional diversity of EF-hand proteins: Evolutionary perspectives. *Protein Science* **26**, 1898–1920 (2017).

9. Burgoyne, R. D. & Geisow, M. J. The annexin family of calcium-binding proteins. *Cell Calcium* **10**, 1–10 (1989).

10. Kligman, D. & Hilt, D. C. The S100 protein family. *Trends Biochem Sci* **13**, 437–443 (1988).

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

12. Zimmer, D. B., Cornwall, E. H., Landar, A. & Song, W. The S100 protein family: History, function, and expression. *Brain Res Bull* **37**, 417–429 (1995).

13. Sudo, T. & Hidaka, H. Regulation of Calcyclin (S100A6) Binding by Alternative Splicing in the N-terminal Regulatory Domain of Annexin XI Isoforms. *Journal of Biological Chemistry* **273**, 6351–6357 (1998).

14. Sedaghat, F. & Notopoulos, A. S100 protein family and its application in clinical practice. *Hippokratia* **12**, 198–204 (2008).

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

16. Donato, R. *et al.* S100B's double life: Intracellular regulator and extracellular signal. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1793**, 1008–1022 (2009).

17. Boye, K. & Mælandsmo, G. M. S100A4 and Metastasis. *Am J Pathol* **176**, 528–535 (2010).

18. Leśniak, W., Słomnicki, Ł. P. & Filipek, A. S100A6 – New facts and features. *Biochem Biophys Res Commun* **390**, 1087–1092 (2009).

19. Rosenberger, S., Thorey, I. S., Werner, S. & Boukamp, P. A Novel Regulator of Telomerase. *Journal of Biological Chemistry* **282**, 6126–6135 (2007).

20. Lagasse, E. & Weissman, I. L. Mouse MRP8 and MRP14, two intracellular calcium-binding proteins associated with the development of the myeloid lineage. *Blood* **79**, 1907–15 (1992).

21. SVENNINGSSON, P. & GREENGARD, P. p11 (S100A10) — an inducible adaptor protein that modulates neuronal functions. *Curr Opin Pharmacol* **7**, 27–32 (2007).

22. Delangre, E. *et al.* S100 Proteins in Fatty Liver Disease and Hepatocellular Carcinoma. *Int J Mol Sci* **23**, 11030 (2022).

23. Oris, C., Kahouadji, S., Durif, J., Bouvier, D. & Sapin, V. S100B, Actor and Biomarker of Mild Traumatic Brain Injury. *Int J Mol Sci* **24**, 6602 (2023).

24. Otterbein, L. R., Kordowska, J., Witte-Hoffmann, C., Wang, C.-L. A. & Dominguez, R. Crystal Structures of S100A6 in the Ca2+-Free and Ca2+-Bound States. *Structure* **10**, 557–567 (2002).

25. Linse, S., Helmersson, A. & Forsén, S. Calcium binding to calmodulin and its globular domains. *J Biol Chem* **266**, 8050–4 (1991).

26. Grabarek, Z., Grabarek, J., Leavis, P. C. & Gergely, J. Cooperative binding to the Ca2+-specific sites of troponin C in regulated actin and actomyosin. *J Biol Chem* **258**, 14098–102 (1983).

27. Donato, R., Sorci, G. & Giambanco, I. S100A6 protein: functional roles. *Cellular and Molecular Life Sciences* **74**, 2749–2760 (2017).

28. Nelson, M. R., Thulin, E., Fagan, P. A., Forsén, S. & Chazin, W. J. The EF-hand domain: A globally cooperative structural unit. *Protein Science* **11**, 198–205 (2002).

29. Yap, K. L., Ames, J. B., Swindells, M. B. & Ikura, M. Diversity of conformational states and changes within the EF-hand protein superfamily. *Proteins* **37**, 499–507 (1999).

30. Sunahori, K. *et al.* The S100A8/A9 heterodimer amplifies proinflammatory cytokine production by macrophages via activation of nuclear factor kappa B and p38 mitogen-activated protein kinase in rheumatoid arthritis. *Arthritis Res Ther* **8**, R69 (2006).

31. Vogl, T. *et al.* MRP8 and MRP14 control microtubule reorganization during transendothelial migration of phagocytes. *Blood* **104**, 4260–4268 (2004).

32. Siegenthaler, G. *et al.* A Heterocomplex Formed by the Calcium-binding Proteins MRP8 (S100A8) and MRP14 (S100A9) Binds Unsaturated Fatty Acids with High Affinity. *Journal of Biological Chemistry* **272**, 9371–9377 (1997).

33. Lindsay, S., Bartolotti, L. & Li, Y. Interactions and conformational changes of annexin A2/p11 heterotetramer models on a membrane: a molecular dynamics study. *J Biomol Struct Dyn* 1–10 (2023) doi:10.1080/07391102.2023.2256877.

34. Rezvanpour, A., Santamaria-Kisiel, L. & Shaw, G. S. The S100A10-Annexin A2 Complex Provides a Novel Asymmetric Platform for Membrane Repair. *Journal of Biological Chemistry* **286**, 40174–40183 (2011).

35. Ecsédi, P. *et al.* Regulation of the Equilibrium between Closed and Open Conformations of Annexin A2 by N-Terminal Phosphorylation and S100A4-Binding. *Structure* **25**, 1195-1207.e5 (2017).

36. Michetti, F. *et al.* The S100B Protein: A Multifaceted Pathogenic Factor More Than a Biomarker. *Int J Mol Sci* **24**, 9605 (2023).

37. Leśniak, W. & Filipek, A. S100A6 Protein—Expression and Function in Norm and Pathology. *Int J Mol Sci* **24**, 1341 (2023).

38. Passey, R. J. *et al.* A null mutation in the inflammation-associated S100 protein S100A8 causes early resorption of the mouse embryo. *J Immunol* **163**, 2209–16 (1999).

39. Rohde, D. *et al.* S100A1: A Multifaceted Therapeutic Target in Cardiovascular Disease. *J Cardiovasc Transl Res* **3**, 525–537 (2010).

40. Schmidt-Hansen, B. *et al.* Extracellular S100A4(mts1) stimulates invasive growth of mouse endothelial cells and modulates MMP-13 matrix metalloproteinase activity. *Oncogene* **23**, 5487–5495 (2004).

41. Stein, U. *et al.* The Metastasis-Associated Gene S100A4 Is a Novel Target of β-catenin/T-cell Factor Signaling in Colon Cancer. *Gastroenterology* **131**, 1486–1500 (2006).

42. Stein, U. *et al.* Intervening in β-Catenin Signaling by Sulindac Inhibits S100A4-Dependent Colon Cancer Metastasis. *Neoplasia* **13**, 131-IN8 (2011).

43. Wang, Y., Kang, X., Kang, X. & Yang, F. S100A6: molecular function and biomarker role. *Biomark Res* **11**, 78 (2023).

44. Li, C., Zhang, F., Lin, M. & Liu, J. Induction of S100A9 Gene Expression by Cytokine Oncostatin M in Breast Cancer Cells Through the STAT3 Signaling Cascade. *Breast Cancer Res Treat* **87**, 123–134 (2004).

45. Emberley, E. D. *et al.* The S100A7-c-Jun Activation Domain Binding Protein 1 Pathway Enhances Prosurvival Pathways in Breast Cancer. *Cancer Res* **65**, 5696–5702 (2005).

46. Arai, K. *et al.* S100A8 and S100A9 Overexpression Is Associated with Poor Pathological Parameters in Invasive Ductal Carcinoma of the Breast. *Curr Cancer Drug Targets* **8**, 243–252 (2008).

47. Fernandez-Fernandez, M. R., Rutherford, T. J. & Fersht, A. R. Members of the S100 family bind p53 in two distinct ways. *Protein Science* **17**, 1663–1670 (2008).

48. Chen, H. *et al.* Functional Role of *S100A14* Genetic Variants and Their Association with Esophageal Squamous Cell Carcinoma. *Cancer Res* **69**, 3451–3457 (2009).

49. Li, C. *et al.* A novel p53 target gene, *S100A9* , induces p53-dependent cellular apoptosis and mediates the p53 apoptosis pathway. *Biochemical Journal* **422**, 363–372 (2009).

50. van Dieck, J. *et al.* Posttranslational Modifications Affect the Interaction of S100 Proteins with Tumor Suppressor p53. *J Mol Biol* **394**, 922–930 (2009).

51. Bloomfield, S. M., McKinney, J., Smith, L. & Brisman, J. Reliability of S100B in predicting severity of central nervous system injury. *Neurocrit Care* **6**, 121–138 (2007).

52. Hammer, H. B. *et al.* Calprotectin (a major S100 leucocyte protein) predicts 10-year radiographic progression in patients with rheumatoid arthritis. *Ann Rheum Dis* **69**, 150–154 (2010).

53. Baillet, A. *et al.* Synovial fluid proteomic fingerprint: S100A8, S100A9 and S100A12 proteins discriminate rheumatoid arthritis from other inflammatory joint diseases. *Rheumatology* **49**, 671–682 (2010).

54. Foell, D., Wittkowski, H. & Roth, J. Mechanisms of Disease: a ‘DAMP’ view of inflammatory arthritis. *Nat Clin Pract Rheumatol* **3**, 382–390 (2007).

55. Richards, T. J. *et al.* Peripheral Blood Proteins Predict Mortality in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* **185**, 67–76 (2012).

56. Yang, Z. *et al.* S100A12 provokes mast cell activation: A potential amplification pathway in asthma and innate immunity. *Journal of Allergy and Clinical Immunology* **119**, 106–114 (2007).

57. Faruqui, T., Singh, G., Khan, S., Khan, M. S. & Akhter, Y. Differential gene expression analysis of RAGE-S100A6 complex for target selection and the design of novel inhibitors for anticancer drug discovery. *J Cell Biochem* **124**, 205–220 (2023).

58. Leclerc, E., Fritz, G., Vetter, S. W. & Heizmann, C. W. Binding of S100 proteins to RAGE: An update. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1793**, 993–1007 (2009).

59. Yatime, L. *et al.* The Structure of the RAGE:S100A6 Complex Reveals a Unique Mode of Homodimerization for S100 Proteins. *Structure* **24**, 2043–2052 (2016).

60. Donato, R. RAGE: A Single Receptor for Several Ligands and Different Cellular Responses: The Case of Certain S100 Proteins. *Curr Mol Med* **7**, 711–724 (2007).

61. Penumutchu, S. R., Chou, R.-H. & Yu, C. Structural Insights into Calcium-Bound S100P and the V Domain of the RAGE Complex. *PLoS One* **9**, e103947 (2014).

62. Moysa, A. *et al.* A model of full-length RAGE in complex with S100B. *Structure* **29**, 989-1002.e6 (2021).

63. Jurewicz, E., Robaszkiewicz, K., Moraczewska, J. & Filipek, A. Binding of S100A6 to actin and the actin–tropomyosin complex. *Sci Rep* **10**, 12824 (2020).

64. Słomnicki, Ł. P., Nawrot, B. & Leśniak, W. S100A6 binds p53 and affects its activity. *Int J Biochem Cell Biol* **41**, 784–790 (2009).

65. Tsoporis, J. N., Izhar, S. & Parker, T. G. Expression of S100A6 in Cardiac Myocytes Limits Apoptosis Induced by Tumor Necrosis Factor-α. *Journal of Biological Chemistry* **283**, 30174–30183 (2008).

66. Yang, F. *et al.* The Role of S100A6 in Human Diseases: Molecular Mechanisms and Therapeutic Potential. *Biomolecules* **13**, 1139 (2023).

67. Nedjadi, T. *et al.* S100A6 binds to annexin 2 in pancreatic cancer cells and promotes pancreatic cancer cell motility. *Br J Cancer* **101**, 1145–1154 (2009).

68. Li, Y. *et al.* The Calcium-Binding Protein S100A6 Accelerates Human Osteosarcoma Growth by Promoting Cell Proliferation and Inhibiting Osteogenic Differentiation. *Cellular Physiology and Biochemistry* **37**, 2375–2392 (2015).

69. Grahn, T. H. M. *et al.* S100A6 is a critical regulator of hematopoietic stem cells. *Leukemia* **34**, 3323–3337 (2020).

70. Michetti, F. *et al.* The S100B story: from biomarker to active factor in neural injury. *J Neurochem* **148**, 168–187 (2019).

71. Abboud, T., Rohde, V. & Mielke, D. Mini review: Current status and perspective of S100B protein as a biomarker in daily clinical practice for diagnosis and prognosticating of clinical outcome in patients with neurological diseases with focus on acute brain injury. *BMC Neurosci* **24**, 38 (2023).

72. Drohat, A. C., Tjandra, N., Baldisseri, D. M. & Weber, D. J. The use of dipolar couplings for determining the solution structure of rat apo-S100B(betabeta). *Protein Sci* **8**, 800–9 (1999).

73. Goyal, A. *et al.* S100b as a Prognostic Biomarker in Outcome Prediction for Patients with Severe Traumatic Brain Injury. *J Neurotrauma* **30**, 946–957 (2013).

74. Mocellin, S., Zavagno, G. & Nitti, D. The prognostic value of serum S100B in patients with cutaneous melanoma: A meta-analysis. *Int J Cancer* **123**, 2370–2376 (2008).

75. McIlroy, M. *et al.* Interaction of Developmental Transcription Factor HOXC11 with Steroid Receptor Coactivator SRC-1 Mediates Resistance to Endocrine Therapy in Breast Cancer. *Cancer Res* **70**, 1585–1594 (2010).

76. Hollingsworth, S. A. & Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **99**, 1129–1143 (2018).

77. Tuckerman, M. E. & Martyna, G. J. Understanding Modern Molecular Dynamics: Techniques and Applications. *J Phys Chem B* **104**, 159–178 (2000).

78. Case, D. A. *et al.* Amber22.

79. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* **11**, 1864–1874 (2015).

80. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* **23**, 327–341 (1977).

81. Scott, W. R. P. *et al.* The GROMOS Biomolecular Simulation Program Package. *J Phys Chem A* **103**, 3596–3607 (1999).

82. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **31**, 671–690 (2010).

83. Tian, C. *et al.* ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* **16**, 528–552 (2020).

84. Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J Chem Phys* **120**, 11919–11929 (2004).

85. The PyMOL Molecular Graphics System.

86. Ghoreishi, D., Cerutti, D. S., Fallon, Z., Simmerling, C. & Roitberg, A. E. Fast Implementation of the Nudged Elastic Band Method in AMBER. *J Chem Theory Comput* **15**, 4699–4707 (2019).

87. Herbol, H. C., Stevenson, J. & Clancy, P. Computational Implementation of Nudged Elastic Band, Rigid Rotation, and Corresponding Force Optimization. *J Chem Theory Comput* **13**, 3250–3259 (2017).

88. Lewis, K., Lindsay, S. & Li, Y. Nudged Elastic Band Study on the N-Terminal Domain Conformational Pathways of Annexin A1 from a Buried State to an Exposed State. *J Phys Chem B* **123**, 10163–10170 (2019).

89. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **9**, 3084–3095 (2013).

90. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**, 2319–2327 (2011).

91. Gowers, R. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. in 98–105 (2016). doi:10.25080/Majora-629e541a-00e.

92. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).

93. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng* **9**, 90–95 (2007).

94. Johnston, B. *et al.* BradyAJohnston/MolecularNodes: v4.1.4 for Blender 4.1.

95. Okazaki, K., Koga, N., Takada, S., Onuchic, J. N. & Wolynes, P. G. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **103**, 11844–11849 (2006).

96. Yu, Z., Jacobson, M. P., Josovitz, J., Rapp, C. S. & Friesner, R. A. First-Shell Solvation of Ion Pairs: Correction of Systematic Errors in Implicit Solvent Models. *J Phys Chem B* **108**, 6643–6654 (2004).

97. 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).

98. Drohat, A. C., Baldisseri, D. M., Rustandi, R. R. & Weber, D. J. Solution Structure of Calcium-Bound Rat S100B(ββ) As Determined by Nuclear Magnetic Resonance Spectroscopy ·. *Biochemistry* **37**, 2729–2740 (1998).

99. Xiong, Y., Shabane, P. S. & Onufriev, A. V. Melting Points of OPC and OPC3 Water Models. *ACS Omega* **5**, 25087–25094 (2020).

100. Sengupta, A., Li, Z., Song, L. F., Li, P. & Merz, K. M. Parameterization of Monovalent Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *J Chem Inf Model* **61**, 869–880 (2021).

101. McClintock, K. A. & Shaw, G. S. A Novel S100 Target Conformation Is Revealed by the Solution Structure of the Ca2+-S100B-TRTK-12 Complex. *Journal of Biological Chemistry* **278**, 6251–6257 (2003).

102. Koch, M., Diez, J. & Fritz, G. Crystal Structure of Ca2+-Free S100A2 at 1.6-Å Resolution. *J Mol Biol* **378**, 933–942 (2008).

103. Malashkevich, V. N. *et al.* Structure of Ca^{2+} -Bound S100A4 and Its Interaction with Peptides Derived from Nonmuscle Myosin-IIA. *Biochemistry* **47**, 5111–5126 (2008).

104. Bertini, I. *et al.* Solution structure and dynamics of S100A5 in the apo and Ca2+-bound states. *JBIC Journal of Biological Inorganic Chemistry* **14**, 1097–1107 (2009).

105. Ozorowski, G., Milton, S. & Luecke, H. Structure of a C-terminal AHNAK peptide in a 1:2:2 complex with S100A10 and an acetylated N-terminal peptide of annexin A2. *Acta Crystallogr D Biol Crystallogr* **69**, 92–104 (2013).

106. Lee, Y.-C. *et al.* Letter to the Editor: NMR Structure of the Apo-S100P Protein. *J Biomol NMR* **29**, 399–402 (2004).

107. Calderone, V., Fragai, M., Gallo, G. & Luchinat, C. Solving the crystal structure of human calcium-free S100Z: the siege and conquer of one of the last S100 family strongholds. *JBIC Journal of Biological Inorganic Chemistry* **22**, 519–526 (2017).